Center for Demography and Ecology

University of Wisconsin-Madison

Detectable Errors in NSFG 1995 Contraceptive and Sexual Nonintercourse Calendars

Steven P. Martin

Lawrence L. Wu

CDE Working Paper No. 99-10

DETECTABLE ERRORS IN NSFG 1995

CONTRACEPTIVE CALENDARS AND

SEXUAL NONINTERCOURSE CALENDARS

Steven P. Martin
University of Wisconsin-Madison

Lawrence L. Wu
University of Wisconsin-Madison

October 1998

Direct correspondence to Steven P. Martin, Department of Sociology, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706, smartin@ssc.wisc.edu. The authors gratefully acknowledge research funding provided by an NICHD core grant to the Center for Demography and Ecology (HD 05876) and by an NICHD grant for research on single mothers (HD 29550).

Abstract

In this study we estimate error rates in two event calendars in the 1995 NSFG – a five year contraceptive use calendar and a five-year sexual nonintercourse calendar. We superimpose each event calendar over respondents' birth histories and count implausible combinations of fertility events, sexual activity, and contraceptive use. We then evaluate how similar errors might affect summary measures based on the event calendars. Finally, we identify characteristics of respondents, interviewers, and interviews that correlate with high detectable error rates.

Our results suggest that researchers should use the NSFG event histories cautiously. We find a small but significant rate of detectable errors in both of the event calendars we study. Unfortunately, events in the event calendars are distributed so that even a few response errors can severely impair some types of studies (such as studies of contraceptive failure). In addition, response errors correlate with some important background characteristics of the respondents; this may hamper some comparative studies of contraceptive use or sexual nonintercourse. Lastly, we identify characteristics of interviews and interviewers that are associated with detectable errors. Such associations may suggest ways to reduce response errors in the future.

Introduction

The 1995 cycle of the NSFG has event histories for many new variables. With these event histories, researchers can ask more questions about fertility patterns than they could with earlier cycles of the NSFG. Unfortunately, the new event histories contributed to a half-hour increase in the average length of each interview (Kelley et al., 1997), and the added burden on respondents might lead to more response errors. In addition, researchers make strenuous demands on data when they base their analyses on superimposed event histories; incorrect reports of the timing of events could weaken or distort findings. For these reasons we set out to assess the accuracy of the NSFG event calendars in the context of specific research questions.

Background

Measurement errors affect research in many disciplines, and different disciplines have developed different ways to think and talk about measurement error (Groves, 1991). The research literature on measurement error is vast, but we can assign most studies to two broad categories. One body of literature on measurement error focuses on the cognitive and other processes that generate correct and incorrect responses (for a review, see Mathiowetz, 1998). Survey designers rely heavily on this literature to reduce response error in their studies.

The other body of literature is mathematical and assesses statistical adjustments for misclassified data. In almost all cases, these statistical adjustments require estimates of the misclassification probabilities acquired from reinterview surveys. When such reinterview data are available, statistical adjustments can profoundly alter the results of a study (e.g. Chua and Fuller, 1987). Unfortunately, such data are seldom available. Thus, a gap has emerged between reinterview - based research that meticulously corrects for response error, and most other studies which try to eliminate response error but do not account for its effects.

Axinn et al. (1997) and many others have promoted event calendars as an effective technique for reducing response error. Event calendars employ memorable events as landmarks to help respondents recall other events. For example, a fertility survey may a collect a birth history, then use it to guide the respondent's recollection of a contraceptive history. The respondent assumedly recalls her contraceptive behavior more accurately in the context of more memorable fertility events. Cognitive studies indicate that landmarks indeed aid recall (Loftus and Marburger, 1983), and event calendars appear to contain fewer response errors than traditional survey questionnaires (Becker and Sosa, 1992; Goldman et al., 1989).

Researchers can study response errors in event calendars by cross-validating two or more event calendars in the same survey. For example, one can superimpose a contraceptive history on a pregnancy history to isolate the few months that coincide with a late pregnancy. Then one can compute an error rate by counting reports of contraceptive use in late pregnancy. Becker and Sosa (1992) call such instances "erroneous superpositions of events," but we prefer the term "detectable errors" because many errors cannot be detected by superposition.

Cross-validation within a survey provides only partial information about response error rates. Unlike a reinterview analysis, an analysis of detectable errors cannot provide values for the whole matrix of possible misclassifications in a two-way table. For example, superimposing contraceptive use on pregnancies tells us only about respondents who report contraceptive use but are not truly using contraceptives; we know nothing about misclassification among respondents who *do not* report contraceptive use. Because of this limitation, researchers have not used detectable superposition errors to statistically control for response error. Instead, researchers use detectable errors as a simple measure of survey quality. Simply put, a measure with many detectable errors is bad, a measure with few detectable errors is good.

Our goal in this paper is to use more of the information in detectable errors. Without knowledge of the full matrix of response errors, and without information about response errors in most months of a survey, we cannot estimate or correct errors as we might if we had reinterview data. However, we can often identify minimum bounds of the effect of response errors. We can also use regression analyses and descriptive data to investigate possible causes of detectable response errors and find ways to reduce them.

The implications of response errors

Before we begin to estimate response error rates, we will briefly discuss the effects of errors in some different kinds of studies. The impact of response errors depends not only on the number of response errors, but also on the nature of the summary measures being estimated, the underlying distribution of true responses, and the distribution of response errors among responses. To demonstrate how error rates and response distributions can interact, we shall present equations for the effect of response errors on three different summary measures – a prevalence measure, a rate, and a simple measure of association. For now, we ignore sampling error and error variance. We only describe the effects of response error on point estimates for observations for the given sample. Bross (1954), Koch (1969), and others have laid out much of the following analysis using slightly different terms, so the next section is a selective review of some basic concepts.

Assume that an event calendar records a series of events as "YES" or "NO" for each month. For example, an interview might ask whether the respondent used birth control pills at all in each month of the calendar. The survey will then contain a distribution of observed responses that match closely but not perfectly with the underlying distribution of true responses as shown in Table 1 below.

< Table 1 about here. >

In Table 1, the distribution of observed "YES" and "NO" responses corresponds to the marginal totals ***g*** and ***h***, respectively. Marginal total ***g*** consists of all correctly reported "YES" events ***a*** and all incorrectly reported "NO" events ***c***. Similarly, marginal total ***h*** consists of all incorrectly reported "YES" events ***b*** and all correctly reported "NO" events ***d***. A good survey will have relatively few response errors ***b*** and ***c***, and the marginal totals ***g*** and ***h*** of observed responses will correspond closely to the marginal totals ***i*** and ***j*** of actual events. We may define two error ratios, the odds of a response error among true "YES" events $e_1 = b / a$, and the odds of a response error among true "NO" events $e_2 = c / d$. These two error ratios account for differences between misreporting of true "YES" events and misreporting of true "NO" events. We use error ratios rather than error rates in the following examples to simplify the algebra.

The effect of response errors in a prevalence estimate

If we know e_1 and e_2 and we have the observed marginals g and h, we can calculate the impact of response errors on a simple study of prevalence. For example, assume we wish to know what proportion of women of childbearing age who have ever taken birth control pills are currently taking birth control pills. For responses from a given month, we can define the effect of response error as the difference between the fraction of respondents reporting pill use and the fraction who truly use the pill in that month. The following equation then defines the effect of response error on a prevalence estimate E_p:

Eq. 1: $$E_p = (g / (g + h)) / (i / (i + j)) = g(i + j) / (g + h)\, i$$

We do not know the numbers of true "YES" and "NO" events i and j, but if we assume knowledge of the error ratios e_1 and e_2 we can substitute and solve:[1]

Eq. 2: $$E_p = g(i + j) / (g + h)\, i = \frac{g}{(g - e_2 h)} \frac{(1 - e_1 e_2)}{(e_1 + 1)}$$

This effect of response error is thus the product of two fractions. The first fraction, $g / (g - e_2 h)$, describes how much the e_2 error odds inflate the observed rate above the true rate. The second fraction, $(1 - e_1 e_2) / (e_1 + 1)$, basically describes how the e_1 error odds *deflate* the observed rate below the true rate. Note that the error odds e_1 and e_2 cancel each other to some extent as errors move observations back and forth across the cells shown in Table 1. If ge_1 is about equal to he_2, the balanced distributions of observations and error odds will minimize the effect of response error (E_p will be close to one). However, if ge_1 is much larger or smaller than he_2, E_p will be much smaller or larger than one, and the observed prevalence $g / (g+ h)$ will be very different from the true prevalence $i / (i + j)$.

Equation 2 above serves an instructive purpose, but we can not easily apply it in this study. As we shall explain later, detectable errors in event calendars give us no basis to estimate e_1, the error rate among true "YES" events. Without knowledge of e_1, we cannot calculate the effect of response error in a prevalence estimate. Bross (1954), however, identifies circumstances where we can expect response errors to distort our findings. When observed responses have a lopsided distribution, the few correct observed responses in the rare category can be overwhelmed by the many incorrect responses moving across from the dominant category, even if the error ratios are low. Thus, unequal distributions of responses should alert us to use caution in making prevalence estimates.

The effects of response errors in a rate estimate

When we use a survey to estimate event rates instead of prevalence, we make greater demands on the data. Consider the example of a contraceptive failure rate. This rate is actually one prevalence divided by another prevalence. Contraceptive failures occur in months when a pregnancy begins, so we know the observed failure rate if we have a count g_c of contraceptive uses in months with a conception and a count g_{tot} of contraceptive use in all months of the calendar. Then we may calculate the effect E_r of response error on the observed failure rate.

---

[1] To substitute and solve, first derive the equations for cells $a, b, c,$ and d in terms of $g, h, e_1,$ and e_2. Then use the equations $i = a+b$ and $j = c+d$ to solve for marginal values.

Eq. 3: $E_r = (g_c / g_{tot}) / (i_c / i_{tot}) = g_c \; i_{tot} / g_{tot} \; i_c$

Again, we do not know the values of i_{tot} and i_c, so we substitute and solve in terms of e_1, e_2, and observed values for h_{tot} and h_c,

Eq. 4: $$E_r = (g_c / g_{tot}) / (i_c / i_{tot}) = \frac{g_c}{g_{tot}} \frac{(g_{tot} - h_{tot} e_2)}{(g_c - h_c e_2)}$$

Equation 4 contains only one error ratio because e_1 drops out. As we shall show, e_2 is the error ratio one can derive from an analysis of response errors.[2] Thus, we will be able to solve the equation for and make estimates of E_r.[3] Note also that the effect of response errors depends on the distributions of observed responses in all months and on the distribution of observed responses in months with a conception. If either distribution is unbalanced, response error can strongly influence the observed failure rate.[4]

The effect of response errors on measures of association

Our final example involves differences of group means. For example, Hyde et al. (1996) surveyed 570 new mothers at one month postpartum and found that women who breastfed were less

[2] The error ratio e_1 still has an effect, but that effect is entirely incorporated within the observed marginal totats.

[3] Note to discussant: The fact that e_1 drops out of Equation 4 is crucial for an analysis of detectable superposition errors, but is relatively unimportant for other types of error analyses. We have tried without success to locate discussions of this topic in the research literature, and would be grateful if you knew of any paper or statistics textbook that treats this issue.

[4] If a respondent stops using a contraceptive early in a month to conceive a pregnancy later in the same month, that case will be mislabeled as a contraceptive failure. The population of observed contraceptive failures would thus consist of true failures, mislabeled cases, and response errors. In all the examples we shall study, any such mislabeled cases would cause us to *underestimate* the effect of error on contraceptive failure rates.

likely to have resumed sexual intercourse than women who never breastfed. The NSFG contains a sexual nonintercourse calendar, so we should be able to replicate this difference between breastfeeding and nonbreastfeeding mothers. Equation 5 represents the effect of response errors on the measured difference in postpartum abstinence between breastfeeding (b) and nonbreastfeeding (nb) women.

$$\textit{Eq. 5:}\quad E_d = \frac{g_b / (g_b + h_b) - g_{nb} / (g_{nb} + h_{nb})}{i_b / (i_b + j_b) - i_{nb} / (i_{nb} + j_{nb})}$$

When we solve by substitution, the equation simplifies.

$$\textit{Eq. 6:}\quad E_d = \frac{g_b / (g_b + h_b) - g_{nb} / (g_{nb} + h_{nb})}{i_b / (i_b + j_b) - i_{nb} / (i_{nb} + j_{nb})} = \frac{1 - e_1 e_2}{(1 + e_1)(1 + e_2)}$$

We cannot solve for E_d if we have no information about e_1, but we may still draw some insights from Equation 6. First, unless e_1 and e_2 are both zero, the effect of response error will always be to reduce the power of the study; that is, $E_d < 1$. Second, and unlike the case for other measures, the effects of e_1 and e_2 act in concert; they do not counterbalance each other. Thus, response errors generally have significant and directional effects on measures of association. Because the observed differences are biased downward relative to the true differences, statistical tests of associations are generally weaker but still valid in the presence of response error (Mote and Anderson, 1965).[5] However, coefficients in causal studies may respond unpredictably to response errors because errors in one variable can affect coefficients of other variables (see Fuller, 1991). Note finally that the marginal

---

[5] Response errors have an additional detrimental effect in that they increase the error variance. This effect, however, will usually be much smaller than the direct effect on observed mean scores for the covariates.

totals ***g*** and ***h*** drop out of Equation 6. This means that unequal distributions of observed categories do not magnify the effects of errors in measures of association.

Assumptions for estimating error rates

Now that we have discussed the implications of response errors for different types of studies, we shall discuss the assumptions that enable us to derive error rates from a count of detectable errors. Whenever possible, we make conservative assumptions to produce minimum estimates of error rates.

We identify detectable errors by superimposing a contraceptive calendar and a sexual nonintercourse calendar over a birth history calendar. In this and other studies of errors of superposition (e.g. Becker and Sosa, 1992), most detectable errors are tied to a birth. The most common detectable error in these studies is contraceptive use during (late) pregnancy. If there is no reason to use contraceptives while pregnant, then either the contraceptive calendar or the birth calendar must be incorrect. At this point we make our first assumption:

1.) *Only one event calendar contains response errors. The other, the birth history, contains no response errors.*

To argue that this assumption is indeed conservative, we consider the alternative. If we superimpose two event calendars which both have error rates, a given time interval could contain an error in one calendar, or the other, or both. However, an erroneous superposition of events can only be discovered in one of those three conditions; which one depends on the actual events in that time interval. Thus, total errors will always outnumber discovered errors when both event calendars are fallible.

The second assumption is fairly obvious, but it is worth considering carefully.

2.) *We can designate events and specific months for which every observed "YES" response is an error.*

Some studies of superposition errors in event histories (e.g. Becker and Sosa, 1992) count as errors events which are improbable but not impossible. For example, some women might use condoms during pregnancy, perhaps to prevent sexually transmitted diseases, or perhaps because they are not yet convinced they are pregnant. Many reports of condom use during pregnancy are no doubt erroneous, but not all are. Thus, we must carefully choose time intervals and events for which there is essentially no chance that an observed "YES" response can be correct. We discuss our attempts to do this in the data and methods section.

< Table 2 here>

At this point we present a new version of Table 1 which identifies observed responses only for events and months in which every observed "YES" response is a detectable error. In Table 2 above, there are no true "YES" events, so we assign all cases in observed categories g' and h' to cells c' and d', respectively. This allows us to calculate the odds of a response error among true "NO" events e_2' *for the month of detectable errors only*.

$$e_2' = g' / h'$$

Now we are ready for assumption three, the crucial one.

3.) *The odds of a response error among true "NO" events is the same in all months of the event calendar.*

Referring to Tables 1 and 2, this implies that $e_2' = e_2$. This assumption is not true, so we must make sure it is conservative. In a sense, a conservative bias is built into any study of response errors based on erroneous superpositions of events. As we have mentioned, the live birth which creates an opportunity to study detectable errors also provides a "landmark" by which a respondent can identify dates of less memorable events occurring about the same time. Indeed, NSFG event calendars specifically instruct respondents to recall birth events in piecing together event calendars. From this fact alone we might conclude that detectable error rates are much lower than response error rates in the general universe of an event calendar.[6]

In another sense, detectable error rates may not represent a conservative estimate of response errors throughout an event calendar. Pregnancy and childbirth is a complex time for contraceptive use. If a respondent has long intervals of consistent contraceptive use, her reports for those intervals should have fewer errors than for intervals of inconsistent contraceptive use such as those surrounding a pregnancy. Of course, intervals of inconsistent contraceptive use – such as method switching, entering or leaving relationships, attempting to become pregnant, pregnancy and childbirth, and breastfeeding – are often the most interesting to a researcher. Still, to estimate minimum error rates across a whole method calendar, we must make the assumption that a pregnancy improves recall at least as much as it impedes recall.

We will attempt to support assumption three by arguing that response errors should be distributed the same way as missing data. We can then examine months with missing data in the NSFG contraceptive calendars. If respondents report fewer months of missing contraceptive data in late

---

[6] Data quality studies which compare detectable errors in event calendars with errors in traditional questionnaires have a critical weakness. Most months in which one can detect superposition errors are adjacent to landmark events, where the advantages of an event calendar are greatest. Thus, an analysis of detectable errors provides only an upper bound estimate of the relative benefits of an event calendar.

pregnancy than in any other part of the calendar, then assumption three is more likely to be conservative.

As a final note about assumption three, notice that the assumption makes no mention of the error ratio among true "YES" events e_1. We cannot make any statements about e_1 because there are no superpositions in the event calendars we study for which an observed "NO" response is a detectable error. Because we lack information on e_1, we cannot directly estimate the effect of response error on measures of prevalence or association.

Assumption four concerns the population "at risk" of a detectable error.

4.) *Respondents who truly experience an event during the interval covered by the calendar report such an event at least once in the event calendar, and respondents who do not experience such an event do not report it at all in the calendar.*

Assumption four excludes respondents whose response errors can not be detected by our study; they may be making response errors, but at some unknown rate. This assumption thus restricts the sample population we examine, and also clarifies our beliefs about the origins of response errors. For the example of pill use, this means that e_2 shows the odds that a woman who sometimes uses birth control pills will misassign pill use to a given month; e_2 provides *no* information about the odds that a respondent who never uses birth control pills will report doing so.

Of course, a few women who never use the pill might have their contraceptive use miscoded, and such miscodings might show up as detectable errors in violation of assumption four. However, our estimates of error effects are robust with respect to such cases. If we omit assumption four, we increase the numbers of respondents in categories h and h' in Tables 1 and 2. The larger number of h'

cases drives down the estimated error rate $\boldsymbol{e_2'}$ in Table 2. However, the proportion of $\boldsymbol{h}$ cases increases even more and unbalances the distribution of observed "YES" and "NO" responses. Thus, if we relaxed assumption four, the estimated effect of response errors on most prevalence or rate estimates would remain high, even though $\boldsymbol{e_2'}$ would be lower.

The fifth and final assumption is the weakest.

5.) *Any variables that correlate with detectable response errors correlate with all response errors.*

Using logistic regressions, we can find background, interviewer and interview characteristics that predict detectable errors. We make assumption five to generalize those findings to all response errors. However, most detectable errors are strongly linked to landmark events (births), while response errors at other intervals are not. Thus, any statistically significant background, interviewer, or interview characteristics may only predict difficulty with cognitive markers, not a general propensity to response error. We thus urge caution in interpreting the covariate analyses of detectable response errors.

To summarize, we have made a series of assumptions to estimate response errors in entire NSFG event calendars based on improbable superpositions in some parts of those calendars. We can assure the soundness of some of our assumptions by careful selection of events and intervals. Other assumptions, though imperfect, are conservative because they define a lower limit on the impacts of response errors. However, the fifth assumption, and to some extent the third and fourth assumptions, are still open to challenge and might distort our conclusions.

Data

Data source

This study uses data from the National Survey of Family Growth (NSFG) cycle V (1995). The NSFG 1995 gathered fertility, social, economic, and family background data from a national sample of 10847 women in the United States age 15 to 44 at interview. The NSFG identifies sampling weights for extrapolation to the U.S. population. We did not adjust for sampling weights in the descriptive analysis, and our logistic analyses include statistical controls for minority groups and other oversampled categories. The 1995 NSFG sample was drawn from a larger sample in the 1993 National Health Interview Survey, and had a response rate of 77 percent.

We used three event calendars from the NSFG in our analyses. The first calendar is a monthly report of contraceptive use. Interviewers asked all respondents who had ever had sexual intercourse to identify the primary contraceptive they used, plus up to three additional methods and overlapping methods, for each month from January 1991 to the interview month in 1995. In this study we recorded the first two methods reported in each month. (Most respondents listed no methods or only one method in a month.) The second calendar is a sexual nonintercourse history for the same time interval. Interviewers asked sexually experienced respondents if they had any intervals of one month or more with no sexual intercourse since January 1991. For respondents who reported a month or more of no sexual intercourse, interviewers asked the starting and ending months of up to four such intervals. The third calendar is a birth history drawn from a pregnancy history in which respondents reported the date, pregnancy length, and many other variables pertaining to each pregnancy. We focus on live birth pregnancies after January 1991.

In the 1995 NSFG instrument, respondents completed the birth history first. Next, respondents used the births as landmarks to complete the sexual nonintercourse history. Finally, respondents used both the birth history and the sexual nonintercourse history as landmarks to complete the contraceptive use portion of the event calendar.

Detectable errors

We identified a first type of detectable error by superimposing the contraceptive history over the birth history. If a woman reported using birth control pills, injectable hormonal contraceptives or morning-after pills, or female barrier methods not including female condoms, at exactly two months before a live birth, then we coded a response error for that month. We did not code male or female condom use during pregnancy as a response error because women might use those barrier methods during pregnancy to prevent sexually transmitted diseases. Similarly, we did not identify response errors involving insertable hormonal contraceptives or IUDs because of the possibility that women did not have the devices removed after they became pregnant. Finally, we did not code contraceptive use in other months of a pregnancy as a response error; we did this to allow for the possibility that women discovered their pregnancies late.

We superimposed the sexual nonintercourse history and the birth history to identify a second type of response error. We coded a response error if a woman reported three or more consecutive months of no sexual intercourse, and if the middle of the three months coincided with a conception that led to a live birth. Sixteen women with a live birth from 1991 to 1995 reported using artificial insemination or in vitro fertilization services, and one of those women conceived a live birth amid three months of reported sexual abstinence. We removed all 16 such cases from our analyses of this type of response error.

The third type of response error is an incompatible superposition of the sexual nonintercourse history and the contraceptive use history. We coded an error when the respondent reported using a condom or female barrier method in the middle month of three consecutive months of no sexual intercourse.

Both the second and the third type of response errors involve sexual nonintercourse calendars, and sexual nonintercourse calendars have a special problem. We can confidently detect an error in a sexual nonintercourse calendar only when the respondent reports no sex in an interval when she did have sexual intercourse, such as a month that corresponds to a live birth conception. However, by the design of the sexual nonintercourse calendar, we expect many more response errors of the opposite type – respondents reporting sex in intervals when they did not have sexual intercourse. We expect such errors because about half of all NSFG respondents skipped the sexual nonintercourse calendar by answering "no" when asked if they had any intervals of no sexual intercourse since 1991. All these women thus had an event calendar that indicated that they had sex every month. We suspect some such reports were incorrect, but detectable error analysis alone cannot support or dismiss our suspicions.

Because we had no direct evidence of the effect of response errors in the sexual nonintercourse calendars, we turned to indirect evidence. Women often stop having sexual intercourse for one or more months during late pregnancy or early postpartum. Hyde et al. (1996) recently performed a survey of pregnancy and postpartum sexuality in which they interviewed 570 couples in the second trimester and one, four, and twelve months postpartum. We age-matched the NSFG sample to the sample in the Hyde et al. survey by excluding 385 teenage mothers and retaining 2799 births in 1991-95 to 2413 women. We then based our main analyses of response errors in the sexual nonintercourse calendar on a comparison of the NSFG and Hyde et al. data on postpartum sexuality.

Explanatory variables

The NSFG contains many variables that we employed to find predictors of response errors. We gathered data in three broad categories based on characteristics of the respondent, the interviewer, and the interview. Characteristics of the respondent include the contraceptives she uses, the circumstances of her birth, current circumstances such as labor force status, marital status and education, and finally background variables such as race, childhood family structure, the respondent's mother's education and age at first birth. Characteristics of the interviewer include race, education, interviewing experience and training, and willingness to answer questions about interviewer characteristics. Characteristics of the interview include type of interview, duration of recall, amount of missing data on other variables, and whether the respondent skipped the sexual nonintercourse calendar by reporting no intervals at all of no sexual intercourse.

Restricting the samples

Many NSFG respondents did not report any intervals of no sexual intercourse, and for each contraceptive method, many respondents never used that method. In part of the descriptive and all of the logistic analysis, we restricted our sample to those women who were at any risk of committing the given type of response error. For example, of 2713 women with a first birth after January 1991, 853 reported an interval of three or more consecutive months of no sexual intercourse and had a sexual nonintercourse history that included the month of conception. Almost half (1322) of women with a live birth after January 1991 used birth control pills in the 1991-1995 interval, and over half (1596) used birth control pills, injectable hormonal contraceptives, morning-after pills, or female barrier methods excluding condoms. These groups represent the samples for our logistic analyses; if we had not limited the samples in this way, our analyses would have confounded the predictors of response errors with predictors of contraceptive use or sexual nonintercourse.

Methods

Our descriptive analyses show counts of discoverable errors under different assumptions about the population at risk of a discoverable error. After we calculate rates of detectable errors, we use those rates to estimate the effects of response errors. For three types of contraceptive use – pill, female barrier method, and injectable/morning after, we estimate the minimum impact of response error on a hypothetical study of contraceptive failure rates. We base our estimation procedures on Equation 4 and the assumptions we outlined in the introduction.

In addition to the descriptive analyses, we perform logistic regressions on the odds of a report of incompatible contraceptive use two months before a live birth. We do logistic regressions on detectable errors for pill use during pregnancy and for any incompatible contraceptive use during pregnancy. Other types of detectable errors have too few events to support a regression analysis. In cases where the detectable error coincides with a birth, respondents with two or more live births have two or more chances for a detectable error. To account for this circumstance we only examine errors associated with the most recent live birth of each respondent.[7] We use the STATA 5.0 statistical package (1997) for all logistic regressions.

After we study contraceptive use during pregnancy, we analyze two types of detectable errors associated with the sexual nonintercourse calendar. The first detectable error superimposes sexual nonintercourse histories on birth histories. After we determine detectable error rates based on sexual abstinence and live birth conceptions, we compare the "prevalence" rates of postpartum sexual abstinence in the NSFG and in the 1996 Hyde et al. study. Then we compare the NSFG and Hyde studies in a simple measure of association; differences in postpartum abstinence among breastfeeding versus nonbreastfeeding women.

The second detectable error superimposes sexual nonintercourse calendars on contraceptive calendars. Unfortunately, both of these calendars are assumedly subject to error, and we cannot interpret these detectable error rates without making strong and unsupported assumptions about the

[7] We also ran logistic regressions on all births and used robust variance estimates (White, 1980) to compensate for multiple observations from some respondents. These runs did not alter any coefficients or their interpretation in a major or systematic fashion. Results of these and other sensitivity analyses are available on request from the first author.

underlying error processes. We briefly display and interpret results for this kind of detectable error in Appendix A.

Results

Descriptive analysis of contraceptive calendars

Table 3 shows rates of detectable errors in the NSFG 1991-95 contraceptive calendars. The detectable error in all cases is incompatible contraceptive use two months before a live birth.

The first row of figures shows the percent of all respondents with a live birth from 1991 to 1995 who reported a detectable error in their contraceptive history. The first row of figures provides a crude picture of the total number of errors, but the error rates in the first row confuse two separate rates – the rates of detectable errors among contraceptive users of a given type, and the rates of contraceptive users of a given type in the general population.

< Table 3 here >

Rows two through four show detectable error rates among users of each type of method.

Cell values in the second through fourth rows show that more than 10 percent of contraceptive users who could have reported a detectable error, did report a detectable error. The detectable error rate is slightly higher for barrier methods than for birth control pills and is much lower among the few women who reported using injectable or morning-after contraceptives. These detectable error rates are higher than the detectable error rates in other studies of event calendar surveys in developing countries (e.g. Becker and Sosa, 1992).

In addition, it appears that NSFG respondents were several times as likely to make a response error as to report that they didn't remember the events of a month. Cells in the second through fourth rows exclude cases that are missing contraceptive use in the second month before a live birth. Row five, which corresponds to row four, shows the total number of live births dropped because of missing contraceptive data two months earlier. Dropped cases number less than four percent of all observations for each method. Thus, in comparison to detectable response errors in other surveys and in comparison to missing data scores, detectable response errors in the NSFG contraceptive calendar are disturbingly common.[8]

Effects of response error on prevalence measures

Table 4 shows observed monthly use rates for different contraceptive methods. We can use these values in the equations for the effect of response error on summary measures, although we cannot solve some equations without an estimate for e_1, the odds of a response error among women who are truly using a contraceptive method in a given month. For all three contraceptive methods we are studying, the distribution of observed responses is fairly balanced in the entire calendar. This means that

[8] We can use missing data scores to test assumption three from the introduction. Row five in Table 3 shows missing data for 2.1 percent of contraceptive use reports at two months before a live birth. If we calculate missing data rates in relation to the most recent live birth (not shown), the value is 2.0 percent. By comparison, the lowest rate of missing data for any calendar month was 2.9 percent in December 1994, the most recent month for which all respondents had interviews. Thus, if response errors are distributed similarly to missing data, detectable error rates are almost certainly lower than overall response error rates in the NSFG contraceptive calendars.

women who use a method at all will report using it in about half the months of the contraceptive calendar.

How then will response errors in the NSFG affect prevalence estimates for overall contraceptive use? Even though we cannot solve Equation 2, we infer from the balanced distribution of observations that a prevalence estimate for any of these contraceptive methods will be fairly accurate. Unless the unknown response error rate e_1 is near zero or is much higher than e_2, most of the response errors will cancel each other.

Effects of response error on contraceptive failure rates

Unfortunately, response errors have a much larger effect on contraceptive failure rates than on simple prevalence estimates. In Table 4, observed contraceptive reports in months involving a live birth conception are (not surprisingly) skewed toward "NO" responses. When we use live birth pregnancies in the NSFG to calculate contraceptive failure rates, we get results shown in Figure 1. Figure 1 also shows the minimum effect of response error on observed contraceptive failure rates. The black columns adjust for the effects of error counts in row four of Table 3, estimated using Equation 4 from the introduction.

< Figure 1 about here >

In Figure 1, contraceptive failure rates adjusted for error are much lower than observed failure rates; more than 40 percent lower in the case of pill use. The effect of error varies because contraceptive methods vary in their observed distributions of use and in their rates of response errors.

Surprisingly, response error has the least effect on failure rates for barrier methods, which had the largest proportion of response errors. In all cases, response errors exaggerate contraceptive failure rates and distort comparisons of failure rates among different contraceptive methods.[9]

Effects of response error on measures of population differences

Finally, we come to the effects of response error on differences in contraceptive use among two hypothetical categories of women. As is the case with prevalence estimates, we do not have enough information to estimate the effects of response error on a difference of population proportions. For the case of pill use, Table 3 gives us a minimum estimate of the known error odds e_2.

$$e_2 = 167 / (1565 - 167) = .12$$

If we assume zero response error among respondents who are truly using the pill in a given month, we can plug this value for e_2 into Equation 6. If e_1 is zero, we calculate that the expected observed difference of proportions for some study of pill use will be .89 as large as the true difference. If e_1 is any value greater than zero, we can expect the observed difference to shrink even more. However, as we shall see in the following section, response errors that correlate with the categories may cause more problems than the general loss of resolving power.

[9] The findings in Figure 1 are strengthened somewhat by assumption four in the introduction. For example, if we completely relax assumption four, the adjusted maximum proportion for pill use would be .00217 rather than .0019. However, to completely relax assumption four, we must assume that a woman's true contraceptive use has no bearing on her likelihood of reporting contraceptive use two months before a live birth. Our forthcoming logistic regressions strongly suggest this is not true. An addition, if we relax assumption four, we would need to account somehow for all the response errors which are occurring but are not counted because they do not involve the contraceptive method we are studying.

Logistic regression analysis of contraceptive calendars

We now know that NSFG contraceptive calendars have generated enough response errors to distort at least one type of analysis. Next we turn to logistic regression analyses of variables that predict response errors in the contraceptive calendars. Table 5 shows descriptive statistics for the explanatory variables in the logistic models. For the first two groups of variables, interview and interviewer characteristics, we are primarily interested in finding ways to reduce response errors in future studies. For the third group of variables, respondent characteristics, we are interested in finding associations between response error and social background that could distort causal analyses of social background and contraceptive use. Some variables could fit in more than one category.

< Table 5 here >

Table 6 displays results for four logistic regression models. Model 1 examines interview characteristics, Model 2 interviewer characteristics, Model 3 respondent characteristics, and Model 4 all characteristics together. The outcome of interest in each model is the odds that a respondent will report using birth control pills two months before her most recent live birth. The coefficient for each variable portrays a multiplicative effect of that variable on the odds of a detectable error. Thus, if a covariate has a coefficient of 2.00, then a unit increase in that covariate doubles the odds of an error.

< Table 6 here >

Model 1 in Table 6 tells us which interviews are most likely to contain detectable response errors in their contraceptive history. As one might expect, longer durations of recall lead to more response errors. Similarly, we find more response errors when respondents refuse or do not recall information about their backgrounds or their target birth. The few Spanish language interviews also contained frequent response errors. Finally, we find more response errors when respondents skipped the sexual nonintercourse calendar by answering "no" when they were asked if they had any intervals of no sexual intercourse. We will discuss this skipping variable again in the results for sexual nonintercourse histories.

Model 2 examines interviewer characteristics. Of the interviewer characteristics we included in our models, a few showed statistically significant effects in one or more models. Race of interviewer seems to count; when the interviewer and respondent are the same race or ethnicity, we find fewer response errors.[10]

The only evidence that interview experience decreases response error comes from the variable for interview month in Model 2. We assume that as the month of interview increased from January to October of 1995, interviewers learned to administer the interviews more and more effectively. In Model 2, interview month confuses effects of interviewer experience and duration of recall. In Model 4, where we separately estimate the effect of duration of recall, we find that the odds of an error declined by about 15 percent each month (coefficient = .85, $p<.01$).

In Model 3, we first look at characteristics associated with the respondent's background and race. Respondents who spent part of their childhood without one or more parents in the household are almost twice as likely as other respondents to have a detectable response error. Our other two measures of social disadvantage, the respondent's mother's level of education and age at first birth,

[10] Interviewers with computer experience appear to have higher rates of response errors in the contraceptive calendars of their interviews. This "computer use" effect may occur because interviewers with computer experience were trained to use skips in the program for the interview instrument (personal correspondence with A. Chandra, 1998) We also expected years of interviewing experience to decrease response errors, but we found if anything the opposite trend. Given that NSFG interviewers are trained to probe "I don't know" responses, experienced interviewers may have more skill at drawing guesses (accurate *and* inaccurate) out of uncertain respondents.

were not statistically significant. However, all three family background coefficients point to higher response error rates for respondents with disadvantaged social backgrounds. The large error variance prevents us from drawing strong conclusions, but these results imply that response error could compromise causal studies of social background and contraceptive use. Coefficients for race variables in Model 3 suggest that nonwhite or Hispanic respondents have more detectable errors than white respondents. However, the effects of race and ethnicity largely disappear in Model 4, where we control for race of the interviewer.

The next cluster of variables in Model 3 identifies circumstances of the most recent birth. Premature births are fairly rare, but they seem to increase the likelihood of a detectable error. Breastfeeding decreases the likelihood of a detectable error. We believe breastfeeding improves reporting of birth control pill use. Most women who breastfeed do not resume using birth control pills for some months after they give birth, and longer intervals of contraceptive nonuse are easier to recall correctly.

Our final group of respondent variables describes the respondent's status at the time of the interview. According to Model 3 and the corresponding part of Model 4, women married at interview may have lower response error rates. The association between marriage and response error may be random, it may indicate that currently married women have had more stable recent contraceptive histories, or it may be further evidence of a correlation between social disadvantage and response error.

Also, women with more children at interview have more frequent response errors, probably because of the cognitive challenge of assigning events and dates to the correct birth.[11]

The strongest predictor of response errors in Model 3 is whether the respondent reported contraceptive use in a majority of months on the calendar. This covariate causes conceptual problems because all detectable errors are reports of contraceptive use, so to some extent this independent variable is simply a very crude measure of the dependent variable. However, we believe that a calendar with many correct months of contraceptive use will be more likely to have a mistaken month of contraceptive use. Thus, unless we include some measure of frequency of contraceptive use, we cannot tell whether other variables predict detectable response errors or simply frequent contraceptive use.[12]

Figure 2 demonstrates two findings about the relationship between contraceptive reports and contraceptive errors. First, detectable errors increase in proportion to the percent of all months in the contraceptive calendar which mention pill use. In other words, more months of pill use seem to create more chances for a detectable error. Second, 57 women who simply reported pill use in every month of the calendar account for almost half (42 percent) of the 136 detectable errors among the 1322 women in the sample. Thus, the few interviews with unchanging contraceptive histories produce a substantial proportion of all detectable response errors.

---

[11] We believe that the coefficient for number of children indicates a cognitive rather than a social effect. Most women in the United States who have one child have a second child soon after, regardless of age or social background. One might thus expect little effect of parity two on response error rates if social processes mediated the effect of parity on response errors. However, sensitivity analyses of separate parity levels suggest that the unit effect of a second birth on response errors is greater than the unit effect of a subsequent birth.

[12] Various sensitivity analyses omit or respecify the variable for frequency of contraceptive use. We find no evidence that these alternative models influence any other coefficients by more than a small amount.

Table 7 presents logistic regression models similar to those in Table 6. Where Table 6 examined only detectable errors involving contraceptive pill use, Table 7 includes other contraceptive methods as well – the pill, barrier methods, injectable contraceptives, and morning-after pills. We will limit our discussion of Table 7 and say that most of the covariate effects for detectable errors involving pill use generalize to other contraceptive methods.

< Table 7 about here >

Descriptive analysis of sexual nonintercourse calendars

We turn now from the contraceptive method calendar to the sexual nonintercourse calendar. Table 8 shows rates of detectable errors in NSFG 1991-95 sexual nonintercourse calendars. Rows two through four indicate detectable errors for about 5 percent of respondents who reported a 3- or more month interval of abstinence. A 5 percent detectable error rate for the sexual nonintercourse history seems better than the 10+ percent detectable error rate for the contraceptive history, but the two rates are not really comparable.

< Table 8 here >

There are several reasons we are reluctant to conclude that the sexual intercourse calendar has fewer errors than the contraceptive calendar. First, we underestimate the response error rate by only counting detectable errors when three consecutive months of no sexual intercourse surround a live birth conception. Second, even women who report intervals of no sexual intercourse report only about ten such intervals in a 50+ month event calendar. Thus, if we took each respondent's sexual nonintercourse calendar and randomly rearranged her responses, the detectable error rate would still be below 20 percent. Finally, we believe that the response errors that can lead to detectable errors are much less common than the response errors that cannot be detected. The undetectable errors are months in which women do not have sexual intercourse but report that they do.

To examine response errors that cannot be detected in the NSFG itself, we compared NSFG sexual nonintercourse data with data from another survey. Hyde et al. (1996) interviewed women during the second trimester of a pregnancy, at one month postpartum, at four months postpartum, and at 12 months postpartum. The Hyde et al. study focused on sexual behavior, and respondents reported sexual nonintercourse for the month preceding each interview. Thus, we believe that the Hyde study provides more accurate data about postpartum sexuality than the NSFG. We restricted the NSFG sample to exclude teen mothers (who were not included in the Hyde et al. study). Then we compared the proportions of women who reported sexual intercourse in the previous month. Figure 3 shows our results.

< Figure 3 about here >

The open squares in Figure 3 show reported sexual intercourse rates for NSFG respondents. As one might expect, fewer NSFG respondents have sexual intercourse the month of and after a birth than in other months. However, the Hyde et al. study (closed squares) suggests that almost all women suspend sexual intercourse for a month or more after a birth. The corresponding proportion for the NSFG is one third or less.

If the Hyde et al. study is accurate, then about half of NSFG respondents with a live birth in 1991-95 failed to report postpartum sexual abstinence in their sexual intercourse calendars. In other words, Figure 3 suggests that e_1 (the odds that a month of sexual abstinence will be misclassified) exceeds 1 for certain critical months. Much of the problem here should be attributed to the entry question into the sexual nonintercourse calendar; 42 percent of NSFG respondents with a live birth from 1991-95 skipped the sexual intercourse calendar by reporting that they had no intervals of no sexual intercourse.

Table 9 shows our final comparison between the NSFG and Hyde et al. sexual nonintercourse histories. We used both studies to estimate differences in postpartum abstinence by whether the mother breastfed her baby. The Hyde et al. study found a statistically significant difference between the two groups. For the NSFG, the difference was smaller but still statistically significant, thanks in part to a larger sample size.

< Table 9 about here >

Without an estimate of the full matrix of response errors in the NSFG (and in Hyde et al., too), we cannot evaluate the relative effects of response and sampling error in each study. Also, while the results are suggestive, we cannot conclude by looking at the differences of proportions that the NSFG has definitely lost power because of response error. However, this result is *not* consistent with an e_1 odds of over 1, as suggested by Figure 3. Such a high response error rate should have decreased the observed differences of proportions in the NSFG by a larger amount.

We suspect that response errors correlated negatively with breastfeeding *because* postpartum abstinence correlates positively with breastfeeding. These correlated errors artificially restored most of the observed differences between breastfeeding and nonbreastfeeding mothers. However, because we cannot prove that this happened, the results in Table 9 weaken our conclusions.

Discussion:

Review of findings

In this study, we found detectable error rates exceeding 10 percent for many types of contraceptives. Starting with several assumptions, we calculated that such response error rates severely distort estimates of contraceptive failure rates. Other types of summary measures were less sensitive to response errors. We also found some evidence that response errors correlate with respondents' background characteristics, particularly family structure. Our findings for the NSFG sexual nonintercourse history are not as complete as for the contraceptive history, but we did find evidence that NSFG respondents underreport intervals of sexual abstinence.

Our analyses depended on several assumptions. In retrospect, we may have to change some methods to better match those assumptions. For example, we assumed that respondents never truly used female barrier methods during late pregnancy. This assumption may be wrong in a few cases, so we may have to drop or seriously qualify our findings for response errors associated with barrier methods. On the other hand, we may have been overzealous in excluding injectable contraceptives and intrauterine devices, so we might wish to include more information on response errors associated with those methods. We must also certainly modify our final assumption that detectable errors and response errors have the same causes.

Recommendations for researchers

Based on these findings, we recommend that researchers who use NSFG event calendars should interpret their findings carefully. The events recorded in those calendars may almost all be correct, but respondents often misplace the times of those events. The effects of response errors will vary for different study questions. To anticipate the effects of response errors, researchers should examine related detectable errors if possible. If researchers do not have the time or means to identify detectable errors, they should at least consider three important factors:

1.) For the event of interest, are "YES" and "NO" observations unequally distributed?
2.) For the event of interest, might the error rate e_1 among "YES" events be very different than the error rate e_2 among "NO" events?
3.) In the analysis, do the error rates e_1 and e_2 compound each other instead of canceling?

If the researcher answers "yes" to any of these questions, he or she should exercise additional caution.

Of course these are not new recommendations, but we believe we have provided new support for them. Event calendars can provide more information about response errors than most researchers currently use. As more studies incorporate event calendars, researchers should check their results using partial information about response errors.

Recommendations for improving the NSFG

Our logistic regression results suggest several ways to reduce response error in event calendars in future cycles of the NSFG. One remedy might be to match the race and ethnicity of the interviewer to that of the respondent. Perhaps more importantly, we recommend a different entry question for the sexual nonintercourse calendar to alleviate underreporting of sexual abstinence.

It is possible to reduce detectable errors by introducing edit checks that, for example, notify interviewers when respondents report contraceptive use during a pregnancy. We discourage overreliance on this remedy. Such a fix would give us more reliable estimates of contraceptive use *during pregnancy* but would serve no other research purpose. At the same time, selective edit checks would deprive researchers of any estimate of the reliability of the rest of the contraceptive calendar.

Simple fixes may reduce response errors somewhat, but the NSFG may require some broader adjustments as well. For example, we suspect that the sexual nonintercourse calendar, because it severely underreports intervals of sexual abstinence, causes more cognitive disruption than cognitive help for respondents trying to remember their contraceptive use. The NSFG instructs respondents to use the sexual nonintercourse calendar to fill in their contraceptive calendars. Mistakes in the sexual nonintercourse calendar thus propagate into the contraceptive calendar, as our logistic regressions strongly indicate. To address this problem, it may be better to eliminate the sexual nonintercourse calendar altogether rather than try to repair it.

This brings us to our final recommendation for the next NSFG. The 1995 cycle contains a cornucopia of data for all sorts of causal analyses, but all this data comes at a cost. We suspect that the burden on the respondent has become hard to bear. Motivation is critical to recall (Cannell et al., 1981), and there may be a motivation problem when more than a tenth of mothers who report pill use also report it during the third trimester. Of course, respondents don’t cause all response errors, although our analysis has often made it sound that way. Still, a simpler and shorter NSFG would also likely have fewer response errors associated with the interviewer and the instrument. Thus, we suggest the NSFG should become smaller to reduce the burden on respondents, interviewers, and programmers. Perhaps the best way to start is to eliminate event calendars like the sexual nonintercourse calendar, which are theoretically important but perhaps too inaccurate to be useful.

Bibliography:

Axinn W.G., J.S. Barber, and D.J. Ghimire. 1997. 'The neighborhood history calendar: A data collection method designed for dynamic multilevel modeling. *Sociological Methodology* 27:355-392.

Becker, S., and D. Sosa. 1992. "An experiment using a month-by-month calendar in a family planning survey in Costa Rica." *Studies in Family Planning* 23(6):386-391

Bross, I. 1954. "Misclassification in 2 x 2 tables." *Biometrics* 10:478-486.

Cannell, C., P.V. Miller, and L. Oksenverg. 1981. "Research on interviewing techniques." In S. Leinhardt, ed., *Sociological Methodology*. San Francisco: Jossey-Bass Publishers.

Chua, T.C., and W.A. Fuller. 1987. "A model for multinomial response error applied to labor flows." *Journal of the American Statistical Association* 82(397):46-51.

Fuller, W.A. 1991. "Regression estimation in the presence of measurement error." pp 617-636 in Biemer et al., eds., *Measurement Errors in Surveys*. New York: Wiley.

Goldman, N., L. Moreno, and C. Westoff. 1989. "Collection of survey data on contraception: an evaluation of an experiment in Peru." *Studies in Family Planing* 20(3): 147-157.

Groves, R.M. 1991. "Measurement error across the disciplines." In Biemer et al., eds., *Measurement Errors in Surveys*. New York: Wiley, pp. 1-28.

Hyde, J.S., J.D. DeLamater, E.A. Plant, and J.M. Byrd. 1996. "Sexuality during pregnancy and the year postpartum." *The Journal of Sex Research* 33(2):143-151.

Kelley J.E. W.D. Mosher, A.P. Duffer, and S.H. Kinsey. 1997. *Plan and Operation of the 1995 National Survey of Family Growth.* Hyattsville MD, National Center for Health Statistics.

Koch, G.G. 1969. "The effect of nonsampling errors on measures of association in 2 x 2 contingency tables." *Journal of the American Statistical Association* 64:852-863.

Loftus, E.F., and W. Marburger. 1983. "Since the eruption of Mount St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events." *Memory and Cognition* 11:114-120.

Mathiowetz, N.A. 1998. "The redesign of the National Longitudinal Survey of Youth 1979; The impact of biannual interviewing on nonresponse and measurement error." Paper presented at the Conference on Redesigning the National Longitudinal Survey of Youth, Bureau of Labor Statistics, Wahington, DC.

Mote, V.L., and R.L. Anderson. 1965. "An investigation of the effect of misclassification on the properties of Chi-squared tests in the analysis of categorical data." *Biometrika* 52:95-109.

White, Halbert. 1980. "Nonlinear Regression on Cross-Section Data." *Econometrica* 48(3):721-746.

Appendix A:

We superimposed contraceptive histories on sexual intercourse histories to count detectable errors involving contraceptive use during intervals of sexual abstinence. Table A1 shows that we found plenty of such cases, but we have no way to interpret them. When both event histories are fallible, the equations we have used do not apply.

< Table A1 here >

We did make one attempt to interpret these data, shown in the bottom row of Table A1. We assume that respondents correctly report the number of months of contraceptive use and the number of months of sexual abstinence. We further assume that respondents assign contraceptive use and sexual abstinence to different months with some amount of error. Row four of Table A1 shows the number of errors we might get if respondents assigned months completely at random. For these detectable errors, our observed error rates are about 40 percent as great as the completely randomized error rates. If we subject other detectable error rates to the same kind of comparison, we get error rates for pill use during pregnancy about 20 percent of random, and error rates for sexual abstinence during conception about 25 percent of random.

Table 1: Relationships between observed responses, correct responses, and response errors in survey data.

		observed response			
		YES	NO	TOTAL	*odds of an error*
true response	YES	a	b	$i = a + b$	$e_1 = b / a$
	NO	c	d	$j = c + d$	$e_2 = c / d$
	TOTAL	$g = a + c$	$h = b + d$		

Table 2: Relationships between observed responses, correct responses, and response errors in survey data for months and events corresponding to detectable errors.

		Observed response			
		YES	NO	TOTAL	*odds of an error*
true response	YES	0	0	0	$e_1 = ?$
	NO	c'	d'	$j' = c' + d'$	$e_2' = c' / d'$
	TOTAL	$g' = c'$	$h' = d'$		

Table 3: Rates of detectable errors in the NSFG 1991-95 contraceptive calendars.
Contraceptive use superimposed on birth history for women with at least one live birth.

	Type of error: Incompatible contraceptive use 2 months before birth			
	Error rates among true "NO" *responses:*			
	All incompatible	Pill only	Injectable/ morning after	Female barrier methods
Denominator of error rate:				
all respondents with one or more live births	**7.2 %** (194 / 2713)	**5.5 %** (148 / 2713)	**0.1 %** (3 / 2713)	**1.6 %** (44 / 2713)
all (pill/hormonal/ barrier) method users with one or more live births	**11.4 %** (194 / 1696)	**11.2 %** (148 / 1322)	**1.4 %** (3 / 217)	**13.3 %** (44 / 332)
most recent live birth of all (pill/ hormonal/barrier) method users	**10.7 %** (181 / 1696)	**10.3 %** (136 / 1322)	**1.4 %** (3 / 217)	**12.7 %** (42 / 332)
all live births to all (pill/ hormonal/barrier) method users	**11.5 %** (220 / 1910)	**10.7 %** (167 / 1565)	**1.4 %** (4 / 288)	**12.0 %** (49 / 407)
	Comparable missing data rates:			
missing data for same month: all live births to all (pill/ hormonal/barrier) method users	**2.1 %** (40 / 1950)	**2.1 %** (33 / 1598)	**0.0 %** (0 / 288)	**3.1 %** (13 / 420)

Table 4: Observed distributions of contraceptive use in NSFG 1991-1995 contraceptive calendars among respondents who reported any use of each type of contraceptive

Contraceptive method:	*Observed contraceptive use reports among contraceptive users:* for all months of the contraceptive calendar		*for months with a live birth conception*	
	YES (= ***g***)	NO (= ***h***)	YES (= ***g***)	NO (= ***h***)
birth control pills	106,419	75,883	264	1,102
female barrier methods*	19,036	27,019	95	250
injectable/morning after	8,434	9,031	9	241
total observations for all NSFG respondents	485,464		2,578	

*Female barrier methods exclude female condoms

Table 5: Descriptive statistics for interview characteristics, interviewer characteristics, and respondent characteristics in the NSFG 1995: Respondents who reported both a live birth in 1991-1995 and specific contraceptive use in the 1991-95 contraceptive calendar.

	Birth control pill users	*All methods for which errors are discoverable*
Interview characteristics		
mean months between birth and interview	21.6	21.9
mean # of birth questions with missing data	.07	.09
mean # of background questions w/ missing data	.06	.06
interview in Spanish	.05	.05
respondent skipped sexual intercourse calendar	.43	.42
Interviewer characteristics		
interviewer is same race or ethnicity as respondent	.65	.65
interviewer has computer experience	.63	.64
mean years of interviewing experience	5.3	5.3
mean interviewer's level of education (scaled)	5.3	5.4
mean age of interviewer	50	50
mean interview month (1 to 10 for Jan. to Oct.)	4.0	4.0
# of interviewer questions with missing data	.04	.04
Respondent characteristics		
level of education (1=hs diploma, 2=college)	1.2	1.2
labor force status (1=part-time, 2=full time)	.93	.89
married at interview	.65	.66
number of children at interview	1.9	2.0
uses contraceptive in majority of calendar months	.43	.47
uses injectable or morning after contraceptives	----	.14
uses female barrier methods	----	.21
baby delivered five or more weeks premature	.08	.08
teen birth	.13	.12
respondent learned of pregnancy after 4th month	.008	.007
pregnancy was unintended	.36	.37
respondent breastfed the baby	.50	.53
respondent's mother's level of education	.90	.91
respondent's mother had a teen first birth	.45	.45
respondent experienced nonintact family	.45	.45
nonhispanic white	.56	.57
Hispanic	.19	.19
nonhispanic black	.23	.23
other race	.02	.02
respondent born outside US	.11	.11
N	1322	1596

Methods for which errors are detectable include birth control pills, female barrier methods, and/or injectable/morning-after hormonal methods.

Table 6: Logistic regressions on the likelihood of a detectable error in the NSFG 1991-95 contraceptive calendars. Odds of a respondent reporting use of birth control pills two months before her most recent live birth.

	1: interview characteristics	*2: interviewer characteristics*	*3: respondent characteristics*	*4: full model*
Interview characteristics				
log months between birth and interview	1.24 (.14)*			1.38 (.17)*
# of birth questions with missing data	1.44 (.16)**			1.50 (.20)**
# of background questions with missing data	2.98 (.83)**			3.03 (1.06)**
interview in Spanish	2.06 (.69)*			7.07 (4.27)**
respondent skipped sexual intercourse calendar	1.76 (.33)**			1.62 (.37)*
Interviewer characteristics				
interviewer is same race or ethnicity as respondent		0.53 (.10)**		0.52 (.14)*
interviewer has computer experience		1.74 (.36)**		1.91 (.47)**
years of interviewing experience		1.02 (.03)		1.05 (.03)
interviewer's level of education		0.94 (.06)		0.93 (.08)
interviewer's age		0.99 (.01)		0.99 (.01)
interview month (1 to 10 for Jan. to Oct.)		0.93 (.04)		0.85 (.05)**
# of interviewer questions with missing data		0.81 (.20)		0.92 (.21)

continued

Table 6 (continued): Logistic regressions on the likelihood of a detectable error in the NSFG 1991-95 contraceptive calendars. Odds of a respondent reporting use of birth control pills two months before her most recent live birth.

	1: interview characteristics	*2: interviewer characteristics*	*3: respondent characteristics*	*4: full model*
Respondent characteristics				
background and race				
respondent's mother's level of education			0.86 (.13)	0.89 (.15)
respondent's mother had a teen first birth			1.42 (.30)	1.30 (.30)
respondent lived in nonintact family at some time			1.82 (.39)**	1.94 (.44)**
nonhispanic white			*omitted category*	
hispanic			1.33 (.43)	0.84 (.34)
nonhispanic black			1.81 (.48)*	1.14 (.37)
other race			2.23 (1.57)	2.16 (1.64)
respondent born outside US			1.13 (.40)	0.59 (.27)
circumstances of the most recent birth				
five or more weeks premature			1.84 (.62)	2.01 (.70)*
teen birth			0.71 (.26)	0.58 (.24)
respondents learned of pregnancy after 4th month			1.75 (1.56)	2.19 (2.09)
pregnancy was unintended			1.13 (.26)	1.24 (.30)
respondent breastfed the baby			0.62 (.14)*	0.56 (.13)*
current status				
current level of education (1=high school diploma, 2=college)			0.99 (.17)	1.14 (.21)
current labor force status (1 = part-time, 2 = full time)			0.88 (.10)	0.83 (.10)
married at interview			0.68 (.17)	0.58 (.15)*
number of children at interview			1.37 (.14)**	1.38 (.14)**
uses contraceptives in majority of calendar months			11.49 (2.99)**	14.12 (4.00)**
log likelihood	-415.4	-425.8	-339.6	-308.6

*Coefficients show the multiplicative effect of each covariate on the odds of a discoverable error. 26 cases were missing data on contraceptive use two months before most recent live birth and were dropped. Final N = 1322 *p < .05 **p<.01*

Sample includes only women who reported birth control pill use in their contraceptive calendar.

Table 7: Logistic regressions on the likelihood of a detectable error in the NSFG 1991-95 contraceptive calendars. Odds of a respondent reporting use of birth control pills, female barrier methods excluding female condoms, or injectable/morning-after hormonal contraceptives two months before her most recent live birth.

	1: interview characteristics	*2: interviewer characteristics*	*3: respondent characteristics*	*4: full model*
Interview characteristics				
log months between birth and interview	1.27 (.13)*			1.26 (.14)*
# of birth questions with missing data	1.40 (.13)**			1.46 (.17)**
# of background questions with missing data	2.09 (.53)**			1.88 (.58)*
interview in Spanish	1.46 (.46)			3.13 (1.64)*
respondent skipped sexual intercourse calendar	1.71 (.28)**			1.73 (.33)**
Interviewer characteristics				
interviewer is same race or ethnicity as respondent		0.58 (.09)**		0.54 (.13)**
interviewer has computer experience		1.55 (.27)*		1.43 (.28)
years of interviewing experience		1.03 (.02)		1.05 (.03)
interviewer's level of education		0.95 (.06)		0.95 (.07)
age of interviewer		1.00 (.01)		1.00 (.01)
interview month (1 to 10 for Jan. to Oct.)		0.94 (.04)		0.87 (.04)**
# of interviewer questions with missing data		1.01 (.17)		0.94 (.20)

continued

Table 7 (continued): Logistic regressions on the likelihood of a discoverable error in the NSFG 1991-95 contraceptive calendars. Odds of a respondent reporting use of birth control pills, female barrier methods excluding female condoms, or injectable/morning-after hormonal contraceptives two months before her most recent live birth.

	1: interview characteristics	*2: interviewer characteristics*	*3: respondent characteristics*	*4: full model*
Respondent characteristics				
background and race				
respondent's mother's level of education			0.85 (.11)	0.89 (.12)
respondent's mother had a teen first birth			1.22 (.23)	1.20 (.23)
respondent lived in nonintact family at some time			1.52 (.28)*	1.58 (.30)*
nonhispanic white			*omitted category*	
hispanic			1.34 (.37)	0.95 (.32)
nonhispanic black			1.77 (.41)*	1.22 (.35)
other race			1.58 (1.07)	1.51 (1.08)
respondent born outside US			1.28 (.38)	0.90 (.33)
circumstances of the most recent birth				
five or more weeks premature			1.36 (.42)	1.46 (.47)
teen birth			0.74 (.25)	0.63 (.22)
respondents learned of pregnancy after 4th month			2.67 (2.05)	2.60 (2.18)
pregnancy was unintended			1.12 (.22)	1.22 (.25)
respondent breastfed the baby			0.71 (.13)	0.72 (.14)*
current status				
current level of education (1=high school diploma, 2=college)			1.00 (.14)	1.10 (.17)
current labor force status (1 = part-time, 2 = full time)			0.89 (.09)	0.84 (.09)
married at interview			0.72 (.16)	0.64 (.15)
number of children at interview			1.22 (.10)*	1.19 (.10)*
uses contraceptives in majority of calendar months			10.12 (2.31)**	11.30 (2.70)**
uses injectable or morning after contraceptives			0.44 (.13)**	0.47 (.15)*
uses female barrier methods			1.79 (.38)**	1.73 (.39)*
log likelihood	-536.7	-553.4	-460.3	-429.8

*Coefficients show the multiplicative effect of each covariate on the odds of a discoverable error. 32 cases were missing data on contraceptive use two months before most recent live birth and were dropped. Final N = 1596 *p < .05 **p<.01*

Sample includes only women who reported birth control pill, female barrier method, injectable or morning-after contraceptive use in their contraceptive calendar.

Table 8: Rates of detectable errors in NSFG 1991-95 sexual intercourse calendars.
Sexual intercourse history superimposed on birth history for women with at least one live birth.

	Type of error: No reported sexual intercourse in the three months surrounding a live birth conception
Denominator of error rate:	*Error rate:*
all respondents with one or more live births	**1.7 %** (46 / 2685)
all women with one or more live births who reported any 3-month intervals of no sexual intercourse	**5.5 %** (46 / 838)
most recent live birth of all women who reported any 3-month intervals of no sexual intercourse	**4.8 %** (40 / 838)
all live births to all women who reported any 3-month intervals of no sexual intercourse	**4.8 %** (47 / 989)

Table 9: Estimated effect of response error on a measure of association using the NSFG 1991-95 sexual intercourse calendar: Percent of women reporting at least one month of postpartum sexual abstinence, by breastfeeding status.

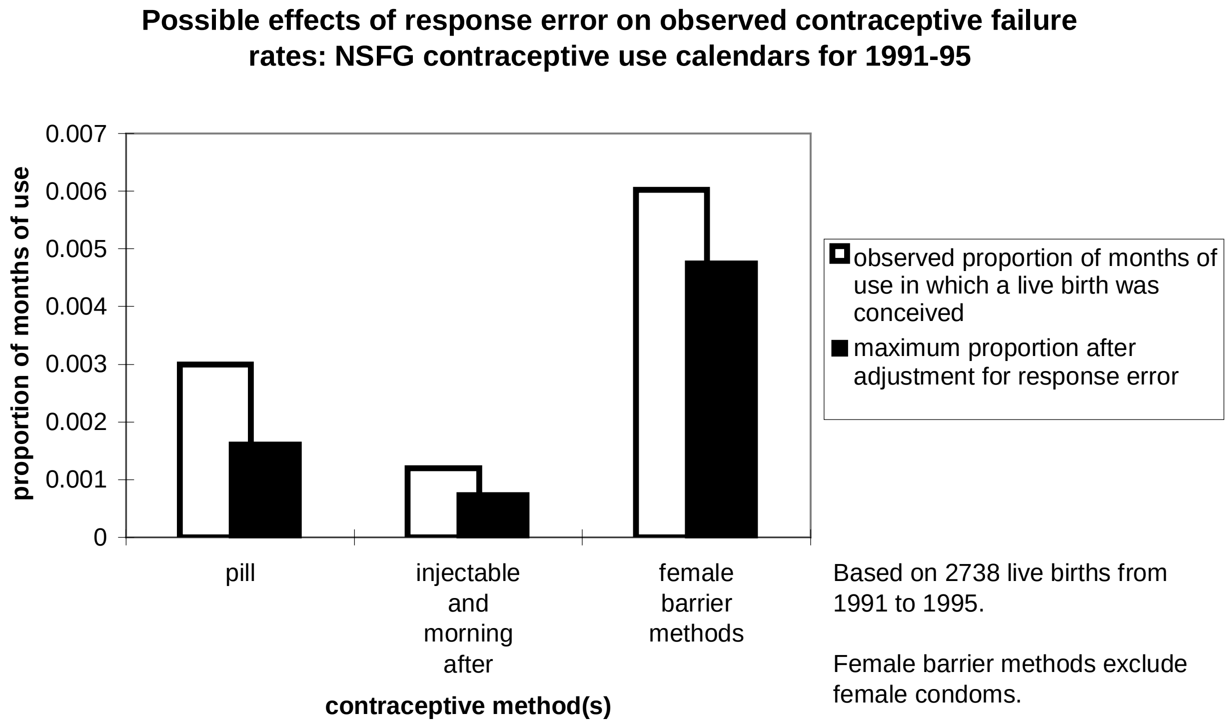
Age at first birth and data source	*Percent of women reporting one or more months of postpartum abstinence*					
	breastfeeding women		*nonbreastfeeding women*		*difference of proportions*	*N*
	est.	*s.e.*	*est.*	*s.e.*		
nonteen first birth						
Hyde et al., 1996	85 %	2 %	71 %	3 %	14 %**	528
NSFG 1991-95	38 %	1 %	27 %	1 %	10 %**	2799
teen first birth						
NSFG 1991-95	27 %	4 %	31 %	3 %	-4 %	385

* $p < .05$ ** $p < .001$ for X^2 test with 1 df
Standard errors are based on binomial distributions.

Table A1: Rates of detectable errors in the NSFG 1991-95 contraceptive use and sexual intercourse calendars. Sexual intercourse history superimposed on contraceptive history.

	Type of error: Incompatible contraceptive use during the middle month of three months of no sexual intercourse		
	All incompatible	Male condom	Female barrier methods
Denominator of error rate:	*Error rate:*		
all respondents	**7.6%** (824/10847)	**7.0%** (761/10847)	**1.1%** (117/10847)
respondents with intervals of both no sex and (condom/barrier) method use	**37.7 %** (824 / 2188)	**37.6 %** (761 / 2023)	**25.7 %** (117 / 456)
all months with (condom/barrier) method use for respondents with intervals of no sex	**10.1 %** (4543 / 44856)	**10.4 %** (4175 / 40120)	**9.4 %** (720 / 7640)
For comparison: expected error rate if reports of method use and no sex are scattered at random in their event calendars	**24.9 %** (11174 / 44856)	**25.1 %** (10069 / 40120)	**22.7 %** (1734 / 7640)

Figure 1



Based on 2738 live births from 1991 to 1995.

Female barrier methods exclude female condoms.

Figure 2

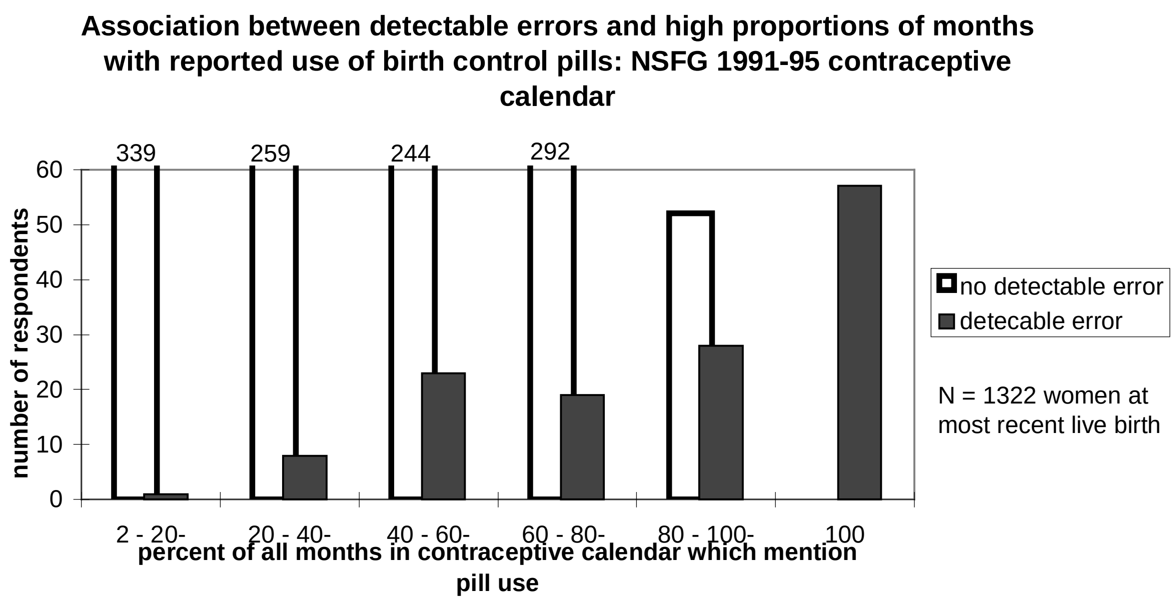
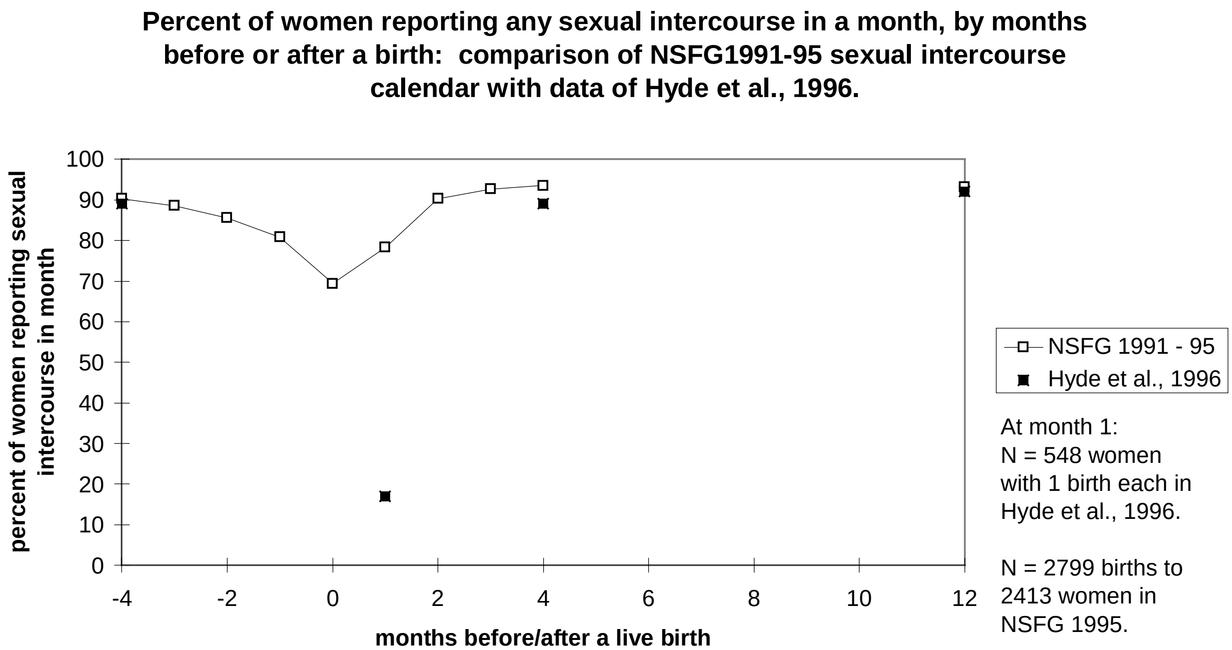

Association between detectable errors and high proportions of months with reported use of birth control pills: NSFG 1991-95 contraceptive calendar
339
259
244
292
60
50
40
30
20
10
0
number of respondents
2 - 20-
20 - 40-
40 - 60-
60 - 80-
80 - 100-
100
percent of all months in contraceptive calendar which mention pill use
no detectable error
detecable error
N = 1322 women at most recent live birth

Figure 3


Percent of women reporting any sexual intercourse in a month, by months before or after a birth: comparison of NSFG1991-95 sexual intercourse calendar with data of Hyde et al., 1996.
percent of women reporting sexual intercourse in month
100
90
80
70
60
50
40
30
20
10
0
-4
-2
0
2
4
6
8
10
12
months before/after a live birth
NSFG 1991 - 95
Hyde et al., 1996
At month 1:
N = 548 women with 1 birth each in Hyde et al., 1996.
N = 2799 births to 2413 women in NSFG 1995.

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