Recall bias in women’s and men’s reports of pregnancy intentions in the National Survey of Family Growth

Kathryn Kost and Laura Lindberg
Guttmacher Institute, 125 Maiden Lane, New York, NY 10038, USA
e-mail: kkost@guttmacher.org

Abstract
There is widespread concern that recall bias in reports of the pregnancy intentions of prior births diminishes the validity of retrospective measurements of these intentions. However, evidence of such bias has not been examined in the primary survey used to study intention status of births in the United States: the National Survey of Family Growth (NSFG).

Self-reported intention status measures for births to women and men from the 2006–2010 and 2011–2013 NSFGs were examined for statistically significant differences in the distribution of intended, mistimed, and unwanted births in successive period moving further back in time, comparing estimates for up to 10 years prior to interview. We tested also for evidence of recall bias in other measures of pregnancy desires, including pregnancy timing.

For both women and men, we find little evidence for concern that respondents’ reports of retrospective pregnancy intentions in the NSFG are strongly biased by length of recall. The distributions of recalled intentions were relatively stable as the length of time from the birth to the interview increased up to 4 years. Births to women and men 49-60 months before interview were more likely to be reported as unwanted as compared to births within one year of the interview. Thus, contrary to what is often expected, some parents may be more likely to recall an earlier birth as unwanted than wanted.

Key words: recall bias, retrospective reporting, pregnancy intentions, men’s intentions, pregnancy desires

Acknowledgments
The authors gratefully acknowledge support for this research from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health under award number R01HD068433. Additional support was provided by the Guttmacher Center for Population Research Innovation and Dissemination under NICHD award number R24HD074034.
Background

Survey questions to ascertain pregnancy intention status—whether an individual had wanted a child (or another child) just before conceiving a pregnancy and whether she/he became pregnant at the right time (for her/him)—were designed to obtain elemental information on an individual’s control of their reproductive life (Campbell and Mosher 2000). And the extent to which births are “intended” or “unintended” is a core measure of public health, used by demographers to study both population-level characteristics of fertility as well as individual-level behaviors that are both influences of and influenced by pregnancy intentions (Kost and Lindberg 2015).

The most commonly used dataset to measure pregnancy intentions in the U.S. is the National Survey of Family Growth (NSFG). Since 1973, the NSFG has enabled researchers to classify women’s pregnancies as wanted, unwanted or mistimed (Campbell and Mosher 1999). And in 2002, a separate NSFG survey program was initiated for men with generally parallel question items. As part of a series of questions to document an individual’s pregnancy history, NSFG respondents are asked to recall their feelings about having a baby right before they became pregnant—specifically, whether they had wanted a (or another) baby at any point in their future. If they had wanted no more, or any, children, the pregnancy is classified as “unwanted.” If they wanted a baby at some point in the future, the pregnancy is classified as “wanted.” Respondents with wanted pregnancies are then asked whether the pregnancy came sooner than they would like, at about the right time, or later than they would like. Pregnancies that came sooner than preferred are classified as “mistimed.” Those that came at the right time or later than wanted are characterized as “wanted” (or, “intended”).

Researchers have long been concerned that parents may report inaccurate recollection of their original pregnancy intentions, so that recall bias diminishes the validity of retrospective survey measurements of pregnancy intentions (Miller 1974; Rosenzweig and Wolpin 1993; Trussell et al. 1999; Williams et al. 1999; Santelli et al. 2003). Intention status is a subjective categorization that may be influenced by the subconscious process of recall or reassessment over time (Guzzo and Hayford 2014). One the one hand, over time, as parents have positive experiences with their child, those who had not intended the pregnancy may change the intention in their memory and report the birth as having been intended—so called ex post rationalization. Similarly, because of social desirability bias, cultural norms could inhibit parents from reporting
a child as having been unintended, and this inhibition may grow as the child ages. In both of these cases, the concern is that recall bias would result in more births being reported as intended in retrospective measures collected at the time of the interview than were intended at the time the pregnancy occurred. In contrast, it is also possible that social desirability could induce individuals to report an intended pregnancy as having been unintended if one perceives that the conditions under which the pregnancy occurred were considered culturally taboo (e.g. unmarried, too young, etc.). If pregnancy intentions reported at interview differ from those the woman or man had prior to the pregnancy, then analyses using these retrospective measures of pregnancy intentions could be biased, and not always in predictable ways.

Some studies have examined the extent of correspondence between prospectively measured childbearing intentions, or ideal family size, and later fertility (Williams and Abma 2000; Yeatman et al. 2013; Yeatman and Sennott 2015). However, future fertility preferences may not be stable, and differentials in relation to later fertility may represent actual changes in preferences and not recall bias (Jones et al. 2015).

An alternate approach to testing the reliability of reports of intendedness is to collect this information about the same birth at multiple points in time. Studies of this sort in the United States have found that aggregate-level estimates of intention status are relatively stable across repeated measurement, although some women do indeed change their intention status reports. Joyce et al. (2002) found that when pregnant women were asked their intention status during pregnancy and again two years later, 84% of women reported the same intention status. Similarly, Poole et al. (2000) found that when pregnant women were asked their pregnancy intentions twice during the pregnancy, 77% of the women did not change their responses. And Guzzo and Hayford (2014) found that 88% of young women did not change the intendedness categorization of their first birth when asked about the same birth in two rounds of the National Longitudinal Study of Adolescent Health collected six to seven years apart. However, studies using panel data from Morocco and India did find substantial shifts in reporting of intention status across two interviews (Bankole and Westoff 1998; Koenig et al., 2006). In most of these studies, women were more likely to shift their reports from unintended to intended than vice versa. But this was not uniform; Guzzo and Hayford (2014) found that women were more likely to switch from reporting a birth as intended to unintended than in the other direction.
Capitalizing on natural variation in the spacing of interviews in the Moroccan data, Bankole and Westoff (1998) directly tested the hypothesis that the extent of recall bias varies with the length of the recall period. They found that over the three-year period of study, women were more likely to change their report of intention status, as more time had passed. Similarly, Koenig et al. (2006) found that the older the age of the child (and thus the longer the length of the retrospective recall period), the more likely women who had prospectively reported a pregnancy as unwanted were to retrospectively report the birth as intended.

To date, we are not aware of any studies of retrospective recall bias in men’s reported pregnancy intentions. Indeed, research on men’s pregnancy intentions in the United States has only begun in earnest in recent years, with the collection of relevant data directly from men (Augustine et al. 2009; Bronte-Tinkew et al. 2009; Edin and Nelson 2013; Johnson and Williams 2005; Jackson et al. 2011; Lindberg and Kost 2013). Analyses of the recently collected NSFG data from men finds that nearly four out of ten of births to men were reported as unintended, with significant variation by men’s demographic traits; these patterns are similar to those estimated for women (Lindberg and Kost 2013). Data quality concerns have focused on men’s underreporting of births more generally (Joyner et al. 2012; Martinez et al. 2006; Lindberg et al. 1998), and not the specific issue of recall bias in measuring intention status. However, men’s reports may have recall bias for reasons similar to women; additionally, their gendered experience of parenthood, such as being less likely than mothers to coreside with their child (Vespa et al. 2013) or to engage in daily care activities (Casper and Bianchi 2001), may further influence reporting patterns. This analysis fills an important gap by separately examining recall bias in the reporting of pregnancy intentions for both women and men.

Much of the research on unintended childbearing in the U.S. relies on data from the National Survey of Family Growth (NSFG), a cross-sectional survey which asks women and men to report retrospectively on the intention status of each pregnancy, as well as on other attitudinal measures of pregnancy desires. If pregnancy intention status reports in the NSFG are biased, researchers using these data should be concerned about the potential for such bias to compromise or even distort findings using the pregnancy intention measures in the NSFG. However, evidence of pregnancy intention recall bias in the NSFG has not been directly examined, either for women’s or men’s reports. In this study, we examined evidence of recall bias for women and men (separately), combining data on births from the 2006–2010 and 2011–
2013 National Survey of Family Growth (NSFG). We test whether respondent’s retrospective reports of pregnancy intention appear to be influenced by recall bias, by estimating differences in these measures by periods of recall up to 10 years. In addition, we test for recall bias in retrospective reports of other measures of pregnancy desires beyond the conventional measure of pregnancy intention status. Finally, we exploit additional measures available only for women and provide, for the first time, an examination of recall bias in retrospective reports of the extent of pregnancy mistiming.

**Data and Methods**

The National Survey of Family Growth (NSFG), conducted by the National Center for Health Statistics (NCHS), is a multi-stage, stratified, national probability survey of the noninstitutionalized population aged 15–44 in the United States (Lepkowski et al. 2013; Groves et al. 2009). The 2006–2010 survey contains a sample of 12,270 women aged 15–44 and a sample of 10,403 men aged 15–44 (NCHS 2011), and the 2011–2013 NSFG includes a sample of 5,601 women aged 15–44 and a sample of 4,815 men aged 15–44 (NCHS 2014). The women’s and men’s samples were collected separately; in other words, the men interviewed have no relationship with the women. Methods of data collection and dissemination of the public-use dataset are reviewed by the Institutional Review Board at the National Center for Health Statistics. Further information about the design of the NSFG is available at [http://www.cdc.gov/nchs/nsfg.htm](http://www.cdc.gov/nchs/nsfg.htm).

Both the 2006–2010 and 2011–2013 surveys include complete birth histories for each woman interviewed. The women’s pooled sample from the 2006–2010 and 2011–2013 NSFGs used for this analysis contained 20,503 non-multiple live births. We limited analysis to births within ten years of the woman’s date of interview, for a final analytic sample of 12,626 non-multiple live births. Because the NSFG is a cross-sectional survey, births beyond ten years could only be reported by the oldest women in the sample; that is, those aged roughly 25 to 44 years, with a greater likelihood of being among the older mothers in this age range. Limiting the analysis to births in the past 10 years therefore limits the extent to which the sample of births available for the longest durations is skewed toward those with the oldest mothers and therefore not strictly comparable to births occurring in the year before the survey.
In both the 2006–2010 and 2011–2013 men’s surveys, respondents were asked to report information about each biological child; men were not asked to report pregnancies ending in abortion or miscarriage. Detailed information on births, including pregnancy intention status, was not collected for births occurring more than five years prior to interview. In addition, unlike the women’s data, there is no pregnancy or birth file for men. We therefore created a birth-level file similar to that provided for women, and, like women, men could contribute more than one birth to this file. The men’s pooled sample from the 2006–2010 and 2011–2013 NSFGs used for this analysis contained 3,993 non-multiple live births of men’s biological children.

For our analyses of the women’s data, we used a conventional pregnancy intention status measure based on a series of questions to assess their feelings right before they became pregnant; these questions are used to classify each birth as intended (wanted and on time or later than wanted), mistimed by less than two years (wanted but occurring up to 23 months sooner than desired), mistimed by two or more years (wanted but occurring more than 24 months sooner than desired), or unwanted (see Appendix for questionnaire items). Men are not asked the follow-up question on extent of mistiming, so our categories for pregnancy intention for men are limited to intended, mistimed and unwanted.

We also examined three questions measured on a Likert-type scale to assess how happy the mother had been once she discovered the pregnancy (1 to 10), how much she had wanted to avoid or have a pregnancy (0 to 10), and how much she had been trying to avoid or become

* Interviewers were instructed to limit collection of information on births to a period of time starting with January of the year five calendar years prior to interview. Thus, some men contributed up to almost six years worth of birth history (e.g., if interviewed in December). We therefore limited the analytical sample to births occurring within 60 months of interview.

† In the NSFG, very few respondents used “Not sure/Don’t know (n=11 births to women and n=7 births to men). They are coded as missing. Those responding “Didn’t care” can be classified as wanted, unwanted or missing, depending upon the focus of the analysis. For this analysis, they are coded as wanted/intended (n=34 births to women and n=52 births to men).
pregnant (0 to 10).‡ These can also be considered measures of pregnancy desire and capture other attributes of pregnancy intentions beyond what is measured with the conventional measure (see Appendix for specific question wording) (Santelli et al. 2009). Male respondents were only asked the Likert-scale question on happiness. Among men not married or not living with the baby’s mother at the time of the birth, both pregnancy intentions and happiness were only measured among those reporting that they found out about the pregnancy before the child was born.§

For women, the Likert-scale questions were asked only of births in the three years prior to interview, so our tests for evidence of recall bias in these measures are limited to these births.** We examined distributions of these additional measures of pregnancy intention by time since interview, measured in months and recoded into three categories: 0–12 months, 13–24 months, and 25–36 months. For men, the happiness scale question was asked of all births that occurred in the five years prior to interview; our analysis for men therefore includes lengths of recall for the following five categories: 0–12 months, 13–24 months, 25–36 months, 37–48 months and 49–60 months.

‡ These questions were based on the psychosocial theories of Warren Miller, which posit multiple dimensions of pregnancy attitudes that affect motivations and behavior before, during, and after pregnancy (Miller 1992, 1994, 1998; Miller and Pasta 2002; Miller et al. 2004).

§ These men were asked “When did you find out that (partner) was pregnant? Was it during the pregnancy or after the child was born?” Births to men who had been unaware of the pregnancy until after the baby was born are excluded from our analyses (n=84).

** Similarly to men, women’s reports were limited to births occurring from January of the calendar year three years prior to interview to the interview date. Again, some respondents would therefore be able to contribute information for a period of up to 47 months prior to interview. We therefore limited analysis to births occurring within 36 months of interview.
We examined distributions of the intention status of births for all births within 10 years of the mothers’ interviews and five years of the father’s interviews to examine whether there was any indication that “older” births (i.e. further back from interview) were more or less likely to be reported as intended than those closer to the interview date. We examined distributions of these measures of pregnancy intention by time since interview, measured in months and recoded into seven categories: 0–12 months, 13–24 months, 25–36 months, 37–48 months, 49–60 months, 61–72 months, and 73–120 months. For the men’s births, analysis of the conventional pregnancy intention status measure is limited to the first five categories. To examine whether births further back in time are more or less likely to be reported as intended, mistimed or unwanted than those that occurred closer to the date of interview, we tested for statistically significant differences in the proportions reported between each successive time period category and the most recent one (i.e. 0–12 months since interview). Given little change in patterns of intention status overall during the period under study (Finer and Zolna 2014), any finding of a significant difference in intention status by time since interview would be evidence of recall bias in the measure.

We tested for potential autocorrelation among births with the same mother or father by including the respondent’s unique identification code as a cluster indicator in complex survey design commands, in addition to the other design variables specified for use with the NSFG. However, estimated variances with specification of this cluster were virtually identical to those without. All analyses are weighted by the population weights provided in the NSFG data sets, and we used complex sample design commands in Stata 14.0 (Stata Statistical Software 2015). Statistically significant differences in proportions (intention status groups and categories of length of mistiming) and in means (scaled measures) were evaluated using t-tests. Only those with \( p \) values less than or equal to .05 were considered significant.

**Findings**

Figure 1a shows the proportions of births in each of the four intention status groups as reported by women by time since interview. We found no evidence of a bias in the reports of births intended, mistimed by less than two years or mistimed by two years or more as the time since

---

\( ^{\dagger\dagger} \) We followed a method described by Lepkowski et al. (2013) to include multistage sample designs in complex surveys.
interview increased. There were no statistically significant differences in the estimates. Births that occurred soon after interview — within one year — were reported to be intended as frequently as were births that occurred within two, three, five, four, five, six or more years after interview. The same holds true for mistimed births — whether mistimed “slightly” (by less than two years) or “greatly” (by two or more years). For unwanted births, the proportion reported as unwanted for births four to five years prior to the survey (17%) was significantly higher than the proportion of births in the year just before interview (12%), with a p value=.031. However, there was no significant difference for birth born six or more years after the interview and the earliest period.

Figure 1b presents the proportions reporting intended, mistimed and unwanted births among men in the five years prior to interview. Similarly to women, there is no evidence of recall bias in the proportions intended or mistimed. And, as we found for women’s reports, the proportion reporting a birth occurring between 49 and 60 months prior to interview as unwanted is higher (13%) than the proportion reporting a more recent birth as unwanted (8%). This difference was statistically significant at p=.039.

Figure 2 shows the mean values of the happiness, wanting and trying scales. Again, only the happiness measure was included in the men’s NSFG questionnaire, but the question was asked for births across a longer time period (5 years) than for women’s (3 years). These scales also show very little variation across time. In fact, while men tend to report higher levels of happiness than women for all births, as noted in earlier research (Lindberg and Kost 2013) the estimated values for both men and women show virtually no change in this measure over time. The degree to which women report having wanted the pregnancy or having tried to get pregnant appears to increase slightly after two years, but differences in these estimates are not statistically significant, even though large sample sizes available for these analyses should have been sufficient to capture statistically significant effects.

Finally, we looked at retrospective reporting of the extent of mistiming among mistimed births by time since interview for women (Figure 3). Among all mistimed births, mothers of those born close in time to the interview report similar numbers of months they would have preferred to wait (33.2) to mothers of births born between 13 and 72 months too early (ranging from a preference of 31.4 to 34.7 months). Mothers reported wanting to wait slightly longer for
births seven to ten years before the survey (36.4), but this and all estimates were not significantly different from the timing preference given for births born within a year of interview.

We also looked at the amount of time women retrospectively reported they had wanted to wait until having a birth among those who reported the birth as only slightly mistimed (less than two years) and among those who reported the birth as greatly mistimed (two years or more). Mirroring findings for all mistimed births, there was no association of reported preference for how long the mother would have preferred to wait before having the birth with the length of time that had passed since the birth occurred for either of these two groups of mistimed births (Figure 3).

**Discussion**

We find no clear evidence that reports of pregnancy intentions in the NSFG survey were affected by the length of time since the birth for either women or men. Among women, the proportions reporting past births as having been intended, mistimed by less than two years, or mistimed by two years or more were relatively stable across births up to 10 years prior to the survey interview. Similarly, women who reported that the birth had been mistimed did not seem to have difficulty recalling how long they had wanted to wait, even if the birth had occurred ten years prior to interview. And this was true of both births that had been slightly mistimed — with women reporting on average that they had occurred about 9.5 months too soon — as well as of births that had been greatly mistimed (on average, occurring just less than four years too soon). Similarly, men’s reports of intended and mistimed births also do not differ by length of recall, although we are limited to births in a period up to five years prior to interview for the men’s data. Whether longer periods of recall would have reflected bias is still unknown.

We also found that men’s and women’s recall of how happy they had been to discover the pregnancy did not vary by the length of time since the birth. For women, this measure is collected only for births occurring within three years of the interview, while the men’s measure extends to five years prior to the interview; for both genders, it is still unknown if recollections of births occurring further back in time would have differed.

We also found that women’s reports of how much they had wanted to have a baby and how much they had been trying to have a baby also did not differ by time since interview. These findings suggest that women have relatively accurate recollections of their childbearing desires
and behaviors. While this is perhaps not surprising for a period as short as three years prior to interview, it validates the available NSFG data and contributes to the value of future research using these retrospectively reported measures.

Finally, we did find that for both men and women, there were slightly more births reported as unwanted for those born four to five years prior to the survey interview than for those born in the year prior to the interview. This finding is surprising given than others have hypothesized the opposite: that unwanted births would be more likely to be reported as wanted with the passage of time in response to social desirability or a response to positive parenting experiences. Instead, this bias towards reporting unwantedness may suggest negative parenting experiences over time for both genders. More research is needed to consider these mechanisms and their implications for understanding the causes and consequences of unwanted fertility (Kost and Lindberg 2015; Lindberg et al. 2014).

Other large-scale surveys such as the Pregnancy Risk Assessment Monitoring System (PRAMS) and the Demographic and Health Surveys (DHS) include slightly different wording for pregnancy intention questions, but the same categories of wanted, mistimed, or unwanted have generally been adopted. If retrospective reporting of intentions is affected by social and cultural factors, there may still be evidence of recall bias in other settings. Our analysis could be extended to these other settings to explore evidence of such bias in other surveys.

Conclusion

Our findings for the U.S. National Survey of Family Growth should provide confidence to other researchers focused on capturing pregnancy intention status in the nation’s premier fertility survey. We found very little evidence of bias related to the length of time for retrospective recall, which suggests that women and men are able to accurately recall and report whether they had wanted have a baby before the infant was conceived. The similar patterns by gender should also assuage concerns that men are less reliable reporters of fertility-related data. Overall, these are important findings for those performing analyses using the pregnancy intention measures in the NSFG.
References


Conventional Measure of Intention Status:

For women: Right before you became pregnant with your (Nth) pregnancy (which ended in (date), did you yourself want to have another baby at any time in the future?

- Yes
- No
- Not sure, don’t know
- Didn’t care

Check question if answered No:
So right before you became pregnant (this time), you thought you did not want to have any children at any time in the future, is that correct?

- Correct
- Incorrect

For men: Please look at Card 58. Right before (MOTHER OF BIOLOGICAL CHILD) became pregnant with (BIOLOGICAL CHILD), did you, yourself, want to have a child at some time in the future? (Applicable if this biological child is 18 years or younger and respondent was living with the child's mother at time of birth or knew about the pregnancy before the birth.)

- Definitely Yes
- Probably Yes
- Probably No
- Definitely No
- Inapplicable
- Don’t Know

A men’s pregnancy intention status variable comparable to the one for women was constructed using this question and the following. The third and fourth “no” categories were coded as unwanted births. Intended and mistimed births were determined using the following question (rules for construction are specified in recoded variables documentation, available at http://www.icpsr.umich.edu/icpsradmin/nsfg/variable/recode_spec/cycle8/male/WANTB01.pdf):

Would you say that the pregnancy came sooner than you wanted, at about the right time, or later than you wanted? [BIOLOGICAL CHILD] (Applicable if respondent responded he "definitely" or "probably" wanted to have a child at the time the child's mother became pregnant)

- Too Soon
- Right time
Timing question (asked only of women respondents):
So would you say you became pregnant too soon, at about the right time, or later than you wanted?
- Too soon
- Right time
- Later
- Didn’t care

If too soon:
How much sooner than you wanted did you become pregnant?
- Enter MONTHS or YEARS

Happiness question, for women:
On this scale, a one means that you were very unhappy to be pregnant and a ten means that you were very happy to be pregnant. Tell me which number on the card best describes how you felt when you found out you were pregnant. (Subject was shown scale cards.)

Happiness question, for men:
Please look at Card 59. On this scale, a zero means that you were very unhappy about this pregnancy, and a ten means that you were very happy about this pregnancy. Tell me which number on the card best describes how you felt when you found out that (MOTHER OF BIOLOGICAL CHILD) was pregnant this time. (Subject was shown scale cards.)

(Applicable if respondent reported a biological child anywhere in the interview, child is 18 years or younger, and respondent was married to or living with child's mother at time of birth or learned of pregnancy before the child was born.)

Wanting question, for women:
0 means you wanted to avoid a pregnancy and a 10 means you wanted to get pregnant. If you had to rate how much you wanted or didn’t want a pregnancy right before you got pregnant (that time), how would you rate yourself? (Subject was shown scale cards.)

Trying question, for women:
0 means trying hard not to get pregnant, and a 10 means trying hard to get pregnant. If you had to rate how much you were trying to get pregnant or avoid pregnancy right before you got pregnant (that time), how would you rate yourself? (Subject was shown scale cards.)
Figure 1a. Women's intention status reports at interview for non-multiple live births by time since birth; pooled data from NSFG 2006-2010 and NSFG 2011-2013.
Figure 1b. Men's intention status reports at interview for non-multiple live births by time since birth; pooled data from NSFG 2006-2010 and NSFG 2011-2013.
Figure 2. Men's and women's mean value on attitude scales for non-multiple live births by time since birth; pooled data from NSFG 2006-2010 and NSFG 2011-2013.
Figure 3. Mean number of months mothers reported wanting to have waited before becoming pregnant for mistimed non-multiple live births by time since birth; pooled data from NSFG 2006-2010 and NSFG 2011-2013.

Mean months mistimed for different time periods:
- Mistimed two or more years: 45.7, 47.6, 47.3, 50.7, 47.9, 48.9, 48.7
- All mistimed: 33.2, 34.7, 33.2, 31.4, 33.1, 34.3, 36.4
- Mistimed less than two years: 9.4, 8.5, 9.0, 9.8, 9.5, 9.6, 9.8

Length of recall period in months:
- <13 months
- 13-24 months
- 25-36 months
- 37-48 months
- 49-60 months
- 61-72 months
- 73-120 months