



Published in final edited form as:

Stud Fam Plann. 2012 September ; 43(3): 213–222.

The Reliability of Calendar Data for Reporting Contraceptive Use: Evidence from Rural Bangladesh

Rebecca L. Callahan [doctoral candidate] and Stan Becker [Professor]

Department of Population, Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, 4th floor, Baltimore, MD 21205

Abstract

Collecting contraceptive-use data by means of calendar methods has become standard practice in large-scale population surveys, yet the reliability of these methods for capturing accurate contraceptive histories over time remains largely unknown. Using data from overlapping contraceptive calendars included in a longitudinal study of 3,080 rural Bangladeshi women, we assessed the consistency of reports from the baseline interview month in 2006 with reports from the same month in a follow-up survey three years later, and examined predictors of reliable reporting. More than one-third of women were discordant in their reports for the reference month in the two surveys. Among women reporting use of any contraceptive method for the reference month in both surveys, 25 percent reported different methods at the two time points. Women using condoms or traditional methods and those with more complex reproductive histories, including more births and more episodes of contraceptive use, were least likely to report reliably.

Retrospective reports in surveys are subject to recall bias, which can affect the quality and usefulness of the resulting data. Concern about the validity of retrospective data is especially strong in the social science and public health fields, in which many studies depend on autobiographical memory to reconstruct life histories. Such concern has led to the development of calendar and timeline methods of data collection that assist respondents in recalling information by placing events in chronological sequence and offering “event cues” (Glasner and van der Vaart 2009). Few studies, however, have assessed the reliability of reports of contraceptive use collected with calendar methods, particularly the reliability among different populations, recall periods, and contraceptive methods. This study expands the literature by examining the consistency of reports regarding contraceptive use among women in rural Bangladesh.

Theoretical Framework

Psychological studies have shown that autobiographical memories are temporally and thematically structured within a hierarchical ordering consisting of extended, summarized, and specific events. As Belli (1998) explains, “extended events” at the top of the hierarchy provide the foundation for autobiographical memories and generally span an extended time period ranging from days to years. Extended events have identifiable starting and end points, such as a period of residence in a particular place or the duration of a relationship. Within extended events, “summarized events” include general events that represent memories of two or more similar events remembered as composites, wherein the specifics of any one event may be lost. Examples include lunches with colleagues (summarized event) during a specific episode of employment (extended event) and, in terms of family planning, use of a contraceptive method before the birth of a particular child. Finally, the lowest level of the

memory hierarchy includes “specific events,” which include rich contextual and perceptual information. Specific events are nested within summarized and extended events and represent specific occasions such as one’s wedding day or the birth of a child.

Questionnaire designs that make use of this hierarchy and allow respondents both “top down” and “parallel” retrieval of information have the potential to elicit more accurate life history data (Belli 1998; van der Vaart 2004). Top down retrieval of memories involves using thematic and temporal information of higher-order memory structures to encourage the remembrance of more specific events. Parallel retrieval also uses the hierarchy of memory but encourages respondents to link interconnected aspects of one’s past. Research on autobiographical memory has produced conflicting results regarding whether experiences are recalled more accurately if individuals are required to begin with the earliest events in a series and work forward, to use backward retrieval beginning with the most recent events, or to use no temporal ordering (Bradburn, Rips, and Shevell 1987; Jobe et al. 1990; Loftus et al. 1992). Unlike traditional surveys that gather retrospective data through the use of discrete questions about specific events generally separated by topic, life history and calendar methods use the inherent hierarchy of memory and encourage participants to report on specific events by asking about incidents and placing them in a broader temporal and thematic stream. Calendars provide a graphic display that allows respondents to more easily sequence events, whether through forward or backward retrieval, and to cross-reference between memory domains and reference points (Glasner and van der Vaart 2009).

Accuracy in reporting autobiographical events is also dependent on factors such as the salience, frequency, similarity, and regularity of events, as well as the length of the retention interval (Brewer 1986; Belli 1998). Additionally, autobiographical memory declines with age and is positively associated with education level (Levine et al. 2002; Piolino et al. 2006; St. Jacques and Levine 2007; Christensen et al. 2008; Gylmour et al. 2008; Angel et al. 2010). In general, vivid and salient events and those that are similar to each other and are experienced at regular intervals are more easily recalled over a longer period than less noteworthy events and those experienced less frequently or at irregular intervals (Brewer 1986; Menon 1993; Belli 1998). Therefore, respondents are more likely to accurately report salient events such as marriage, childbirth, or timing of sterilization than more mundane events. Similarly, regular and frequent contraceptive use such as daily pill use, regular injections, or consistent use of an IUD could be expected to be more accurately recalled than coitally dependent methods such as condom use and withdrawal, which are practiced more sporadically and infrequently. Research has indicated that calendar methods are more effective than traditional survey questions when the recall task is more difficult—that is, when respondents are asked to report on less salient events and incidents that occurred further in the past (van der Vaart 2004; van der Vaart and Glasner 2007). Calendar methods also have the potential to reduce social desirability bias and lead to more truthful reporting, as evidenced by a recent study comparing data concerning sexual behavior (including contraceptive use) collected using a life history calendar with similar data collected using standard survey questions (Luke, Clark, and Zulu 2011).

Origin and Current Use of Contraceptive Calendar

As understanding of memory structure and retrieval has improved, use of calendar techniques has become more widespread in social science and health behavior research. Calendar instruments and terminology have not been standardized, however, and study of their effectiveness in overcoming recall bias has been limited (Glasner and van der Vaart 2009). Use of calendar methods for querying respondents about contraceptive experience has become standard practice in large population surveys, including the National Survey of Family Growth (NSFG) in the United States and the Demographic and Health Survey (DHS)

in developing countries. The reliability of these methods for capturing accurate contraceptive histories over time, however, remains largely unknown.

The first use of a calendar in the reproductive health field appears to have been the coding of contraceptive methods in the US-based National Fertility Survey of 1965 (OPR 2012). Subsequent National Fertility Surveys in 1970 and 1975 and later the NSFG employed three- and five-year month-to-month calendars to gather fertility-exposure information (Goldman, Moreno, and Westoff 1989). Surveys in Latin America dating back to the late 1960s used a 12-month “sexual activity table” to gather information concerning family planning (Gaslonde and Carrasco 1982), and one of the earliest uses of a 30-month calendar, similar to the current version of the tool, was developed and used in the 1978 and 1980 Community Outreach Surveys in the Philippines (Laing 1984). The DHS began to incorporate a calendar module into selected surveys in the mid-1980s (Macro International 1990).

The DHS generally includes the calendar module in countries with relatively high contraceptive prevalence. The module included in the current round of the DHS (Phase 6, 2008–13) collects information concerning pregnancies and births, contraceptive use, and reasons for contraceptive discontinuation. The calendar included in earlier rounds, however, also gathered information regarding sexual unions, breastfeeding experience and duration, amenorrhea status, and sexual abstinence (Macro International 1990). Data collected via the calendar are widely used in analyses of contraceptive-use dynamics, including discontinuation, as well as of abortion, breastfeeding, and abstinence trends (Becker and Ahmed 2001; Steele, Goldstein, and Browne 2004; Hossain 2005; Baschieri and Hinde 2007; Creanga et al. 2007; Bradley, Schwendt, and Khan 2009; Ali and Cleland 2010).

Previous Assessments of Calendar Data

A small number of studies have attempted to assess the quality of data concerning the practice of family planning that was collected using calendar methods. As part of the 1986 Demographic and Health Surveys of Peru and the Dominican Republic, experimental field evaluations of a calendar data collection tool were conducted comparing the calendar with the standard questionnaire used at the time for collecting data on the proximate determinants of fertility, including contraceptive use (Goldman, Moreno, and Westoff 1989; Westoff, Goldman, and Moreno 1990). Results from these two randomized experimental studies indicated that the standard questionnaire and experimental calendar produced similar estimates of ever and current use of contraceptives. In Peru, estimates from the calendar more accurately captured duration of past contraceptive use than did the standard questionnaire when compared with actual contraceptive use data from the 1981 Contraceptive Prevalence Survey (CPS). Estimates of the overall contraceptive prevalence rate (CPR) for the period five years prior to the survey were lower, however, for both the standard questions and the calendar, compared with the CPR from the 1981 CPS. In the Dominican Republic, the experimental calendar reports (compared with the standard questions) more closely matched the previous 1983 CPS data in terms of prior duration of individual method use and overall past prevalence. The variation in results between the two countries might be explained by the shorter interval of reporting in the Dominican Republic calendar (three years, versus five in Peru) and the heavy reliance on traditional methods in Peru, which are generally reported less completely than are modern methods. In both countries, the standard version produced data with considerably more heaping of reported duration of use at annual intervals, resulting in a longer mean duration of use in the aggregate. Becker and Diop-Sidibé (2003) observed similar reductions in heaped responses in the data concerning the duration of contraceptive use collected with a calendar, compared

with single questions in the body of the questionnaire in subsequent DHS surveys in five countries.

The calendar was also evaluated as part of the 1986 survey of Maternal and Child Health and Family Planning in Costa Rica (Becker and Sosa 1992). Results of the randomized study showed that compared with traditional questions, the calendar resulted in less overlap of incongruous contraceptive-use episodes and pregnancies. The calendar also captured more pregnancy losses, more contraceptive-use events, and a higher proportion of breastfed infants.

Another assessment of the quality of contraceptive use data collected using a calendar was performed with longitudinal data from the 1995 Morocco DHS panel study (Strickler et al. 1997). Matched data from 1,694 ever-married women who were interviewed in both the 1992 and 1995 DHS were included in the analysis. The calendar study sample consisted of 61 percent of the ever-married respondents from 1992. Urban and older women were more likely to be lost to follow-up. The marginal distribution of contraceptive-use status (using, not using, or pregnant) reported in the 1995 calendar for the 1992 reference month was fairly consistent with the 1992 reports. At the individual level, however, 17 percent of women reported a different status for the 1992 reference month in the two surveys (Kappa statistic = 0.72). Consistency in reporting was highest among women reporting sterilization and oral contraceptive use in 1992 and lowest among those reporting condom use. The authors concluded that calendar data are fairly reliable in the aggregate and that differences in reporting do not appear to affect contraceptive-prevalence estimates. Reports at the individual level are less reliable, however, especially information related to reasons for discontinuation.

Although the calendar module appeared to perform fairly well in capturing prior contraceptive-use behavior and patterns in the aggregate in the studies described above, questions remain concerning its accuracy and reliability among different populations, over varying recall periods, and for long-term and permanent methods versus temporary, short-term methods. All calendar reliability studies undertaken thus far have been in contexts with relatively high levels of schooling among girls and women. Nevertheless, the DHS currently uses the calendar in a large number of countries with low female literacy. Similarly, the relative effectiveness of the methodology among respondents of different ages and socioeconomic backgrounds and of urban versus rural residence remains unanswered. Only the Moroccan study examined the reliability of calendar data with regard to demographic characteristics. Of the three studies that analyzed the reliability of calendar data, the evaluations in Morocco and the Dominican Republic included a recall period of three years, whereas the Peru study compared results to five-year-old reports. Notably, the Peru study showed lower consistency in reporting than the other studies in terms of overall contraceptive prevalence. These three studies also found variation in the reliability of the calendar data by contraceptive method and, in the case of Morocco, in pattern of method use. Long-term method use, especially sterilization, was reported much more accurately in the calendar than was use of temporary and traditional methods. These results support the idea that consistent behaviors, such as regular contraceptive injections, and specific events, such as being sterilized, are more easily recalled than are more sporadic family planning behaviors such as periodic abstinence and condom use. The results also indicate that the calendar may perform better or worse depending on the contraceptive-method mix in a country.

The present study takes advantage of a unique longitudinal survey among women in rural Bangladesh that includes a contraceptive calendar at two time points during a three-year period. The calendar in the follow-up survey covers the period between the two surveys,

including the baseline interview month, making it possible to compare women's reports of contraceptive use and pregnancy and birth outcomes provided at two different time points for a particular reference month. In this analysis, we describe the overall and method-specific concordance of reports in data collected with a calendar in the follow-up survey with that collected during the baseline interview month. We also explore predictors of reliable reporting at the individual level. We hypothesize that (1) younger age and more schooling are associated with more reliable reports; (2) women with more complex histories (those who have used more methods in the past and/or have had more births) are less likely to report reliably than women with less complex histories; and (3) use of long-term methods is positively associated with reliable reporting.

Data and Methods

The data used in this analysis are drawn from two household surveys carried out in 2006 and 2009 in 128 villages in three of the six divisions of Bangladesh (Chittagong, Dhaka, and Rajshahi). The questionnaires collected baseline and follow-up data for an experimental project designed to assess the relative effects of separately and jointly introducing additional micro-credit and essential-health-services interventions on the use of health services, economic well-being, and women's empowerment. The baseline questionnaire was completed in 2006 by 3,933 currently married women; 3,687 (94 percent of the original sample) completed the follow-up questionnaire three years later. The response rate for the follow-up interviews is unusually high because we instituted tracking of households and of women who had moved after 2006.¹ To be included in our sample, respondents needed to be less than 50 years of age at baseline (508 of those interviewed were not), to have completed the follow-up survey (190 did not), and to be less than 50 years of age at follow-up (155 were not). Thus, our final sample consisted of 3,080 women (3,933 – 853). Additional details of the study design and survey sampling are provided elsewhere (Amin, Shah, and Becker 2010).²

Both surveys included socioeconomic, demographic, and maternal and child health questions similar to those in the DHS. Women were asked about their knowledge of contraceptive methods and whether they had ever used a method. Women who were married and not pregnant at the time of the survey were asked whether they were currently using any method and, if so, which one. The surveys also included a calendar in which interviewers recorded monthly data on pregnancy and contraceptive use history, source of contraceptives, and marital status for the 40 months prior to the baseline interview and the 43 months prior to the follow-up. The first column in the calendar had 15 possible response categories, including pregnancy, birth, termination, hysterectomy, and specific contraceptive methods. A zero was added for every month in the calendar during which a woman was not pregnant, did not have a birth outcome, and was not using a method. For each episode of contraceptive use, beginning with the most recent, women were asked when they started using the method,

¹When an interviewer discovered that a household no longer resided in the village, he/she asked neighbors whether they knew the whereabouts of the household and, more specifically, whether they could provide a cell phone number for them. Similarly, in the instance where a household was present but the previously interviewed woman was no longer residing there, the interviewer asked for contact information for the woman. Though households and women who had migrated outside Bangladesh were not tracked, attempts were made to locate and interview those women living within a reasonable distance of the sample villages. Specifically, a tracking team was organized at the end of the second month of interviews and was assigned to track cases that teams had listed. In this way, the tracking team successfully tracked 79 women, and an additional 21 women were located in their original village (these women had been away from the village at the time the interview team visited).

²Prior to the baseline survey, a census was conducted in all 128 villages to categorize the households into three strata: (1) those not eligible for microcredit, (2) those eligible and who had accessed microcredit, and (3) those eligible but who had not accessed microcredit. For the survey, a stratified random sample was taken with these three strata in each village among all households in which an ever-married woman resides. The sample sizes chosen were 4, 12, and 15 women from strata 1, 2, and 3, respectively. From the sample and census information and the interview response rates, sampling weights were derived for each household and woman. The sampling weights are used in the present analysis where noted.

how long they had used it continuously, and where they obtained it. Pregnancies and births were used as reference points. For example, women were asked, “How long after the birth of (name) did you begin using the method?” Women were also asked where and/or from whom they obtained each of the methods they used in the past, which helped them remember the context in which they began using. The calendar data do not show evidence of heaping, indicating that most women were apparently able to identify when they started and stopped using a method. The calendar was completed by all women younger than age 50 at the time of each interview.

Measuring Reliability

Although we are unable to assess the validity of calendar reports, we assume that a woman’s report for the month of interview (in this instance the baseline interview month in 2006) is the best approximation of the truth because she is reporting on her current status.³ The two calendars overlap for a period of three to five months, depending on the dates of the interviews. To assess reliability, we first performed a simple cross tabulation to compare the reports for the month of interview in 2006 with the calendar reports for the same month in the follow-up survey.⁴ We calculated a Kappa statistic measuring overall and method-specific concordance of reports between the two surveys.⁵ We then widened the comparison window in the follow-up survey and compared concordance of the baseline calendar report for the month of interview in a five-month window centered on the baseline interview month in the follow-up calendar, spanning the two months prior and the two months following the baseline-interview month.

Predictors of reliable reporting were explored using logistic regression in which the outcome variable was set to zero if the calendar responses in the two surveys were discordant, and set to one if concordant. Model covariates included age, parity, schooling, number of contraceptive methods ever used, and type of method used at baseline (long- or short-acting). A measure of household wealth was also included based on a previously constructed asset index (Amin, Shah, and Becker 2010).⁶ Sampling weights were used in the regression analysis to account for the stratified sampling design.

³The presumption that a woman’s report of her current status is more accurate than a report of past use may not always be appropriate. For example, when contraceptive use is a sensitive topic and women do not feel comfortable reporting their family planning behaviors, the retrospective report might be more accurate. Given the widespread practice of contraception in this setting, however, we believe the assumption is appropriate.

⁴The starting month of the calendar in both survey rounds is the month of interview. A woman’s response to the standard survey question of whether she is using a method and, if so, which method she is using is recorded in the first month of the calendar.

Therefore, for the month of interview, the standard survey question and the calendar report match exactly for all women.

⁵The Kappa statistic is calculated as follows:

$$K = \frac{(\text{proportion of observed agreement}) - (\text{proportion of agreement expected by chance alone})}{1 - (\text{proportion of agreement expected by chance alone})}$$

We follow the prevailing interpretation of Kappa values initially proposed by Landis and Koch (1977): a Kappa statistic greater than 0.80 indicates excellent reliability, a value between 0.41 and 0.80 represents moderate to substantial agreement, a value between 0.01 and 0.40 indicates slight to fair agreement, and a value of 0.00 indicates agreement no better than chance alone.

⁶Information concerning assets was collected in the household questionnaire. Binary indicators included presence or absence of electricity, a wardrobe, table, chair, clock, bed, radio, television, and bicycle, and at least one of a motorcycle, sewing machine, or telephone; brick, cement, or tin walls; and modern toilet or pit latrine. In addition, the ratio of the number of individuals in the household to the number of rooms in the house was included. Principal components analysis was used to combine the asset indicators and household density figure into an asset index that was assigned to each respondent (Filmer and Pritchett 2001).

Results

Less than six percent of women who completed the calendar during the baseline survey and who would have been eligible for the calendar questions during the follow-up survey (younger than age 50 at time of second survey) were lost to follow-up. Table 1 indicates that the 192 women who did not complete the second calendar were significantly younger and of lower parity than the women who completed both calendars. Similar proportions of both groups reported ever attending school, ever practicing contraception, and practicing contraception at the time of the baseline survey. Women who were lost to follow-up were in households with a lower asset score than women who completed both calendars, but the difference was only marginally statistically significant ($p = 0.066$).

Seventy percent of nonpregnant married women aged 13–49 reported currently using a method of contraception at the time of the 2006 baseline survey (Table 2). Contraceptive prevalence and the distribution of methods for the baseline interview month as captured in the 2009 calendar deviated only slightly from the 2006 figures.

Overall, 64 percent of women had identical reports for the baseline interview month in the baseline and follow-up surveys (Kappa 0.55) (not shown). In the follow-up survey, when the window of comparison was widened to include two months on either side of the baseline interview month, the percentage of women with concordant reports increased only to 67 percent (not shown). Reliability varies considerably by type of report, however.

Table 3 compares the 2006 and 2009 reports for the 2006 interview month grouped by category (not using contraceptives, using contraceptives, and pregnant/pregnancy outcome). The Kappa statistic of 0.56 indicates moderate to good reliability; however, 23 percent of respondents reported a different status in the two surveys for the reference month (not shown). Women who reported use of a contraceptive method at the time of interview reported their status most reliably in the follow-up calendar (85 percent agreement), whereas women reporting nonuse and those reporting a pregnancy or pregnancy outcome at baseline reported less reliably in the second calendar (65 percent and 68 percent reported the same status, respectively). When the window for comparison was widened to two months on either side of the baseline interview month, the percentage of concordant responses increased slightly, to 86 percent among women reporting contraceptive use at baseline and to 70 percent among women reporting pregnancy or pregnancy outcome (not shown). Removing the 73 pregnancies that did not end in a live birth did not change the level of report concordance. Additionally, in 2006 no difference in reporting reliability was seen between women in the first three months of pregnancy and those later in their pregnancy.

Table 4 displays the correspondence of reports between baseline and follow-up among women who reported use of any contraceptive method in the baseline interview month in both calendars. Reporting was most reliable for women using sterilization, with more than 98 percent of women reporting the same method in the two calendars. The least reliable reports were for coitally dependent methods, with only 12 percent of withdrawal users and 38 percent of condom users reporting reliably. Eighty-four percent of pill users and 69 percent of injectables users (the two most commonly used methods in this sample) provided concordant reports. Interestingly, the most common discordant report among pill users was injectables use, and the most common discordant report among injectables users was pill use. Examination of both the baseline and follow-up calendars for these women showed that the majority had switched between these two methods during the six-year period covered by the two calendars.

As noted above, only 38 percent of condom users at baseline reported consistently in the follow-up survey; the same percentage of condom users reported pill use for the reference

month. Women could have been using the pill and condoms simultaneously; the calendar only allowed interviewers to record one method per month, so dual method use could not be coded. Women practicing periodic abstinence at baseline were much more likely to report reliably than women practicing withdrawal—47 and 13 percent concordant, respectively.

Predictors of reliable reporting are shown in Table 5. In both crude and adjusted logistic regression analyses, only parity and number of different contraceptive methods ever used were significantly associated with reliability of reporting. The odds of reporting reliably in the calendar were reduced by 27 percent with each additional birth. Similarly, compared with women who had never used a method of contraception, women who had used two contraceptive methods in their lifetime had less than half the odds of providing reliable calendar reports, and those who had used four or more methods had 70 percent lower odds of reliable reporting.

Discussion

In this sample of rural Bangladeshi women, the data collected with the month-by-month calendar produce estimates of contraceptive prevalence similar to those from standard questions about current contraceptive use three years in the past. The results presented here align with those of previous studies that have found moderate to good reliability of reporting associated with the contraceptive calendar at the aggregate level. As expected, reports of long-term and regularly used methods such as sterilization and hormonal methods are reported more reliably than methods used less frequently and at irregular intervals such as coitally dependent and traditional methods. Similarly, overall reports of pregnancies and birth outcomes show less concordance than those of any method use or no use, even after removing women whose pregnancies did not end in a live birth. Because of the potential difficulty in identifying early pregnancy, we also compared the reliability of reporting among women within the first three months of pregnancy with those in later pregnancy in 2006, but found no difference. Expansion of the time window for matches by two months on either side of the interview month resulted in only a small increase in concordant reports. More matches would likely be found if the amount of calendar overlap allowed for a greater expansion of the window.

Compared to results from Morocco, our results show poorer correspondence of reports by category of response (not using contraceptives, using a method, or pregnant/pregnancy outcome) and by specific method (Strickler et al. 1997). The percentage of women in our sample with inconsistent reports by method is more than double the Morocco results (24 and 11 percent, respectively). Reports of condom and traditional method use show the greatest differences in consistency in the two studies. As in Morocco, the pill is the most commonly used method in our sample, and the majority of discordant reports among users of all methods except withdrawal were of pill use. The common misreporting of pill use reflects the fact that many women switched from the pill to other methods or discontinued use in the last six years.

Not surprisingly, women who have used a greater number of methods during their lifetime and those who have had multiple pregnancies were more likely to report inconsistently in the two calendars. As was found in Morocco, the complexity of a woman's reproductive history emerged as the most important predictor of her reporting reliability. Though we had hypothesized that women with less schooling would report less reliably than women with more education, no differences were seen by school attendance. The lack of an education effect may be the result of the overall low levels of educational attainment among this sample of women: 38 percent of women never attended school, and among those who had received any schooling only 35 percent completed sixth grade. Household wealth also had

no effect on the reliability of reporting. Because our sample consisted of rural women only, we could not compare reporting by urban and rural residence. Finally, though we expected that long-term method use would be associated with more reliable reporting, it did not reach statistical significance in the logistic regression model, probably because of the limited number of women reporting use of a long-term method (male or female sterilization, IUD, or implant).

The purpose of this analysis was to compare the consistency of women's reports of prior contraceptive use and pregnancy events using a month-by-month calendar to reports given three years earlier. The data used here do not allow for a comparison of the effectiveness of the contraceptive calendar with that of other methods of soliciting contraceptive and birth-history data. The few studies that have compared the calendar method with other types of questions about past contraceptive use have found that the calendar performs just as well or better than questions included in the body of the questionnaire (Goldman, Moreno, and Westoff 1989; Westoff, Goldman, and Moreno 1990; Becker and Sosa 1992; Becker and Diop-Sidibé 2003). Additional comparisons of this type should be carried out in different settings to strengthen the evidence base regarding the contraceptive calendar. Our analysis shows that the calendar can provide fairly reliable reports of reproductive histories over a three-year period. Most women in this sample reported the same reproductive event (use of a contraceptive method, pregnancy, birth, and so forth) for a particular month when asked three years later, though the degree of concordance varies by event. Given the challenges often encountered in rural Bangladesh and the rest of the region regarding reporting of age and date (Bairagi 1982; Friedman 1993; Pullum 2006; Pardeshi 2010), the results we obtained are encouraging.

Our results support the continued inclusion of the calendar in surveys such as the DHS. Researchers analyzing calendar data to examine contraceptive-use dynamics should be aware of their limitations, however, especially in settings with high rates of contraceptive discontinuation and switching. In these settings, a large proportion of women will have complex reproductive histories, and as our results and those from the Moroccan study show, women with more complicated histories are less likely to report reliably. Unfortunately, the calendar included in our study did not overlap for a sufficiently long period to allow us to look at the reliability of the reported duration and sequence of contraceptive-use events. Results from the Moroccan study, however, showed that only 67 percent of women reported the same number of contraceptive-use segments in the two-year-plus overlap period of the two calendars, and only 59 percent reported the same duration of use for the overlap period (Strickler et al. 1997). Additional studies should be designed to evaluate the reliability of data reporting the duration and sequencing of reproductive-history events among different populations. Additionally, interviewer training should include strategies for identifying and helping respondents who have more complicated histories to report as reliably as possible. Such strategies might include spending extra time with these respondents in completing the calendar and providing additional memory cues for dating specific reproductive events.

The increasing use of calendar data for studies of contraceptive-use dynamics, coupled with the cost of the training associated with implementing the calendar in current DHS surveys, warrant investment in additional research on the effectiveness of the method. The biggest obstacle to conducting research on the reliability of data collected with the contraceptive calendar, however, is a lack of appropriate longitudinal datasets that include the calendar. The DHS program should consider conducting additional panel surveys—similar to those carried out in Morocco—that include overlapping contraceptive calendars. Whereas a panel survey with a five-year follow-up would be useful for assessing the reliability of calendar reports during this period (which is the standard length of the calendar included in most DHS surveys), a panel survey with a shorter follow-up period is necessary to have a

sufficient period of overlap in the baseline and follow-up calendar to assess the reliability of reports regarding duration and sequence of use. A panel survey conducted three years apart with a five-year calendar in each would be ideal because it would allow a two-year period of overlap to study the reliability of reports of contraceptive-method type, duration, and switching. The family planning field would also benefit from a better understanding of how measurement error in the calendar affects the results and interpretation of studies of contraceptive use dynamics. For this, it would be helpful to have simulation studies that model measurement error in the calendar data and specify how much measurement error is too much.

Acknowledgments

This research was supported by the National Institute of Child Health and Human Development. Associates for Community and Population Research was responsible for data collection.

References

- Ali, Mohamed M.; Cleland, John. Oral contraceptive discontinuation and its aftermath in 19 developing countries. *Contraception*. 2010; 81(1):22–29. [PubMed: 20004269]
- Amin, Ruhul; Shah, Nirali M.; Becker, Stan. Socioeconomic factors differentiating maternal and child health-seeking behavior in rural Bangladesh: A cross-sectional analysis. *International Journal for Equity in Health*. 2010; 9(1):9. [PubMed: 20361875]
- Angel, Lucie; Fay, Séverine; Bouazzaoui, Badiâa; Baudouin, Alexia; Isingrini, Michel. Protective role of educational level on episodic memory aging: An event-related potential study. *Brain and Cognition*. 2010; 74(3):312–323. [PubMed: 20875697]
- Bairagi R, Aziz KMA, Chowdhury MK, Edmonston B. Age misstatement for young children in rural Bangladesh. *Demography*. 1982; 19(4):447–458. [PubMed: 7173466]
- Baschieri, Angela; Hinde, Andrew. The proximate determinants of fertility and birth intervals in Egypt: An application of calendar data. *Demographic Research*. 2007; 16(3):59–96.
- Becker, Stan; Ahmed, Saifuddin. Dynamics of contraceptive use and breastfeeding during the postpartum period in Peru and Indonesia. *Population Studies*. 2001; 55(2):165–179.
- Becker, Stan; Diop-Sidibé, Nafissatou. Does use of the calendar in surveys reduce heaping? *Studies in Family Planning*. 2003; 34(2):127–132. [PubMed: 12894757]
- Becker, Stan; Sosa, Doris. An experiment using a month-by-month calendar in a family planning survey in Costa Rica. *Studies in Family Planning*. 1992; 23(6):386–391. [PubMed: 1293862]
- Belli RF. The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory*. 1998; 6(4):383–406. [PubMed: 9829098]
- Bradburn, Norman M.; Rips, Lance J.; Shevell, Steven K. Answering autobiographical questions: The impact of memory and inference on surveys. *Science*. 1987; 236(4798):157–162. [PubMed: 3563494]
- Bradley; Sarah, EK.; Schwandt, Hilary M.; Khan, Shane. DHS Analytical Studies No. 20. ICF Macro; Calverton, MD: 2009. Levels, trends, and reasons for contraceptive discontinuation.
- Brewer, William F. What is autobiographical memory?. In: Rubin, David C., editor. *Autobiographical Memory*. Cambridge University Press; New York: 1986. p. 25–49.
- Christensen, Helen; Anstey, Kaarin J.; Leach, Liana S.; Mackinnon, Andrew J. Intelligence, education, and the brain reserve hypothesis. In: Craik, Fergus I. M.; Salthouse, Timothy A., editors. *Handbook of Aging and Cognition*. Psychology Press; New York: 2008. p. 133–187.
- Creanga, Andreea A.; Acharya, Rajib; Ahmed, Saifuddin; Tsui, Amy O. Contraceptive discontinuation and failure and subsequent abortion in Romania: 1994–99. *Studies in Family Planning*. 2007; 38(1):23–34. [PubMed: 17385380]
- Filmer, Deon; Pritchett, Lant H. Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography*. 2001; 38(1):115–132. [PubMed: 11227840]

- Friedman, William J. Memory for the timing of past events. *Psychological Bulletin*. 1993; 113(1):44–66.
- Gaslonde, Santigao; Carrasco, Enrique. World Fertility Survey Occasional Papers No. 23. International Statistical Institute; Voorburg, Holland: 1982. The impact of some intermediate variables on fertility: Evidence from the Venezuela National Fertility Survey, 1977.
- Glasner, Tina; van der Vaart, Wander. Applications of calendar instruments in social surveys: A review. *Quality and Quantity*. 2009; 43(3):333–349.
- Goldman, Noreen; Moreno, Lorenzo; Westoff, Charles F. Collection of survey data on contraception: An evaluation of an experiment in Peru. *Studies in Family Planning*. 1989; 20(3):147–157. [PubMed: 2734811]
- Gylmore, Maria M.; Kawachi, Ichiro; Jencks, Christopher S.; Berkman, Lisa F. Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology & Community Health*. 2008; 62(6):532–537. [PubMed: 18477752]
- Hossain, Mian B. Analyzing the relationship between family planning workers' contact and contraceptive switching in rural Bangladesh using multilevel modeling. *Journal of Biosocial Science*. 2005; 37:529–554. [PubMed: 16174345]
- Jobe, Jared B.; White, Andrew A.; Kelley, Catherine L.; Mingay, David J.; Sanchez, Marcus J.; Loftus, Elizabeth F. Recall strategies and memory for health-care visits. *Millbank Quarterly*. 1990; 68(2): 171–189.
- Laing, John. E. Natural family planning in the Philippines. *Studies in Family Planning*. 1984; 15(2): 49–61. [PubMed: 6710549]
- Landis, J. Richard; Koch, Gary G. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [PubMed: 843571]
- Levine, Brian; Svoboda, Eva; Hay, Janine F.; Winocur, Gordon; Moscovitch, Morris. Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*. 2002; 17(4):677–689. [PubMed: 12507363]
- Loftus, Elizabeth F.; Smith, Kyle D.; Klinger, Mark R.; Fiedler, Judith. Memory and mismemory for health events. In: Tanur, Judith M., editor. *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*. Russell Sage Foundation; New York: 1992. p. 102-137.
- Luke, Nancy; Clark, Shelley; Zulu, Eliya Ml. The relationship history calendar: Improving the scope and quality of data on youth sexual behavior. *Demography*. 2011; 48(3)
- Macro International. Model 'A' Questionnaire with Commentary for High Contraceptive Prevalence Countries, DHS-II Basic Documentation No. 1. Columbia, MD; 1990. <<http://measuredhs.com/pubs/pdf/DHSQ2/DHS-II-Model-A.pdf.pdf>>
- Menon, Geeta. The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research*. 1993; 20(3):431–440.
- Office of Population Research (OPR), Princeton University. [Accessed February 2012] National Fertility Survey. 1965. <<http://opr.princeton.edu/Archive/nfs/>>
- Pardeshi, Geeta S. Age heaping and accuracy of age data collected during a community survey in the Yavatmal District, Maharashtra. *Indian Journal of Community Medicine*. 2010; 35(3):391–395. [PubMed: 21031103]
- Piolino, Pascale; Desgranges, Béatrice; Clarys, David, et al. Autobiographical memory, auto-noteic consciousness, and self-perspective in aging. *Psychology and Aging*. 2006; 21(3):510–525. [PubMed: 16953713]
- Pullum, Thomas W. DHS Methodological Reports No. 5. ICF Macro; Calverton, MD: 2006. An assessment of age and date reporting in the DHS surveys, 1985-2003.
- Steele, Fiona; Goldstein, Harvey; Browne, William. A general multilevel multistate competing risks model for event history data with an application to a study of contraceptive use dynamics. *Statistical Modeling*. 2004; 4:145–159.
- Jacques, St.; Peggy, L.; Levine, Brian. Ageing and autobiographical memory for emotional and neutral events. *Memory*. 2007; 15(2):129–144. [PubMed: 17534107]

- Strickler, Jennifer A.; Magnani, Robert J.; McCann, H. Gilman; Brown, Lisanne F.; Rice, Janet C. The reliability of reporting of contraceptive behavior in DHS calendar data: Evidence from Morocco. *Studies in Family Planning*. 1997; 28(1):44–53. [PubMed: 9097385]
- van der Vaart, Wander. The time-line as a device to enhance recall in standardized research interviews: A split ballot study. *Journal of Official Statistics*. 2004; 20(2):301–317.
- van der Vaart, Wander; Glasner, Tina. Applying a timeline as a recall aid in a telephone survey: A record check study. *Applied Cognitive Psychology*. 2007; 21(2):227–238.
- Westoff, Charles F.; Goldman, Noreen; Moreno, Lorenzo. Dominican Republic Experimental Study. Institute for Resource Development/Macro Systems; Columbia, MD: 1990.

Table 1

Percentage of women, by demographic characteristics and contraceptive practices, at time of baseline survey, according to whether completed both survey calendars, Rural Bangladesh, 2006

Characteristic	Completed both calendars (N = 3,080)	Lost to follow-up (N =192)
Age (mean)	30.4	26.7***
Parity (mean)	3.03	2.04*
Asset score (mean)	0.292	-0.542
Ever attended school	61.9	54.0
Ever practiced contraception	86.4	78.8
Currently practicing contraception	72.8	51.0
Current method used (percent of all current users)		
Pill	47.8	64.2
Injectables	20.8	11.0
Female sterilization	8.1	10.2
Periodic abstinence	9.6	5.4
Condom	6.7	6.0
Implant	1.1	1.9
Male sterilization	0.2	0.0
Withdrawal	1.6	1.2
IUD	2.6	0.0

Notes: P-values based on Wald tests for continuous variables and chi square tests for categorical variables. Adjusted for sample design.

* Significant at p .05;

*** p .001.

Table 2

Percentage of nonpregnant, currently married women aged 13–49, by reported contraceptive use and type of method used during the baseline interview month, according to survey, Rural Bangladesh

	Baseline survey (N = 2,695) ^a	Follow-up survey (N = 2,678) ^a
Contraceptive prevalence	69.5	70.5
Current method used		
Pill	47.3	49.1
Injectables	23.5	22.2
Female sterilization	9.6	10.6
Periodic abstinence	9.7	11.7
Condom	4.1	2.1
Implant	1.6	1.5
Withdrawal	1.4	0.6
Other ^b	2.6	2.3

^aVariation in sample size results from different reports of marital status in the baseline and follow-up surveys. Only currently married women completed the calendar.

^b"Other" consists of male sterilization and IUD.

Table 3

Percentage distribution of women's reports in 2009 of reproductive status during the month of the 2006 baseline interview, by reproductive status reported in 2006, Rural Bangladesh

Current status reported in 2006 baseline survey (n)	Status reported in 2009 follow-up survey for 2006 baseline interview month (n)			Total (3,080)
	Not using contraceptives (919)	Using contraceptives (1,899)	Pregnant or pregnancy outcome ^a (262)	
Not using contraceptives (974)	64.7	28.9	6.4	100.0
Using contraceptives (1,876)	13.0	84.8	2.2	100.0
Pregnant/pregnancy outcome ^a (230)	19.6	12.2	68.3	100.0
Total (3,080)	29.8	61.7	8.5	100.0

Note: Kappa = 0.56 ($z = 38.24$, $p < 0.005$ for test that Kappa = 0).

^aPregnancy outcome includes births, miscarriages, abortions, and stillbirths.

Table 4

Percentage distribution of women who reported use of any contraceptive method during the reference month in both surveys, by method reported as currently used in 2006, according to method reported in 2009 for month of 2006 baseline interview, Rural Bangladesh

Current method reported in 2006 baseline survey (n)	Method reported in 2009 follow-up survey for 2006 baseline interview month (n)								All methods (1,590)
	Pill (780)	Injectables (363)	Female sterilization (187)	Periodic abstinence (147)	Condom (34)	Implant (27)	Withdrawal (9)	Other (43)	
Pills (751)	84.2	8.5	0.4	5.2	0.5	0.4	0.4	0.7	100.0
Injectables (381)	20.5	68.8	1.1	7.6	0.8	0.3	0.3	0.8	100.0
Female sterilization (180)	1.1	0.6	98.3	0.0	0.0	0.0	0.0	0.0	100.0
Periodic abstinence (124)	25.0	17.7	0.8	46.8	2.4	0.8	2.4	4.0	100.0
Condom (61)	37.7	11.5	0.0	11.5	37.7	0.0	1.6	0.0	100.0
Implant (29)	17.2	3.5	3.5	0.0	0.0	72.4	0.0	3.5	100.0
Withdrawal (24)	25.0	16.7	4.2	37.5	4.2	0.0	12.5	0.0	100.0
Other (40)	7.5	5.0	0.0	12.5	0.0	2.5	0.0	72.5	100.0
All methods (1,590)	49.1	22.8	11.8	9.3	2.1	1.7	0.6	2.7	100.0

Note: Kappa = 0.65 ($z = 46.06$, $p < 0.001$ for test that Kappa = 0). "Other" consists of male sterilization and IUD.

Table 5

Crude and adjusted odds ratios predicting reliable reporting of contraceptive use, pregnancy, and pregnancy outcomes between report at baseline and report for the baseline interview month from the follow-up survey, Rural Bangladesh

Covariate	Odds Ratio	
	Crude	Adjusted
Age	1.00	1.04
Parity	0.84**	0.73**
Household asset index	1.02	1.02
Ever attended school	0.84	0.75
Number of methods used in lifetime		
0 (r)	1.00	1.00
1	0.86	1.11
2	0.40*	0.48*
3	0.50*	0.61
4+	0.26*	0.30*
Use of long-term method at baseline ^a	2.99	3.07

Note: Adjusted for sample design.

* Significant at $p = 0.05$;

*** $p = 0.01$.

^a Long-term methods include female and male sterilization, IUD, and implants.