

Characteristics Associated with Reliability in Reporting of Contraceptive Use: Assessing the Reliability of the Contraceptive Calendar in Seven Countries

Philip Anglewicz,¹ Dana Sarnak, Alison Gemmill, and Stan Becker

Although the reproductive calendar is the primary tool for measuring contraceptive dynamics in low-income settings, the reliability of calendar data has seldom been evaluated, primarily due to the lack of longitudinal panel data. In this research, we evaluated the reproductive calendar using data from the Performance Monitoring for Action Project. We used population-based longitudinal data from nine settings in seven countries: Burkina Faso, Nigeria (Kano and Lagos States), Democratic Republic of Congo (Kinshasa and Kongo Central Provinces), Kenya, Uganda, Cote d'Ivoire, and India. To evaluate reliability, we compared the baseline cross-sectional report of contraceptive use (overall and by contraceptive method), nonuse, or pregnancy with the retrospective reproductive calendar entry for the corresponding month, measured at follow-up. We use multivariable regressions to identify characteristics associated with reliability or reporting. Overall, we found that the reliability of the calendar is in the moderate/substantial range for nearly all geographies and tests (Kappa statistics between .4 and .7). Measures of the complexity of the calendar (number of contraceptive use episodes, using the long-acting method at baseline) are associated with reliability. We also found that women who were using contraception without their partners/husbands' knowledge (i.e., covertly) were less likely to report reliably in several countries.

Philip Anglewicz, Dana Sarnak, Alison Gemmill, Stan Becker, Department of Population, Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. E-mail: panglew@jhu.edu.

© The Authors. Studies in Family Planning published by Wiley Periodicals LLC on behalf of Population Council.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium provided the original work is properly cited.

of retrospectively collected data would be to interview the same women multiple times (pg.).

To date, there have been only three studies with the longitudinal design and calendar data necessary to achieve this objective (Callahan and Becker ; Strickler et al. ; Tumlinson and Curtis), using data from only three countries: Kenya, Bangladesh, and Morocco. Of these studies, only one used nationally representative data; the Bangladesh study only included data for rural residents (Amin et al.), and the Kenya study was among urban residents. All three studies involved populations with relatively high contraceptive use and low fertility compared to most countries in sub-Saharan Africa (SSA): the contraceptive prevalence rate was percent in Morocco, percent in Bangladesh, and percent in Kenya (Callahan and Becker ; Strickler et al. ; Tumlinson and Curtis). Only Tumlinson and Curtis () used longitudinal calendar data from a country in SSA, many of which have among the highest fertility rates and lowest rates of contraceptive use in the world (United Nations).

The three studies that have used longitudinal data to evaluate the reliability of calendar data generally agree on some features of the calendar: () the reliability of the calendar falls within the moderate to substantial agreement category (Kappa statistics between . and .), and () women with more complex reproductive histories are less reliable in reporting their calendar information (Callahan and Becker ; Strickler et al. ; Tumlinson and Curtis). However, there is also substantial disagreement and notable gaps in this research. First, there is variation in reliability across settings, with Kappa's ranging from . to . across countries and ways of testing reliability (Callahan and Becker ; Strickler et al. ; Tumlinson and Curtis). Second, the characteristics associated with reliability are not consistent across studies: measures like age are not consistently associated with reliability across studies, and other measures, like household wealth and urban/rural residence, were not tested in all studies. Third, although all three studies hypothesize that the use of long-acting reversible methods (LARCs) is associated with greater reliability, this was only found in one (Tumlinson and Curtis), potentially due to small sample sizes of LARC users in the other two studies.

The limitations of previous research are well-documented. Bradley et al. () note that •Few studies to date have examined the quality of the contraceptive information collected

the percent of women who reported each status in their P retrospective calendar for the P survey month out of the total number of women who reported that status in the P survey. The total percent agreement was calculated as the total number of women who were concordant across the categories out of the total number of women. Kappa statistics were also computed as a means to evaluate how likely the concordance departs from chance (Landis and Koch).

In addition to the analysis of agreement in the exact month, we also examined agreement within a +

TABLE Percent agreement and kappa statistics for concordance in reporting of reproductive status in the same reference month (the \times Concordance Analysis), – Performance Monitoring for Action data from nine geographies

Kenya	% Agreement/Kappa	Unweighted n
Nonuse	.	,
Pregnancy outcome	.	,
FP use	.	,
Total % agreement	.	,
Kappa statistic	.	,
Unweighted n=	.	,
Nigeria-Lagos	% Agreement/Kappa	Unweighted n
Nonuse	.	,

TABLE (Continued)

Kenya	% Agreement/Kappa	Unweighted n
Rajasthan	% Agreement/Kappa	Unweighted n
Nonuse	.	,
Pregnancy outcome	.	,
FP use	.	,
Total % agreement	.	,
Kappa statistic	.	,
Unweighted n=	.	,

NOTE: Agreement is defined as the weighted percent of women who reported the respective reproductive status at the P₁ survey and also reported the same status in P₂ calendar for the same month of P₁ survey.

The results of our multivariable analysis are shown in Table 3, which includes sociodemographic and calendar complexity measures that are associated with reliability. Several results are consistent across geographies: across all sites except Nigeria-Kano, women who reported using a long-acting method at P₁ had greater odds of being reliable in their report than women not using long-acting methods in P₁. In all sites except Nigeria-Kano and Rajasthan, women with more use episodes had lower odds of concordant reporting. Across most sites, higher parity had a negative relationship with concordance (except Nigeria-Lagos, DRC-Kinshasa, and Uganda).

Other results were consistent in a subset of geographies. In DRC-Kongo, Cote d'Ivoire, and Rajasthan, older age was associated with higher odds of being concordant. In another three sites, Nigeria-Kano, DRC-Kongo, and Burkina Faso, higher education was associated with lower concordance. Also, having more pregnancies in the P₁ calendar was associated with lower odds of concordance in Kenya, Nigeria-Kano, and Burkina Faso. Finally, we see that the relationship with wealth varies across settings: increased wealth was associated with higher concordance in Nigeria-Kano and DRC Kongo but lower concordance in Cote d'Ivoire.

Finally, in Table 4, we see the results for the association between $\text{cafo.2(14.9(r)6elia)73(($

TABLE Percent agreement and kappa statistics for concordance in reporting of reproductive status in the same reference month (the

TABLE Adjusted odds ratios and confidence intervals for the effects of covert use and EA-level community norms on reliable reporting in the reproductive calendar for nine geographic regions, – Performance Monitoring for Action data

Setting and covariates	Adjusted Odds Ratios	% Confidence Intervals	
Kenya			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
Nigeria-Lagos			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
Nigeria-Kano			
Reported covert use at P or P	N/A	N/A	N/A
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
DRC-Kinshasa			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
DRC-Kongo Central			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
Burkina Faso			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
Uganda			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
Cote d'Ivoire			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.
India-Rajasthan			
Reported covert use at P or P	.	.	.
EA-level CPR	.	.	.
EA-level % agree that family planning is for married women only	.	.	.

NOTE: Models also control for age, parity, wealth, education, residence (where available), pregnancies in P calendar, use episodes in the P calendar, and using a long-acting method in P. Boldfaced odds ratios and CIs are significant at $p < .05$. Nigeria-Kano was excluded from the analysis of covert use due to small sample size. The survey question that defines covert use (‘‘does your partner know you are using?’’) were asked of any woman who reported the use of any method except male condom, withdrawal, or male sterilization. The binary measure of covert use included in the model is woman reported covert use at either P or P = woman never reported covert use. Therefore, this sample is limited to users at P or P of nonmale-controlled methods.

DISCUSSION

In this research, we used longitudinal panel data from nine geographies in seven countries to evaluate the reliability of the reproductive calendar. These geographies capture considerable variation in family planning characteristics, with a range of mCPR among all women from . . . percent to . . . percent and varying method mixes. To begin, we compared reports of contraceptive use, nonuse, and pregnancy from two separate datasets for the same women for

the same month. We evaluated the extent of agreement for three broad categories (nonuse, pregnancy, and contraceptive use) and more-specific categories (nonuse, pregnancy, and contraceptive use for each specific method). We then identified characteristics associated with greater reliability in reporting these items, including both sociodemographic characteristics, measures of calendar complexity, and measures of the willingness of women to report contraceptive use.

Overall, across both the broader () and more specific approaches (), we find that reliability generally falls within the moderate to substantial agreement category, with some results (Rajasthan) in the category of excellent reliability (Landis and Koch). As expected, the reliability is generally higher for the broader approach, but the differences in Kappa statistics are not substantial between these approaches and fall between . and . for all geographies. Also, as expected, the extent of agreement increases with an expanded reporting period of .3(t)-4.7(hin)m76m .0041uxgemotas pp endxgTe,e b89.8ub

with tils b899(u)85.7(er).25-.5(p)-947(e)-.8(3.9(s5-.5(.)]TJ 1.6367 -1.2791 TD -.0002 Tc [Lp)-9.1(o-)9.2(o17.2ke)-4.3inh)8.1(g)-29 9.2(o)12.1fs

their husbands may also not consistently report to an interviewer, which suggests that the tendency to reliably report contraceptive use may depend not only on the respondent's ability to remember their use patterns but also their willingness to report contraceptive use. We do not find consistent evidence for the other measures in this category, the EA-level average contraceptive use or the EA-level average percentage of women who believe family planning is only for married women; as these measures are only significantly associated with reliability in one or two settings. Nonetheless, the more consistent relationship between reliability

data collection (and we found minimal data entry errors), and () provide the potential for correcting the data if necessary. An experiment in which REs were randomly assigned to use the paper aid compared with no aid would be necessary to identify the effect of the aid on calendar reliability. Second, because prior research has suggested that the reliability of calendar data decreases with larger recall periods (Bradley et al.), PMA chose to implement a two-to-three-year calendar; and it is very likely that data quality would be worse with a one-year calendar instead (as DHS and other surveys do). Third, PMA devoted a considerable amount of time training REs in the calendar data collection, including extensive pilot testing of the approach prior to data collection, and video instructions that REs could keep on their phones for future reference. Fourth, previous analysis of PMA data has demonstrated the value of the RE approach, which suggests that REs yield better quality data than using interviewers who are not from the study sites (Anglewicz et al. ; Saha). Based on this analysis, it is reasonable to expect that the impact of social desirability bias might be greater with outsider interviewers. Finally, because we found that the use of long-acting methods is associated with greater reliability, the increase in these methods in recent years (Tsui et al.) suggests that calendar data will be more reliable over time, although one would also want to consider the fertility rate and extent of method switching when using the calendar approach.

Finally, we revisit the tradeoffs between study designs. The longitudinal panel approach allows a comparison of reports for the same women over time, which permits one to identify characteristics associated with reporting patterns; as well as the opportunity prospectively measure contraceptive use. In contrast, a cross-sectional design only allows a population-level comparison but is less costly than a longitudinal panel. If evaluating the reliability of the calendar data is a goal of the study, the longitudinal approach is preferable.

FUNDING

Funding for this study was provided by the Bill & Melinda Gates Foundation under grant number OPP . Funders were not involved in any aspect of the study design, data collection, and analysis, nor the interpretation and writing of the manuscript.

REFERENCES

Ali, Mohamed M., John Cleland, and Iqbal H. Shah. Causes and Consequences of Contraceptive Discontinuation: Evidence from Demographic and Health Surveys. Geneva: World Health Organization.

Amin, Ruhul, Nirali Shah, Becker Stan. Socioeconomic Factors Differentiating Maternal and Child Health-Seeking Behavior in Rural Bangladesh: A Cross-Sectional Analysis. International Journal for Equity in Health. ...

Anglewicz, Philip, Pierre Akilimali, Linnea Perry Eitmann, Julie Hernandez, and Patrick Kayembe. The Relationship

Characteristics Associated with Reliability in Reporting of Contraceptive Use

Sa, Sally. . •The Effect of Innovative Survey Design on Demographic and Reproductive Health Data in Sub-Saharan Africa: A Multi-Country Analysis. Doctoral dissertation, Johns Hopkins University.

Strickler, Jennifer, Robert Magnani, Gilman McCann, Lianne Brown, and Janet Rice. . •The Reliability of Reporting of Contraceptive Behavior in DHS Calendar Data: Evidence from Morocco. *Studies in Family Planning* ():

Tsui, Amy, Win Brown, Qingfeng Li. . •Contraceptive Practice in Sub-Saharan Africa. *Population and Development Review*