Is Client Reporting on Contraceptive Use Always Accurate? Measuring Consistency and Change with a Multicountry Study

Amy O. Tsui, Carolina Cardona, ⓘ Varsha Srivatsan,
Funmilola OlaOlorun, Elizabeth Omoluabi, Pierre Akilimali,
Peter Gichangi, Mary Thiongo, PMA Agile Team, Scott Radloff,
and Philip Anglewicz ⓘ

The consistency of self-reported contraceptive use over short periods of time is important for understanding measurement reliability. We assess the consistency of and change in contraceptive use using longitudinal data from 9,390 urban female clients interviewed in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso. Clients were interviewed in-person at a health facility and four to six months later by phone. We compared reports of contraceptive use at baseline with recall of baseline contraceptive use at follow-up. Agreement between these measures ranged from 59.1 percent in DR Congo to 84.4 percent in India. Change in both contraceptive method type (sterilization, long-acting, short-acting, nonuse) and use status (user, nonuser, discontinuer, adopter, switcher) was assessed comparing baseline to follow-up reports and retrospective versus current reports within the follow-up survey. More change in use was observed with panel reporting than within the cross section. The percent agreement between the two scenarios of change ranged from 64.8 percent in DR Congo to 84.5 percent in India, with cross-site variation. Consistently reported change in use status was highest for nonusers, followed by users, discontinuers, adopters, and switchers. Inconsistency in self-reported contraceptive use, even over four to six months, was nontrivial, indicating that studying measurement reliability of contraceptive use remains important.

Amy O. Tsui, Carolina Cardona, Varsha Srivatsan, Scott Radloff, Philip Anglewicz, Department of Population, Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA. E-mail: ccardon4@jhu.edu. Funmilola OlaOlorun, University of Ibadan, Ibadan, Nigeria. Elizabeth Omoluabi, Center for Research, Evaluation Resources and Development, Ile-Ife, Osun, Nigeria. Elizabeth Omoluabi, Statistics and Population Studies Department, University of the Western Cape, Bellville, South Africa. Pierre Akilimali, University of Kinshasa, Kinshasa, Democratic Republic of Congo. Peter Gichangi, Mary Thiongo, International Center for Reproductive Health, Nairobi, Kenya. Peter Gichangi, Technical University of Mombasa, Mombasa, Kenya. Peter Gichangi, Department of Public Health and Primary Care, Faculty of Medicine and Health Sciences, Ghent University, Gent, Belgium. PMA Agile Team,The following are other members of the PMA Agile Team: Titilope Akinlose, United Nations Foundation, Washington DC, USA; Georges Guiella and Bazie Fiacre, Institut Supérieur des Sciences de la Population, Joseph Ki-Zerbo University, 03 BP 7118 Ouagadougou, Burkina Faso; P. R. Sodani and Anupam Mehrotra, Indian Institute of Health Management and Research, Jaipur, India; Sani Oumarou and Souleymane Alzouma, Institut National de la Statistique du Niger, Niamey, Niger.

INTRODUCTION

Self-reports of sensitive or personal behaviors cannot be avoided when studying sexual activity, contraceptive use or abortion practice, or illicit behaviors. Measures of these outcomes are necessary for understanding their risk factors and determinants. In the family planning field, reliance on national household surveys of women of childbearing age has provided a wealth of insights for decades but requires measurement of reproductive behaviors that may not be accurately reported.

In the early years of developing survey measurements of reproductive and contraceptive behaviors, more effort was devoted to establishing reliability and validity than in recent years (Bignami-Van Assche 2003). For example, Anderson and Cleland (1984) compared measurement of current contraceptive use between the World Fertility Surveys and the Contraceptive Prevalence Surveys, noting that ambiguity with how "current" is interpreted, excluding unmarried women as respondents on contraceptive use, and harmonizing the denominator of exposed women were some reasons behind disparate estimates. Pebley, Goldman, and Choe (1986) compared data from Korean surveys in the 1970s to assess consistency of reported contraceptive use to conclude that adequate interviewer training, having contraceptive method awareness questions precede questions on use, and asking about use within defined intervals all improved reporting reliability. An experimental contraceptive calendar introduced in the 1986 Peru Demographic and Health Survey (DHS) enabled assessment of data quality (Goldman, Moreno, and Westoff 1989) and was subsequently included in the core DHS female questionnaire. Two more recent studies of reporting reliability added further insights, both comparing baseline responses about contraceptive use at a particular point in time with retrospective reports using a calendar. The first is based on a panel of Moroccan DHS respondents followed up after three years (Strickler et al. 1997) and the second on a panel of rural Bangladeshi women also reinterviewed after three years (Callahan and Becker 2012). Both studies found substantial discordance in current use reports at the individual level, while aggregated use measured at the population level was more similar. In recent years, there have been fewer assessments of self-reports of contraceptive use, but with annual measurement of contraceptive use in selected countries through the Performance Monitoring for Action (PMA) surveys (Zimmerman et al. 2017), there is renewed opportunity to assess the reliability of self-reported contraceptive use.

Measurement errors during the data collection process can affect the validity and reliability of the key indicators and will be important in program monitoring of contraceptive behavior and validation of routine health information data (Nock, Zeller, and Carmines 1982). Response rates and reporting errors are affected by individual respondent and interviewer characteristics and the content and nature of the questions themselves. Sensitive, personal questions may be subjected to larger biases in response and as a result be less reliable (Knodel and Piampiti 1977). Large surveys employ probability samples to reduce measurement errors by design, while also addressing reliability, stability, and internal consistency concerns by using standardized well-tested questionnaire wordings (Kimberlin and Winterstein 2008). Another factor that may impact survey outcomes is the means of survey administration. Direct face-to-face surveys may offer the benefits of accuracy of screening responses but also can introduce interviewer effects and social desirability bias in responses to sensitive

questions. Self-administered surveys offer respondent privacy and appear best suited for sensitive questions (Hewett, Mensch, and Erulkar 2004). Surveys administered via a telephone, although an economical approach to data collection for large samples, are much likely to be shorter than those conducted face to face and may be affected by selection bias, low response rates, and diminished ability for the interviewer to explain survey questions (Mahfoud et al. 2014) as well as network communications quality.

Reducing measurement error in self-reported behaviors is often achieved through the instrument development process. One approach to understanding survey response error is to examine its sources of variation, those due to measurement, such as questionnaire structure, and those due to nonmeasurement, such as respondent social desirability bias. Consistency in respondent self-reporting of key behaviors reflects measurement reliability, whether from the same individual at different times, indicative of stability, or the replicated use of the same question set across participant samples, suggesting internal consistency for construct measurement.

Consistency in self-reports of sexual behavior other than contraceptive use has been investigated in surveys of populations at risk of HIV or sexually transmitted infections (Bignami-Van Assche 2003; Hamilton and Morris 2010; Hewett, Mensch, and Erulkar 2004) and of adolescents (Goldberg et al. 2014; Alexander et al. 1993; Eggleston, Leitch, and Jackson 2000; Lauritsen and Swicegood 1997; Sieving et al. 2005). These latter studies found strong consistency in reporting of age at first intercourse. Hamilton and Morris (2010) compared reported number of sexual partners across a range of large sample US Surveys, finding relatively small differences in the means but differentials by gender and race. Hewett, Mensch, and Erulkar (2004) randomized Kenyan adolescent female respondents by mode of survey administration (face-to-face interview and audio computer-assisted self-interview or ACASI) and observed the latter generated higher prevalence of high-risk sex and coerced sex. Bignami-van Assche (2003) assessed consistency in responses to survey questions by participants in a Malawi longitudinal study; some of whom were inadvertently reinterviewed a second time. Consistency was high across all questionnaire items, less so for males than females. The number of points of measurement and duration between them can affect levels of consistency. Lauritsen and Swicegood (1997) found 28–32 percent inconsistent reporting, including inaccurate recall, by youth of age at first intercourse between the first survey and last report up to seven years later. The type of measure can also influence the degree of consistency: Rosenbaum et al. (2017) found semen exposure in self-administered vaginal swabs among adolescent females who report recent sexual abstinence. Similar discordance with biomarker measures have been found for self-reported drug use and urine sample assays (Harrison and Hughes 1997).

Beyond the aforementioned studies, there are relatively few recent ones that examine the consistency in self-reported contraceptive use at the same point in time. One that analyzed weekly diary reports of pregnancy intentions and contraceptive use at each intercourse (Moreau et al. 2013) found that consistent use was higher when motivation for pregnancy avoidance was high. A second study (Sieving et al. 2005) found that the reliability of duration of contraceptive use among adolescent girls was high in reports over a six-month period with variation by frequency of intercourse and recall of condom use.

Because self-reported contraceptive use is a central measure in family planning research, it is important to continually assess the accuracy and consistency of self-reports. Survey-based instruments, such as recall using calendars, enhance the importance of accurate measurement. Expanded use of panel surveys to study change in contraceptive behaviors, such as Karp et al. (2021), further reinforce the need to assess the quality of personal reports. Gaps in knowledge remain, for example, in the degree of consistency over time, especially at short intervals and where methods are used in combination serially or simultaneously. At the population level, trends in contraceptive use status can mask change at the individual level. As fertility levels decline, the time spent being sexually active before, between and after births may be more or less protected by contraception and other family planning means to avoid unintended births. This heightens the significance of understanding the dynamics of contraceptive use over different intervals of time and obtaining answers to whether survey respondents, who are interviewed multiple times, recall their past behaviors accurately and consistently.

This study addresses the following two research questions using longitudinal data from six large samples of female clients in urban Democratic Republic of Congo (DR Congo), India, Kenya, Niger, Nigeria, and Burkina Faso, who consented to being reinterviewed by phone after an in-person exit interview at the attended health facility:

(1) How consistently does a client report her contraceptive use status and method type at baseline interview compared to her retrospective report of that use at follow up four to six months later?

(2) How much change in contraceptive use status and method type is observed based on comparing female clients' reports: (1) across surveys (baseline and follow-up), and (2) within the follow-up survey only (cross section)? Related to this, how similar or consistent are the two distributions?

METHODS

Data

Data for this research come from panel surveys of female clients between the ages of 18–49 who attended an urban health facility. The surveys were conducted by in-country research institutions collaborating with the Performance Monitoring and Accountability (PMA) Agile project (www.pmadata.org/technical-areas/pma-agile). PMA Agile was a continuous data-monitoring and evaluation system that collected quarterly estimates of family planning and reproductive health care readiness measures from a probability sample of up to 200 public and private health facilities in each survey site. Semiannual surveys were conducted with a systematic sample of female and male clients—ten per facility—to assess their family planning consumption behaviors. Data were collected in urban sites using resident enumerators; the surveys are conducted at low cost with rapid turnaround (Tsui et al. 2020). Countries included in this analysis are the DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso. PMA Agile has one city site in DR Congo and Niger, three city sites in India, three county town sites in Kenya, three urban sites in Nigeria, and two city sites in Burkina Faso. For DR Congo the site is located in Kinshasa and for Niger the site is Niamey. For India the sites

are in Ferozabad/Shikohabad/Tundla (Uttar Pradesh), Indore (Madhya Pradesh), and Puri (Odisha); for Kenya in Kericho, Uasin Gishu, and Migori towns; for Nigeria in Kano, Lagos, and Ogun; and for Burkina Faso in Ouagadougou and Koudougou. More information about the sampling approach, study design, response rates, and other features of PMA Agile can be found in Tsui et al. (2020).

We use data from client exit interview (CEI) surveys and CEI follow-up surveys conducted by phone approximately four to six months later. Exit interviews were conducted with female clients systematically selected from a probability sample of public and private health facilities in the sites. Upon completion of the exit interview, participants were consented for a follow-up phone survey and asked to provide up to two telephone numbers where they could be reached. A mobile airtime card or recharge with a value of about 1 USD was provided to each respondent completing the baseline interview.

In five countries, the baseline CEI survey was conducted in 2018 (Niger took place in 2019). The total sample of female clients with completed baseline interviews was 13,316: 1,226 in DR Congo, 1,596 in India, 4,431 in Kenya, 936 in Niger, 3,615 in Nigeria, and 1,512 women in Burkina Faso. Consent for follow-up by telephone was obtained from 11,978 clients (90 percent) of whom 11,150 (93 percent) had telephone access. Of these, 9,390 (84 percent) completed follow-up interviews: 751 of 876 in DR Congo (86 percent), 659 of 1,002 in India (66 percent), 3,941 of 4,274 in Kenya (92 percent), 515 of 667 in Niger (77 percent), 2,326 of 2,947 in Nigeria (79 percent), and 1,198 of 1,384 in Burkina Faso (87 percent). Overall, 71 percent participants from the baseline sample were successfully reinterviewed through phone follow-up. Usually, the same interviewer at baseline conducted the follow-up interview. Mobile phone airtime of 1 USD, transmitted electronically, was again provided to the followed-up clients.

All survey data used for analysis were deidentified. Ethical approval was not sought for this analysis of secondary data.

Measures

In the baseline interview (T1), the client's current contraceptive use status was determined based on her responses to her reason for the visit. For a family planning client, her current contraceptive method was the one she was either prescribed or dispensed. If she came for nonfamily planning care, she was asked her current use status, "Are you or your partner currently taking any steps or using any method to avoid or prevent becoming pregnant?," irrespective of where she received this service. If she answered in the affirmative, she was asked about her (or her partner's) current method, including traditional methods. This point of measurement is defined as Time 1 current or T1c. At the follow-up interview (T2), the female client was asked to recall her contraceptive use and method at the time of the baseline exit interview: "We interviewed you at [facility name] about 4 months ago. At that time were you or your partner then using a method to avoid or to prevent becoming pregnant?" If she answered yes, she was then asked about the method she was prescribed or was using (this referent being Time 1 retrospective, or T1r). She was next asked about her current contraceptive use and method, repeating the same question wordings as in the baseline, for her T2 use status.

FIGURE 1   Diagram of urban female client's contraceptive reporting at baseline (T1c) and follow-up survey (T1r and T2)
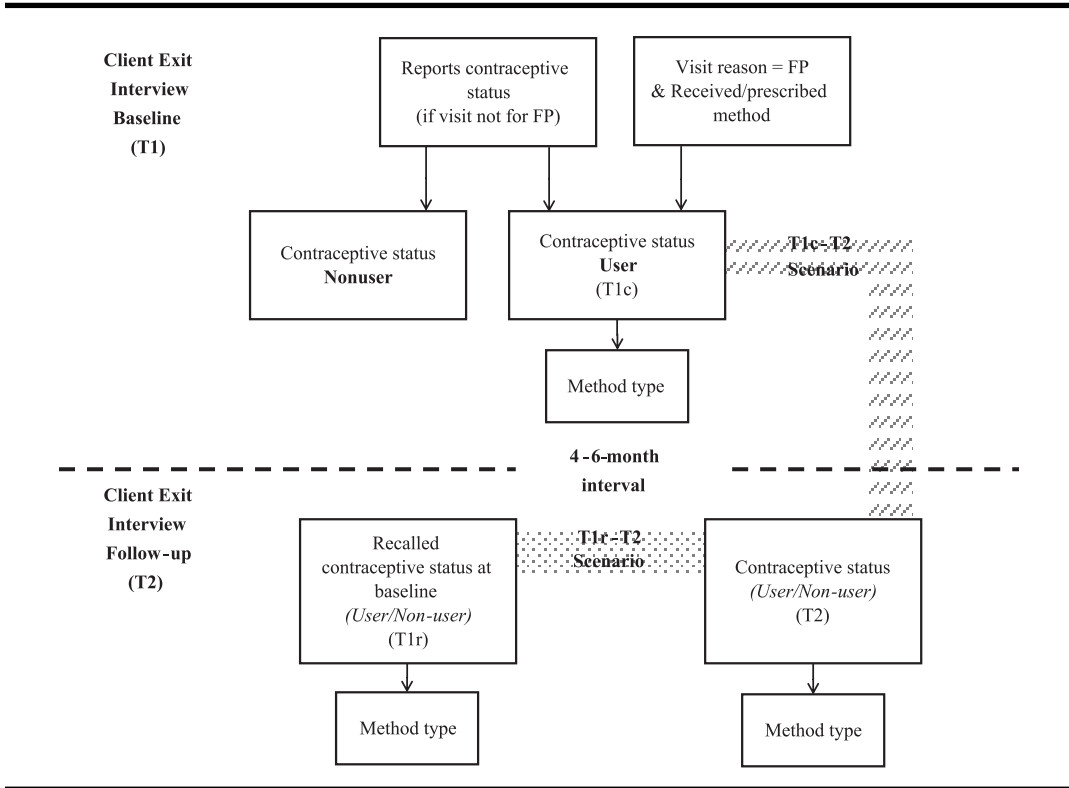
To assess change, we classify the female client's use status as a (1) continuing user or (2) continuing nonuser if she reports using or not using consistently at both T1 and T2. We classify her as a (3) discontinuer if she reports using at T1 and not using at T2, or as (4) an adopter, if she reports not using at T1 but becomes a user at T2. Last, she is classified as (5) a method switcher if she reports using a different method at T1 compared to T2.

It can be challenging for women to recall their contraceptive status from four to six months ago accurately; retrospectively reported baseline contraceptive use may suffer from recall bias. It is also possible that we are measuring actual behavior change (adoption, discontinuation, and switching) in both scenarios and not inconsistent reporting. We cannot disentangle these two sources of variation, unfortunately and will instead focus on agreement among pairwise reports from T1c, T1r, and T2. These comparison pathways are shown in Figure 1, distinguishing the consistency of reporting the baseline method T1c–T1r with the dashed line from the shaded change scenarios of T1c–T2 and T1r–T2. Consistency in change reporting involves then agreement between the T1c–T2 and T1r–T2 scenarios of method type and status.

To assess the compositional heterogeneity of the client samples, we examine selected characteristics of the facility where the in-person exit interview took place and the reason for her visit. The facility type is classified into broad categories of medical college or hospital, private health clinic, public health clinic, and pharmacy. The reason for her health visit

TABLE 1 Percent distribution of selected characteristics of urban female clients in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso

Characteristics	DR Congo (*n* = 755)	India (*n* = 653)	Kenya (*n* = 3,940)	Niger (*n* = 515)	Nigeria (*n* = 2,324)	Burkina Faso (*n* = 1,198)
Visit reason						
General health & other	314 (41.8)	342 (52.4)	2114 (53.7)	219 (42.5)	804 (34.6)	433 (36.1)
Maternal health	132 (17.6)	117 (17.9)	267 (6.8)	86 (16.7)	457 (19.7)	242 (20.2)
Child health	274 (36.5)	165 (25.3)	1325 (33.6)	127 (24.7)	928 (39.9)	384 (32.1)
Family planning	31 (4.1)	29 (4.4)	234 (5.9)	83 (16.1)	135 (5.8)	139 (11.6)
Facility type						
Medical college hospital	429 (57.4)	165 (25.3)	283 (7.2)	28 (5.4)	496 (21.3)	140 (11.7)
Private health clinic	134 (17.9)	62 (9.5)	849 (21.5)	50 (9.7)	1131 (48.7)	657 (54.8)
Public health clinic		165 (25.3)	1851 (47.0)	279 (54.2)	395 (17.0)	264 (22.0)
Pharmacy	185 (24.7)	261 (40.0)	957 (24.3)	158 (30.7)	302 (13.0)	137 (11.4)
Age group						
18–24	218 (29.0)	150 (23.0)	1232 (31.3)	145 (28.2)	448 (19.3)	394 (32.9)
25–34	326 (43.4)	329 (50.4)	1810 (45.9)	238 (46.2)	1188 (51.1)	564 (47.1)
35–49	207 (27.6)	174 (26.6)	898 (22.8)	132 (25.6)	688 (29.6)	240 (20.0)
Parity group						
None-One	326 (43.4)	281 (43.0)	1501 (38.1)	166 (32.3)	891 (38.4)	586 (48.9)
Two	149 (19.8)	196 (30.0)	888 (22.5)	108 (21.0)	497 (21.4)	247 (20.6)
Three plus	276 (36.8)	176 (27.0)	1551 (39.4)	240 (46.7)	934 (40.2)	365 (30.5)
Schooling						
None/Primary	157 (21.0)	151 (23.1)	1556 (39.5)	204 (40.1)	235 (10.2)	322 (26.9)
Post-Primary/secondary	395 (52.7)	369 (56.5)	1432 (36.3)	190 (37.3)	1110 (48.1)	503 (42.1)
College/university	197 (26.3)	133 (20.4)	952 (24.2)	115 (22.6)	964 (41.7)	371 (31.0)
Household wealth						
Poorest 1–3	251 (33.7)	206 (31.5)	885 (22.5)	138 (28.0)	518 (22.8)	631 (53.3)
4	171 (23.0)	130 (19.9)	992 (25.2)	91 (18.5)	463 (20.4)	184 (15.5)
5	207 (27.8)	160 (24.5)	1416 (35.9)	167 (33.9)	650 (28.6)	234 (19.8)
Richest 6–10	115 (15.5)	157 (24.0)	647 (16.4)	96 (19.5)	638 (28.1)	135 (11.4)
Marital status						
In union	490 (65.3)	601 (92.2)	2785 (70.7)	448 (87.0)	2023 (87.0)	952 (79.5)
Not in union	260 (34.7)	51 (7.8)	1154 (29.3)	67 (13.0)	301 (13.0)	246 (20.5)

NOTES: Each country site's sample is of urban female clients interviewed at baseline and follow-up.
Public and private managing authority of facilities in DR Congo/Kinshasa could not be distinguished and the figures refer to clients interviewed at both types of health clinics.
Percentages may not sum to 100.0 exactly due to rounding error.

is classified as for family planning, maternal health, child health, or general health/other services. The client's age is grouped into 18–24, 25–34, and 35–49 years. Her number of own children is classified as 0–1, 2, and 3 or more. Client schooling is measured with ordinal categories taking the underlying distribution in each country into account. Across the sites, the categories have been combined into three, representing (1) never attended or attended primary school, (2) attended secondary or vocational school, and (3) attended tertiary school or higher. Household well-being is measured using client's self-report, using a Cantril-like economic ladder question (Cantril 1966) scaled 1–10 with 1 being the lowest step for the poorest and 10 being the highest step for the richest. Economic ladder categories are (1) poorest (steps 1–3), (2) poorer (step 4), (3) richer (step 5), and (4) richest (steps 6–10). Marital status is classified as married or living with someone and not in-union, with the latter including clients who are single/never married, divorced, or widowed.

Analyses

We first describe the composition of the client samples to assess the compositional heterogeneity across our six countries in Table 1. We then compare two outcome measures of the female client's self-reported contraceptive use at the baseline in-person interview (T1c) with her retrospective report of that use (T1r) at the follow-up telephone interview to examine

consistency. The first outcome (Table 2) is the specific method she reports (sterilization, IUD, implant, injectable, pill, emergency contraception, condom, other modern or traditional method, and none), and the second outcome (Table 3) combines those methods into four categories (permanent/sterilization, long-acting, short-acting, and nonuse). These comparisons are made separately for each of the six sites panels. By ordering and grouping contraceptive methods by their effectiveness, we further test the sensitivity of reporting consistency.

We compute Cohen's kappa statistic comparing the distribution between T1c and T1r to assess response agreement. This coefficient accounts for the chance of agreement in a nominal variable by computing the proportion of agreement after chance agreement is removed. The difference in the proportion of units in which women's reporting agreed minus the proportion of units for which agreement is expected by chance, is divided by the proportion of units for which agreement is not expected by chance (Cohen 1960). The strength of agreement of a kappa statistic can be interpreted as almost perfect if the statistic falls between 0.81 and 1.00, substantial between 0.61 and 0.80, moderate between 0.41 and 0.60, fair between 0.21 and 0.40, and slight between 0.00 and 0.20 (Landis and Koch 1977).

To assess the percent agreement in *change* over the four months, we next compare the client's baseline report of current use with her follow-up report of current use at time 2 (time 1c with time 2, or T1c-T2) using the four-method category (Table 4, left panel). We also compare her retrospective recall of the baseline method reported at the follow-up interview with the method she reports currently using at the follow-up interview (time 1r with time 2, or T1r-T2), again with the four-method-category (Table 4, right panel). This comparison by specific methods is available in Online Appendix Tables T1 and T2.

We then assess consistency in reported *change* based on continuity or shift in the client's use *status* by comparing T1c-T2 against T1r-T2. We test the level of agreement between these two distributions with Cohen's kappa (Table 5). The last analysis examines the percent agreement (and 95 percent confidence intervals) in use status comparing T1c-T2 against T1r-T2 (Table 6) across the six urban samples. These percentages are based on pairwise agreement among subsamples of clients who report being in the same use status at both times (T1c-T2 and T1r-T2) or at least one of them (T1c-T2 or T1r-T2).

RESULTS

A total of 9,390 female clients between the ages of 18 and 49, who consented to follow-up, were reached by phone and had complete information on contraceptive use consistency between and within follow-up and baseline. The client analysis samples across the countries are: DR Congo 755, India 653, Kenya 3,940, Niger 515, Nigeria 2,324, and Burkina Faso 1,198.

The percent distribution of urban contraceptive users by method at baseline varies by country sample (see Figure 2). In India, the majority of women rely on sterilization, 42 percent for female sterilization and less than one percent for male sterilization, followed by male condoms reported by 32 percent of users. In Kenya and Nigeria, most users rely on injectables, 46 and 30 percent, respectively, while in Niger most clients use the pill (57 percent). However, the second method most used is different for these countries. In Kenya 31 percent and in Niger 22 percent are implant users, in Nigeria 26 percent are male condom

TABLE 2 Percent agreement between contraceptive method reported at baseline (T1c) and method reported retrospectively at follow-up (T1r) by urban female clients in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso

Method reported at baseline (T1c)	Method reported retrospectively at follow-up (T1r)									Total percent agreement	Cohen kappa statistic and SE
	Sterilization	IUD	Implant	Injectable	Oral pill	Emergency contraception	Condom	Other methods*	Nonuse		
DR Congo (n = 755)											
Sterilization	**0.53**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	59.07	0.337 [0.025]
IUD	0.00	**0.00**	4.11	0.79	0.13	0.26	0.93	0.00	2.38		
Implant	0.00	0.53	**0.00**	0.13	0.26	0.00	0.00	0.00	0.00		
Injectable	0.00	0.00	0.13	**3.18**	0.66	0.53	0.40	0.00	2.12		
Oral Pill	0.00	0.00	0.26	0.00	**1.59**	0.53	0.66	0.00	1.72		
Emergency contraception	0.00	0.00	0.13	0.00	0.40	**2.91**	0.26	0.00	3.84		
Condom	0.00	0.00	0.00	0.26	0.40	0.40	**5.03**	0.53	7.28		
Other Methods*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.40**	0.26		
Nonuse	0.13	0.00	0.93	1.19	0.26	1.46	5.83	0.66	**45.43**		
India (n = 653)											
Sterilization	**15.31**	0.00	0.00	0.00	0.00	0.00	0.46	0.00	3.98	84.37	**0.753** **[0.023]**
IUD	0.00	**2.60**	0.00	0.00	0.00	0.00	0.00	0.00	0.46		
Implant	0.00	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00		
Injectable	0.00	0.00	0.00	**2.91**	0.00	0.00	0.00	0.00	0.00		
Oral pill	0.00	0.00	0.00	0.00	**4.59**	0.00	0.15	0.00	0.77		
Emergency contraception	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	0.46		
Condom	0.00	0.00	0.00	0.31	0.15	0.00	**11.03**	0.00	3.52		
Other methods*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00		
Nonuse	1.23	0.31	0.00	0.31	0.77	0.00	2.60	0.15	**47.93**		
Kenya (n = 3,940)											
Sterilization	**0.76**	0.00	0.13	0.10	0.03	0.00	0.03	0.00	0.15	67.35	**0.564** **[0.010]**
IUD	0.03	**1.73**	0.23	0.20	0.05	0.00	0.03	0.00	0.20		
Implant	0.05	0.20	**14.31**	2.74	0.79	0.03	0.05	0.00	3.50		
Injectable	0.03	0.20	2.28	**23.10**	0.96	0.03	0.41	0.05	5.03		
Oral pill	0.00	0.05	0.63	0.94	**2.94**	0.03	0.10	0.00	0.94		
Emergency contraception	0.00	0.00	0.13	0.15	0.10	**0.43**	0.18	0.00	0.36		
Condom	0.00	0.00	0.18	0.76	0.08	0.08	**2.66**	0.03	1.60		
Other methods*	0.00	0.00	0.00	0.00	0.03	0.00	0.00	**0.05**	0.03		
Nonuse	0.10	0.36	2.51	3.17	1.19	0.00	0.79	0.66	**21.37**		

(Continued on next page)

TABLE 2 (Continued)

Method reported at baseline (T1c)	Method reported retrospectively at follow-up (T1r)									Total percent agreement	Cohen kappa statistic and SE
	Sterilization	IUD	Implant	Injectable	Oral pill	Emergency contraception	Condom	Other methods*	Nonuse		
Niger (n = 515)										75.34	**0.631** [**0.028**]
Sterilization	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.58		
IUD	0.00	**1.55**	0.00	0.39	0.00	0.00	0.00	0.00	0.58		
Implant	0.00	0.00	**8.54**	0.19	0.58	0.00	0.00	0.00	2.91		
Injectable	0.00	0.00	0.19	**5.44**	0.97	0.00	0.00	0.00	2.14		
Oral pill	0.00	0.58	0.97	0.19	**22.72**	0.00	0.00	0.19	8.16		
emergency contraception	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00		
Condom	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.19		
Other methods*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00		
Nonuse	0.19	0.19	0.39	0.39	4.08	0.00	0.00	0.38	**37.09**		
Nigeria (n = 2,324)										71.87	**0.471** [**0.017**]
Sterilization	**0.09**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04		
IUD	0.00	**0.60**	0.00	0.09	0.00	0.04	0.04	0.00	0.65		
Implant	0.00	0.04	**4.95**	0.26	0.17	0.00	0.04	0.00	1.64		
Injectable	0.00	0.00	0.39	**6.45**	0.73	0.04	0.09	0.00	3.10		
Oral pill	0.00	0.04	0.13	0.39	**2.24**	0.00	0.26	0.00	2.50		
Emergency contraception	0.00	0.00	0.04	0.04	0.04	**0.09**	0.04	0.00	0.77		
Condom	0.00	0.00	0.09	0.13	0.22	0.04	**3.23**	0.00	5.68		
Other methods*	0.00	0.00	0.00	0.00	0.04	0.00	0.00	**0.00**	0.30		
Nonuse	0.04	0.56	1.51	4.26	1.59	0.17	1.85	0.04	**54.22**		
Burkina Faso (n = 1,198)										74.12	**0.633** [**0.018**]
Sterilization	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
IUD	0.00	**2.75**	0.08	0.08	0.17	0.00	0.08	0.00	0.42		
Implant	0.00	0.17	**11.69**	0.42	0.25	0.00	0.25	0.00	2.25		
Injectable	0.00	0.00	0.75	**10.85**	0.42	0.00	0.17	0.00	2.42		
Oral Pill	0.00	0.08	0.50	0.58	**5.09**	0.00	0.25	0.00	2.67		
Emergency contraception	0.00	0.00	0.50	0.00	0.08	**0.17**	0.08	0.00	0.00		
condom	0.00	0.25	0.00	0.17	0.33	0.00	**5.34**	0.00	2.75		
Other methods*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.17		
Nonuse	0.00	0.67	1.92	1.50	1.84	0.17	3.42	0.00	**38.23**		

NOTE: Cell percentages sum to 100.0 within country panels. * Other Methods include other modern and traditional methods. Bold values indicate a moderate or higher kappa statistic. Strength of agreement: almost perfect [0.81–1.00]; substantial [0.61–0.80]; moderate [0.41–0.60]; fair [0.21–0.40]; slight [0.00–0.20].

TABLE 3 Percent agreement between contraceptive method type reported at baseline (T1c) and method type reported retrospectively at follow-up (T1r) by urban female clients in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso

Method reported at baseline (T1c)	Method reported retrospectively at follow-up (T1r)				Total percent agreement	Cohen kappa statistic and SE
	Sterilization	Long-acting method	Short-acting method*	Nonuser		
DR Congo (*n* = 755)						
Sterilization	**0.53**	0.00	0.00	0.13	68.75	**0.420 [0.031]**
Long-acting method	0.00	**4.64**	2.52	2.38		
Short-acting method*	0.00	0.53	**18.15**	15.23		
Nonuser	0.13	0.93	9.40	**45.43**		
India (*n* = 653)						
Sterilization	**15.31**	0.00	0.46	3.98	84.98	**0.753 [0.023]**
Long-acting method	0.00	**2.60**	0.00	0.46		
Short-acting method*	0.00	0.00	**19.14**	4.75		
Nonuser	1.23	0.31	3.83	**47.93**		
Kenya (*n* = 3,940)						
Sterilization	**0.76**	0.13	0.15	0.15	71.70	**0.567 [0.011]**
Long-acting method	0.08	**16.47**	3.86	3.71		
Short-acting method*	0.03	3.48	**33.10**	7.94		
Nonuser	0.10	2.87	5.81	**21.37**		
Niger (*n* = 515)						
Sterilization	**0.00**	0.00	0.00	0.58	76.70	**0.619 [0.030]**
Long-acting method	0.00	**10.10**	1.17	3.50		
Short-acting method*	0.00	1.75	**29.51**	10.49		
Nonuser	0.19	0.58	4.66	**37.09**		
Nigeria (*n* = 2,324)						
Sterilization	**0.09**	0.00	0.00	0.04	73.97	**0.468 [0.018]**
Long-acting method	0.00	**5.59**	0.65	2.28		
Short-acting method*	0.00	0.69	**14.07**	12.35		
Nonuser	0.04	2.07	7.92	**54.22**		
Burkina Faso (*n* = 1,198)						
Sterilization	**0.00**	0.00	0.00	0.00	39.48	**0.623 [0.020]**
Long-acting method	14.69	**1.25**	0.00	2.67		
Short-acting method*	2.09	23.54	**0.00**	8.01		
Nonuser	2.59	6.93	0.00	**38.23**		

NOTE: Cell percentages sum to 100.0 within country panels. * Short-acting methods include traditional methods. Total percent nonagreement is sum of off-diagonal values. Bold values indicate a moderate or higher kappa statistic. Strength of agreement: almost perfect [0.81–1.00]; substantial [0.61–0.80]; moderate [0.41–0.60]; fair [0.21–0.40]; slight [0.00–0.20].

users. In Burkina Faso, clients reported using IUDs and injectables about equally 28–29 percent. Contraceptive users in the DR Congo have a different method distribution than in the other five countries, with the majority relying on male condoms (32 percent) and a smaller fraction on IUDs (20 percent). The method choice distribution, which will also be affected by facility supply, can play a role in continuity and consistency of reported use.

In Table 1 we present the composition of the six panel samples of female clients according to selected client characteristics—reason for facility visit, type of facility, age group, parity group, schooling, household wealth, and marital status. The percentages of urban clients to present for family planning are relatively small, ranging from 4.1 percent in the DR Congo to 16.1 percent in Niger samples. Most clients present for general health or child health, followed by maternal health reasons. Most clients were interviewed at hospitals and public health clinics in the DR Congo, Kenya, and Niger samples, while private health clinics were a major source of client interviews in Nigeria and Burkina Faso samples. A high proportion of clients were interviewed at pharmacies in the Indian cities. The age composition of clients was relatively equal across sites with about half of females being 25–34 years of age. Clients had a higher number of children (three or more) in Niger (46.7 percent), Nigeria (40.2 percent), and Kenya (39.4 percent) with those attending in Burkina Faso, DR Congo, and India having

TABLE 4 Percent agreement in change in contraceptive method reported at baseline (T1c), retrospectively reported at follow-up (T1r) and method reported being used at follow-up (T2) by urban female clients in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso

	Method reported currently used at follow-up (T2)							
	Reported at baseline (T1c)				Retrospectively reported at follow-up (T1r)			
	Sterilization	Long-acting method	Short-acting method*	Nonuser	Sterilization	Long-acting method	Short-acting method*	Nonuser
DR Congo (n = 755)								
Sterilization	**0.53**	0.00	0.00	0.13	**0.66**	0.00	0.00	0.00
Long-acting method	0.00	**4.50**	2.52	2.52	0.00	**5.43**	0.13	0.53
Short-acting method*	0.00	1.46	**17.35**	15.10	0.00	0.66	**22.25**	7.15
Nonuser	0.13	1.72	12.72	**41.32**	0.00	1.59	10.20	**51.39**
India (n = 653)								
Sterilization	**18.53**	0.00	0.46	0.77	**16.54**	0.00	0.00	0.00
Long-acting method	0.15	**1.53**	0.15	1.23	0.15	**1.68**	0.15	0.92
Short-acting method*	0.77	0.77	**13.48**	8.88	0.61	0.46	**16.23**	6.13
Nonuser	3.52	0.46	6.58	**42.73**	5.67	0.61	4.29	**46.55**
Kenya (n = 3,940)								
Sterilization	**0.86**	0.18	0.10	0.05	**0.96**	0.00	0.00	0.00
Long-acting method	0.08	**16.09**	4.01	3.93	0.00	**19.44**	1.42	2.08
Short-acting method*	0.08	4.80	**30.91**	8.76	0.03	2.72	**34.49**	5.69
Nonuser	0.13	4.11	7.84	**18.07**	0.15	3.02	6.95	**23.05**
Niger (n = 515)								
Sterilization	**0.19**	0.00	0.00	0.39	**0.19**	0.00	0.00	0.00
Long-acting method	0.00	**8.93**	1.94	4.08	0.00	**9.90**	0.58	1.94
Short-acting method*	0.00	1.94	**24.47**	15.34	0.00	0.39	**26.60**	8.74
Nonuser	0.19	0.78	6.21	**35.53**	0.19	1.36	5.44	**44.66**
Nigeria (n = 2,324)								
Sterilization	**0.09**	0.00	0.00	0.04	**0.13**	0.00	0.00	0.00
Long-acting method	0.00	**5.25**	1.03	2.24	0.00	**7.14**	0.30	0.90
Short-acting method*	0.04	1.16	**15.58**	10.33	0.00	0.30	**19.97**	2.37
Nonuser	0.04	2.71	10.93	**50.56**	0.04	1.68	7.27	**59.90**
Burkina Faso (n = 1,198)								
Sterilization	**0.00**	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00
Long-acting method	6.59	**5.18**	2.42	4.42	8.10	**6.09**	1.00	4.17
Short-acting method*	1.25	0.83	**18.70**	12.85	0.58	0.25	**21.95**	8.93
Nonuser	2.75	2.50	8.51	**33.97**	1.92	2.17	6.68	**38.15**

NOTE: Cell percentages sum to 100.0 within country panels. * Short-acting methods include traditional methods.

TABLE 5 Percent agreement in change in contraceptive use status as reported for T1c-T2 (baseline to follow-up) compared to T1r-T2 (within follow-up), among urban female clients in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso

T1c-T2 scenario	Total	T1r-T2 scenario					Total percent agreement	Cohen kappa statistic and SE
		User	Nonuser	Discontinuer	Switcher	Adopter		
DR Congo (*n* = 755)								
Total	100.0	24.6	51.4	7.7	4.5	11.8	64.8	0.508 [0.022]
User	15.1	11.9	0.0	0.0	0.1	3.1		
Nonuser	41.6	0.0	38.8	2.5	0.1	0.1		
Discontinuer	10.9	0.0	6.9	4.0	0.0	0.1		
Switcher	18.1	5.6	5.7	1.2	3.6	2.1		
Adopter	14.3	7.2	0.0	0.0	0.7	6.5		
India (*n* = 653)								
Total	100.0	33.4	46.6	7.0	2.5	10.6	84.5	0.770 [0.020]
User	32.4	28.0	0.0	0.0	0.2	4.3		
Nonuser	42.7	0.0	42.1	0.6	0.0	0.0		
Discontinuer	10.9	0.0	4.4	6.4	0.0	0.0		
Switcher	3.4	0.8	0.0	0.0	2.1	0.5		
Adopter	10.6	4.6	0.0	0.0	0.2	5.8		
Kenya (*n* = 3,940)								
Total	99.9	51.9	23.0	7.8	7.2	10.1	70.2	0.579 [0.010]
User	42.0	38.5	0.0	0.0	1.2	2.3		
Nonuser	18.1	0.0	16.0	2.1	0.0	0.0		
Discontinuer	12.7	0.0	7.1	5.7	0.0	0.0		
Switcher	15.0	7.9	0.0	0.0	4.7	2.4		
Adopter	12.1	5.4	0.0	0.0	1.3	5.4		

TABLE 5   (Continued)

T1c-T2 scenario	Total	T1r-T2 scenario					Total percent agreement	Cohen kappa statistic and SE
		User	Nonuser	Discontinuer	Switcher	Adopter		
Niger (n = 515)								
Total	100.1	29.4	65.8	3.6	1.2	0.0	76.3	**0.662 [0.025]**
User	18.2	**28.2**	0.0	0.0	0.4	3.5		
Nonuser	50.5	0.0	**34.6**	1.0	0.0	0.0		
Discontinuer	12.4	0.0	10.1	**9.7**	0.0	0.0		
Switcher	5.2	3.1	0.0	0.0	**1.4**	1.0		
Adopter	13.8	4.1	0.0	0.0	0.6	**2.5**		
Nigeria (n = 2,324)								
Total	100.1	34.4	38.1	13.1	3.6	10.8	72.0	**0.558 [0.012]**
User	27.2	**0.6**	2.9	0.0	0.1	1.6		
Nonuser	34	0.3	**15.4**	0.0	0.0	2.5		
Discontinuer	17.3	0.0	0.0	**1.9**	10.5	0.0		
Switcher	7.8	0.0	0.0	1.3	**49.2**	0.0		
Adopter	13.8	0.3	8.4	0.0	0.1	**5.0**		
Burkina Faso (n = 1,198)								
Total	100.0	35.9	44.7	10.7	1.7	7.0	75.6	**0.669 [0.016]**
User	7.6	**25.1**	0.0	0.0	0.3	1.8		
Nonuser	35.5	0.0	**30.6**	3.4	0.0	0.0		
Discontinuer	19.8	0.0	7.6	**9.7**	0.0	0.0		
Switcher	29.9	3.8	0.0	0.0	**2.6**	1.3		
Adopter	7.2	5.4	0.0	0.0	0.7	**7.7**		

NOTE: Cell percentages sum to 100.0 within country panels. Inconsistent any is defined as those clients with an unmatched consistency report between T1c-T2 scenario and T1r-T2 scenario within country panels, not in the diagonal of the matrix. (See Figure 1 for scenario definitions.) Bold values indicate a moderate or higher kappa statistic. Strength of agreement: almost perfect [0.81 – 1.00]; substantial [0.61 – 0.80]; moderate [0.41 – 0.60]; fair [0.21 – 0.40]; slight [0.00 – 0.20].

TABLE 6   Percent agreement in pairwise reported contraceptive use status between T1c-T2 (baseline to follow-up) and T1r-T2 (within follow-up) by user status among client subsamples in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso

| | Percent agreement and 95% confidence interval (CI) | | | | | | | | | |
Country	User Percentage [95% CI]	N	Nonuser Percentage [95% CI]	N	Discontinuer Percentage [95% CI]	N	Switcher Percentage [95% CI]	N	Adopter Percentage [95% CI]	N
DR Congo	0.429 [0.361–0.496]	409	0.716 [0.673–0.760]	144	0.273 [0.188–0.357]	110	0.188 [0.123–0.252]	210	0.331 [0.254–0.408]	148
India	0.741 [0.686–0.796]	308	0.893 [0.858–0.928]	24	0.560 [0.445–0.675]	75	0.583 [0.371–0.796]	247	0.380 [0.283–0.477]	100
Kenya	0.694 [0.675–0.713]	990	0.636 [0.606–0.666]	692	0.384 [0.344–0.423]	584	0.267 [0.234–0.300]	2184	0.320 [0.284–0.355]	663
Niger	0.718 [0.655–0.780]	235	0.757 [0.702–0.813]	33	0.467 [0.371–0.563]	107	0.212 [0.065–0.359]	202	0.217 [0.109–0.324]	60
Nigeria	0.523 [0.485–0.560]	1422	0.804 [0.783–0.824]	133	0.141 [0.102–0.179]	320	0.098 [0.047–0.149]	685	0.278 [0.234–0.321]	414
Burkina Faso	0.689 [0.645–0.732]	498	0.735 [0.696–0.774]	105	0.468 [0.405–0.530]	248	0.295 [0.207–0.384]	437	0.455 [0.386–0.525]	202

NOTE: Clients subsampled if they reported user status consistently at both times (T1c-T2 = T1r-T2) or inconsistently at either time (T1c-T2 or T1r-T2).

FIGURE 2   Method mix among urban contraceptive users at T1c, T1r, and T2 in DR Congo, India, Kenya, Niger, Nigeria, and Burkina Faso
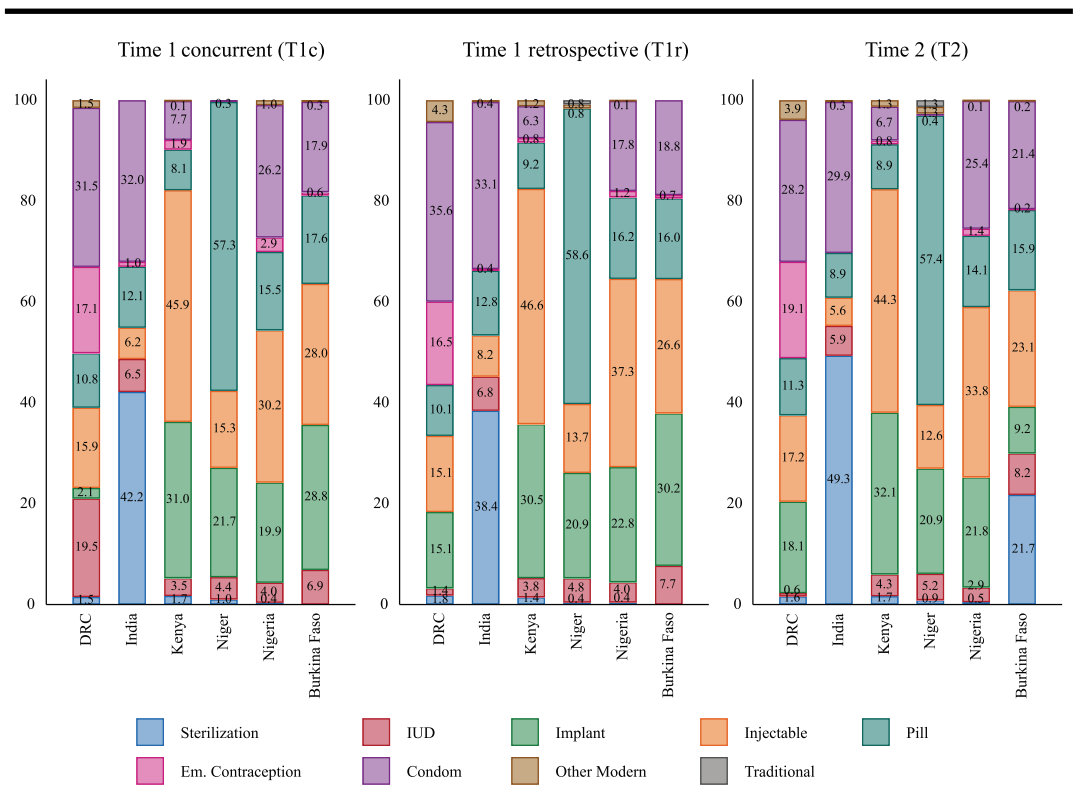
smaller numbers of 0–1 children, 48.9, 43.4, and 43.0 percent respectively. A higher proportion of clients had primary or no schooling levels in Kenya (39.5 percent) and Niger (40.1 percent), with those having postsecondary schooling highest in the Nigeria sample

(41.7 percent). A majority of the samples was married or in-union, ranging from 92.2 percent in India to 65.3 percent in DR Congo.

Table 2 presents the percent agreement between clients' reports of the method they currently use (T1c), and they retrospectively recalled (T1r) by site. Each panel sums to 100.0 percent with the percentage values on the diagonal indicating agreement, and the off-diagonal elements disagreement. Nonuse tends to have the highest agreement, ranging from 54.22 percent in the Nigeria sample to 47.93 percent (India), to 45.43 percent (DR Congo), to 38.23 percent (Burkina Faso), to 37.09 percent (Niger) to a low of 21.37 percent (Kenya). Low use of sterilization is reflected in small percentage agreements, except in India (15.31 percent), and a similar pattern is observed for IUDs. Consistent reporting on implant use is observed in Kenya (14.31 percent), Burkina Faso (11.69 percent), and Niger (8.54 percent) samples, with consistency in injectable reporting also high in the Kenya (23.10 percent) and Burkina Faso (10.85 percent) samples. Oral pill use is reported consistently by client users in Niger (22.72 percent) and condom in India (11.03 percent). The total percent agreement, based on the sum of the diagonal values, is not perfect, despite the short interval, ranging from 59.07 percent in DR Congo to 84.37 percent in India.

These patterns are reinforced in comparing agreement by client reporting using the four method type categories as seen in Table 3. Disagreement is observed, however, with the larger percentages for DR Congo, Niger, and Nigeria clients reporting short-acting method use in T1c but nonuse in T1r, that is, 15.23, 10.49, and 12.35 percent respectively. Likewise, high off-diagonal values are seen for Burkina Faso clients reporting using long-acting methods in T1c and sterilization in T1r (14.69 percent) as well as short-acting methods in T1c and long-acting methods in T1r (23.54 percent).

With the exception of the DR Congo, the kappa values in Table 2 have a moderate or better strength of agreement and reveal varying levels of consistent reporting of the specific baseline method, ranging from a high value of 0.753 in the India sample to a low of 0.337 for DR Congo. Next highest are 0.633 and 0.631 for Burkina Faso and Niger, respectively, with 0.564 for Kenya and 0.471 for Nigeria. The kappa values in Table 3 for consistent reporting by method type are very similar to those in Table 2 with the DR Congo value of 0.420 being moderate this time.

Table 4 addresses the question of how much short-term change is observed in contraceptive method use, based on four categories, when judged by the baseline method as currently (T1c) or retrospectively (T1r) reported. The left panel of columns compares method type from T1c to T2 and the right panel from T1r to T2. Examining the percent agreement values on the diagonals for each country, we see greater agreement in the T1r-T2 reports than T1c-T2 ones. This suggests that clients are more likely to report to the interviewer that they are still using the same method at follow-up as recalled for baseline, than when compared to the method they reported at baseline. Because nonuse values are smaller in the T1c-T2 comparison (left panel), this implies clients are reporting more current use at baseline than four months later. A source of the shift is from baseline reports of short-acting method use to subsequent nonuse; for example, in the T1c-T2 comparison, we see percentages of 15.1 percent (DR Congo), 8.9 percent (India), 8.8 percent (Kenya), 15.3 percent (Niger), 10.3 percent (Nigeria), and 12.9 percent (Burkina Faso) for clients reporting discontinuation. The same percentages based on the T1r-T2 comparison are lower, that is, 7.2, 6.1, 5.7, 8.7, 2.4, and 8.9 percent respectively.

The extent of change in contraceptive behavior may not be large over four months, especially for clients reporting use of permanent or long-acting methods. However, given research showing significant discontinuation of short-acting methods (Ali, Cleland, & Shah, 2012) or of increased contraceptive availability (Ahmed et al., 2019), we expect to observe contraceptive switching, adoption and discontinuation even after four months. We are interested in assessing both the extent of T1c-T2 and T1r-T2 change as well as how similar or consistent the reported change is. If the T1c-Tr comparison is highly consistent, we would expect T1c-T2 and T1r-T2 to also be consistent.

Table 5 examines change and consistency in reported use and nonuse, discontinuation, adoption, or switching by comparing the T1c-T2 and T1r-T2 scenarios. First, we discuss their marginal distributions as these reflect measured change, and then we examine their internal agreement. We see that based on panel reports (T1c-T2) a smaller percentage of clients are classified as continuing users, and a larger percentage as continuing nonusers, than when assessed in the cross section (T1r-T2). For example, 15.1 percent of the client panel in DR Congo reported themselves as users both in T1c and T2, whereas 24.6 percent reported accordingly based on the T1r and T2 comparison. For nonusers, the percentage reporting to be nonusers at both times is 41.6 percent based on T1c-T2 and 51.4 percent based on T1r-T2. This pattern holds true for all country panels except India where the difference for users and nonusers is small. The largest gap among continuing users is registered in Burkina Faso, 28.3 percentage points, and in Niger among nonusers, 15.3 percentage points. A similar pattern is observed for reporting of change in baseline method between T1c-T2 and T1r-T2 (Table 4). It appears that although only separated by about four months, retrospective reporting of one's contraceptive use results in higher prevalence than when based on actual in-time, longitudinal reporting. Longitudinal data measurement also appears to result in greater percentages reporting discontinuation, switching and adoption than within a cross-sectional round. In terms of cross-site variation, greater change in discontinuation is seen in Burkina Faso and Nigeria samples, while for switching, levels are higher in the DR Congo, Kenya, and Burkina samples.

Consistency in reported use status between T1c-T2 and T1r-T2 is observed with the percent agreement findings in Table 5. For DR Congo, 38.8 percent of the client panel consistently reported being a nonuser over time and 11.9 percent a user, with 4.0, 3.6, and 6.5 percent reporting being a discontinuer, switcher, or adopter. These total 64.8 percent, with 35.2 percent inconsistently reporting their change over four months. For example, 6.9 percent of DR Congo clients are classified as discontinuers based on T1c-T2 and as nonusers based on T1r-T2. The total percent disagreement (off diagonal totals) ranges from the high of 35.2 percent for DR Congo to a low of 15.5 percent (India) and between 23.7 percent for Niger and 29.8 percent for Kenya. Many of the off-diagonal values reflect client reports of shifts toward nonuse (discontinuation), adoption and switching whereas they retrospectively report themselves as users or nonusers at baseline. The degree of consistency in change reported with the two distributions is also revealed in the kappa values, all of which have a strength of agreement of moderate or better. These follow a similar pattern to the kappa values seen in Table 3, with the highest observed for the India sample (0.770), followed by Burkina Faso (0.669) and Niger (0.619), and the lowest in the DR Congo (0.420).

In Table 6, we further our examination of how consistent the reported change is by looking individually at the percent agreement for the five user statuses and across sites. The

percent agreement is tabulated for clients classified in each of the five user statuses either in the T1c-T2 or T1r-T2 distributions. This avoids biasing the percent agreement upward with the many remaining clients who will consistently not be in most of the pairwise comparisons. The results affirm again that nonusers have the most consistent reporting, with the highest percent agreement values, ranging from 0.893 in India to 0.636 in Kenya. After nonusers, clients who are users show the next highest percent agreement, ranging from 0.741 in India to 0.429 in DR Congo. Discontinuation also tends to be consistently reported either in the T1c-T2 or T1r-T2 change measure in India, Niger and Burkina Faso but least in Nigeria. The percent agreement for switching or adopting between the two change distributions is weak to modest across all sites, except India.

DISCUSSION

Survey interviews are the principal mode of data collection for individual-level contraceptive practice in low-income countries (Khan et al. (2007). The validity and reliability of women's reports of their status as users are critical to obtaining an accurate profile of their pregnancy management behaviors, not only for program planning purposes but also for scientific measurement and understanding. Our descriptive study documents that females interviewed four to six months apart, first upon exit from a health facility visit and then by phone, provide two different profiles of consistency in reported baseline method and subsequent change in use status. When we examined reported change across T1c to T2 rounds and then by T1r to T2, the former revealed more status changes for users than the latter comparison, except in India. For T1r-T2 transitions, being a continuing nonuser outsized most other categories, raising the question of whether use at baseline was more accurately reported under face-to-face circumstances than four months later by telephone. We next compared the reporting consistency of these two scenarios (T1c-T2 vs. T1r-T2). Only 0.6 percent of Nigerian female clients consistently reported themselves as contraceptive users both in T1c-T2 and T1r-T2, with the counterpart percentages being 11.9 percent for DR Congo, 25.1 percent for Burkina Faso, 28.0 percent for India, 28.2 percent for Niger, and 38.5 percent for Kenyan clients. Our measures of consistent reporting were highest in India and lowest in DR Congo overall. This is partially explained by the high use of sterilization among contraceptives users in India and the high use of short-term contraceptive methods among contraceptive users in DR Congo.

Because we have observed nontrivial amounts of inconsistent reporting, this suggests continuing attention to measurement reliability. Inconsistency may be the result of unwillingness to disclose methods at the baseline interview, misreporting or other recall bias. Our findings have the following implications for survey measurement of contraceptive behavior: (1) reports of nonuse are likely to be reliable; (2) currently obtained reports of use are next most likely to be reliable; (3) change in use is more robust in a panel study than when assessed with retrospective reporting; and (4) the extent of change in starting, stopping, and switching methods is dependent on any change being measured reliably. This third finding has implications in the reliability of the information collected in the contraceptive calendar, which is widely used to measure contraceptive use dynamics in low-income countries.

Our study design offers a number of strengths, including its demonstration of the feasibility of interviewing urban female health clients about contraception by phone in low-income

countries. A large share, 71 percent, of the female clients were successfully reinterviewed, after being consented for follow-up, providing contact telephone numbers (either theirs or a family member's phone number), and then being reached and completing the interview by phone four to six months after the baseline interview. Another study strength is its longitudinal design that enabled assessing contraceptive use change and in particular the rate of adoption and discontinuation for some country sites. Across the six countries between 2.5 and 7.7 percent consistently reported going from nonuse at baseline to adopting a contraceptive method at follow-up, and between 1.9 and 9.7 percent consistently reported stopping use. In addition, while studies have monitored fertility preferences in low-income countries at short intervals, for example, Sennott and Yeatman (2012), to our knowledge there are no longitudinal studies conducted in low-income countries assessing shifts in women's contraceptive use with short intervals of time between surveys.

The inconsistent reporting of one's contraceptive status just four months earlier raises the question of which report to accept as the actual situation—the one at baseline or at follow-up. Levels of change were observed to be higher between baseline and follow-up rounds than within the follow-up round, suggesting that inaccurate recall of use four months earlier may be a factor. We cannot eliminate the possibility that the mode of survey administration, face-to-face or telephone, was a factor. It would also be important to establish if the baseline method was intentionally or unintentionally misreported as we are unable to differentiate change from misreporting.

Our study has its limitations, one being the limited generalizability of the baseline samples to all females seeking health care services. While systematic random sampling protocols were in place, their full implementation could not be assessed because the number of clients approached and recruited to participate was not uniformly recorded, limiting our ability to record accurately client flows per facility. Thus, we were unable to calculate client selection probabilities and weight the data accordingly. A second possible source of measurement bias was that contraceptive use at baseline was captured differently for family planning clients than for clients visiting for other services. The family planning client's reported method prescribed or dispensed was assigned as her current contraceptive method, while this was directly asked to other clients. However, the percentage of clients seeking family planning services at baseline was less than 16 percent in all countries. A third limitation is the low follow-up rates in India (59 percent) and DR Congo (68 percent) due to phone access and the inability to reach and reinterview many women by phone, which can potentially introduce selection bias to our estimates. Female clients in India reported low phone access, for example, phones were owned by their husbands and landline phones were scarce, while clients in DR Congo were difficult to recontact. Nonetheless, the other country samples had recontact rates of 73 percent or higher and the ability to reinterview the same female clients adds to the validity of the results. Last, our analysis does not attempt to assess whether the mode of survey interview affected response reliability, as interviews were all face-to-face at baseline and by phone at follow-up. This can affect the assessed consistency of self-reported contraceptive use. Respondents could be more or less accurate about their use four months earlier or when followed up.

Our study's focus has been on the accuracy of clients' self-reported contraceptive use and assessing consistency and change both between baseline and follow-up and within follow-up rounds across six urban samples. By doubly comparing self-reported use under the two

scenarios, our findings add to the research literature on reporting consistency of sexual and reproductive behaviors. The literature has largely focused on the reliability of reported sexual behaviors across panel rounds or waves, for example, Goldberg et al. (2014), Alexander et al. (1993), Sieving et al. (2005), or between cross-sectional samples and calendar-based retrospective reports (Strickler et al. 1997; Callahan and Becker 2012) but not when retrospectively recalled within a round. The study's cross-national samples, while not generalizable to the population of health-seeking females at large, are robust in size and exhibit patterns that would be expected in these settings. For example, consistency in being a continuing nonuser in India comports with the high proportion of client interviews taking place at pharmacies. Given the predominant method in the country is female sterilization, women are unlikely to be obtaining this method at such locations and their spouses are likely purchasing the condoms. Similarly, low contraceptive prevalence in Niger, Nigeria, and DR Congo indicates that most women are not using and thus there is a higher probability of continued nonuse being consistently reported by female clients both at four months and within the follow-up survey.

Potential misreporting raises concerns of validity and reliability of measured indicators (Nock, Zeller, and Carmines 1982). The inconsistency in reporting that we observe, which may well have acceptable reasons behind it, has implications for studying contraceptive use dynamics specifically and sensitive behaviors generally. Panel data are invaluable for studying the determinants of behavioral change. At the same time, continued evaluation of the validity of measures and indicators that are of significant scientific and programmatic importance is warranted.

CONFLICT OF INTEREST

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or nonfinancial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

AUTHOR'S CONTRIBUTIONS

AOT and PA conceived the idea for this work, obtained funding, conceptualized the study methods, reviewed analyses, and edited the manuscript. CC conceptualized the study methods, cleaned data, conducted analyses, reviewed analyses, wrote the first draft of the manuscript, and edited it. VS cleaned data and conducted analyses. FO, EO, AK, PG, MT, SR reviewed the manuscript and are co-principal investigators of PMA Agile. All authors reviewed and provided substantial input to revisions. AOT and CC had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

ETHICS APPROVAL STATEMENT

All survey data used for analysis were de-identified. Ethical approval was not sought for this analysis of secondary data.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from PMA Agile (www.pmadata.org/technical-areas/pma-agile) upon request.

REFERENCES

Ahmed, Saifuddin, Yoonjoung Choi, Jose G. Rimon, Souleymane Alzouma, Peter Gichangi, Georges Guiella, Peter Kayembe, Simon P. Kibira, Fredrick Makumbi, Funmilola OlaOlorun, Elizabeth Omoluabi, Easmon Otupiri, Sani Oumarou, Assefa Seme, Solomon Shiferaw, Philip Anglewicz, Scott Radloff, and Amy Tsui. 2019. "Trends in Contraceptive Prevalence Rates in Sub-Saharan Africa Since the 2012 London Summit on Family Planning: Results from Repeated Cross-Sectional Surveys." *The Lancet Global Health* 7(7): e904-e911.

Alexander, Cheryl S., Mark R. Somerfield, Margaret E. Ensminger, Karin E. Johnson, and Young J. Kim. 1993. "Consistency of Adolescents' Self-Report of Sexual Behavior in a Longitudinal Study." *Journal of Youth and Adolescence* 22 (5).

Ali, Mohamed M., John G. Cleland, and Iqbal H. Shah. 2012. *Causes and Consequences of Contraceptive Discontinuation: Evidence from 60 Demographic and Health Surveys*. Geneva: World Health Organization.

Anderson, John E., and John G. Cleland. 1984. "The World Fertility Survey and Contraceptive Prevalence Surveys: A Comparison of Substantive Results." *Studies in Family Planning* 15 (1): 1–13.

Bignami-Van Assche, Simona. 2003. "Are We Measuring What We Want to Measure? Individual Consistency in Survey Response in Rural Malawi." *Demographic Research Special* 1: 77–108. https://doi.org/10.4054/DemRes.2003.S1.3.

Callahan, Rebecca L., and Stan Becker. 2012. "The Reliability of Calendar Data for Reporting Contraceptive Use: Evidence from Rural Bangladesh." *Studies in Family Planning* 43(3): 213–222. https://doi.org/10.1111/j.1728-4465.2012.00319.x. https://pubmed.ncbi.nlm.nih.gov/23185864

Cantril, Hadley. 1966. *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scale." *Educational and Psychological Measurement* 20(1): 37–46.

Eggleston, Elizabeth, Joan Leitch, and Jean Jackson. 2000. "Consistency of Self-Reports of Sexual Activity among Young Adolescents in Jamaica." *International Family Planning Perspectives* 26(2): 79–83.

Goldberg, Shoshana K., Anigail A. Haydon, Amy H. Herring, and Carolyn T. Halpern. 2014. "Longitudinal Consistency in Self-Reported Age of First Vaginal Intercourse among Young Adults." *Journal of Sex Research* 51(1): 97–106. https://doi.org/10.1080/00224499.2012.719169. https://www.ncbi.nlm.nih.gov/pubmed/23237101.

Goldman, Noreen, Lorenzo Moreno, and Charles F. Westoff. 1989. "Collection of Survey Data on Contraception: An Evaluation of an Experiment in Peru." *Studies in Family Planning* 20(3): 147–157.

Hamilton, Deven T., and Martina Morris. 2010. "Consistency of Self-Reported Sexual Behavior in Surveys." *Archives of Sex Behavior* 39 (4): 842–60. https://doi.org/10.1007/s10508-009-9505-7. https://www.ncbi.nlm.nih.gov/pubmed/19588240.

Harrison, Lana, and Arthur Hughes. 1997. "The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates." US Department of Health and Human Services, National Institutes of Health, National Institute on Drug Abuse, Division of Epidemiology and Prevention Research. 167.

Hewett, Paul C., Barbara S. Mensch, and Annabel S. Erulkar. 2004. "Consistency in the Reporting of Sexual Behaviour by Adolescent Girls in Kenya: A Comparison of Interviewing Methods." *Sexually Transmitted Infections* 80(Suppl 2): ii43-8. https://doi.org/10.1136/sti.2004.013250. https://www.ncbi.nlm.nih.gov/pubmed/15572639.

Karp, Celia, Shannon N Wood, Georges Guiella, Peter Gichangi, Suzanne O Bell, Philip Anglewicz, Elizabeth Larson, Linnea Zimmerman, and Caroline Moreau. 2021. "Contraceptive Dynamics during COVID-19 in Sub-Saharan Africa: Longitudinal

Evidence from Burkina Faso and Kenya." *BMJ Sexual & Reproductive Health*. Published Online First: February 12, 2021. https://doi.org/10.1136/bmjsrh-2020-200944.

Khan, Shane, Vinod K. Mishra, Fred Arnold, and Noureddine Abderrahim. 2007. "Contraceptive Trends in Developing Countries." DHS Comparative Reports No. 16. Calverton, MD: Macro International Inc.

Kimberlin, Carole L., and Almut G. Winterstein. 2008. "Validity and Reliability of Measurement Instruments Used in Research." *American Journal of Health-System Pharmacy* 65(23): 2276–2284. https://doi.org/10.2146/ajhp070364. https://www.ncbi.nlm.nih.gov/pubmed/19020196.

Knodel, John, and Sauvaluck Piampiti. 1977. "Response Reliability in a Longitudinal Survey in Thailand." *Studies in Family Planning* 18(9): 55–66.

Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33(1): 159–174.

Lauritsen, Janet L., and C. Gray Swicegood. 1997. "The Consistency of Self-Reported Initiation of Sexual Activity." *Family Planning Perspectives* 29(5): 215–221.

Mahfoud, Ziyad, Lilian Ghandour, Blanche Ghandour, Ali H. Mokdad, and Abla M. Sibai. 2014. "Cell Phone and Face-to-face Interview Responses in Population-based Surveys." *Field Methods* 27(1): 39–54. https://doi.org/10.1177/1525822X14540084.

Moreau, Caroline, K. Hall, James Trussell, and Jennifer Barber. 2013. "Effect of Prospectively Measured Pregnancy Intentions on the Consistency of Contraceptive Use among Young Women in Michigan." *Human Reproduction* 28(3): 642–650. https://doi.org/10.1093/humrep/des421. https://www.ncbi.nlm.nih.gov/pubmed/23241838.

Nock, Steven L., Richard A. Zeller, and Edward G. Carmines. 1982. "Measurement in the Social Sciences: The Link Between Theory and Data." *Contemporary Sociology* 11(1): 79–80. https://doi.org/10.2307/2066656.

Pebley, Anne R., Noreen Goldman, and Minja Kim Choe. 1986. "Evaluation of Contraceptive History Data in the Republic of Korea." *Studies in Family Planning* 17(1): 22–35.

Rosenbaum, Janet E., Jonathan M. Zenilman, Eve Rose, Gina M. Wingood, and Ralph J. DiClemente. 2017. "Semen Says: Assessing the Accuracy of Adolescents' Self-Reported Sexual Abstinence Using a Semen Y-Chromosome Biomarker." *Sexually Transmitted Infections* 93(2): 145–147.

Sennott, Christie, and Sara Yeatman. 2012. "Stability and Change in Fertility Preferences Among Young Women in Malawi." *International Perspectives on Sexual and Reproductive Health* 38 (1): 34–42. https://doi.org/10.1363/3803412. https://www.ncbi.nlm.nih.gov/pubmed/22481147.

Sieving, Renee, W. Hellerstedt, C. McNeely, Ruth Fee, J. Snyder, and M. Resnick. 2005. "Reliability of Self-Reported Contraceptive Use and Sexual Behaviors among Adolescent Girls." *Journal of Sex Research* 42(2): 159–166. https://doi.org/10.1080/00224490509552269. https://www.ncbi.nlm.nih.gov/pubmed/16123846.

Strickler, Jennifer A., Robert J. Magnani, H. Gilman McCann, Lisanne F. Brown, and Janet C. Rice. 1997. "The Reliability of Reporting of Contraceptive Behavior in DHS Calendar Data: Evidence from Morocco." *Studies in Family Planning* 28(1): 44–53.

Tsui, Amy, Philip Anglewicz, Titilope Akinlose, Varsha Srivatsan, Peter Akilimali, Souleymane Alzouma, Fiacre Bazie, Peter Gichangi, Georges Guiella, Peter Kayembe, Anupam Mehrotra, Funmilola OlaOlorun, Elizabeth Omoluabi, Sani Oumarou, PR Sodani, Mary Thiongo, Meghan Byrne, Kurt Dreger, Michelle Decker, Carolina Cardona, Pierre Muhoza, Carolyn Combs, Alain K. Koffi, and Scott Radloff. 2020. "Performance Monitoring and Accountability: The Agile Project's Protocol, Record and Experience." *Gates Open Research* 4: 30. https://doi.org/10.12688/gatesopenres.13119.2. https://gatesopenresearch.org/articles/4-30/v2.

Zimmerman, Linnea, Hannah Olson, PMA Principal Investigators Group, Amy Tsui, and Scott Radloff. 2017. "PMA2020: Rapid Turn-Around Survey Data to Monitor Family Planning Service and Practice in Ten Countries." *Studies in Family Planning* 48 (3): 293–303.

ACKNOWLEDGMENT