REPORT

The Validity of Women’s Reports of Family Planning Service Quality in Cambodia and Kenya

Ann K. Blanc, Katharine J. McCarthy, Charlotte Warren, Ashish Bajracharya, and Benjamin Bellows

Population-based indicators of the coverage of key elements of high-quality family planning services are tracked via household surveys with female respondents, yet little work has been done to establish their validity. We take advantage of existing data sets from Cambodia and Kenya to compare women’s responses at exit interviews following a health facility visit against the observations of a trained third-party observer during the visit. The results, which treat the observations as the reference standard, show that indicators that measure contraceptive methods received are accurately reported while indicators of whether the woman received her preferred method and whether information was “discussed” or “explained” during counseling are less reliably reported. Studies designed explicitly to assess the validity of family planning questions in household surveys, especially questions in large survey programs critical for monitoring demographic trends and programmatic coverage, are needed.

BACKGROUND

Recognizing that providing clients with high-quality services is key to meeting the demand for family planning, maximizing client satisfaction, and delivering services within a human rights framework, increasing effort is being put into defining, measuring, and tracking indicators of quality (Jain et al. 2018; Tumlinson 2016; Track20 2019). Indicators used to track progress on the achievement of both national and international goals for improvements in women’s reproductive health include several population-based measures of the quality of

©2021 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.
family planning services that women receive. These indicators can also be used to assess the quality of family planning services at the subnational level, track changes in quality over time, and to measure the impact of quality on contraceptive choice and continuation and other outcomes (Ramamo and Mohanam 2003). For example, FP2020, a global partnership committed to expanding access to contraception, tracks the Method Information Index (MII), which reflects the quality of family planning counseling, and is made up of three questions asked of contraceptive users about their experience in obtaining their current method (FP2020, 2019). A composite index of family planning program quality recently proposed by Jain (2018) incorporates the MII as well as a Method Success Index that is constructed from population survey questions on reproductive intentions and contraceptive use.

Population-based indicators of the coverage1 of some of the elements of high-quality family planning services are tracked via household surveys with female respondents of reproductive age. In contrast, health facility-based indicators are useful for assessing the quality of services provided by a sample of facilities but do not provide information on the coverage of interventions (e.g., family planning counseling) among a defined population and geographic location. Several questions on the content of family planning consultations are included in the Demographic and Health Surveys (DHS) (Round 7) women’s questionnaire, including questions about whether the woman was told about side effects or other potential problems and whether she was told about multiple methods of contraception (DHS 2018). The Performance Monitoring and Accountability (PMA) 2020 surveys include these questions as well as a question about whether the woman received her preferred method during the visit (PMA 2020 2019).

There is a large literature on indicator definitions and measurement of family planning service quality (e.g., Simmons and Elias 1994; see Tumlinson 2016 for a recent review), but little work has been done to establish the criterion validity of specific indicators (i.e., comparison of survey responses to an external “gold standard” measure of what actually happened) (Munos et al. 2018). Recent reviews have highlighted the gap in information on the validity and reliability of service quality measures and have called for additional work in this area (Sprockett 2016; Marchant et al. 2019). Previous studies have assessed validity by comparing the results from indicator data collected via different methodologies, using quality measures to predict contraceptive outcomes, and employing “mystery clients” or “simulated clients” to compare their experience against the reports of actual clients. Although each of these approaches provides valuable information on the validity of indicator data, it is also important to acknowledge that every data collection method has disadvantages and comparisons against a true “gold standard” may not be possible.

Bessinger and Bertrand (2001) compared observations of family planning client–provider interactions and client exit interviews in three countries (Ecuador, Uganda, Zimbabwe) with the purpose of “lending credibility” to a package of tools to monitor quality (the Quick Investigation of Quality) (MEASURE Evaluation 2016). A range of indicators was compared, including those related to interpersonal relations between clients and providers, choice of

---

1 Population-based coverage is defined as the extent to which a population in need of an intervention (e.g., family planning clients) receives the intervention. More recently, recognition of the need to take account of intervention quality has led to efforts to define “effective coverage” or “quality-adjusted” measures and advocate for their routine inclusion in monitoring frameworks (Amouzou et al. 2019).
methods, and information given to clients. In general, the study found responses from the two sources were “highly comparable” based on the percentage agreement, although a few indicators performed poorly. Agreement was highest on the indicators that measured interpersonal relations, such as whether the client was greeted with respect/courtesy and whether the visit was conducted in privacy. Given the consistency of the results, the authors concluded that there may be no need to implement both exit interviews and observations of client-provider interactions, depending on local conditions and the extent of the quality assessment needed. A similar analysis comparing observations of client-provider interactions with exit interviews in Haiti, Malawi, and Senegal focused on family planning counseling indicators (Assaf et al. 2016). Four indicators were included in the analysis: whether the client had questions or concerns about her current method, how to use a method provided during the visit, side effects of the method, and when to return for a follow-up visit. Overall, clients reported higher levels of counseling than observers. Correspondence between the two reports was low: Kappa statistics for agreement were in the low to fair range (0.03–0.35) and the percentage agreement ranged from 29 to 74 percent. The authors recommended that self-reported indicators on counseling be viewed with caution because “clients tended to over-report receipt of services.”

Another study that focused on family planning counseling about side effects compared direct observations to exit interviews in health facility surveys, and direct observations to reports by women in household surveys (Choi 2018). The study is based on Service Provision Assessments (SPA) of a representative sample of health facilities and nationally representative DHS in four countries—Haiti, Senegal (two rounds), Malawi, and Tanzania. Responses to the exit interviews consistently indicated higher estimates of counseling about side effects than observations in all countries. Sensitivity ranged from 74 to 91 percent while specificity was lower (35–55 percent). The analysis also included a comparison of estimates based on observations in health facilities with estimates derived from DHS conducted around the same time. The household survey estimates were not statistically different from the observations in seven of 10 comparisons. The author concludes that population-based survey data on counseling should not “necessarily be considered to be of poor quality” but also calls for further studies to evaluate indicators of quality of care beyond counseling about side effects.

Recent efforts have also assessed the validity of composite measures of quality to predict contraceptive needs and/or use at a later time. Holt et al. (2019) developed and validated the Quality of Contraceptive Counseling scale in Mexican public health facilities. The scale was significantly associated with whether family planning clients felt they needed more information about contraceptive methods and whether they were using a method one to three months later (although contraceptive use was only marginally significant). In two cohorts recruited from franchised private clinics in Uganda and Pakistan, Chakraborty et al (2019) found the cumulative probability of modern method continuation at 12 months was higher for clients who reported higher baseline values of an MII metric at exit. Jain et al. (2019) assessed the predictive validity of composite measures of process quality on contraceptive continuation in two states in India using longitudinal data. Two variations of multi-item indices of process quality were strongly associated with contraceptive continuation after three months. The authors argue that these results support the validity of quality measures, but also endorse additional testing of “their feasibility in cross-sectional surveys.”
Although the above studies are important contributions to measuring the quality of family planning services, the comparison of client reported receipt of services against a “gold standard” is also critically needed. To our knowledge, the only methodology used to validate family planning indicators against a “gold standard” has been comparing reports of a simulated client to results from provider interviews, observations, and exit interviews. This method—sending a data collector pretending to be a family planning client to a facility—is arguably less likely to be biased by measurement error introduced by recall error, social desirability reporting, or the Hawthorne effect (Tumlinson 2016). A 2014 study by Tumlinson and others tested this hypothesis in a sample of public and private facilities in Kisumu, Kenya by assessing whether simulated clients provided a more accurate measurement of quality than either observations of provider–client interactions or interviews with providers and clients (Tumlinson et al. 2014). The hypothesis of the researchers was that the biases of other methods—courtesy bias, recall bias, the Hawthorne effect—overestimate quality and that the “gold standard” simulated client method would yield a more negative assessment of quality. In general, the hypothesis was confirmed with negative provider behavior being underreported, although some of the indicators associated with method choice and client relations were consistently reported by observers and exit interviews with clients.

This paper extends the evidence base on measuring the quality of family planning services by assessing the validity of a set of population-based indicators that include both process (e.g., counseling) and outcome indicators (e.g., receiving a method). We compared women’s responses at exit interview following a visit to a health facility against observations of a trained third-party observer during the visit using data from Cambodia and Kenya. In our analysis of the validity of women’s reports, we treat the observations as the reference under the assumption that they are an essentially accurate, if not perfect, reflection of the family planning visit.

DATA AND METHODS

Observations of family planning visits were conducted by trained third-party observers using a structured checklist in health facilities located in Cambodia and Kenya. Women’s reports of care received were collected via exit interview prior to her leaving the health facility following a family planning visit. The data were originally collected as part of a quasi-experimental pretest-posttest evaluation of a voucher and accreditation intervention led by the Population Council and, in Cambodia, collaborating with the National Institute of Public Health. The primary objective of the evaluation was to assess the influence of the voucher program (henceforth “Voucher program”) on reproductive, maternal, and newborn health service utilization. For purposes of the Cambodia Voucher program evaluation, health facilities accredited by the Voucher program were purposively selected and then matched to comparison facilities with similar characteristics. In the Kenya Voucher evaluation, 21 of 56 Voucher facilities were randomly selected and matched to 20 comparison facilities with similar characteristics. For purposes of the current study, data from both groups of facilities were utilized. Matched exit interviews and observations were conducted in 81 health facilities—40 public sector facilities in Cambodia, and 41 public and private or faith-based facilities in Kenya. In Kenya, the majority of facilities were hospitals (61 percent), followed by health
centers (31 percent), and nursing homes, dispensaries or clinics (8 percent). In Cambodia, all but two health facilities were health centers; two were former district hospitals.

Family planning clients at study health facilities were randomly selected for observation until six consultations in each facility were observed. To be eligible for inclusion in the study, clients needed to be aged 18–45 and have given consent for their consultation to be observed and to be interviewed following the consultation. The exit interview was conducted by an interviewer who was not the same person who observed the consultation. Two rounds of cross-sectional data were collected in each country between 2010 and 2012, which were pooled for each country. In total, there are 475 family planning visits in Cambodia and 573 in Kenya. Full details on the methodology of the evaluations are provided in published protocols (Bellows et al. 2011; Warren et al. 2011).

Study Settings

In Kenya, voucher facilities in the study were located in Kisumu, Kiambu, Kitui counties, and Nairobi while the matched comparison facilities were in Nyandarua, Uasin Gichu, and Makueni counties. In the most recent national DHS prior to data collection in Kenya (2008–2009), 39 percent of currently married women reported using a modern method of family planning. Injectable are by far the most common method (21 percent) followed by the pill (7 percent) and female sterilization (5 percent). Overall, 57 percent of users of modern methods obtained their method from a public sector facility. Sixty three percent of those who used a public sector facility reported being informed about potential side effects while 56 percent said they were told what to do if they experienced side effects. About two-thirds (67 percent) were told about other methods that could be used (KNBS and ICF Macro, 2010).

In Cambodia, voucher and comparison health facilities were located in eight provinces (Kampong Thom, Kampot, Prey Veng, Kampong Cham, Kep, Siem Reap, Svay Reing, and Oddor Mean Chey). The 2010 DHS in Cambodia showed that 35 percent of currently married women were using a modern method of family planning nationally. The most common methods were the pill (15 percent), injectables (10 percent), and the Intra-uterine device (IUD) (3 percent). The public sector is the largest source of these methods, but there is substantial private sector provision of pills and injectables. Women who received their current family planning method at a public sector facility were highly likely to report that they were informed about side effects (86 percent), they were informed about what to do if they experienced side effects (85 percent), and that they were informed about multiple methods (82 percent) (Directorate General for Health and ICF Macro, 2011).

Sample Size

For purposes of this secondary analysis, we anticipated indicator prevalence would range between 50 and 80 percent coverage. We assumed levels of moderate to high sensitivity (60–70 percent) and specificity (70–80 percent), given that women were asked to recall the interventions immediately after the visit. Sample size for anticipated sensitivity and specificity levels was calculated using Buderer’s formula (Buderer 1996). We set \( \alpha = 0.05 \) for both accuracy parameters assuming a normal approximation to a binomial distribution. Based on these
specifications, a sample size of 400 women per country is sufficient to estimate 60 percent sensitivity and 70 percent specificity with ±7 percent precision.

**Ethical Review**

Ethical clearance for the Voucher study was granted by the Population Council’s Institutional Review Board (IRB) (approval number 496 for Cambodia, 470 for Kenya), the National Ethics Committee for Health Research (NECHR) (approval numbers 173 and 186) in Cambodia, and the Kenya Medical Research Institute (KEMRI) Ethical Review Board (approval number 164). An exemption from full review to conduct secondary analysis of de-identified data was obtained from the Population Council IRB prior to analysis.

**Validation Analysis**

In each data set, a unique client identification code was recorded in both the exit interview and observation record and matched. The number of cases available for analysis for each variable differs depending on the extent of missing data. Questions about whether interventions occurred were coded as 1 if the response was “Yes” and all other responses were coded as 0.

We constructed two-by-two tables and calculated sensitivity (the “true” positive rate) and specificity (the “true” negative rate) for each indicator. Following recommendations based on previous validity studies of population-based indicators, we calculate measures of both individual level and population level reporting accuracy (under the assumption that the observer reports are “true”) (Munos et al., 2018). For individual-level accuracy, we estimated the area under the receiver operating characteristic (ROC) curve (AUC) and corresponding 95 percent confidence intervals following a binomial distribution. The AUC can be interpreted as “the average sensitivity across all possible specificities” (Macaskill et al., 2010). The AUC varies between 0 and 1. An AUC of 0.5 indicates that the indicator is no better than a random guess and an AUC of 1 represents perfect accuracy (100 percent sensitivity and 100 percent specificity). An AUC value of 0.7 or higher was used as the cutoff criteria for high individual-level reporting accuracy (Munos et al., 2018).

To assess population-level validity for each indicator, we calculated the degree to which an indicator would be over- or underestimated in a household survey using the inflation factor (IF). The IF is the ratio of the indicator’s estimated population-based survey prevalence to the indicator’s “true” (observed) prevalence. To estimate the population-based survey prevalence, we applied the indicator’s estimated sensitivity and specificity to its “true” observed prevalence, using the following equation:

\[
\text{Estimated population survey prevalence} = (\text{"true" (observed) prevalence} \times \text{sensitivity}) \\
+ [(1 - \text{"true" (observed) prevalence}) \times (1 - \text{specificity})].
\]

We used an IF cutoff between 0.75 and 1.25 as the benchmark for low population-level bias (Munos et al., 2018). Indicators based on a small number of “true” (observed) positive or “true” negative cases that resulted in estimated precision for sensitivity or specificity of 15 percentage points or more are reported in the data tables, but not discussed in the text. The summary measures AUC and IF are also suppressed for these indicators due to a high degree
of uncertainty around the estimate. Analyses were performed using R Studio Version 1.1.383 (RStudio Inc., Boston, MA).

The first set of indicators we tested measure whether a woman received any method during her visit and, if so, the specific method she received. An indicator was constructed for each specific method of family planning by coding “1” if the woman received the method and “0” if she did not receive the specific method. We also assessed whether the client received her preferred method (among clients who received a method). Women were asked whether the method they received was their preferred method. Observers recorded whether the woman mentioned a preferred method during the consultation and whether she received that method. Four indicators of the quality of family planning counseling were also collected in the study: whether the provider discussed the client’s prior use of family planning, two or more family planning methods, how the chosen method works, and the advantages and disadvantages of the chosen method (for specific interview questions and observer instructions, see Table 1).

RESULTS

The family planning clients in each sample had similar social and demographic profiles (Table 2). In both Cambodia and Kenya, over 80 percent of family planning clients were aged 20–39; very few adolescents were family planning clients. The majority of clients in both countries had a primary education while a substantial minority had secondary education. In Cambodia, all family planning clients in the sample were married or cohabiting. In Kenya, most clients were married or cohabiting but around 12 percent were not currently married. All clients in both countries had at least one prior birth, but most had two or more births at the time of the visit. Although younger, unmarried women were not excluded from eligibility for the study, married, older and women who have had at least one child are likely the typical family planning client for a public-sector family planning program in these settings.

The percentage of clients who received any method according to the observers was 89 percent in Kenya and 98 percent in Cambodia (Table 3). This indicator had high sensitivity (SE 97.7, CI (95.9, 98.8); SE 98.9, CI (97.4, 99.6), respectively) while specificity was lower (SP 71.2, CI (57.9, 82.2); SP 50.0, CI (18.7, 81.3), respectively). The indicator met both criteria for high validity in Kenya. In Cambodia, summary criteria are not reported due to high uncertainty about the estimate. In Kenya, the most common methods that clients received were injectables, implants, and pills (66, 15, and 13 percent, respectively). In Cambodia, pills and injectables predominated (46 and 45 percent, respectively). Other methods, including IUDs and condoms, were received far less frequently. Overall, clients accurately reported the methods they received, especially for those that are most common. The indicators for pills and injectables met the criteria for high validity in both countries. In Kenya, the indicator for whether an implant was received also met both criteria. The less common methods were less well-reported.

Women slightly over-reported the extent to which they received their preferred method. This indicator had high sensitivity in both Kenya and Cambodia (95.3 percent (CI: 92.6, 97.3), 99.7 percent (CI: 98.1, 100.0)). In contrast, specificity was low (16.1 percent (7.6, 28.3), 33.3
| Indicator                  | Exit interview question                                                                 | Observer checklist                                                                 |
|---------------------------|-----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Received a FP method      | Did you receive a family planning method today?                                          | Which method(s) does she actually receive? [WRITE METHOD OR NONE]                    |
| Method received           | Which method(s) did you receive today?                                                   | Does the client mention a preferred method?                                         |
| Received preferred method | Is this/are these your preferred method(s)?                                               | Does the client receive her preferred method(s)?                                    |
|                           | For the method(s) you received today, did the provider…                                  | For the method(s) the client receives, does the provider… [OBSERVE AND CIRCLE “1” FOR “YES” IF DISCUSSED; OTHERWISE CIRCLE “0”] |
| Discussed how method works| Discuss with you how the chosen method works?                                            | • Explain how method works                                                           |
| Explained advantages/disadvantages | Explain the advantages/disadvantages of the chosen method? | • Explain advantages/benefits                                                        |
|                           |                                                                                         | • Explain disadvantages                                                              |
| Discussed previous use of FP | Discuss your previous use of family planning?                                           | Are the following areas discussed/ mentioned during the consultation? [OBSERVE AND CIRCLE “1” FOR “YES” IF MENTIONED; OTHERWISE CIRCLE “0”] |
|                           |                                                                                         | • Previous use of FP                                                                 |
| Discussed 2+ FP methods   | Discuss with you two or more available methods?                                         | Which methods are discussed during the consultation? [OBSERVE AND CIRCLE “1” FOR “YES” IF DISCUSSED; OTHERWISE CIRCLE “0”] |
|                           |                                                                                         | • Progestin only pill                                                                |
|                           |                                                                                         | • Combined pill                                                                      |
|                           |                                                                                         | • IUCD                                                                              |
|                           |                                                                                         | • Male condom                                                                       |
|                           |                                                                                         | • Female condom                                                                     |
|                           |                                                                                         | • Injectables                                                                        |
|                           |                                                                                         | • Sterilization                                                                     |
|                           |                                                                                         | • Emergency contraception                                                           |
|                           |                                                                                         | • Implants                                                                          |
|                           |                                                                                         | • LAM                                                                               |
|                           |                                                                                         | • Natural Family Planning methods                                                    |
|                           |                                                                                         | • Other (specify)                                                                   |
TABLE 2  Sample characteristics of family planning clients

|                     | Cambodia |                     | Kenya |                     |
|---------------------|----------|----------------------|-------|----------------------|
|                     | Total = 475 |                      | Total = 573 |                      |
| **Round**           |          |                      |       |                      |
| Baseline            | 240      | 50.5                 | 392   | 68.4                 |
| Follow-up           | 235      | 49.5                 | 181   | 31.6                 |
| **Age group**       |          |                      |       |                      |
| 15–19               | 6        | 1.3                  | 31    | 5.4                  |
| 20–29               | 199      | 41.9                 | 330   | 57.6                 |
| 30–39               | 192      | 40.4                 | 159   | 27.7                 |
| 40–49               | 75       | 15.8                 | 39    | 6.8                  |
| Do not know         | 0        | 0.0                  | 3     | 0.5                  |
| Missing             | 3        | 0.6                  | 11    | 1.9                  |
| **Level of education** |          |                      |       |                      |
| None/Pre-primary    | 57       | 12.0                 | 25    | 4.4                  |
| Primary             | 274      | 57.7                 | 340   | 59.3                 |
| Secondary           | 121      | 25.5                 | 195   | 34.0                 |
| Tertiary            | 23       | 4.8                  | 5     | 0.9                  |
| Missing             | 0        | 0.0                  | 8     | 1.4                  |
| **Marital status**  |          |                      |       |                      |
| Never married       | 0        | 0.0                  | 45    | 7.9                  |
| Married/co-habiting | 472      | 100.0                | 497   | 86.7                 |
| Separated/divorced/ | 0        | 0.0                  | 23    | 4.0                  |
| widowed             |          |                      |       |                      |
| Missing             | 3        | 0.0                  | 8     | 1.4                  |
| **Number of births**|          |                      |       |                      |
| One                 | 2        | 0.4                  | 99    | 17.3                 |
| Two                 | 77       | 16.2                 | 110   | 19.2                 |
| Three               | 121      | 25.5                 | 126   | 22.0                 |
| Four or more        | 100      | 21.1                 | 102   | 17.8                 |
| Missing             | 0        | 0.0                  | 0     | 0.0                  |

percent (9.9, 65.1)), suggesting that few women who did not receive their preferred method reported that they did not receive their preferred method. The AUC for this indicator was low in Kenya and did not meet the high validity benchmark. However, the IF value for Kenya was close to 1.0, suggesting that while individual-level validity is low, the population-level estimate is acceptable. In Cambodia, summary indicators were suppressed due to low precision in the specificity estimate.

Tables 4 and 5 show the cross-tabulation of responses on method received, comparing responses from the client and the observer. Agreement in reporting is shown in the cells along the diagonal. Results show that in Kenya there is close, but not perfect, agreement on the method received. There are only 14 cases of 42 in which the observer reported that the client did not receive a method, but the client reported that she did. There are nine of 37 cases in which the client reported that she did not receive a method, but the observer reported that she did. In Cambodia, the largest number of discrepancies occurred among cases where the observer reported that the woman received oral contraceptives (211 cases) but the client reported that she received another method including no method (2 cases), injectable (12 cases), and IUD (5 cases).

We assessed four indicators of the content of family planning counseling (Table 6). Only one indicator (whether the provider discussed two or more methods) met both of the criteria for high validity in one country (Kenya). Most of the IF values were above 1.0 indicating that women tended to over-report whether counseling occurred. It is notable that, even with likely over-reporting, the reported level of counseling for these indicators was low. After adjusting
### TABLE 3 Validation results for indicators on method received, Kenya and Cambodia

| Indicator                                      | Study | N   | TP  | FP  | TN  | FN  | Obsv. prev. (%) | Est. prev. (%) | Sensitivity % (95% CI) | Specificity % (95% CI) | AUC (95% CI) | IF   | Met both criteria? | No. of studies met both criteria |
|------------------------------------------------|-------|-----|-----|-----|-----|-----|----------------|---------------|------------------------|------------------------|-------------|------|-----------------|-------------------------------|
| Received a family planning method              | KE    | 538 | 468 | 17  | 42  | 11  | 89.0           | 90.1          | 97.7 (95.9, 98.8)       | 71.2 (57.9, 82.2)       | 0.84 (0.79, 0.90) | 1.01 | Y              | *                            |
|                                                | CA    | 461 | 446 | 5   | 5   | 5   | 97.8           | 97.8          | 98.9 (97.4, 99.6)       | 50.0 (18.7, 81.3)       | NA          |                  | NA                           |
| Received injectable                            | KE    | 486 | 314 | 14  | 152 | 6   | 65.8           | 67.4          | 98.1 (96.0, 99.3)       | 91.6 (86.3, 95.3)       | 0.95 (0.93, 0.97) | 1.03 | Y              | **                           |
|                                                | CA    | 230 | 102 | 8   | 119 | 1   | 44.8           | 47.8          | 99.0 (94.7, 100.0)      | 93.7 (88.0, 97.2)       | 0.96 (0.94, 0.99) | 1.07 | Y              |                 |
| Received IUD                                   | KE    | 484 | 16  | 8   | 460 | 0   | 3.3            | 5.0           | 100.0 (79.4, 100.0)     | 98.3 (96.7, 99.3)       | 0.99 (0.99, 1.00) | 1.50 | N              |                 |
|                                                | CA    | 230 | 7   | 0   | 218 | 5   | 5.2            | 3.0           | 58.3 (27.7, 84.8)       | 100.0 (98.3,100.0)      | NA          |                  | NA                           |
| Received oral contraceptive pill               | KE    | 484 | 53  | 4   | 419 | 8   | 12.6           | 11.8          | 86.9 (75.8, 94.2)       | 99.1 (97.6, 99.7)       | 0.93 (0.89, 0.97) | 0.93 | Y              | **                           |
|                                                | CA    | 460 | 192 | 9   | 239 | 20  | 46.1           | 43.7          | 90.6 (85.8, 94.1)       | 96.4 (93.2, 98.3)       | 0.93 (0.91, 0.96) | 0.95 | Y              |                 |
| Received implant                               | KE    | 484 | 65  | 5   | 408 | 6   | 14.7           | 14.5          | 91.5 (82.5, 96.8)       | 98.8 (97.2, 99.6)       | 0.95 (0.92, 0.98) | 0.99 | Y              | *                            |
|                                                | CA    | 460 | 11  | 0   | 444 | 5   | 3.5            | 2.4           | 68.8 (41.3, 89.0)       | 100.0 (99.2,100.0)      | NA          |                  | NA                           |
| Received condom                                | KE    | 421 | 348 | 47  | 9   | 17  | 86.7           | 93.8          | 95.3 (92.6, 97.3)       | 16.1 (7.6, 28.3)        | 0.56 (0.51, 0.61) | 1.08 | N              |                 |
| Received preferred method (of those who received a method) | KE    | 421 | 348 | 47  | 9   | 17  | 86.7           | 93.8          | 95.3 (92.6, 97.3)       | 16.1 (7.6, 28.3)        | 0.56 (0.51, 0.61) | 1.08 | N              |                 |
|                                                | CA    | 310 | 297 | 8   | 4   | 1   | 96.1           | 98.4          | 99.7 (98.1,100.0)       | 33.3 (9.9, 65.1)        | NA          |                  | NA                           |

NOTE: Benchmark for high validity is AUC > 0.7 and IF (inflation factor) between 0.75 and 1.25.

In Cambodia, data are shown for follow-up round only. Implant in CA had <2% prevalence and was not recorded by observers and is not shown. Condom in KE had <1% prevalence and is not shown.

KE = Kenya; CA = Cambodia, N = total matched N, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives, Obsv. prev.: observed ("true") prevalence, Est. prev.: estimated prevalence that would be obtained from a household survey applying the sensitivity, specificity and prevalence observed in this study.

Gray values are indicators based on a small number of “true” (observed) positive or “true” negative cases which resulted in estimated precision for sensitivity or specificity of 15 percentage points or more.

The * and ** in the last column indicate the number of studies that met both criteria (* = 1, ** = 2).
TABLE 4  Method received according to observer and client, Cambodia

| Observer recorded | No method | Injectable | IUD | Pill | Implant | Condom | Total |
|-------------------|-----------|------------|-----|------|---------|--------|-------|
| No method         | 5         | 0          | 1   | 4    | 0       | 0      | 10    |
| Injectable        | 0         | 102        | 0   | 1    | 0       | 0      | 103   |
| IUD               | 0         | 0          | 7   | 0    | 5       | 0      | 12    |
| Pill              | 2         | 12         | 5   | 192  | 0       | 0      | 211   |
| Implant           | 0         | 0          | 0   | 0    | 0       | 0      | 0     |
| Condom            | 1         | 1          | 3   | 0    | 0       | 11     | 16    |
| Total             | 8         | 115        | 16  | 197  | 5       | 11     | 352   |

NOTE: IUD and injectable suppressed in R0 due to likely coding discrepancy. Bold values indicate consistent responses between observer and client.

TABLE 5  Method received according to observer and client, Kenya

| Observer recorded | No method | Injectable | IUD | Pill | Implant | Condom | Total |
|-------------------|-----------|------------|-----|------|---------|--------|-------|
| No method         | 28        | 5          | 6   | 0    | 3       | 0      | 42    |
| Injectable        | 1         | 314        | 0   | 4    | 1       | 1      | 321   |
| IUD               | 2         | 0          | 16  | 0    | 0       | 0      | 18    |
| Pill              | 2         | 3          | 1   | 53   | 1       | 0      | 60    |
| Implant           | 4         | 4          | 1   | 0    | 65      | 0      | 74    |
| Condom            | 0         | 1          | 0   | 0    | 0       | 0      | 1     |
| Total             | 37        | 326        | 24  | 57   | 70      | 1      | 516   |

Bold values indicate consistent responses between observer and client.

for sensitivity and specificity, the estimated survey prevalence ranged from 60 to 73 percent in Kenya and from 32 to 64 percent in Cambodia.

To examine whether reporting accuracy varied by respondent characteristics, we stratified results by age group, level of education, and parity. Although the small number of women below age 20 and at low parity in the samples may influence the results, we did not find systematic differences in the accuracy of self-reports by women by respondent characteristics.

DISCUSSION AND CONCLUSION

Overall, five of 10 indicators in Kenya and two of six in Cambodia met the high validity criteria for both individual and population-based validity. Compared to indicators of other types of service delivery, such as labor and delivery care (Blanc et al. 2016a, 2016b; McCarthy et al. 2016; Stanton et al. 2013), the family planning indicators demonstrated comparatively high accuracy. For purposes of tracking population-based indicators of quality at the national or subnational level, the IF provides a useful measure of reporting accuracy at the aggregate level. Individual discrepancies in reporting between observers and interviewees can cancel out and result in an acceptable estimate at the population level. The IF met our standard for high validity in seven of nine indicators in Kenya and five of six indicators in Cambodia.

However, the IF is also influenced by the level of coverage of an intervention, so it is important to assess its value in the context of the observed and "true" prevalence estimates, as shown previously (Blanc et al., 2016b). These results do not vary systematically by characteristics of the respondent, a finding that is replicated in previous studies with similar methodology.
| Indicator                                      | Study | N  | TP  | FP  | TN  | FN  | Observed prevalence (%) | Est. survey prevalence (%) | Sensitivity % (95% CI) | Specificity % (95% CI) | AUC (95% CI) | IF   | Met both criteria? | No. studies met both criteria |
|-----------------------------------------------|-------|----|-----|-----|-----|-----|--------------------------|----------------------------|-------------------------|------------------------|--------------|------|------------------|-----------------------------|
| Discussed prior use of family planning        | KE    | 497| 286 | 79  | 49  | 83  | 74.2                     | 73.4                       | 77.5 (72.9,81.7)         | 38.3 (29.8,47.3)        | 0.58 (0.53, 0.63) | 0.99 | N                |                             |
|                                               | CA    | 469| 157 | 93  | 139 | 80  | 50.5                     | 53.3                       | 66.2 (59.8,72.2)         | 59.9 (53.3, 66.3)        | 0.63 (0.59, 0.67)    | 1.06 | N                |                             |
| Discussed 2+ family planning methods          | KE    | 449| 199 | 91  | 119 | 40  | 53.2                     | 64.6                       | 83.3 (77.9,87.8)         | 56.7 (49.7, 63.5)        | 0.70 (0.66, 0.74)    | 1.21 | Y                | *                           |
|                                               | CA    | 464| 104 | 98  | 218 | 44  | 31.9                     | 43.5                       | 70.3 (62.2,77.5)         | 69.0 (63.6, 74.0)        | 0.70 (0.65, 0.74)    | 1.36 | N                |                             |
| Explained how chosen method works             | KE    | 491| 174 | 164 | 115 | 38  | 43.2                     | 68.8                       | 82.1 (76.2,87.0)         | 41.2 (35.4, 47.2)        | 0.62 (0.58, 0.66)    | 1.59 | N                |                             |
|                                               | CA    | 448| 136 | 55  | 182 | 75  | 47.1                     | 42.6                       | 64.5 (57.6,70.9)         | 76.8 (70.9, 82.0)        | 0.71 (0.66, 0.75)    | 0.91 | N                |                             |
| Discussed advantages/ disadvantages of chosen | KE    | 491| 155 | 140 | 150 | 46  | 40.9                     | 60.1                       | 77.1 (70.7,82.7)         | 51.7 (45.8, 57.6)        | 0.64 (0.60, 0.69)    | 1.47 | N                |                             |
| method                                        | CA    | 448| 227 | 61  | 97  | 63  | 64.7                     | 64.3                       | 78.3 (73.1,82.9)         | 61.4 (53.3, 69.0)        | 0.70 (0.65, 0.74)    | 0.99 | N                |                             |

NOTE: Benchmark for high validity is AUC > 0.7 and IF (inflation factor) between 0.75 and 1.25. KE = Kenya; CA = Cambodia, N = total matched N, TP = true positives, FP = false positives, FN = false negatives, TN = true negatives. Obsv. prev.: observed ("true") prevalence, Est. prev.: estimated prevalence that would be obtained from a household survey applying the sensitivity, specificity and prevalence observed in this study.
Drawing from previous research on the validity of reporting by women in structured interviews, we hypothesize that indicators that measure objective events, such as whether the woman received a method during her visit, will be more accurately reported than indicators that might be more subjective, such as whether a provider discussed specific topics with the client (Bessinger and Bertrand 2001; Glick 2009). As expected, women’s reports on whether they received a method and, for the most common methods, the specific method received met our criteria for high accuracy. The less common methods were less accurately reported, but small cell counts contribute to high uncertainty about the estimates. Detailed cross-tabulations of observer reports by client reports revealed a high, but not perfect, level of consistency. Aside from possible errors in recording, the differences may result from different interpretations of what it means to “receive” a method. For example, if the woman had to go from her consultation to the pharmacy in the facility to get pills, she may have said that she did not receive the method. Or she may have received a referral for a procedure (e.g., implant or IUD insertion) that would be done at another facility or on another day.

Although the evidence we are able to provide on the preferred method indicator is limited by a small number of observations in some table cells (in Cambodia), the relatively low validity of the indicator suggests a need for clarity on what this indicator is capturing and what it is intended to capture. As noted by Tumlinson et al. (2014), a woman may enter into a family planning counseling session with a preference for a specific method but may change her preference after receiving additional information about that method or other methods. She may also change her preference if she is told that her initial preferred method is not available. Some women may be reluctant to express a preference if they believe they do not know enough about the methods, that their information may be incorrect, or have experienced stock outs of their preferred method in the past. The low levels of specificity may indicate courtesy bias (i.e., a reluctance to report that she did not receive what she came for) but may also result from weakly held preferences. A further concern about this indicator is that it is complex because it relies on collecting several pieces of information—whether the client has a preference, whether she expresses it during the consultation, and whether she gets her preferred method. Although further testing of this indicator would be useful, reliance on a simpler indicator that requires only one question about preferences might be desirable (e.g., whether the provider asks about her preference).

One of the strengths of this study compared to previous studies is that we were able to assess the accuracy of reporting about several indicators of family planning counseling. The counseling indicators performed less well than the indicators on receiving a method. Given that these types of questions refer to whether specific issues were “discussed” or “explained” by the provider, they are likely to be more subject to variations in reporting than questions about whether a method was received. Social desirability bias may be a factor in women’s reporting about the interaction with the provider. The relatively low levels of specificity compared to sensitivity in the Kenyan data may reflect women’s reluctance to report that a provider did not discuss or explain issues during counseling.

---

2 In general, it is difficult to assess the validity of indicators of interventions with very high or very low prevalence due to small numbers of cases falling into some cells of the $2 \times 2$ tables for calculating sensitivity and specificity (Munos et al. 2018).
not give her information that she assumes the provider was supposed to give. For Cambodia, specificity was lower than sensitivity for two of the four indicators and specificity of these indicators on counseling was higher than in Kenya but still did not exceed 77 percent. Similar results have been reported in other studies. For example, Choi (2018) found that specificity was significantly lower than sensitivity for indicators of counseling about side effects in five country studies. Assaf et al. (2016) also found that family planning clients in three countries reported higher levels of counseling than observers on four indicators. For counseling indicators, it may be that for most purposes what matters is only the client’s perception of whether she received certain information or was asked about her preferences. However, objectively verifiable indicators of service quality are also needed to determine whether service providers are delivering the quality they are intended to deliver. Overall, most of the few studies conducted to date, including this one, suggest that women over-report positive aspects of service quality in survey interviews (Assaf et al. 2016, Tumlinson et al. 2016, Choi 2018).

The limitations of this study are important to keep in mind when interpreting its results and their implications. While using an external observer may be as close to a true “gold standard” measure as we can achieve at present, it is still likely to reflect some level of error or bias. Even well-trained observers may miss some component of a visit or mis-record it. There is some level of subjective interpretation about whether a topic was “discussed” or “explained.” The presence of an observer during a visit with a provider may itself positively influence women’s recall of the visit. Additionally, when using third-party observations, there is often concern about the Hawthorne effect, that is, the person observed is more likely to be on their best behavior rather than performing in a typical manner. If this study were concerned about capturing the quality of care provided on a typical day, this effect could cause measurement error. However, since this study is designed to measure whether clients can accurately self-report select indicators of quality, as observed by a third party, the Hawthorne effect is not a concern for this type of validation study.

A further limitation is that, since this study was based on secondary analysis of existing data collected for another purpose, it did not exactly replicate the conditions under which population-based surveys like the DHS usually take place. Women’s reports of their experience during a family planning visit in a survey interview usually involve reporting on an event (e.g., a family planning visit) that took place months or years prior to the interview. As such, a deterioration in women’s recall over time might be a factor in the extent to which they are able to report with accuracy. However, to the extent that this type of recall bias has been examined under conditions more closely replicating a population-based survey, it does not appear to be substantial or systematic for indicators that are immediately recalled with high accuracy (McCarthy et al. 2016, Carter et al. 2021, Choi 2018). In fact, social desirability bias might be a more important factor in exit interviews since they take place in a facility immediately after a visit, but presumably would not be an issue in a household survey interview. Women’s slight over-reporting of the extent to which they received their preferred method may reflect this bias.

This validity study took advantage of existing data to assess a relatively small number of indicators of family planning service quality. More studies of this type could be undertaken by ensuring that when facility-level observational data are collected (as in the DHS SPA), they can be linked at the individual level to exit interviews with women. Overall, the...
limited literature on this topic shows that comparisons of family planning indicators derived from different sources need to be undertaken with caution and that the available data collection methodologies have diverse strengths and weaknesses (Tumlinson et al. 2014). For example, household surveys of women may not be well-suited for answering questions about whether family planning providers are adhering to protocols around counseling, while they may be much better suited to measuring women’s knowledge of side effects of methods or experience using methods. Studies designed explicitly to assess the validity of questions currently or potentially to be added to household survey questionnaires, especially those that are included in large survey programs and are critical for monitoring progress, are needed. For example, current contraceptive use is a key indicator but is known to have some biases in reporting (Hall et al. 2010). Validation of this indicator against a “gold” standard (e.g., a biomarker) would provide important information, particularly for global efforts that rely on internationally comparable indicators. Other indicators that have been less well-studied and would benefit from additional validation work are those that measure interpersonal relations between clients and providers (e.g., being greeted, privacy, confidentiality). Employing both quantitative methodologies (as in this study) and qualitative approaches, such as cognitive interviewing (Sudhinaraset et al. 2018), to understand women’s interpretation of survey questions and their descriptions of their experiences in health facilities would add to a greater understanding of the validity of family planning service quality and enhance the ability of programs to improve services for women.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Dataverse at https://dataverse.harvard.edu/dataverse/popcouncil.

REFERENCES

Amouzou, Agbessi, Hannah Hogan Leslie, Malathi Ram, Monica Fox, Safia S. Jiwani, Jennifer Requejo, Tanya Marchant, Melinda Kay Munos, Lara M. E. Vaz, William Weiss, Chika Hayashi, Ties Boerma, and On behalf of the Countdown Coverage Technical Working Group. 2019. “Advances in the Measurement of Coverage for RMNCH and Nutrition: From Contact to Effective Coverage.” BMJ Global Health 4:i114–i124.

Assaf, Shireen, Wenjuan Wang, and Lindsay Mallick. 2016. "Provider Counseling and Knowledge Transfer in Health Facilities of Haiti, Malawi, and Senegal.” DHS Analytical Studies No. 60. ICF International, Rockville, MD.

Blanc, Ann K., Claudia Diaz, Katharine J. McCarthy, and Karla Berdichevsky. 2016a. “Measuring Progress in Maternal and Newborn Health Care in Mexico: Validating Indicators of Health System Contact and Quality of Care.” BMC Pregnancy and Childbirth 16(1): 255.

Blanc, Ann K., Charlotte Warren, Katharine J. McCarthy, James Kimani, Charity Ndliga, and Saumya RamaRao. 2016b. "Assessing the Validity of Indicators of Maternal and Newborn Health Care in Kenya." Journal of Global Health 6 (1): 010405. doi:10.7189/jogh.06.010405.

Bellows, Benjamin, Charlotte Warren, Saphonn Vonthanak, Chhea Chhorvany, Hean Sokho, Chean Men, Ashish Bajaracharya, Ubaidur Rob, and Tung Rathavy. 2011. "Evaluation of the Impact of the Voucher and Accreditation Approach on Improving Reproductive Behaviors and Status in Cambodia." BMC Public Health 11(1): 667.

Bessinger, Ruth E. and Jane T. Bertrand. 2001. Monitoring quality of care in family planning programs: a comparison of observations and client exit interviews. International Family Planning Perspectives. 27 (2): 63–70.

Buderer, N. M. 1996. "Statistical Methodology: I. Incorporating the Prevalence of Disease into the Sample Size Calculation for Sensitivity and Specificity.” Academic Emergency Medicine 3 (9): 895–900.
Carter, Emily, Karen Chang, Luke Mullany, Subarna Khatri, Steven Le Clerq, Melinda Munos, and Joanne Katz. 2021. Reliability of Maternal Recall of Delivery and Immediate Newborn Care Indicators in Sarlahi, Nepal. *BMC Pregnancy Childbirth* 21: 82.

Chakraborty, Nirali M., Karen Chang, Benjamin Bellows, Karen A. Grépin, Waqas Hameed, Amanda Kalamar, Xaher Gul, Lynn Atuyambe, and Dominic Montagu. 2019. “Association between the Quality of Contraceptive Counseling and Method Continuation: Findings from a Prospective Cohort Study in Social Franchise Clinics in Pakistan and Uganda.” *Global Health Science and Practice* 7 (1): 1–16.

Choi, Yoonjoung. 2018. Estimates of side effects counseling in family planning using three data sources: Implications for monitoring and survey design. *Studies in Family Planning*, 49: 23–39.

Demographic and Health Surveys Program (DHS). 2018. “Women’s Individual Questionnaire.” https://dhsprogram.com/pubs/pdf/DHSQ7/DHS7-Womens-QRE-EN-17Dec2018-DHSQ7.pdf.

FP2020. 2019. https://www.familyplanning2020.org/measurement-hub#what-we-measure.

Glick, Peter. 2009. “How Reliable are Surveys of Client Satisfaction with Healthcare Services? Evidence from Matched Facility and Household Data in Madagascar.” *Social Science & Medicine* 68 (2): 368–379.

Hall, Kelli Stidham, Katharine O’Connell White, Nancy Reame, and Carolyn Westhoff. 2010. “Studying the Use of Oral Contraception: A Review of Measurement Approaches.” *Journal of Women’s Health* 19 (12): 2203–2210.

Holt, Kelsey, Icela Zavala, Ximena Quintero, Danielle Hessler and Ana Langer. 2019. Development and validation of the client-reported quality of contraceptive counseling scale to measure quality and fulfillment of rights in family planning. *Studies in Family Planning* 50 (2): 137–158.

Jain, Aparna, Kumudha Aruldas, Arupendra Mozumdar, Elizabeth Tobey, and Rajib Acharya. 2019. “Validation of Two Quality of Care Measures: Results from a Longitudinal Study of Reversible Contraceptive Users in India.” *Studies in Family Planning* 50 (2): 179–193.

Jain, Anrudh K. 2018. “A New Composite Index to Measure National-level Quality of Family Planning Programs.” *International Perspectives on Sexual and Reproductive Health* 44 (2): 63–72.

Kenya National Bureau of Statistics – KNBS and ICF Macro. 2010. Kenya Demographic and Health Survey 2008–2009. Calverton, MD: KNBS and ICF Macro.

Macaskill, P., Constantine Gatsonis, Jonathan Deeks, R. Harbord, and Y. Takwoingi. 2010. *Cochrane Handbook for Systematic Review of Diagnostic Test Accuracy*: Chapter 10: Analysing and Presenting Results, Version 1.0. 1–61.

Marchant, Tanya, Zulfiquar A. Bhutta, Robert Black, John Grove, Catherine Kyobutungi, and Stefan Peterson. 2019. “Advancing Measurement and Monitoring of Reproductive, Maternal, Newborn and Child Health And Nutrition: Global and Country Perspectives.” *BMJ Global Health* 4:e001512.

MEASURE Evaluation. 2016. *Quick Investigation of Quality (QIQ): A User’s Guide for Monitoring Quality of Care in Family Planning*, 2nd edition. Chapel Hill, NC: University of North Carolina.

McCarthy, Katharine J., Ann K. Blanc, Charlotte Warren, James Kimani, Brian Mdawida, and Charity Ndwidga. 2016. “Can Surveys of Women Accurately Track Indicators of Maternal and Newborn Care? A Validity and Reliability Study in Kenya.” *Journal of Global Health* 6 (2): 020502. doi: 10.7189/jogh.06.020502.

McCarthy, Katharine J., Ann K. Blanc, Charlotte Warren, Ashish Bajracharya, and Benjamin Bellows. Forthcoming. “How Respondent and Facility Characteristics Influence the Accuracy of Self-Reported Maternal and Newborn Care: A Meta-Analysis.” Under review.

Munos, Melinda K., Ann K. Blanc, Emily D. Carter, Thomas P. Eisele, Steve Gesuale, Joanne Katz, Tanya Marchant, Cynthia K. Stanton, Harry Campbell, and Improving Coverage Measurement Group. 2018. “Validation Studies for Population-Based Intervention Coverage Indicators: Design, Analysis, and Interpretation.” *Journal of Global Health* 8 (2): 020804. doi: 10.7189/jogh.08.020804.

Directorate General for Health, and ICF Macro. 2011. *Cambodia Demographic and Health Survey 2010*. Phnom Penh, Cambodia: National Institute of Statistics.

PMA 2020. 2019. https://www.pma2020.org/questionnaires.

RamaRao, Saumya, and Mohanam Raji. 2003. “The Quality of Family Planning Programs: Interventions, Concepts, Measurements, and Effects.” *Studies in Family Planning* 34 (4): 227–248.
Simmons, Ruth, and Christopher Elias. 1994. “The Study of Client-Provider Interactions: A Review of Methodological Issues.” *Studies in Family Planning*. 25 (1): 1–17.

Sprockett, A. 2016. “Family Planning Quality Assessment Tools used in Low- and Middle-Income Countries: Review for Application in Clinic-Based Services” In Leisher SH, Sprockett A, Longfield K, and Montagu D (eds.) *Quality Measurement in Family Planning: Past, Present, Future: Papers from the Bellagio Meeting on Family Planning Quality*. Oakland, CA: Metrics for Management. pp. 95–106.

Stanton, Cynthia K., Barbara Rawlins, Mary Drake, Matias dos Anjos, David Cantor, Lidia Chongo, Leonardo Chavane, Maria da Luz Vaz, and Jim Ricca. 2013. “Measuring Coverage in MNCH: Testing the Validity of Women’s Self-Report of Key Maternal and Newborn Health Interventions during the Peripartum Period in Mozambique.” *PLoS One* 28 (5): e60694.

Sudhinaraset, May, Patience A. Afulani, Nadia Diamond-Smith, Ginger Golub, and Aradhana Srivastava. 2018. “Development of a Person-Centered Family Planning Scale in India and Kenya.” *Studies in Family Planning* 49 (3): 237–258.

Track20. 2019. http://www.track20.org/.

Tumlinson, Katherine, Ilene S. Speizer, Siân L. Curtis, and Brian W. Pence. 2014. “Accuracy of Standard Measures of Family Planning Service Quality: Findings from the Simulated Client Method.” *Studies in Family Planning* 45 (4): 443–470.

Tumlinson, Katherine. 2016. “Measuring Quality of Care: A Review of Previously Used Methodologies and Indicators.” Working Paper 2 of the Measuring and Monitoring Quality of Services and Quality of Care Project, Population Council, New York.

Warren, Charlotte, Timothy Abuya, Francis Obare, Joseph Sunday, Rebecca Njue, Ian Askew, and Ben Bellows. 2011. “Evaluation of the Impact of the Voucher and Accreditation Approach on Improving Reproductive Health Behaviors and Status in Kenya.” *BMC Public Health* 11: 177. doi:10.1186/1471-2458-11-177.

**ACKNOWLEDGMENTS**

This work was supported by the Bill & Melinda Gates Foundation through the Improving Measurement and Program Design (IMPROVE) Grant (OPP1172551) to the Johns Hopkins Bloomberg School of Public Health. The authors acknowledge and thank the Voucher study team who collected and allowed the use of the data for the present secondary analysis. They also appreciate the input received from the Core Group of the IMPROVE project and helpful comments from two reviewers.