Comparing simulated client experiences with phone survey self-reports for measuring quality of information given in family planning counseling: The case of depot medroxyprogesterone acetate – subcutaneous (DMPA-SC) in Nigeria [version 2; peer review: 2 approved with reservations]

Previously titled: Comparing simulated client experiences with phone survey self-reports for measuring quality of family planning counseling: The case of depot medroxyprogesterone acetate – subcutaneous (DMPA-SC) in Nigeria

Nadia Diamond-Smith 1, Emily Treleaven 1, Elizabeth Omoluabi 2, Jenny Liu 1

1 University of California San Francisco Medical Center, San Francisco, California, 94158, USA
2 Center for Research, Evaluation Resources and Development, Ile-Ife, Osun, Nigeria

Abstract

Background: The quality of family planning services can have important implications for uptake and continued method use. The aim of this analysis is to examine aspects of quality related to information provided for a new injectable contraceptive method, DMPA-SC (depot medroxyprogesterone acetate – subcutaneous, known as Sayana Press®), service provision and contraceptive services more broadly in Nigeria.

Methods: We compared self-reports from follow-up phone surveys with users to simulated client interactions that were designed to measure the same concepts. Through mixed-methods, we sought to more deeply understand the biases associated with different data collection methods that ultimately lead to different conclusions regarding quality of information provided in contraceptive services, and to further assess to what extent these methods were suitable for detecting differences in quality across sub-groups using the case of married versus unmarried women.

Results: We found that simulated clients reported lower levels of informational quality across all comparable quality indicators than phone survey respondents attending the same facilities. Both methods were able to detect differential treatment by marital status.

Conclusions: A mixed-methods approach can provide differential insights into informational quality of family planning services, especially when aiming to understand both objective and subjective aspects of quality.

Keywords

contraception, quality, Sayana Press, mystery clients
Corresponding author: Nadia Diamond-Smith (nadia.diamond-smith@ucsf.edu)

Author roles: Diamond-Smith N: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation;
Treleaven E: Formal Analysis, Writing – Original Draft Preparation; Omoluabi E: Investigation, Methodology, Writing – Review & Editing; Liu J: Conceptualization, Funding Acquisition, Investigation, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Bill & Melinda Gates Foundation [OPP1133271] This work was also supported by the Children's Investment Fund Foundation, London, UK. Program officers from these Foundations contributed to the overall objectives of the research portfolio and provided feedback on research protocols and instruments. However, the authors had sole responsibility and authority for the collection, analysis and interpretation of data, the writing of the report, and in the decision to submit this article for publication. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Diamond-Smith N et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Diamond-Smith N, Treleaven E, Omoluabi E and Liu J. Comparing simulated client experiences with phone survey self-reports for measuring quality of information given in family planning counseling: The case of depot medroxyprogesterone acetate – subcutaneous (DMPA-SC) in Nigeria [version 2; peer review: 2 approved with reservations] Gates Open Research 2019, 3:1092 (https://doi.org/10.12688/gatesopenres.12935.2)

Background

Global efforts to reduce unintended pregnancies and abortions and to increase utilization of contraceptive services include a range of programs and interventions specifically targeted towards improving quality of care\(^1\). However, what constitutes quality care and how to measure it is complex. The Bruce-Jain framework (1994) for understanding the quality of contraceptive services included six domains: choice of method; information to client; technical competence; interpersonal relations; mechanisms to encourage continuity; and constellation of services\(^2\). Other frameworks have expanded upon this framework to include additional domains, such as access/accessibility and patient-centeredness\(^3\).

Furthermore, a standard method for measuring quality in contraceptive services does not exist. Many measures primarily focus on technical aspects of quality that can be assessed objectively (e.g., facility stocks and infrastructure, provider competence)\(^4\). Even studies that ask actual clients often rely on more objective questions, and struggle to capture their experiences of interpersonal or person-centered quality. Moreover, objective aspects of quality and clients' subjective perceptions may not align\(^5\). Tumlinson et al. (2014) detail the advantages and disadvantages of commonly-used methods for measuring the quality of contraceptive services, including facility audits, direct observations, provider interviews, and client interviews, highlighting the fact that different methods yield different insight into different aspects of quality. For example, facility audits might identify issues related to stockouts, while direct observation may be useful for assessing provider competence. User interviews may be especially useful for assessing subjective aspects of quality, including satisfaction and client perceptions of quality. However, these methods are not without their limitations for measuring service quality, including courtesy bias, lack of reliability in self-reports, the Hawthorn effect, and recall bias.

Another methodological approach becoming increasingly common is the simulated client or “mystery client” method, wherein trained actors are sent to seek contraceptive services from a provider following a scripted interaction guide. These simulated clients then record the details of their interactions in a standard survey or interview format after they leave the facility. Ideally, this method allows for more standardization than asking actual clients or providers themselves as the actors can be trained to look for certain indicators of quality, and to have a standardized expectation of quality\(^6\). Recently, Tumlinson et al. (2014) used simulated clients to test the validity of client exit surveys and provider interviews (self-reported measures), and direct observations (conducted by a third party) in measuring quality; they found low specificity and positive predictive values of quality in all of these approaches. These results suggest that using different methodological approaches is necessary for understanding the complexity of women’s experiences of quality of care for contraceptive services, and that exploring the differences arising from such methods may yield new insights into how to improve both the clinical aspects and the patient-centeredness of care received.

Given these insights, we chose to use mixed methods—namely simulated clients and client self-reports via a follow-up phone survey—to assess the quality of services received by women seeking contraceptive services and particularly for obtaining DMPA-SC (depot medroxyprogesterone acetate – subcutaneous, also known as Sayana Press®), a new injectable contraceptive recently introduced into the Nigeria market. Furthermore, because this new delivery method for DMPA was specifically designed to reduce barriers to contraceptive uptake for underserved populations, including adolescent and young women\(^7\), we sought to additionally examine differences in contraceptive services experienced by unmarried women compared to married women. The overarching aim was to understand how these different methodological approaches (simulated clients and phone surveys) compare for measuring one domain of quality (arguably the most objective, informational), and if both are able to detect inequitable treatment by client characteristics.

Despite recent increases, contraceptive use is generally low in Nigeria, reaching 11.1% in 2013 for modern methods\(^8\). Injectable contraceptives are increasingly popular, and comprise the largest increase in the contraceptive prevalence rate among married women in the past three years\(^9\). However, these gains are not equal. Although unmarried Nigerian women are more likely to use modern contraception, they are also at increased risk for unintended pregnancies, suggesting that consistency in use of effective methods remains poor\(^10\). In general, adolescents and young adults have substantial unmet need for sexual and reproductive health services and related information\(^11\) and many avoid seeking services due to poor quality of care.

Unlike other countries where DMPA-SC has been introduced and studied as small pilot projects\(^12\)-\(^14\), DMPA-SC was introduced in Nigeria in 2015 through multiple channels—facilities, drug shops, and specially-trained community-based distributors—in the private health care sector which accounts for the majority of all health services provided, and in particular for contraceptive services\(^15\). This presented an opportunity to assess actual user experiences with DMPA-SC from the broader population of new and continuing contraception users under real world conditions. Our data collection focused on DMPA-SC service provision from selected private sector providers across seven states in South West Nigeria (where the introductory efforts were concentrated) in 2016. Since this time DMPA-SC distribution has expanded to the government sector. In order to more fully understand the experiences of married and unmarried users, we chose to use multiple methods to assess various aspects of quality of service provision. In this paper, we directly compared...
the quality of care along specific measures collected from both the simulated clients sent to providers selling DMPA-SC and the follow-up phone surveys with women who recently obtained DMPA-SC at the same facility. To understand whether these varied methodological approaches could identify the same differences in the quality of care among population sub-groups, we further analyzed the responses by marital status.

**Methods**

**Sampling frame**

Data was collected between March and October 2016 in seven states in south-western Nigeria (Ekiti, Kwara, Lagos, Ogun, Ondo, Osun, Oyo). These states were targeted for the initial private sector introduction of DMPA-SC.

From March to May 2016, we recruited a convenience sample of providers who had purchased at least 25 units of DMPA-SC from the distributor. In addition to being on the list provided by the distributor and having purchased the amount of DMPA-SC state previously, there were no specific inclusion criteria, and providers included private clinics or hospitals, pharmacies or retail drug outlets, private providers employed at government clinics or hospitals, and specially-trained community-based distributors (qualified as licensed Community Health Extension Workers). In total, 205 providers consented to participate in the study. Each was asked to help recruit women who purchased DMPA-SC from them for follow-up survey and interview. Inclusion criteria included the woman having purchased any type of injectable contraceptive from that provider. Providers were instructed to ask clients who received an injectable contraceptive (of any type) if they were willing to be called for a short phone interview and to record their contact information if they consented. Providers received a small incentive (1500 Naira or ~US$4.25 of mobile phone credits) for keeping the list of potential respondents. After dropping 76 providers who did not record any injectable contraceptive user willing to be contacted and two public facilities that were misclassified as private providers, 127 providers constituted the sampling frame for the resulting phone survey and simulated client interactions.

**Phone survey of DMPA-SC users**

All women purchasing DMPA-SC at an enrolled provider site and who consented to be contacted, as described above at the time of purchase (N=994) were called to complete a survey administered over the phone, usually lasting about 15–20 minutes. The survey was administered by a trained, bilingual (English and Yoruba, the dominant local language) interviewer over the phone. Respondents were compensated with 200 Naira (~US$0.57) of mobile phone credits for completing the survey. A total of 541 women completed the phone survey, about half within one month of their visit\(^1\). Respondents were asked about demographic and socioeconomic characteristics, prior contraceptive use, as well as the quality measures discussed below.

**Simulated clients**

From the 127 providers who provided phone survey respondents, we purposively selected a subsample of 60 providers with which to conduct simulated client interactions. Providers were first stratified by channel type, and then further stratified into four categories based on DMPA-SC client volume: (1) high: 30+ clients; (2) medium: 10–29 clients; (3) low: 1–9 clients; and (4) very low: 0 clients. Within each channel, providers were then selected based on the following procedures, which were slightly different for each channel due to the particularities of the resulting sample.\(^1\) For hospitals/clinics, we selected 13 facilities that had sold any DMPA-SC and that recorded the highest volumes of other (i.e., non-DMPA-SC such as Depo Provera or Noristerat) injectable sales, and two high volume facilities for other injectables but which had low volumes of DMPA-SC. For pharmacies and retail drug outlets, we selected all facilities that had sold any DMPA-SC and that recorded the highest volumes of other injectables until a total of 15 facilities were identified in each channel. For community-based distributors, we purposively selected a mix of high, medium and low volume agents proportional to the total number of distributors in each volume category, while also aiming to achieve representation across the six states in which they were found (one state did not have a qualifying distributor in the sampling frame at the time). When data collection was nearly completed, a status review found three facilities where simulated client interactions could not be completed (one clinic and two community-based distributors) after multiple attempts. These providers were replaced with a provider selected from the same channel within the same DMPA-SC volume stratum.

At each selected provider, two different simulated client interactions were conducted reflecting two profiles of women: (1) a sexually active, unmarried adolescent woman aged 18 without children, seeking a contraceptive method for pregnancy prevention (“unmarried”), and (2) a married woman aged 28 with two children, seeking a contraceptive method for birth spacing (“married”). Each provider was approached two times about one week apart—one interaction for each of the two profiles. A total of eight trained simulated client actors (four for each profile) were sent to selected providers to follow a scripted interaction (Please see in Extended data\(^2\)). The actors were standardized in terms of age-appropriate attire typical of a middle-income woman. Actors were trained to approach the provider stating that she was interested in getting contraception and to ask for someone who could help her. The provider was then allowed to lead the counseling conversation. Only at the end of the session was the actor instructed to ask specifically about DMPA-SC if it was not already mentioned by the provider. Simulated clients did not purchase a method. Providers were not informed of the visit ahead of time. After the visit, simulated clients immediately (within 30 minutes) completed a short survey administered by another member of the research team about her interaction with the provider. The survey was conducted in a location where the provider could not see the

\(^1\)Our original sampling protocol called for selecting the 15 highest-volume DMPA-SC facilities within each channel. However, for all channels except community-based distributors, the number of facilities that had sold any DMPA-SC was insufficient to achieve the targeted 15 per category. Therefore, we adjusted the sampling strategy by channel to accommodate these limitations.
simulated client (e.g., in the car down the street), but as soon as possible to optimize recall. Of the 117 completed interactions in which the actor was able to successfully engage the provider with her initial inquiry for contraceptive information, 112 interactions were completed pairwise for the married and unmarried profiles at the same provider location. We restrict our analysis to this subset of pairwise, completed interactions.

Measures of quality and analysis
For the analysis of phone survey responses and simulated client interactions, we focused on indicators for the technical aspects of quality of care that were captured in both the phone survey and simulated client survey. We chose these measures to facilitate the direct comparison because they are arguably more objective, relying less on users’ or clients’ subjective interpretation of how the interaction unfolded. In all, there were eight items included in the technical competency domain, which primarily focused on if the provider had asked the client about past contraceptive use, experiences of side effects, complicating health factors, childbearing goals, pregnancy status, expected side effects of DMPA-SC, instructions for dealing with problems, and information on length of pregnancy prevention protection (see Table 3). For each item, questions in the phone and simulated client surveys were worded and structured as similarly as possible. Each user or simulated client actor was asked to respond “yes” or “no” to each item question, and a dummy variable indicator was constructed for each “yes” response.

For each data source, we calculate the overall sample response frequency for each technical competency quality item measure. We also calculate the responses by marital status. Simple chi-square tests were then used to explore differences by marital status or profile within each data collection approach. Data was analyzed using STATA version 15.

Ethical approval
This study received ethical approval from the Institutional Review Board at the University of California, San Francisco (IRB# 15–18,353), and from the National Health Research Ethics Committee of the Ministry of Health, Government of Nigeria (NHREC/01/01/2007–06/01/2016).

Results
The phone survey included 541 women who had visited a private facility for DMPA-SC and were using it (see Underlying data). As shown in Table 1, almost two-thirds (65.8%) of phone survey participants visited a community-based distributor to obtain DMPA-SC, followed by private hospital/clinic/maternity homes (11.1%), and government hospital/clinic/maternity home (private providers employed at public facilities) (11.7%). About 6% of women attended a pharmacy, and 3.7% attended a retail drug shop. Few women attended private doctors or nurses. The majority of women were currently married (93.0%) and age 25 or older (only 9% were under 25). Most switched from another modern method, with just over one-quarter being new users of contraception. Simulated clients completed visits for both profiles at a total of 122 facilities (Table 2). The majority of these facilities or providers were located in urban areas (72.1%), with most of the remainder in peri-urban areas (24.6%). Less than half of providers visited were female (42.6%).

Phone survey responses compared to simulated client findings
Both methods record a perceived difference in quality of care received by married and unmarried family planning clients, with unmarried younger women receiving less favorable treatment. The main difference between both methods is in the magnitude of this difference. Overwhelmingly, clients interviewed in the phone survey reported higher levels of quality and less variation across item measures compared to simulated clients (Table 3). Compared to the simulated clients, about twice as many women in the phone survey reported being asked

<table>
<thead>
<tr>
<th>Table 1. Phone survey sample characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>State</strong></td>
</tr>
<tr>
<td>Ekiti</td>
</tr>
<tr>
<td>Lagos</td>
</tr>
<tr>
<td>Kwara</td>
</tr>
<tr>
<td>Ogun</td>
</tr>
<tr>
<td>Ondo</td>
</tr>
<tr>
<td>Osun</td>
</tr>
<tr>
<td>Oyo</td>
</tr>
<tr>
<td><strong>Type of facility or provider attended</strong></td>
</tr>
<tr>
<td>Community-based distributor</td>
</tr>
<tr>
<td>Retail drug outlet</td>
</tr>
<tr>
<td>Pharmacy</td>
</tr>
<tr>
<td>Private hospital/clinic/maternity home</td>
</tr>
<tr>
<td>Government hospital/clinic/maternity home</td>
</tr>
<tr>
<td>Private doctor/nurse</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td><strong>Age group</strong></td>
</tr>
<tr>
<td>&lt;25</td>
</tr>
<tr>
<td>25–34</td>
</tr>
<tr>
<td>35+</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
</tr>
<tr>
<td>Not currently married</td>
</tr>
<tr>
<td>Currently married</td>
</tr>
<tr>
<td>Don’t know/No response</td>
</tr>
<tr>
<td><strong>Prior family planning use</strong></td>
</tr>
<tr>
<td>New user</td>
</tr>
<tr>
<td>Switched from other modern method</td>
</tr>
<tr>
<td>Switched from traditional method</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>
if they had used contraception before, or asked if they wanted more/any children. Phone survey respondents were more than three times as likely as simulated clients to report being asked about other health problems, and asked about or tested for pregnancy. Phone survey respondents were almost three times as likely as simulated clients to report that they were described side effects of DMPA-SC. The most similar element of quality between the different method approaches was reporting being told how long DMPA-SC protects against pregnancy (82.3% of simulated clients compared to 99.1% of phone survey respondents). While at least 70% of phone survey respondents reported experiencing each measure of quality asked, length of protection of pregnancy was the only measure of quality to surpass this level of coverage among simulated clients.

**Differences in quality of care between unmarried and married clients (simulated clients)**

The simulated client data yielded similar results from the user interviews about lower quality of care received by unmarried contraceptive clients. Unmarried simulated clients reported being asked if they were currently pregnant or wanted to have any children in the future significantly less frequently than married simulated clients. They were less likely than married simulated clients to report being asked if they had used contraception before. They were also significantly less likely than married simulated clients to report being told how long DMPA-SC protects against pregnancy.

**Discussion**

To assess the quality of one component of contraceptive counseling in the provision of DMPA-SC delivered through private sector providers in South West Nigeria, we compared measures of technical quality from users’ experiences self-reported in a follow-up phone survey and with simulated client visits. Our findings suggest that different methodological approaches can yield similar results but with varying intensity about the quality of information provided. Both methods were able to detect systematic differences in technical quality for women of different marital statuses. Across all measures, phone survey respondents perceived higher levels of technical quality than simulated clients, however, the direction of the difference in quality of care between young unmarried clients and older married clients was similar: younger unmarried women received lower technical quality of care for contraceptive services. In sensitivity analyses, we restricted the analyses to the same set of providers who saw both simulated clients and phone survey respondents; responses were similar in magnitude and direction.

These findings support previous research suggesting that asking women directly about their experiences can lead to higher quality scores than more objective approaches (such as simulated clients), even for aspects of technical quality that may be less open to subjective rating and only require a yes/no response. While actual users’ responses may better capture their perceived experiences, these results may be less useful to program implementers because they are determined by the socio-cultural expectations governing such interactions, which in this case, are reflective of a population that may have low expectations of quality of care. In contrast, a key advantage of the simulated client approach is the standardization of perspectives on quality, which ultimately yield greater variation in quality measurements with which to identify where improvements in service provision are needed.

In fact, our findings coupled with our experiences in training our simulated client actors suggest that self-reported user perceptions may reflect more reflexive reactions or reactions based on social expectation, rather than a more detached, independent assessment of quality that researchers or program implementers intend. Many of our simulated client actors similarly

---

**Table 2. Sample of providers visited by simulated clients.**

<table>
<thead>
<tr>
<th>State</th>
<th>Providers N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ekiti</td>
<td>18 (14.8)</td>
</tr>
<tr>
<td>Lagos</td>
<td>44 (36.1)</td>
</tr>
<tr>
<td>Kwara</td>
<td>12 (9.8)</td>
</tr>
<tr>
<td>Ogun</td>
<td>8 (6.6)</td>
</tr>
<tr>
<td>Ondo</td>
<td>10 (8.2)</td>
</tr>
<tr>
<td>Osun</td>
<td>10 (8.2)</td>
</tr>
<tr>
<td>Oyo</td>
<td>20 (16.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of facility</th>
<th>Providers N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community-based distributor</td>
<td>30 (24.6)</td>
</tr>
<tr>
<td>Retail drug outlet</td>
<td>30 (24.6)</td>
</tr>
<tr>
<td>Pharmacy</td>
<td>30 (24.6)</td>
</tr>
<tr>
<td>Private hospital/clinic/maternity home</td>
<td>32 (26.2)</td>
</tr>
<tr>
<td>Government hospital/clinic/maternity home</td>
<td>0</td>
</tr>
<tr>
<td>Private doctor/nurse</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location of facility</th>
<th>Providers N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>88 (72.1)</td>
</tr>
<tr>
<td>Peri-urban</td>
<td>30 (24.6)</td>
</tr>
<tr>
<td>Rural</td>
<td>4 (3.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender of provider</th>
<th>Providers N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>52 (42.6)</td>
</tr>
<tr>
<td>Male</td>
<td>70 (57.4)</td>
</tr>
<tr>
<td>Total</td>
<td>122</td>
</tr>
</tbody>
</table>

They were also less likely to report having been told of any likely side effects of DMPA-SC, or told what to do if they experienced side effects compared to their married counterparts.
perceived higher levels of quality in role play during initial training sessions than after eventual calibration. When mock patient-provider interactions were analyzed more systematically to standardize ratings, actors’ perceptions of quality declined, particularly among unmarried profile actors. Compared to the older women actors assuming the married profile, our younger women actors did not seem to expect to receive higher quality care, and thus initially rated the mock interactions to be higher quality than older women for the same observed encounter. Thus, questions about the quality of clinical care and counseling may be less salient for unmarried women seeking contraceptive services who do not expect to receive higher quality of care or who may have few reference points with which to assess relative quality of care. Other studies have found that younger contraceptive users in Nigeria have limited contact with contraceptive services\(^1\). However, when expectations of quality are standardized across individuals, unmarried women may be more likely to recognize relatively worse quality of care, potentially explaining the larger magnitude of differences by marital status found with simulated clients as compared to phone survey respondents. Thus, users’ assessments of the quality of contraceptive service provision should be interpreted in light of prevailing social and cultural expectations, which may differ across subgroups. This also suggests that simulated clients, and technical aspects of quality, might not be as objective as we had initially assumed. Careful training is required, and thought should go into differential biases/expectations of the actors themselves that they may bring into their work.

Additionally, women’s expectations of quality may not necessarily align with programmatic or international standards of quality. While our results suggest that user self-reports may provide a better summary depiction of their lived experiences, it may behoove researchers to question if the common aspects of quality adopted by researchers and practitioners are salient to the women represented by the sample. While this study was only able to directly compare quality measures for a limited subset of technical competency measures purposely chosen for that purpose, the relevant dimensions of quality for a particular population may be highly context-dependent. For example, recent research on person-centered quality of care for childbirth in Kenya found that even rural and urban populations of the same country identified different important indicators for quality\(^2\).

### Table 3. Proportion of respondents reporting technical competency quality measures in the phone survey compared to simulated client reports.

<table>
<thead>
<tr>
<th>Quality measure (Did the provider…)</th>
<th>Phone Survey</th>
<th>Simulated Clients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All (N=536)</td>
<td>Unmarried (N=33)</td>
</tr>
<tr>
<td>…ask if you had ever used family planning before? (Yes)</td>
<td>89.4</td>
<td>75.8</td>
</tr>
<tr>
<td>…ask if you had ever experienced any side effects? (Yes)</td>
<td>84.9</td>
<td>75.8</td>
</tr>
<tr>
<td>…ask if you had had any other health problems (e.g., infections, high blood pressure)? (Yes)</td>
<td>74.8</td>
<td>42.4</td>
</tr>
<tr>
<td>…ask if you wanted to have any/more children in the future? (Yes)</td>
<td>87.4</td>
<td>75.8</td>
</tr>
<tr>
<td>…do anything to find out if you were pregnant?/Ask if you were pregnant? (Yes)</td>
<td>91.6</td>
<td>84.9</td>
</tr>
<tr>
<td>… describe any side effects of Sayana Press? (Yes)</td>
<td>71.4</td>
<td>48.5</td>
</tr>
<tr>
<td>… tell you what to do if you had problems with the Sayana Press injection? (Yes)</td>
<td>71.9</td>
<td>54.6</td>
</tr>
<tr>
<td>… tell you how long it protects against pregnancy?/Tell you the accurate amount of time? (Yes)</td>
<td>99.1</td>
<td>100.0</td>
</tr>
</tbody>
</table>

*Respondents with unknown marital status excluded (N=5)
These differences in quality ratings between methods were similar to those found by Tumlinson et al. between their simulated clients and user self-reports in exit interviews. Thus, even though our phone surveys were conducted with some delay after the service was rendered rather than immediately afterward vis-a-vis exit interviews, recall bias for more technical quality measures did not appear to affect overall user ratings.

Several study limitations should be noted. First, while simulated clients went to the same facilities as phone survey respondents, this does not mean that they saw the same provider. This is most likely especially true for the larger facilities, but also the pharmacies. Thus, the comparison between these two methods may not be reflective of interactions with the same provider, only the same facility. Phone survey interviews with actual users were conducted within a few weeks of the provider interaction, which may introduce recall bias, unlike with our simulated client data obtained from clients debriefed within 30 minutes of their interaction with providers. It is likely that phone survey clients might be more likely to recall especially positive or negative interactions, potentially skewing the results in either extreme. As with most data collection, courtesy bias is a concern from the phone survey respondents and on the flip side, simulated clients may have been primed in their training to be overly negative. Although simulated client interactions and profiles were standardized across actors as much as possible, providers may still perceive actors to be members of better educated and wealthier classes given their mannerisms and speech. Due to the large geographic area covered by the DMPA-SC program under study, local actors from targeted communities could not be recruited for all selected localities. Additionally, due to IRB restrictions, simulated clients did not purchase a method and it is possible that clients would receive additional information after purchasing a method.

Respondents for the phone survey were disproportionately drawn from clients of community-based distributors, who were more likely to provide services to unmarried women, than other provider types. This could lead to different types of provider-client interactions that could have impacted reports of quality, due to cultural acceptance of contraceptive use among unmarried women or different power dynamics. Yet, most of our results still hold when the phone survey sample is restricted to the same set of providers included in the simulated client interactions, although this sample size is too small to make any claims of significance. The providers from which we recruited phone survey respondents were a convenience sample; a design which randomly selected providers would have reduced this bias. Only a 54% of women responded to the request to participate, resulting in selection bias. Perhaps women who had disproportionately bad or good experiences or certain types of women (older, married, etc) were more or less likely to respond. A relatively small proportion of our respondents were young and unmarried, a prime focus of our analysis, and we do not know if this is reflective of the actual population seeking contraceptives or selection bias. If younger, unmarried women who were willing to participate were more likely than older married women to participate if they had very good or very bad experiences this could be problematic for our findings. The small sample of unmarried women thereby weakens our analysis and the implications of our findings. Higher incentives, additional attempts at recruitment or different recruitment approaches could have reduced selection bias. Finally, we only focused on one measure of quality—technical quality—thus our findings are not generalizable to other domains of quality. As discussed above, we did this to try to make the simulated client and phone survey findings more comparable under the assumption that this domain was the least subjective.

Conclusions
Measuring the quality of contraceptive provision is essential for improving the experiences of clients, and subsequently increasing the uptake and continuation of contraceptive use. However, measuring the quality of contraceptive services has proven to be difficult as different approaches have specific limitations and advantages. Our study has compared two different approaches to measuring one domain of quality of contraceptive services in the private healthcare sector—technical quality mostly related to information provided and questions asked. Both approaches confirmed that unmarried clients receive lower technical quality of care compared to married clients. The simulated client method is more time- and resource-intensive than exit or follow-up interviews. However, this method has the potential to provide a more standardized perspective on the quality of the provider-client interaction, while still collecting valuable information on important differences in user experiences across subgroups. It is not affected by recall bias unlike phone interviews conducted several weeks after the interaction. It can also allow for better control of the effects of social expectations unlike all other direct interviews of actual clients, including client exit. Depending on the research and intervention agenda, having standardized, comparable data might be most beneficial, for example, when comparing programs or across populations. On the other hand, other research might benefit more from understanding women’s lived experiences of provider interactions and quality, as well as their expectations and perceptions of specific aspects of quality, for example when aiming to improve a specific program. These findings can help future researchers select the most appropriate methodological approach for their research goals.

Data availability
Underlying data
Harvard Dataverse: Monitoring and evaluation data from the introduction of DMPA-SC in Nigeria’s public and private sectors. https://doi.org/10.7910/DVN/XCLNIS

This project contains the following underlying data:

• DKT Data Registry

  • Customer register data
  • Customer registers codebook
  • Phone Survey codebook
  • Phone survey data
This projects contains the following extended data:

- Provider Survey codebook
- Provider survey data
- User Survey_Round 1 (phone user survey at baseline)
- User Survey_Round 2 (phone user survey at 3 months)
- Mystery client guide (guide to actors being mystery clients)
- Mystery client debrief (questionnaire administered following mystery client attending a facility)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Acknowledgements

We would like to thank the research assistants at Akena Associates, as well as the participants who gave their time to this study. We would also like to thank Maia Sieverding.

References

Open Peer Review

Current Peer Review Status: ? ?

Version 2

Reviewer Report 10 January 2020

https://doi.org/10.21956/gatesopenres.14249.r28390

© 2020 Hardee K. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Karen Hardee
1 Hardee Associates, Arlington, VA, USA
2 What Works Association, Morristown, NJ, USA

The authors have incorporated most of my original comments. In addition to further explanation about the differences in the samples of simulated clients and phone respondents, the paper is generally more clear about the aspect of quality of care that is included in their study. However, reading through the revised version of the paper, there is still some confusion about what is being studied. I suggest being clear at the beginning what aspect of quality of care is covered in the paper and refer to that aspect throughout the rest of the paper, including in the discussion.

The paper includes a range of terms, including “quality of information,” “informational quality,” “to assess the quality of services,” “assess various aspects of quality of service provision,” “quality of counseling,” “technical competency domain,” “measures of technical quality,” “client-provider interaction,” “quality of care,” “systematic differences in quality,” “levels of quality,” “differences in quality care,” “lower quality of care for contraceptive services.”

The authors cite the Bruce/Jain framework early in the paper – it would be good to explain what aspect of the framework they are assessing – it seems that it is the domain of information given to clients (with a few items on information elicited from clients). That domain is different in the Bruce/Jain framework from the domain of technical competence – which is why it is confusing to see that term used in this paper. Or the authors could say they are studying counseling, and primarily eight items related to the information given to and elicited from clients. This would clarify the point that they are studying these items since they are more objective than some of the more subjective aspects of client-provider interaction. I would avoid using the term technical competency since that is confusing with the “technical competence” domain of the Bruce/Jain framework.

The section on measures of quality and analysis in the methods section has a sentence that needs editing (suggested added words underlined):

“In all, there were eight items included in the study (delete: the technical competency domain), which primarily focused on if the provider had asked the client about past contraceptive use, experiences of side effects, complicating health factors, childbearing goals, pregnancy status, and if the provider told the
client about expected side effects of DMPA-SC, gave instructions for dealing with problems, and provided information on length of pregnancy prevention protection (see Table 3)."

Please note that Table 1 refers to the Phone survey sample characteristics, so it is not clear why it is referenced in the sentence about the aspects of quality of care covered. Did the authors intend to have a table (which would be helpful) with the aspects of quality of care that are included in this study?

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Rights-based family planning; gender and family planning, including male engagement; population and climate change; implementation research; research to policy.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

**Siân L. Curtis**

Maternal and Child Health, Gillings School of Global Public Health, Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

This paper compares results from follow up phone interviews with DMPA-SC clients with results from mystery client interviews on eight items related to contraceptive counselling. The quality of family planning care is an area of high interest yet its measurement remains challenging so this paper addresses an important topic. There are few studies that have directly compared different approaches to measuring specific dimensions of quality of family planning care so I found this study interesting and a useful contribution to that literature. However it has a number of limitations that should be considered when reading the study and that could be addressed more specifically in the discussion.

1. The overall study design is pragmatic but has some structural limitations – samples are convenience samples and the response rate for the phone survey is around 54% (541 of 994 women called). This is not a bad response rate for a phone survey but means there is likely to be notable selection bias in who responded to the survey. The potential for selection bias and its implications should be more explicitly acknowledged in the limitations.

2. There are several important differences between the phone survey respondents and the mystery clients, which limit the comparability of the findings. While these are discussed to some degree, they are important. Specifically:
a) The phone interview clients are all women who obtained DMPA-SC while the mystery clients visited providers to obtain information on contraception options and did not accept a method (at least I think they didn’t – I don’t think that is explicitly stated in the paper). Is it possible that additional information might be provided to clients on their selected method once they have selected one and are moving forward with obtaining it which could account for at least some of the large differences in the results from survey respondents and mystery clients?

b) As noted by the other reviewer, the characteristics of survey respondents are quite different from those of mystery clients (most actual clients went to community-based distributors, few were under 25 and few were unmarried compared to mystery clients). More discussion of the implications of these differences for the study findings is warranted.

c) In the discussion it states that the phone survey respondents were interviewed within a few weeks of their interaction with the DMPA-SC provider. It would be useful to state that in the methods section too (p.4). The mystery clients were interviewed within 30 minutes so there is more scope for recall error among phone respondents, as noted in the discussion but the issue of recall errors warrants more attention (see also comment 3).

3. The authors explain on p.4-5 that they selected the eight items on technical counselling because they reflect arguably more objective aspects of quality that might be less subject to the client’s subjective interpretation of the interaction (unlike items related to client satisfaction and how they were treated for example). Yet in the Discussion, the authors focus mostly on the role of client expectations of quality as an explanation for the different responses from phone respondents and mystery clients, and don’t discuss recall issues much (see also comment 2c above). Does this mean that those questions are not objective after all? Is there room for interpretation in whether an item was covered with the mystery clients having a different interpretation of whether they were asked about other health problems etc.? I would be interested to know more about the training of mystery clients and how that changed the interpretation of these supposedly more objective items in the role play (p.6).

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly
**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Family planning evaluation and measurement; measurement of quality of care; contraceptive use dynamics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 04 Dec 2019

**Nadia Diamond-Smith**, University of California San Francisco Medical Center, San Francisco, USA

Thank you for your thoughtful comments, we have replied to each below:

1. The overall study design is pragmatic but has some structural limitations – samples are convenience samples and the response rate for the phone survey is around 54% (541 of 994 women called). This is not a bad response rate for a phone survey but means there is likely to be notable selection bias in who responded to the survey. The potential for selection bias and its implications should be more explicitly acknowledged in the limitations. We have added the following in the limitations section “The phone survey respondents were also a convenience sample, with only a 54% of women responding to the request to participate. This leads to selection bias, perhaps with women who had disproportionately bad or good experiences being more likely to respond. A design which randomly selected women being asked to participate and higher incentives could have reduced these biases.”

2. There are several important differences between the phone survey respondents and the mystery clients, which limit the comparability of the findings. While these are discussed to some degree, they are important. Specifically:a) The phone interview clients are all women who obtained DMPA-SC while the mystery clients visited providers to obtain information on contraception options and did not accept a method (at least I think they didn’t – I don’t think that is explicitly stated in the paper). Is it possible that additional information might be provided to clients on their selected method once they have selected one and are moving forward with obtaining it which could account for at least some of the large differences in the results from survey respondents and mystery clients?

   We have clarified in the methods that simulated clients did not purchase a method, and also added the following to the limitations “Additionally, due to IRB restrictions, mystery clients did not purchase a method and it is possible that clients would receive additional information after purchasing a method.”

b) As noted by the other reviewer, the characteristics of survey respondents are quite different from those of mystery clients (most actual clients went to community-based distributors, few were under 25 and few were unmarried compared to mystery clients). More discussion of the implications of these differences for the study findings is warranted.

   We have added the following to the limitations section “This could lead to different types of
provider-client interactions that could have impacted reports of quality, due to cultural acceptance of contraceptive use among unmarried women or different power dynamics.” And “The providers from which we recruited phone survey respondents were a convenience sample; a design which randomly selected providers would have reduced this bias. Only a 54% of women responded to the request to participate, resulting in selection bias. Perhaps women who had disproportionately bad or good experiences or certain types of women (older, married, etc) were more or less likely to respond. A relatively small proportion of our respondents were young and unmarried, a prime focus of our analysis, and we do not know if this is reflective of the actual population seeking contraceptives or selection bias. If younger, unmarried women who were willing to participate were more likely than older married women to participate if they had very good or very bad experiences this could be problematic for our findings. The small sample of unmarried women thereby weakens our analysis and the implications of our findings. Higher incentives, additional attempts at recruitment or different recruitment approaches could have reduced selection bias.”

c) In the discussion it states that the phone survey respondents were interviewed within a few weeks of their interaction with the DMPA-SC provider. It would be useful to state that in the methods section too (p.4). The mystery clients were interviewed within 30 minutes so there is more scope for recall error among phone respondents, as noted in the discussion but the issue of recall errors warrants more attention (see also comment 3).

We have added the following to the limitations section “It is likely that phone survey clients might be more likely to recall especially positive or negative interactions, potentially skewing the results in either extreme.” And we have added the length of time data was collected from simulated clients and phone survey respondents in the methods.

3. The authors explain on p.4-5 that they selected the eight items on technical counselling because they reflect arguably more objective aspects of quality that might be less subject to the client’s subjective interpretation of the interaction (unlike items related to client satisfaction and how they were treated for example). Yet in the Discussion, the authors focus mostly on the role of client expectations of quality as an explanation for the different responses from phone respondents and mystery clients, and don’t discuss recall issues much (see also comment 2c above). Does this mean that those questions are not objective after all? Is there room for interpretation in whether an item was covered with the mystery clients having a different interpretation of whether they were asked about other health problems etc.? I would be interested to know more about the training of mystery clients and how that changed the interpretation of these supposedly more objective items in the role play (p.6).

Yes, this is indeed what we struggled with in our analysis and discussions—if our assumptions that these quality items were more objective holds. We have expanded our discussion of this throughout the paper, and we believe that this section of the discussion addresses your other question about simulated clients specifically “In fact, our findings coupled with our experiences in training our simulated client actors suggest that self-reported user perceptions may reflect more reflexive reactions or reactions based on social expectation, rather than a more detached, independent assessment of quality that researchers or program implementers intend. Many of our simulated client actors similarly perceived higher levels of quality in role play during initial training sessions than after eventual calibration. When mock patient-provider interactions were analyzed more systematically to standardize ratings, actors’ perceptions of quality declined, particularly
among unmarried profile actors. Compared to the older women actors assuming the married profile, our younger women actors did not seem to expect to receive higher quality care, and thus initially rated the mock interactions to be higher quality than older women for the same observed encounter. Thus, questions about the quality of clinical care and counseling may be less salient for unmarried women seeking contraceptive services who do not expect to receive higher quality of care or who may have few reference points with which to assess relative quality of care. Other studies have found that younger contraceptive users in Nigeria have limited contact with contraceptive services (14). However, when expectations of quality are standardized across individuals, unmarried women may be more likely to recognize relatively worse quality of care, potentially explaining the larger magnitude of differences by marital status found with simulated clients as compared to phone survey respondents. Thus, users’ assessments of the quality of contraceptive service provision should be interpreted in light of prevailing social and cultural expectations, which may differ across subgroups. This also suggests that simulated clients, and technical aspects of quality, might not be as objective as we had initially assumed. Careful training is required, and thought should go into differential biases/expectations of the actors themselves that they may bring into their work.

**Competing Interests:** none

---

Karen Hardee  
1 Hardee Associates, Arlington, VA, USA  
2 What Works Association, Morristown, NJ, USA

This paper addresses an important topic – examining aspects of quality of care related to a new injectable contraceptive method (DMPA-SC). The authors used mixed methods (simulated clients and client self-reports via follow up phone call) to assess the quality of services received by women related to DMPA-SC, which has been recently introduced in Nigeria. The paper compares responses from simulated clients with responses from actual clients via phone interview. The findings from the study confirm results from other quality of care studies that have found that different aspects of quality of care are measured through different means and with different groups (e.g. clients vs. providers; actual clients vs. simulated clients). The paper also assesses the difference in quality of care received by married and unmarried women. This aspect of the paper is more problematic.

Specific comments:
- Page 3 of 9: The authors say that the aim of the study was to “assess how the quality of service provision could be improved in order to encourage greater uptake and method continuation.” The authors should specify if the aim was to encourage greater uptake of DMPA-SC or greater uptake
of some method of contraception. If the former, I am concerned about depriving clients of their right to choose the contraceptive method they want to use. Please clarify.

- Page 3 of 9: The authors note that DMPA-SC was introduced into the mass market without any pilot feasibility or acceptability studies done. Yet the authors note that the introductory efforts were concentrated in South West Nigeria. A brief from HP+ brief also notes: “Currently, DMPA-SC availability is concentrated at the facility level across pilot states in public and private sectors.” Source: http://www.healthpolicyplus.com/ns/pubs/8197-8351_DMPASCIntroductionandScaleUpinNigeria.pdf

- Please clarify and also explain what then is meant by “a mass market” approach.

- Why did the authors choose to conduct their research in the private sector? And if the study is limited to the private sector, why were government clinics and hospitals included in the convenience sample?

- The methodology section states that the convenience sample of providers was drawn from March to May of 2015. When were the follow up phone calls made to the clients? When were the simulated client visits made? I note from an earlier paper that 2 calls were made to the actual clients. Which of those phone calls was used for this study?

- Since the findings are what clients and simulated clients said about the treatment they received, I suggest that the findings be reported as such – so rather than saying the clients or simulated clients received such and such care, the paper should say that the clients or simulated clients reported receiving such and such care.

- Page 4 of 9: The word “state” should be “stated” in the second paragraph of column 1, line 5.

- Page 5 of 9: The text in column one, paragraph 1, line 8, mentions Table 1 – but the text leading up to the mention is about the eight items included in the technical competency domain that were included in the study. Table 1, shown in the second column on page 5 of 9 is titled “Phone Survey Sample Characteristics”. This seems to be a mistake.

- I am having trouble deciphering the findings that compare groups when the samples are so different. The authors note that the survey respondents predominantly said they had been served by community-based distributors (66%), while only 25% of simulated client visits were to CBDs. The majority of simulated clients visited facilities in urban areas (72%), yet the distribution of the phone sample of women by urban and rural is not given. The authors should say how the different distributions of actual clients and simulated clients may have shaped the results they found. Also, half of the simulated clients were, by study design, age 18, yet in the sample of actual clients only 9% were under age 25, and only 6.1% were unmarried. That gives a very small sample size of unmarried women in the phone follow up. Given that the treatment of married vs. unmarried women is the central focus on the paper, the authors should justify how their sampling is reasonable to make the comparison and also say how might this difference have influenced the findings.

- The authors note in the introduction that there are many dimensions of quality of care and that it is difficult to measure the non-technical dimensions. An earlier paper showed that 14 items were
collected from the actual clients in the phone calls (the 8 questions related to information given, 3 questions related to interpersonal relations, and 3 questions related to choice). Yet, the information reported in this paper is only the 8 items of information given – what were clients told by providers? Why were only the 8 items related to information given used in this paper? I’m wondering how this small slice of measurement of quality can help improve quality of DMPA services – other than to say that providers should make sure to give the same information and ask the same questions to all clients.

I urge the authors to consider revising the title of the paper since the analysis does not really measure “the quality of family planning counseling” – what about “the quality of information given in family planning counseling” – that is a more accurate reflection of what was studied.

Page 6: The authors note that the clients may have been giving courtesy responses of favorable treatment. They should consider that the simulated clients might have been conditioned to be “tough” in their review of the care they received and could thus potentially have given more negative reviews than the actual clients. The description of the training for the simulated clients described on page 6 of 9 suggests that they were conditioned to provide more negative reviews – potentially a Hawthorn effect for them.

Page 7 of 9, first column, first paragraph, line 8: There is an “of” missing between “quality of contraceptive service provision”.

Page 7, Table 3: Were the simulated clients not asked if they had ever experienced side effects? Why are those cells empty?

The authors say that the purpose of this study was to assess how the quality of services could be improved, yet they say nothing about this in the discussion or conclusion of the paper. Given the questions about the comparability of the samples, the very small sample sizes of unmarried women in the phone follow up, and the mismatch between the service channels that actual women used (predominantly CBD) and those the simulated clients visited (an even distribution of channels), it is not clear how these findings could be used to improve services. Also, given that only 8 items related to the technical quality of counseling were included in the study, what do the authors say about the utility of the approach of using simulated clients for program improvement? Would the authors suggest not asking actual clients about the quality of care they received?

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Rights-based family planning; gender and family planning, including male engagement; population and climate change; implementation research; research to policy.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 04 Dec 2019

Nadia Diamond-Smith, University of California San Francisco Medical Center, San Francisco, USA

Thank you for your helpful comments we have addressed each below:

Specific comments:

- Page 3 of 9: The authors say that the aim of the study was to “assess how the quality of service provision could be improved in order to encourage greater uptake and method continuation.” The authors should specify if the aim was to encourage greater uptake of DMPA-SC or greater uptake of some method of contraception. If the former, I am concerned about depriving clients of their right to choose the contraceptive method they want to use. Please clarify.

  Thank you for this comment, this study is part of a larger project assessing the role-out of DMPA-SC, but we explored all method uptake and in no way were encouraging only uptake of one method. However, we have removed this sentence as actually the aim of the paper is more methodological and now the sentence reads: “The overarching aim was to understand how these different methodological approaches (simulated clients and phone surveys) compare for measuring one domain of quality (arguably the most objective, informational), and if both are able to detect inequitable treatment by client characteristics.”

- Page 3 of 9: The authors note that DMPA-SC was introduced into the mass market without any pilot feasibility or acceptability studies done. Yet the authors note that the introductory efforts were concentrated in South West Nigeria. A brief from HP+ brief also notes: “Currently, DMPA-SC availability is concentrated at the facility level across pilot states in public and private sectors.” Source: http://www.healthpolicyplus.com/ns/pubs/8197-8351_DMPASCIntroductionandScaleUpinNiger

  Please clarify and also explain what then is meant by “a mass market” approach.

  We have removed this wording as it was confusing and meant just to indicate that there were multiple distribution channels, as the sentence now reads.

- Why did the authors choose to conduct their research in the private sector? And if the study is limited to the private sector, why were government clinics and hospitals included in the convenience sample?

  We have clarified that we collected data from private providers based in government facilities. In 2016 DMPA-SC was only available in the private sector, which we have clarified. We have also
added “Since this time DMPA-SC distribution has expanded to the government sector.”

- The methodology section states that the convenience sample of providers was drawn from March to May of 2015. When were the follow up phone calls made to the clients? When were the simulated client visits made? I note from an earlier paper that 2 calls were made to the actual clients. Which of those phone calls was used for this study?

The data from the first round of surveys, most within 1 month of the visit was used for this study. Another one of our papers describes the phone survey in more detail https://www.sciencedirect.com/science/article/pii/S0010782418301525, we have added this information and cited the paper.

- Since the findings are what clients and simulated clients said about the treatment they received, I suggest that the findings be reported as such – so rather than saying the clients or simulated clients received such and such care, the paper should say that the clients or simulated clients reported receiving such and such care.

Changed.

- Page 4 of 9: The word “state” should be “stated” in the second paragraph of column 1, line 5.

Apologies, we cannot find this.

- Page 5 of 9: The text in column one, paragraph 1, line 8, mentions Table 1 – but the text leading up to the mention is about the eight items included in the technical competency domain that were included in the study. Table 1, shown in the second column on page 5 of 9 is titled “Phone Survey Sample Characteristics”. This seems to be a mistake.

Changed.

- I am having trouble deciphering the findings that compare groups when the samples are so different. The authors note that the survey respondents predominantly said they had been served by community-based distributors (66%), while only 25% of simulated client visits were to CBDs. The majority of simulated clients visited facilities in urban areas (72%), yet the distribution of the phone sample of women by urban and rural is not given. The authors should say how the different distributions of actual clients and simulated clients may have shaped the results they found. Also, half of the simulated clients were, by study design, age 18, yet in the sample of actual clients only 9% were under age 25, and only 6.1% were unmarried. That gives a very small sample size of unmarried women in the phone follow up. Given that the treatment of married vs. unmarried women is the central focus on the paper, the authors should justify how their sampling is reasonable to make the comparison and also say how might this difference have influenced the findings.

We agree that the small sample of young/unmarried women in the phone survey is unfortunate, since we were particularly interested in the experiences of these women in the study. Despite this, the actual Ns of unmarried simulated clients (n=56) and unmarried phone respondents (n=33) is not that different, albeit small. We added the following The providers from which we recruited phone survey respondents were a convenience sample; a design which randomly selected providers would have reduced this bias. Only a 54% of women responded to the request to participate, resulting in selection bias. Perhaps women who had disproportionately bad or good experiences or certain types of women (older, married, etc) were more or less likely to respond. A relatively small proportion of our respondents were young and unmarried, a prime focus of our
analysis, and we do not know if this is reflective of the actual population seeking contraceptives or selection bias. If younger, unmarried women who were willing to participate were more likely than older married women to participate if they had very good or very bad experiences this could be problematic for our findings. The small sample of unmarried women thereby weakens our analysis and the implications of our findings. Higher incentives, additional attempts at recruitment or different recruitment approaches could have reduced selection bias.

We also didn't ask rural/urban status over the phone because this is a more complicated concept for people to report and may not align with our interpretation. We also wanted to ensure privacy so didn't ask more location details. This information would have been useful.

The authors note in the introduction that there are many dimensions of quality of care and that it is difficult to measure the non-technical dimensions. An earlier paper showed that 14 items were collected from the actual clients in the phone calls (the 8 questions related to information given, 3 questions related to interpersonal relations, and 3 questions related to choice). Yet, the information reported in this paper is only the 8 items of information given – what were clients told by providers? Why were only the 8 items related to information given used in this paper? I'm wondering how this small slice of measurement of quality can help improve quality of DMPA services – other than to say that providers should make sure to give the same information and ask the same questions to all clients.

Thanks for reading the other paper too! We felt that the other domains were more challenging to measure and interpret from mystery clients (choice and interpersonal relations) and thus we focused on these more objective, though perhaps not completely objective, measures related to information given. You are correct in pointing out that really we are just comparing differences in information given. I think by changing the title, as you suggest, and the terminology throughout the paper, we have now made it clear that this is in fact our focus and not try to claim a broader type of quality that we are comparing here. We have also added a comment in the limitations section on this topic. Finally, we have clarified that this is really a methods paper, comparing two methodological approaches for measuring 1 domain of quality, and seeing how they both preform in looking at differences across sub groups. We are not trying to make a comment on how to improve the quality of care in this paper.

I urge the authors to consider revising the title of the paper since the analysis does not really measure “the quality of family planning counseling” – what about “the quality of information given in family planning counseling” – that is a more accurate reflection of what was studied.

Page 6: The authors note that the clients may have been giving courtesy responses of favorable treatment. They should consider that the simulated clients might have been conditioned to be “tough” in their review of the care they received and could thus potentially have given more negative reviews than the actual clients. The description of the training for the simulated clients described on page 6 of 9 suggests that they were conditioned to provide more negative reviews – potentially a Hawthorn effect for them.

We have added 'It is likely that phone survey clients might be more likely to recall especially positive or negative interactions, potentially skewing the results in either extreme. As with most data collection, courtesy bias is a concern from the phone survey respondents and on the flip side, simulated clients may have been primed in their training to be overly negative.'

Page 7 of 9, first column, first paragraph, line 8: There is an “of” missing between “quality of contraceptive service provision”. Changed
<table>
<thead>
<tr>
<th>Page 7, Table 3: Were the simulated clients not asked if they had been asked if they had ever experienced side effects? Why are those cells empty?</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Unfortunately in an effort to keep the debrief short and reduce recall bias we did not ask this.</em></td>
</tr>
<tr>
<td>The authors say that the purpose of this study was to assess how the quality of services could be improved, yet they say nothing about this in the discussion or conclusion of the paper. Given the questions about the comparability of the samples, the very small sample sizes of unmarried women in the phone follow up, and the mismatch between the service channels that actual women used (predominantly CBD) and those the simulated clients visited (an even distribution of channels), it is not clear how these findings could be used to improve services. Also, given that only 8 items related to the technical quality of counseling were included in the study, what do the authors say about the utility of the approach of using simulated clients for program improvement? Would the authors suggest not asking actual clients about the quality of care they received?</td>
</tr>
<tr>
<td><em>Thank you for this great point—in fact our aim was really a methodological one, to compare two data collection approaches for collecting information on 1 aspect of contraceptive counseling and to how these methods compare when exploring differential care by marital status. We are not really saying anything about the quality of care and how it can/should be improved in this methods paper.</em></td>
</tr>
</tbody>
</table>

**Competing Interests:** none