When properly implemented, use of simulated clients ("mystery clients") can provide insight into actual experiences of real clients and evaluate quality of care. Successful implementation calls for recruiting mystery clients who represent the facility’s clientele, have strong recall of recent events, and are comfortable being undercover data collectors. Developing training protocols and checklists to standardize mystery client behavior and responses is also key.

ABSTRACT
The use of simulated clients or “mystery clients” is a data collection approach in which a study team member presents at a health care facility or outlet pretending to be a real customer, patient, or client. Following the visit, the shopper records her observations. The use of mystery clients can overcome challenges of obtaining accurate measures of health care quality and improve the validity of quality assessments, particularly in low- and middle-income countries. However, mystery client studies should be carefully designed and monitored to avoid problems inherent to this data collection approach. In this article, we discuss our experiences with the mystery client methodology in studies conducted in public- and private-sector health facilities in Kenya and in private-sector facilities in Uganda. We identify both the benefits and the challenges in using this methodology to guide other researchers interested in using this technique. Recruitment of appropriate mystery clients who accurately represent the facility’s clientele, have strong recall of recent events, and are comfortable in their role as undercover data collectors are key to successful implementation of this methodology. Additionally, developing detailed training protocols can help ensure mystery clients behave identically and mimic real patrons accurately while short checklists can help ensure mystery client responses are standardized. Strict confidentiality and protocols to avoid unnecessary exams or procedures should also be stressed during training and monitored carefully throughout the study. Despite these challenges, researchers should consider mystery client designs to measure actual provider behavior and to supplement self-reported provider behavior. Data from mystery client studies can provide critical insight into the quality of service provision unavailable from other data collection methods. The unique information available from the mystery client approach far outweighs the cost.

INTRODUCTION
Accurate data collection is essential to evaluating potential problems in service delivery, particularly in low- and middle-income countries. Several methodological tools can be used to gather such information, such as facility-level surveys including provider interviews or the client exit interview. Another tool is the simulated client design. In this approach, a study team member or assistant pretends to be a real customer (mystery shopper) or client/patient (standardized patient) who seeks services or care according to standardized prearranged scripts. During the health care visit, the provider is unaware that the encounter is for research purposes. Following the session, the undercover data collector—henceforth referred to as the mystery client—then reports her observations to the study team.

The mystery client methodology has been used successfully in several studies to assess the quality of health care delivery and identify areas for quality improvement.1–3 Mystery clients are considered to be a highly reliable and valid method for assessing provider communication skills and behavior.4,5 Mystery clients have been used extensively in medical education to train and
assess the competency of doctors.6,7 This methodology has been used in evaluating provider behavior more generally since 1985,5–10 but is currently used less frequently than other methods such as surveys.

Mystery clients allow insight into actual transactions and experiences of real customers or patients. By relying on data recalled by mystery clients, this method overcomes some key concerns with traditional structured interviews or direct observation methods.11 For example, in direct observation, provider behavior is confounded by patient behavior. In contrast, mystery clients are trained to display standardized behavior, isolating how providers respond to observationally similar patrons. Similarly, directly observing providers may result in them changing their behavior toward what they believe will impress or please their observers, in a Hawthorne effect. In structured questionnaires of real patrons, there may be a desire for the respondent to be polite or positive toward the provider (courtesy bias) or they may simply make mistakes in recalling information (recall bias).12–14

In addition, patients and health care providers may not be eager to participate in surveys—patients may not feel well and providers may be in a rush to deliver medicines to the patient—resulting in response bias from refusal rates as high as 38%.15 Finally, surveys of providers may instead reflect their best intentions or how they have been trained to deliver services rather than what they actually do in practice, a phenomenon called the “know-do gap.”16 However, mystery client studies may present special challenges as compared to surveys or interview-based designs—particularly in low- and middle-income countries.

In this article, we discuss our experiences from studies conducted in Kenya and Uganda using mystery client designs and identify lessons learned to inform successful implementation of this approach to assess quality of care. In Kenya, within a sample of 19 urban public and private health facilities, mystery clients were employed to assess the validity of standard data collection tools used to measure the quality of family planning service delivery in large-scale demographic surveys17; this study revealed harmful provider practices not detected by the standard tools.18 In Uganda, mystery clients purchased anti-malarial medicines at nearly 500 informal private-sector outlets to measure counterfeit drug rates; the study included both urban and rural areas.15 We augment our experiences with findings and approaches from recent research from other mystery client studies.

APPROACH

Prior to the Visit: Ethical Approval

In general, the ethical issues involved will be study-specific. Using mystery clients may be classified as “deceptive research” depending on the design details, for example, if providers are not informed about the study and do not provide informed consent in advance. Therefore, institutional review boards (IRBs) should be consulted and approve all protocols and activities to ensure ethical concerns are addressed. Study team members must be aware that they are acting as observers and must keep all experiences and data strictly confidential both during and after study completion. Shoppers may sign non-disclosure agreements as a condition of employment to enforce the importance of confidentiality. There may also be special challenges in observing public health facilities due to long lines and possible political sensitivity around the evaluation of government facilities. For example, it may also be necessary to obtain ethical approval from the provincial- and district-level ministries of health. Consulting with these agencies may also help relevant stakeholders feel invested in and appreciate the results of the study. Such buy-in may also help prevent concerns of trespassing on government property.

To avoid discontent among respondents (i.e., health care providers), IRBs may require principal investigators (PIs) to disclose the study design in advance and gain consent of providers or facility supervisors without specifying the date of the visit ahead of time. Alternatively, IRBs may require debriefing providers after the transactions are completed. IRBs may also require that, if debriefed, providers maintain the right to have their data withdrawn because the information was collected by deception. If a sufficient number of providers withdraw as a result of forewarning or debriefing, then the generalizability of results may be in question—particularly if either low- or high-quality providers differentially withdraw.

In Uganda, the IRB agreed to waive both forewarning and debriefing requirements. This waiver was approved because (1) the societal benefits exceeded potential harms to providers due to a dearth of evidence on drug quality, and (2) the deception itself was minor and protocol sufficiently mimicked real-life interactions. Thus, potential harm to subjects was minimal. Signing informed consent before or after the study may also have provided identifying information, risking confidentiality if data were compromised. By contrast, in Kenya, the IRB required informed...
Hiring mystery clients with strong recall ability, which can be assessed with memory tests, is important to ensuring accurate data collection.

As a result, the average educational levels of mystery clients exceeded those of the average drug shop clientele, potentially affecting the interpretation of results. In the Kenya study, which took place in central Kisumu, mystery clients had to be fluent in the local language of Dholuo to avoid suspicion on the part of local health care providers; as such, data collectors from Nairobi willing to reside for the duration of the study period in Kisumu were ineligible because they spoke Swahili, not Dholuo. PIs may additionally find it helpful to train and hire several more study team members than needed in the event of attrition or illness; this was necessary in the Uganda study but not in Kenya where data collection activities were completed within 2 weeks.

Similarly, accurate data collection also implies that providers are unaware the shopper is also a research team member. For example, if mystery clients are intended to represent a particular income or regional group, their language, dress, and even hairstyle must be standardized to match the target demographic’s typical appearance. It also will be important to remove accessories that are correlated with wealth while conducting visits (e.g., wedding rings, watches, and cell phones, among others). This may be challenging; for example, standards of dress may differ between urban and rural areas. In situations where language or dialect varies substantially across areas, researchers may consider using multiple study teams or adjusting protocol to reflect differing norms. In the Uganda study, 2 different teams with different language abilities were used. If using multiple teams is not possible, then the PIs should alternatively consider restricting to a more homogeneous geographic area. Finally, depending on the script content, shoppers may also need to appear outwardly healthy to ensure that the providers do not get confused about what ailment to treat.

These safeguards appear to have worked in our studies in Kenya and Uganda. In Uganda, during a separate survey conducted at the same outlets that mystery clients visited, providers reported that, on average, they suspected the mystery client only 3% of the time. Although there may be concerns that shoppers may stand out more in rural areas, this rate did not statistically differ between urban and rural areas. In Kenya, in only 1 visit out of 134 did a mystery client report that the provider questioned her authenticity.

Adequate time for training study team members—including supervisors—is also important to ensure all shoppers follow protocols...
exactly and consistently. In our studies, mystery clients received extensive training over 3 to 5 days on the study protocol, including the survey instrument, research objectives, outlet or facility locations, transportation logistics, dress code, confidentiality, and other study policies. Training should also emphasize that shoppers should not prompt providers in an effort to help them perform “better.” Training periods may vary depending on the complexities of the intervention. While our studies used 24–40 training hours, one study reported 250 hours of training.20

Generally, mystery clients should feel comfortable in their role as an “undercover” data collector; toward this end, extensive time should be set aside during training for role play and practice using the data collection instrument. Role play activities conducted in front of all mystery clients can also help to ensure shopper behavior is standardized. If possible, create time to pilot test the data collection instrument and allow the mystery clients the opportunity to test out their new role in a location away from the study sites. In Uganda, pilot shopping rounds were both supervised and conducted independently, giving confidence to both shoppers and supervisors that shoppers were prepared for full-study data collection. During this time, the trainer can also make sure that shoppers are interpreting and answering questions in the same way to ensure standardization and inter-rater reliability. In both Kenya and Uganda, a short checklist was used to help the mystery clients recall aspects of their visit and provide a standardized evaluation.21 Questions should be as objective as possible to prevent speculation.

Finally, training must emphasize protocols that protect mystery clients from harm. For example, providers can sometimes recommend intramuscular injections even when not clinically indicated. Similarly, project leaders must ensure that less obvious risks, such as taking a temperature with an unsterilized thermometer, also do not occur. As part of their training, mystery clients need to be prepared with various culturally appropriate and plausible strategies and answers to avoid any procedure that could put them at risk. In addition to physical harm, emotional harm and stress can also be minimized by preparing answers for both expected and unexpected questions. In Kenya, all mystery clients were trained to use culturally appropriate excuses to avoid unwanted services such as injections, implantable contraception, or the intrauterine device; excuses included telling the provider they changed their mind or that they first needed to ask their husband, think about it, acquire sufficient funds, or compare with another facility. In Uganda, shoppers were prepared to answer questions ranging from details on the patient’s illness to the shopper’s personal background. Standardizing shopper answers also ensures that providers view each shopper identically, regardless of what occurs during the transaction. Conducting focus groups with real customers or patients, interviews with local experts, or receiving support from an anthropologist may help PIs anticipate expected questions and ensure realistic responses in line with cultural norms. In the Ugandan study, the exact language, translations, and excuses used were discussed and validated by the mystery clients during training. In extreme circumstances where risks cannot be sufficiently minimized or providers may not accept refusals to treat, the research design should be modified or a different methodology used. For example, in any study ascertaining the quality of service delivery around actual insertion of an intrauterine device or the implantation or injection of a contraceptive method, the mystery client methodology would not be appropriate.

Planning the Visit: Sampling Frame

Building a sampling frame can also be particularly challenging in low- and middle-income countries. Many health care outlets lack outward signage; official records may be incomplete; administrative boundaries uncertain. Developing a standardized protocol for building the sample frame of eligible outlets and creating the study plan is important to reduce bias and minimize unexpected events during fieldwork. If data are to be collected more than once from the same location, as is common in audit study designs, study team members may create maps of all study areas, along with a physical description of the outlets or buildings, to help them find the same outlets again. We followed this approach in Uganda. If ethical approvals allow, GPS coordinates can facilitate this process. In Uganda, we validated mapped location information by consulting with local key informants, such as village chiefs or motorcycle taxi drivers, to cross-reference geographic details. The location of health care facilities selected for the Kenya study were widely known to both the local population and the data collectors, who also resided in the area.

Planning the Visit: Pre-Study Visits

Even if a reliable sample frame already exists, visits to study sites prior to actual data collection are
Clear protocols and rules need to be established if mystery clients are to conduct financial transactions.

During the Visit: Standardized Shopper Behavior

During the visit, all shopper behavior—from words spoken to shopper demeanor—should be completely standardized to be able to make comparisons across providers or visits. Ensuring data are collected according to guidelines over time requires ongoing effort and close monitoring. In both Kenya and Uganda, we reviewed protocol and common responses to questions during daily team meetings to ensure consistency over time and allow for a discussion of how shoppers were feeling. In our studies, supervisors conducted audits and additional checks to ensure that the correct facilities were visited, while being careful not to compromise the mystery client’s covert status.

During the Visit: Collecting Data on Prices

Prices for services or products in developing countries are often not posted and receipts are uncommon. If the research design requires shoppers to make financial payments, then developing strategies to ensure shoppers honestly and correctly report prices is of utmost importance. One option is to trust study team members to honestly report the prices they paid and allow them to bargain for the best price in an effort to balance minimizing study costs with gaining insight to true client experiences. While common in studies of bargaining, this type of protocol could be expensive and may still lead to incorrect data if shoppers switch to cheaper treatments (without notifying the study team). Our approach in both Kenya and Uganda was to instead have multiple shoppers visit the same outlet or facility; this created the impression that the study team could cross-reference prices between shoppers, even though in Uganda, measuring price differences between shoppers at the same provider was the outcome of interest. Researchers considering this approach should ideally have visits conducted by different shoppers several hours (or even days) apart. Multiple visits to the same provider, facility, or outlet are typically feasible but may be difficult if the provider is frequently absent from work or if the facility or outlet has limited operating hours. On occasion, official hours of operation may differ from actual hours.

Similarly, clear protocols and rules need to be established if shoppers are to conduct financial transactions. In Uganda, we set a budget of 10,000 UGX (Ugandan shillings) per transaction. The bills used were in small denominations of used-looking money to minimize the likelihood that providers lacked adequate change to complete a transaction and ensure that the shopper remained inconspicuous. The money was placed in an envelope with the shop and transaction ID labeled on the outside. After the transaction, the shoppers returned the balance to the envelope and accounts reconciled with the price paid. If shoppers required additional funds for a purchase, they were required to return to supervisors. Shoppers were not allowed to use their own money under any circumstances. Separate funds and budgets were used for transportation expenditures. Similar procedures were observed in Kenya.

In addition, collecting data from real customers, or drug pricing data from vendors, may provide better data and allow for cross-referencing of price data. Additional data collection can also validate assumptions of the experimental design. In the Ugandan context, collecting additional data from real customers validated the assertion that mystery clients behaved similarly to real customers. Another mystery client study examined whether providers altered drug prices or quality if they perceived customers to be of different income levels. To validate provider impressions of shopper income levels, the researchers took daily photographs of shopper outfits that were then evaluated by real customers to test whether mystery clients’ appearance reasonably represented socioeconomic status.22

Changing Study Protocols and Rules

Developing good communication is important to ensuring study team members perform as
Mystery Client Implementation Strategies

CONCLUSION

In summary, the use of mystery clients offers a unique and critical opportunity to accurately measure the quality of service provision. When selecting mystery clients, it is important to choose people who will present realistically and have excellent recall. In addition, careful attention should be paid to strategies that allow mystery clients to feel comfortable in their role and to prevent unwanted exams or procedures. Researchers should appreciate details of data collection and listen to concerns of their team members. Despite these challenges, the unique information available from this type of methodology far outweighs the cost.

Acknowledgments: Partial support for the development of this article was provided by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health: P2C HD047879 and T32 HD007163. With respect to references to research conducted in Kenya, support was provided by the Bill & Melinda Gates Foundation, the Carolina Population Center, and the American Psychological Foundation. Research conducted in Uganda was supported by the National Science Foundation and various organizations within the University of Michigan (Department of Economics, Department of Afro-American and African Studies, the ASC, Rackham Graduate School, Center for the Education of Women, and Center for International Business Education Research).

Competing Interests: None declared.

REFERENCES

1. Hardee K, McDonald OP, McFaran L. Quality of care in family planning clinics in Jamaica. Do clients and providers agree? West Indian Med J. 2001;50(4):322–327. CrossRef. Medline

2. Huntington D, Luttenmaier C, Obeng-QuadiDO. User’s perspective of counseling training in Ghana: the “mystery client” trial. Stud Fam Plann. 1990;21(3):171–177. CrossRef. Medline

3. Population Council. Quantitative assessment of the quality of care: effects of provider training. In: Population Council INOPAL updates. New York: Population Council; 1992.

4. Rethans JJ, Gorter S, Bakken L, Morrison L. Unannounced standardized patients in real practice: a systematic literature review. Med Educ. 2007;41(6):537–549. CrossRef. Medline

5. Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. JAMA. 2007;298(9):993–1001. CrossRef. Medline

6. Rethans JJ, Sturmans F, Dreg R, van der Vleuten C. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. Br J Gen Pract. 1991;41(344):97–99. CrossRef. Medline

7. Cox M, Irby DM, Epstein RM. Assessment in medical education. N Engl J Med. 2007;356(4):387–396. CrossRef. Medline

8. Schuler SR, McIntosh EN, Goldstein MC, Pande BR. Barriers to effective family planning in Nepal. Stud Fam Plann. 1985;16(5):260–270. CrossRef. Medline

9. Woodward CA, McConvey GA, Neufeld V, Walsh A. Measurement of physician performance by standardized patients. Refining techniques for undetected entry in physicians offices. Med Care. 1985;23(8):1019–1027. CrossRef. Medline

10. Huntington D, Schuler SR. The simulated client method: evaluating client-provider interactions in family planning clinics. Stud Fam Plann. 1993;24(3):187–193. CrossRef. Medline

11. Simmonds R, Elias C. The study of client-provider interactions: a review of methodological issues. Stud Fam Plann. 1994;25(1):1–17. CrossRef. Medline

12. Bessinger RE, Bertrand JT. Monitoring quality of care in family planning programs: a comparison of observations and client exit interviews. Int Fam Plan Perspect. 2001;27(2):63–70. CrossRef

13. MEASURE Evaluation. Quick investigation of quality (IQiQ): a user’s guide for monitoring quality of care in family planning. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill; 2001. https://www.measureevaluation.org/resources/publications/ms-01-02. Accessed January 3, 2017.

14. Whitmecar R, Mila R, Hussain B, Keong M. Evaluating rural Bangladeshi women’s perspectives of quality in family planning services. Health Care Women Int. 199917(5):393–411. CrossRef. Medline

15. Fitzpatrick A. Customer information, price, and quality: evidence from a randomized audit study. Paper presented at: Midwest International Economic Development Conference (MIEDC) 2011;54(1):347–357. CrossRef. Medline

16. Mohanam M, Vera-Hernández M, Das V, et al. The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India. JAMA Pediatr. 2015;169(4):349–357. CrossRef. Medline

17. Tumlinson K, Speizer IS, Curtis SI, Pence BW. Accuracy of standard measures of family planning service quality: findings from the simulated client method. Stud Fam Plann. 2014;45(4):443–470. CrossRef. Medline

18. Tumlinson K, Speizer I, Archer L, Behets F. Simulated clients reveal factors that may limit contraceptive use in Kisumu, Kenya. Glob Health Sci Pract. 2013;1(3):407–416. CrossRef. Medline

19. Lu F. Insurance coverage and agency problems in doctor prescrip- tions: evidence from a field experiment in China. J Dev Econ. 2014;106:156–167. CrossRef

20. Das J, Holla A, Das V, Mohanan M, Tabak D, Chan B. In urban and rural India, a standardized patient study showed low levels of
provider training and huge quality gaps. Health Aff. 2012;31(12):2774–2784. CrossRef. Medline

21. León FR, Quiraz G, Brazzoduro A. The reliability of simulated clients’ quality-of-care ratings. Stud Fam Plann. 1994;25(3):184–190. CrossRef. Medline

22. Bennett D, Yin W. The market for high-quality medicine. Paper presented at: The Bureau for Research and Economic Analysis of Development (BREAD) Conference 2013; April 26-27, 2013; Providence, RI. http://ibread.org/bread/sites/default/files/0413conf/bennett.pdf. Accessed January 3, 2017.

Peer Reviewed

Received: 2016 Aug 16; Accepted: 2016 Nov 22; First Published Online: 2017 Jan 26

Cite this article as: Fitzpatrick A, Tumlinson K. Strategies for optimal implementation of simulated clients for measuring quality of care in low- and middle-income countries. Glob Health Sci Pract. 2017;5(1):108-114. https://doi.org/10.9745/GHSP-D-16-00266

© Fitzpatrick and Tumlinson. 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 author and source are properly cited. To view a copy of the license, visit http://creativecommons.org/licenses/by/3.0/. When linking to this article, please use the following permanent link: https://doi.org/10.9745/GHSP-D-16-00266