

Monitoring Quality of Care in Family Planning Programs: A Comparison of Observations And Client Exit Interviews

By Ruth E. Bessinger and Jane T. Bertrand

Context: *Monitoring quality of care in family planning programs is important, but the complexity and subjectivity of the topic create many challenges.*

Methods: *The Quick Investigation of Quality (QIQ) was developed to monitor quality of care by using observations of client-provider interactions, exit interviews with clients and facility audits. The QIQ was field-tested in multiple countries in 1998–1999. Using linked data for 583 clients in Ecuador, 539 in Uganda and 736 in Zimbabwe, this analysis examines the comparability of results from observations and exit interviews.*

Results: *For a given indicator, levels of agreement between data from observations and interviews varied across countries, but within a country, results were consistent between instruments. For the three countries combined, agreement was good to excellent (kappas, 0.47–0.98) on 13 of the 14 indicators examined; observations and exit interviews yielded consistent responses in 63–99% of cases. Agreement was highest on the indicators that measured interpersonal relations. Inconsistencies reflected primarily that clients received information outside of the observed client-provider interaction.*

Conclusions: *Observations and client exit interviews provide very similar results for many indicators. However, while some programs may opt for one instrument over the other because of resource constraints, the combination provides a fuller assessment of quality of care.*

International Family Planning Perspectives, 2001, 27(2):63–70

During the 1990s, quality of care became a central focus for the international family planning and reproductive health community. Over the past decade, work in this area has been guided by the Bruce-Jain framework, which outlines six elements of quality: choice of method, information to the client, technical competence, interpersonal relations, mechanisms to encourage continuity and constellation of services.¹ Organizations have adopted variations on this theme, such as the International Planned Parenthood Federation's (IPPF's) Client Bill of Rights, later amended to the Client and Provider Bill of Rights.

*The QIQ was developed in collaboration with the Monitoring and Evaluation Subcommittee of the Maximizing Access and Quality (MAQ) initiative, funded by the U.S. Agency for International Development (USAID).

†The list of indicators was compiled from several sources, including a checklist developed by MAQ, the IPPF Client Bill of Rights, IPPF's list of indicators on "Quality of Care from the Gender Perspective" and the EVALUATION Project's *Handbook of Indicators for Family Planning*.

‡While other instruments were considered during the development of the QIQ, only these three were selected, to make the data collection as simple and low-cost as possible. For example, while provider surveys have yielded valuable information in the context of situation analysis (source: reference 5), they were not essential in the current context; moreover, researchers opted for measuring what providers actually do in counseling and clinical procedures, rather than what they say that they do.

With this increased focus on quality, a parallel interest has arisen in developing means of measuring quality, for several reasons. First, client-provider interactions can be understood as intervening elements in a causal chain through which organized family planning efforts meet or generate demand for fertility regulation.² Learning more about these processes with the aim of improving them can have important programmatic payoffs. Second, many programs have undertaken activities to improve quality of care in their facilities. Without measurement tools, it is impossible to know whether these activities have achieved their objectives. Third, by investing resources in measuring quality, management sends a message to staff that "quality is important." As such, measurement reinforces initiatives to improve quality.

The challenge in measuring quality is the complexity and subjectivity of the topic. Although the Bruce-Jain framework outlines six elements of quality, literally hundreds of possible "subelements" might be measured. A task force created to explore the measurement of quality in 1990 identified more than 200 indicators of quality in family planning services.³ This group recommended experimentation with these different indicators, using

the approach of "let 100 flowers bloom." Subsequently, the EVALUATION Project convened a working group of researchers to study quality of care, and this group reduced the list to 42 process indicators.⁴

The successor project, MEASURE Evaluation, developed and field-tested a low-cost, practical approach to monitoring quality of care, later named the Quick Investigation of Quality (QIQ).⁵ The project used an approach based on the Delphi method, in which experts share their knowledge through questionnaires and feedback: It asked specialists in quality of care, family planning service delivery or program evaluation to select from a list of some 80 indicators[†] the 25 that in their opinion most directly affected quality outcomes in terms of clients' behavior. To collect data on these 25 indicators, three instruments were developed: a facility audit with selected questions to the program manager, observation of client-provider interactions and selected clinical procedures, and exit interviews with clients departing from the facility (and previously observed).[‡] These instruments were field-tested in four countries (Ecuador, Turkey, Uganda and Zimbabwe) between October 1998 and March 1999 to determine the feasibility of data collection and the reliability of the data. A detailed description of the development and implementation of the QIQ methodology and full results from the field test are reported elsewhere.⁶

With the conduct of facility-based surveys on the rise, the field-test data provide an excellent opportunity to address a question of critical importance: Do observations of client-provider interactions and exit interviews with clients yield consistent results? This question is particularly pertinent in the context of QIQ for two reasons. First, consistency in results

Ruth E. Bessinger is evaluation analyst, ORC Macro International, Calverton, MD, USA; Jane T. Bertrand is professor and chair, Department of International Health, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. Funding for this study was provided by the U.S. Agency for International Development (USAID) through the MEASURE Evaluation project at the University of North Carolina at Chapel Hill, cooperative agreement HRN-A-00-97-00018-00. The views expressed in this article are those of the authors and do not necessarily reflect the views of USAID.

between the two instruments would lend credibility to the QIQ package of instruments. Second, if it is sufficient to administer only one of the instruments, rather than both, the cost implications for future facility-based surveys could be important.

It should be stressed that the two instruments differ in terms of the type of information they are best suited to capture. Observation is useful in measuring the accuracy and thoroughness of information imparted during counseling and in assessing providers' technical competence (which clients are generally not able to do). By contrast, the exit interview is the only instrument used in QIQ that taps clients' perspectives on the services they have received. However, information from the two can be compared when clients report on providers' actions during counseling and clinical examinations.

Both methods have limitations. The reliability of data from observation can be an issue because observers may interpret the same set of provider actions differently.⁷ Observation also introduces the potential bias that service providers will perform better than they might under usual conditions. Indeed, the Kenya Situation Analysis yielded evidence that providers' performance increased during the first three days in a week of observation, then declined, suggesting that it was not possible to remain on "best behavior" indefinitely.⁸

Observation of client-provider interactions is also limited in that it includes only a part of the client's visit and does not cover, for example, group counseling sessions, where clients may receive important information. Moreover, this type of data collection requires more skilled personnel than a standard interview, since the observer must have adequate clinical background to judge whether procedures are performed correctly. And she must be quick enough to record a series of events that often do not occur in the same order as they are listed on the data collection form.

Exit interviews have their own set of problems, the most serious of which is courtesy bias. Respondents may give what they consider socially acceptable answers, especially if they believe that the interviewer works for the clinic or that unfa-

vorable comments could negatively affect the services they will receive in the future.⁹ In addition, clients may have such low expectations of services that even when the quality of services is poor, it exceeds their expectations and they report positively on their experience. A further concern is recall bias, which occurs when a respondent cannot accurately recount what happened during the session.

Methodology

Assessment Instruments

In this article, we assess two of the three QIQ instruments: observation of client-provider interactions and exit interviews with clients. For the first instrument, a trained observer (who usually wore a white coat to blend into the service delivery environment) obtained consent from the provider and the client to be present during individual counseling and clinical examination. She used an observation guide with a structured checklist of items related to quality of care (e.g., whether the provider asked particular questions, whether specific points of information were covered and whether certain clinical procedures were used in the provision of particular contraceptives).

As the client left the facility after her visit, an interviewer approached her to ask if she could talk with the woman about the visit and her satisfaction with the services she received. The interviewer explained that she did not work for the clinic; that all responses would remain confidential; and that the woman's answers would in no way affect the services she received in the future. Assuming that the client gave her consent, the interviewer proceeded to ask her a series of questions (which usually took about 20 minutes).

Three of the countries that participated in the QIQ field test (Ecuador, Uganda and Zimbabwe) implemented both client-provider observations and client exit interviews. The instruments used in the three countries were almost identical, although questionnaires for the exit interviews were translated into local languages. Clinically trained staff, nurses and midwives conducted the client-provider observations in Uganda and Zimbabwe; physicians conducted observations in Ecuador. In all three countries, social workers and sociologists conducted the exit interviews. Data collection staff in each country, all of whom were female, underwent a one-week training session on the instruments and methodology; training also included a pilot test of the instruments.

The types of facilities included in the study differed across the three countries, but these differences should not affect the research question. In Ecuador, the sample consisted of all 43 family planning facilities run by two nongovernmental organizations: Asociación Pro-Bienestar de la Familia Ecuatoriana (APROFE) and Centro Médico de Orientación y Planificación Familiar (CEMOPLAF). The Uganda study used a probability sample of 72 public facilities located in 10 districts receiving support from the Delivery of Improved Services for Health (DISH) project and in three districts not receiving support from that initiative. In Zimbabwe, all 39 facilities receiving support from the Family Planning Service Expansion and Technical Support Project (SEATS) were surveyed.*

Analysis

Unique identifying information recorded on the observation and exit interview forms for each client was used to link the data from the two instruments for each woman; linked data from the three countries were combined for the purposes of this analysis. In all three countries, 98–99% of clients who were observed and interviewed had data that could be linked from both instruments. A total of 1,858 family planning clients are included in this analysis—583 from Ecuador, 539 from Uganda and 736 from Zimbabwe.

Eleven of the 25 quality of care indicators were measured by both observation and exit interview (Figure 1). Of these, two (provider demonstrates good counseling skills and client is empowered) were composite indicators, measured by aggregating responses to several questions; these are not included in the analysis because comparable questions from both instruments for all components of the indicator were not available for all three countries. Five additional indicators of quality for which questions appeared on both instruments have been included, giving us a total of 14 indicators of quality of care for analysis.

Some questions (e.g., whether the provider gave instructions on when to return) were virtually the same on the observation form and in the exit interview and were strictly objective. Others (e.g., whether the provider treated the client with respect), while the same on both forms, required a subjective judgment, and answers could well differ between the observer and the client. In yet other cases, questions addressed a similar topic, but were not the same. For example, one item

*DISH, a project funded by USAID through a bilateral agreement with the Uganda Ministry of Health, was designed to change reproductive-related behaviors by increasing availability and improving the quality of integrated health services. The aim of SEATS, which is carried out by John Snow Inc. with funding from USAID, is to expand the development of, access to and use of quality family planning services in underserved populations.

on the observation form asked whether the provider “gave accurate information on how to use the method.” The parallel item in the exit interview asked the client to provide correct information on how to use the method. However, if a client cannot provide the correct information, this does not necessarily mean that the provider did not supply it during the visit.

The first step of the analysis consisted of comparing frequencies on indicators that were available from both instruments in all three countries. To simplify presentation of the data, we have organized the indicators into five of the six elements of the Bruce-Jain framework. (No indicators available from both instruments captured technical competence.)

We calculated simple agreement on each indicator as the proportion of responses in which the observation and exit interview results were in agreement. Kappa coefficients were calculated to correct for the proportion of responses that would be in agreement as a result of chance alone. Since kappa values become low when the prevalence deviates from 50%, and many of the indicators were highly skewed toward positive responses, we report kappa coefficients adjusted for prevalence and bias.¹⁰ Kappa coefficients ranging from 0.00 to 0.39 indicate poor agreement, the 0.40–0.74 range indicates fair to good agreement, and values of 0.75–1.00 indicate excellent agreement.¹¹

Finally, we combined data from all three countries and present percentage agreement and kappa coefficients. We also assessed evidence of bias or systematic error using McNemar’s test for bias. Bias was considered to be present if one instrument consistently rated the indicator higher (or lower) than did the other instrument.

Results

Client Characteristics

Clients’ characteristics, which may influence their ability to accurately report information from the visit, varied somewhat among the three countries. Overall, almost one-half of the women were aged 24–35, and age patterns were similar across countries. Educational levels, however, varied; they were highest in Ecuador, where 67% of clients had attended at least secondary school, and lowest in Uganda, where only 40% of clients had advanced beyond primary school.

Reasons for coming to the clinic were similar. Slightly more than one-quarter of clients in each country were new family planning clients (defined as those who were either coming to the facility for a family planning method for the first time,

restarting a method after not using it for more than six months, switching methods or making their first visit to the facility). However, the contraceptive methods they received differed substantially. In Ecuador, the IUD predominated, with 43% of clients receiving a device during their visit; other frequently prescribed methods were the injectable (21%) and the pill (17%). In Uganda, 71% of clients received the injectable and 22% the pill. In Zimbabwe, most clients received the pill (62%) or the injectable (35%).

Interpersonal Relations

Virtually all clients in every country were treated with respect (Table 1, page 66), and results on this indicator were highly consistent between observations and client exit interviews (kappas, 0.98–0.99). Results regarding whether counseling and the pelvic examination took place in private were also similar on both instruments; consistency across instruments was good to excellent for Ecuador and Zimbabwe (kappas, 0.74–0.94), and was lower but still good in Uganda (0.63–0.65). Where disagreement occurred, clients were typically less likely than observers to report that privacy was adequate. In Ecuador, for example, 99% of observers recorded that the pelvic examination was conducted in privacy, compared with 93% of clients.

Providers were supposed to ask returning clients whether they had any concerns or problems. In Ecuador and Uganda (Zimbabwe had no data on this indicator), consistency between responses from observations and exit interviews

Figure 1. Quality of care indicators measured by the Quick Investigation of Quality, by type of indicator, according to measurement instrument, 1998–1999

Indicator	Client exit interviews	Observations	Facility audit
Provider			
Demonstrates good counseling skills (composite)	3	3	
Assures client of confidentiality		3	
Asks client about reproductive intentions	3	3	
Discusses with client which method she would prefer	3	3	
Mentions STDs or HIV/AIDS (initiates or responds)	3	3	
Discusses dual method use	3	3	
Treats client with respect/courtesy	3	3	
Tailors key information to the client’s particular needs	3		
Gives accurate information on the method accepted	3	3	
Gives instructions on when to return	3	3	
Follows infection control procedures outlined in guidelines		3	
Recognizes/identifies contraindications consistent with guidelines		3	
Performs clinical procedures according to guidelines		3	
Other staff			
Treat clients with dignity and respect	3		
Client			
Participates actively in discussion and selection of method, is “empowered” (composite)	3	3	
Receives her method of choice	3	3	
Believes the provider will keep her information confidential	3		
Facility			
Has all (approved) methods available			3
Has basic items needed for delivery of methods available through service delivery point			3
Offers privacy for pelvic examination/IUD insertion	3	3	3
Has mechanisms to make programmatic changes on the basis of client feedback			3
Has received a supervisory visit in past _ months			3
Has adequate storage for contraceptives and medicines			3
Has state-of-the-art clinical guidelines			3
Has acceptable waiting time	3		3

Notes: The two composite indicators were not included in the analysis because measures that they are based on differed across countries. Five indicators that do not appear on this list were also included in the study: provider sees client in privacy for counseling, asks client if she has any concerns or problems, tells client how to use the method, gives information on side effects and explains that the method does not protect against HIV/AIDS.

was fair to good (kappas, 0.54 and 0.61, respectively). In Ecuador, observers noted that the providers asked 84% of clients if they had any problems or concerns, whereas 87% of respondents answered affirmatively in exit interviews. In Uganda,

Table 1. Percentage of visits exhibiting quality care, by measurement instrument; percentage agreement between instruments; and kappa coefficients indicating strength of agreement—all by country, according to quality of care indicator

Indicator	Ecuador (N=583)				Uganda (N=539)				Zimbabwe (N=736)			
	Observations	Interviews	Agreement	Kappa	Observations	Interviews	Agreement	Kappa	Observations	Interviews	Agreement	Kappa
Interpersonal relations												
Provider treats client with respect/courtesy	99	99	99	0.99	99	99	99	0.98	99	99	99	0.98
Provider sees client in private for counseling	99	87	88	0.74	87	88	82	0.63	97	95	92	0.85
Facility offers privacy for pelvic examination/IUD insertion*	99	93	93	0.85	95	82	82	0.65	98	98	97	0.94
Provider asks client if she has any concerns or problems†	84	87	76	0.54	87	86	81	0.61	u	u	u	u
Choice of methods												
Provider discusses client's fertility intentions‡	53	63	62	0.23	63	53	63	0.26	40	35	70	0.40
Provider discusses with client which method she would prefer§	99	98	98	0.95	74	85	69	0.38	90	94	87	0.76
Client receives her method of choice§	80	84	91	0.82	76	81	82	0.64	89	87	95	0.88
Information given clients												
Provider tells the client how to use the method**	85	97	85	0.70	94	93	89	0.77	85	82	83	0.64
Provider gives accurate information on the method accepted**	84	75	75	0.50	94	100	94	0.78	85	96	83	0.66
Provider gives information on side effects**	71	80	75	0.49	86	77	78	0.57	70	65	70	0.41
Appropriate constellation of services												
Provider mentions STDs or HIV/AIDS	13	27	79	0.57	22	30	69	0.38	12	15	84	0.68
Provider explains that method does not protect against HIV/AIDS††	19	34	73	0.46	40	59	63	0.27	10	52	55	0.08
Provider encourages dual method use††	19	36	74	0.48	26	49	62	0.25	46	57	79	0.62
Mechanisms for continuity												
Provider gives instructions on when to return	94	96	91	0.81	94	94	91	0.81	83	72	70	0.41

*Clients receiving a pelvic examination or IUD insertion: Ecuador (N=403), Uganda (N=34) and Zimbabwe (N=138). †Returning clients only: Ecuador (N=353) and Uganda (N=300). ‡New clients only: Ecuador (N=177), Uganda (N=137) and Zimbabwe (N=198). §New clients with a method preference: Ecuador (N=109), Uganda (N=88) and Zimbabwe (N=150). **New clients who received a method: Ecuador (N=145), Uganda (N=111) and Zimbabwe (N=191). ††New clients who received a nonbarrier method: Ecuador (N=130), Uganda (N=109) and Zimbabwe (N=169). Notes: u=unavailable. Kappa coefficients, which correct for the proportion of responses that would agree because of chance alone, are adjusted for prevalence and bias.

the proportions were 87% and 86%, respectively.

Choice of Methods

To assist new clients in selecting the most appropriate family planning method, providers should ask them about their fertility intentions. Observers noted whether the provider and client discussed her desire for more children or the timing of the next birth; staff conducting exit interviews asked each woman if the provider asked her whether she would like to have more children. In each country, results from observations and exit interviews were comparable (53% and 63% in Ecuador, for example), but agreement between instruments was poor (kappas, 0.23–0.40). The lack of agreement may have stemmed from differences in the items used to calculate the indicator: While the observation guide contained two items that captured the ideas of both spacing and limiting, the exit interview asked only about limiting.

Consistency on whether the provider discussed the client's preferred method during the visit was excellent for Ecuador and Zimbabwe (kappas, 0.95 and 0.76, respectively), and very similar frequencies on responses were found from the two instruments within each country. In Uganda,

however, consistency was poor: Sixty-nine percent of responses were in agreement on this indicator, and the kappa value was 0.38. Rephrasing of the question in the exit interview in Uganda may have led to the inconsistent responses.

In Ecuador and Uganda, the proportion of women who stated during exit interviews that they received their preferred method (84% and 81%, respectively) was slightly higher than the proportion recorded during observations (80% and 76%, respectively). Results were more similar in Zimbabwe: 89% from observations and 87% from exit interviews. Agreement was excellent in Ecuador and Zimbabwe (kappas, 0.82 and 0.88, respectively), and good in Uganda (0.64).

Information Given Clients

Whether clients received information on how to use their chosen method was gathered in two ways during the exit interview. Clients were first asked whether the provider told them how to use the method. They were also asked questions about how the method is used, to assess whether they had correct information about it. For example, pill users were asked, "How often do you take the pill?" By contrast, observers noted only whether

providers gave clients correct information on how to use their selected method. For example, to receive a check for this item, providers must have told clients receiving the pill that it has to be taken every day.

For the indicator on whether the provider told new clients how to use the method, consistency across instruments in all three countries was good to excellent (kappas 0.64–0.77). In Ecuador, almost all discrepant responses were for clients for whom observers recorded that the information was not provided but who reported during exit interviews that they received this information. Information on how to use the method may have been given to clients during a supplemental counseling session not covered by the client-provider observation. No such pattern for discrepant responses was apparent in Uganda and Zimbabwe.

In Uganda and Zimbabwe, the proportions of new clients whom observers considered to have received accurate information on using their method (94% and 85%, respectively) were lower than the proportions who could accurately respond to the interview question on how to use the method (100% and 96%, respectively). The opposite was true in

Ecuador, where observers recorded that 84% of clients were told how to use the method, but only 75% of clients could correctly answer the question posed during the exit interview. Agreement ranged from fair to excellent, depending on the country. A number of reasons could explain the lack of consistency. For example, differences may reflect knowledge acquired during a previous visit (particularly among returning clients in Uganda or Zimbabwe obtaining the pill or injectable). Alternatively, they may be associated with the amount or type of knowledge required for a particular method. Perhaps clients in Ecuador either did not know to check IUD strings or were too embarrassed to mention this during the exit interview. It should also be noted that these questions are fundamentally different from the truly paired questions; therefore, a lack of consistency may not be surprising.

A comparison of the results on whether new clients received information on the side effects of their selected method shows only fair agreement between observations and exit interviews in each country (kappas, 0.41–0.57). The level of agreement ranged from 70% to 78%; the spread between instruments was approximately 10 points in Ecuador and Uganda, and five points in Zimbabwe.

Appropriate Constellation of Services

The QIQ instruments also captured topics other than family planning that were discussed during counseling sessions. Frequencies from exit interviews on these indicators were generally higher than those from observations. In Ecuador, observers reported that 13% of clients received information on HIV or other sexually transmitted diseases (STDs), whereas 27% of clients said they received such information; in Uganda, these proportions were 22% and 30%, respectively. In Zimbabwe, results from both instruments were similar: 12% from observations and 15% from exit interviews. Agreement ranged from poor to fair (kappas, 0.38–0.68). The majority of discrepant responses are for clients who were recorded as not receiving information on HIV and other STDs during observation, but who reported receiving this information when asked during the exit interview. Note that this indicator captures only whether the topic was discussed and does not explore the content of the discussion.

Evidence from the fieldwork suggested that clients received information during their visit to the health facility from other sources in addition to the provider. In

Table 2. Percentage of visits exhibiting quality care, by measurement instrument; percentage agreement between instruments; kappa coefficients indicating strength of agreement; and assessment of bias—for all three countries combined, by quality of care indicator

Indicator	Observations	Interviews	Agreement	Kappa	Evidence of bias*
Provider actions with client (objective)					
Provider discusses with client which method she would prefer†	88	93	85	0.71	No
Provider gives instructions on when to return	89	86	83	0.66	Yes
Provider asks client if she has any concerns or problems‡	86	86	79	0.57	No
Provider discusses client's fertility intentions§	51	49	65	0.30	No
Information given to client (objective)					
Provider tells client how to use the method**	87	90	85	0.69	No
Provider mentions STDs or HIV/AIDS	15	23	78	0.56	Yes
Provider gives information on side effects**	74	73	74	0.48	No
Provider encourages dual method use††	32	48	74	0.47	Yes
Provider explains that method does not protect against HIV/AIDS††	21	48	63	0.26	Yes
Provider gives accurate information on the method accepted**	87	91	84	0.67	Yes
Interpersonal relations (subjective)					
Provider treats client with respect/courtesy	99	99	99	0.98	No
Client receives her method of choice†	83	84	90	0.80	No
Provider sees client in private for counseling	95	91	88	0.75	Yes
Facility offers privacy for pelvic examination/IUD insertion‡‡	99	94	93	0.86	Yes

*Based on McNemar's test for bias, $p < .05$. †New clients with a method preference (N=347). ‡Returning clients (N=653). §New clients only (N=512). **New clients who received a method in Ecuador (N=447). ††New clients who received a nonbarrier method (N=408). ‡‡Clients receiving a pelvic examination or having an IUD inserted (N=575). Note: Kappa coefficients, which correct for the proportion of responses that would agree because of chance alone, are adjusted for prevalence and bias.

Ecuador, information was provided in a separate counseling session (either one-on-one or in groups) conducted by social workers and health educators prior to clients' visits with the provider. In Uganda, about 50% of new clients attended group talks that covered family planning methods and the prevention of HIV and other STDs before seeing the provider. Group talks were also a frequent occurrence at facilities in Zimbabwe. As the client-provider observation did not include information given to clients in these other settings, it is not surprising that the frequencies for the indicators measuring whether information was provided during the visit are higher on the exit interview than what was found during the observation.

For new clients, we also compared indicators that measured whether the provider mentioned that the accepted method (other than condoms) does not protect against HIV, and whether she encouraged dual method use. In all three countries, the frequencies of positive responses for both indicators were higher on the exit interview than on the observation. The spread for the first indicator was about 15 points in Ecuador and Uganda, and more than 40 points in Zimbabwe. Agreement between the two instruments was fair in Ecuador (kappa=0.46), poor in Uganda (0.27) and very poor in Zimbabwe (0.08). Approximately 75% of the discrepant results in Ecuador and Uganda and 97% in Zimbabwe

we reflect instances in which the observers did not mark that this information was given yet the clients reported receiving it.

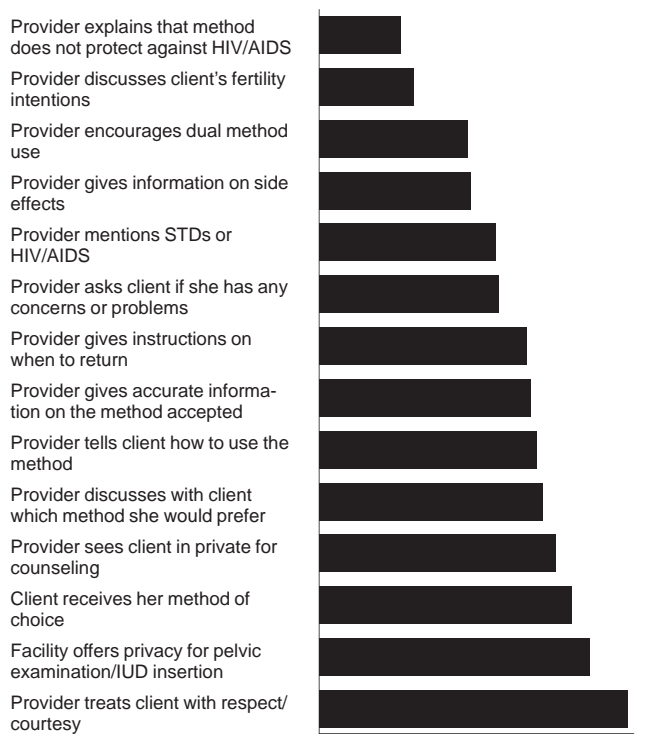
In Uganda and Ecuador, similar patterns were found for whether the provider encouraged dual method use, while in Zimbabwe, differences were much smaller than on the previous indicator. Again, most of the discrepant results (more than 75%) reflected negative responses on the observation and positive responses on the exit interview. This indicator, too, may have been affected by information on STD prevention that clients received in counseling sessions and group talks that were not covered by the observations.

Mechanisms for Continuity

An important indicator for continuity of care is whether providers give clients any instruction regarding their return to the facility. Agreement on this indicator from the two instruments was excellent for Ecuador and Uganda (kappa, 0.81 for each), where providers discussed return visits with nearly all clients. In Zimbabwe, agreement on this indicator was fair (0.41), and observers somewhat more frequently than clients said that such a discussion took place (83% vs. 72%).

Indicator Agreement by Question Type

In Table 2, we present data for all three countries combined. For this table, we have reorganized the indicators to reflect

Figure 2. Kappa coefficients indicating the comparability of 14 indicators measured through observation and client exit interviews

the type of question and degree of comparability of the questions between the instruments. In addition to measures of agreement, we present an assessment of bias—i.e., whether one instrument consistently rated the indicator higher (or lower) than did the other instrument.

The first indicators are objective measures of the provider's actions with the client. Agreement was fair to good (kappas, 0.57–0.71) on three of these indicators and poor on the fourth (0.30). The only indicator for which we found evidence of bias was whether the return visit was discussed. This finding reflects that in Zimbabwe, clients greatly underreported the occurrence of such discussions.

A second set of indicators measure the information exchange that occurred between the client and provider on different topics; we also considered these to be objectively measured. All but one of these indicators had fair to good agreement (kappas, 0.47–0.69). As we noted earlier, clients frequently reported receiving information not recorded during the observation, probably because this occurred during an unobserved part of the visit.

This set of indicators includes one that was calculated from questions that were less than comparable on the two instru-

ments—whether the provider gave the client accurate information on how to use the method she chose. On this indicator, we saw some evidence of better results from the exit interview, primarily because often in Uganda and Zimbabwe, new clients correctly reported key information on how to use their chosen method, yet observers did not record that accurate information was provided. While a client's knowledge of her method may have been obtained during the visit with the provider, in many cases, she may have already had correct information or obtained it from other sources at the health facility.

The third set of indicators measure interpersonal involvement; we deemed these to involve more subjectivity. Surprisingly, agreement

between results from observations and exit interviews was excellent for all of these (kappas, 0.75–0.98), and was actually higher than for the more objective indicators. The two indicators that assessed whether privacy was adequate revealed bias: Fewer clients than observers reported adequate privacy for counseling or examinations. This difference may suggest that observers' perceptions of what constitutes privacy differ from clients', possibly because of observers' familiarity with the health care system and its norms. Effective training of observers can reduce interrater reliability, but it cannot eliminate this difference in perception. For the remaining two indicators, the responses showed no evidence of bias.

Kappa coefficients for the 14 indicators for the three countries combined are presented in Figure 2. Agreement ranged from poor to excellent. Both kappa coefficients and percentage agreement (which ranged from 63% to 99%—see Table 2) gave very similar findings for the indicators.

Discussion

Overall, the results obtained from observations and client exit interviews were highly comparable for most indicators. To the extent that discrepancies occurred, the

major reason for these discrepancies was that clients received information from sources other than the observed client-provider interaction. Such other sources as group talks and supplemental counseling sessions need to be taken into consideration in interpreting the results of this study and in using these instruments in the future.

The consistently high ratings for indicators measuring interpersonal relations may reflect that providers were on their best behavior, since they knew they were under study (the Hawthorne effect). This upward bias, however, should have affected the responses from observations and exit interviews equally—i.e., observers would have recorded better behavior on the part of providers, and clients would have reported the same during exit interviews. (Whether the high ratings on these subjective measures are due to the Hawthorne effect cannot be addressed with the current data.) Moreover, despite the presence of observers, many objective indicators suggest serious deficits in quality in many areas.

A major drawback of exit interviews is courtesy bias. One would expect that indicators measuring subjective states such as attitude, opinions or feelings would be more susceptible to courtesy bias than more objective measures. We did not find this to be true in our data. In fact, agreement was highest on the indicators that we considered more subjective. It is possible, however, that providers were on their best behavior because of the observers' presence, and clients were truthfully reporting good interpersonal relations. Whether the clients would have been as truthful in a situation where providers were rude or unresponsive is not known.

Other sources of error are also possible. Recall bias may account for a client's "forgetting" that a specific instruction or particular information was provided during the visit. Given that the client was interviewed immediately following the visit, she may not have had time to think about the session and process all of the information that she received. This can be seen with indicators that are measured similarly in both instruments: Whether the provider discussed the return visit with the client is an example. Though the questions are relatively straightforward, clients sometimes reported that this was discussed while it was not, and vice versa.

We also considered whether interviewee fatigue and a desire to terminate the interview quickly may have introduced two

types of error. A client may have provided any response to hurry along the interview, resulting in an increase in discordant responses in the latter half of the exit interview. Or she may have provided what she thought was the correct response (often a yes), in hopes of quickly terminating the interview; this would have resulted in a bias toward more positive responses in the second half of the interview. After examining the data from the three countries, we found no relationship with percentage agreement on the instruments and whether the questions appeared earlier or later in the interview. Neither did we see evidence of more biased responses if the questions appeared in the latter half of the exit interview.

Variations in clients' characteristics are another potential source of error that may explain differences in agreement by country. Uganda, which had the lowest percentage agreement for all of the indicators, also had the clients with the lowest levels of education. In an analysis not presented in this article, we found that agreement on many indicators was slightly lower among clients who had not attended secondary school than among those who had. However, this is not sufficient to explain the discrepancies seen.

A final consideration was whether the stigma associated with STDs and HIV may have prohibited some clients from mentioning that these topics were discussed during the session. Results from this analysis do not support this potential bias. A larger proportion of clients than of observation records reported that information on these topics was discussed, because clients may have remembered receiving information in group talks and previous counseling sessions.

As we noted previously, the measurement of quality can be difficult because of the complexity and subjectivity of the topic. Given these difficulties, some error in measurement of quality of care indicators is expected. This measurement error is acceptable as long as it is minimized and inconsistencies can be reasonably explained. Although agreement between instruments was high for most indicators in this study, it was poor for a few. While the source of error could usually be explained, the inconsistencies underscore the need to understand the local context and implementation of the instruments when interpreting the results.

The level of quality differed by country on a number of indicators, but comparability on the instruments for a given country was high. In many cases, results from

observations and exit interviews were extremely close. As a monitoring tool, either method could be used to calculate many of these indicators—as long as a distinction is made about the source of information, the main cause of discrepancies in this study. Our results show that similar conclusions on the quality of care may be reached regardless of the data collection instrument and methodology used. Where inconsistencies occur, judgment on the "correct" answer will often have to be made after taking local conditions into account.

Given the comparability of many of the indicators, it could be argued that there is no need to use both observations and exit interviews; programs may reduce the costs and complexity of data collection by implementing only one of the two instruments. For example, a program that focuses its efforts on improving providers' interpersonal skills as a way to increase client satisfaction may opt to implement only the exit interview. Alternatively, a program that wants to assess the information given to clients by the provider during counseling and clinical examination may choose to implement only the observation.

The QIQ, however, was designed to capture a short list of quality indicators for monitoring family planning programs, and the combined use of its three data collection instruments is recommended so that the full set of indicators can be measured. Selecting only the client-provider observation would eliminate indicators that capture clients' perspective on the care they receive. Selecting only the exit interview, on the other hand, would not permit an assessment of the provider's technical competence during counseling and clinical examination. Neither of these permits assessment of the indicators that the facility audit (not discussed in this article) measures: factors that influence the facility's readiness to provide quality services, such as supplies in stock, conditions of the facility and types of records kept. Therefore, although one instrument may be selected over another where resources are limited, there is a cost in the breadth of indicators that will be available to measure quality.

References

1. Bruce J, Fundamentals of quality of care: a simple framework, *Studies in Family Planning*, 1990, 21(2):61–90.
2. Simmons R and Elias C, The study of client-provider interactions: a review of methodological issues, *Studies in Family Planning*, 1994, 25(1):1–17.
3. Report of the Subcommittee on Quality Indicators in Family Planning Service Delivery, submitted to U.S. Agency for International Development's Task Force on Standardization of Family Planning Program Performance

Indicators, New York, 1990.

4. Bertrand J, Magnani RJ and Knowles J, *Handbook of Indicators for Family Planning Program Evaluation*, Chapel Hill, NC, USA: Carolina Population Center, 1994.
5. Miller R et al., *The Situation Analysis Approach to Assessing Family Planning and Reproductive Health Services: A Handbook*, New York: Population Council, 1997.
6. Bertrand JT, Sullivan T and Rice J, Quick investigation of quality: measuring quality of care in family planning programs, Chapel Hill, NC, USA: Carolina Population Center, 2001; and Sullivan T and Bertrand J, Monitoring quality of care in family planning programs by the quick investigation of quality (QIQ): country reports, MEASURE Evaluation Technical Report, Chapel Hill, NC, USA: Carolina Population Center, 2000, No. 5.
7. Askew I, Mensch B and Adewuyi A, Indicators for measuring the quality of family planning services in Nigeria, *Studies in Family Planning*, 1994, 25(5):268–283; and Huntington D, Miller K and Mensch B, The reliability of the situation analysis observation guide, *Studies in Family Planning*, 1996, 27(5):277–282.
8. Ndhlovu L, Quality in the context of Kenya's integrated reproductive health services, dissertation, University of London, London, 1999.
9. Williams T, Schutt-Ainé J and Cuca Y, Measuring family planning service quality through client satisfaction exit interviews, *International Family Planning Perspectives*, 2000, 26(2):63–71.
10. Byrt T, Bishop J and Carlin JB, Bias, prevalence and kappa, *Journal of Clinical Epidemiology*, 1993, 46(5):423–429.
11. Fleiss JL, *Statistical Methods for Rates and Proportions*, second ed., New York: John Wiley and Sons, 1981.

Resumen

Contexto: Es importante controlar la calidad de atención de los programas de planificación familiar, pero la complejidad y subjetividad del tema presenta muchos desafíos.

Métodos: El sistema de Quick Investigation of Quality (QIQ) fue diseñado para controlar la calidad de la atención de los servicios de planificación familiar mediante la observación de las relaciones entre el proveedor y la cliente, entrevistas a las clientes al final de su visita, y la auditoría de las instalaciones. El QIQ fue probado in situ en muchos países, en 1998 y 1999. Mediante el uso de datos comparables de 583 clientes del Ecuador, 539 de Uganda y 736 de Zimbabwe, este análisis examina las relaciones de los resultados obtenidos mediante observaciones y entrevistas.

Resultados: Con respecto a un indicador determinado, fueron muy diferentes los niveles de acuerdo entre los datos de las observaciones y entrevistas entre un país y otro, aunque al comparar estos datos dentro de cada país, los resultados fueron congruentes entre los diversos instrumentos. El nivel de acuerdo fue entre bueno y excelente en los tres países combinados (kappas, 0.47–0.98) en 13 de los 14 indicadores examinados; las observaciones y las entrevistas realizadas presentaron resultados congruentes en el 63–99% de los casos. El acuerdo presentó su nivel más elevado entre los indicadores que medían las relaciones per-

sonales. Las incongruencias reflejaron principalmente el hecho de que las clientes recibían información por fuera de la interacción observada entre clientes y proveedores.

Conclusiones: Las observaciones y las entrevistas a las clientes ofrecieron resultados similares con respecto a muchos indicadores. Sin embargo, en tanto que algunos programas optan por un instrumento en particular en lugar de otros debido a las restricciones de recursos, la combinación ofrece una evaluación más compleja con respecto a la calidad de la atención.

Résumé

Contexte: Le contrôle de la qualité des prestations offertes dans le cadre des programmes de planning familial est, certes, important, mais la complexité et la subjectivité de la ques-

tion sont sources de nombreuses gageures.

Méthodes: L'ERQ (enquête rapide de qualité) a été mise au point pour contrôler la qualité des prestations sur la base de l'observation des rapports entre clientes et prestataires, d'entrevues organisées à la sortie des clientes et d'audits d'établissement. L'ERQ a été testée sur le terrain, dans plusieurs pays, en 1998–1999. Au départ des données liées de 583 clientes en Équateur, 539 en Ouganda et 736 au Zimbabwe, cette analyse examine la comparabilité des résultats obtenus des observations et des entrevues à la sortie.

Résultats: Pour un indicateur donné, les niveaux de concordance entre les données des observations et des entrevues variaient d'un pays à l'autre. Au sein d'un même pays, les résultats étaient toutefois cohérents d'un instrument à l'autre. Pour les trois pays combi-

nés, la concordance s'est révélée bonne à excellente (kappas, 0,47–0,98) sur 13 des 14 indicateurs examinés; les observations et les entrevues à la sortie ont produit des réponses convergentes dans 63% à 99% des cas. La concordance s'est avérée la plus élevée sur les indicateurs de mesure des relations interpersonnelles. Les divergences reflétaient principalement la réception, par les clientes, d'informations obtenues en dehors du rapport client/prestataire observé.

Conclusions: Les observations et entrevues de clientes à la sortie produisent des résultats fort comparables pour de nombreux indicateurs. Si leurs ressources limitées imposent à certains programmes le choix d'un instrument seulement, on soulignera toutefois que la combinaison des deux permet une meilleure évaluation de la qualité des prestations.