# Evaluation Review

## Challenging the Courtesy Bias Interpretation of Favorable Clients' Perceptions of Family Planning Delivery

Federico R. León, Rebecka Lundgren, Ana Huapaya, Irit Sinai and Victoria Jennings

The online version of this article can be found at:

Published by:
**S**SAGE Publications

http://www.sagepublications.com

**Additional services and information for *Evaluation Review* can be found at:**

**Email Alerts:** http://erx.sagepub.com/cgi/alerts

**Subscriptions:** http://erx.sagepub.com/subscriptions

**Reprints:** http://www.sagepub.com/journalsReprints.nav

**Permissions:** http://www.sagepub.com/journalsPermissions.nav

**Citations** (this article cites 11 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://erx.sagepub.com/cgi/content/abstract/31/1/24#BIBL

# Challenging the Courtesy Bias Interpretation of Favorable Clients' Perceptions of Family Planning Delivery

Federico R. León
Rebecka Lundgren
*Georgetown University Institute for Reproductive Health*
Ana Huapaya
*Instituto de Salud Reproductiva, Lima, Peru*
Irit Sinai
Victoria Jennings
*Georgetown University Institute for Reproductive Health*

Favorable client perceptions of provider's interpersonal behavior in contra-
ceptive delivery, documented in clinic exit questionnaires, appear to contra-
dict results from qualitative evaluations and are attributed to clients' courtesy
bias. In this study, trained simulated clients requested services from Ministry
of Health providers in three countries. Providers excelled in courteousness/
respect in Peru and Rwanda; in India, providers were less courteous and
respectful when the simulated clients chose the pill. Privacy and two-way
communication were less prevalent in all three countries. The findings chal-
lenge the courtesy bias interpretation. Global results from qualitative studies
may have expressed the views of the minority of clients who are not treated
well by providers.

**Keywords:**  *courtesy bias; family planning delivery; service test; client exit
interview*

Interpersonal relations—the affective content of the transaction between
family planning provider and client—are an important aspect of quality of
care. Providers' understanding, respect, honesty, and two-way communication
skills, as well as the privacy and individual counseling afforded, may strongly
influence clients' satisfaction with the services and the probability of a return
visit (Bruce 1990). Therefore, program administrators need periodic information

24

on the strengths and weaknesses of providers' interactions with clients, to reinforce good practices and introduce improvements.

Information on interpersonal relations can be obtained using various methodologies, including exit interviews, qualitative studies, and direct observations. These methodologies yield apparently conflicting results. This study uses another methodology—the service test—to assess interpersonal relations of providers in India, Peru, and Rwanda and reconcile the various findings.

Qualitative studies—nondirective retrospective interviews and focus groups with clients—often yield a negative view of provider interpersonal relations. For example, Mernissi (1975, 424) quoted a Moroccan client as saying,

> If you make a mistake, if you mispronounce a word, the name of a syrup or a pill, the nurse laughs at you, calls her colleagues to tell them the story and points at you. You feel the floor crumble away under your feet.

Scrimshaw (1976) described insults to Ecuadorian women's modesty resulting from the lack of privacy, and Mundigo (1973, 114) quoted as follows a Honduran client who was asked to contribute suggestions for service improvement: "There should be more understanding for us humble people because we go there with great fear."

Negative qualitative findings have also emerged in Asia. In Nepal, Schuler et al. (1985, 264) reported rude treatment of clients by providers. In the words of a couple, "When we asked about other methods he (the provider) got angry and said, 'Why are you mocking me when I have already told you that the operation is the best method for you? Do not behave like fools!'" More recent qualitative studies found similar patterns. For example, Schuler, Choque, and Rance (1994) reported that market

women expected hostile treatment from Bolivian providers, and Schuler and Hossain (1998) observed that Bangladeshi providers were rude to clients of lower socioeconomic status.

However, other methodologies often present highly positive views of providers' interpersonal relations. Evaluators regard interviews with clients exiting the clinic as a useful technique for assessing services from the client perspective. To tap into interpersonal relations, questionnaires include items such as "Did the provider treat you with respect and courtesy?" The client responses are almost always positive, even in countries where qualitative studies present a negative picture of provider interpersonal relations. For example, in a study in Morocco, 94% of the clients interviewed as they left the clinics reported that the providers had treated them courteously (Brown et al. 1995). The percentage was 99 in a study in Ecuador, Uganda, and Zimbabwe (Bessinger and Bertrand 2001). Williams, Schutt-Ainé, and Cuca (2000) found in a study in five Latin American and Caribbean countries that 0.0% of clients identified provider personal courtesy as a problem.

The qualitative studies were not designed to provide population estimates. However, their findings appear to have been interpreted as inconsistent with the quantitative analysis of clients' responses to clinic exit questionnaires, and this has led to the development of the courtesy bias concept. Thus, the emergence of this concept in the literature can be explained as follows: Given the apparent contradiction between the results of the two methodologies, evaluators concluded that one of them must be invalid. Researchers have suggested that the proximity of the providers leads to the positive responses of clients in exit interviews, to please the interviewer (Simmons and Elias 1994; Bertrand et al. 1995). This courtesy bias would mask the pattern of poor provider-client interpersonal relations and would reinforce the status quo (Hull 1996).

The courtesy bias concept has been further developed. Williams, Schutt-Ainé, and Cuca (2000) coined the term *differential courtesy bias* to refer to the situation that would arise if clients found it harder to express negative views for certain types of questions than for others. This extended concept can explain post facto any result. For example, Bessinger and Bertrand (2001) found in three countries that a high percentage of clients viewed interpersonal relations favorably but providers' exchange of information considerably less positively. From a differential courtesy bias angle, this result raises the possibility that clients may have wanted to please the interviewers in one respect but not the other.

Driven by such influential opinion leaders as Simmons and Elias (1994), Bertrand et al. (1995), Hull (1996), and Williams, Schutt-Ainé, and Cuca

(2000), the concept of courtesy bias prevails in the family planning field. The net result is that the family planning community, despite the quantitative empirical evidence showing favorable client perceptions of provider behavior, disbelieves evidence that the overwhelming majority of providers around the world treat their clients well. Yet no direct empirical findings have been offered to support the courtesy bias interpretation. Moreover, this interpretation is contradicted by research results based on various methodologies.

First, courtesy bias cannot explain exit interview findings such as those of Kim et al. (1992), who found in Nigeria that almost all women attended by trained nurses felt that they were listened to carefully, whereas only two thirds of the clients of untrained nurses felt the same way. A simple interpretation of the Kim et al. findings is that the trained nurses actually provided a better quality of care than did the untrained nurses and that this was reflected in the clients' responses.

Second, another methodology—direct observations of the client-provider transaction—suggests the validity of the client exit interview. In direct observation, a trained observer accompanies provider and client during the consultation. Bessinger and Bertrand (2001) found in their study in Ecuador, Uganda, and Zimbabwe that virtually all direct observations indicated that the providers treated clients with respect and courtesy. Because this coincided with the responses of most of the clients interviewed as they exited the clinic, the correlation between client and observer perceptions was extremely high (k = 0.98). Studies using the situation analysis technique, which yields reports from direct observers, also confirm that most clients are treated by providers in a friendly manner (Mensch et al. 1994).

A common problem with these studies is that direct observations are assumed to cause a Hawthorne effect on providers, that is, the exhibition of atypically high levels of task effort (Simmons and Elias 1994; Bertrand et al. 1995; Brown et al. 1995). The empirical evidence shows that family planning providers improve their task performance when they expect to be under observation (Miller et al. 1991; Ndhlovu 1999; León et al. 2003). Bessinger and Bertrand (2001) considered, therefore, that the reports from the clients and observers of their study could have reflected maximal rather than typical provider performance.

However, a study that used a fourth methodology suggests that Hawthorne effects are not necessarily responsible for the generally favorable client reports. The service test consists of a client script that includes a contraceptive history, marital status, method preferences, and other personal details of an imaginary client. A trained simulated client pretends to be the client in the script as she requests services and responds to questions from

a provider, who believes she or he is attending to a real client. When the simulated client leaves the clinic, she completes an observational checklist in a yes-no format (León et al. 2001). To the extent that the provider believes that she or he is attending to a real client, the service test yields results that are free of Hawthorne effects.

In a study in 24 regions of the Peru Ministry of Health, León, Ríos, and Zumarán (2005) trained simulated clients to choose the pill and assessed provider performance on eight behavioral dimensions (interpersonal relations, needs assessment, use instructions for the chosen method, etc.) several months after requesting their consent to be visited by simulated clients. The settings were predominantly urban. The service test's interpersonal relations dimension generated the highest score, and this was closer to the maximum possible score than to the average on the other dimensions (entailing information exchange). This finding suggests that providers actually offered good interpersonal relations but less satisfactory information exchange. Thus, a technique that was free of Hawthorne effects contributes evidence confirming that clients typically receive positive interpersonal treatment at clinics.

The service test findings, however, were limited to one country. To throw further light on this issue, this study uses the service test methodology to examine providers' interpersonal relations across geographic boundaries. The study also focuses on the simulated clients' ability to differentiate facets of interpersonal relations. Some qualitative materials (Schuler et al. 1985; Schuler, Choque, and Rance 1994; Schuler and Hossain 1998) suggest that providers are patronizing and authoritarian, involving the client in hierarchical relationships with one-way communication. This study tests the hypothesis that providers will excel in respect/courtesy (as shown in exit interviews) but not in two-way communication with clients (as suggested by qualitative studies).

## Method

The study was conducted in India, Peru, and Rwanda as part of a larger effort to assess the impact of introducing the standard days method (SDM), a new fertility awareness–based method of family planning (Arévalo, Jennings, and Sinai 2002), into ongoing services. The study subjects in all three sites were Ministry of Health providers. Modern contraceptive prevalence varies greatly among the various countries. At the time of data collection, the SDM had not yet been introduced in the study sites, and the participating providers had not yet been trained to offer it. The data for this study came from the baseline of the larger project.

The study sites are culturally and economically diverse. The site selected in India was the state of Jharkhand. The population is mostly (80%) rural, poverty is endemic, and only 25% of married women of reproductive age use a modern method of contraception (International Institute for Population Sciences and ORC Macro 2001). Within Jharkhand, the Ranchi district was selected, and within it, three blocks: Burmu, Kanke, and Ormanjhi. The research sites in Peru were the provinces of Jaén in the Cajamarca department (58% urban) and Moyobamba in the San Martín department (64% urban). Modern contraceptive prevalence in Cajamarca is 41.1% and in San Martín 57.5% (Peru Instituto Nacional de Estadística e Informática, 2001). The sites in Rwanda were the poverty-ridden provinces of Byumba (91% rural) and Kibungo (87% rural). Modern contraceptive prevalence in these areas is only 3.3% and 5.3%, respectively (Rwanda National Census Service 2003).

Providers in the seven study sites participated in a contraceptive update workshop 2 to 5 weeks before data collection began. The training addressed action mechanisms, use instructions, contraindications, side effects, and warning signs of the methods included in the family planning programs of each country. The workshops used the accepted family planning training curricula in each country. Good interpersonal relations with clients were recommended but not emphasized. At the end of the workshop, the trainees in the three countries gave their consent to be visited by simulated clients. The workshop lasted 2 days each in Peru and India and 9 days in Rwanda.

## Service Test Scripts and Checklist

Client scripts in each country reflected the most prevalent methods used: oral contraceptive in Rwanda, depot medroxy-progesterone acetate (DMPA or Depo-Provera) in Peru, and female sterilization (tubal ligation) in India. A fourth script pertained to the SDM.

The following service test "pill user" profile was used in the three countries: "new in town; wife of a small trader; 25 years old; two children (3 and 2 years old); not breastfeeding; mutually faithful relationship; no family violence; healthy; wishes to have children in the future; used condoms in the past, but husband disliked using it every time the couple had intercourse; currently using condoms inconsistently (not always and incorrectly); wants to switch to another method; knows little or nothing about other methods; does not trust natural family planning; afraid of inserting anything into uterus; would not want to be injected (fear of needles); will choose pill if given the option; on Day 4 of her menstrual cycle; will reject pelvic exam (ashamed); when the provider describes the pill, will ask, 'How long can I use this method?' "

The DMPA script, that is, one in which the client was instructed to choose DMPA, used in Peru and Rwanda, differed from the pill script in one important aspect—the woman's menstrual status. Whereas the pill script indicated the 4th day of her cycle, the DMPA script indicated the 20th day. Given this cycle day, it was expected that the simulated client would not be offered an injection that day but would be asked to come back when she got her period. Nonetheless, full DMPA counseling was expected. The female sterilization script used in India depicted an older woman on the 4th day of the cycle who would choose tubal ligation. In this script, the client asked, "How long can I think about it before I make a decision?" The provider was expected to counsel the client and refer her to a facility qualified to perform the sterilization. The SDM script, used in the three countries, depicted a young woman on her 4th day of menstruation who is afraid of the side effects of hormonal contraceptive methods. Simulated clients using the SDM script were instructed to reject all other methods offered by the providers except the related rhythm method that is occasionally offered in Peru and Rwanda. Otherwise, they would tell the provider that they preferred to consult with their husbands before receiving any contraceptive method.

To measure interpersonal relations, a single observational checklist was used along with the four scripts. It included the following nine items: Counseling was individual, the counseling session was interrupted, there were strangers hearing what I said, the provider treated me amiably, I felt he or she cared for my health, he or she looked annoyed, he or she treated me respectfully, he or she asked me if I had any questions, and he or she responded to my questions. The items were scored 1 (yes, or observed) or 0 (no, or not observed). Then, three items (the counseling session was interrupted, there were strangers hearing what I said, and he or she looked annoyed) were recoded (1 = 0, 0 = 1) so that the 1s corresponded to good interpersonal relations in the nine items.

The checklists also included other aspects of providers' behavior that are of less interest in the context of this article: need diagnosis, method options offered, and, with respect to the method chosen by the client, information exchanged on contraindications, use instructions, action mechanisms/advantages/disadvantages, side effects/warning signs, and follow-up. Each included from 7 to 13 items, following the same yes-no format.

## Recruitment and Training of Simulated Clients

To ensure adequate cognitive abilities in data collectors, we recruited women with a high school education and preselected a group on the basis of interviews and psychological tests. In Peru, all the simulated clients were

recruited locally (in Moyobamba and Jaén), and most were university students. They had the same ethnic appearance as most clients of the health services and went to the field dressed in the manner of the typical client. The Rwandan clients were recruited in the capital city of Kigali, and most were nurses, not working for the Ministry of Health. In India, village health workers identified women of the Ranchi district from the same socioeconomic status as that of typical public sector clients who could convincingly play the client scripts in Burmu, Kanke, and Ormanjhi. To control for courtesy bias, we ensured that simulated clients were not associated with the Ministry of Health.

Training of simulated clients was extensive and lasted 5 days. The first 3 days were dedicated to introductory presentations and role-playing exercises in the classroom using the client scripts and checklist. Each candidate was trained on only one script. The simulated clients were instructed to approach the clinics asking for family planning services; they were taught to avoid volunteering information and just respond to provider's questions. The concept of courtesy bias was not presented in the training, to avoid focusing the attention of the simulated clients on a particular aspect of the consultation. However, written instructions to trainers who played the role of providers specified various levels of interpersonal relations, and each trainee conducted role plays at each level and received feedback. Training was in the local language for each location.

The remaining 2 days of the training were dedicated to realistic practices in the field, including supervised visits to facilities. At the end of training, participants were ranked by the trainers, and those with the highest ranking were selected to be simulated clients.

## Data Collection

The data were collected from October 2004 to January 2005. Each Peruvian and Rwandan clinic was visited by a pill client, a DMPA client, and an SDM client, in random order. Similarly, each Indian clinic was visited by one simulated client for each script—pill, female sterilization, and SDM. Nine simulated clients participated in India (one per block for each script) and six in Peru and Rwanda, respectively (one per province for each script). The simulated clients worked in teams, two or three supervised three-person teams (one simulated client per script) operating, one per province or block. The simulated clients did not ask to see a specific provider. Rather, they asked for family planning services and received counseling from the provider assigned to them. To avoid observational bias

within each country, each team visited facilities in each block or province; that is, each team worked in all the blocks/provinces.

In rural settings, the simulated clients had to be dropped off far from the clinic and in some cases walk several kilometers to avoid detection. Nonetheless, comments from Peruvian providers suggested that a few simulated clients were discovered in rural areas. In most cases, the clinic had only one provider who received the three visits. In a number of instances, the simulated clients encountered clinics that could not attend them: The provider was absent because he or she was visiting communities as part of a health campaign or because of personal or other reasons. In these cases, the simulated client returned to the clinic on a different day. In India, some simulated clients did not receive services because the provider offered no family planning methods. These cases were excluded from the analyses. Most providers in Peru and Rwanda refused to provide DMPA to the client and asked her to return at the onset of her menses. In these cases, a consultation took place, and the simulated clients completed the checklist.

## Analysis

Each of the nine interpersonal relations items received a 1 or 0 score. Within each script, the proportion of 1s (i.e., the item average) was calculated for each item. These proportions were then averaged across various items. We use various significance tests (binomial test, sign test) to measure the prevalence of certain responses across locations and further analyze the various items that constitute interpersonal relations. Logistic regression is used in the special case of India. To estimate the interpersonal relations parameter across locations, we added, for each case, the nine item scores and divided by nine. Then we calculated the mean and its 0.95 confidence interval. Similar operations were performed for subgroups of items.

## Results

Table 1 presents the average item scores for the nine interpersonal relations items. As expected, results show that a majority of providers treated clients well, with "yes" response proportions ranging from .53 through .88. These results are statistically significant (item mean > .50, binomial test; Siegel 1956) for the pill script ($p < .01$, one-tailed, $n = 7$), the SDM script ($p < .01$, one-tailed, $n = 7$), and the combined sterilization and DMPA scripts ($p < .01$, one-tailed, $n = 7$). The significance test requires that each

**Table 1**
**Average Item Scores for Nine Interpersonal**
**Relations Items, per Service Test Script and Location:**
**India, Peru, and Rwanda, 2004-2005**

| Country | Location[a] | Pill | SDM | Sterilization/ DMPA[b] | Mean Over Scripts |
|---|---|---|---|---|---|
| | | | Service Test Script | | |
| India | Burmu (*n* = 18.3) | .53 | .63 | .83 | .66 |
| | Kanke (*n* = 28.7) | .66 | .75 | .81 | .74 |
| | Ormanjhi (*n* = 24.0) | .68 | .76 | .78 | .74 |
| Peru | Jaén (*n* = 32) | .72 | .73 | .70 | .72 |
| | Moyobamba (*n* = 30) | .79 | .79 | .79 | .79 |
| Rwanda | Byumba (*n* = 19.3) | .77 | .82 | .88 | .82 |
| | Kibungo (*n* = 19.7) | .67 | .75 | .68 | .70 |
| Mean over locations | | .70 | .75 | .78 | .74 |
| Lower and upper limits of 0.95 confidence interval | | .66, .73 | .72, .78 | .75, .80 | .72, .76 |

Note: SDM = standard days method; DMPA = depot medroxy-progesterone acetate.
a. *n* is the average number of visits by the simulated clients. For example, Burmu appears with an *n* = 18.3, which is the average for the number of visits received by providers under the pill (*n* = 18), SDM (*n* = 19), and sterilization (*n* = 18) scripts. In the other locations, the differences were equally trivial or nonexistent.
b. Sterilization in the case of India and DMPA in the case of Peru and Rwanda.

of the seven averages be above .50 to reject the null hypothesis that the proportion is less than .50 and thus is very informative concerning the extent to which the phenomenon transcends locations. However, it is not very instructive concerning the magnitude of the proportion. This magnitude is provided by the mean average over locations and the corresponding 0.95 confidence interval. The average ranged from .70 (pill script) through .78 (combined DMPA/sterilization scripts), and the grand mean reached .74, which indicates that good interpersonal relations with clients are present in three of every four instances in clinic environments in the three countries. These results lend empirical support to the hypothesis that family planning providers in developing countries offer reasonably good interpersonal relations in their interactions with clients.

Our second hypothesis stated that the simulated clients would be able to discriminate various facets of interpersonal relations and that the providers

**Table 2**
**Average Item Scores for Providers' Courteousness, Privacy,**
**and Two-Way Communication, per Location:**
**India, Peru, and Rwanda, 2004-2005**

| Country | Location | Interpersonal Relations Facet | | |
|---|---|---|---|---|
| | | Courteousness | Privacy | Two-Way Communication |
| India | Burmu ($n$ = 18.3) | .75 | .74 | .37 |
| | Kanke ($n$ = 28.7) | .87 | .69 | .50 |
| | Ormanjhi ($n$ = 24.0) | .84 | .74 | .51 |
| Peru | Jaén ($n$ = 32) | .94 | .69 | .51 |
| | Moyobamba ($n$ = 30) | .96 | .89 | .53 |
| Rwanda | Byumba ($n$ = 19.3) | 1.00 | .65 | .69 |
| | Kibungo ($n$ = 19.7) | .92 | .53 | .42 |
| Mean over locations | | .90 | .71 | .51 |
| Lower and upper limits of 0.95 confidence interval | | .88, .92 | .68, .74 | .47, .54 |

Note: Each data point is an average of the results for the pill, SDM, and sterilization scripts in India and for the pill, SDM, and DMPA scripts in Peru and Rwanda. SDM = standard days method; DMPA = depot medroxy-progesterone acetate.

would excel in respect and courtesy though not necessarily in two-way communications with clients.

Table 2 compares three facets. We measured courteousness through the average score of two service test items: "The provider treated me amiably" and "treated me respectfully." Privacy was measured by the average score of "the counseling session was not interrupted," and "there were no strangers hearing what I said." Two-way communication was measured by the average score of "asked me if I had any questions" and "responded to my questions."

In an overwhelming majority of cases in Peru and Rwanda (more than 90%), the simulated clients reported that they had received counseling from a courteous provider. In India, the percentage was not as high but still substantial, ranging from .75 in Burmu through .87 in Kanke. This result (mean > .50) was statistically significant ($p$ < .01, one-tailed, $n$ = 7), and the global average reached .90.

The simulated clients depicted privacy as significantly less prevalent. The comparison of the respective columns in Table 2 yields a statistically significant finding using the sign test (Siegel 1956) over the seven locations

**Table 3**
**Average Item Scores for Three Interpersonal Relations Items**
**per Location: India, Peru, and Rwanda, 2004-2005**

| Country | Location | Interpersonal Relations Item | | |
| | | Individual Counseling | Caring for Client's Health | Not Looking Annoyed |
|---|---|---|---|---|
| India | Burmu (*n* = 18.3) | .73 | .53 | .96 |
| | Kanke (*n* = 28.7) | .85 | .73 | 1.00 |
| | Ormanjhi (*n* = 24.0) | .71 | .78 | .97 |
| Peru | Jaén (*n* = 32) | .79 | .49 | .91 |
| | Moyobamba (*n* = 30) | .92 | .46 | .93 |
| Rwanda | Byumba (*n* = 19.3) | .91 | .92 | .91 |
| | Kibungo (*n* = 19.7) | .85 | .76 | .92 |
| Mean over locations | | .83 | .65 | .94 |
| Lower and upper limits of 0.95 confidence interval | | .79, .86 | .61, .69 | .92, .96 |

Note: Each data point is an average of the results for the pill, SDM, and sterilization scripts in India and for the pill, SDM, and DMPA scripts in Peru and Rwanda.
SDM = standard days method; DMPA = depot medroxy-progesterone acetate.

(courteousness greater than privacy, $p < .01$, one-tailed, $n = 7$). This finding shows that interpersonal relations can yield systematically different results depending on what facet is measured. Nonetheless, the observations, ranging from .53 in Kibungo through .89 in Moyobamba, showed that privacy prevailed in the three countries (using the binomial test, $p < .01$, one-tailed, $n = 7$), and the global average reached .71.

Two-way communication was even less prevalent, with scores ranging from .37 in Burmu through .69 in Byumba, with an average of .51. This facet did not prevail across the seven locations ($p < .77$, one-tailed, $n = 7$, binomial test) and was significantly less prevalent than was courteousness ($p < .01$, one-tailed, $n = 7$, sign test). Beyond statistical significance, courteousness was sizably greater compared with two-way communication; the ratio ranged from 1.45:1 in Byumba to 2.19:1 in Kibungo. These results confirm, then, that simulated clients are able to discriminate various facets of interpersonal relations. Providers can excel in respect and courtesy, without good two-way communication skills.

Table 3 presents the average scores over scripts for the other interpersonal relations items: "counseling was individual," "I felt (the provider) cared for my health," and "he or she did not look annoyed." Individual

counseling ($p < .01$, one-tailed, $n = 7$) and not looking annoyed ($p < .01$, one-tailed, $n = 7$) significantly prevailed in a majority of cases, whereas caring for client's health ($p < .23$, one-tailed, $n = 7$) did not. This represents further evidence that the simulated clients reported overwhelming "yes" responses only for certain interpersonal relations items.

## The Case of India

The figures for India differ from those of Peru and Rwanda in that there is more variability between service test scripts. Providers scored high on the courteousness items when counseling simulated clients who were using the sterilization and SDM scripts, but they only scored moderately on the pill script. This finding suggests that India providers may offer differentiated treatment depending on the client's method preference. Table 4 shows the detailed scores of the interpersonal relations items for the three scripts.

According to the respective coefficients from logistic regressions that considered the item score as dependent variable (yes = 1, no = 0) and the type of script as predictor (e.g., SDM = 1, pill = 0), a significantly greater presence of "yes" responses is observed in the case of the SDM script compared with the pill script for "there were no strangers hearing what I said," "the provider treated me amiably," "treated me respectfully," and "asked me if I had any questions." The opposite occurred for "responded to my questions." The difference in summated scores over the nine interpersonal items between these two scripts was significant according to the Mann-Whitney $U$ test, a nonparametric comparison that is insensitive to nonnormality of the distributions of scores. (The distributions of summated scores were nonnormal according to the Kolmogorov-Smirnov test.) Similar are the results for the comparison of the pill and sterilization scripts. Beyond statistical significance, the differences between scripts were quite large in several instances. With respect to "treated me amiably," the ratio was 2.40 (sterilization script) to 1 (pill script). The providers generally referred sterilization clients to a scheduled tubal ligation clinic. On the other hand, they did not deliver the pill, or advice on where to obtain it if not available at the clinic, to 39% of the clients who chose this method.

# Discussion

We used the service test to examine providers' interpersonal relations behavior and evaluate the courtesy bias concept. Our results show that trained simulated clients in seven locations in India, Peru, and Rwanda

**Table 4**
**Average Item Scores per Service Test Script, Coefficients
From the Logistic Regression of Item Scores on Scripts,
Summated Score Over Items per Script, and Comparison of
Summated Scores—India, 2004-2005**

| Item | Item Score per Client Script and Logistic Regression Coefficients | | | | |
|---|---|---|---|---|---|
| | Pill ($n = 70$) | SDM ($n = 74$) | Exp($B$) | Sterilization ($n = 69$) | Exp($B$) |
| The counseling session was individual | .73 | .74 | 1.08 | .84 | 1.96 |
| The counseling session was not interrupted | .79 | .77 | 0.91 | .83 | 1.29 |
| There were no strangers hearing what I said | .54 | .72 | 2.12* | .65 | 1.58 |
| The provider treated me amiably | .40 | .92 | 17.00** | .96 | 33.00** |
| I felt he or she cared for my health | .51 | .62 | 1.55 | .96 | 20.78** |
| He or she did not look annoyed | .99 | .96 | 0.34 | 1.00 | —[a] |
| Treated me respectfully | .73 | .97 | 13.41** | .99 | 25.33** |
| Asked me if I had any questions | .19 | .36 | 2.52** | .43 | 3.37** |
| Responded to my questions | .83 | .43 | 0.16** | .59 | 0.30** |
| Summated score | 5.70 | 6.49 | $U = 1,247$** | 7.25 | $U = 2,056$** |

Note: The pill script is compared with the SDM script and with the sterilization script by
means of logistic regression (item scores) and the Mann-Whitney $U$ test (summated scores).
SDM = standard days method.
a. Exp(B) could not be properly calculated because one cell had a zero. A $c^2$ test
produced a nonsignificant result.
*$p < .05$, two-tailed. **$p < .01$, two-tailed.

were treated well by providers. These results, which are consistent with the
service test findings of León, Ríos, and Zumarán (2005) in Peru, suggest
that good interpersonal relations with clients prevail in clinic environments
of developing countries in Africa, Asia, and Latin America. However, qual-
ity of care in the clinics studied was far from perfect—the overall mean of
the interpersonal relations score was only .74 (Table 1). But although in

theory less-than-optimal treatment of even a single client is unacceptable, our results clearly establish that good interpersonal relations prevail in clinical settings in developing countries.

As expected considering the qualitative and quantitative literatures, the desirable provider behaviors were observed more frequently in the specific facet of provider courteousness (amiability and respect) than in other facets of interpersonal relations. Peruvian and Rwandan simulated clients in this study reported that they received services from a respectful and amiable provider in an overwhelming majority of cases (more than 90%). The percentage reported in India, although not quite so high, was substantial (range 75% to 87%), and the grand mean for the seven locations reached .90. This is significantly lower than the level of courteous method provision that is often reported in client exit interviews (e.g., Brown et al. 1995; Williams, Schutt-Ainé, and Cuca 2000; Bessinger and Bertrand 2001) but still overwhelmingly high.

Our findings challenge the concept of courtesy bias that has been developed to explain the favorable clients' perceptions of providers' respect and courtesy documented by clinic exit interviews. Simulated clients trained to supply discriminating observations in the interpersonal area offered generally excellent appraisals of provider amiability and respect. They also supplied responses that in themselves expressed a discriminating power—different levels of accomplishment for the various facets of interpersonal relations (two-way communication and privacy significantly lower than respect and courtesy). If simulated clients were trying to be pleasant, they would have reported high levels of providers' good interpersonal relations across the board.

The concept of differential courtesy bias proposed by Williams, Schutt-Ainé, and Cuca (2000) could be applied to explain the depiction of the providers as amiable and respectful but without good two-way communication skills, as well as the contrasts observed in other behavioral items. That is, the simulated clients might have found it harder to express negative views for certain provider behaviors than for others. However, this concept cannot account for another key finding of this study—the differential results reported to a single set of questions for the various service test scripts used in India. In this country, simulated clients using the pill script reported that providers treated them less positively, in comparison with those using the SDM and sterilization scripts. This was reflected in the summated scores for the interpersonal area as well as in the scores for individual items, including "the provider treated me amiably" and "treated me respectfully." Differential courtesy bias does not apply to these situations

and should be ruled out as an explanation of the differences in provider behavior observed in this study.

A possible reason for the India observations concerning simulated clients who chose the pill is method shortage. This is a frequent occurrence in India. Kahn, Gupta, and Patel (1999) noted shortages and erratic supplies of temporary contraceptive methods in the Indian state of Uttar Pradesh, where 33% of rural facilities reported that oral contraceptives were not regularly supplied, and nearly half of the 2,428 public facilities surveyed reported running out of contraceptive supplies at least once in the previous year. The situation was similar in Jharkhand at the time of our study according to the observations of project staff; thus, at least 39% of the simulated clients enacting the pill script did not receive a package of pills during their visits. Perhaps some of the providers were frustrated by the recurrent lack of contraceptive supplies that was made salient by the simulated clients' choice of the pill, and this was reflected in the simulated clients' perceptions of their interpersonal behaviors. Another possible reason is that the Indian providers may be generally reluctant to provide a user-dependent method that requires an ongoing effort on their part to provide follow-up counseling and resupply to clients.

On the other hand, the Indian providers showed excellent interpersonal relations with the simulated clients instructed to choose sterilization. After the International Conference on Population and Development held in El Cairo in 1994, the government of India abandoned a quota system that promoted sterilization and adopted a target-free approach for family planning. However, tubal ligation continues to be the preferred choice of family planning offered by providers and adopted by users (Koenig, Foo, and Joshi 2000). Indian providers are more experienced counseling women on tubal ligation than on other contraceptive methods and may have derived satisfaction from the visit of a client who was fit for sterilization and chose this method. It is not surprising, therefore, that the Indian providers demonstrated their best interpersonal relations with these clients.

Providers in the study expected to receive visits from simulated clients. Moreover, in some rural settings, providers were so familiar with the community residents that they immediately detected the simulated clients as strangers. Therefore, our results may have been influenced by the Hawthorne effect—a situation in which providers perform better because they know they are being observed. However, we expect that this effect was minimal for three reasons. First, the simulated client visits occurred several weeks after the providers gave their consent to participate in the study. Second, Rwanda, where the percentage of rural clinics visited was extremely high, reported interpersonal relations behaviors that were on par with those

observed in the mostly urban clinics visited in Peru. Third, the fact that about half of the Indian providers did not treat amiably the clients who requested pills suggests that the providers were not attempting to show their best behavior. Our findings, then, strongly suggest that the majority of providers in many settings around the world treat clients in a friendly and courteous manner under typical (observation-free) service circumstances.

What, then, are the implications of these findings for the resolution of the apparent inconsistency observed between the results of clinic exit questionnaires and the accounts from qualitative studies that gave rise to the concept of courtesy bias? Our service test results uphold the first set of findings: A substantial majority of clients in India, Peru, and Rwanda indeed received family planning services in the context of good providers' interpersonal behaviors. A minority of clients, approaching 50% when the client chose the pill in India, were not well treated by providers. This suggests that the qualitative studies that reported providers' maltreatment of clients may have reflected the service experiences of a minority of clients. That is, the apparent inconsistency between the qualitative and quantitative findings in the literature would be explained by study results that expressed the views of a majority of cases (positive views in exit interviews) versus the views of a minority of cases (negative views in qualitative studies). The latter would have attained centrality because of their intrinsic dramatic power rather than their being quantitatively representative. It is also possible that some of the qualitative studies worked with cases outside the mainstream. In addition, the client exit interviews focused almost exclusively on providers' respect and courtesy, whereas the qualitative studies gave more emphasis to other aspects of providers' interpersonal relations (authoritarianism, etc.). Our results indicated that providers tend to excel in respect/courtesy but not in two-way communication and other aspects of interpersonal relations.

Because the service test presents a good level of interjudge agreement (León et al. 2005), the findings reported are reliable. To ensure the validity of results, the simulated clients were independent of the health system, which facilitated the expression of objective observations. Also, their training was especially rigorous in the attention it paid to the discrimination between various levels of interpersonal relations. As for the rigor of the analysis, the use of multiple tests could have led to an inflated number of significant findings. Many of these, however, were at the stringent .01 level, and all were consistent and easily interpretable.

A final word of caution is opportune. Demonstrating that a majority of clients receive good interpersonal relations should not make us oblivious to the situation of the minority of clients who are maltreated by providers. Evaluation methodologies should be able to detect these cases and generate

corrective actions. The service test detected rather important minorities of cases that were not treated well by providers. In contrast, the extremely positive findings of Brown et al. (1995), Bessinger and Bertrand (2001), and Williams, Schutt-Ainé, and Cuca (2000) suggest the presence of Hawthorne effects in their studies, due to the undergoing evaluation activities perceived by providers. A reassessment of instruments used in quality of care evaluation in the family planning field may be in order.

# References

Arévalo, M., V. Jennings, and I. Sinai. 2002. Efficacy of a new method of family planning: The standard days method. *Contraception* 65:333-38.

Bertrand, J. T., K. Hardee, R. J. Magnani, and M. A. Angle. 1995. Access, quality of care and medical barriers in family planning programs. *International Family Planning Perspectives* 21 (2): 64-69, 74.

Bessinger, R. E., and J. T. Bertrand. 2001. Monitoring quality of care in family planning programs: A comparison of observations and client exit interviews. *International Family Planning Perspectives* 27 (2): 63-70.

Brown, L., M. Tyane, J. Bertrand, D. Lauro, M. Abou-ouakil, and L. deMaria. 1995. Quality of care in family planning services in Morocco. *Studies in Family Planning* 25 (3): 154-68.

Bruce, J. 1990. Fundamental elements of the quality of care: A simple framework. *Studies in Family Planning* 21 (2): 61-91.

Hull, V. J. 1996. *Improving Quality of Care in Family Planning: How Far Have We Come?* South and East Asia Regional Working Paper No. 5. New York: Population Council.

International Institute for Population Sciences and ORC Macro. 2001. *National Family Health Survey (NFHS-2), India, 1998-99: Jharkhand*. Mumbai, India: International Institute for Population Sciences.

Kahn, M. E., R. B. Gupta, and B. C. Patel. 1999. The quality and coverage of family planning services in Uttar Pradesh: Client perspectives. In *Quality of care within the Indian family planning program: The challenge ahead,* ed. M. E. Koenig and M. E. Kahn, 49-69. New York: Population Council.

Kim, Y. M., J. Rimon, K. Winnard, C. Corso, I. V. Mako, S. Lawal, S. Babalola, and D. Huntington. 1992. Improving the quality of service delivery in Nigeria. *Studies in Family Planning* 23 (2): 118-27.

Koenig, M. A., G. H. C. Foo, and K. Joshi. 2000. Quality of care within the Indian family welfare program: A review of recent evidence. *Studies in Family Planning* 31 (1): 1-18.

León, F. R., C. Brambila, M. de la Cruz, J. Garcia Colindres, C. Morales, and B. Vasquez. 2005. Provider compliance with the balanced counseling strategy in Guatemala. *Studies in Family Planning* 36 (2): 117-26.

León, F. R., V. Espinoza, A. Espinoza, and B. Meza. 2003. Expectancy of being observed and quality of care offered. Paper presented at the 131st annual meeting of the American Public Health Association, San Francisco.

León, F. R., R. Monge, A. Zumarán, I. García, and A. Ríos. 2001. Length of counseling sessions and amount of relevant information exchanged: A study in Peruvian clinics. *International Family Planning Perspectives* 27 (1): 28-33, 46.

León, F. R., A. Ríos, and A. Zumarán. 2005. Training ´ trainee interactions in a family plan-
    ning intervention. *Evaluation Review* 29 (6): 576-90.

Mensch, B. S., R. A. Miller, A. A. Fisher, J. Mwita, N. Keyonzo, F. Y. Ali, and C. Ndeti. 1994.
    Family planning in Nairobi: A situation analysis of the city commission clinics. *International
    Family Planning Perspectives* 20 (2): 48-54.

Mernissi, F. 1975. Obstacles to family planning in urban Morocco. *Studies in Family Planning*
    6 (12): 418-25.

Miller R. A., L. Ndhlovu, M. M. Gachara, and A. A. Fisher. 1991. The situation analysis study
    of the family planning program in Kenya. *Studies in Family Planning* 22 (3): 131-43.

Mundigo, A. 1973. Honduras revisited: The clinic and its clientele. In *Clinics, contraceptives,
    and communication*, ed. J. Mayone Stycos, 98-115. Des Moines, IA: Meredith Corporation.

Ndhlovu, L. 1999. Quality in the context of Kenya's integrated reproductive health services.
    Diss., University of London.

Peru Instituto Nacional de Estadística e Informática. 2001. *Peru: Encuesta demográfica y de
    salud familiar 2000*. Lima, Peru: Instituto Nacional de Estadística e Informática.

Rwanda National Census Service. 2003. *3rd general census of population and housing:
    Report on the preliminary results*. Kigali, Rwanda: National Census Service.

Schuler, S. R., M. E. Choque, and S. Rance. 1994. Misinformation, mistrust, and mistreatment:
    Family planning among Bolivian market women. *Studies in Family Planning* 25 (4): 211-21.

Schuler, S. R., and Z. Hossain. 1998. Family planning clinics through women's eyes and
    voices: A case study from rural Bangladesh. *International Family Planning Perspectives*
    24 (4): 170-75, 205.

Schuler, S. R., E. N. McIntosh, M. C. Goldstein, and B. R. Pande. 1985. Barriers to effective
    family planning in Nepal. *Studies in Family Planning* 16 (5): 260-70.

Scrimshaw, S. 1976. Women's modesty: One barrier to the use of family planning clinics in
    Ecuador. In *Culture, natality, and family planning*, ed. J. Marshall and S. Polgar, 167-87.
    Chapel Hill: Carolina Population Center, University of North Carolina.

Siegel, S. 1956. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Simmons, R., and C. Elias. 1994. The study of client-provider interactions: A review of
    methodological issues. *Studies in Family Planning* 25 (1): 1-17.

Williams, T., J. Schutt-Ainé, and I. Cuca. 2000. Measuring family planning service quality
    through client satisfaction exit interviews. *International Family Planning Perspectives* 26 (2):
    63-71.

**Federico R. León** is a consultant at the Georgetown University Institute for Reproductive
Health in Washington, D.C.

**Rebecka Lundgren** is the director of operations research at the Georgetown University
Institute for Reproductive Health.

**Ana Huapaya** is a consultant at the Instituto de Salud Reproductiva in Lima, Peru.

**Irit Sinai** is an evaluation specialist at the Georgetown University Institute for Reproductive
Health.

**Victoria Jennings** is the director of the Georgetown University Institute for Reproductive
Health.