# ARE CLIENT SATISFACTION SURVEYS USEFUL? EVIDENCE FROM MATCHED FACILITY AND HOUSEHOLD DATA IN MADAGASCAR

Peter Glick

Cornell University
3M12 MVR Hall
Ithaca, NY, 14853
Telephone 607 254-8782
Fax 607 255-0178
E-mail pjg4@cornell.edu

September 2007

## SUMMARY

Client satisfaction surveys in developing countries are increasingly being promoted as a means of understanding health care quality and the demand for these services. However, concerns have been raised about the reliability of responses in such surveys: for example, 'courtesy bias' may lead clients, especially if interviewed upon exiting clinics, to provide misleadingly favorable responses. This study uses unique data from Madagascar to investigate these and other issues. Identical questions about satisfaction with local health care centers were asked in user exit surveys and in a population based household survey; the latter would be less contaminated by courtesy bias as well as changes in provider behavior in response to being observed. We find strong evidence that reported satisfaction is biased upward in exit surveys for subjective questions regarding (for example) treatment by staff and consultation quality, but is not biased for relatively objective questions about facility condition and supplies. The surveys do provide useful information on the determinants of consumer satisfaction with various dimensions of provider quality. Still, to obtain reliable estimates of consumer perceptions of health service quality, household based sampling appears to be far superior to the simpler exit survey method.



**Keywords**: Health services quality, patient satisfaction, survey methods, Madagascar

## 1. INTRODUCTION

Ensuring the quality of health care services has in recent years become a major concern of developing country policy and research. This is the result of a recognition that severe resource constraints and institutional factors (in particular, poor incentive structures) often result in public health care of very low quality, which in turn reduces both utilization of and the benefits from these services (World Bank, 2004). These developments in turn have motivated an expansion of efforts to measure and monitor health service quality via surveys of health care providers and their clients (Lindelow and Wagstaff, 2003).

Among these tools, client surveys are intended to measure user satisfaction with, or perceptions of, services or specific aspects of services. They potentially provide an important means of judging health sector performance and whether new policies are having their desired effects. In addition to measuring levels of client satisfaction, client surveys potentially also can identify facility attributes or practices that increase satisfaction, hence the willingness to use a service (and ultimately, health outcomes). When carried out, as they almost always are, in the form of user exit surveys at the point of service, they are simple and inexpensive to administer: one merely has to sample health facilities and develop simple protocols for sampling and interviewing their patients.

However, while the advocacy and use of client satisfaction surveys has grown substantially in both developed and developing country research (MEASURE, 2001), there have long been concerns about the usefulness of the responses in such surveys as a means of understanding either client satisfaction or service quality (Sitzia and Wood, 1997; Avis *et al.,* 1997). It is typical for user exit surveys to show uniformly high satisfaction with services (Carr-Hill, 1992; Lindelow and Wagstaff, 2003; Bitran, 1994). A likely source of this outcome is thought to be 'courtesy bias' whereby respondents are reluctant to express negative opinions to a stranger, leading to overestimation of satisfaction. The potential for courtesy bias is expected to be especially strong when respondents are interviewed at the facility right after they receive care; particularly among a poorly educated population, respondents are likely to associate the interviewer with the facility in some way and thus be particularly eager not to provide 'disappointing' unfavorable responses. Unexpectedly high levels of satisfaction in exit surveys may also reflect a 'Hawthorne effect' whereby care providers perform better when they know they are being observed or their patients are being interviewed. Another reason may simply be that clients judge services against low expectations formed from prior experiences (Like and Zyzanski, 1987; Brody *et al.,* 1989).

It is possible that more accurate measures of consumer perceptions can be generated by asking about specific aspects of facility quality or areas for improvement rather than asking general questions about overall satisfaction (Simmons and Elias, 1994; Bessinger and Bertrand, 2001). It has also been suggested that courtesy bias will be higher for more subjective questions, such as those asking respondents to rate their satisfaction with their interactions with facility personnel, than for questions pertaining to more objective attributes of facilities, such as cleanliness (Williams *et al.,* 2000).

There is a notable absence of quantitative analysis of these bias issues, which in large part reflects data limitations: it is difficult to assess bias in exit surveys if there is no comparison group of consumers who are not interviewed in such surveys.[1] Beyond these basic concerns with the reliability of client satisfaction surveys, there is a lack of studies of client satisfaction using regression techniques that are the standard tools of empirical economics. A search of the literature yielded only one multivariate analysis of the determinants of health care client satisfaction in a developing country (Agha *et al.,* 2003). The lack of attention by economists reflects a traditional aversion to analyzing the determinants of subjective outcomes in favor of revealed preferences (i.e., demand behavior). Lately, however, interest in analyzing different forms of subjective data has intensified (Hammermesh, 2004; Frey and Stuster, 2002).

This study estimates the determinants of satisfaction with, and perceptions about the quality of, the care provided in health centers in Madagascar. It uses unique data to investigate the reliability concerns with user exit surveys just noted and begins to rectify the paucity of multivariate analysis of client satisfaction. These data, collected for the Study on the Efficiency and Equity of Health Care in Madagascar (known by its French acronym EEEFS), contain detailed information on health care providers and households as well as on consumer attitudes about these providers. What is particularly uncommon is that comparable perceptions data are collected from user exit surveys, which are the standard means of collecting such information, and from population-based household surveys. Since in the latter individuals are interviewed at a remove in both space and time from their visits to center, their responses are expected to be unaffected, or significantly less affected, by courtesy bias. These visits also by and large would not have coincided with the presence of survey personnel at the centers, so Hawthorne effects on provider behavior are also not an issue for the household survey data. Hence the EEEFS provides an ideal basis for assessing the extent of bias in user exit surveys, as well as the usefulness of client satisfaction surveys more generally.

Therefore we address the following questions: *First*, is there evidence of courtesy bias (or Hawthorne effects) in exit survey data? *Second*, does such bias involve estimates only of levels of reported satisfaction or perceptions, or also estimates of the determinants of these outcomes? If the former, user exit surveys may still have some value for understanding the factors determining client satisfaction and perceptions, even if they produce misleading estimates of the mean levels of these outcomes. *Third*, is courtesy bias more pronounced for relatively subjective questions (for example, concerning the overall level of satisfaction with the service), and less important for questions about objectively describable provider attributes? If so, this would suggest that exit surveys might still provide reliable information for some types of questions.

*Fourth*, with regard to more general concerns over the usefulness of client satisfaction surveys, do the satisfaction and perception measures have strong 'validity', in

[1] One British study, however, found qualitative evidence of courtesy bias using follow-up surveys of clients initially interviewed in exit satisfaction surveys (Avis *et al.*, 1997).

the sense that they measure what they are supposed to measure? We assess this by determining if the responses to a given type of satisfaction question correlate strongly (and in the expected direction) with the observed health facility attributes that should be relevant to satisfaction along that dimension, while being unrelated to levels of 'irrelevant' attributes.[2]

Section 2 below describes in detail the methodology for analyzing client perceptions, followed in Section 3 by a discussion of the EEEFS data. The findings for client perceptions are presented in Section 4. Lastly, Section 5 summarizes the main findings and the methodological implications of this study. Among the more unambiguous conclusions drawn is that courtesy bias (or other sources of bias) in exit surveys is substantial. If researchers and policymakers want to know how consumers feel about health service quality, more elaborate household survey-based approaches appear necessary.

## 2. METHODOLOGICAL APPROACH

We estimate specifications of the following forms:

$$(1) \qquad S_{ijc} = \alpha X_i + \beta Q_{jc} + \delta SURVEY_i + u_{ijc}$$

$$(2) \qquad S_{ijc} = \alpha X_i + \beta Q_{jc} + \delta SURVEY_i + \zeta Q_{jc} * SURVEY_i + u_{ijc}$$

where the *i,j,* and *c* subscripts denote, respectively, the individual, the health care provider, and the community. The dependent variable $S_i$ is an index of satisfaction or quality perceptions; $X_i$ is a vector of individual characteristics of the patient and survey respondent; $Q_j$ is a vector of characteristics of provider *j*; and $SURVEY_i$ is an indicator of whether the individual was interviewed in the exit survey or the household survey (for which cases SURVEY takes the values of zero and 1, respectively). As discussed further in the next section, individuals were asked to rate their satisfaction or perceptions in terms of ordinal categories. Therefore our preferred model is ordinal probit.

The parameter $\delta$ captures the mean difference in reported satisfaction for household survey respondents relative to exit survey respondents. A significant negative value for this coefficient—i.e., more favorable ratings from those interviewed upon leaving the health centers—is taken to indicate the presence of courtesy bias, or alternatively, a Hawthorne

[2] Economists would refer to this concurrence of regression coefficient signs and significance levels with theoretical expectations as indicating "theoretical validity" of the measure. Psychologists and those in other disciplines who have traditionally analyzed satisfaction surveys would say that it establishes "construct validity." (See Sitzia, 1999).

effect whereby providers perform better when survey personnel are present at the facility. Note that the latter interpretation would be plausible only for outcomes related to practitioner or staff behavior that can be readily adjusted, such as actions taken during the consultation and the 'welcome' given to patients. Others, such as drug availability and the condition of equipment, are less amenable to short-term manipulation by providers. This distinction, it may be observed, by and large corresponds to that between subjective assessments (since many of these involve reflections on aspects of staff behavior) and more objective measures of facility characteristics.

The coefficient on the interaction term $Q_j$*$SURVEY_i$ in the second model indicates whether the effect of specific facility characteristics differ depending on the location of the interview. As noted, it is possible for courtesy bias to affect the level of satisfaction expressed at the point of service delivery, while estimates of the responses of satisfaction to specific factors (i.e., the slope coefficients) are not biased.[3]

We are also able to test whether, as some have suggested, courtesy bias is greater for more subjective questions. The EEEFS exit surveys and household surveys contain subjective (e.g, satisfaction with staff courtesy) as well as relatively objective (e.g., cleanliness of the facility) questions. If the hypothesis is correct, and assuming that courtesy bias, if present, is associated with user exit surveys but not (or less so) with household survey responses, the responses to more objective questions should not depend on whether the interview was conducted at home or outside the facility. For subjective questions the location of the interview should matter more.

Interpretation of the coefficients of eqs. (1) and (2) are potentially complicated by sample selection. Unlike courtesy bias, the problem of selectivity bias is quite often not acknowledged, let alone dealt with, in studies using client satisfaction surveys. Yet the potential for problems is obvious, since client satisfaction questions are necessarily restricted to those who have sought treatment. In our population-based household survey data, about half of those reporting an illness or injury in the previous two weeks did not seek formal care. One might expect those who do seek treatment, or seek it at a particular provider, to feel more strongly that it is worthwhile, hence to be more likely to express satisfaction than the population overall. In terms of the econometrics, if sample selection on unobservable individual or community level factors occurs, both the slopes and constant term parameters in the regressions may be biased.

One can correct for selectivity bias by modeling the process through which individuals select into the sample of health care users – and further, of users of specific provider categories – employing standard approaches of Heckman (1979) or Lee (1983). Identification in sample selection models requires the availability of variables that affect

[3] However, bias in slope coefficients will necessarily occur if overall levels of satisfaction are sufficiently upwardly biased. Categories of satisfaction are in effect 'top coded' ('satisfied' or 'very satisfied' usually being the highest possible response); if the bias pushes the vast majority of respondents to the highest levels, there will be little variation left, so estimates of the effects of characteristics on outcomes will be attenuated.

treatment choice but not satisfaction conditional on this choice. Such instruments could include, plausibly, the distances to different providers. In our data distance to various local providers was collected only in the household survey. In the user exit survey individuals were asked only how long they traveled to the provider at which they were interviewed, which is potentially endogenous to preferences for that provider.[4]

An alternative approach to controlling for selection bias is to use fixed effects methods. We cannot apply fixed effects at the level of the individual since individuals are not observed in multiple illness episodes or at multiple providers. However, community and provider fixed effects are feasible. In the former, Eqs. (1) and (2) are transformed by subtracting individual values from the cluster level means of the dependent and independent variables. The community-level component of the disturbance term $u_{ijc}$ is differenced out by this procedure. Hence selection bias caused by unmeasured community level factors that influence both the decision to seek medical care and satisfaction with the care received will be largely if not totally eliminated. One such factor, for example, may be the degree of local health awareness, reflecting the presence of NGOs or community health workers or access to mass media messages.

There are several important limitations of the community fixed effects approach. Estimation is restricted to communities with multiple surveyed providers to provide within cluster variation in provider characteristics, which means dropping eight (mostly rural) communities out of 80 total in which there was just one relevant local health facility. In addition, of course, the within cluster estimator can only deal with selectivity due to heterogeneity at the community level. It does not control for individual level heterogeneity that may affect the choice of provider within a community; it assumes that this choice is exogenous and uses differences across providers within a community to estimate the effects of provider characteristics.

We may turn then to provider level fixed effects, that is, we may subtract eqs. (1) and (2) from the provider level means. This will control for selection by individuals into the sample of clients for a given provider and thus may eliminate much of the potential bias associated with individual preferences. The significant disadvantage of within provider estimates is that they exclude the direct effects of provider characteristics, since any provider specific factors difference out. However, the effects of individual-varying covariates can be estimated. This includes factors such as schooling and wealth and—of particular interest for this study—the coefficient $\delta$ and $\zeta$, i.e., the effects of being interviewed survey at home relative to at the clinic, and the interaction of this indicator with provider characteristics (since there remains variation in both variables). Thus the provider fixed effects allows us to assess, with controls for both community and individual level

[4] This limitation is not inherent in the design of user surveys, obviously, as respondents could also be asked how far specific alternative care providers are from their homes.

selectivity, the presence of courtesy bias as a level effect and as an influence on slope coefficients.[5]

For each of our patient satisfaction/perception indicators, therefore, we present results from the ordered probit (level) models and community and provider fixed effects models. For the fixed effects estimators, we are not able to use the nonlinear ordered probit form. As is well known, when the number of observations per group (cluster or provider in this case) is small and the number of groups large, estimation of most nonlinear fixed effects models yields inconsistent parameter estimates due to the incidental parameters problem (see Wooldridge, 2002). Therefore to estimate the fixed effects models we must treat the ordinal satisfaction indicators as linear dependent variables.

This treatment is not necessarily innocuous, especially since we are dealing with measures of satisfaction or more generally, of well being. It imposes a cardinal interpretation on ordinal categories (e.g., the difference in satisfaction between stage 2 and 3 is the same as between 1 and 2), implying that it is possible to make cardinal interpersonal comparisons of satisfaction (see Ferrer-i-Carbonell and Frijters, 2004). However, Ferrer-i-Carbonell and Frijters show, in the context of individual level fixed effects, that this assumption is less problematic than assuming the absence of heterogeneity when such heterogeneity is present, so that linear fixed effects to control for heterogeneity (of which selection is one manifestation) are still more reliable.

Because of the linear treatment of the dependent variables, the fixed effects estimates are not directly comparable to those from the nonlinear ordered probits. Still, the results should indicate in qualitative fashion the importance of courtesy or other exit survey biases and the relative roles of different provider characteristics in the determination of consumer perceptions of quality. Finally, we note that the fixed effects models will control not only for selectivity but also for correlations of community or provider level unobervables with the included regressors. Without individual level fixed effects, however, we cannot control for remaining heterogeneity at the individual level in the sense of unobservable factors that correlate with both regressors and outcomes.[6]

## 3. DATA AND VARIABLES

The *Etude sur l'Efficience et Equite des Formation Sanitaires a Madagascar* (EEEFS) was designed to address a range of questions about the performance and equity of the Malagasy

[5] With regard to the use of the term 'provider fixed effects', it should be made clear that the method controls not only for some provider fixed effect but also for common unobservable factors among clients of that provider.

[6] For the same reason, we cannot deal with potential biases in the estimates from the measurement error problem noted by Bertrand and Mullainathan (2001) whereby the propensity for misreporting one's true feelings (hence the degree of measurement error in the dependent variable) is causally determined by education or other explanatory variables.

health care system. The survey was carried out in 2003 by the Word Bank in partnership with the Malagasy statistics institute (INSTAT) (See World Bank/INSTAT, 2005).[7] The EEEFS used the sampling frame of a representative national household survey from earlier in the same year: 80 of 303 clusters (corresponding generally to the *fokontany* administrative level or more loosely, a 'community') in that survey were randomly chosen to be resurveyed, and the original sample households in these clusters were revisited. The EEEFS comprised several integrated survey instruments: a household survey (adding additional modules on health care to the earlier household survey); a user exit survey; detailed surveys of selected public and private health facilities; and a health district survey. The user exit surveys were implemented in part to permit oversampling users of local health centers, since it was recognized that the numbers in the household survey reporting recent health care visits would be fairly small.

The user survey and household survey health care module used essentially identical formats for key blocks of questions. This includes, notably, questions about the care they received and their perceptions of the facility, as well as questions about patients and their households. With a few exceptions 10 patients were interviewed in each cluster for the exit survey. In most clusters one public and one private provider each was surveyed, with seven clients interviewed at the public facility and three at the private facility. This division reflects the overall population shares of public and formal private health care users as observed in prior representative household surveys. Through this procedure, the sample of public and private interviews is thus approximately self-weighted.[8] However, the small numbers of private patients precludes separate analysis on the public and private client samples.

The exit surveys were implemented at health care providers that had also been chosen to receive facility surveys, administered by a medical professional. The main local public health provider for the community was always surveyed, whether a clinic (*Centre de santé de base*/CSB) or (much less often) a hospital (*Centre hospitalier de district*/CHD). In addition, the questionnaire was administered to the most important private provider used by local residents, if there was one, and in several occasions, to a second public provider. All in all, 153 facilities in the 80 clusters were interviewed. In addition to detailed information on personnel, equipment, supplies and medicines, physical condition, and absenteeism and incentives, the visits to these facilities also included direct observations of practitioners with their patients to assess compliance with medical protocols for care. Via identification codes for providers, the facility information was linked to information in the user surveys and the household surveys. The sample of clients of local surveyed health facilities

[7] The project was directed by Mead Over of the World Bank, in collaboration with Waly Wane (World Bank) and Mamisoa Razakamanantsoa (INSTAT).

[8] Still, the fact that the same number of clients (10) were interviewed in every cluster means that there was oversampling of health care users in communities where the total number of patients in the population is lower than average (because the population is smaller or rates of utilization lower) and undersampling where the number is high. We use the sampling weights for the same clusters from the original household survey with adjustments to compensate for this over or under sampling of clients from the user surveys. Ultimately, therefore, the weighted sample provides a by and large nationally representative sample of health care users.

consists of 1072 individuals; 790 from the user exit surveys and 282 from the household survey.

*Dependent variables*

Users of health care services were asked to rate their satisfaction with the 'welcome' provided (i.e., courtesy and respect shown) by the staff, with the consultation itself, and with the quantity and quality of medicines provided. Respondents were also asked to indicate their level of satisfaction with the quality of the service relative to the costs they had to pay. For each of the foregoing questions respondents were provided with a series of ordinal categories from which to choose, typically 'very satisfactory', 'satisfactory', 'mediocre', and 'unsatisfactory'. Other questions had respondents rate the exterior and interior appearance of the center in terms of cleanliness as well as to rank the availability and condition of medical equipment (See Table 1). For each of the outcomes used in this study, examination of the frequencies of the responses suggested that the number of categories be reduced to three, usually by aggregating the two lowest responses.

*Independent variables*

Since perceptions of quality may be influenced by personal characteristics such as education and wealth, all the models include a common set of such controls. To represent household resources, we employ a wealth index created by factor analysis using information on housing and other durable assets, following the methodology of Sahn and Stifel (2003). Also among the controls are dummies for the illness or injury that motivated the consultation.[9] Note that the questions on assets and other characteristics were identical in the user and household surveys.

With respect to health provider characteristics, the influential framework of Donabedian (1980) distinguishes among three kinds of indicators of quality: 'structure', referring to characteristics such as facility, equipment and staff; 'process', referring to the care practices of the practitioners in the facility; and 'outcomes' referring to the health outcomes resulting from the care provided. The EEEFS facility survey provides detailed information on structural characteristics. Since many of these indicators are highly correlated, some form of data reduction is necessary for regression analysis. We construct indicators following the approach used in most previous research (e.g., Peabody *et al.,* 1994). For an indicator of facility infrastructure we take the mean of 0-1 indicators for several specific attributes (including, among others, having a pump or faucet, electricity, and a working refrigerator). Similarly, for cleanliness or appearance we calculate the mean of binary indicators for dirtiness, humidity damage, and decay of walls, floors and ceilings, and for insects and condition of toilet facilities.

[9] While we thus distinguish by broad class of complaint (respiratory, diarrhea, injury, etc) we are not able to use information on the severity of the illness.

For medicine availability, we create a variable for the share of six basic medicines (antibiotics, aspirin, etc.) for which the facility experienced supply shortages in the last 90 days. Price is represented by the median of the consultation cost for the facility reported by household and user survey clients. Several other characteristics, such as the number of beds in the facility and hours per week of operation, are entered separately. Appendix Table 1 provides descriptions and means of these variables.

With respect to process indicators, as noted above, the facility survey included direct clinical observations (DCO) of practitioners with patients indicating, for example, whether the practitioner detected symptoms and made the correct diagnoses. We do not use these data for the main part of our analysis for several reasons. First, the facility survey collected this information only at clinics, not hospitals. Second, to insure comparability across providers, direct clinical observations were restricted to prenatal care and treatment of young children, whereas the data on perceptions from both exit survey and household survey sources encompasses the full spectrum of patients. Instead we use information on process from the user and household surveys. Clients were asked who examined them (doctor, nurse, or other staff); if they were given a physical examination (*palpé*); if they were provided instructions on their medication (duration of treatment, frequency and dosage); and if the practitioner discussed several health subjects with them (nutrition, family planning, HIV/AIDS, malaria, and vaccinations).[10] These are objective measures that are likely to be well understood by respondents, so the risk of measurement error or biased reporting seems slight.[11]

Table 1 shows the distribution of responses to each of our satisfaction and perception indicators. Levels of satisfaction—with the consultation, the service relative to cost, and staff courtesy—are generally but not universally high. For example, 83% of respondents they are 'satisfied' or 'very satisfied' with the consultation. On the other hand, less than 60% say they are satisfied or very satisfied with the service relative to its cost.

In Table 2 we consider the means of dependent and independent variables by different stratifications of the data. Considering first the division into public and private providers, it should be noted that while public sector providers are dominated by basic care centers (*Centres de sante de base*, accounting for 90% of public clients), the sample of private care providers is much more varied. The largest categories are confessional or church organized centers and private doctors (32% and 30% of private care users, respectively). Smaller numbers go to private clinics, employer-run clinics, informal practitioners or traditional healers, and NGOs.[12] Despite this somewhat eclectic mix, satisfaction and quality perceptions indicators are higher across the board for private

[10] While using these data enable us to use a larger sample of patients (and facilities), we are not standardizing on a specific type of patient or complaint, unlike the DCO for prenatal and child care,. However, this is dealt with in the estimation by the inclusion of controls for nature of illness/injury and sex and age of the patient.
[11] Note that recall error should not be a major issue: user exit surveys take place right after the visit, while for the household surveys, the information on health care usage pertains to visits made in the last two weeks only.
[12] These are weighted shares of the client sample (using the weights described in Fn. 8) and thus represent the population shares of clients by provider type.

providers. While this may reflect in part self-selection of individuals into private and public sectors based on preferences or perceptions of quality, the structural indicators as recorded by survey personnel in the facility surveys (appearance, infrastructure, availability of medicines, etc.) are also generally superior for private facilities, and a larger share of private patients received a physical examination.

The table also distinguishes the household and user exit survey samples. With regard to individual characteristics, in most respects the samples are alike.[13] It is noteworthy that mean reported satisfaction is higher in the exit survey group, especially with respect to assessments of value relative to cost and availability of medicines. This suggests the possibility of courtesy or Hawthorne effect biases, which we now examine in a multivariate framework.

## 4. DETERMINANTS OF CLIENT SATISFACTION AND PERCEPTIONS

Tables 3 to 7 present regression results for the following indicators: evaluation of the consultation, of the service relative to cost, of the attitude of the staff, of quantity of medicines provided, and of condition of the facility interior. To save space we do not show the results for exterior facility condition and condition of equipment, which were very similar to those for interior condition. For each outcome we present five models, starting with an ordered probit including a set of individual level controls plus the set of facility characteristics that *a priori* should be directly relevant to the outcome at hand. In the second ordered probit model we add the full series of facility indicators. The third column shows the community fixed effects regression using the same right hand side variables as the previous model. The final two columns show the provider fixed effects models: a basic specification including individual level covariates (including the household survey dummy and the provider process variables), and an expanded model including interaction of the survey indicator with key facility indicators as just defined.

*Survey effect*

The effect of being interviewed in the household survey (relative to the exit survey) is shown in the first row of each table. As seen in Tables 3 though 5 there are strongly significant negative effects of a household survey interview on satisfaction with the consultation, with the service relative to its cost, and with the attitude of the staff. These effects are robust to controlling for community and provider level fixed effects (cols. 3-5)[14]. For satisfaction with availability of medicines (Table 6), there is a consistently significant

[13] More respondents in the exit surveys report unclassified complaints or not having an illness, in part because a larger share of individuals in the user exit surveys report coming in for preventative or routine care (e.g., prenatal care). Restricting the analysis to individuals in both surveys reporting an illness or injury did not alter the results.

[14] Meaning, robust in a qualitative sense; recall from Section 2 that the fixed effects models are based on a linear approximation of the ordinal dependent variable so are not comparable in terms of magnitudes to the ordered probit estimates.

negative household survey effect in the ordered probits but not in the fixed effects models. Hence for relatively 'subjective' indicators (all of the above except possibly medicine availability), there is strong evidence of courtesy bias, or possibly Hawthorne effects, in user exit interviews: respondents in such surveys are more likely to report satisfaction than those interviewed at home.[15]

Moreover, these differences are often large. For example, calculations based on the estimates[16] indicate that controlling for respondent and provider characteristics, being questioned in an exit interview raises the probability of reporting being 'very satisfied' with the consultation by about 15 percentage points relative to being interviewed at home. This is a substantial effect considering that the mean probability of this outcome is 28 percent. For satisfaction with staff courtesy, the analogous effect is about 12 percentage points.[17]

In contrast, there is no survey effect on perceptions of the condition of the interior environment of the center (shown Table 7) and of the external environment and condition of equipment, in any of the specifications used. As noted, perceptions of structural aspects of quality such as these are probably less subjective than assessments of overall satisfaction and satisfaction with process-related characteristics involving staff behavior and attitudes. The lack of differences in household and exit survey responses for the former accords with the hypothesis that courtesy bias (and Hawthorne effect bias) will more strongly affect estimates of highly subjective indicators.

*Provider characteristics*

The ordered probit results in Table 3 indicate that satisfaction with the consultation is higher when the respondent saw a doctor or nurse and if he or she was given a physical examination. The physical exam effect but not the doctor/nurse effect is robust to controls for community and provider fixed effects (with respect to the latter, recall that these variables, unlike other quality measures, vary within provider). It should be noted that the great majority of patients—about 90%—were seen by a doctor or nurse, so the differencing may serve to eliminate the variation needed to estimate this impact. An additional process indicator, for whether the practitioner provided detailed instructions on medications, was only available in the user exit survey. In estimations on the user survey sample (not presented) this measure had a positive and marginally significant (t=1.62) effect on

[15] More precisely, the results indicate a greater bias in user surveys, since responses on the household surveys may also suffer to some extent from courtesy bias.

[16] Using the estimates in column 2 and calculated as the difference in probability of reporting being 'very satisfied' and reporting being either just 'satisfied' or less than satisfied, evaluated at the means of the regressors.

[17] There is one other explanation for more negative subjective assessments in household interviews: these clients have had some time to reflect on their treatment, including whether it has improved their condition. Whether or not this makes their quality assessments more 'accurate' (not a useful concept for subjective responses in any case), it should make them more meaningful for understanding health care choices; such choices undoubtedly are based more on *ex post* assessments than on first impressions formed at the time of treatment and captured in exit interviews.

satisfaction with the consultation. On the other hand, there is no effect of provision of nutrition and other health information. This was the case even if the sample was restricted to cases where the respondents were women, or women bringing children in for care, for whom the type of information referred to in the question would be most relevant.

The process-related variables just discussed, which are reported by patients, necessarily differ from the indicators recorded in direct clinical observations of practitioners with patients made by medically trained survey personnel. The latter focus on whether the practitioner asked the appropriate questions during the consultation, made the correct diagnoses, and prescribed the appropriate treatment. To what extent do the observable measures that clients value (according to our regressions) correspond to the measures in the DCO, which presumably more closely reflect actual process quality?

Our ability to answer this question is somewhat limited because, as noted earlier, the DCO involved only consultations for children under 5 years old (as well as visits for prenatal care), and also excluded hospitals. Hence this exercise is limited to *Centres Sante de Base* as well as private care providers. For each child seen by the practitioner, the doctor-observer for the DCO module recorded the presenting symptoms and whether the practitioner made the correct diagnosis and prescribed the appropriate treatment. We calculated for each child the share of illness conditions (there was usually only one) correctly diagnosed and the share of correct treatments prescribed. To derive individual provider scores for diagnostic and treatment ability that control for potential differences across providers in the types of ailments seen, we regressed these diagnosis and treatment 'share correct' variables on dummies for presenting symptoms (as recorded by the observer) in provider random effects regressions.[18] The random effects yielded by these regressions are measures of each provider's propensity or 'ability' to correctly diagnose and treat patients. Next, on the exit and household survey client data, we regressed each of the client-reported process variables on the provider average of the two random effects (which were highly correlated) along with controls for client characteristics including age, sex, and nature of the illness. The models were run on the full client sample as well as on the sample of patients under 5 years old to enhance comparability with the DCO information.

The results for both samples, shown in Appendix Table 2, indicate a significant association of the DCO indicator of provider diagnostic and treatment skill with just one of the three respondent-reported process variables, health information provision. This is the one such variable that had no apparent impact on client satisfaction. The variables that do increase satisfaction (physical exam, explain medications) are not statistically associated with the DCO indicator of good practice. These findings should be regarded as somewhat tentative, given that in some cases the interviewed clients of a given center would not have been seen by the same individual observed for the DCO (though we would expect some

[18] That is, regressions allowing for correlations of errors among children seen by the same provider (typically five in number). The random effects regressions also include the order in which the patient was seen, since practitioner efforts to 'be on their best behavior' may wane over the period of observation. (See Leonard and Masatu, 2005).

correlation of practices among health professionals in the same center). Still, the pattern underscores the fact that aspects of medical practice that patients can recognize and approve of may not correspond closely to actual levels of process quality, which involves behaviors or ability (e.g., making a correct diagnosis) that are difficult or impossible for patients to judge.[19] This, of course, is an aspect of the asymmetrical information problem inherent in health care markets. It raises the possibility that providers interested in increasing satisfaction with, hence demand for, their services, might achieve this by altering their behavior in ways that do not involve significant improvements in actual process quality, especially if such improvements would be less easily recognized by consumers.

We turn now to the other satisfaction or perception outcomes in Tables 4-7. Satisfaction with service quality relative to cost is negatively affected by the level of the consultation cost in the ordered probit models (Table 4, cols 1 and 2) and is generally positively affected by the process variables (physical examination, doctor/nurse). The cost effect is not robust to adding community fixed effects controls, however (col. 3). Satisfaction relative to cost is also positively affected by the hours per day that the facility is open, though the effect is imprecisely estimated in the community fixed effects model.

Satisfaction with staff courtesy (Table 5) is positively affected by having been given a physical examination and having seen a doctor or nurse (cols. 1 and 2). The former effect remains strongly significant in the community as well as provider fixed effects models. The strong impact of having a physical exam is sensible in that patients are likely to view a longer or more thorough visit with a practitioner as an aspect of respectful treatment.[20]

Turning to perceptions of structural aspects of quality, we consider first drug availability. A higher share of basic drugs in short supply in the last 90 days reduces patient assessment of drug availability in the ordered probits (Table 6, cols. 1 and 2). In the community effects model the estimate turns positive, which may reflect strong collinearity with the other facility regressors in the differenced model; in a community fixed effects regression including only the process measures and the drug shortage indicator the coefficient on the latter is negative and quite large though imprecisely estimated. Our shortage measure, which combines information on supplies of various medicines over a three-month period, is obviously only an approximate representation of whether the individual is able to obtain the specific drug for his or her condition at the time of the visit.

For perceptions of the interior and exterior of the facility and of the condition of equipment, the ordered probit results are very similar. Each of these outcomes is positively

[19] Indeed, models entering the provider ability random effects directly into the satisfaction regressions, with or without the inclusion of the respondent-observed process variables, yielded insignificant coefficients.

[20] It might be expected that one attraction of private facilities is a more courteous staff, since private staff behavior is presumably more motivated by profit or wage incentives than is the case in public facilities. However, a dummy for private provider had no significant independent effect on satisfaction with welcome, controlling for the process variables in our data. Elsewhere, the private sector dummy had significant positive effects on satisfaction with consultation and with medicine availability but did not affect the other outcomes examined.

and highly significantly associated with values of the appearance index and the infrastructure index created from the facility survey (see Table 7 cols. 1 and 2 for the results for interior condition). For all three outcomes the appearance index, which is most relevant for these questions, is robust to controlling for community fixed effects.

The foregoing results suggest that client responses accord with specific measures of quality that *a priori* should be relevant to the questions asked: process indicators for satisfaction with the consultation itself, facility condition indicators for evaluations of cleanliness of facility and condition of equipment, etc. However, our criteria for validity should be stronger that this: the responses should in addition *not* be strongly correlated with other aspects of quality that, based on a reasonable theoretical framework, are not relevant to the question at hand.[21] It is possible, for example, that consumers respond to questions about different aspects of services based only on their general impressions of provider quality. This obviously would reduce the usefulness of specific questions about a range of characteristics.

Examination of the full set of results suggests, however, that this is not the case. For example, having a physical exam matters for satisfaction with the consultation and drug shortages reduce satisfaction with medicine quantity, but these perceptions by and large are not affected by the physical condition (appearance) of the facility (see cols. 2 and 3 of Table 3.) Individuals' assessments of the courtesy shown by staff are not influenced by facility appearance and the quality of infrastructure. The drug shortage indicator and consultation cost are generally insignificant in regressions other than for satisfaction with drug availability and service relative to cost, respectively. There are exceptions to this general pattern as well as a handful of improbably signed coefficients. For example, the number of beds, which is related to the size and capacity of the center, is negatively associated with several satisfaction measures; possibly this captures negative experiences in larger or more impersonal facilities. However, overall it seems that the responses on satisfaction or perception questions reflect consideration of the appropriate set of service characteristics.

*Individual characteristics*

There are very few cases in Tables 3-7 where characteristics of the patient or respondent such as age, gender, or wealth show significant associations with reported satisfaction or perceptions. A lack of wealth or education effects on satisfaction is perhaps surprising given many earlier findings of negative effects at least for education (though largely based on U.S. studies, see Hall and Dornan, 1990). We might have expected the better educated or well off to be more discriminating judges of service quality. In a developing country population with very low average education, however, it is possible that there is a large positive incremental effect of schooling on one's understanding of the value of formal health care, counteracting any tendency for negative effects on satisfaction.

[21] This idea corresponds loosely to the concept of divergent validity (See Sitzia, 1999).

*Interactions of survey and provider characteristics*

The second provider fixed effects model (last column of each table) presents our preferred specification for assessing whether the effect of provider characteristics on reported satisfaction depends on whether the respondent was interviewed at the health center or at home. Few of the interactions with the household survey dummy approach statistical significance. For satisfaction with staff courtesy, the positive effect of having been given a physical examination is larger for household survey respondents than for those interviewed at the facilities (t=1.64). For satisfaction with drug availability, the negative effect of drug shortages is smaller for household survey respondents (t=1.64). Therefore while there is strong evidence of courtesy bias—or possibly, Hawthorne effects—in exit surveys with respect to *levels* of satisfaction (captured by survey-specific intercept terms), there is less evidence that the effects of various aspects of provider quality (i.e., the slope coefficients) suffer from bias.

## 5. DISCUSSION

The unusual data from Madagascar used for this analysis suggest that client satisfaction responses in user exit surveys are positively biased, sometimes substantially so. That this is not the case for all questions—in particular, the bias does not appear for perceptions of the quality of facility infrastructure—might be considered a somewhat positive finding; perhaps exit surveys just need to be confined to certain kinds of questions. However, the questions that seem to be answered with the least bias in exit surveys are those for which user surveys are probably the least useful. If consumer perceptions in exit surveys accord closely with observable structural factors such as facility condition, questions on these perceptions would not add to what could be learned from direct observation using facility surveys. What is more uniquely available from clients are their subjective perceptions of process—the behavior of practitioners, the attitude of staff—and their overall satisfaction with service quality. But these unavoidably subjective responses are more strongly subject to biases in exit surveys.

This concern aside, the user satisfaction data in the present case are informative. They are able to differentiate satisfied from less satisfied consumers and illuminate the service characteristics that increase user satisfaction or perceptions of quality. Regression results show responses to be affected by some factors and not by others in accordance with expectations for the question being asked. Among the more notable findings are the importance for satisfaction with the consultation of process variables such as being given a physical examination and receiving adequate instructions on medicines; such factors also seem to affect clients' perceptions of how respectfully they were treated at the center. Consultation cost in some if not all specifications has a significant negative impact on the perceived value of the service relative its cost. A possibly surprising finding is that the characteristics of the patient or of his or her household seem to matter very little for reported satisfaction.

We find as well that the patient-observed aspects of process quality that increase satisfaction with the consultation are uncorrelated with measures of provider skills (with respect to diagnoses and treatment) obtained from direct clinical observation. This underscores the fairly obvious point that client evaluations of process are no substitute for expert evaluations of medical practice. It suggests further that providers interested in increasing consumer satisfaction, hence demand for their services, potentially would try to achieve this objective by altering their behavior in ways that do not improve actual process quality.

From a practical point of view, the results raise a dilemma for researchers and policymakers wishing to use client perceptions surveys as a means of monitoring and improving the delivery of health services. This analysis indicates that household survey data with expanded questions on health care are a more reliable source of this information than user exit surveys (though the findings also suggest that the latter may still provide reliable estimates of the determinants, if not levels, of satisfaction). In addition to not suffering, or suffering less, from positive courtesy bias, household survey data also permit the modeling the decision to seek care and the choice of provider, which is a necessary input into standard selectivity correction techniques.[22] However, population-based household surveys are significantly more costly and logistically difficult to carry out than exit interviews. An additional problem is that standard sample sizes per cluster in random population surveys are usually too small to capture adequate numbers of individuals who have visited local health care providers within a short reference period. A way to deal with the second problem though not the first would be to oversample users of health care services when carrying out a population-based survey. That is, procedures at the enumeration stage could insure a minimum number of such households to be selected for interview, combined with random sampling of the cluster overall to insure representation of non-using households and to provide sample weights for reweighting in the analysis.

[22] With appropriate reweighting, choice based samples (i.e., samples of users) can be used to estimate choice among providers. But this necessarily excludes the choice of no care, since only users are captured in such surveys, and in any case, population-based data would be needed to calculate the appropriate weights for each option for the choice-based sample estimations.

## REFERENCES


Agha, S., Mehryar Karim, A., Balal, A. Sosler, S. A quasi-experimental study to assess the performance of a reproductive health franchise in Nepal. Commercial Market Strategies Country Research Series No. 14, 2003.

Avis, M., Bond, M., Arthur, A. Questioning patient satisfaction: an empirical investigation in two outpatient clinics. *Social Science and Medicine* 1997; **44**(1): 85-92.

Bertrand, M., Mullainathan, S. Do people mean what they say? Implications for subjective survey data. MIT Economics Working Paper No. 01-04, 2001. Available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=260131.

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

Bitran, R. Efficiency and quality in the public and private sectors in Senegal. *Health Policy and Planning* 1994; **10**(3): 271-283.

Brody D.S., Miller, S.M., Lerman, C.E., Smith, D.G., Lazaro, C.G., Blum, M.J. The relationship between patients' satisfaction with their physicians and perceptions about interventions they desired and received. *Medical Care* 1989; **27**(11): 1027-1035.

Carr-Hill, R.A. The measurement of patient satisfaction. *Journal of Public Health Medicine* 1992; **14**(3): 236-249.

Donabedian, A. *Explorations in Quality Assessment and Monitoring, Volume I: The Definition of Quality and Approaches to Its Assessment*. Health Administration Press: Ann Arbor, MI, 1980.

Ferrer-i-Carbonell, A., Frijters, P. How important is methodology for the estimates of the determinants of happiness? *The Economic Journal* 2004; **114**(497): 641-659.

Frey, B., Stuster, A. What can economists learn from happiness research? *Journal of Economic Literature* 2002; **40**(2): 402-435.

Hall, J.A., Dornan, M.C. Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. *Social Science and Medicine* 1990; **30**(7): 811-818.

Hamermesh, D.S. Subjective outcomes in economics. *Southern Economic Journal* 2004; **71**(1): 2-11.

Heckman, J. Sample selection bias as specification error. *Econometrica* 1979; **47**(1): 153-61.

Lee, L.F. Generalized econometric models with selectivity. *Econometrica* 1983; **51**(2): 507-512.

Leonard K.L., Masatu, M.C. The use of direct clinician observation and vignettes for health services quality evaluation in developing countries. *Social Science and Medicine* 2005; **61**(9): 1944-51.

Like, R., Zyzanski, S.J. Patient satisfaction with the clinical encounter: social psychological determinants. *Social Science and Medicine* 1987; **24**(4): 351–357.

Lindelow, M., Wagstaff, A. Health facility surveys: an introduction. Policy Research Working Paper 2953, January 2003. Online at http://www-wds.worldbank.org/

MEASURE Evaluation. Quick investigation of quality (QIQ). In *A User's Guide for Monitoring Quality of Care in Family Planning.* MEASURE Evaluation Manual Series, No.2. Measure Evaluation, Carolina Population Center, University of North Carolina at Chapel Hill, 2001.

Peabody, J.W., Rahman, O., Fox, K., Gertler, P. Quality of care in public and private primary health care facilities: structural comparisons in Jamaica. *Bulletin of PAHO* 1994; **28**(2): 122-141.

Sahn, D, Stifel, D. Exploring alternative measures of welfare in the absence of expenditure data. *Review of Income and Wealth* 2003; **49**(4): 463-489.

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

Sitzia, J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *International Journal for Quality in Health Care* 1999; **11**(4): 319-328.

Sitzia, J., Wood, N. Patient satisfaction: a review of issues and concepts. *Social Science and Medicine* 1997; **45**(12): 1829-1843.

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

Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA, 2002.

World Bank Development Research Group. *Making Services Work for Poor People: A Research Program on Public Services*. World Bank /Development Research Group: Washington, DC, 2004.

World Bank/INSTAT (Intitut National de la Statistique Madagascar). L'efficience et l'equite aupres des formations sanitaire Malgache : Resultats d'un enquete. Region Afrique Department du developpment humain**,** 2005.

**Table 1 - Distribution of responses to satisfaction and perception questions (percent)**

| | *Satisfaction with:* | | | | *Perception of:* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Consultation | Quality of service relative to cost | 'Welcome' (courtesy of staff) | Availability of drugs | Exterior cleanliness | | Interior cleanliness | | Condition of equipment | |
| Very satisfactory | 28.8 | 19.7 | 31.0 | 19.2 | Clean | 47.3 | Sparkling clean | 9.8 | Good | 28.9 |
| Satisfactory | 53.9 | 40.6 | 55.8 | 40.2 | Medium | 33.8 | Clean | 56.4 | Medium | 51.1 |
| Medium\mediocre\ unsatisfactory | 17.3 | 39.7 | 13.2 | 40.6 | Mediocre\ dirty | 19.0 | Medium\mediocre\ dirty | 33.9 | Medicore\ poor | 20.0 |
| No. of observations | 1032 | 794 | 1029 | 702 | | 1021 | | 1026 | | 947 |

**Table 2 - Health care clients: Variable means by survey and facility type**

| Variable | Household survey | Exit survey | Private facility | Public facility |
|---|---|---|---|---|
| *Dependent variables* [a] | | | | |
| Satisfaction with consultation | 1.93 | 2.21 | 2.31 | 2.06 |
| Satisfaction with service relative to cost | 1.53 | 1.94 | 2.00 | 1.73 |
| Satisfaction with welcome (staff courtesy) | 2.03 | 2.25 | 2.42 | 2.11 |
| Satisfaction with drug availability | 1.56 | 1.96 | 2.15 | 1.63 |
| Perception of interior facility condition | 1.64 | 1.82 | 2.02 | 1.68 |
| Perception of exterior facility condition | 2.11 | 2.38 | 2.64 | 2.19 |
| Perception of condition of equipment | 1.93 | 2.17 | 2.51 | 1.97 |
| *Independent variables* | | | | |
| Seen by doctor/nurse | 0.76 | 0.92 | 0.93 | 0.85 |
| Given physical exam | 0.74 | 0.77 | 0.91 | 0.72 |
| Given health information | 0.57 | 0.58 | 0.48 | 0.60 |
| Medicines explained | -- | 0.86 | 0.95 | 0.83 |
| Consultation cost | 0.03 | 0.08 | 0.29 | 0.01 |
| Appearance index | 0.84 | 0.91 | 0.94 | 0.87 |
| Infrastructure index | 0.46 | 0.53 | 0.60 | 0.48 |
| Share drugs short | 0.43 | 0.31 | 0.16 | 0.40 |
| Number of doctors/nurses | 2.68 | 4.10 | 2.99 | 3.76 |
| Share doctor/nurses absent | 0.10 | 0.13 | 0.06 | 0.14 |
| Number of beds | 4.54 | 4.89 | 1.16 | 5.76 |
| Hours open per day | 8.12 | 8.31 | 8.10 | 8.29 |
| Respondent years of school[b] | 3.39 | 4.51 | 5.05 | 3.85 |
| Respondent age[b] | 41.37 | 32.73 | 34.37 | 36.17 |
| Male patient | 0.44 | 0.44 | 0.41 | 0.45 |
| Patient age | 22.57 | 23.04 | 23.93 | 22.58 |
| Asset index | -0.41 | -0.06 | 0.09 | -0.26 |
| *Complaint*: Infection | 0.12 | 0.10 | 0.11 | 0.11 |
| Malaria | 0.32 | 0.29 | 0.29 | 0.30 |
| Diarrhea | 0.16 | 0.13 | 0.11 | 0.15 |
| Injury | 0.05 | 0.07 | 0.07 | 0.06 |
| Other illness | 0.29 | 0.33 | 0.38 | 0.30 |
| Other reason for visit | 0.06 | 0.09 | 0.05 | 0.09 |
| Number of observations | 262 | 770 | 286 | 746 |

Notes:

[a] Shows mean values of satisfaction/perception indices. Each index takes values of 1 (lowest) to 3 (highest). See Table 1 for exact definitions of these responses.

[b] In the exit surveys, where respondent and patient differ this is usually because the patient is a child. For the household survey, reflecting standard instructions to interviewers, it is assumed that individuals over 15 responded for themselves to questions about their health care consultations. For children 15 or younger the household head is assumed to be the respondent.

The sample of facilities is the same for household and exit survey samples. The mean values of facility characteristics differ because the distributions of the household and exit survey respondents across these facilities are not the same.

## Table 3 - Determinants of client satisfaction with consultation

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ordered probit: key facility characteristics | Ordered probit: full facility characteristics | Community fixed effects: full facility characteristics | Provider fixed effects | Provider fixed effects: interactions with household survey |
| Household survey interview [a] | -0.518 | -0.526 | -0.237 | -0.232 | -0.192 |
| | (3.11)*** | (3.09)*** | (3.89)*** | (3.79)*** | (1.23) |
| Rural | -0.204 | -0.204 | | | |
| | (0.96) | (0.92) | | | |
| Respondent years of school | -0.016 | -0.018 | -0.010 | -0.011 | -0.012 |
| | (0.93) | (1.02) | (1.18) | (1.33) | (1.38) |
| Asset index | -0.133 | -0.151 | 0.020 | 0.013 | 0.013 |
| | (1.54) | (1.80)* | (0.34) | (0.21) | (0.22) |
| Male patient | 0.049 | 0.041 | 0.017 | 0.013 | 0.012 |
| | (0.46) | (0.39) | (0.32) | (0.25) | (0.24) |
| Patient age <5 | 0.169 | 0.179 | 0.002 | 0.004 | 0.007 |
| | (1.25) | (1.33) | (0.04) | (0.06) | (0.12) |
| Patient age 5-15 | -0.079 | -0.086 | -0.034 | -0.036 | -0.033 |
| | (0.51) | (0.58) | (0.36) | (0.37) | (0.34) |
| Patient age > 60 | 0.343 | 0.377 | -0.045 | -0.047 | -0.054 |
| | (1.85)* | (1.97)** | (0.45) | (0.45) | (0.51) |
| Seen by doctor/nurse | 0.460 | 0.470 | 0.198 | 0.160 | 0.240 |
| | (2.06)** | (1.82)* | (1.21) | (1.14) | (2.07)** |
| Given physical exam | 0.644 | 0.619 | 0.294 | 0.290 | 0.253 |
| | (4.09)*** | (3.76)*** | (4.24)*** | (4.06)*** | (3.20)*** |
| Given health information | -0.015 | 0.020 | 0.034 | 0.030 | 0.029 |
| | (0.09) | (0.12) | (0.34) | (0.30) | (0.29) |
| Consultation cost | | -0.108 | 0.239 | | |
| | | (0.36) | (0.96) | | |
| Appearance index | | -0.201 | -0.167 | | |
| | | (0.24) | (0.14) | | |
| Infrastructure index | | 0.361 | -0.152 | | |
| | | (0.88) | (0.33) | | |
| Share drugs short | | -0.070 | -0.223 | | |
| | | (0.26) | (1.07) | | |
| Number of beds | | -0.024 | -0.023 | | |
| | | (2.29)** | (1.32) | | |
| Hours open per day | | 0.021 | 0.006 | | |
| | | (1.07) | (0.22) | | |
| hhsurvey*saw doc/nurse | | | | | -0.153 |
| | | | | | (0.94) |
| hhsurvey*physical exam | | | | | 0.126 |
| | | | | | (0.96) |
| Observations | 992 | 987 | 688 | 688 | 688 |

**Notes to Tables 3-7:**

T-statistics in parentheses. Standard errors are adjusted for within cluster correlations. * significant at 10%; ** sig.at 5%; *** sig. at 1%

[a] Base is being interviewed in user exit survey

Excluded category for age is 15-64.

Not shown: Intercept, controls for condition (malaria, infection, diarrhea, injury, other reason for visit) and province (Faritany) dummies.

Sample sizes for community fixed effects are smaller than in the ordered probits because communities with only one surveyed facility are dropped. For provider fixed effects, sample sizes reflect the requirement that there be observations from both household and user exit surveys for each provider.

## Table 4 - Determinants of client satisfaction with service quality relative to cost

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ordered probit: key facility characteristics | Ordered probit: full facility characteristics | Community fixed effects: full facility characteristics | Provider fixed effects | Provider fixed effects: interactions with household survey |
| Household survey interview | -0.760 | -0.748 | -0.311 | -0.292 | -0.168 |
| | (3.89)*** | (3.82)*** | (3.85)*** | (3.58)*** | (0.81) |
| Rural | -0.419 | -0.330 | | | |
| | (2.02)** | (1.50) | | | |
| Respondent years of school | 0.009 | 0.007 | 0.002 | 0.000 | -0.002 |
| | (0.59) | (0.46) | (0.28) | (0.04) | (0.26) |
| Asset index | -0.057 | -0.125 | -0.039 | -0.042 | -0.041 |
| | (0.64) | (1.37) | (0.70) | (0.80) | (0.77) |
| Male patient | -0.091 | -0.121 | 0.043 | 0.032 | 0.028 |
| | (0.89) | (1.29) | (0.72) | (0.55) | (0.48) |
| Patient age <5 | 0.101 | 0.105 | -0.069 | -0.067 | -0.060 |
| | (0.45) | (0.50) | (0.82) | (0.77) | (0.68) |
| Patient age 5-15 | -0.011 | 0.015 | -0.075 | -0.077 | -0.060 |
| | (0.05) | (0.07) | (0.70) | (0.69) | (0.55) |
| Patient age > 60 | 0.118 | 0.244 | -0.309 | -0.287 | -0.301 |
| | (0.43) | (0.97) | (1.91)* | (1.71)* | (1.75)* |
| Seen by doctor/nurse | 1.005 | 0.658 | 0.399 | 0.415 | 0.572 |
| | (4.26)*** | (3.48)*** | (1.36) | (1.50) | (1.68)* |
| Given physical exam | 0.265 | 0.191 | 0.152 | 0.116 | 0.059 |
| | (2.04)** | (1.35) | (1.68)* | (1.41) | (0.68) |
| Given health information | -0.328 | -0.290 | -0.065 | -0.078 | -0.079 |
| | (1.07) | (1.09) | (0.44) | (0.54) | (0.54) |
| Consultation cost | -0.884 | -1.070 | -0.140 | | |
| | (2.37)** | (2.60)*** | (0.26) | | |
| Appearance index | | 1.175 | -2.067 | | |
| | | (1.38) | (0.88) | | |
| Infrastructure index | | 0.763 | 0.261 | | |
| | | (1.58) | (0.32) | | |
| Share drugs short | | -0.170 | -0.354 | | |
| | | (0.67) | (0.71) | | |
| Number of beds | | -0.021 | -0.028 | | |
| | | (1.76)* | (0.76) | | |
| Hours open per day | | 0.054 | 0.095 | | |
| | | (2.75)*** | (1.51) | | |
| hhsurvey*saw doc/nurse | | | | | -0.246 |
| | | | | | (1.29) |
| hhsurvey*physical exam | | | | | 0.166 |
| | | | | | (0.82) |
| hhsurvey*cost | | | | | -0.576 |
| | | | | | (1.08) |
| Observations | 766 | 765 | 516 | 516 | 516 |

Note: See notes to Table 3. T-statistics in parentheses. Standard errors are adjusted for within cluster correlations.
* significant at 10%; ** significant at 5%; *** significant at 1%

## Table 5 - Determinants of client satisfaction with staff courtesy

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ordered probit: key facility characteristics | Ordered probit: full facility characteristics | Community fixed effects: full facility characteristics | Provider fixed effects | Provider fixed effects: interactions with household survey |
| Household survey interview | -0.394 | -0.385 | -0.161 | -0.162 | -0.264 |
| | (2.86)*** | (2.72)*** | (2.56)** | (2.61)** | (1.43) |
| Rural | -0.093 | -0.061 | | | |
| | (0.48) | (0.30) | | | |
| Respondent years of school | -0.010 | -0.015 | -0.006 | -0.008 | -0.009 |
| | (0.57) | (0.85) | (0.81) | (1.12) | (1.22) |
| Asset index | -0.076 | -0.117 | 0.013 | 0.008 | 0.011 |
| | (1.15) | (1.77)* | (0.27) | (0.16) | (0.22) |
| Male patient | 0.087 | 0.071 | 0.025 | 0.031 | 0.031 |
| | (0.71) | (0.59) | (0.53) | (0.65) | (0.66) |
| Patient age <5 | 0.021 | 0.030 | -0.001 | 0.001 | 0.004 |
| | (0.18) | (0.26) | (0.01) | (0.02) | (0.08) |
| Patient age 5-15 | 0.025 | -0.002 | 0.046 | 0.058 | 0.058 |
| | (0.16) | (0.01) | (0.56) | (0.68) | (0.69) |
| Patient age > 60 | 0.318 | 0.404 | 0.100 | 0.100 | 0.086 |
| | (1.80)* | (2.37)** | (0.99) | (0.95) | (0.82) |
| Seen by doctor/nurse | 0.544 | 0.425 | 0.075 | 0.108 | 0.147 |
| | (2.57)** | (1.74)* | (0.48) | (0.83) | (0.94) |
| Given physical exam | 0.557 | 0.484 | 0.213 | 0.202 | 0.129 |
| | (3.62)*** | (2.91)*** | (3.17)*** | (2.99)*** | (1.65) |
| Given health information | 0.062 | 0.153 | -0.002 | 0.005 | 0.007 |
| | (0.42) | (0.94) | (0.02) | (0.04) | (0.07) |
| Consultation cost | | 0.252 | 0.858 | | |
| | | (0.92) | (2.59)** | | |
| Appearance index | | 0.312 | -1.049 | | |
| | | (0.39) | (0.64) | | |
| Infrastructure index | | 0.579 | 0.639 | | |
| | | (1.59) | (0.84) | | |
| Share drugs short | | -0.157 | -0.570 | | |
| | | (0.53) | (1.72)* | | |
| Number of beds | | -0.037 | -0.002 | | |
| | | (3.06)*** | (0.05) | | |
| Hours open per day | | 0.027 | 0.009 | | |
| | | (1.23) | (0.23) | | |
| hhsurvey*saw doc/nurse | | | | | -0.075 |
| | | | | | (0.47) |
| hhsurvey*physical exam | | | | | 0.222 |
| | | | | | (1.64) |
| Observations | 998 | 996 | 691 | 691 | 691 |

Note: T-statistics in parentheses. Standard errors are adjusted for within cluster correlations.
* significant at 10%; ** significant at 5%; *** significant at 1%

## Table 6 - Determinants of client satisfaction with quantity of medicine

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ordered probit: key facility characteristics | Ordered probit: full facility characteristics | Community fixed effects: full facility characteristics | Provider fixed effects | Provider fixed effects: interactions with household survey |
| Household survey interview | -0.541 | -0.611 | -0.150 | -0.109 | -0.287 |
| | (3.46)*** | (3.61)*** | (1.62) | (1.19) | (1.92)* |
| Rural | -0.061 | 0.176 | | | |
| | (0.26) | (0.71) | | | |
| Respondent years of school | 0.025 | 0.017 | 0.000 | -0.002 | 0.000 |
| | (1.30) | (1.00) | (0.02) | (0.18) | (0.00) |
| Asset index | -0.078 | -0.114 | 0.073 | 0.066 | 0.057 |
| | (1.01) | (1.27) | (1.08) | (1.02) | (0.92) |
| Male patient | 0.291 | 0.190 | 0.012 | 0.015 | 0.015 |
| | (2.57)** | (1.66)* | (0.16) | (0.21) | (0.21) |
| Patient age <5 | -0.393 | -0.434 | -0.111 | -0.110 | -0.112 |
| | (2.33)** | (2.51)** | (1.35) | (1.34) | (1.35) |
| Patient age 5-15 | -0.091 | -0.079 | -0.006 | 0.030 | 0.029 |
| | (0.69) | (0.57) | (0.07) | (0.36) | (0.37) |
| Patient age > 60 | -0.067 | -0.048 | 0.009 | 0.002 | 0.015 |
| | (0.16) | (0.09) | (0.04) | (0.01) | (0.06) |
| Seen by doctor/nurse | | -0.030 | 0.080 | 0.096 | 0.099 |
| | | (0.10) | (0.45) | (0.51) | (0.55) |
| Given physical exam | | 0.120 | 0.165 | 0.169 | 0.178 |
| | | (0.49) | (1.81)* | (1.92)* | (2.04)** |
| Given health information | | -0.289 | -0.154 | -0.172 | -0.169 |
| | | (1.06) | (1.14) | (1.32) | (1.28) |
| Consultation cost | | -0.689 | -0.714 | | |
| | | (1.67)* | (2.10)** | | |
| Appearance index | | 0.694 | 2.921 | | |
| | | (1.10) | (2.76)*** | | |
| Infrastructure index | | 1.565 | 0.697 | | |
| | | (2.30)** | (2.03)** | | |
| Share drugs short | -1.132 | -0.579 | 0.615 | | |
| | (4.20)*** | (1.76)* | (1.67)* | | |
| Number of beds | | -0.001 | -0.031 | | |
| | | (0.08) | (1.45) | | |
| Hours open per day | | 0.011 | 0.074 | | |
| | | (0.34) | (2.86)*** | | |
| hhsurvey*share drugs short | | | | | 0.486 |
| | | | | | (1.64) |
| Observations | 661 | 625 | 458 | 487 | 487 |

Notes: [a] for sample prescribed medication.
See notes to Table 3. T-statistics in parentheses. Standard errors are adjusted for within cluster correlations.

* significant at 10%; ** significant at 5%; *** significant at 1%

## Table 7 - Determinants of client perceptions of facility condition (interior)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Ordered probit: key facility characteristics | Ordered probit: full facility characteristics | Community fixed effects: full facility characteristics | Provider fixed effects | Provider fixed effects: interactions with household survey |
| Household survey interview | -0.142 | -0.085 | -0.030 | -0.057 | 0.172 |
| | (1.16) | (0.68) | (0.58) | (1.09) | (0.43) |
| Rural | -0.388 | -0.425 | | | |
| | (1.49) | (1.75)* | | | |
| Respondent years of school | 0.002 | 0.001 | 0.003 | 0.003 | 0.003 |
| | (0.09) | (0.06) | (0.41) | (0.49) | (0.46) |
| Asset index | -0.154 | -0.182 | -0.015 | -0.020 | -0.022 |
| | (1.36) | (1.71)* | (0.44) | (0.58) | (0.63) |
| Male patient | 0.124 | 0.083 | 0.034 | 0.048 | 0.047 |
| | (1.81)* | (1.18) | (0.98) | (1.37) | (1.37) |
| Patient age <5 | -0.199 | -0.176 | -0.036 | -0.052 | -0.052 |
| | (1.70)* | (1.58) | (0.76) | (1.13) | (1.13) |
| Patient age 5-15 | 0.037 | -0.006 | -0.068 | -0.081 | -0.078 |
| | (0.43) | (0.06) | (1.26) | (1.51) | (1.46) |
| Patient age > 60 | 0.159 | 0.219 | 0.007 | 0.019 | 0.020 |
| | (0.66) | (1.05) | (0.08) | (0.20) | (0.21) |
| Seen by doctor/nurse | | -0.021 | -0.131 | -0.101 | -0.099 |
| | | (0.05) | (1.38) | (0.98) | (0.96) |
| Given physical exam | | -0.067 | -0.010 | 0.021 | 0.020 |
| | | (0.43) | (0.14) | (0.31) | (0.31) |
| Given health information | | -0.072 | -0.015 | 0.035 | 0.034 |
| | | (0.54) | (0.21) | (0.48) | (0.47) |
| Consultation cost | | 0.250 | -0.391 | | |
| | | (0.84) | (0.84) | | |
| Appearance index | 4.448 | 4.628 | 3.121 | | |
| | (4.99)*** | (5.22)*** | (2.60)** | | |
| Infrastructure index | 1.261 | 1.781 | 0.731 | | |
| | (3.82)*** | (4.68)*** | (1.54) | | |
| Share drugs short | | 0.220 | 0.426 | | |
| | | (0.94) | (1.26) | | |
| Number of beds | | -0.048 | -0.027 | | |
| | | (3.76)*** | (0.76) | | |
| Hours open per day | | -0.015 | | | |
| | | (0.82) | | | |
| hhsurvey*appearance | | | | | -0.279 |
| | | | | | (0.50) |
| hhsurvey*infrastructure | | | | | 0.041 |
| | | | | | (0.17) |
| Observations | 1035 | 995 | 690 | 690 | 690 |

Notes: See notes to Table 3. T-statistics in parentheses. Standard errors are adjusted for within cluster correlations.
* significant at 10%; ** significant at 5%; *** significant at 1%

**Appendix Table 1 - Definitions of Facility Structure and Process Indicators**

| | |
|---|---|
| Seen by doctor/nurse | Patient was treated by doctor or nurse |
| Given physical exam | Patient was given a physical examination |
| Given health info | Share of the following subjects for which practitioner provided information: nutrition, family planning,HIV/AIDS,malaria,vaccinations |
| Medicines explained | =1 if practitioner explained duration, dose, and frequency of medicine to be taken, 0 otherwise |
| Consultation cost | Provider level median of per visit consultation cost (excluding drugs) |
| Appearance index | 1 minus the mean of 0-1 indicators for: floor,walls, and ceiling (FWC) dirty, FWC damp, FWC decaying, insects evident, no toilet or dirty toilet |
| Infrastructure index | Share of the following present: radio or phone; electricity; working refrigerator; pump; incinerator; benches; examination table |
| Share drugs short | Share of the following medicines for which shortages were experienced in the last 90 days: chloraquine (for malaria); Mebendazole (intestinal parasites); oral rehydration salts; pain relievers (acetomeniphen or asprin); iron tablets; Cotrimoxazole (antibiotic) |
| Number of beds | Number of beds in the facility |
| Hours per day | Average hours open per day |

Note: The indicators 'saw doctor/nurse' through 'consultation cost' are based on the household and user exit survey data; the remaining indicators are from the facility survey.

**Appendix Table 2 - Regressions of respondent-reported process variables on provider diagnostic and treatment skills**

| | Full client sample | | | Children under 5 | | |
|---|---|---|---|---|---|---|
| | Given physical exam | Medicines explained[a] | Given health information | Given physical exam | Medicines explained[a] | Given health information |
| male patient | -0.028 | -0.035 | -0.013 | 0.058 | -0.049 | 0.041 |
| | (0.79) | (0.73) | (0.35) | (0.81) | (0.88) | (0.76) |
| patient age <5 | -0.004 | 0.032 | -0.028 | -- | -- | -- |
| | (0.09) | (0.96) | (0.58) | -- | -- | -- |
| patient age 5-15 | 0.082 | 0.009 | -0.072 | -- | -- | -- |
| | (1.65) | (0.17) | (1.14) | -- | -- | -- |
| paitent age > 60 | 0.106 | 0.139 | -0.239 | -- | -- | -- |
| | (1.56) | (3.09)*** | (3.44)*** | -- | -- | -- |
| infection | 0.153 | 0.056 | 0.064 | 0.218 | -0.037 | 0.061 |
| | (2.20)** | (1.10) | (1.27) | (2.16)** | (0.44) | (0.61) |
| malaria | 0.075 | 0.032 | 0.043 | -0.027 | -0.040 | -0.006 |
| | (1.35) | (0.97) | (0.66) | (0.25) | (0.48) | (0.04) |
| diarrhea | 0.123 | 0.053 | 0.141 | 0.208 | -0.036 | 0.202 |
| | (1.86)* | (1.22) | (2.09)** | (2.48)** | (0.29) | (2.05)** |
| injury | 0.078 | -0.227 | 0.091 | 0.134 | 0.056 | 0.162 |
| | (0.95) | (1.80)* | (1.02) | (0.72) | (1.31) | (0.98) |
| Provider diagnosis/treatment skill[b] | 0.226 | 0.038 | 0.351 | 0.306 | 0.192 | 0.473 |
| | (1.54) | (0.28) | (2.53)** | (1.46) | (0.92) | (2.94)*** |
| Constant | 0.677 | 0.839 | 0.549 | 0.633 | 0.926 | 0.487 |
| | (15.54)*** | (27.02)*** | (16.21)*** | (9.43)*** | (18.35)*** | (5.60)*** |
| Observations | 853 | 607 | 873 | 233 | 157 | 238 |

Notes:

[a] User exit survey only

[b]Derived from direct clinical observation of practitioners with patients under 5. It is the mean of provider level random effects from regressions of diagnostic performance and treatment prescription performance on patient characteristics and presenting condition. See text for details.

T-statistics in parentheses. Standard errors adjusted for clustering at provider level.
* significant at 10%; ** significant at 5%; *** significant at 1%