

9

Survey Tools for Assessing Performance in Service Delivery

Jan Dehn, Ritva Reinikka, and Jakob Svensson

It has become increasingly clear that budget allocations, when used as indicators of the supply of public services, are poor predictors of the actual quantity and quality of public services, especially in countries with poor accountability and weak institutions. At least four breaks in the chain can be distinguished between spending—meant to address efficiency and equity concerns—and its transformation into services (Devarajan and Reinikka 2002). First, governments may spend on the wrong goods or the wrong people. A large portion of public spending on health and education is devoted to private goods, where government spending is likely to crowd out private spending (Hammer, Nabi, and Cercone 1995). Furthermore, most studies of the incidence of public spending in health and education show that benefits accrue largely to the rich and middle-class; the share going to the poorest 20 percent is almost always less than 20 percent (Castro-Leal and others 1999). The first three chapters in this volume discuss benefit incidence analysis.

Second, even when governments spend on the right goods or the right people, the money may fail to reach the frontline service provider. A study of Uganda in the mid-1990s, using a Public Expenditure Tracking Survey (PETS)—the topic of this chapter—showed that only 13 percent of nonwage recurrent expenditures for primary education actually reached the primary school (Reinikka 2001). The considerable variation in grants received across schools was determined more by the political economy than by efficiency and

Box 9.1 Public-Sector Agencies, Measurability, PETS, and QSDS

The organizational structure of public sector agencies involves multiple tiers of management and frontline workers. Multiplicity is also a key aspect of the tasks they perform and the stakeholders they serve. For example, primary education teaches young children to read and write, and it also teaches social skills, instills citizenship, and so forth. The different tasks and interests at each tier may compete with each other for limited resources in a finite time period. Moreover, the output of public service agencies is often difficult to measure, and systematic information on specific inputs and outputs is rarely available in developing countries. In many cases management information systems are unreliable in the absence of adequate incentives to maintain them. On closer observation, the characteristics of public service agencies and the nature of their tasks explain why traditional tools for public expenditure analysis alone may not be adequate for evaluating performance. Because the Public Expenditure Tracking Survey (PETS) and Quantitative Service Delivery Survey (QSDS) can bring together data on inputs, outputs, user charges, quality, and other characteristics directly from the service-providing unit, more can be learned about the linkages, leakage, and the way spending is transformed into services.

Over and above the problem of vague output measures, the existence of multiple principals reduces the agent's incentives, because activities often desired by the principals to realize their respective

equity considerations. Larger schools and schools with wealthier parents received a larger share of the intended funds (per student), while schools with a higher share of unqualified teachers received less (Reinikka and Svensson 2002).

Third, even when the money reaches the primary school or health clinic, the incentives to provide the service may be weak. Service providers in the public sector may be poorly paid, hardly ever monitored, and given few incentives from the central government bureaucracy, which is mostly concerned with inputs rather than outputs. The result can be a high absenteeism rate among frontline workers. The Quantitative Service Delivery Survey (QSDS)—the other instrument featured in this chapter—is a useful tool for getting at these issues. A survey in Bangladesh, described later, showed that the absenteeism rate was 74 percent for doctors in primary health care centers (Chaudhury and Hammer 2003).

goals are substitutes for each other. Similarly, when some task outcomes are verifiable and others are not, it may not be optimal to provide explicit incentives for any tasks, as the agent would otherwise divert all effort from unverifiable to verifiable tasks. In education, for example, exam results would be disproportionately emphasized over aspects that lend themselves less easily to monitoring and measurement. Incentive schemes are most suitable when outcomes are clearly defined, observable, and unambiguous, and become weak when neither outcomes nor actions are observable, such as in a typical government ministry. Public service providers also often lack competitors. Although the introduction of competition does not in itself guarantee better performance, it places greater emphasis on other management devices.

PETS and QSDS can increase the observability of both outputs and actions and thereby provide new information about the complex transformation from public budgets to services. Tailored to the specific circumstances, these tools can help identify incentives and shed light on the interactions to which these incentives give rise, such as collusion and bribery. They can also illuminate the political economy, such as the effect of interest groups. The novelty of the PETS-QSDS approach lies not so much in the development of new methods of analysis per se, but in the application of known and proven methods (microsurveys) to service providers.

Sources: Bernheim and Whinston (1986); Dixit (1996, 1997, 2000).

Fourth, even if the services are effectively provided, households may not take advantage of them. For economic and other reasons, parents pull their children out of school or fail to take them to the clinic. These demand-side failures often interact with the supply-side failures to generate a low level of public services and human development outcomes among the poor.

This chapter argues that microeconomic-level survey tools are useful not only at the household or enterprise level but also at the service provider level to assess the efficiency of public spending and the quality and quantity of services. The two microlevel surveys discussed here, the PETS and the QSDS, both obtain policy-relevant information on the *agent* (say, a district education office) and the *principal* (say, the ministry of finance or a parent-teacher association) (box 9.1). Similarly, repeat PETSs or QSDSs can be used as tools to evaluate the impact of policy changes.

Key Features of PETS and QSDS

Government resources earmarked for particular uses flow within legally defined institutional frameworks, often passing through several layers of government bureaucracy (and the banking system) down to service facilities, which are charged with the responsibility of exercising the spending. But information on *actual* public spending at the frontline level is seldom available in developing countries. A PETS—frequently carried out as part of a public expenditure review—tracks the flow of resources through these strata to determine how much of the originally allocated resources reaches each level. It is therefore useful as a device for locating and quantifying political and bureaucratic capture, leakage of funds, and problems in the deployment of human and in-kind resources, such as staff, textbooks, and drugs. It can also be used to evaluate impediments to the reverse flow of information to account for actual expenditures.

The primary aim of a QSDS is to examine the efficiency of public spending and incentives and various dimensions of service delivery in provider organizations, especially on the frontline. The QSDS can be applied to government as well as to private for-profit and not-for-profit providers. It collects data on inputs, outputs, quality, pricing, oversight, and so forth. The facility or frontline service provider is typically the main unit of observation in a QSDS in much the same way as the firm is in enterprise surveys and the household is in household surveys. A QSDS requires considerable effort, cost, and time compared to some of its alternatives—surveys of perceptions, in particular (box 9.2). As the example of Uganda, discussed later, demonstrates, the benefits of quantitative data can easily offset the cost.

Both tools explicitly recognize that an agent may have a strong incentive to misreport (or not report) key data. This incentive derives from the fact that information provided, for example, by a health facility partly determines its entitlement to public support. In cases where resources (including staff time) are used for other purposes, such as shirking or corruption, the agent involved in the activity will most likely not report it truthfully. Likewise, official charges may only partly capture what the survey intends to measure (such as the user's cost of service). The PETS and QSDS deal with these data issues by using a multiangular data collection strategy—that is, a combination of information from different sources—and by carefully considering which sources and respondents have incentives to misreport and identifying data sources that are the least contaminated by these incentives. The triangulation strategy of data collection serves as a means of cross-validating the information obtained separately from each source.

Box 9.2 PETS, QSDS, and Other Tools to Assess Service Delivery

The Public Expenditure Tracking survey (PETS) and Quantitative Service Delivery Survey (QSDS) are distinct from other existing survey approaches, such as facility modules in household surveys or empirical studies to estimate hospital cost functions. Living Standards Measurement Study household surveys have included health facility modules on an ad hoc basis (Alderman and Lavy 1996). A number of the Demographic and Health Surveys carried out in more than 50 developing countries have also included a service provider component. Similarly, the Family Life Surveys implemented by RAND have combined health provider surveys with those of households. The rationale for including a facility module in a household survey is to characterize the link between access to and quality of public services and key household welfare indicators. Because the perspective in these surveys is that of the household, they pay little attention to the question of why quality and access are the way they are. In most cases facility information collected as a part of community questionnaires relies on the knowledge of one or more informed individuals (Frankenberg 2000). Information supplied by informants is therefore heavily dependent on the perception of a few individuals and not detailed enough to form a basis for analysis of service delivery parameters, such as operational efficiency, effort, or other performance indicators. To the extent that the information is based on perceptions, there may be additional problems attributable to the subjective nature of the data and its sensitivity to respondents' expectations. By contrast, the PETS-QSDS approach emphasizes and quantitatively measures provider incentives and behavior.

Unlike household surveys, the hospital cost function literature has a clear facility focus analogous to that applied to private firms in enterprise surveys. This literature typically looks at cost efficiency—mostly in hospitals in the United Kingdom and the United States—although work on hospital performance has also been conducted in developing countries (Wagstaff 1989; Wagstaff and Barnum 1992; and Barnum and Kutzin 1993). Perhaps more relevant, though, is the budding literature on cost efficiency and other performance indicators in clinics and primary health facilities in developing countries (McPake and others 1999; Somanathan and others 2000). In the PETS-QSDS approach, the main departure from the cost function literature is the explicit recognition of the close link between the public-sector service provider and the rest of the public sector. Providers of public services typically rely on the wider government structure for resources, guidance about what services to provide, and how to provide them. This dependence makes them sensitive to systemwide problems in transfer of resources, the institutional framework, and the incentive system, which private providers do not face.

Source: Lindelöw and Wagstaff (2003).

PETS and QSDS can also complement each other. Their combination allows for the evaluation of wider institutional and resource-flow problems on the performance of frontline service providers. With more precise (quantitative) measures, it will be easier for policymakers in developing countries to design effective policies and institutional reforms.

Design and Implementation

Like other microlevel surveys, PETS and QSDS require careful design and implementation. At least some members of the study team should have prior experience with surveys. The intuitive appeal of PETS and QSDS can belie the complexity involved in their planning and implementing.¹ This section outlines the steps involved in successful design and implementation of PETS and QSDS in light of the experience to date. Most steps are common to both surveys, given that PETS typically includes a facility component and that QSDS needs to relate government facilities to the public sector hierarchy.

Consultations and Scope of the Study

During the initial phase, the survey team needs to consult with in-country stakeholders, including government agencies (ministry of finance, sector ministries, and local governments), donors, and civil society organizations. Broad-based consultations are useful for:

- reaching agreement on the purpose and objectives of the study; and choosing the sector(s) for the study
- identifying key service delivery issues and problems (research questions) in the chosen sector(s)
- determining the structure of government's resource flow, rules for resource allocation to frontline facilities, and the accountability system
- obtaining a good understanding of the institutional setting of government and of private for-profit and not-for-profit providers
- checking data availability at various tiers of government or other provider organizations and at the facility level
- assessing available local capacity to carry out the survey and to engage in data analysis and research
- choosing the appropriate survey tool

A survey requires considerable effort, so it is advisable to limit the number of sectors to one or two. Until now, PETS and QSDS have mostly been carried out in the “transaction-intensive” health

and education sectors with clearly defined frontline service delivery points (clinics and schools), but there is no reason to limit the use of these tools to health and education.

Rapid Data Assessment

A rapid data assessment may be needed to determine the availability of records at various layers of the government—as well as in the private sector—particularly at the facility level. Some studies have failed because the availability of records in local governments and facilities was not assessed beforehand. It is important to verify the availability of records early on, even if it means a delay and some extra up-front costs.

The consultations at the design stage often take place in the capital city, so it is easy to visit facilities in its vicinity to check on records, with the proviso that they may not be representative of facilities in remote locations. It may be important to assess data availability in more than one location. A simple questionnaire is usually sufficient for a rapid data assessment.

Questionnaire Design

It is important to ensure that recorded data collected at one level in the system can be cross-checked against the same information from other sources. A PETS or QSDS typically consists of questionnaires for interviewing facility managers (and staff) as well as separate data sheets to collect quantitative data from facility records. It also collects data from local, regional, and national provider organizations in three ownership categories: government, private for-profit providers, and private not-for-profit organizations. The combination of questionnaires and datasheets is usually flexible enough to evaluate most of the problems under study. A beneficiary survey can also be added.

As mentioned earlier, a crucial component of PETS-QSDS is the explicit recognition that respondents may have strong incentives to misreport (or not report at all) certain information. As a general guideline, information should be collected as close as possible to the original source. Data is thus typically collected from records kept by the facility for its own use (for example, patient numbers can be obtained from daily patient records kept by staff for medical use, drug use can be derived from stock cards, and funding to schools can be recorded from check receipts). It is also important to keep in mind that some information (for instance on corruption) is almost impossible to collect directly (especially from those benefiting from it). Instead, different sources of information have to be combined.²

To be comparable a core set of questions must remain unchanged across waves of surveys, across sectors, and across countries. Six core elements for all facility questionnaires have been identified:

- *Characteristics of the facility.* Record the size; ownership; years of operation; hours of operation; catchment population; competition from other service providers; access to infrastructure, utilities and other services; and range of services provided. Information about income levels and other features of the population living in the vicinity of the facility may also be useful.

- *Inputs.* Because service providers typically have a large number of inputs, it may not be feasible to collect data on all of them. Some inputs are typically more important than others. For example, labor and drugs account for 80–90 percent of costs in a typical primary health care facility. In addition, there may be important capital investments. The key point in the measurement of inputs is that they need to be valued in monetary terms. Where monetary values are not readily available, this requires that quantities be recorded carefully and consistently and price information (for example, wages and allowances for labor) be assembled for each key input.

- *Outputs.* Examples of measurable outputs include numbers of inpatient and outpatients treated, enrollment rates, and numbers of pupils completing final exams. Unlike inputs, outputs rarely convert to monetary values (public services are often free or considerably subsidized). Efficiency studies frequently use hybrid input-output measures, such as cost per patient.

- *Quality.* Quality is multidimensional, and an effort should be made to capture this multidimensionality by collecting information on different aspects of quality. Examples of these aspects of quality include observed practice, staff behavior and composition, availability of crucial inputs, and provision of certain services such as laboratory testing. Information collected from users can also capture aspects of quality.

- *Financing.* Information should be collected on sources of finance (government, donor, user charges), amounts, and type (in-kind versus financial support).

- *Institutional mechanisms and accountability.* Public facilities do not face the same competitive pressures as private facilities. Instead, they are subject to supervision and monitoring by central, regional, or local government institutions, civil society, political leaders, and the press. That means collecting information on supervision visits, management structures, reporting and record-keeping practices, parent or patient involvement, and audits.

Variations to this basic template can include modules to test specific hypotheses. Box 9.3 discusses sampling issues for PETS and QSDS.

Box 9.3 Sample Frame, Sample Size, and Stratification

Many developing countries have no reliable census on service facilities. An alternative is to create a sample frame from other sources (administrative records of some kind). A list of public facilities is often available from the central government or donors active in the sector. However, creating a reliable list of private providers may be a considerable undertaking or may simply not be feasible. An alternative is to mimic the two-stage design that is typically used in household surveys. In other words information on the private facilities is, at the first stage, gathered from randomly drawn sampling units (for example, a district or the catchment population of a government facility). At the second stage, the required number of private facilities is drawn from a list of facilities in the sampling unit.

When determining sample size, a number of issues must be considered. First, the sample should be sufficiently large and diverse enough to represent the number and range of facilities in the specified categories. Second, subgroups of particular interest (for example, rural and private facilities) may need to be more intensively sampled than others. Third, the optimal sample size is a tradeoff between minimizing sampling errors and minimizing nonsampling errors (the former typically decrease and the latter increase with the sample size). Arguably, in a Public Expenditure Tracking Survey (PETS) or Quantitative Service Delivery Survey (QSDS), nonsampling error (caused by poor survey implementation) is more of a concern than sampling error, as the data are often in a highly disaggregated form and hence labor-intensive to collect. Enumerator training and field testing of the instrument are therefore critical in obtaining high-quality data. Finally, these objectives must be achieved within a given budget constraint.

The above considerations often lead to a choice of a stratified random sample. Stratification entails dividing the survey population into subpopulations and then sampling these subpopulations independently as if they were individual populations. Stratification reduces sampling variance (increases sampling efficiency) and ensures a sufficient number of observations for separate analysis of different subpopulations. Stratification is an opportunity for the surveyor to use prior information about the population to improve the efficiency of the statistical inference about quantities that are unknown.

Sampling issues become more complicated when PETS and QSDS are combined. In PETS one may want to sample a relatively large number of local government administrations. But sampling a large number of districts reduces the number of facilities that can be sampled within each district for a given budget. From the perspective of QSDS, it is desirable to have more facilities within fewer districts in order to characterize the intradistrict variation among facilities.

Sources: Alreck and Settle (1995); Grosh and Glewwe (2000); Rossi and Wright (1983).

Training, Field Testing, and Implementation

Once the survey instruments (questionnaires and data sheets) are drafted according to the specific needs of the study, the next steps are piloting the questionnaire and then training the enumerators and their supervisors. Experience has shown that training is a crucial component and a significant amount of time has to be allocated for it. After completion of the training, survey instruments should be field tested. Supervision of enumerators is critical during implementation of the survey. It is also good practice to prepare a detailed implementation manual for the survey personnel. Using local consultants to conduct the PETS or QSDS is likely to be more cost-effective as well as beneficial for capacity building. In-country consultants are likely to have a comparative advantage over their international counterparts regarding knowledge of local institutions.

All instruments should be field tested on each type of provider in the sample (government, nongovernmental organization, and private), because different providers may have different practices of record-keeping. In case the field test leads to major modifications in the questionnaire, the modified questionnaire should be retested before finalization. The field-testing procedure takes between two weeks and one month to complete. More time is required if the final questionnaire is in more than one language, because changes made in one language need to be translated to the other.

Data Entry and Verification

Cost may limit the study team's ability to monitor the data collection process continuously. In this case the team should do spot checks during the early stages of data collection to discover possible problems and make the necessary adjustments in time. The team will also need to scrutinize the completed questionnaires and the data files, and, where necessary, request return visits to facilities or to various levels of the government. The output from this stage is the complete data set. It is also important to prepare comprehensive documentation of the survey soon after its completion.

Analysis, Report, and Dissemination

The analysis is typically done either by the study team or by the survey consultant in collaboration with the team. The reports and analysis should be widely disseminated to encourage debate and discussion to facilitate the alleviation of the problems highlighted in the survey.

Findings from PETS

Several countries have implemented public expenditure tracking surveys, including, Ghana, Honduras, Papua New Guinea, Peru, Rwanda, Tanzania, Uganda, and Zambia. This section summarizes findings from these studies, focusing on the Uganda and Zambia PETS in education, and a health and education PETS in Honduras. In the first two, leakage of public funds—defined as the share of resources intended for but not received by the frontline service facility—is the main issue. That was the main issue for the other PETSs carried out in Africa as well. The Honduras study diagnoses incentives that negatively affect staff performance, as manifested in ghost workers, absenteeism, and job migration.

Capture of Public Funds

Uganda, in 1996, was the first country to do a PETS. The study was motivated by the observation that despite a substantial increase in public spending on education, the official reports showed no increase in primary school enrollment. The hypothesis of the study was that actual service delivery, proxied by primary enrollment, was worse than budgetary allocations implied because public funds were subject to capture (by local government officials) and did not reach the intended facilities (schools). To test this hypothesis, a PETS was conducted to compare budget allocations to actual spending through various tiers of government, including frontline service delivery points in primary schools (Ablo and Reinikka 1998; Reinikka 2001).

Adequate public accounts were not available to report on actual spending, so a survey collected a five-year panel data set on provider characteristics, spending (including in-kind transfers), and outputs in 250 government primary schools. Initially, the objective of the PETS was purely diagnostic, that is, to provide a reality check on public spending. Subsequently, it became apparent that a quantitative tool like the PETS can provide useful microeconomic data for analyses of service provider behavior and incentives.

As mentioned earlier, the Ugandan school survey provides a stark picture of public funding on the frontlines. On average, only 13 percent of the annual capitation (per student) grant from the central government reached the schools in 1991–95. Eighty-seven percent either disappeared for private gain or was captured by district officials for purposes unrelated to education. Most schools (roughly 70 percent) received very little or nothing. The picture looks slightly better when constraining the sample to the last year of the survey period. Still, only 20 percent of the total capitation grant from the

central government reached the schools in 1995 (Reinikka and Svensson 2002). About 20 percent of teacher salaries were paid to ghost teachers—teachers who never appeared in the classroom. Subsequent PETSs in Tanzania and Ghana showed similar problems of capture in health care and education, although of a somewhat smaller magnitude (Government of Tanzania 1999, 2001; Xiao and Canagarajah 2002).

Following publication of the findings, the central government made a swift attempt to remedy the situation. It began publishing the monthly intergovernmental transfers of public funds in the main newspapers, broadcasting information about them on radio, and requiring primary schools to post information on inflows of funds for all to see. This tactic not only made information available to parent-teacher associations, but also signaled to local governments that the center had resumed its oversight function. An evaluation of the information campaign (using a repeat PETS) reveals a large improvement. Although schools are still not receiving the entire grant (and there are delays), capture was reduced from an average of 80 percent in 1995 to 20 percent in 2001 (Reinikka and Svensson, 2003a).³ A before-and-after assessment, comparing outcomes for the same schools in 1995 and 2001—and controlling for a broad range of school-specific factors, such as household income, teacher education, school size, and degree of supervision—suggests that the information campaign can explain two-thirds of this massive improvement (Reinikka and Svensson 2003a). This finding is likely to be an upper bound on the effect, since the effect of the information campaign cannot be distinguished from other policy actions or changes that simultaneously influenced all schools' ability to claim their entitlement.

A key component in the information campaign was the newspaper publication of monthly transfers of public funds to the districts. Thus schools with access to newspapers have been more extensively exposed to the information campaign. Interestingly, in 1995 schools with access to newspapers and those with no newspaper coverage suffered just as much from local capture. From 1995 to 2001, both groups experienced a large drop in leakage, which is consistent with the before-and-after findings. However, the reduction in capture is significantly higher for the schools with newspapers. On average these schools increased their funding by 10 percentage points more than the schools that lacked newspaper coverage. The results hold also when controlling for differences in income, school size, staff qualifications, and the incidence of supervision across the two groups.

With a relatively inexpensive policy action—provision of mass information—Uganda has managed dramatically to reduce capture

of a public program aimed at increasing primary education. Being less able to claim their entitlement from the district officials before the campaign, poor schools benefited most from the information campaign.

According to a recent PETS in primary education in Zambia—unlike in Uganda in the mid-1990s—rule-based allocations seem to be reaching the intended beneficiaries well: more than 90 percent of all schools received their rule-based nonwage allocations, and 95 percent of teachers received their salaries (Das and others 2002). But rule-based funding accounts for only 30 percent of all funding. In discretionary allocations (70 percent of the total), the positive results no longer hold: fewer than 20 percent of schools receive *any* funding at all from discretionary sources. The rest is spent at the provincial and district level. Similarly, in the case of overtime allowances (which must be filed every term) or other discretionary allowances, 50 percent were overdue by six months or more.

In conclusion, the PETS carried out in Africa found leakage of nonwage funds on a massive scale. Salaries and allowances also suffer from leakage but to a much lesser extent. Given that availability of books and other instructional materials are key to improving the quality of schooling, the fact that between 87 percent (Uganda) and 60 percent (Zambia) of the funding for these inputs never reach the schools makes leakage a major policy concern in the education sector. Furthermore, there is clearly scope for better targeting of interventions to improve public sector performance. Instead of instituting more general public sector reforms, the PETS in Uganda shows that it may be more efficient to target reforms and interventions at specific problem spots within the public hierarchy. For example, the PETS pointed to the fact that nonwage expenditures are more prone to leakage than salary expenditures. The PETS also demonstrated that leakage occurred at specific tiers within the government. This knowledge can be exploited to effect more efficient interventions.

Ghost Workers, Absenteeism, and Job Migration

Honduras used the PETS to diagnose moral hazard with respect to frontline health and education staff (World Bank 2001). The study demonstrated that issues of staff behavior and incentives in public service can have adverse effects on service delivery, such as ghost workers, absenteeism, and job capture by employees, even when salaries and other resources reach frontline providers. One hypothesis was that the central payroll office had no means of ensuring that public employees really exist (ghost workers). Another concern was that employees were not putting in full hours of work (absenteeism). Yet another question was whether workers were working

where they were supposed to be working (migration of posts). Migration of posts was considered to pose a major problem, because the Honduran system of staffing does not assign posts to individual facilities but rather to the central ministry. Given that the central ministry has discretion over the geographic distribution of posts, the system provides an incentive to frontline staff to lobby the ministry to have their posts transferred to more attractive locations, most often to urban areas. The implication is that posts migrate from rural areas and primary health care and primary school jobs toward cities and higher levels of health care and schooling. Such migration is neither efficient nor equitable.

The PETS set out to quantify the incongruity between budgetary and real assignments of staff and to determine the degree of attendance at work. The PETS used central government information sources and a nationally representative sample of frontline facilities in health and education. Central government payroll data indicated each employee's place of work. The unit of observation was not the facility but the staff member, both operational and administrative, and at all levels of the two sectors from the ministry down to the service facility level.⁴

In health, the study found that 2.4 percent of staff did not exist (ghost workers). For general practitioners (GPs) and specialists, 8.3 percent and 5.1 percent of staff, respectively, were ghost workers. Second, absenteeism was generic, with an average attendance rate of 73 percent across all categories of staff in the five days before the survey date. Thirty-nine percent of absences were without justifiable reason (such as sick leave, vacations, and compensation for extra hours worked). That amounts to 10 percent of total staff work time. Third, many health care providers, especially GPs and specialists, held multiple jobs. Fifty-four percent of specialist physicians had two or more jobs, of which 60 percent were in a related field. Fourth, 5.2 percent of sampled staff members had migrated to posts other than the one to which they were assigned in the central database, while 40 percent had moved since their first assignment. The highest proportions of migrators were found among GPs. Migration was typically from a lower-level to a higher-level institution, although there was also some lateral migration. Job migration was found to reflect a combination of employee capture and budget inflexibility.

In education, 3 percent of staff members on the payroll were found to be ghosts, while 5 percent of primary school teachers were unknown in their place of work. Staff migration was highest among nonteaching staff and secondary teachers. Absenteeism was less of a problem than in the health sector, with an average attendance rate

of 86 percent across all categories of staff. Fifteen percent of all absences were unaccounted for. Multiple jobs in education were twice as prevalent as in health, with 23 percent of all teachers doing two or more jobs. Finally, 40 percent of all education sector workers had administrative jobs, suggesting perhaps a preference for nonfrontline service employment or deliberate employment creation on the part of the government.

In conclusion, the Honduras study illustrates well that efforts to improve public sector service delivery must consider not just resource flows, but also incentives the staff has to perform.

PETS and QSDS as Research Tools

For a careful policy evaluation, it is important to design the PETS and QSDS instruments in such a way that the data have enough observations (say, facilities) for robust statistical analyses.⁵ Unless the policy change affects a subset of facilities, it is generally not possible to evaluate its effectiveness using only cross-sectional data. Hence, a panel data set is required. The first round of baseline QSDSs includes health care in Bangladesh, Chad, Madagascar, Mozambique, Nigeria, and Uganda.

The time dimension of the rounds of surveys depends on the speed at which policy changes translate into outcomes, that is, the time it takes for the policy change to be reflected in actual changes in spending, the speed at which the spending changes affect actual service delivery, and the time it takes the changes in service delivery to produce changes in outcomes. Several years of data may be needed, either by returning to the facility each year or, in the case of ex post policy evaluations, by collecting data on several time periods at once during the same visit. For example, five years of data was collected from schools in the Uganda PETS during one survey (Ablo and Reinikka 1998; Reinikka 2001).

The not-for-profit sector plays an important role in provision of social services in many developing countries. In the health sector, religious organizations are particularly prevalent. One of the purposes of the Uganda QSDS (box 9.4) was to examine the effect of not-for-profit providers on the quantity and quality of primary health care. To find this effect, one needs to know how the not-for-profit actors are motivated as service providers: are they altruistic or are they maximizing perks (Reinikka and Svensson 2003b)? The Uganda QSDS provides data that can be used for such an evaluation, since the survey collected data from government, private for-profit providers, and private not-for-profit (religious) providers.

Box 9.4 QSDS of Dispensaries in Uganda

A Quantitative Service Delivery Survey (QSDS) of dispensaries (with and without maternity units) was carried out in Uganda in 2000. A total of 155 dispensaries were surveyed, of which 81 were government facilities, 30 were private for-profit, and 44 were operated on a nonprofit basis. The survey collected data at the level of the district administration, the health facility, and the beneficiary to capture the links between these three levels. Comparisons of data from different levels permitted cross-validation (triangulation) of information. At the district level, a *district health team questionnaire* was administered to the district director of health services that included data on health infrastructure, staff training, supervision arrangements, and sources of financing for one fiscal year. A *district health facility data sheet* was also used to collect detailed information on the 155 health units, including staffing, salaries, the supply of vaccines and drugs to the facilities, and the monthly statistics from each facility on the number of outpatients, inpatients, immunizations, and deliveries.

At the facility level, a *health facility questionnaire* gathered a broad range of information relating to the facility and its activities. Each facility questionnaire was supplemented with a *facility data sheet* to obtain data from the health unit records on staffing, salaries, daily patient records, type of patients using the facility, immunization, and drug supply and use. These data were obtained directly from the records kept by facilities for their own use (medical records), rather than administrative records submitted to local government. Finally, also at the facility level, an *exit poll* was used to interview around 10 patients per facility on cost of treatment, drugs received, perceived quality of services, and reasons for preference for this unit instead of alternative sources of health care.

Source: Lindelöw, Reinikka, and Svensson (2003).

In the cross-section, religious not-for-profit facilities were found to hire qualified workers below the market wage. Moreover, these facilities are more likely to provide pro-poor services and services with a public good element and to charge strictly lower prices for services than do for-profit units. Religious not-for-profit and for-profit facilities both provide a better quality of care than their government counterparts, although government facilities are better equipped. These findings are consistent with there being a religious premium in working in a religious, nonprofit facility (that is, staff in such facilities are prepared to work for a salary below the market rate) and with religious nonprofits being driven (partly) by altruistic concerns.

Detailed knowledge of the institutional environment not only is important for identifying the right questions to ask, but can also assist in identifying causal effects in the data. The Uganda QSDS is an example. The year of the survey, the government of Uganda initiated a program stipulating that each not-for-profit unit was to receive a fixed grant for the fiscal year. However, because this was a new, and partly unanticipated, program and because communications in general were poor, some not-for-profit facilities did not receive their first grant entitlement until the following fiscal year. This *de facto* phasing of the grant program provides a near natural policy experiment of public financial aid. Analysis of the QSDS data reveals that financial aid leads to more testing of suspected malaria and intestinal worm cases—an indication of quality—and lower prices, but only in religious not-for-profit facilities. The estimated effects are substantial.

Another interesting pattern in the data is related to prescription antibiotics. Preliminary analysis shows that antibiotic prescriptions are generally very high. In fact, almost half of the patients report receiving some antibiotic. In some cases, patients receive several types at the same time. Government facilities are significantly more likely to provide antibiotics than private providers, and the effect is particularly strong in government facilities without qualified (medical) staff. Work is under way to distinguish between three (complementary) explanations for these patterns: the provision of antibiotics is a substitute for effort; the provision of antibiotics is higher in government units because the opportunity cost of providing antibiotics is lower; and patients demand antibiotics when treated and when the provider has a weak (bargaining) position, it (over)provides antibiotics.

A QSDS-type survey was conducted in Bangladesh, where unannounced visits were made to health clinics with the intention of discovering what fraction of medical professionals were present at their assigned post (Chaudhury and Hammer 2003). The survey quantified the extent of this problem on a nationally representative scale. The first notable result is that, nationwide, the average number of unfilled vacancies for all types of providers is large (26 percent). Regionally, vacancy rates are generally higher in the poorer parts of the country. Absentee rates for medical providers in general are quite high (35 percent), and these rates are particularly high for doctors (40 percent; at lower levels of health facilities, the absentee rate for doctors increases to 74 percent). When exploring determinants of staff absenteeism, the authors find that whether the medical provider lives near the health facility, has access to a road, and has electricity are important.

Linkages to Other Tools

Facility-level analysis can be linked upstream to the public administration and political processes through public official surveys and downstream to households through household surveys and thereby can combine supply of and demand for services. Linking the PETS-QSDS with the household surveys would include the demand for services or outcomes, and linking it with public official surveys would include political economy and administrative aspects. Taken together, such data would allow a much more comprehensive analysis of service delivery performance and its determinants. The PETS in Zambia and Laos (the latter is currently in the field) include a household survey, while the ongoing QSDS in Nigeria incorporates a survey of local officials. Reports on these studies will become available during 2003.

Benefit incidence analysis, common in many developing countries, combines household data on consumption of public goods with information on public expenditures. A unit subsidy per person is determined, and household usage of the service is then aggregated across key social groups to impute the pattern and distribution of service provision. A methodological problem in incidence analysis, however, is the practice of using budgeted costs as proxies for service benefits—see chapter 2 in this volume. The PETS approach permits a better measurement of these benefits by relaxing the assumption that budgeted resources are automatically translated into actual services. Specifically, a PETS or QSDS can provide a “filter coefficient” for public expenditures, which can be used to deflate budget allocations. For example, such a coefficient in primary education in Uganda was 0.2 for nonsalary spending and 0.8 for spending on teacher salaries in the mid-1990s (Reinikka 2001). In Zambia this coefficient was around 0.4 for nonwage public spending (average for rule-based and discretionary spending) in 2002. For salaries it was 0.95, apart from allowances for which the coefficient ranged between 0.85 and 0.5 (Das and others 2002). These examples are national averages. They can be further refined, because evidence from the PETS indicates that poorer and smaller schools tend to receive less funding (per student)—indeed sometimes no nonwage funding at all—than better-off and larger schools (Reinikka and Svensson 2002).

As mentioned above, the Zambia PETS includes a separate household survey. In addition to PETS providing filter coefficients for benefit incidence analysis, the combination of the household survey and PETS allows an innovative analysis of funding equity: gauging the extent to which public funding can be regarded as progressive or

regressive. The Zambia study finds, for example, that rule-based nonwage spending is progressive, while discretionary nonwage spending is regressive in rural areas and progressive in urban areas. Salary spending is regressive (Das and others 2002).

Conclusions

In countries with weak accountability systems, budget allocations are a poor proxy for services actually reaching the intended beneficiaries. PETS and QSDS are new tools for measuring the efficiency of public spending and analyzing incentives for and the performance of frontline providers in government and the private sector. Together these tools can provide a better understanding of behavior of frontline providers and, by linking them to other surveys, the relationship between providers, policymakers, and users of services can be studied.

Studies carried out so far point to ways to improve public sector performance. First, interventions can be targeted far better at vulnerable types of expenditures, such as nonwage recurrent spending, and at weak tiers in the public sector hierarchy. This ensures more accurate interventions and a more efficient use of resources. Second, efforts to improve service delivery must consider not just resource flows, but also the institutional framework and incentives. Adequate resources are not sufficient to guarantee performance if, as in Honduras, these resources migrate away from where they were intended to be used. Third, information dissemination, both to vulnerable tiers in the public hierarchy and to end-users, as done in Uganda, can be a potent way to mitigate problems arising from the information asymmetries that characterize most public sectors.

Notes

1. Information on survey design, sampling, implementation, and costs as well as sample questionnaires are available at www.publicspending.org (tools).
2. Another approach is to observe providers over a longer period of time on the assumption that the agent's behavior will revert to normal due to economic necessity. But this can be expensive, limiting the sample size. The study by McPake and others (1999), which used this approach, included only 20 health facilities.
3. Similar improvements are reported in Republic of Uganda in 2000 and 2001.

4. The health sample frame consists of 14,495 staff members in 873 workplaces. The education sample frame had 43,702 staff members in 9,159 workplaces. The total sample is 1,465 staff nationwide with 805 staff members from health and 660 staff members from education. These are clustered within 35 health establishments and 44 education establishments. The samples were stratified by type of facility and by type of employee. Population weighting was used to determine how many of each type of employee to draw from each type of facility. Two questionnaires were used for each institution from which individual staff members were sampled. One questionnaire was for the institution's manager and one was for each individual employee working in the sampled institution on the day of the visit. If the individual was not there, close colleagues filled in the required information about the employee.

5. In some cases, diagnostic PETs have been carried out with, say, 20–40 facilities (Government of Tanzania 1999, 2001), which is not enough for statistical analysis.

References

The word *processed* describes informally reproduced works that may not be commonly available through library systems.

- Ablo, Emmanuel, and Ritva Reinikka. 1998. "Do Budgets Really Matter? Evidence from Public Spending on Education and Health in Uganda." Policy Research Working Paper 1926. World Bank, Development Research Group, Washington, D.C.
- Alderman, Harold, and Victor Lavy. 1996. "Household Responses to Public Health Services: Cost and Quality Tradeoffs." *World Bank Research Observer* 11(1): 3–22.
- Alreck, Pamela L., and Robert B. Settle. 1995. *The Survey Research Handbook*, 2nd ed. Chicago: Irwin.
- Barnum, Howard, and Joseph Kutzin. 1993. *Public Hospitals in Developing Countries: Resource Use, Cost, Financing*. Baltimore: Johns Hopkins University Press.
- Bernheim, B. Douglas, and Michael D. Whinston. 1986. "Common Agency." *Econometrica* 54(4): 923–42.
- Castro-Leal, F., J. Dayton, Lionel Demery, and Kalpana Mehra. 1999. "Public Spending in Africa: Do the Poor Benefit?" *World Bank Research Observer* 14(1): 49–72.
- Chaudhury, Nazmul, and Jeffrey S. Hammer. 2003. "Ghost Doctors: Absenteeism in Bangladeshi Health Facilities." Policy Research Working Paper 3065. World Bank, Development Research Group, Washington, D.C.

- Das, Jishnu, Stefan Dercon, James Habyarimana, and Pramila Krishnan. 2002. "Rules vs. Discretion: Public and Private Funding in Zambian Basic Education. Part I: Funding Equity." World Bank, Development Research Group, Washington, D.C. Processed.
- Devarajan, Shantanayan, and Ritva Reinikka. 2002. "Making Services Work for Poor People." Paper presented at the African Economic Research Consortium, Nairobi, May. Processed.
- Dixit, Avinash. 1996. *The Making of Economic Policy: A Transaction-Cost Politics Perspective*. Cambridge, Mass.: MIT Press.
- . 1997. "Power of Incentives in Public Versus Private Organizations." *American Economic Review Papers and Proceedings* 87(2): 378–82.
- . 2000. "Incentives and Organizations in the Public Sector: An Interpretative Review." Revised version of paper presented at the Conference on Devising Incentives to Promote Human Capital. National Academy of Sciences, Irvine, Calif., December 17–18. Processed.
- Frankenberg, Elizabeth. 2000. "Community and Price Data." In Margaret E. Grosh and Paul Glewwe (2000).
- Government of Tanzania. 1999. "Tanzania Public Expenditure Review: Health and Education Financial Tracking Study. Final Report, Vol. III." Report by Price Waterhouse Coopers, Dar es Salaam. Processed.
- . 2001. "ProPoor Expenditure Tracking." Report by Research on Poverty Alleviation and Economic (REPOA) and Social Research Foundation to Tanzania PER Working Group. March. Dar es Salaam. Processed.
- Grosh, Margaret E., and Paul Glewwe. 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*. Washington, D.C.: World Bank.
- Hammer, Jeffrey, Ijaz Nabi, and James Cercone. 1995. "Distributional Effects of Social Sector Expenditures in Malaysia, 1974 to 1989." In Dominique van de Walle and Kimberly Nead, eds., *Public Spending and the Poor: Theory and Evidence*. Baltimore, Md.: Johns Hopkins University Press.
- Lindelöw, Magnus, and Adam Wagstaff. 2003. "Health Facility Surveys: An Introduction." Policy Research Working Paper 2953. World Bank, Development Research Group, Washington, D.C.
- Lindelöw, Magnus, Ritva Reinikka, and Jakob Svensson. 2003. "Health Care on the Frontline: Survey Evidence on Public and Private Providers in Uganda." Human Development Working Paper Series. World Bank, Africa Region, Washington, D.C. Processed.
- McPake, Barbara, Delius Asiimwe, Francis Mwesigye, Mathias Ofumbi, Lisbeth Ortenblad, Pieter Streefland, and Asaph Turinde. 1999. "Informal Economic Activities of Public Health Workers in Uganda: Implications for Quality and Accessibility of Care." *Social Science and Medicine* 49(7): 849–67.

- Reinikka, Ritva. 2001. "Recovery in Service Delivery: Evidence from Schools and Health Centers." In Ritva Reinikka and Paul Collier, eds., *Uganda's Recovery: The Role of Farms, Firms, and Government*. World Bank Regional and Sectoral Studies. Washington, D.C.: World Bank.
- Reinikka, Ritva, and Jakob Svensson. 2002. "Explaining Leakage of Public Funds." CEPR Discussion Paper 3227. Centre for Economic Policy Research, London, United Kingdom.
- . 2003a. "The Power of Information: Evidence from an Information Campaign to Reduce Capture." World Bank, Development Research Group, Washington, D.C. Processed.
- . 2003b. "Working for God? Evaluating Service Delivery of Religious Not-for-Profit Health Care Providers in Uganda." Policy Research Working Paper 3058. World Bank, Development Research Group, Washington, D.C.
- Republic of Uganda. 2000. "Tracking the Flow of and Accountability for UPE Funds." Ministry of Education and Sports. Report by International Development Consultants, Ltd., Kampala.
- . 2001. "Study to Track Use of and Accountability of UPE Capital Grants." Ministry of Education and Sports. Report by International Development Consultants, Ltd., Kampala, October.
- Rossi, Peter H., and J. D. Wright, eds. 1983. *Handbook of Survey Research, Quantitative Studies in Social Relations*. New York: Wiley.
- Somanathan, A., K. Hanson, B. A. Dorabawila, and B. Perera. 2000. "Operating Efficiency in Public Sector Health Facilities in Sri Lanka: Measurement and Institutional Determinants of Performance." Sri Lankan Health Reform Project (USAID). Processed.
- Wagstaff, Adam. 1989. "Econometric Studies in Health Economics: A Survey of the British Literature." *Journal of Health Economics* 8: 1–51.
- Wagstaff, Adam, and Howard Barnum. 1992. "Hospital Cost Functions for Developing Countries." Policy Research Working Paper 1044. World Bank, Development Research Group, Washington, D.C.
- World Bank. 2001. "Honduras: Public Expenditure Management for Poverty Reduction and Fiscal Sustainability." Report 22070. Poverty Reduction and Economic Sector Management Unit, Latin America and the Caribbean Region. Washington, D.C.
- Xiao, Ye, and Sudharshan Canagarajah. 2002. "Efficiency of Public Expenditure Distribution and Beyond: A Report on Ghana's 2000 Public Expenditure Tracking Survey in the Sectors of Primary Health and Education." Africa Region Working Paper Series 31. World Bank, Washington, D.C.