

**Part II.H
The Service
Delivery
Environment**

- 1. Access**
- 2. Quality of Care**
- 3. Integration of Services**
- 4. Gender Equity/Sensitivity**

THE SERVICE DELIVERY ENVIRONMENT

Rationale for Evaluating the Service Delivery Environment

The preceding sections in this *Compendium* have focused on background factors (status of women, the policy environment) and functional areas that support the delivery of reproductive health services (management, training, commodities and logistics, behavioral change communication, and research/evaluation). Each of these areas contributes to determining the **service delivery environment**. The service delivery environment is the situation prospective clients find when they seek services, both in terms of tangible factors (e.g., the physical plant, personnel, equipment, and supplies) and the intangibles (e.g., treatment received from the staff). The stronger the input from each of these functional areas, the better will be the services available to clients. Whereas the term “service environment” implies clinical services, the concept equally applies to BCC programs that are educational in nature.

Program evaluation in the early days tended to focus either on counting activities performed (e.g., the number of persons trained, brochures distributed) or results obtained in terms of service delivery (e.g., number of clients, number of clinic visits, number of supplies distributed), rates of use (e.g., contraceptive prevalence rate, breastfeeding rate), or long-term outcomes (e.g., fertility rates, infant mortality rate). Curiously, evaluators considered the actual functioning of the program itself somewhat of a “black box.” In general, evaluation did not probe the quality of the services and their availability to the population of the catchment area.

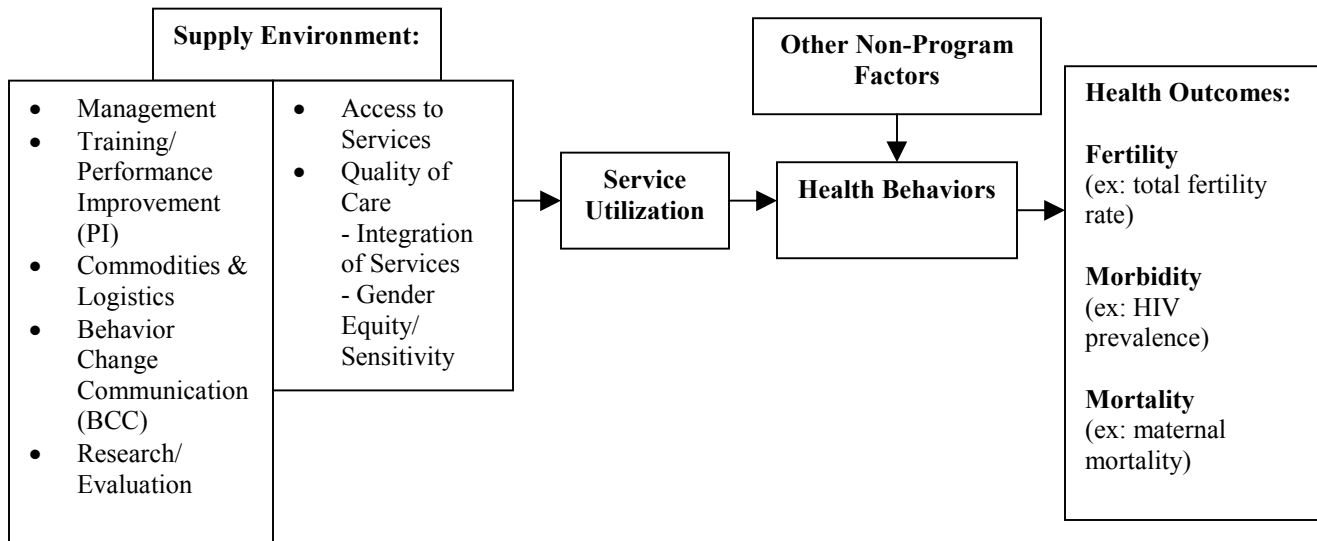
This situation has changed markedly in the past decade as evaluators have increasingly focused on the two defining characteristics of the service delivery environment: access to services and quality of care.¹ Of the

two, quality of care has been evaluated in far more detail than access, despite widespread recognition of the importance of the latter.

The rationale for evaluating access and quality is two-fold. First, evaluation of these topics serves to focus staff attention on the need to strive for improvements in these areas. Second, this type of evaluation measures the objectives the different functional areas – management, training, commodities and logistics, behavioral change communication, and research/evaluation – are working to achieve: better services and programs. Some have argued that evaluators should assess these functional areas in terms of behavioral change among clients/participants in the program or in the population at large. This argument fails to recognize that although the functional areas contribute to achieving program objectives, they do so by creating an improved service delivery environment, which in turn increases service utilization and desired behavior. This segment of the conceptual framework (Figure I.1) is reproduced in Figure II.H.1. Thus, it is logical to evaluate the functional areas in terms of their effect on the service delivery environment – in particular, their effect on access and quality of care (factors within their manageable control) – and not on more distant results also affected by numerous non-program factors. Although this *Compendium* does not discuss the study designs needed to make these linkages, many of the organizations working in this field have undertaken evaluations that link specific activities (e.g., in training and performance improvement) with the results they produce in terms of the service delivery environment.

¹ The *Handbook of Indicators for Family Planning Program Evaluation* (Bertrand, Magnani, and Knowles, 1994) includes a third element: program image. However, this aspect has received little attention from evaluators, and thus it is not included in this volume.

Figure II.H.1 Service Delivery Environment



Because of the importance accorded to quality of care, multiple approaches have been developed to measure it across different areas of RH. In the *Compendium*, we present indicators corresponding to three alternative approaches:

- The Service Provision Assessment (SPA): a comprehensive set of instruments containing a standardized list of indicators to cover multiple areas of RH: family planning, STI/HIV/AIDS, safe motherhood, and child survival;
- The Quick Investigation of Quality (QIQ): a “short list” of 25 indicators (a subset of the SPA) that measures family planning only; and
- Quality Assurance (QA): an approach that uses the indicators to measure improvement on the issues specific to a given clinic or to a set of clinics as part of a systematic process.

Integration of services is one aspect of quality related to “constellation of services” in the Bruce/Jain framework (Bruce, 1990). Many potential permutations (combinations) of services might be offered. However, given the limited experimentation to date with integration indicators, we present illustrative indicators for the integration of family planning with STI/HIV/AIDS only.

The final topic in this section on the service delivery environment involves gender. Although sensitivity to

gender issues has always formed part of good client-provider interaction, this subject has received renewed attention over the past decade. In addition, this section includes the related concept of gender equity in the organizational context. (The latter would work equally well in the section on management, but appears here instead because the two concepts are closely linked.)

Instruments for Measuring the Service Delivery Environment: Facility-based Surveys

Evaluators developed the tools for measuring the service delivery environment years after they had designed instruments for measuring results among clients (through program statistics) or among members of the population at large (through large-scale representative surveys).

The Service Availability Module (known as “SAM”) was the first attempt to systematically document the functioning of the service delivery environment on a national scale. The SAM, developed as an optional module to the Demographic Health Survey (DHS), was first carried out in Colombia in 1986. Since then, more than 40 SAMs have been implemented in conjunction with the DHS. The SAM measured the extent of access that the surveyed population had to reproductive and child health services. The typical SAM collected information from a group of community informants on distances to the nearest facilities of various types that offered reproductive and child health services. In most

countries, the SAM team then visited the nearest facility of each type within a specified distance to verify the distance information and to ascertain whether or not a facility offered certain basic services (e.g., immunization services for children, family planning). In a small number of countries, the SAM questionnaire was expanded to cover a more comprehensive inventory including information on the infrastructure, equipment, supplies, drugs, and staffing at the facility. Despite attempts to refine the SAM, it never gained the same importance as did the DHS household survey among policy makers or evaluators.

The Situation Analysis (SA) (Miller et al., 1997), also a facility-based survey, did capture the interest and attention of the international reproductive health community. Developed by the Population Council in connection with its operations research project in Africa, Situation Analysis was designed to measure the functioning and quality of care in family planning facilities. Situation analysis was conducted in approximately 21 African countries and in 16 developing countries from other regions of the world between 1989 and 2000.²

The Situation Analysis created widespread awareness of the value of facility-based surveys for evaluating the service delivery environment. As of 1999, the SAM and Situation Analysis were combined into a new set of instruments – the Service Provision Assessment (SPA) – that incorporates elements of both. The SPA is far more comprehensive than are either the Situation Analysis or the SAM, in that it covers family planning, safe motherhood, newborn care, child survival, and STI/HIV/AIDS with multiple data collection instruments. As of 2001, the SPA had been carried out in one country (Kenya) and was in the final preparatory stages in three other countries (Rwanda, Ghana, and Egypt).

Methodological Challenges of Evaluating the Service Delivery Environment

The challenges for evaluating access to services are quite different from those for evaluating quality of care (and its related components – integration of services and gender sensitivity). Indeed, each type of access presents different methodological issues. Thus, we present the challenges for measuring access as part of the Purpose and Issues section for each indicator of access.

As for quality of care, most of the indicators are derived from one of four sources: facility audits, observation, client exit interviews, or review of medical records/clinic registries. The methodological issues that surface in connection with the measurement of quality relate to the concept of quality, data collection techniques, or sampling bias.

- **The opinions of actual clients may differ from that of international experts on “what is important” in terms of quality of care.**

Privacy and confidentiality, often major concerns of clients, may carry far more weight in the client’s decision to use the services than other items on the list do. International experts who define the items on the instruments generally try to encompass a large range of issues, and in doing so may give less weight to the key issues for clients. A similar problem may occur when different stakeholders disagree about which indicators are most important. Developing consensus among stakeholders becomes an important first step in designing the evaluation.

- **Data collected by two or more observers may have low inter-rater reliability.**

If observers are well-trained, direct observation is generally the best method to measure compliance with standards of care. However, inadequate training of observers can result in low inter-rater reliability that seriously compromises the validity of the findings. To improve inter-rater reliability, multiple observers (multiple may refer to only two) should watch and rate the same client-provider sessions in a role-play situation or through videotapes of actual counseling sessions. Those responsible for training the observers can compare results and provide feedback as to how to code items with low inter-rater reliability.

- **Providers may perform better than usual when observed (i.e., the Hawthorne effect).**

The Hawthorne effect refers to the tendency for persons to perform differently (usually better) when they know they are being observed (Rossi, Freeman, and Lipsey, 1999). Thus, the presence of an observer in the

² Precisely quantifying the number of Situation Analysis studies conducted to date is difficult, because several countries have adapted or used the SA on their own.

room during the counseling and clinical sessions may have caused providers to be especially attentive to their duties. Although some providers will perform better while being observed, the experience to date suggests that observation still effectively identifies shortcomings in provider performance. If a provider does not know a certain fact or is incompetent at a certain procedure, the presence of the observer will not change that reality. And if a provider is rated as discourteous during an observed session, one can surmise that he/she is equally rude, if not ruder, when not observed.

- **Direct observation is not always feasible.**

Although the direct observation of care is the preferred method to measure level of compliance with clinical standards, in reality it is seldom used for monitoring maternal care because of the following limitations: emergency care is difficult to observe; deliveries happen often at night when observers are absent; deliveries can last many hours and is thus very time-consuming for an observer; and the opportunity to observe rare events is low. As a result, a review of medical records often becomes the only feasible means to assess health worker performance according to clinical standards. Unfortunately, medical records are often incomplete or contain insufficient information. Using the mystery client approach to collect data holds promise for many, but not all (e.g., surgical or invasive procedures) scenarios.

- **Clients may not accurately remember the events that took place during the counseling and clinical sessions (recall bias).**

Clients may not remember exactly what occurred during the session with the provider. The reliability of their responses may vary with the provider action in question. For this reason, the client exit interview included in the QIQ contains a limited number of largely factual questions (e.g., which method did you want when you came here?). The results of the QIQ field test show a relatively high degree of consistency between the reports of the observer and of the client as to what occurred during the counseling session, and thus suggest that client recall on these items was satisfactory (Bessinger and Bertrand, 2001).

- **Clients may report that they are satisfied with services, even if they are not (courtesy bias).**

Studies have shown that clients are likely to report that they feel satisfied with the services they have received and will not speak negatively about the clinic or clinic staff during exit interviews (Williams, Schutt-Aine, and Cuca, 2000). Hence, results from the client exit interview tend to be positively skewed on the question of satisfaction. Interviewers should be trained to emphasize to respondents that their responses in the interview will not jeopardize their care at the clinic. Where funds permit, persons not associated with the clinic can usefully interview clients in the client's own homes (away from the clinic) to reduce bias. Similarly, interviews (one-on-one or in focus groups) with non-users of the service can reveal more about their perceptions of the service.

- **The unit of analysis differs for the different data collection instruments.**

The challenge of this indicator applies in particular to the SPA and QIQ. The unit of analysis for both the client exit interview and for the observation is the client; however, the unit of analysis for the facility audit is the clinic. Evaluators have two possible ways to address this discrepancy in sample size: (1) carry out the analysis at the client level, and assign the same facility audit measure to each client who visited a particular facility; or (2) carry out the analysis at the facility level by averaging client-level results. Unfortunately, each scenario has its drawbacks. If evaluators collect data at the client level, they will find no variation in estimates for the clients who attended a given facility. If the evaluators aggregate data to the facility level, they may lose important information, and thus decrease the precision of the estimates.

- **Evaluators have difficulty appropriately estimating the sample size when the client volume differs substantially for the different reproductive health services to be evaluated.**

Often the evaluator wishes to collect data on different RH services within a given facility or within a set of facilities. However, because the client volume may dif-

fer by service (e.g., family planning versus antenatal care in sub-Saharan countries), the evaluators will have difficulty establishing a sampling strategy that will yield the appropriate number of cases for evaluating both services. To address this challenge, evaluators can collect information on the client volume in each service and can weigh the data during analysis.

- **Standards defining quality of care may not be available or consistent across countries.**

The illustrative indicators presented for evaluating the quality of maternal and neonatal care services as part of the QA approach require standards or guidelines as a reference for measurement. Some countries lack clinical standards for their programs or they are not evidence-based. In this case, evaluators can use standards developed elsewhere and relevant to their situation or they can develop local standards from scratch. The latter is very time consuming, but ensures that the evaluation is more relevant to the local context (i.e., it is feasible with the resources available). The evaluation exercise provides a good opportunity to address the issues of availability of evidence-based standards.

If the purpose of the evaluation is to compare quality of maternal care among countries, then standards need to be the same. Some maternal care standards (such as the use of a partograph as an early warning of insufficient uterine action and of cephalopelvic disproportion) are universally accepted. However, standards may differ by country (e.g., the number of recommended prenatal care visits). When standards are inconsistent, and evalua-

tors lack scientific basis to validate some and to reject others, they can still compare by modifying the indicator (e.g., “percent of women who completed three prenatal care visits on time,” to “percent of women who completed the recommended number of prenatal care visits”).

- **Special (periodic) studies of quality fail to address the need for regular monitoring.**

Useful as periodic studies of quality (such as the SPA and IQI) are to program managers, they fail to provide an ongoing measurement of performance. Relatively few programs have regular, ongoing systems to monitor quality systematically. The Latin American Maternal Mortality (LAMM) system, presented in connection with the Quality Assurance approach, is an exception, in that it uses existing data routinely collected through the health information system or medical records. However, this type of effort requires continuous training of health facility staff to assure the quality of data collected.

The indicators in this section are organized as follows:

- Access;
- Quality of care;
 - SPA (four areas of reproductive health);
 - IQI (family planning);
 - QA (illustrative examples for maternal and neonatal care);
- Integration of services; and
- Gender equity/sensitivity.