PN-ABJ-972

The Cono Sur PRICOR II Project
Subagreement 87/11/3300 to
U.S.A.I.D. Coop. Agrmnt. DPE-5920-A-00-5056

PERU COUNTRY STUDY

Final Report

Volume 1

SYSTEMS ASSESSMENT MODEL

WITH

INSTRUMENTS AND INDICATORS

The PRISM Group
Projects in Rural Industry, Agriculture, Science and Medicine, Inc.
10480 Little Patuxent Parkway
Suite 500, Columbia MD 21044
Telephone: (301) 730-4750
FAX: (301) 740-5603

The Cono Sur PRICOR II Project
Subagreement 87/11/3300 to
U.S.A.I.D. Coop. Agrmnt. DPE-5920-A-00-5056

PERU COUNTRY STUDY

Final Report

Volume 1

SYSTEMS ASSESSMENT MODEL

WITH

INSTRUMENTS AND INDICATORS

The PRISM Group
Projects in Rural Industry, Agriculture, Science and Medicine, Inc.
10480 Little Patuxent Parkway
Suite 500, Columbia MD 21044
Telephone: (301) 730-4750
FAX: (301) 740-5603

The work upon which this presentation is based was performed under a subagreement with the Center for Human Services under its Cooperative Agreement No. DPE-5920-00-A-5056-00 with the U.S. Agency for International Development.

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

1. Introduction

a. The Assessment of Primary Health Care (PHC) Service Delivery: Needs and goals

A key problem in Primary Health Care (PHC) programs in most countries is the existing management of the public sector health organizations charged with delivering them. Specifically:

> ***Assuring the quality of services delivered to a health system's users is essential if PHC programs in the Third World are to have a significant impact.***
>
> ***The effectiveness with which health care organizations are managed makes a difference in how well they deliver their services.***
>
> ***Even with the best of intentions, the existing capacity of most public sector health systems to manage a wide-flung network of PHC service outlets is extremely limited.***
>
> ***Therefore, these "peripheral" service outlets have great difficulty delivering services, such as PHC, that are of high quality.***
>
> ***and***
>
> ***Therefore, PHC programs in the Third World are having less than the desired impact.***

Stated in practical terms, the problem is how to ensure that each peripheral service unit maintains an adequate standard of performance quality in the face of severe constraints on supplies, transportation, communications, and trained personnel. Such constraints, themselves, are certainly a nother key problem in PHC but they are not, in general, the important factors underlying ineffective management.

We believe that the lack of reliable information and managerial expertise are more important causes. The current project addresses the first directly and the second indirectly. In order to control an activity, it is first necessary to be able to measure it. That is the primary goal of the PRISM PRICOR project as we understand it.

The operations manager's perspective on the assessment of service delivery, be it primary health care or hamburgers, can be stated simply: *"How can I assure that the clients of my unit are getting the best service possible to meet their needs when they want it and how they want it?"* Assuring the quality of the service delivery process at the point of contact with the client is essential if primary health care programs in the Third World are to have a significant impact.

It seems obvious that a large part of the solution to problems with service delivery must lie with the operations managers themselves -- health center

directors, program supervisors and others. They must constantly measure their unit's performance and manage it so as to maintain an acceptable level. But, if pressed to manage effectively, these individuals will, in turn, be pressing for better ways to assess performance and seeking consistent guidelines on how to use such assessments to correct problems and improve performance.

It is surprising, therefore, that so little systematic effort has been made to develop reliable and valid ways for measuring and assessing either PHC services or the design and management of peripheral service organizations.

It seems almost as if the "simple" health care content of PHC has been confused with the system for delivering PHC. This is a mistake. The latter, since it is highly dispersed, variable, and incentive- and resource-limited, is anything but simple.

Traditional evaluation plans based on mainly input and outcome indicators reveal surprisingly little about whether a health system is actually functioning as it is supposed to. Ascertaining the "cost-effectiveness" of a poorly managed service without a detailed consideration of process, for example, forces the analyst to accept the service "as it was delivered" -- as an externality to the analysis.

The current approach to evaluation sometimes provides information about the "what" but rarely give any insight into the "why". It leaves managers, at both strategic and operations levels, to their own devices to find solutions to problems: which is precisely the hardest part of the process.

There is a need to focus more carefully on the process of health care and describe, in concrete terms, how important activities can be measured. The objective is to describe how service delivery personnel actually provide child survival services, addressing quality of care, counselling, outreach, supervision, training, logistics, management information and other elements of implementation.

This is the raticnale for the PRICOR systems analysis approach and, thus, forms the basis for The PRISM Group's effort in the PRICOR II Cono Sur Project. The terms of the subcontract under which this work has been done also called for PRISM to produce, if it chose, an alternative to the existing PRICOR approach to systems analysis. We have exercised this option and have developed the PRISM Systems Assessment Model, or SAM, which is described below.

In both the SAM and the original PRICOR model, characterization of service delivery is not to be accomplished by unstructured expert observation; it is to be done, rather, through a well-defined methodology that can be applied by regular operations managers. This suggests the development of standardized, replicable analytical instruments that can be employed widely.

The development of such instruments is a sophisticated process but the ultimate products must be relatively simple to use. The final "kits" will consist of proven

indicators, step-by-step application protocols, and clear suggestions about what approach to take as the analysis proceeds and problems are encountered. This is the practical goal to which the effort described in subsequent sections has been aimed.

In the process, The PRISM Group, has also carried out a systems assessment of actual service delivery in two PHC programs (immunizations and control of diarrhea) of the Ministry of Health in Lima's Cono Sur. The results of these assessments are also part of this final report (Volumes 3 & 4). Nevertheless, and with rare exceptions, it isn't surprising new insights into hitherto unrecognized problems that have been sought from this effort, *but a practical methodology that will routinely and unambiguously pinpoint common problems and deal with them before they seriously threaten the efficiency or effectiveness of service delivery.*

The PRISM Group is convinced, after an intensive review of existing approaches, that such a practical methodology is not immediately at hand but that it could be developed from what is available if the willingness exists to make an all-out effort to do it.

b. Summary description of the Ministry of Health, the Cono Sur and the PRISM PRICOR II Cono Sur Project

The PRISM Group has undertaken a country study in Peru to assess primary health care (PHC) programs under the terms of a subcontract to U.S.A.I.D. Cooperative Agreement #DPE-5920-A-00-5056-00: Primary Health Care Operations Research (PRICOR) Project. The *PRISM Cono Sur PRICOR II Project* was done in collaboration with the Peru Ministry of Health and focused on PHC service delivery in 14 health centers located in the southern peri-urban fringe (the "Cono Sur") of Lima, Peru. Specifically, the project carried out assessments of the immunization and diarrhea control programs in these health centers.

1) Basic structure of the Peruvian Ministry of Health (PMOH)

The standard organizational chart of the Peruvian Ministry of Health of Peru, or PMOH, shows a typical pyramidal arrangement which portrays the formal power arrangements within the system

The central ministry sets policy, allocates the budget, establishes national coordinators for each program, and appoints the directors of the next level down, the UDES (Unidad Departmental de Salud). There are 28 such departmental units and they function with a reasonable amount of autonomy.

The UDES' management role is still poorly defined, however, in part because of an on-going political movement to create regional governments in the country which will, it is hoped, supplant many central government functions in coming years -- including health care.

Immediately below the UDES is a variety of intermediate administrative structures: the UTES (Unidad Territorial de Salud) and the EEP (Entidad Ejecutivo Presupuestal).

The UTES is basically a mini-UDES used in rural departments with large areas and poor communications to centralize the necessary core of support activities at a level more accessible than the UDES headquartered in the departmental capital.

The EEP is a budgetary entity with, supposedly, little or no managerial authority. Its main historical justification has been to maintain budgets for true peripheral service units such as health centers separate from those of the hospitals.

In the past, the Ministry channeled all resources via the area hospitals with the result that virtually all resources stayed at the area hospitals. The EEP concept attempts, fairly successfully, to guarantee that a reasonable share of resources do indeed go to health centers and health posts.

At the bottom of the pyramid, one finds the support hospitals on one branch and health centers and health posts on the other. There are currently over 600 health centers and 2000 health posts in the Ministry system. Reference hospitals and other facilities managed directly from the UDES or central ministry.

Health posts tend to be administratively attached to the nearest health center though some may report directly to the EEP, UTES or UDES if they are physically closer to the latter than any health center. The health centers and their associated health posts are what is referred to as "peripheral services".

Programs, such as control of diarrhea, immunizations, and so forth exist as separate entities with identifiable budgets and independent coordinators down to the health center level. Once there, however, most program activities are often intermixed and handled on a group basis by the available staff.

The PMOH is the institution mandated with fostering community health services and primary health care. In light of this mandate, Ministerial Resolution No. 048-87-SA/DM of February 6, 1987, defines the following programs for the PMOH in the current fiscal year (Jan-Dec, 1987):

HEALTH OF WOMEN AND CHILDREN
- *Attention during pregnancy*
- *Attention to well children*
  - *growth and development monitoring*
  - *breast feeding*
  - *supplementary feeding programs*
- *Expanded Program of Immunizations*
- *Control of Diarrheal Disease*
- *Control of Acute Respiratory Infections*
- *Attention to school-age children*
- *Oral health program*

HEALTH OF ADULTS
Family planning
Control of Tuberculosis

FOOD SUPPLY AND NUTRITION
Food supply program

ENVIRONMENTAL SANITATION
Basic sanitation
potable water program
rural sewage
Protection of food
Environmental protection
Malaria control
Occupational health

MEDICAL SERVICES
General medical services
Quaternary health services

SUPPORT SERVICES
Development of administrative services
Development of human resources
Production of medical/health supplies
Physical plant

FUNCTIONAL INTEGRATION
Integration of PMOH and Social Security

2) Geography of the Cono Sur

The Cono Sur is an area of peri-urban slums with a population of approximately 700,000 persons in three political districts (Villa El Salvador, Villa Maria de Triunfo, and San Juan de Miraflores) in the southern urban margin of metropolitan Lima. The region, 150 square kilometers in area, was first inhabited in 1950 when large numbers of immigrants from the sierra (high andean plateau) began arriving in Lima to occupy diverse sectors of what is now one of the three districts of the Cono Sur.

The Cono Sur is located in low hills and plains stretching inland from the ocean and consisting entirely of desert sand at an altitude no greater than 450 meters above sea level. The climate is typical of tropical deserts with an annual rainfall of less than 150 mm and a high relative humidity. Temperatures range from highs of 35-37°C to nightly lows of 12-14°C.

Population density in the Cono Sur averages 4,500 per square kilometer (1986); the distribution by age group and district is given in Table 1. Approximately 70 percent of the population live in "permanent" housing of wood or brick, and of those, about 65 percent have sanctioned water and sewage connections. Unemployment is officially estimated to be 15 percent, but another 45 percent

report themselves to be underemployed. Approximately 25 percent of the work-force is engaged in construction activities of one form or another.

The 35 percent of the population which lives in recent "invasions" is housed in shanty-like one and two room constructions of woven matting or cardboard. These recent invasion areas, all less than five years old, are usually without household water or sewage service. Most households in the three districts of the Cono Sur do have access to electricity. Households in even the most recently settled sites quickly acquire electricity by illegally tapping adjacent power lines.

3) Health statistics for the Cono Sur

The health situation in the Cono Sur is typical of that in many peri-urban, relatively new "invasions". Because of easier access to health care during medical emergencies and an improved economic situation, people living in the Cono Sur actually experience improvements in morbidity and mortality compared to their previous lives in the Sierra.

Volume 4 of this final report series presents an epidemiologic surveillance for 1988 which was done by one of the PRISM PRICOR Project team members, Dr. Wilfredo Gutierrez. These data have been obtained, tabulated and reported by Dr. Gutierrez, epidemiologist of the Hospital de Apoyo Maria Auxiliadora, with the collaboration of Dr. William Spira and others from the PRISM PRICOR team. Dr. Gutierrez has produced a similar annual report for the past four years.

4) Health services in the Cono Sur

Health care services in the Cono Sur are provided by a poorly coordinated, mixed system of government, private sector, and PVO-supported facilities. The PMOH system consists of 10 basic health centers, 4 maternity/infant hospitals (12-20 beds), 60-70 functioning health posts (depending on the definition of "functioning"), and Hospital de Apoyo "Maria Auxiliadora".

There are also approximately 40 PVO-run health posts (most run by the Catholic Church), two social security (IPSS) "policlinicos", one clinic run by the Army, and approximately 75 private medical practices. About 50 percent of these consultorios are operated in the afternoons and evenings by physicians who also work "full-time" (until 2 p.m.) for the PMOH. An unknown, but significant, percentage of preventive and curative health care services is also provided directly by the numerous pharmacies distributed throughout the region.

Until 1986, the health centers and posts were directed through the Hospital Maria Auxiliadora. During this period, resource allocation has been skewed toward the hospital and larger health centers, and communities have, not unexpectedly, channeled their demand to those facilities with the best resources. As a result, the periphery has suffered all the problems characteristic of an under-supported

effort: poor utilization, faulty triage of highest risk individuals, uncoordinated allocation of scarce resources and a lack of planning and evaluation of on-going activities.

In 1986, the Cono Sur was placed under the direction of the UDES-Lima Sur, one of the four UDES in Lima. with the expectation that resources would be allocated in a way that strengthens community services and community participation. However, the UDES-Lima Sur directorate had no protocols to follow in setting up its infrastructure. It has been given heavy responsibilities with little time for planning of reasonable or even minimally functional, organizational models for the complex system envisioned by the national planners.

The major health problems in the Cono Sur fall into areas which are being given high priority by both the PMOH and USAID: acute respiratory infections, diarrheal disease, nutrition/supplementary feeding, immunizable diseases, family planning, and tuberculosis. Malaria, whose control is a high priority in other regions of Peru, is not a problem in the Cono Sur.

5) Human Resources

The Cono Sur has some 250 paid health workers (called either Sanitarios or Auxiliares de Enfermeria and referred to below as CHWs) who work in the health posts and centers. Taken together, they account for three times more person-contact time with the community than the 60 physicians and nurses who work largely in the health centers. They are viewed as the front-line workers for the PHC programs, and are responsible for delivering basic PHC services including immunizations and ORT. The health worker's role extends beyond service delivery to health education and promotion tasks since the PMOH has no program of volunteer health promoters.

Health workers are officially charged with executing the bulk of the Ministry's public health program. In their capacity as the interface between the Ministry and the public, they must satisfy the numerous demands of program managers for each of the Ministry's public health programs.

c. Summary description of the Systems Assessment Model

We view "systems assessment" from the perspective both of the health systems manager as well as that of the analyst. The following definition is intended to capture this concern for both sound theory and practical utility:

PHC systems assessment is:

the systematic and selective measurement of structure, process and outcome indicators encompassing the performance of primary health care service organizations;

evaluated within

an analytical framework which specifies relevant and testable relationships between the three classes of indicators;

and directed toward

the identification of effective actions that can be taken by operations management to correct deficiencies or otherwise improve individual or organization performance.

Our purpose is to develop an analytical framework, a set of measurement instruments and a process for implementation that are both scientifically valid and practically useful at a management level for assessing and improving primary health care systems on an on-going basis. Many of the ideas and approaches that are outlined below have been developed by others and, to them, we owe a debt of gratitude and the responsibility of recognizing their contributions.

The SAM is intended to become an organic part of the operations management process. It is intended to create a context in which centers that are performing poorly in any specific aspect can learn from and model themselves on centers that are performing the same work well. This process is to continue iteratively as the health centers improve their performance.

The ability to do this comes from the specificity and the immediate relevance of the measurements that are used. It is not our intention to initiate change based on a priori thinking. The SAM calls for changes, not on the basis of someone's theory, but on the basis of empirical evidence gained from the health centers themselves.

The SAM is complex and it is particularly complex in its operational form. This does not mean that it is impractical. It does mean that its successful

implementation requires a carefully thought-out process model. We have been developing exactly such an implementation context using the paradigm of the "Master Teacher" that has been used with impressive results in the field of education. With this as a basis, we have designed a sustainable SAM-based quality control program with a high potential for being integrated into the Ministry's routine operations management system.

2. Frames of Reference

Analysis is shaped by the experiences and beliefs of the analysts, and subsequent debate over a project and its findings will be more coherent if the premises arising from this combined set of experience and belief are laid out. What follows in this section on frames of reference is a listing and justification of our premises.

a. General premises underlying the project

1) The value of existing public sector health organizations

We begin with a declaration of our belief that the existing public sector health organizations operating in the Third World are worth improving. It is not uncommon to hear that, for good health care to truly come to most underdeveloped countries, the established order of Ministry-run health care will have to be either abolished or bypassed.

We do not believe this to be true, either in principle or as a practical option. Public sector organizations will continue to be the dominant provider of health services in the Third World in the foreseeable future. On the other hand, so little systematic research has been done on the workings of government-run primary health care systems that it is, in any case, premature to declare them dinosaurs.

In our group's experience with primary health care organizations in a number of countries, we have encountered health centers that deliver excellent services on limited budgets and, even in the bad ones, service personnel who, individually, perform extremely well in difficult circumstances. There is an extraordinary amount of variation in health care service delivery in these systems, both between and within centers. We need to build from this existing capacity for good service in these organizations, identifying centers of excellence within the system and using them as role-models in training and motivating their peers to equal their performance.

2) Models of successful service delivery can be found within the organization itself

If models of successful service delivery already exist in the system, there is little need for prescriptive solutions originating from the outside. Prescriptive solutions can cause dissension, and very frequently are ignored. When the solution is, instead, based upon empirical evidence that participants in the system have helped to define and collect, the resulting solution has a much greater likelihood of being accepted.

3) The value of systematic knowledge

For knowledge in the field of PHC delivery to advance, some systematic approaches must be taken. Comparison is at the heart of knowledge and, therefore, of analysis. Information must be generalizable to have any real value in the acquisition of knowledge; unless it can be compared with information collected elsewhere, it is merely anecdotal.

As a corollary to this, we believe that while differences between countries are large, there still exists a large body of generalizable knowledge that represents the core system of any and every PHC delivery organization. The more systematically these experiences are collected, recorded and analyzed, the more applicable the knowledge gained from the experience will be.

4) The value of information as a change agent

An important premise to our effort is that information catalyzes change. To some extent we see our goal as getting the right information to the right people at the right time. The direction of analysis lies in identifying the three "rights". In this context, analysis is applied research in the truest sense of the word - the goal is completely pragmatic, but since we do not know a priori what information is most important to collect (either from the manager's perspective or in relation to certain goals or outcomes), research is needed to determine what information is most important.

We also believe that it is justified to have a high level of concern for the reliability and validity of the indicators and data collection instruments to be used. We cannot simply assume that we are measuring what we intend to measure; we have to have some basis other than just saying that this is so. This is discussed in some detail in subsequent sections.

5) The need for involving the organization's own people

Health care service delivery research can not be done in a vacuum.

> ***"Unlike research in the physical and biological sciences, people within the organization, including the managers, (must be) an active part of the research process." [Shortell & Kaluzny, P. 506]***

This is true even if the researchers have no intention of improving or changing the system. For those of us working in evaluation and assessment, who do hope to improve the working of the system, it is absolutely essential. This in no way compromises the "purity" of the analysis: if anything, it enhances it. It forces the analyst to construct a conceptual model of the organization that is understandable to the managers in the system.

This model then provides a basis for interaction, it ...

> ***... provides an opportunity to transform the role of manager from that of an advocate of untested solutions to often imaginary problems to that of a diagnostician who identifies problems for action, considers alternative courses, and uses assessment data to evaluate options as they affect the organization. [Shortell & Kaluzny, P. 506]***

6) Focus on the needs of operations management and the workers themselves

We have chosen the health center directors as the primary constituency for the PRISM SAM. We believe that the primary constituency of systems analysis should be the managers of the key operational units in the PHC service delivery system who need tools to monitor their unit and identify areas needing improvement. In Peru, as in most other national public sector health systems, the key unit is the health center with its associated health posts.

Effective management at the health center level is, by far, the most important determinant in successful service delivery. Furthermore, systems analysis aimed primarily at this level can be "aggregated" with relative ease to serve the needs of higher level management as well. By concentrating most of our attention on the process of service delivery, we aim to provide a set of useful tools to health system managers to use in their efforts to promote better health care.

The level of analysis employed in the PRISM SAM is the health center, the lowest level in the PMOH system that provides integrated PHC services as an independently functioning operational unit. Even though the health center is the focus, the SAM must take into account certain relationships between the health center and its supra-organization and between the health center and the community it serves. Flow of resources to the health center and the geographical location of physical facilities are obvious examples of such relationships.

b. Basic concepts of the general Systems Assessment Model

1) Systems theory and boundary functions

The standard organizational model of the PMOH and its programs looks like a hierarchy of simple linkages in which reports and requests for support flow upward and orders, resources and demands for more information flow downward. Much thought and effort within the Ministry, with and without technical assistance from the outside, has gone into organizing itself to better fit the premises of such a model in order to become more efficient and effective.

The common difficulties with this model stem from its tendency to:

1) focus most attention to the top of the pyramid, hence on policy and strategic planning rather than operations;

2) lead one to assume that relationships between system elements are more directly linear than they really are (e.g., that program inputs will lead to easily predictable program outputs);

and

3) channel efforts to resolve performance problems into yet another reshuffling of the organization chart.

Moreover, this is a poor model for dealing with performance of the service functions of the organization at its periphery. It gives no sense of the organization operating in its societal niche, existing in order to serve certain needs of the society within which it functions. This, in other words, is not at all the client's view of the organization.

Our consideration of characteristics that ought to be part of an adequate organizational model has led to the incorporation of a number of principles that form much of the basis of contemporary systems theory.

The first is that systems exist partially or wholly within systems. Health centers and health posts exist as systems within EEP's and UTES, which in turn exist as systems within UDES, which exist as systems within the Ministry, which exists as a system within a larger social system.

Moreover, each system exists only partially within the next larger unit since each has direct contact with other systems, as well, that are clearly outside the larger unit. This contact could only occur if some part of the smaller system were "outside" the larger system.

The most important example of this is the peripheral services unit which, though within super-organizational units in some senses, clearly forms an important contact with the clients it serves in other senses.

A second principle is that all such systems are open: that each system receives and discharges something from and to other systems -- usually those that are contiguous.

Systems must be open to some extent to survive; genuinely closed systems cease to function after a period of time. But systems cannot be wholly open either; there must be some measure of discrimination and filtering of what goes in and out or there is no system, only chaos.

A system "is a set of components (also systems) interacting with each other and enclosed by a boundary which selects both the kind and rate of flow of inputs and outputs to and from the system" (Berrien, 1983).

A third principle is that the functions of a system are dependent on its structure. This is not to say that a variety of different system behaviors cannot be associated with what appears to be the same structure, only that structure sets a limit on such behaviors and makes some more likely than others in response to specific conditions.

A fourth principal is that a system functions so as to maintain critical states at an optimal equilibrium irrespective of environment. The science of cybernetics (or control-information theory) introduces a number of concepts that are important to understanding how systems actually maintain control of their internal functions. These include feedback loops, setpoints, attenuation and amplification, and other ideas which are crucial, not just to the design of efficient organization information systems, but to the implementation of effective management, as well.

A corollary to the fourth principal is that an organization such as the PMOH is a complex of flexible "structures" that flow and shift around unavoidable perturbations in order to reestablish the "really important" equilibria: most of which are unknown or unproven. Unlike the rigid, linear model of the traditional hierarchy, such a view should keep us from being too smug in our belief that interventions will produce straightforward, predictable results.

Finally, the fifth principal is that a system functioning as the boundary, or peripheral, component of a larger system determines the real-time functioning of that larger system. This is the theoretical reasoning behind the commonsense wisdom that a service organization of whatever kind will succeed or fail depending on how well it manages its point of contact with the client or customer.

This fifth principle gives strong support to our decision to focus almost exclusively on the process of health care service delivery at the periphery of the PMOH. It is not just that this is the point at which the stated goals of the organization are

actualized (though that is reason enough to be concerned with it), but also that what goes on at the periphery has a profound effect on the functioning of the PMOH all the way back to its center.

2) Basic approach to assessment: structure-process-outcome

Our approach to integrating the four organizational frames of reference and the systems theory just described into a systems assessment model borrows heavily from the U.S. domestic literature on management, organization theory and behavior, information systems and cybernetics, education, and psychometrics.

The work of Donabedian (1966; 1982) on quality of care plus that of Van de Ven and Ferry (1980) on organization assessment, and Shortell and Kaluzny (1988) on health care management have been particularly noteworthy as sources of much of what has been incorporated into the theoretical framework of the current SAM.

A fundamental decision that we made in giving the SAM an operational definition was to use the three classes of indicators defined by Donabedian (1966) as the basis for the development of our instruments and assessment protocol. These, as was indicated in the definition of systems assessment given previously, are structure, process, and outcome:

> Structural indicators are those organizational attributes which determine an organization's potential or capacity for effective work (e.g., proportion of registered nurses, average educational level of health auxiliaries).
>
> Process indicators have dimensions of quality and quantity and relate to members' activities in carrying on their work. Process indicators apply both to direct services such as care and patient education and to support services such as supervision, logistics, etc.
>
> Outcome indicators refer to the status of the objects on whom the work is performed. Changes in characteristics that can be attributed to the work performed upon them can be interpreted as impact. The most common examples of outcome indicators are morbidity and mortality.

3) The definition of quality in service delivery

The health of any population can be thought of in terms of the following equation:

Population's Health = Function (Health Programs + Epidemiological Factors)

Systems assessment is aimed at evaluating an on-going health program against a set of established norms and goals. Its focus is almost exclusively on the "Health

Programs" factor in the above equation and leaves most epidemiological factors aside. The outcomes that are relevant to a systems assessment are not measures of a population's health but rather measures of system performance that proceed directly from program goals and that are expressed in operational terms.

This emphasis does not ignore the importance of epidemiological factors where they are relevant. For example, to explain health in a community one needs to know a lot about the community, its activities, and its environment. Such information is also needed for the development of new and better approaches to the prevention and treatment of disease and in the design of health programs to implement such approaches. Likewise, if the norms and goals of a health program are, themselves, under assessment, then epidemiological factors must be taken into account.

We do not believe that epidemiological factors are as relevant, however, when the focus is on quality of care. Donabedian argues that measures of process often serve as better criteria for assessing quality of care than do measures of outcome. We agree with this point of view, particularly when the criteria are to be used to aid decision-making at the operations management level. This priority placed on the process of service delivery and on the working of the peripheral health care organization does not, in any way, deny the importance of outcome (or even structural) criteria as elements in a complete understanding and assessment of service delivery.

Process and outcome measures are not mutually exclusive; they are, in fact, complementary. But as quality of care criteria, outcome measures have a value and application quite different from that of process measures.

Process measures are better suited to situations in which management control of performance is the primary issue. Drawing on the vocabulary of cybernetics, we might say that standards of good practice can be established as "setpoints" to which actual performance can be compared. Monitoring performance provides "feedback" that can be compared to the setpoint in order to establish when a deviation from the desired state exists. This provides the basis for management action.

Outcome measures, in this context, are the criteria used to establish the setpoints. If an existing standard of good practice is found to be associated with undesirable outcomes, than that standard is modified or replaced. This changes the setpoint but does not alter the essential loop of monitoring-feedback-management action just described.

This is a reasonably accurate picture of how performance changes really are introduced in practice. Usually someone notes that treatment failures or undesirable outcomes seem to be associated with a normatively good treatment and begins to question the assumed relationship between that treatment and the desired outcome. He/she then studies the matter in a scientifically rigorous

fashion and, if the original doubt is ultimately confirmed to the satisfaction of enough of the professional community, the standard of good practice changes.

Nevertheless, the basic relationship between the standard itself and the management of performance based on that standard remains unchanged throughout. In other words, outcomes are reasonable criteria for establishing what quality of care is but process assessement provides better criteria for establishing if quality care is being delivered according to existing standards.

The normative approach now used in establishing standards of good practice in primary health care only rarely establishes a trustworthy linkage between process and outcome in terms that are measurable. The result is a heavy emphasis on outcomes as evaluation criteria on one hand and on poorly tested statements of what constitutes good practice on the other.

4) Differentiating the SAM from other approaches

Primary health care service delivery has been studied by a variety of approaches which can be distinguished from systems assessment, as we use the term. The most important of these approaches include traditional systems analysis and operations research. We will discuss only the former in detail in this section.

Operations research has a variety of definitions ranging from true industrial operations research with its sophisticated mathematical modelling to the simplified definition used by the original PRICOR Program (Blumenfeld, S.N., Operations research methods: A general approach in primary health care. PRICOR Monograph Series: Methods Paper 1. May 1985). We will distinguish systems assessment from the current PRICOR systems analysis and, by extension, the ancillary PRICOR operations model in the section immediately after this one.

Systems Analysis. Obviously, the general model on which the SAM is based is a systems analysis model. Structure-process-outcome is not, on the surface, all that different from the traditional systems analysis formulation of input-process-output.

Yet, traditional systems analysis has tended to focus heavily on input-output relationships and treat process as a black box. The problem with this, as discussed previously, is that, even when problems or constraints can be clearly shown to exist, the operations manager, who usually has limited control over inputs, is left with little that he/she can do to remedy the situation. It is the significantly greater emphasis on process that provides the main difference between systems assessment and traditional systems analysis.

Two common assessment methodologies that are often, but not necessarily, grounded in traditional systems analysis are intervention studies and program evaluation. Their usual, heavy emphasis on input-output over process would

make them difficult tools, at best, for operations management but each approach has other characteristics that even further reduce their applicability to the routine management of peripheral health services.

Intervention studies (quasi-experimental studies, demonstration projects) usually measure inputs such as personnel or money, and outputs such as vaccination coverage, ORT knowledge, etc. These variables are compared in two or more situations (control and experimental) where the experimental condition involves some change to the system. This change is almost always structural (e.g., a new type of supervisory position is created, a new training module is implemented, or health care workers are given bicycles).

If a distinct control group is included, the studies are usually labelled quasi-experimental. If a group serves as its own control (i.e., baseline data is collected before the intervention is introduced) then the study is usually called a demonstration project.

In either case, intervention studies are inherently expensive and time-consuming. One must wait until an effect could have occurred before testing for it and also collect a lot of data, often about matters not of direct concern but nonetheless needed to control for differences. This predicates a large sample size and relatively sophisticated statistical and analytical skills to make sense of all the data.

Because of these limitations, intervention studies have little practical operations management utility. Managers need to get fast answers to immediate questions using the resources they have readily at hand. One could not conceivably tell a health center director faced with many management problems each day that he should run "little" quasi-experimental studies to resolve them.

Program evaluation tends to be faster and less expensive than quasi-experimental studies, though still focused on input-output rather than process. Usually, program evaluation is instituted by an agency outside the organization that has provided support to a given program and wants to know what has been done and how well. The most frequently used method of evaluating programs is to send an individual or team of consultants either to carry out the evaluation directly or to provide technical assistance (TA) to responsible individuals within the organization who do it.

Typically, the evaluators are experts in the field(s) pertaining to the program to be evaluated. They are briefed by the agency and sent to the project site for some predetermined amount of time, frequently as short as a week or less. There they work with the program managers but have some latitude (based on their expertise) to design the evaluation or TA themselves.

It is common in these evaluations to meet with any personnel that the evaluators or the program management deem necessary, then spend some time on site (leading to the common image of "two men and a Land Rover"). The success of

this approach depends on the skills, acuity and accumulated knowledge of the team. In this sense it is almost an art and its practitioners can be seen as forming a guild of artisans skilled in their craft.

The strength of program evaluation is that it often puts good, knowledgeable people in a position where they can use their experience to assess situations and solve problems. Its biggest fault (aside from sending people who are not competent) is that it is neither a standardized nor replicable process.

Experts most often identify problems within their narrow area of expertise. This is fine if the evaluation or technical assistance requested is narrowly defined and the person or team sent matches the request. If, however, the request is broad -- and it often is -- the decision of whom to send in effect decides what will be found.

Most primary health care systems are quite accommodating in this regard; they suffer from so many ills that finding one or more in one's own area of interest is usually relatively easy. Of course, it begs the question of whether the problems identified were actually the most urgent: the ones that, if fixed, will improve service delivery by the largest amount.

Even more significant in the long run is the fact that there are surprisingly few "lessons learned". The learning process goes on inside the heads of the consultants themselves and they are often not in place to see whether their suggestions are taken, or, if taken, work. Further, there is no standardized vocabulary or on-going linkage between the many, separate program evaluations that take place around the world. Most of what is seen and noted in these discrete efforts is lost to the field, buried in a report that is distributed to only a small coterie of people directly involved with the program.

Systems assessment is designed to be a process that is much more integrated with and organic to the normal functioning of the organization than is program evaluation. In this regard, systems assessment parallels the concept of monitoring elaborated by World Bank researchers in their work on monitoring and evaluation in agricultural extension agencies (Training & Visit Extension Program):

> ***"... monitoring is defined as an internal function, an integral part of good management which is required whether outside funding is involved or not. ... While the monitoring function provides managers with feedback on the nature and extent of progress achieved to date in implementing development activities, compared with what had been planned, an evaluation of a particular program will seek to explain and if possible measure the level of efficiency of its implementation in relation to costs and accrued benefits, to reassess the relevance of the objectives, and eventually to measure its contribution to overall development. Such information, while of some interest to managers, is of greater use to central authorities and funding agencies." (Casley and Kumar, 1987; Murphy and Marchant, 1988)***

5) Differentiating the SAM from the existing PRICOR systems analysis model (operations research)

The PRICOR Project has demonstrated that systems analysis doesn't have to treat process as a black box and that a detailed consideration of performance is a sound approach on both theoretical and practical grounds. The SAM we have developed shares with the existing PRICOR approach this special emphasis on process indicators in the measurement of performance. In a number of important respects, however, it represents a significant departure from the existing model:

First, the SAM represents a comprehensive model of the Health Center as a complete operating system comprising six distinct modules (which is described in detail below). The model describes relationships between these six modules and creates a fairly rigorous framework for analysis. The original PRICOR approach embodies a defined process model (i.e., how to carry out an evaluation), but no system model beyond the undifferentiated input-process-output model.

Second, the SAM is designed from the point of view of the operations manager (e.g., the health center director): primarily to meet his or her needs at the health center level for routine performance information linked to activities and infrastructure that are under his or her control. The original PRICOR approach is unclear as to its primary constituency but appears to take a perspective similar to that of a program evaluation team. The measures used are aimed at a variety of different levels of analysis (i.e., the health center, the supra-organization, and the community) without being clearly differentiated in this regard.

Third, the development of analytical instruments based on the SAM flows directly from the comprehensive systems model and the operations management perspective. This process is described in detail in subsequent sections.

In contrast to this, the large set of items used in the existing PRICOR approach (and compiled in the PRICOR Thesaurus) were produced by groups of international experts working without an explicit system model who produced items that are conceptually isolated from one another (so that they can only be treated as independent measures) and expressed in terms that were deliberately kept as general as possible.

Fourth, SAM implementation assumes the participation of one or more working groups from the system under study in all its phases. There is a heavy commitment to the rapid feedback of summary information to these working groups in order that they might advise on subsequent iterations of the systems analysis. As a result, many problems are resolved during systems analysis and will not require dedicated operations research projects as a follow-up. The SAM is ultimately intended to become a routine part of the PMOH management/health information system.

The original PRICOR model is closer to a program evaluation team approach in which a single cycle of systems analysis is carried out. The strategy of PRICOR is

to use the concrete problems identified in the systems analysis to develop small scale, rapid, inexpensive operations research studies that deal with highly specific, circumscribed issues. Client working groups, feedback loops, and structural supports for routine implementation of systems analysis are not explicitly a part of this approach.

c. Style of analysis

The broad issues on which this work is based were dealt with in the two preceding sections. This section on style of analysis discusses a number of premises that are more directly linked to methodology and are important to understanding how and why we went about measuring as we did. These premises affected a variety of aspects of indicator and instrument development: selection of items, phrasing of items and scoring ranges, organization of instruments, protocol for administering an instrument, etc.

1) Purpose: Monitoring/training versus evaluation/rating

The basic purpose which we have defined for systems assessment is to permit local managers and other trained personnel to monitor performance of key activities so that they might better focus in-service training where it is really needed to improve such performance. Rather than aiming assessment at rating and comparing the performance of workers as individuals, we have stressed that assessment should focus on the performance of the unit as a whole.

Assessment is intended to provide health center managers with a detailed picture of what tasks are being performed exceptionally well or poorly in their unit rather than pinpointing which workers are performing better or worse than the average. Our hope is that this will channel subsequent management action into better feedback and training as the chief method for improving performance.

More feedback on current performance and more in-service training are given the highest priority by a majority of health care workers in job surveys done within the PMOH. This is probably true of many other health systems in the world, as well. Assessment can be legitimately presented as a performance monitoring process: a first, logical step in an active in-service training program. However, to be accepted as such by the workers themselves, it is important to minimize the potential it has to identify individuals for punitive action.

In developing the SAM, we have made an assumption that we can get honest reporting when required from a majority of the individuals in a unit that is being assessed. This greatly simplifies some of the data collection effort, if true. The assumption, however, is predicated on individual participants feeling two things: that the assessment is not threatening to them personally and that they will receive some benefit from the effort they put into it. Presenting assessment as a

necessary step in a program of active in-service training and as a process that does not focus on individual differences meets these criteria.

Two practices developed during pilot testing of instruments, proved particularly useful in instilling this point of view in the health center workers who participated in the assessment. First, individual performance assessment was done with as high a degree of anonymity as could be maintained. Second, monitoring visits always dedicated at least 15% of the total time on-site to intensive in-service sessions in which quickly-tabulated findings were presented to the workers and local managers as immediate feedback from the assessment.

Ultimately, this evolved into the concept of the "Monitoring and Training Visit (MTV)" which has become an integral part of the PRISM process model for systems assessment and is described in some detail in a following section.

2) Domain of analysis: direct and support services at the health center level

The domain of analysis that constitutes the systems assessment of primary health care services includes both direct (i.e., care and counselling) and support services at the health center level. In this context, immediate outcomes of direct services in the community are also included as are structural criteria bearing directly on the provision of services. The domain of analysis is described in greater detail in the following section on the operational systems model.

3) Maximum versus typical performance measures: simulation exercise versus direct observation

The most significant innovation we have introduced in assessment methodology is the introduction of simulation exercises (SIMULEX) as a means of measuring performance. SIMULEX, or role-playing, has been introduced as an alternative to observations of actual patient encounters. The instruments developed for use with SIMULEX have, in fact, been designed to serve in either context.

While observation of actual encounters has undeniable strengths, it also has serious disadvantages in that: 1) observations are made in uncontrolled and non-standard situations so comparisons between them are difficult to make; 2) it requires extensive, often very expensive, training of multiple observers in order to achieve acceptable standards of reliability; 3) observing many types of encounters depends on waiting (perhaps long periods) for unscheduled clinic visits; 4) it is often impossible to collect "negative" observations of the health worker (e.g., that he/she notes that the child does not have a rash or a cough or a broken arm); and 5) procedural reactivity (the effect of the observation process on subject behavior) undercuts, to an unknown extent, the assumption that typical performance is being observed.

Role-playing is an effective way of collecting information on health services performances because it approximates real life situations and the assessment function of the exercise can be integrated with health care worker training. The validity and reliability of the data collected through role-playing is generally good due to the ability to control for ambiguity and extraneous factors (i.e. every participant is presented with the same situation which is designed to have one relatively clear-cut proper response.)

Our approach has been to employ SIMULEX with standardized vignettes to test the performance of health service delivery personnel in basic care-giving and educational activities. The evaluation is done within a non-threatening context in which the exercise is treated as the first stage of a personalized in-service training session. It is made clear to the subject that he or she is being asked to perform as well as possible so that the observer/trainer can see what the person's real strengths and/or weaknesses are in the topic activity. Such simulation exercises carried out in this way avoid most, if not all, of the theoretical and practical weaknesses of direct encounter observation.

The data obtained from simulation exercises clearly represents maximal as opposed to typical performance. Two points are important, however. The first is that inadequate maximal performance (a fairly common result in our testing so far) can be taken as an excellent indicator of inadequate typical performance. This has been confirmed both by direct encounter observations and by interviews with the supervisors of these individuals. Workers who routinely fail to do something right in their day-to-day activity are unlikely to be able to change when challenged by the reasonably fast-paced simulation exercise we have designed. A detailed comparative analysis of simulation exercises with actual encounter observations is presented in volume 2 of this series.

The second point is that maximal performance data is not interpreted in isolation. The complete battery of instruments for performance appraisal includes the simulation exercise, a verbal examination of content knowledge and diagnostic reasoning, a self-report questionnaire, and a checklist-controlled on-site observation of the service area.

Simulation exercises are interpreted, in general, as performance tests of tasks that are important in their own right rather than as measures of specific abilities. We assume that most adults have the ability to learn and do all of the expected PHC activities. The question to be answered is are they proficient enough at the given task under consideration?

4) Performance scores: domain-referenced versus criterion-referenced and norm-referenced; task assessment versus performance assessment

When we talk about performance assessment, it is important to distinguish clearly between three equally reasonable frames of reference we might be dealing with. That is to say that performance assessment can mean:

Frame A - measuring an individual worker's performance in a given task and rating that worker against his/her peers

Frame B - measuring an entire work group's performance based on the group average in a given task area and rating that unit against other units

Frame C - measuring the performance of a given task by individuals, work groups, or all workers in the system and rating that task against other tasks

Each of these has important, and differing, implications for the process of carrying out a performance assessment. For example, an assessment heavily grounded in Frame A is going to be perceived by the individual worker as more threatening to him/her than is one clearly based on Frame C.

The approach we have chosen explicitly excludes Frame A for precisely the reason just given. SIMULEX is perceived by workers as a variant of a familiar training exercise and we are going to great lengths to maintain that viewpoint (e.g., guaranteeing anonymity of participants, focusing mainly on data aggregated at the unit level).

As part of immediate feedback after performance analysis, the observer does go over with the individual worker the tasks he/she did exceptionally well or poorly, but there is no explicit rating of individuals at any time.

The greatest stress in the current assessment model is laid on Frame C: on detailing important tasks and identifying those commonly done well or poorly by the service providers as a group. This is closely allied to the concept of targeted in-service training as the most appropriate control mechanism for improving performance. Within this frame, it is important that performance data be as specific and detailed as possible.

We utilize Frame B in emphasizing the management support aspect of the current assessment model. Thus, we feel that it is appropriate and feasible to produce unit ratings that can be compared to those of other units. This is valuable both for the local and middle management teams. Within this frame, it is clearly necessary to reduce the data load by dealing with indices rather than individual items so that comparisons between units can be made.

As described in more detail below, a manager has two sets of data at his/her disposal: 1) performance indices showing the mean group score plus some indication of the extremes within the unit, and 2) overall item scores showing which item(s) within a given index are exceptionally good or bad. The manager is able to "zoom" from the first to the second with facility and gain a clear picture of performance in short time.

The reports provided as the result of assessment are anonymous with respect to unit members as individuals, but they establish the range of performance within the unit and the characteristics of service delivery. We expect that local managers should be able to identify by their own observation those workers in their unit who are performing exceptionally.

The purpose of performance assessment in the SAM is to identify weaknesses in performance on a unit- or system-wide basis. Thus, we are not interested in scoring the performance of a given worker against his/her peers as we would be in mastery testing or in traditional performance appraisal.

We are not, in other words, looking at the SAM performance assessment instruments as an alternative way of measuring how well, overall, a person is doing his or her job at a given moment. We are using these instruments to identify those specific tasks within an activity that many workers are doing incorrectly so that everyone, workers and managers alike, may be sensitized to them and educated to perform them correctly. The implications of this distinction for both the validation of these instruments and for their application in real-life settings cannot be over-emphasized.

5) Descriptive versus evaluative measures

Descriptive measures focus on facts (e.g., characteristics, behaviors, etc.) that exist or could exist in an organization. They are intended to be value-free. Evaluative measures are value-laden; they ask for an opinion about an organization's characteristics (e.g., are they good or bad, strong or weak). It is possible for a question to have both descriptive and evaluative properties.

This issue is germane to those measures that have been included in questionnaires dealing with the design and functioning of the unit and of the individual's job. The indicators for much of this were taken from Van de Ven and Ferry (1980) and, so, reflect their original strategy:

> ***"In developing the organizational assessment instruments, an attempt was made to define and measure all organizational context and design dimensions in descriptive terms, whereas organizational performance, morale, and job satisfaction were intended to be evaluative."***

This strategy was carried over by us to those measures, newly developed for the current study, that fell into the same categories.

6) Time perspective of questions

It is well-known that judgement made about an organization can vary greatly if one focuses on practices over a reasonably long period of time versus the immediate past or if one considers "normal" practice during periods of usual activity versus focusing on times crisis and extremes. It is left to the analyst to specify what time frame he/she wishes to use but, if none is specified, each respondent is forced to select one in uncontrolled fashion and this leads rapidly to uninterpretable results for the investigator. In general, the time frame ought to be explicitly stated for each question.

The time perspective of questions varies with the type of instrument being used. For performance tests and direct observations, the time perspective is current. For descriptive measures, the time perspective may be current or the past 3-6 months. For evaluative measures, the time perspective is the past 3-6 months.

The main reason for this time frame is that the indicators used are intended to describe existing organizational characteristics, not predict future patterns. For performance testing and direct observations, the nature of the questioning and observer training guarantee this time frame.

For measures that represent worker responses, it is desirable to channel respondents into a recognition of the normal pattern of behavior. This calls for an historical perspective over some reasonably long period of time. On the other hand, the unreliability of historical recall is also well-established. A time perspective of 3-6 months seems to represent a reasonable balance between establishing a time frame long enough to recognize patterns but short enough so as not to introduce significant distortions.

The time to complete the assessment of a single health center should be kept to less than two weeks so that it can safely be assumed that individual phenomena will not have changed significantly during the period of data collection.

Seasonality of health problems must be taken into account when comparing performance between different health centers evaluated at different times of the year. Seasonality also suggests the value in keeping the time perspective of questions within the last 3 months whenever possible.

7) Objective versus subjective measures

The terminology used in discussing objective and subjective measures is taken from Payne and Pugh (1976):

> ***"Objective [measures require only] a direct assessment of organizational properties without any conceptual transformation. Subjective [measures require an] indirect assessment of organizational properties by instruments that measure group perceptions; here a member is a respondent to instruments with statements***

> *such as, 'The jobs in this organization are clearly defined ...' or 'The employees here are constantly checked for rule violation'."*

An extremely common view about objective versus subjective measures is that the former are somehow more reliable and valid than the latter. This is simply not true. Our position on this issue was stated clearly and forcefully by Van de Ven and Ferry (1980) when they wrote:

> *"... we concur with John Campbell's (1977) stance, 'Any objective measure is a subjective measure once removed.' Documents, records, and archives can be very useful to measure objective properties of organizations. However, the belief that they are generally more reliable or valid than subjective measures is patent nonsense, particularly when one considers the sloppy ways many organizations score or keep track of their reporting systems, the fudging of data that occurs daily, the shifts in administrative reporting directives, the need to look good to higher executives and funding sources, and the need to prevent law suits. ... Indeed, there are many instances where subjective measures that ask respondents directly and in confidence what goes on within the organization may yield more accurate data than objective measures obtained from records compiled by the organization being assessed."*

The point of view taken in this study is to place reasonable trust in the perceptions of the workers, themselves, to indicate problem areas. Of course, no data, either subjective or objective, should be taken at face value until their accuracy and limitations have been ascertained. With regard to maximizing the accuracy of respondent's answers in context of a systems assessment, it is clear that a number of steps must be taken to establish a proper emotional context for honest responses. These include: establishing rapport and trust, guaranteeing anonymity, providing obvious received value, and demonstrating integrity.

8) Open versus closed questions: number of points on answer scales

We have avoided the use of open-ended questions in the final versions of the systems assessment instruments. Data in a non-standardized format is virtually impossible to tabulate during the actual assessment process. More fundamentally, limiting and standardizing responses is crucial to the use of these questions as measures in a replicable fashion. Thus, the majority of measures included in the assessment instruments have been cast as Likert-type items with five scale points.

The choice of five points on the rating scale is based primarily on published studies concerning the effects of number of scale points on the reliability of measures. In particular, we accept the Monte Carlo study of Lissitz and Green (1975) which showed that the increase in reliability with increasing scale points was significant up to 3 but that numbers greater than 5 added little.

In general, our experience with using 5-point Likert scales in Peru has been extremely good. On one hand, informant groups brought together to design

questions had little difficulty in developing clearly distinguishable anchors or cues to represent the intended meanings of each of the five points. On the other, study participants who responded to the questions reported little to no difficulty or confusion in "knowing" what the ranges meant or where their judgement fell within them.

The use of Likert-type scales yields a greater variance than would be possible with simple Yes/No-style questions and still permits the analyst to aggregate responses into a dichotomous if desired simply by setting the threshold between what will be interpreted as a positive and negative response.

9) Anchors and cues for answer scales

Our guide in the development of anchors and cues for answer scales has been Guilford (1954), who pointed out that anchors serve two purposes: they reinforce the idea of a continuous range from a minimum to a maximum and they help the respondent place his/her, often, qualitative judgement as a value on that quantitative range.

The criteria we used for constructing anchors were those given by Guilford:

Clarity -	***Short phrases, simple, unambiguous***
Consistency -	***Cue is consistent with the meaning of the question and with the other cues***
Precision -	***Cue refers to a clear portion of the overall range that does not overlap the ranges of other cues***
Variety -	***Avoid using the same terms in many or all cues; use different terms for obviously different scale levels***
Objectivity -	***Avoid cues that imply good or bad, worthy or unworthy, and desirable and undesirable***
Uniqueness -	***Cues are unique to a given question***

The criteria "variety" and "uniqueness" appeared, at least in practice, to overlap in meaning. In any case, we found that attempting to introduce a high a degree of variety and uniqueness in an instrument ended up producing significant confusion and irritation in respondents during pilot testing. This was due to a combination of factors: 1) the basic format used is not familiar to many Peruvians, which introduced a heightened tension; and 2) respondents found that they had to continuously interpret new sets of anchors, which greatly slowed their completion of the questionnaires.

As a compromise to this reaction, we have reduced variation by using the same set of anchors for all questions of a similar type, while maintaining variety and uniqueness between different question types.

For the measures taken from Van de Ven and Ferry (1980; indicated in the section below that details and describes the indicators), the anchors they used were generally translated directly into Spanish. In some cases, we found during pilot testing of the instruments, that the direct translations were not as general and fixed as could be desired, and the direct translation was replaced with another anchor (usually one suggested by one of the informant groups participating in instrument development).

10) Task-referenced versus concept-referenced indicators

A number of the indicators used, especially those taken from Van de Ven and Ferry, are measures of constructs that have been developed and tested as such using factor analysis and other techniques. In such cases, the items that are used have been selected as a parsimonious set of measures of the concept. Their value as independent measures is intended to be of less importance than their value when aggregated with other items to produce the indicator. In some instances, nevertheless, individual measures have a great deal of practical relevance as well. We have called indicators of this type concept-referenced indicators.

We use other indicators, however, that are simple aggregates of dissimilar items grouped together because they make up a given task that is important in itself. The purpose of the indicator is not to operationalize a concept per se but to provide a convenient "overall score" for the items that it includes. There is no intention, at present, to select items as if they were operational referents of a construct, no matter how broad. The reason for using them is mainly practical: they provide a convenient way of summarizing the overall performance of selected groups of tasks. We call such indicators "task-referenced indicators.

11) Level of analysis: individual versus group

In measuring various dimensions of structure, process, and outcome, the individual is the unit of analysis and no aggregation of the data to the group level is performed. This is appropriate for data in which the individual, as a respondent, provides data that apply directly to his or her activities or judgement.

When collective properties of group structure and behavior are wanted, however, it is necessary to consider the problem of how to obtain reliable information from individuals that serves to measure the collective property.

We have used the typology developed by Lazarsfeld and Menzel (1969) of properties of groups which they used to suggest how data obtained from individuals can be aggregated to measure characteristics of a group or an entire organization. They distinguished three type of data that can be obtained: member, relational, and global.

Member data are characteristics of individual members that can be obtained without reference to collective characteristics. Relational data refer to the linkage between members and are obtained by questions about the ties each member has with other members of the group. Global data are characteristics about the group or organization that are not collected by aggregating information about individual members.

In measuring the design of the health center as a work unit, for example, the individual employees are members and the health center is the collective. Member data include measures of job authority, job satisfaction, and educational level. Relational measures include conflict and methods of conflict resolution among members, and frequency of communications. Global measures include measures of unit size and of relations between the work unit and other entities, such as the community it serves and the UTES.

While it is generally accepted that constructing measures of central tendency (the most commonly used type in the systems assessment) can be done by treating the latter as linear combinations of individual values, there are few general prescriptions for avoiding problems in moving from individual to collective variables. We have attempted, therefore, to follow the advice of Van de Ven and Ferry in this matter:

> *"These issues are better dealt with on a specific variable-by-variable basis in which the intended and observed meanings of each variable at the individual and group levels are evaluated and made explicit. To the extent that these functional relations are made conceptually explicit, one can better detect and correct aggregation errors and know more concretely the meanings of the data when they are analyzed."*

12) Respondent versus informant roles of participants

We have tried to maintain the same distinction with respect to people from whom data is obtained as did Van de Ven and Ferry: respondents are persons who provide member and relational data about themselves; informants are persons who provide global data about the unit. The use of the latter means asking individuals to act as reporters about characteristics of the organization or about other people's behavior that is clearly removed from themselves.

Van de Ven and Ferry make a strong point against mixing respondent and informant roles in the same indicator. We have followed this advice throughout the design of our systems assessment instruments as well. Subjects are asked to answer as either respondents or informants to all of the items which serve as measures for a given indicator. It has been necessary sometimes to switch roles between indicators in the sane instrument. In such cases, instructions make it clear that a role change is being called for.

The sampling done to obtain subjects also depends on the role demanded. For instruments in which members are being asked to serve as respondents (e.g., job design indicators, performance indicators), the selection is made so as to have a representative sample of the members under investigation. For instruments in which members are asked to serve as informants (e.g., unit overview indicators), we select a limited set of people who are the most knowledgeable about the properties of interest and who have direct access to other global information sources that may be needed.

d. The Operational Model

1) The basic systems model

Figure 1 illustrates the basic systems model with which we are working:

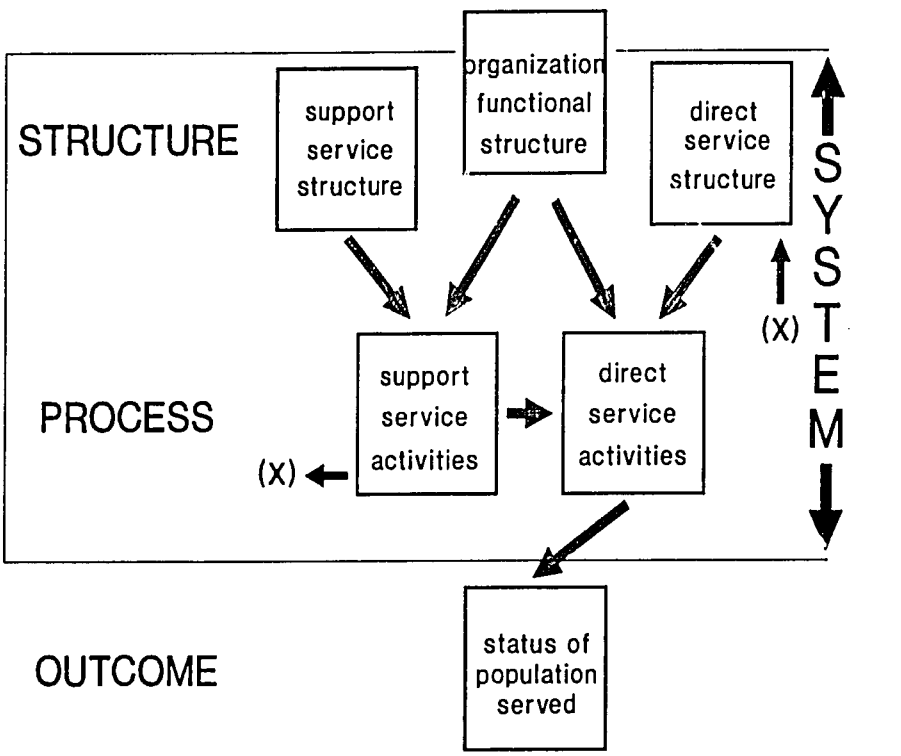

Figure 1. Basic Systems Assessment Model

As we indicated in introducing this model in an earlier section, the SAM follows the distinctions made by Donabedian (1966; 1980) between structure, process and outcome. We will define these elements in somewhat more detail here than was provided previously:

> Structure is defined as "the relatively stable characteristics of the providers of care, of the tools and resources they have at their disposal, and of the physical and organizational settings in which they work. The concept of structure includes the human, physical and financial resources that are needed to provide medical care" (Donabedian, 1980).
>
> Process in health care is usually defined as the set of activities that go on within and between practitioners and patients. We have broadened this scope to include as processes the set of activities that go on between support staff and managers, and practitioners. The former is called direct services and the latter support services. Process indicators have dimensions of quality and quantity and relate to members' activities in carrying on their work.
>
> Outcome is defined by Donabedian to mean "a change in a patient's current and future health status that can be attributed to antecedent health care." Outcome indicators refer to the status of the objects on whom the work is performed. Changes in characteristics that can be attributed to the work performed upon them have also been interpreted as impact; in our model, outcome and impact are synonymous. The most common examples of outcome indicators are morbidity and mortality but the set also includes health-related knowledge acquired by the patient as well as changes in health-related behaviors.

We have limited the use of the three terms, structure-process-outcome, to these definitions. The most significant constraint is that we do not permit frame shifts that would allow one, for example, to treat a variable as a process in one context and an outcome in another. We do not deny that this is a perfectly valid action within the realm of systems theory. Obviously, any process variable (other than a "prime mover") is the result of other actions and is, therefore, an outcome in the latter context.

In order to keep the terminology unambiguous, however, we are freezing the frame on the system as presented in Figure 1. Only those variables that represent characteristics influenced by direct service activities will be called outcomes and only direct and support service activities will be considered as process.

2) Organizational elements and indicators in the SAM

Process indicators (service activities). As Figure 1 illustrates, the focal point of the PRISM SAM is the set of activities associated with service delivery, both *Support Service Activities* and *Direct Service Activities.* Support Services

comprises four elements: supervision, logistics, information system and training. Direct Services comprises two elements: care-giving and promotion/education. All process indicators in the PRISM SAM are contained in these two modules.

We have emphasized that systems assessment is heavily focused on the measurement of the actual processes of service delivery. We have also commented before that service delivery practices are a reflection of normative standards and, as such, are accepted as means to the end of better health, are codified through protocols and manuals, and are implemented through training.

In the SAM, direct service delivery practices can be treated as independent variables, if we choose, in analyses against selected service delivery outcomes to determine if a given process really does have the impact assumed in the standards of good practice.

More typically, direct service delivery practices will be treated as dependent variables. These are the tasks which the system was designed to perform. At this level whether or not they are the "best" means to some end (better health for the population) is of secondary interest.

Much effort and money are being invested in the delivery of direct services, yet few studies have actually looked at whether a health system is delivering these services effectively and, if so, what aspects of the health care system contribute most to the successful delivery of services? In the SAM, direct services practices can be treated as dependent variables in analyses in which support services and organizational attributes are treated as independent variables.

In general, the process indicators used in the SAM are measures of the quality of services. Coverage is the only process measure used for the quantity of service provided. Other measures of quantity, such as number of person-days per month dedicated to a given program are classified as structural variables. We recognize that other models of health service delivery frequently treat coverage as an outcome. It should be emphasized, therefore, that coverage is a measure of the quantity of service output or work done and is, therefore, a process indicator in the SAM.

The elements of support service that we are concerned with are: supervision, training, logistics and information systems. These are often cited as the aspects of primary health care systems most needing improvement but have rarely been studied systematically in relation to their effect on direct service delivery.

"Supervisors are not doing their jobs", or "Transportation is never available when needed" are characteristic findings that are repeated in countless evaluations of health systems. Such comments may be true, but they don't explain how supervisors contribute overall to the successful delivery of services or if lack of transportation was the most critical constraint on good service delivery.

If support services do contribute to successful service delivery, it is necessary to know their relative importance under differing circumstances. Training may prove to more important than supervision in the delivering of vaccinations, while regular supervisory visits may be more important for successful delivery of ORT.

There is an inherent problem in assessing support services and that is the lack of a precise characterization of what constitutes proper supervision, proper training, etc., under different circumstances. Nevertheless, there is agreement on the broad functions of each element and on the fact that the quality of direct services is somehow dependent on them.

To counter these problems of ambiguity in the SAM, and because our measures of such activities are still imprecise, we limit analysis of support services to the extremes of direct service delivery: i.e., to those units who are performing their direct service activities extremely well versus those who are performing them extremely poorly. It seems reasonable that, if support services do contribute to direct service quality, then differences should be discernible in the quantity and/or quality of support services between units whose direct service performance is dramatically different.

Structural indicators. Structural indicators, or organizational attributes, in Fig. 1 are divided into three elements: *Organizational Functional Structure, Support Service Structure, and Direct Service Structure.*

Organizational Functional Structure contains all general indicators of a unit's basic organizational, resource, and management characteristics (e.g., distribution of unit authority, number of job titles in unit). It also includes indicators of intra-unit and inter-unit relationships (e.g., intra-unit communication, inter-unit resource dependence). Inter-unit relationships consider the health center as the focus in its dealing with either the UTES (or other administrative office to which it reports) or the community.

Support and Direct Services Structure, the other two modules, contain indicators that relate to the job design or job incumbent characteristics of the work positions and personnel assigned to either support or direct services, respectively. The activities addressed in these structural modules parallel those in the process modules: i.e., supervision, logistics, information systems and training in Support Service Structure, and care-giving and promotion/education in Direct Service Structure.

Outcome indicators. The indicators contained in the final module, Status of Population Served, include only those characteristics in the population served that can be thought of as related to the direct service activities under study. Outcomes in the PRISM SAM refer strictly to what are often called "impacts" in other systems models (as noted previously, coverage measures are treated as process indicators: i.e., quantity of work done).

Relationships. The three structural and two process modules, together, make up the "system" on which the SAM is focused. Outcome is, by definition, outside the system and a result of actions by the system as they flow through the direct services module. On the other end, the Organizational Functional Structure has been drawn with an ambiguous boundary with respect to the system to emphasize the fact that a number of indicators have been included which reflect relationships between the system and the outside (e.g., the UTES and the community).

The six modules presented in Figure 1 make up the domain of analysis of the SAM. The arrows connecting modules in Figure 1, and their directions, indicate the potential relationships being hypothesized in the domain of analysis. The intention in establishing these hypotheses is to create a basis for testing associations once data are available.

In selecting system assessment measures (i.e., in operationalizing the concepts which comprise the SAM), we begin by selecting items that have significant content-validity based on international and local consensus criteria. The systems model, however, creates a context in which to establish the predictive validity of key indicators in the structural and process modules.

While we do not exclude the possibility of accepting empirical generalizations based solely on simple statistical associations, we anticipate that predictive validity will usually be established within the framework of relationships established by the theoretical model. In establishing predictive validity, process measures can function both as independent variables (to outcome indicators as dependent variables) and as dependent variables (to structural or other process indicators as independent variables).

3. Methodology of Instrument Construction and Evaluation

a. Selection and definition of indicators and development of measures

The following description of the process for selecting and defining indicators and developing measures presents an overview. Detailed descriptions of key elements in this process are provided in subsequent sections.

Each of the six modules in the system model is described by a limited set of indicators. For example, the Direct and Support Service Structure modules are both described by 15 indicators having to do with job design (e.g., task difficulty) and job incumbent characteristics (e.g., expertise).

Ideally, each of these indicators will possess a unique meaning in the set and this meaning will be consistent when the concept is applied to different PHC programs or to the same program in different countries.

Indicators will be either concept-referenced or task-referenced, as described in a prior section. Concept-referenced indicators are assumed to be measuring a

unique dimension and are, whenever possible, adapted from proven, published methodologies. Task-referenced indicators are developed as a means of summarizing a set of practically related activities but are not assumed, necessarily, to be measuring a unique dimension.

Indicators range from a single item (e.g., years of education) to a score for the performance of a complex task tested by a simulation exercise (e.g. ability to carry out a physical examination to establish dehydration status).

Each indicator is fully operationalized by establishing the measures to be included and the rules for calculating and interpreting the indicator. The specific manner for doing this is determined by local conditions, including the relevant characteristics of the specific program being measured. This step includes the designation of which items are to be used as measures, how they are to be presented (e.g., the phrasing of a question), and from whom the data is to be collected.

Items are chosen in close collaboration with the local health service providers in the organization being assessed. For performance indicators in direct services, international and national norms of good performance in specific program areas are interpreted for local conditions by working groups of physicians, nurses, nurse-midwives, and health auxiliaries operating under the guidance of PRISM PRICOR Project staff. This is a highly interactive process with many iterations.

The project staff's function in this is to ensure that the integrity of the system model and its concepts are maintained during this process, that significant deviations from the PMOH's norms are limited to cases of absolute necessity, and that all accepted items are clearly defined with respect to their role as measures and unambiguous in their presentation.

Instruments are prepared and pilot tested in health centers other than those involved in the actual systems assessment.

The reliability and validity of indicators is evaluated using accepted psychometric and edumetric techniques. Modifications are brought back to the working groups for their concurrence at all stages of the evaluation and final writing of the instruments is done with these groups.

The indicators that have been included in the final assessment model are distributed among nine distinct forms (questionnaires, examination, simulation exercise checklists, etc.). Each indicator was placed in a form that best fit the operational definition chosen for that indicator. These forms and their associated indicators are described in detail in subsequent sections.

There are four stages in indicator development that deserve additional comment: literature and document review, working groups, forms design criteria, and pilot testing.

1) Literature and document review

Indicators and the items used to measure them were developed over the course of 15 months (December,1987 - March, 1989) through continuing discussions: 1) among project staff; 2) with focus/informant groups; 3) with program heads in the PMOH; and 4) with national and international experts.
These discussions centered around published or written materials on important health service activities in child survival programs, especially control of diarrhea and immunizations. Materials reviewed included the following:

The PRICOR Thesaurus (Center for Human Services) - the Thesaurus was reviewed as a first step in the selection of indicators and, again, in the selection of specific items. This was done in conduction with the focus/informant working groups.

Niños (bi-monthly journal of primary health care and medicine) and the Niños reprint library (A.B. PRISMA)

Program and training handbooks (UNICEF, PAHO)

Norms and training guides of the PMOH

Selected reprints from the international scientific literature

A number of books were particularly useful and deserve special mention. These are:

Handbook of Industrial and Organizational Psychology. M.D. Dunnette (ed.). Wiley-Interscience, 1983

Educational and Psychological Measurement and Evaluation. K.D. Hopkins and J.C. Stanley. Prentice-Hall, 1981

Health Care Management: A text in organization theory and behavior. S.M. Shortell and A.D. Kaluzny. Wiley Medical, 1988

Health Care in Peru: Resources and Policy. D.K. Zschock (ed.). Westview Press, 1988

Assessing and improving health care outcomes. J.W. Williamson. Ballinger, 1978

Exploration in Quality Assessment and Monitoring (3 volumes). A. Donabedian. Health Administration Press, 1980

Donde no hay doctor. D. Werner. La Fundacion Hesperian, 1980

Van de Ven and Ferry (1980) have developed a methodology for measuring various attributes of organizations entitled the "Organization Assessment Instruments" (OAI). The instruments measure various aspects of Job Design, Job Context, Organizational Structure and Intraunit Relationships. We are using many of these indicators in the current systems assessment model.

2) Participation of organization representatives in focus/informant (F/I) groups

A basic approach used throughout the development of systems assessment instruments has been to draw heavily (but not uncritically) on the expertise and insights of the people actually working in the system under study. Such people played a valuable role in interpreting indicators and measures in extreme detail according to local practices and understanding.

The key creative process in this development stage came in the interchange between the two groups of "experts" we assembled: the project's own systems assessment experts and the people from the system serving as members of focus/informant (F/I) groups.

Sir William Osler enjoined us to "Listen to the patient, he is telling you the diagnosis". Through many months of weekly to bi-weekly meetings with people from the Cono Sur, we let the people who know best tell us about what was wrong and right with their system and how we could best measure it. We were particularly impressed with the wealth of detail that was elicited when we asked for operational definitions to indicators that were part of the SAM.

Beginning in March, 1988, the project formed F/I groups of approximately 8-10 members with the following groups of health workers in the Cono Sur: health auxiliaries, nurses, nurse-midwives, general physicians, and managers (health center heads, EEP heads, and UDES representatives). In addition, three working groups of mothers from communities in the Cono Sur were formed for brief periods to get input from the most important user population.

The health worker F/I groups met at weekly or bi-weekly intervals from March to October, 1988 and on an as-needed basis thereafter The nurse, auxiliary, and nurse-midwife groups were re-established on a more regular, bi-weekly basis in April, 1989 to assist with the development of the final set of instruments and have been meeting regularly from that date to the present time (August).

Instruments and indicators were divided among three F/I groups that were given primary responsibility for developing them. These three were the nurses, the nurse-midwives, and the auxiliaries. The other groups were used in an ancillary fashion and all groups were given a chance to comment on certain particularly critical indicators (e.g., history-taking, physical examination, and treatment strategy in the various care simulation exercises).

Initially, project staff presented rough drafts of indicators and measures gleaned from the PMOH norms, the PRICOR Thesaurus, and other written sources to begin the iterative process. They did not, however, directly specify conclusions. The F/I groups ultimately were given the final word on the inclusion of items and on their final form. The project staff person who functioned as group leader played an active role in managing this process but that role was indirect.

A special note needs to be included here about language. All de novo formulation of items was done from a Spanish-language draft version. Those items originally taken from sources in English (e.g., the PRICOR Thesaurus) were translated by project staff prior to being presented to the F/I group. No effort was made to retain the original translation of these items if the group decided that an alternative was better.

The only exception to this policy was made for those measures from an English source that, taken as a unit, operationalized a proven, concept-referenced indicator. The indicators developed by Van de Ven and Ferry are the most important example of these. For these indicators, every effort was made to retain a Spanish translation as true to the English version as possible, irrespective of local arguments for changes. The idea that these sets of items were specially tested and proven measures for the concept was carefully explained to the F/I groups to justify their acceptance essentially as written.

Project staff laid out the theoretical context for given indicators and pointed out when the F/I group was choosing a course that would violate a significant aspect of the frame of reference. They then took tentative conclusions of the group and subjected them to critical analysis (in conjunction with other project staff, with other F/I groups, and with outside experts - such as PMOH program directors). The results of such detailed analysis were returned to the original F/I group with enough explanation so that its members could appreciate for themselves the implications of choosing a particular item or formulation. In this way, both project staff and the F/I groups were able to deepen their understanding of the system and, thereby, arrive at more profound insights on which to assess potential indicators and measures than either group left to itself would have been able to develop.

Solution-generation was a true collaborative effort in this development; the ever-present tendency among outside experts (ourselves) to short-circuit the process by simply "providing the answers" was counter-balanced somewhat by formally giving the system's personnel the final word on when the solution was adequate.

The most important psychological results of this arrangement are: 1) that the system personnel (and the managers) quite rightly feel that they have been the ones who "really" decided on the aspects of their organization to be assessed; and 2) that they believe that they will be left with the tools to continue improving their systems long after the original group of assessment experts has gone.

The most telling result, however, is that the final instruments, indicators, and measures that were produced benefitted from a substantial amount of insider detail that could only come from the workers and managers themselves but without sacrificing a sound theoretical base.

3) Forms design criteria

It was apparent to us from the beginning of the project that the face validity of our instruments would be an important consideration in gaining the full cooperation of health center managers and staff in carrying out a full assessment of service delivery at their unit. The development process ensured that the indicators would, in fact, be valid, reliable, and meaningful. It was important, as well, that the physical format of the instruments be such that the people we had to deal with believed them to be as sophisticated and powerful as we had designed them to be.

In addition to face validity, a second requirement was that data logging on such instruments be done so as to minimize the effort needed to tabulate responses and produce intermediate indicator scores for a unit while the assessment team is still on-site.

A great deal of care was, therefore, exercised in the design, layout and printing of the final instruments. All were printed with an HP Laserjet II Plus using the WordPerfect 5.0 word-processing package to produce a high-quality "professional" appearance. Standardized formatting was established between all forms so that once learned, it would be obvious to whomever used any of the instruments in the set.

4) Pilot testing

The final version of each indicator was pilot tested using groups of volunteers (health auxiliaries and nurses) from a health area in the Lima area, but in a different UDES than the Cono Sur. Two levels of pilot testing were employed. The first was simply to check the language used (clarity, ease of understanding) and the procedures for collecting data (e.g., standardized questionnaire presentation, running a simulation exercise, etc.). At this level, the responses of no more than 10-15 subjects were used.

For the job/unit design instrument (developed almost exclusively from indicators published by Van de Ven and Ferry) it was important to obtain responses from a substantially larger group of subjects for the reason that many of these formulations were "alien" to local experience. It was a fairly subtle task to modify them sufficiently to be clear to our subject population without losing their original meaning. This required two rounds of testing with 40-50 subjects to achieve an acceptable formulation for each item.

The second level of pilot testing was employed for the content-knowledge examinations, in which the score on a test of basic knowledge (30-40 items) served as the measure for this indicator. In order to construct a test with 20-25 items, it was desirable to be able to calculate endorsement rates and correlations with total score for each item to assist in the final selection. While these criteria were not the only determinants (since we were testing acquired knowledge, the

emphasis was on content validity), they were important considerations, as was the calculation of reliability (coefficient alpha).

We desired to get a sample size that was sufficiently large to provide adequate estimates of these variables. Coefficient alpha presents the most demands on sample size; Nunnally (1978) recommends a minimum of 300 subjects while Guilford (1956) suggests that a minimum of 200 is sufficient. Unfortunately, such numbers were not feasible with the human resources available to us as subjects. We settled on 100-120 as a feasible compromise for this purpose.

b. Validation of indicators

We feel strongly that the validity of the indicators to be used in systems assessment is a matter of concern and one that has not been given the attention it is due in many similar efforts to date. There are a number of definitions of validity and we will describe below those that are most relevant and important to the SAM.

We can say that a test is valid if it measures what it claims to measure. Having said that, however, does not end the matter but opens it up to the difficult question of how we know whether a test measures what it purports to. There is no single way of establishing validity because the very meaning of the term has a variety of facets and each must be dealt with if validity is to be established.

1) Content validity: Domain-referenced evaluation by structured groups

The performance testing we include in the SAM covers most of the process indicators in the model, which are, in turn, the major part of the overall assessment effort. It is fair to say that performance testing is the form of greatest interest to us in developing and using the SAM.

Performance testing is a form of attainment testing and this fact narrows the task of establishing validity to a primary focus on content validity:

> *"If the items of a test can be shown to reflect all aspects of the subject being tested, then it is per se valid, given that the instructions are clear." (Kline, 1986)*

For performance-oriented instruments, validity depends primarily on the adequacy with which a specified domain of content is sampled. The following principles for establishing content validity are taken from Nunnally (1978):

> ***"Rather than test the validity of measures after they are constructed, one should ensure validity by the plan and procedures of construction."***
>
> ***"The validity of the measure is judged by the character of the plan and by the apparent skill with which the plan has been carried out. If it is agreed by most***

potential users of the test, or at least by persons in positions of responsibility, that the plan was sound and well carried out, the test has a high degree of content validity."

"The two major standards for ensuring content validity [are]: (1) a representative collection of items and (2) 'sensible' methods of test construction."

"... the selection of content usually involves questions of values ... with nearly all measures based on content validity, values determine the relative stress on different types of content. Of course, where values are important, there are differences in values among people; consequently, usually there is some disagreement about the proper content coverage of particular tests."

and

"... inevitably content validity rests mainly on appeals to reason regarding the adequacy with which important content has been sampled and on the adequacy with which the content has been cast in the form of the test items."

Along with these basic principles, Nunnally offers a number of forms of circumstantial statistical evidence that suggest that content validity is present:

"... at least a moderate level of internal consistency among the items within a test would be expected ... this is not an infallible guide, however, because with some subject matter it is reasonable to include materials that tap somewhat different abilities."

"... comparing performance on a test before and after a period of training"

"... correlating scores on different tests purporting to measure much the same thing"

All of the task-referenced indicators developed for the SAM depend mainly on the propriety of their content as the basis for their validity. The process of developing these indicators, described previously, was designed to bring together most of the indicator's potential users to carry out a rigorous comparison between the indicator content and the content of the domain to be assessed by that indicator. Accepting that, in some sense, an indicator is a measure of some construct, then these task-referenced indicators must be considered as measures of extremely broad constructs (i.e., constructs with at least 4-5 conceptually distinct terms).

The conclusion of our F/I groups and PMOH experts, both before and after the systems assessment was carried out, is that the plan for the indicators we have developed and the plan for their execution are sufficient to ensure their validity. We note, again, that using F/I groups actively in all stages of assessment instrument design, development, and testing pays off handsomely in terms of ensuring the content validity of these measures.

2) Face validity

The traditional definition of face validity states that a test is face valid if it appears to measure what it purports to measure from the perspective of the subjects and that face validity is important only in so far as adults will generally not participate in tests that appear to be useless or silly.

The face validity of systems assessment instruments is not a trivial matter, however, since the assessment involves members of the organization much more intimately in the process than the kinds of tests that are usually considered when talking about face validity. This greatly raises the expectation of these members that the tests being used are useful and accurate, while giving them far more opportunity to discover flaws.

Maintaining a high level of face validity has meant giving substantially more explanation than usual to organization members concerning the purpose of the tests, the background thinking that went into their development, and the overall context in which they are being applied.

3) Intrinsic validity

Intrinsic validity of an instrument refers to how consistently it measures what it is intended to measure. The primary procedures used to evaluate the intrinsic validity of the instruments in the SAM include two reliability indices: self-consistency and inter-observer stability, convergent validity of selected dimensions with parallel measures, and factor analysis.

a) Reliability indices

The extent to which a test would give consistent results if applied more than once to the same people under standard conditions is an obvious and fundamental factor in determining that it is measuring what it claims to measure.

Coefficient alpha. Both Cronbach (1970) and Nunnally (1978) regard coefficient alpha as the most important index of test reliability. Coefficient alpha is an example of an internal consistency measure: one which depends on the amount of agreement between the different items on a single administration of a test.

Coefficient alpha can be thought of as an extension of the split-half method of scoring reliability (i.e., method in which items in a test are divided randomly into two halves and the two scores are correlated). Coefficient alpha in this context is the overall mean of all split-half reliability coefficients. Another way of describing it is that

> ***"coefficient alpha is the estimated correlation of the test with another test of the same length from the same item universe, and its square root is the estimated correlation of the test with true scores (Kline, 1986).***

The calculation of coefficient alpha is relatively simple:

$$\text{coefficient alpha} = (k/k\text{-}1)\left[1-(\Sigma\sigma_i^2/\sigma_y^2)\right]$$

where k = number of items in the test; $\Sigma\sigma_i^2$ = sum of the item variances; and σ_y^2 = variance of the test

Coefficient alpha is aimed at detecting measurement error due to a lack of internal consistency in responses to items within an indicator. If coefficient alpha is low for an indicator, the items in that indicator are not likely to be all operational referents of the same construct. Though coefficient alpha has importance for all the indicators used in the SAM, it has less for those indicators which we have classified as task-referenced. As discussed earlier, such indicators measure very broad constructs and we would not be expect to encounter more than moderate values (e.g., 0.35-0.55 or less) for coefficient alpha.

For concept-referenced indicators, on the other hand, coefficient alpha is a critical piece of evidence in establishing that no serious errors have been made in sampling the content of the universe of operational referents of a given construct. Van de Ven and Ferry (1980) established criteria for the expected ranges of coefficient alpha for indicators measuring constructs of various breadths. Assuming that the number of items in the indicator was 3, the expected range of coefficient alpha was 0.35-0.55 for a broad construct (3 or more terms), 0.55-0.70 for a moderately broad construct (2 terms) and 0.70-0.90 for a narrow construct (1 term). All of the indicators borrowed from Van de Ven and Ferry for use in the SAM met these criteria in their original study.

Inter-observer stability. For those instruments to be filled out by observers rather than by the subjects themselves, a second aspect of reliability becomes an important issue: inter-observer stability of scores. Traditionally, this index is often calculated as a simple correlation of scores between observers. In this study, we have used coefficient alpha again, this time as an index for stability over multiple observers. The calculation remains the same but is made on data obtained from several observers watching exactly the same thing. Thus, low reliability can only be attributed to inter-observer error.

b) Parallel measures: convergent validity

Convergent validity is defined as present when several dissimilar methods of measurement of the same concept correlate well with the test being developed. Within context of the SAM, the instrument for which convergent validity is particularly important is the simulation exercise (SIMULEX) for which the most common parallel measure is direct observation. Convergent validity between these two measures is established by three methods.

Two of these focus on a comparison of item scores for those items that serve as measures. In the case of SIMULEX versus direct observation, the items (indeed, the form) are the same in both situations. The only difference is that one set represents observations made during a role-playing exercise while the other represent observations of the same person while actually performing these duties in his/her health center or post. The observer is the same in both instances.

Since our primary interest in the SAM is the identification of weaknesses in task performance, tests of convergent validity that focus on the items themselves is particularly germane. The first of these is simple correlation of item scores under the conditions just described.

The second test of convergent validity is item agreement. This is calculated from only positive observations, which is to say that a comparison is made between a persons performance of each item for all item pairs in which at least one of the two methods yielded a positive score. Item agreement is defined as the number of pairs in which both were positive divided by the total number of pairs in which either was positive. An item agreement score of 1.0 is a perfect positive correlation and equal to 0.0 is a perfect negative correlation. Agreement equal to 0.5 means that no relationship exists at all. Generally, item agreement equal to 0.8 is taken as evidence for strong convergence on an item between two methods or tests.

The third method for establishing convergent validity is the use of coefficient alpha to test inter-method consistency in a fashion that parallels that used to test inter-rater stability, except that inter-method error that accounts for lowered reliability rather than inter-rater error.

c) Factor analysis: construct validity

The indicators taken from Van de Ven and Ferry (1980) represent concept-referenced measures whose intrinsic validity has been establi hed by factor analytic methods. The methods and results of this process are detailed in the work cited.

We recognize the need to develop factor-analytic constructs for many aspects of service delivery in future and set this as one of our long-term goals. Nevertheless, such an effort calls for sample sizes in the thousands and this is much beyond the capacity of the current project. Our immediate goal, therefore, is to produce a basic framework and operational model using indicators that "satisfice". We anticipate that this effort will delineate activities or characteristics of service delivery for which constructs can be identified, operationalized, and rigorously analyzed.

4) Extrinsic validity

Extrinsic validity refers to how useful an instrument is in practical terms in meeting the analytical purpose for which it was developed. The primary method for evaluating extrinsic validity was to determine how well observed correlations corresponded to the theoretically predicted relationships among various indicators. In this context, the relationships suggested in the original systems model of Figure 1 were the fundamental basis for such analyses.

Analyses of variance on various indicators are carried out to assess how well they discriminate between different organizational units or job types. Since this phase of the SAM development effort is still preliminary to a full implementation, the need for and desirability of extensive extrinsic validation (such as multiple regression to determine variation in performance by related indicators) was reduced.

4. Definition and Description of the Basic Instruments

The basic set of instruments for the SAM developed to date includes seven formats. Three of these are ***general formats***, two (DWF, JDQ) because the items they contain are expressed in general terms applicable to any program and one (CMI) because it contains items covering all programs in a single form. The other four are ***program-specific formats*** since the items used as measures vary by program so that each program has its own form.

General formats:

Unit Design/Function Worksheet - DFW
Job/Unit Design Questionnaire - JDQ
Community Member Interview - CMI

Program-specific formats:

Care/Counseling Services Simulation Exercise - CSX
Job Knowledge Examination - JKE
On-site Observation Checklist - OSC
Performance Self-Report - PSR

Each of these formats contains a prescribed set of indicators designed to describe the direct service activities of the health center, the structural context for these activities, and their outcomes in the community. Additional formats are under consideration or in development that focus more immediately on support service activities and their structural context. For the present, however, the instrument set is heavily oriented toward direct services.

The indicators themselves will be described in detail in a subsequent section, but a summary listing will be presented here to give an overview of the contents of each format. ***Indicators*** are of the two types: ***concept-referenced*** and ***task-referenced.***

In addition, certain instruments contain measures of dimensions of the overall organization context and structure of the health center or of individual characteristics of subjects. The measures of these dimensions are distinguishable from the two defined indicator types in that they are not operational referents of a given concept nor are they measures of attainment. To maintain this distinction, we have used the term ***contextual factors*** when referring to these dimensions.

Some indicators and contextual factors are measures of the health center as a whole unit; these are called ***global.*** However, the majority of indicators and contextual factors are used to measure specific programs and these are, logically, called ***program-specific.*** Program-specific indicators are measured by items that are, themselves, either ***constant,*** meaning that the same set of items is used as a measure irrespective of the program, or ***divergent,*** meaning that different items are used as measures of the indicator for each program.

The definition of an instrument also includes the specification of a standardized, formal ***situation*** in which each of these formats is presented to subjects providing the information. The situation refers to the interaction between the assessment specialist and the subject(s), how answers are established, and who fills in the form.

We are not listing the specific items used to measure each indicator for reasons of space and clarity. All items used for a given indicator can be easily studied in the copy of the appropriate form that has been included in Volume 5 of this final report. This, plus a review of the notes in the manual that accompany each instrument in the same volume, represent the complete measurement of the indicator. At present, Volume 5 is in Spanish. In future, an English version of this Volume will be produced as well.

The section immediately following this provides verbal descriptions and discussions of many of the most important indicators used in the SAM. This section is still in development but has details on most of the more complex measures.

a. DFW: Unit Design/Function Worksheet

The DFW is filled out by the assessment specialist as he/she carries out a small group interview with the health center's director and management team. The DFW serves as a checklist to guide the interview. The frame of reference concerning answers is that they reflect the majority opinion of the HC management team or are the result of document review done by the management group in the presence of the assessment specialist.

The DFW provides most of the organization functional information for the assessment. It also contains the indicators covering basic characteristics of the relationship between the unit being studied and both its UTES and its catchment community.

THE CONTEXTUAL FACTORS INCLUDED IN THE DFW ARE:

Geographic siting of unit and ancillary health posts (GLOBAL)
Accessibility (GLOBAL)
Size of unit and number of sections (GLOBAL)
Organization of unit and number of personnel (GLOBAL)
Official management structure of unit and programs/services (GLOBAL)
Coverage goals (PROGRAM-SPECIFIC)
Unit supply system (PROGRAM-SPECIFIC)
Transport system (GLOBAL)
Unit communication system (GLOBAL)
Unit training system (PROGRAM-SPECIFIC)
Demographic profile of catchment population (GLOBAL)
Relative service availability and alternative providers (GLOBAL)

Two identical sets of indicators are included in the DFW concerning relationships between the unit and outside entities. One set refers mainly to relations between the health center and the UTES, while the other focuses on relations with the community it serves. These indicators are all concept-referenced and have been translated and adapted from Van de Ven and Ferry (1980).

THE INDICATORS INCLUDED IN THE DFW ARE:

Coordination (GLOBAL; 4 items for PMOH/ 8 items for community)
Formalization of the relationship (GLOBAL; 4 items)
Communication (GLOBAL; 8 items)
Consensus/Conflict (GLOBAL; 5 items)
Conflict resolution (GLOBAL; 4 items)
Influence in/of unit (GLOBAL; 4 items)
Effectiveness of the relationship (GLOBAL; 6 items)

b. JDQ: Job/Unit Design Questionnaire

The JDQ is filled out by the subjects directly in a guided questionnaire situation. This refers to the following process:

1) *the assessment specialist introduces the questionnaire to a group of subjects, explaining clearly how to interpret the answer scales and how to mark their responses;*

2) *he/she gives them a chance to read through it quickly and answers general questions;*

3) *he/she reads each item and its cues slowly and clearly, then leaves 10 - 15 seconds for people to respond;*

4) *at the end of the questionnaire, the assessment specialist permits individuals to ask for further explanation on questions they did not understand and leaves 15 minutes for people to go back and review their responses.*

The JDQ contains most of the indicators that cover direct services structure and some that cover support services activities.

The JDQ is filled out by unit members responsible for: A) program direction or supervision and B) direct service delivery. The frame of reference concerning answers is that they reflect the point of view of the subject acting as a respondent rather than an informant.

Indicators, with one exception, are program-specific, but the items are all constant. The JDQ also contains a number of contextual factors in the form of individual characteristics of the participating subjects.

Responses to each item are recorded separately for each program. The subject is told which programs to consider when giving responses; these are based on his/her recent level of work effort in each program. Average values for each indicator are calculated for each program but responses are tabulated separately for management and for the worker group. An average value for the unit as a whole is then calculated from these two separate averages (giving each equal weight).

THE INDICATORS (ALL PROGRAM-SPECIFIC) INCLUDED IN THE JDQ ARE:

Job specialization (calculated from a matrix)

Unit standardization (6 items)
Job standardization (4 items)

Work interchangeability (4 items)

Job priority (3 items)

(Distribution of unit authority)
Program director authority (3 items)
Supervisor authority (3 items)
Unit employee authority (3 items)
Unit collegial authority (3 items)
Outsider (PMOH) authority (3 items)
Community authority (3 items)

Job authority (4 items)
Job pressure (3 items)
Job accountability (7 items)
Job feedback (8 items)
Task difficulty (3 items)

(Incentives)
Unit incentives (8 items)
Job Incentives (4 items)

Communications in unit (3 items)
Unit conflict (5 items)
Methods of unit conflict resolution (4 items)

Satisfaction with unit support systems (7 items)
Job satisfaction (7 items)
Job training (5 items)

Job logistics support (12 items)

Growth need strength of job incumbent (12 items)

Perceived unit performance (8 items)

The only contextual factors in the JDQ are included under the category "Other/Individual differences" and include: length of time working for the PMOH and the current health center; gender; age; number of dependents; educational level in years and degrees; can subject name program director and supervisor.

c. CMI: Community Member Interview

The CMI is filled out by the assessment specialist as he/she carries out a person-to-person interview with a mother selected from the catchment community of the health center. The CMI serves as a checklist to guide the interview. The frame of reference concerning answers is that they reflect the immediate response of the subject, unprompted and uninterpreted by the interviewer.

The CMI contains all of the indicators that deal with coverage plus all of the indicators that cover outcomes in the community.

Mothers selected as subjects for the CMI are chosen from two populations: those within a 1 km radius of the health center and those from the populations furthest away (in terms of travel time/difficulty) from the center. The second selection criterion is that mothers must have at least one child less than 18 months of age living with them. The first criterion is designed to elicit responses from the groups with greatest and least physical access to the health center. The second is designed to focus on the mothers who are likely to have the most need for health center services in the recent past.

The frame of reference for questions concerning the child's health care experience is that the interview will focus on the youngest child in the family, who must be less than 18 months of age.

The CMI has been designed to cover all programs within a single format and contains global indicators, program-specific indicators with constant items, and program-specific indicators with divergent items.

THE CONTEXTUAL FACTORS (ALL GLOBAL) INCLUDED IN THE CMI ARE:

Address
Mother's name
Mother's age
Mother's level of education

Number of children living at home
Number of children by age among <5 yrs
How many children mother wanted to have
How many children more mother wants to have
Household sanitary status (4 items)

HEALTH SYSTEM PERFORMANCE INDICATORS INCLUDED IN THE CMI ARE:

Health services access (GLOBAL; 6 items)

(Coverage - PROGRAM-SPECIFIC; focus child or mother)

Latest diarrhea treatment at HC/PC (4 items)
Facilities used during latest diarrhea (10 items)

Facilities used during latest IRA (10 items)

Has child's Carnet (1 item)
Growth/dev. is correctly filled out in Carnet (3 items)
Frequency of well-child exams (2 items)

Has child's vaccination record (1 item)
Child's vaccinations correctly recorded (2 items)
Child's vaccinations up to date for age (4 items)

Time since last PAP examination (1 item)
Facilities used during latest delivery (7 items)
Currently using contraceptive measure (1 item)
Mother's tetanus vaccination in latest pregnancy (1 item)

Participation in talks given by HC/PC (7 items)
Practices during latest diarrhea (9 items)
Practices during latest IRA (6 items)

OUTCOME INDICATORS INCLUDED IN THE CMI ARE:

(Morbidity - PROGRAM-SPECIFIC; refer to focus child)
Diarrhea morbidity - day before (4 items)
Diarrhea morbidity - last 2 weeks (4 items)
IRA morbidity - day before (7 items)
Malnutrition morbidity (2 items)
Other infectious disease morbidity (6 items)

Mortality in past year (PROGRAM-SPECIFIC; 4 items)

(Basic knowledge)

- Basic ideas about diarrhea (7 items)
- Signs to take child with diarrhea to HC/HP (10 items)
- Ideas about treating diarrhea (7 items)
- Basic ideas about IRA (12 items)
- Signs to take child with IRA to HC/HP (7 items)
- Ideas about treating IRA (6 items)
- Basic ideas about growth & development (5 items)
- Purpose of vaccinations (1 items)
- Doses of vaccines to protect (5 items)
- Optimal age of vaccination (4 items)
- Basic ideas about Maternal Health (5 items)
- Signs during pregnancy to go to HC/HP (8 items)
- Signs after delivery to go to HC/HP (6 items)
- What is PAP test (1 item)
- Knowledge of contraceptive methods (10 items)
- Potential complications with contraception (5 items)
- Signs during contraceptive use to go to HC/PC (4 items)

(Satisfaction - for <u>either</u> health center or health post. Note: Subject is asked to focus on the facility used "most often" in the past 3-6 months and all questions asked in that context)

- Satisfaction with access (5 items)
- Satisfaction with health services rendered (7 items)
- Satisfaction with efficiency / fairness (4 items)
- Satisfaction with humaneness ...
- ... at admission (7 items)
- ... during triage (7 items)
- ... while receiving health services (7 items)

(Community participation)

- Participation of organizations (9 items)
- Community health activities (10 items)

d. CSX: Care/Counseling Services Simulation Exercise

The CSX is a checklist-controlled simulation, or role-playing exercise (SIMULEX). The CSX instrument consists of a detailed checklist and a guide (with scripts) to performing and observing the simulation exercise. A different CSX is needed for each program to be evaluated.

The CSX is performed by participating unit members responsible: a) for program direction or supervision, and b) for direct services delivery.

The SIMULEX protocol is kept as simple as possible. One of the assessment team members acts as a surrogate mother with child needing attention. A second

member acts as a new health auxiliary to whom the subject is to demonstrate what is to be done to deal with the problem or need presented. The team assures that all supplies and equipment necessary for proper service delivery are at hand at the SIMULEX site. A doll is used in certain instances to simulate the child.

The subject is presented with a situation, or ***vignette***, that closely approximates one of the common or most important service situations he/she faces in the program being assessed. Since his/her role calls for "teaching" the surrogate health auxiliary, it is stressed that he/she should explain every step in as much detail as practicable. The surrogate-student/observer stands at the side and unobtrusively scores the exercise while continuing to monitor the effort and asking questions when necessary.

Each subject is debriefed immediately after a SIMULEX exercise in a short training session that points out what he/she did exceptionally well and what areas need improvement.

Indicators are program-specific, some with constant but most with divergent items to measure them in each program.

Scores are reported separately for each subject and as averages for each of the two job categories. The frequency of correct responses on individual items is tabulated separately to identify those items with exceptionally high or low values.

In the following listing, it has been necessary to add a level of detail to permit us to differentiate between programs. The codes used to indicate the different programs are as follows: CED - diarrhea; IRA - acute respiratory infections; PAI - immunizations; CRE - growth & development; PFM - family planning (FP) / maternal health (MH). FP and MH have been combined because care is, in fact, usually given in the same visit by health workers responsible for both programs. The codes used come from the Spanish acronyms used in the PMOH (which are also those used in Volume 5).

Indicators for some programs are far more extensive than are those for others. This is a result of the nature of the service being provided. History-taking, the first indicator listed below, for example, is a short, simple process during immunizations but represents a significant effort for maternal health visits. The operational model deals with this by admitting ***sub-categories*** for aggregating measures within such complex indicators.

The number of items per indicator will be shown in a simple table to the right of the indicator. If the indicator is complex, a "*" will appear indicating that a fuller explanation is necessary. For the FP/MH form, some indicators are in the process of being operationalized and we do not have a final set of measures. This will be indicated by a "?"

THE INDICATORS IN THE CSX (ALL PROGRAM-SPECIFIC) ARE:

History-taking

CED	IRA	PAI	CRE	PFM
8	9	3	*	*

- ***CRE:*** **Child's history** (13 items); **Family history** (15 items)

  PFM: **General** (10 items); **Physiologic** (6 items); **Obstetric** (9 items); **Contraceptive** (3 items)

Physical examination

CED	IRA	PAI	CRE	PFM
12	11	1	*	?

- ***CRE:*** **Weight** (6 items max.); **Height** (4 items max.); **Other measurements** (4 items); **Direct physical exam** (28 items); **Psychomotor development** (6 items)

Paraclinical services (not currently in use)

Diagnosis

CED	IRA	PAI	CRE	PFM
2	2	6	0	?

Treatment strategy

CED	IRA	PAI	CRE	PFM
4	3	0	0	?

(Treatment technique)

Preparation

CED	IRA	PAI	CRE	PFM
11	0	*	0	?

- ***PAI:*** **Sterility** (71 items); **Expiration** (8 items); **Preparation/handling** (30 items); **Cold chain maintenance** (20 items)

Administration

CED	IRA	PAI	CRE	PFM
12	0	*	0	?

- ***PAI:*** **Application technique** (28 items)

Problem-handling

CED	IRA	PAI	CRE	PFM
9	0	0	0	?

Promotion/educ. content

Case-specific content	CED	IRA	PAI	CRE	PFM
	*	11	22	8	?

- *CED:* Treatment principles (8 items); Preparation/ use of ORS (9 items)

General content	CED	IRA	PAI	CRE	PFM
	*	7	4	9	?

- *CED:* Signs of dehydration (9 items); Prevention of diarrhea (4 items)

Promotion/educ. strategy	CED	IRA	PAI	CRE	PFM
	8	5	5	2	?

Documentation	CED	IRA	PAI	CRE	PFM
	0	0	2	10	?

Affect	CED	IRA	PAI	CRE	PFM
	7	7	7	7	7

Task satisfaction	CED	IRA	PAI	CRE	PFM
	6	6	6	6	6

Humaneness satisfaction	CED	IRA	PAI	CRE	PFM
	6	6	6	6	6

e. JKE: Job Knowledge Examination

The JKE is a written examination consisting of 30-45 questions (multiple-choice, true-false, and list-matching) designed to test the basic knowledge of health workers in the programs in which they have been working most actively in the 3-6 months before the assessment.

The JKE instrument consists of an examination form, an answer guide, and a manual for running the examination. A different JKE is needed for each program to be evaluated.

The JKE is designed to supplement the CSX for the same program by testing basic knowledge over a wider range of issues and factual situations than can practically be covered in a simulation exercise.

The JKE is taken by participating unit members responsible: a) for program direction or supervision, and b) for direct services delivery.

The assessment team member running the examination explains the reasons for having an examination (i.e., to identify common problem areas among the members of the unit) and reviews the types of questions that will be asked. The assessment specialist gives the group 10 minutes to read quickly through the examination (asking them not to mark answers during this period) and explains words or phrases that are unclear.

He/she then reads each question aloud with the set of possible answers and gives the subjects 15 seconds to select and answer. The group proceeds together through the examination following the lead of the assessment specialist. At the end, the group is given 15 minutes to return to left questions and to review answers. Each examination is scheduled to take 45-60 minutes.

The score of correct answers on the JKE is the measure of the only indicator calculated for this instrument. It is program specific and is measured by divergent items for each program.

Scores are reported separately for each subject and as averages for each of the two job categories (worker and manager). The frequency of correct responses on individual questions in the examination is tabulated separately to identify those items with exceptionally high or low values for the unit.

THE ONLY INDICATOR FOR THIS INSTRUMENT IS:

Basic knowledge (30-45 items)

f. OSC: On-site Observation Checklist

The OSC is filled out by the assessment specialist during an unannounced visit to sites at which direct services are provided. The observations are controlled by the checklist. Multiple types of sites may be included in a unit evaluation (e.g., the health center is a different site-type than its group of ancillary health posts).

Indicators in the OSC cover facilities and material, and record-keeping aspects of health services delivery. Indicators are program-specific so a different OSC is needed for each program to be evaluated. Some items are constant and some are divergent.

Scores are reported individually for a site or as the average for each site-type. These individual or averaged scores may be kept program-specific or aggregated into a global value for the individual site or the site-type.

In the following listing, it has been necessary to add a level of detail to permit us to differentiate between programs. The codes used to indicate the different programs are as follows: CED - diarrhea; IRA - acute respiratory infections; PAI - immunizations; CRE - growth & development; PFM - family planning (FP) / maternal health (MH). FP and MH have been combined because care is, in fact, usually given in the same visit by health workers responsible for both programs. The codes used come from the Spanish acronyms used in the PMOH (which are also those used in Volume 5).

Indicators for some programs are far more extensive than are those for others. This is a result of the nature of the service being provided. History-taking, the first indicator listed below, for example, is a short, simple process during immunizations but represents a significant effort for maternal health visits. The operational model deals with this by admitting ***sub-categories*** for aggregating measures within such complex indicators.

The number of items per indicator will be shown in a simple table to the right of the indicator. If the indicator is complex, a "*" will appear indicating that a fuller explanation is necessary. For the FP/MH form, some indicators are in the process of being operationalized and we do not have a final set of measures. This will be indicated by a "?"

THE INDICATORS (ALL PROGRAM-SPECIFIC) INCLUDED IN THE OSC ARE:

Facilities	CED	IRA	PAI	CRE	PFM
	8	2	4	5	4

Equipment	CED	IRA	PAI	CRE	PFM
	7	6	*	16	37

* *PAI:* **Refrigerator** (13 items); **Main Ice Box** (4 items); **Aux. Ice Box** (3 items); **Manual** (1 items);

Supplies	CED	IRA	PAI	CRE	PFM
	8	9	22	8	23

Preparedness ...

... Care-assoc.	CED	IRA	PAI	CRE	PFM
	6	4	4	4	4

... Prom./Educ.-assoc.	CED	IRA	PAI	CRE	PFM
	5	8	7	7	4

Patient record ...

... maintenance	CED	IRA	PAI	CRE	PFM
	3	3	3	3	3

... item completion	CED	IRA	PAI	CRE	PFM
	7	7	7	7	7

Daily register ...

... maintenance	CED	IRA	PAI	CRE	PFM
	7	7	7	7	7

... item completion	CED	IRA	PAI	CRE	PFM
	4	4	4	4	4

Monthly register ...

... maintenance	CED	IRA	PAI	CRE	PFM
	3	3	3	3	3

g. PSR: Performance Self-Report

The PSR is filled out by the subjects directly in a guided questionnaire situation (see details above under JDQ).

The PSR contains three, particularly broad, task-referenced indicators that cover current problems affecting direct and support services delivery. There are two criteria for inclusion of a specific problem as a measure: 1) consensus by the F/I Groups that a problem is pervasive and important, and 2) decision by project staff that the problem is not (and cannot be) covered by an already existing measure.

The PSR, thus, represents an effort to elicit reports from the subjects themselves on service problems they may be having or errors they may be committing. From a theoretical standpoint, the three indicators represent a flexible interface to aspects of the service process that might otherwise be missed. In any case, this is the context in which they are interpreted.

The PSR is filled out by unit members responsible for: A) program direction or supervision and B) direct service delivery. The frame of reference concerning answers is that they reflect the point of view of the subject acting as a respondent rather than an informant.

Indicators, with one exception, are program-specific, and most of the items are divergent.

Responses to each item are recorded separately for each program. The subject is told which programs to consider when giving responses; these are based on his/her recent level of work effort in each program. Average values for each indicator are calculated for each program but responses are tabulated separately for management and for the worker group. An average value for the unit as a whole is then calculated from these two separate averages (giving each equal weight).

The number of items included to measure an indicator in a given program is intended to vary as problems are introduced to or removed from the set. The range of items currently for all programs is shown in parenthesis.

THE INDICATORS (ALL PROGRAM-SPECIFIC) INCLUDED IN THE PSR ARE:

Problems in care activities (7-12 items)
Problems in promotion/educ. activities (3-8 items)
Problems in support activities (0-3 items)

h. Future instruments

A number of instruments are being designed and developed for future implementation. The first two have reached a stage at which a specific set of indicators is being suggested. The others are still at the general concept stage.

1) Supervision Simulation Exercise and Interview

To measure competence in supervisory activities in a SIMULEX resembling that used to measure competency in direct services delivery. The indicators suggested to date are:

Listen effectively
Identify problems
Evaluate problems
Produce solutions
Give directions
Give information
Use effective training methods
Testing for understanding
Comportment (Affect)
Task satisfaction
Humaneness satisfaction
Supervisory awareness of performance
Supervisory awareness of resource availability

2) Observed Teaching Exercise

To measure competency in formal teaching situations. This is intended to be a SIMULEX-like exercise but using a real group of people to whom a class will be given. The intention is to measure how effective the teaching is of persons who are responsible for running training sessions or giving community talks. Indicators suggested to date are:

Message content-Directions
Message content-Information
Emphasis on priority messages
Teaching strategy
Test for understanding
Maintaining attentiveness
Use of time
Lesson development
Comportment (Affect)
Task satisfaction
Humaneness satisfaction

3) Support systems service assessment (laboratory, etc.)

4) Critical incidents checklist

5) Observer's overall evaluation sheet

5. Verbal descriptions of selected indicators

We are in the process of preparing a verbal description of each indicator included in the SAM which will discuss its purpose, contexts for its use, how it is to be interpreted, the history of its development and testing, etc.

This development is still in its early stages but the material completed to date is presented below, arranged by format.

a. DFW: Unit Design/Function Worksheet

Geographic siting of unit and ancillary health posts - Distances to hospital, UTES, UDES; distances to all catchment populations

Accessibility - Hours of service; cost of services; off-hour availability

Size of unit and number of sections - Diagram of building; number of rooms; space allocation for services/programs

Organization of unit and number of personnel - Organization chart; number of personnel by category in unit

Coverage goals - Planned; degree met; written goals

> *Coverage is defined as the extent of care-giving and/or promotion/education activities as an absolute number, relative to program goals, and relative to health problem profile in target group.*
>
> *These are measured by a review of pertinent registers from the proceeding year. The denominator of persons in the catchment population is obtained from census statistics gathered as part of the DFW worksheet.*

b. JDQ: Job/Unit Design Questionnaire

Job specialization - Job specialization involves the number of different tasks performed by the job incumbent and the scope or breadth of these tasks (Van de Ven & Ferry, V&F, p. 386).

> *Job specialization is measured by asking the job incumbent, acting as respondent in a global context only, to indicate which services he/she performs for each of the programs actively carried out in the unit. The approach used here has been developed specifically for the current SAM. The standardized list is presented as a matrix and the job incumbent is asked to score each service in each program with the amount of time he*

spends per week. The number of activities that occupy 5 or more hours of the respondents work time per week multiplied by the number of Programs that do the same is then used as the measure of job specialization.

Unit standardization - Clarity of unit performance standards; preciseness of unit rules, policies, procedures; degree performance criteria quantified; percent unit rules, procedures written out; extent rules violated; strictness of rule enforcement.

Job standardization - Job standardization is the degree to which the roles and tasks that make up a job are clearly detailed and the rules and procedures clearly established to guide the job incumbent in work performance (V&F, p.386).

Job standardization is measured as the average of the following six items asked of the job incumbent as respondent: Number of written job rules; detail of job rules; percent time have SOP's; extent follow SOP's; clarity of job performance standards; extent job description specifies performance standards

Job priority - Job priority is the importance given to the job done for a given program in its competition for time and resources with other programs. This variable has been developed for the current SAM.

Job priority is measured as the average of the following three items asked of the job incumbent as respondent: Compared to what you do in other programs, your job in this program merits how much of ... your time ... support services ... emphasis from "the system"

Distribution of unit authority - Unit employee authority, unit and program supervisor authority, unit collegial authority, external PMOH authority, and community authority measured as: Say on unit tasks; say on performance criteria; say on performance appraisal; say on rules, policies, procedures

Job authority - Job authority is defined as the amount of discretion or influence that the job incumbent exercises in making job-related decisions regarding: (a) what tasks, projects, and assignments constitute the roles and responsibilities of the job; (b) how the work is to be done in terms of what procedures and rules to follow; (c) how work exceptions and problems are to be handled; and (d) what performance criteria are established and to be attained in performance appraisals (V&F, p. 387).

Job authority is measured as the average of the following four items asked of the job incumbent as respondent: Decide what tasks to perform; decide work rules and procedures; decide how to handle exceptions; decide work quotas and standards

Job pressure - Job pressure refers to the amount of work load assigned to a job incumbent, the lead time available to perform it, and the extent to which the job incumbent can control the pace of his/her work. High amounts of job pressure imply that the job incumbent can exercise little job discretion (V&F, p. 387).

> *Job pressure is measured as the average of the following four items asked of the job incumbent as respondent: Heaviness of work load; control over work pace; work lead time; difficulty achieving performance standards*

Job accountability - Job accountability is the degree to which the job incumbent feels personally responsible and feels that he or she is, in fact, asked to answer for his or her work decisions and behavior (V&F, p. 388).

> *Job accountability is measured as two variables: perception of how much the "system" actually holds the job incumbent accountable -- the average of the first two items; and personally felt accountability -- the average of the last four items. All items are asked of the job incumbent as respondent: Held accountability - for work decisions and for achieving standards; Felt accountability - Fairness of job appraisal standards; take credit or blame for work results; feel personally responsible for work; don't care if work done right*

Job feedback - Job feedback is the degree to which the job incumbent receives information about the procedures and results of his/her work efforts. This can be feedback from the job itself (simply by assessing the procedures and the results of one's own work) and feedback from others (supervisors and co-workers) (V&F, p. 389).

> *Job feedback is measured as the average of the following seven items asked of the job incumbent as respondent: Feedback from job; feedback from co-workers; feedback from supervisor - frequency of meeting with supervisor; time since last meeting with supervisor; frequency with which supervisor "gets back" with solutions to problems; degree supervisor discusses performance standards; frequency of practical suggestions from supervisor; supervisor is more "critic" than "teacher"*

Incentives - Expectation of rewards refers to the degree to which the job incumbent anticipates that good job performance will result in some reward (V&H, p. 389). Expectation of sanctions refers to the degree to which the job incumbent anticipates that poor job performance will result in some punishment (V&H, p. 389).

Expectation of rewards is measured as the average of the following three items asked of the job incumbent as respondent: Recognition for good job; chance of promotion for good job

Expectation of sanctions is measured as the average of the following three items asked of the job incumbent as respondent: Reprimand for poor work; chance of demotion for poor work

Task difficulty - Task difficulty refers to the ability of the job incumbent to understand the characteristics of the work encountered: in other words, the analyzability and predictability of the work (V&F, p. 392).

Task difficulty is measured as the average of the following four items asked of the job incumbent as respondent: Difficulty of knowing work correct; unsure of work outcomes; frequency problems arise; time spent solving problems; access to expert advice when needed (from supervisor, from other unit members)

Unit conflict - Frequency of supervisor-subordinate conflict; frequency of conflict among unit members; frequency of conflict with other units; members get ahead at expense of others; agreement on unit performance criteria

Methods of unit conflict resolution - by avoiding issues; by smoothing over issues; by confronting issues; by hierarchy

Satisfaction with unit support systems - Satisfaction with unit support systems is defined as the degree to which the job incumbent feels that the other elements within the unit succeed in providing the support expected to the work that he or she is doing. This is a variable newly added for the current SAM.

Satisfaction is the average of five items asked of the job incumbent as respondent: Job receives adequate management/planning; job receives adequate supervisory support; job receives adequate logistics support; job receives adequate training support; job receives adequate information/feedback support

Job satisfaction - Job satisfaction is an affective reaction or feeling by the job incumbent on how happy or satisfied he or she is with the various key aspects of his or her job (modified from V&F, p. 390).

Job satisfaction is measured as the average of the following nine items asked of the job incumbent as respondent: satisfied with job; satisfied with immediate supervisor; satisfied with pay; satisfied with co-workers; satisfied with past career; satisfied with career potential; often thinking of quitting;

satisfied with status in the community; satisfied with physical work environment

Job training - Job training is the amount of educational preparation for the job in terms of formal education, length of job-entry orientation and training, and the amount of time spent by the job incumbent in on-the-job training and reading necessary for upgrading and remaining current in the knowledge needed to perform the job (modified from V&F, p. 386).

Expertise of job incumbent is assessed with a series of independent measures that are not averaged to yield a composite score. The specific items developed by Van de Ven and Ferry are not appropriate for the systems to which the current SAM is being applied and new items have been developed. The following six items are asked of the job incumbent as respondent, the first two in a global context only and the last four in either a global or program-specific context: Years of formal schooling; highest educational degree; length of job-entry training; time in self-generated OJT; frequency of systematic OJT; most recent systematic OJT

Growth need strength of job incumbent - Growth need strength refers to the degree to which the respondent desires to fulfill self-actualization needs from his or her job. This construct is thought to moderate the relationships of job design characteristics with job satisfaction and motivation. Individuals high on growth needs respond positively to complex, challenging jobs while individuals low on this factor tend to find such jobs unsatisfying and unmotivating (V&F, p. 393).

The "forced choice" index developed by Richard Hackman for the Job Diagnostic Survey is being used in the current SAM as it was by V&F. Growth need strength is measured as the average of the following twelve items asked of the job incumbent as respondent in a global context only:

Prefer creativity over pay
Prefer pleasant people over important decisions
Prefer loyalty over responsibility
Prefer no discretion over financial trouble
Prefer unfriendly workers over routine job
Prefer no-skill job over critical supervisor
Prefer learning over supervisor respect
Prefer no challenge over chance of layoff
Prefer fringe benefits over job skill development
Prefer poor work conditions over little freedom
Prefer personal skill use over teamwork
Prefer isolated job over no challenge

Other/Individual characteristics - A number of individual difference characteristics of job incumbents are important for interpreting job outcomes and performance.

These do not form a single composite score that has any intrinsic meaning and are treated as individual measures.

The following items asked of the job incumbent as respondent: Tenure in organization; tenure in current position; tenure in current program; age of job incumbent; sex of job incumbent; number of dependents

c. CSX: Care/Counselling Simulation Exercise

Clinical assessment is the ability of a care-giver to gather, examine and correctly note the pertinent and important signs and symptoms associated with a patient being examined. This overall concept consists of three dimensions, each of which is treated as a distinct indicator: **History taking**, **Physical examination**, and **Paraclinical services**. The definition of each of these variables is as follows:

History taking - History taking covers all verbal aspects of clinical assessment, including the asking of appropriate open and closed questions of patient characteristics and symptoms, and success in eliciting patient disclosure of pertinent information).

Physical examination - Physical examination includes all physical contact between the care-giver and the patient involved in the evaluation of signs pertinent to the complaint or reason for the encounter.

Use of para-clinical services - Use of para-clinical services refers to the ordering, obtaining specimens for, and reviewing the results of ancillary diagnostic procedures carried out by a laboratory or other expert who is not the immediate care-giver. NOTE: This indicator has not currently been implemented in the SAM.

The three indicators are taken to be measures of a health worker's maximal ability in this area. All three are determined by observation of the job incumbent's performance in a simulation exercise (role-playing).

A trained observer uses a standardized checklist of items which represent consensus criteria of best practice for the care-giving encounter under consideration. The number of items will vary with the program being evaluated.

The simulation exercises are used simultaneously as a training encounter for the health worker and are presented to him or her in this light. To further reduce the evaluative aspect of the exercise for the subject, he or she is shown that the scoring is being recorded in a way that does not directly provide individual identification (e.g., through the use of an ID code known only to the individual).

The simulation exercise calls for the health worker physically to carry out the steps he or she thinks necessary using appropriate surrogates for a real patient. Depending on the exercise, this will be a member of the evaluation team, a doll, photographs, or verbal analogues. In some exercises, actual patients or care-recipients may be used if a consistent patient group can be scheduled without difficulty (e.g., for well-child examination, routine immunization, routine pre-natal examination, etc.)

The simulation exercise and criteria checklist are program-specific and several different sets may be used to cover the care-giving aspects of a given program (e.g., the Maternal Health module has three simulation exercises associated with care-giving). The consensus criteria are based on international and local norms of recommended practice as interpreted by working groups of experienced health workers from the system under study.

The number of items included for a particular indicator may vary greatly depending on the nature of the exercise. Criteria are scored on a 4-point scale: 0 - not performed; 1 - performed but so incorrectly as to be useless; 2 - performed with some skill and understanding but not correctly; and 3 - performed correctly. Each indicator is measured as the cumulative score obtained for these criteria by the job incumbent in the chosen simulation exercise.

All measures are exercise-specific but could be averaged across several exercises, if more than one in a program, to produce a program-specific score, or across programs to produce a global score. Each indicator is considered separately in evaluating a health worker's overall performance in clinical assessment.

Job incumbents engaged primarily in supervision or training will also be evaluated as direct service care-givers along with actual direct care providers.

The clinical assessment indicators are not intended to capture aspects of a workers performance for every major variation which may present in patients coming for treatment. A comprehensive evaluation of these indicators would, otherwise, entail a great many simulation exercises. This is an unmanageable course for practical application as a routine evaluation tool.

Each simulation exercise has been limited to one of the major, clinically significant situations likely to be encountered and this is used to measure each indicator of clinical assessment as a general ability. The specific knowledge needed to carry out clinical assessment of other important situations is evaluated by the inclusion of test items in the Content knowledge examination. Both types of indicators -- practical general ability in the simulation exercise and content knowledge of variations -- are used

to evaluate a health worker's ability in this area. Poor scores in one or both types of indicator indicate a serious problem that is likely to affect the care given to actual patients.

Diagnosis - Diagnosis refers to the critical analysis of data obtained from clinical assessment in order to identify or determine the nature of the clinical problem or state present in the patient being examined. In context of the CSX, this variable is limited to an indicator of practical diagnostic proficiency: the ability to come up with a correct diagnosis in a real-life or simulated situation based on the data at hand.

Practical diagnostic proficiency is measured as the next step after clinical assessment in the care-giving simulation exercise. The details of the exercise and measurement process are as described in the section on clinical assessment. A trained observer uses a standardized checklist of items which represent consensus criteria of best practice for the care-giving encounter under consideration. The number of items will vary with the program being evaluated.

It is clear that clinical assessment profoundly affects Practical diagnostic proficiency in these simulation exercises. Making a mistake in any part of the former can change the diagnosis. Given incomplete or incorrect criteria, the health worker may make a diagnosis that is not wrong based on the knowledge he or she possesses, but one that is, nevertheless, incorrect in light of the established situation represented in the exercise.

One way of dealing with this is to have the observer evaluate the diagnosis based on the clinical assessment criteria actually generated by the health worker even though these criteria are incorrect. To us, this places an unacceptable analytical burden on the observer and, far more importantly, is antithetical to one of the main reasons for using simulation exercises in the first place: to present substantially the same problem situation to each of the health workers being evaluated so as to have a comparable basis for assessing their performance relative to a norm.

The approach used is for the observer to "step into" the exercise at the point at which the clinical assessment is finished and tell the worker if his or her criteria are correct. If they are, the observer is simply to note that fact without additional comment. If they are not, the observer is to prompt the worker on his or her mistakes without a great deal of discussion. The observer is always to interact, whether or not the criteria are correct, in order not to introduce the interaction itself as a difference.

The number of items included for a Practical diagnostic proficiency may vary greatly depending on the nature of the exercise. Individual diagnosis items are scored on a 3-point scale: 0 - not mentioned or completely wrong; 1 - mentioned but partially correct; and 2 - mentioned and correct.

Individual item scores are weighted as necessary to reflect the relative proportion of the diagnostic effort they represent (e.g., the diagnosis of dehydration in the management of diarrhea is easily half or more of the total effort to which clinical assessment is dedicated while other diagnoses, such as chronic diarrhea, are based on a single data item).

The final indicator is the sum of these weighted scores. All measures are exercise-specific but could be averaged across several exercises, if more than one in a program, to produce a program-specific score, or across programs to produce a global score.

The comments about not trying to capture all aspects of performance of clinical assessment during the simulation exercise also apply to Practical diagnostic proficiency. In this case, the effort is to test the general ability of the worker when faced with a simulation of a real-life situation. The test of diagnostic reasoning ability in the JKE contains items covering the majority of important different situations which the worker might face. The two indicators for diagnosis are used separately to evaluate a health worker's performance in this area.

Treatment strategy - Treatment refers to the ability of a care-giver to select and apply remedies or therapy in response to a given diagnosis with the object of affecting a cure. It also refers to preventive care given to maintain a desired state. This overall variable consists of two dimensions, each of which is treated as a distinct indicator: Treatment strategy and treatment technique.

The definition of the first of these is as follows: Treatment strategy covers the selection of the optimum action(s) to be taken in response to a given diagnosis. It measures the appropriateness of the treatment without regard to how that treatment is implemented.

Treatment technique - Treatment technique refers to the technical skills demonstrated during the implementation of the selected treatment. It measures the detailed operational familiarity with the physical reality of actually applying a given treatment rather than the content knowledge of the verbal description for that treatment.

***Treatment strategy** and **Treatment technique** are measured as the next two steps after diagnosis in the care-giving simulation exercise. The details of the exercise and measurement process are as described in the section on clinical assessment. A trained observer uses a standardized checklist of items which represent consensus criteria of best practice for the care-giving encounter under consideration. The number of items will vary with the program being evaluated and the indicator, calculated as the cumulative score for all observed items, is, therefore, always program-specific.*

The discussion in the section on diagnosis regarding the need for the observer to step into the exercise and ensure that the clinical assessment criteria are correct before proceeding to observe the diagnosis also applies here. Evaluation of Treatment technique assumes that the health worker is employing the correct strategy. At both stages, observer interaction is called for to keep the simulation exercise on the chosen path.

As with clinical assessment and diagnosis, the measurement of Treatment is limited to one clinically significant situation. A health workers' potential for handling variant situations is tested verbally by the inclusion of appropriate questions in the CDE.

Counselling strategy - Counselling strategy is defined as the use of specific strategies for patient/guardian education in an attempt to increase the clarity and persuasiveness of the messages included in the counselling effort.

This indicator is measured as the average of the following items assessed by the observer at the end of the care-giving simulation exercise. The measure of this indicator is program-specific but may be averaged across several programs to produce a global score for an individual or unit:

The health worker ...

- ... *physically demonstrates important points*
- ... *has patient/guardian carry out important activities*
- ... *uses language appropriate to patient's/guardian's level of understanding*
- ... *has patient/guardian repeat important points of information*
- ... *makes use of available visual aids when appropriate*
- ... *provides authentic praise for appropriate behavior*
- ... *is open to questions or requests from patient/guardian at any time*
- ... *requires the patient/guardian to restate important points of information in his/her own words*
- ... *asks specific knowledge or thought questions of the patient/guardian to assess his/her understanding of the messages given*
- ... *has the patient/guardian carry out important activities without assistance to determine if the skill has actually be learned*

Counselling content - Counselling content refers to the presentation to a patient, as an integral part of a care-giving session, of specific information specified by the norms of the program being evaluated. This overall variable consists of two dimensions, each of which is treated as a distinct indicator: Case-specific messages and general messages. The definition of each of these variables is as follows:

Case specific - This indicator covers those messages which are called for in dealing with the case immediately at hand. This includes giving directions and instructions related to the clinical examination, current treatment, future treatment,

and followup. It also includes giving information and orientation about the specific nature of the existing illness and its treatment.

General - This indicator covers those messages which, according to program norms, should be presented as an educational effort during all care-giving encounters, and which are not particularly linked to the immediate case at hand. Such messages include giving information and orientation related to, and attempting to persuade the patient concerning, the general characteristics of an illness (e.g., what is diarrhea), to noting signs and symptoms, and to prevention.

Both indicators are measured as the next two steps after treatment in the care-giving simulation exercise. The details of the exercise and measurement process are as described in the section on clinical assessment. A trained observer uses a standardized checklist of items which represent consensus criteria of best practice for the care-giving encounter under consideration. The number of items will vary with the program being evaluated and the indicator, calculated as the cumulative score for all observed items, is, therefore, always program-specific.

Verbal intervention on the part of the observer is likely to be unusually disruptive of the counselling effort. Thus, the observer should not step in to the exercise during this phase. Any orientation done at the start of the treatment phase is assumed to be sufficient for ensuring a fair measurement of these indicators as well, since counselling is almost always linked intimately with the treatment effort. Some counselling may have already taken place during earlier stages and may need to be recalled.

The measurement of these indicators is limited to one clinically significant situation. A health workers' potential for handling variant situations is tested verbally by the inclusion of appropriate questions in the JKE.

Comportment (Affect) - Comportment or Affect is defined as conveyed affect, or emotion carried in the body behavior and/or manner of expression through voice quality or verbal content, rated by a third-party observing an interaction between a health worker and a patient/guardian.

This overall variable consists of four dimensions, each of which is treated as a distinct indicator:

Bored/uninvolved --Interested/concerned
Angry/irritated -- Friendly/warm
Anxious/nervous -- Calm/relaxed
Arrogant/superior -- Respectful/egalitarian

The four dimensions of comportment are each measured as a single Likert scale representing an overall assessment by the observer at the end of the

counselling effort in the in the care-giving simulation exercise. The details of the exercise and measurement process are as described in the section on clinical assessment. The indicator is program-specific but may be averaged across several programs to produce a global score for an individual or unit.

Task satisfaction - Task satisfaction is defined as the degree of patient/guardian satisfaction with the health worker's performance in task-associated behaviors during the simulation exercise. We use Roter's definition of task behavior for this indicator: "those technically based skills used in problem-solving which compose the basis of expertness for which the health worker was consulted" (Roter et al., 1987).

Task satisfaction is the average of the following five items (adapted from Roter, above) asked of the patient/guardian or surrogate participating in the simulation exercise at the end of the exercise. The details of the exercise and measurement process are as described in the section on clinical assessment. The measure of Task satisfaction is program-specific but may be averaged across several programs to produce a global score for an individual or unit:

This health worker...

... answered my questions about the problem(s)
... clearly explained what the trouble is
... told me exactly what he was doing
... told me why certain tests or procedures were being done
... clearly explained why I should do the things he asked me to do

Humaneness satisfaction - Humaneness satisfaction is defined as the degree of patient/guardian satisfaction with the health worker's affective manner during the simulation exercise (Roter et al., 1987).

Humaneness satisfaction is the average of the following six items (adapted from Roter, above) asked of the patient/guardian or surrogate participating in the simulation exercise at the end of the exercise. The details of the exercise and measurement process are as described in the section on clinical assessment. The measure of Humaneness satisfaction is program-specific but may be averaged across several programs to produce a global score for an individual or unit:

The health worker...

... sometimes interrupted me
... sometimes talked down to me
... seemed annoyed
... acted as though he were doing me a favor by talking to me
... seemed to be in a hurry
... made me feel important

d. JKE: Job Knowledge Examination

Content knowledge refers to the job incumbent's factual knowledge of a program area, including relevant biology, physiology, epidemiology, diagnostic paradigms, treatment modalities, control principles, promotion, etc.

The test also includes items designed to test diagnostic reasoning ability, which refers to the higher level ordering and processing of knowledge about signs and symptoms of a clinical condition and their relationship to the actual nature and extent of the condition in order to prioritize, filter, and select data-items and relationships so as to reach the correct conclusion with maximum efficiency.

> *Content knowledge is measured as the cumulative score on a content-valid verbal (written/oral) multiple-choice examination (25+ questions) of the subject matter applicable to the program of interest. The measure of Content knowledge is program-specific but may be averaged across several programs to produce a global score.*

e. OSC: On-site Observation Checklist

Materiel availability - Materiel availability refers to the availability of facilities, equipment, and supplies basic to the provision of the direct services called for in the program being evaluated. This concept is measured by three indicators of the same name: **facilities**, **equipment**, and **supplies**.

> ***Facilities, equipment,*** *and* ***supplies*** *are measured as cumulative scores on a standardized checklist of items which represent consensus criteria of items that should be present for a site to be considered fully ready to provide direct services in a particular program. The indicators are calculated as cumulative scores of items observed during an unannounced visit to the unit. The number of items will vary with the program being evaluated. Items serving as measures of a particularly important logistic subsystem (e.g., "cold chain", "transportation") will also be tabulated as an indicator of this subsystem as well. These measures are program-specific but may be averaged across several programs to produce a global score. In addition to simply being present or absent, some equipment will also be scored for whether it is in operative condition or not.*

Preparedness - Preparedness is defined as objective evidence that a unit is ready to deliver its direct services immediately were the need to arise.

> ***Preparedness*** *is measured as the cumulative score on a standardized checklist of items which represent consensus criteria of items that must be on-hand and ready for a site to be considered fully ready to provide direct services without delay when the need arises. The indicator is calculated as the cumulative score of items observed during an unannounced visit to the unit. The number of items will vary with the program being evaluated. The measure of preparedness is program-specific but may be averaged across several programs to produce a global score.*

Record-keeping - Three levels of record-keeping are assessed. In general, record-keeping refers to the care and accuracy with which the patient care logs and daily/monthly registers for direct services activities are maintained in the unit under evaluation.

> *Record-keeping for each type of data log or register is measured with two indicators:* ***maintenance*** *and* ***item completion*** *. Each was measured after reviewing a set number (at least 20) of the most recent entries. The overall measure for the indicator is obtained by averaging the measures for each register encountered. These measures are program-specific but may be averaged across several programs.*

f. PSR: Performance Self-Report

Items in the PSR, while grouped by category, do not function as measures of indicators, per se. The items are intended to cover important, specific aspects of performance which either: 1) reflect the most-stressed norms, or 2) reflect activities that experienced workers in the system believe may not always be done in actual practice.

6. The process model

In the process of developing the SAM, we have frequently found ourselves on the horns of a dilemma. On one hand, instruments that are simple enough to be used with minimal training by any health center management team are not going to capture enough of the management situation at any moment to be very useful. On the other hand, the final assessment package must be integrated into the existing, operational infrastructure of the Ministry if it is to achieve sustainable impact. Therefore, it cannot be so complex in its demands that it requires a permanent staff of international technical advisors.

As we have progressed, we have rejected any model of PHC management information systems that calls for simplifying to the point that only anecdotes are collected. We have, as a result, developed an approach to assessment that is complex enough to serve real operations management needs. We have kept the indicator set as parsimonious as possible considering the domain to be assessed.

There are approximately 140 indicators (excluding contextual factors) in the SAM. The organization of the formats is designed to facilitate rapid, on-site tabulation of individual measures into these indicators. It is possible, as well, to use any single instrument alone and, even, single indicators if the situation demands it. This is a flexible, yet sophisticated response to the operations management needs of the PMOH.

Nevertheless, we recognize that this package will be routinely implemented in the PMOH only if it is accompanied by a process model (the practical framework for implementing the SAM) that is within the financial, political, and human resource constraints of the PMOH. The remaining sections in this volume are dedicated to describing this process model while, at the same time, showing how the SAM will be implemented in the PMOH in the next three years.

The basic concept for implementing the SAM has been derived from a powerful model taken from education: the Master Teacher. Master teachers are individuals who have proven in their career that they are truly the best at their profession: that they really know how to teach. A master teacher program takes such individuals and uses them as role-models to help other teachers improve their service delivery. Persons designated as master teachers are also often given opportunities for additional training to further enhance their value as high-level trainers.

We stated earlier our belief that the PMOH has much of the expertise it needs within its own body of health professionals to improve service quality. These individuals can be identified by our assessment procedures and named as "Masters" in the same sense that Master Teachers are designated. The model we propose creates this cadre of Masters (physicians, nurses, nurse-midwives, and health auxiliaries) who function as teams dedicated to the improvement of service delivery at the health center level and who link in-service training programs with assessment.

For the PMOH, the current infrastructure could support such teams at the UDES level, each with 10-15 health centers as its responsibility.

The central event in this process model is the visit by the team to the health center, both to observe and to teach. We have called this the "Monitoring and Training Visit" or MTV plan. Instead of discussing MTV in the abstract, we have chosen to present it in the descriptions of three concrete efforts to use it in the PMOH:

> *MTV was pilot tested during the CDD Assessment carried out as part of the Cono Sur PRICOR Project, which is described in detail in Volume 3 of this final report.*
>
> *We are using MTV methodology to carry out the National Management Assessment of the PMOH, which is currently underway and funded by the USAID Mission in Peru as a buy-in to the Cono Sur PRICOR Project. This will be described next.*
>
> *MTV also forms the basis for one element of the triad comprising the health information side (HISPRO) of the new PMOH Health & Management Information System. The PRISM Group and its collaborators are developing and implementing this system for the Ministry under a new contract with USAID. The final section in this volume is a description of the theoretical underpinnings for HISPRO (taken from the contract itself). MTV and HISPRO are the next logical steps in the development of an integrated operations management information system for primary health care organizations.*

7. PROJECT: National management assessment of peripheral service delivery in the Peru Ministry of Health

a. Goals

The project has four specific aims:

> *... to carry out a management assessment, based on the PRISM PRICOR Systems Assessment Model, of the most important aspects of primary health care service delivery and coverage at the health center level in a sample of 8 UDES of the Peru Ministry of Health (PMOH) and in a sample of "policlinicos" of the Peruvian Institute of Social Security;*
>
> *... to link this assessment to immediate feedback and training at the operations level (health centers and UDES) using mini-workshops targeted on the weaknesses and strengths actually found in each unit's performance;*
>
> *... to produce a national database of baseline data on the performance and coverage of key peripheral services that program directors may use for more effective strategic planning and resource allocation; and*
>
> *... to serve as a pilot demonstration of a practice-based, one-year training program in Peripheral Health Services Management which could produce 40 or more certified health professionals annually to fill future needs for health center directors and program coordinators in the public sector.*

b. Advantages of and future prospects for the project

This project will obtain a statistically adequate sample of baseline operational data on performance in 8 of the PMOH's 28 UDES and in policlinicos in the three regions of Peru. This database will be of value to many aspects of planning and evaluation at the national level and may reveal enough consistency in patterns of performance that planning models can also be prepared at the regional or UDES level.

The data will link process indicators to outcomes in a large enough sample to identify factors related to good performance and will also provide in-depth knowledge of user and worker satisfaction. This will assist the PMOH in its strategic planning, resource allocation and targeting of specific training.

Another important expectation of the proposed project is that it will permit us to test key aspects of the model of quality control that we are developing at the level of the health center. A central goal of the PRICOR Project is to develop a model for simple, continuous quality control that is applicable to any health center in the system. A proven model of this type would be of great importance to the PMOH in its effort to decentralize services.

The project will also permit the PMOH to test its routine statistical data against a valid outcome and process sample.

The project will provide an intensive training of a cohort of 40 health professionals in the theoretical and practical aspects of sound peripheral services management through a highly participatory program. These individuals are a potential source from which to choose new health center directors.

The site visits also offer a one month practical training experience for 4 workers in each UDES in management and assessment techniques and an intensive training effort to health centers and UDES targeted to the performance areas in which they most need refresher training.

At relatively low cost, the PMOH and IPSS can implement the proposed project as an annual evaluation. Once the protocol has been worked out in the pilot project, a routine effort should be fairly easy to standardize.

The PHSM part of the project could be developed into a 1 year management training program based at a Peruvian university in collaboration with the PMOH. It is possible that close ties could be established with one or more international schools of management to develop a truly world-class degree in Peripheral Health Services Management. Such a program does not currently exist and should be an exciting prospect for one or more of the international agencies or private foundations to support. It may be possible to obtain an institutional development grant to cover the complete costs of the academic side of this program, leaving the PMOH the beneficiary of a large body of health professionals trained in management for the future.

c. Proposed program approach

1) Basic concept

The basic concept is to create a 40-member evaluation group comprising health professionals who have completed their obligation to the SERUM program. The evaluation group will be trained in the management assessment of health services using the Systems Analysis Model being developed by the PRISM PRICOR Project. The group will then be divided into four 10-member teams and each team will be sent out to two PMOH UDES to carry out a 1-month assessment of key peripheral services in each.

> *NOTE: It may be impossible for the assessment teams to cover those departments declared in emergency due to the terrorist threat. This is a difficult aspect that we are not competent to deal with in the context of the proposed project. If an assessment of service delivery can be done safely under certain conditions, this will need to be negotiated.*

Each UDES assessment will consist of an intensive effort to collect performance data from a statistically valid sample of health centers and communities and analyze it both at the health center and UDES levels. Special emphasis will be placed on the assessment of Child Survival Programs since these are the main focus of the current PRISM PRICOR Project and the U.S.A.I.D. health agenda.

Assessment visits at health centers and at the UDES will end with a series of training workshops based on the results of the assessment so that the PMOH personnel involved can get immediate feedback on the strengths and weaknesses uncovered in their service delivery.

Following an approximately 3 month period of site visits, the teams will be regrouped at the PMOH to work closely with program directors in order to make effective use of the information gathered in strategic planning and resource allocation.

2) Collaborative effort

The proposed project will be a collaborative effort between the PRISM PRICOR Project and the PMOH.

.) Human resources for assessment team

Assessment Team. The forty proposed team members will be recently graduated health professionals (20 physicians and 20 nurses or nurse-midwives) who have finished their service obligation to the PMOH (the SERUM program). These members will be chosen competitively from applicants from all of the professional schools in Peru.

The 40 participants will be considered as the class of a 4-month program in Peripheral Health Services Management (PHSM) which will be built around the assessment effort. This program will lead to a certification in PHSM that the PMOH may recognize in the future as a pre-requisite for persons to be appointed as health center directors or program coordinators.

Additional PMOH personnel during the site visits. During site visits, the assessment team will be augmented by 5 people from the UDES so that local PMOH personnel might have an opportunity to become familiar with some of the approaches being used to assess performance. In addition, these people will serve as a source of immediate knowledge about local conditions that the assessment team will need in order to complete their work.

In assessment visits to individual health centers, the assessment team will be further assisted by at least one person from that health center who will be assigned to this duty during the length of the visit.

Additional personnel to assist in the assessment will be assigned by the UDES or health center on an as-needed basis.

4) Training of team members

Formal training for the 40 PHSM trainees will be carried out during two months by an expert team created for the purpose from the existing PRISM PRICOR project group.

The PHSM training will be based on the PRISM System Assessment Model which has been developed and tested as part of the current PRICOR project. This will include a thorough introduction to organization theory and behavior as well as techniques of organization, performance, and effectiveness assessment. Techniques and protocols for assessing coverage by simple census, and cluster and lot quality assurance samples will also be included.

5) Framework of Analysis

The eight UDES selected to make up the national sample will be selected collaboratively between the PMOH and USAID. We anticipate that the final selection may reflect a regional distribution of: Coastal - 3; Sierra - 3; and Selva - 2.

The core effort of the 1-month assessment of each UDES will involve 1-week site visits to eight health centers (and their associated health posts). This sample will be treated statistically as a lot quality assurance sample of the UDES.

Four health centers will be selected at random from those within "easy" access of the UDES or support hospital and four will be selected from those in "isolated" areas. These categories need to be defined operationally and it may be necessary to limit the definition of "isolated" to centers within two day's (16 hours) travel or less from the UDES office. We will also include 2-4 policlinicos in each UDES to generate a sample of 8 policlinicos in each of the three regions being studied.

Health center assessments will be done using the Monitoring & Training Visit (MTV) model which stresses both assessment and in-service training as a unified approach to controlling the quality of primary health care services.

The assessment of health centers will involve:

An assessment of organizational structure based on an Organization Design/Function Questionnaire and Job/Unit Design Questionnaires.

Performance assessment of direct service delivery (care and education) and of support services (supervision, logistics, information system, training) using instruments developed and validated during the PRICOR project:

Simulation Exercises (short role-playing exercises designed to demonstrate the best performance a worker is capable of under conditions of direct observation)

Basic knowledge examinations

On-Site Observation Checklists

Performance Self-Reports

Assessment of community health status and user satisfaction based on accepted epidemiologic surveillance and market survey techniques.

The assessment will be limited in terms of the number of programs that are included in the complete analysis. With the resources available and the development of the PRISM SAM thus far, the following programs will be most feasible to include for performance assessment:

Control of Diarrhea / ORT
Immunizations
Acute respiratory infections
Growth monitoring and Nutrition
Family Planning
Maternal health

The statistical design for the assessment of each of the eight health centers in the UDES will include the following:

Organizational assessment questionnaires administered to all personnel of the health center

Performance assessments carried out for at least 8 health auxiliaries and all direct supervisors

On-site Observations made at the health center and at 4 associated health posts (or as many as possible if there are less than that number active)

Community surveys on health status and utilization of PMOH services carried out in up to 15 families nearest the health center and up to 15 at distant locations: up to 30 in all

The statistical design in will produce the following numbers at the UDES and PMOH levels:

	UDES	REGION	NATIONAL
No. of health centers	8	24	64
No. of health posts	32	96	256
No. of auxiliaries	64	192	512
No. of families	240	720	1,920
IPSS-policlinicos	2-4	8	24

6) Schedule for an UDES Assessment Visit: 6 weeks duration

Week 1: Initial data collection/Orientation of team members from UDES

Data collection

Geographic siting
Demographic data
Coverage: Record review - monthly reports
Selection of 8 health centers and policlinicos

Orientation/training of UDES team members: the five persons assigned from the UDES will receive a one-week introduction to the procedures being used to carry out the assessment and the principles underlying the approach.

One policlinico assessment

Weeks 2-5: Health Center visits

The assessment team will divide into two mini-teams consisting of 2-3 physicians and 2-3 nurses/nurse-midwives from the PHSM group plus one person from the UDES.

Each mini-team will visit two health centers for 1-week each. This will require 2 mini-teams to complete the 8 health centers in 4 weeks.

Each health center assessment will consist of approximately 5 days of data collection at the center, ancillary posts, and the community by the mini-team and its health center aides. This data collection will follow the framework of analysis specified above.

Immediate feedback will be an important aspect of a health center evaluation. All data collection instruments will be designed so that critical scores can be tabulated quickly at the time they are completed. Feedback will include:

a review of the performance of each health worker whose work was assessed - done with that health worker immediately after he/she has completed the performance review process (role-playing exercise)

a meeting with health workers and supervisors in the last days of the visit to present findings concerning areas of weakness or strength in service delivery performance, coverage, or user satisfaction in the community; this is to be based on the preliminary tabulation of data from these sources.

a meeting with the health center management/ supervisory staff to:

discuss health center management on the basis of concrete examples developed during the assessment; and

show them how to carry out an in-center training program targeted to their weaker areas of service delivery - using the performance checklists and manuals developed by the PRISM PRICOR Project as training tools to guide practical exercises

Week 6: Data analysis and report production at UDES

Half of the team will carry out an assessment of the second IPSS policlinico if one exists.

Data entry will be carried out using a portable microcomputer (one sent with each team) with reports produced on an accompanying portable printer. Data entry and analysis will be done using CADRE (Borland, Inc.). Copies of all reports will be distributed to each of the health centers in the UDES (including those not assessed) and to all key persons at the UDES.

Prior to leaving the UDES, the team will carry out a 1-day workshop for the UDES director and health center directors/supervisors in which data from the UDES as a whole are used to pinpoint common weaknesses or strengths in service delivery, unmet needs in the community, levels of user satisfaction with PMOH services, etc. These themes will be supported by concrete examples taken from the assessment just completed and the team will lead discussions aimed at eliciting appropriate management responses.

d. Work of assessment team after site visits have been completed

Upon returning from the field portion of the course, PHSM participants will participate in a 2-week workshop to prepare reports on their activities. A limited number of the 40 (10 or less) will be invited to participate in a subsequent 3 month work-study program in analysis:

assisting in the preparation of a cleaned, accessible database of baseline data;

preparing individual UDES site reports to be sent to the UDES director and key PMOH personnel;

working with PMOH program directors in each of the programs involved in the assessment to prepare a policy working paper based on the information aggregated at the departmental, regional, and national level; and

serving as expert resources to individuals, committees, working groups, etc. within the PMOH who require current information on any of the aspects covered in the national evaluation.

8. Overview of the Health Information System (HIS) for the PMOH Health and Management Information System

The HIS model is based on the premise that the MOH will continue to move to develop primary health care services delivery within context of the system model for District Health Systems being promulgated by WHO/PAHO.

Major problems with the MOH's existing information system are those that plague the information systems of virtually every public sector health system in the Third World. The list could be expressed as follows:

Data Missing:

... on target populations (size)
... on outcomes and impacts
... on socio-environmental determinants
... on performance
... on non-govt health providers

Data Quality/Quantity:

... collected but not used
... late, incorrect, incomplete

Limited information processing skills

Data not used for local decision-making

In order to correct these problems, data must be collected in a consistent, trustworthy manner and processed via clear protocols to produce indicators about whose reliability, validity, and utility there is general agreement. To be functional, such an information system must serve users' perceived needs at various levels: those of peripheral health workers, direct supervisors, UDES management, and community health workers and/or committees if they exist.

Information should not be collected unless it has first been determined how it will be used specifically in decision-making and all work necessary to define the indicator(s) needed has been accomplished.

Given these fundamental premises, it is not only feasible but obligatory for management at all levels to use information and feedback as a change agent and tool for organizational development in bringing a fully functional, decentralized primary health care system into existence.

BASIC REQUIREMENTS

The goal is to develop and implement an efficient, decentralized HIS that uses microcomputers based at each of the 28 UDES and the central MOH to improve the speed and capacity of data collection, analysis and dissemination of information to all management levels in the MOH.

The requirements include:

(1) a unified statistical data logging system (e.g., forms, protocol, and codebook) for use at the service units (health posts, health centers, and hospitals) and laboratories of the MOH;

(2) a routine, active community surveillance system at the health center level that collects priority data on health status indicators, common health problems, program coverage, and consumer demand for services within each center's area of responsibility;

(3) a routine, active quality control system that monitors key indicator data on the quality of services actually delivered by units of the MOH (e.g., care - preventive and curative, education - individual and community, promotional efforts) and targets the need for corrective response;

(4) a distributed database management system based at the UDES level to collate and archive these primary datasets and produce intermediate, truncated datasets for the central level;

(5) a modular reporting system, based both in the UDES and central levels, capable of producing summary reports that can be flexibly tailored to meet planning and evaluation needs of program and unit managers, and containing software and protocols to support the direct design and modification of reports by these managers.

The implementation strategy for development of the system includes:

(1) training of key UDES and central MOH personnel in the operation of the HIS software package; in the protocols for data collection, processing and distribution; and in the modular reporting system;

(2) training of key UDES and local health personnel responsible for routine, active epidemiologic surveillance and service monitoring;

(3) technical assistance to key UDES and local health personal during the MOH effort to train health personnel at the health center level in routine active epidemiologic surveillance and service monitoring;

(4) on-going technical evaluation of data collection and reporting; in particular, validating statistics collected by the routine system via

independent data collection carried out in collaboration with the Field Epidemiology Training Program (FETP) in conjunction with the CDC resident advisor and FETP staff.

(5) training and technical assistance to the effort to create a permanent, interdisciplinary working group within the central Ministry, representing the interests of community and service personnel as well as program and unit management, and functioning to oversee the operation of the information system and coordinate changes after the current technical assistance contract has ended.

INDICATORS

From most perspectives, the main reason for implementing an information system is the output. For the HIS, this refers to the major indicators it is intended to produce.

Information should not be collected unless one has already defined the use(s) to which it will be put. The HIS, thus, begins with a discussion of the indicators it will produce in context of the functions each is intended to support.

Assessment of needs and impact

The first and most fundamental management function is the strategic assessment of the health needs in the population served and the outcome/impact of program activities carried out to date. The focus of this assessment will be individual families at the health center level; health center catchment areas or districts at the UDES level; and districts, UDES or regions at the national level.

In reality, needs assessment and impact assessment are two sides of the same activity, that of collecting information regarding the health status of the system's catchment population. Two statistics of critical importance as denominators used in the calculation of a great number of such health indicators are population estimates and live births.

Population estimates. In order to produce valid estimates for many indicators, it is necessary first to have accurate estimates for the population of the unit being analyzed stratified by sex and by key age groups. The unit in this context can be a health center area of responsibility, district, UDES, region or the country. Ideally, population estimates will be available for a number of prior years as well as for the current year.

Live births. This important statistic serves as both a denominator and a numerator in the most fundamentally important indicators used in assessing primary health care needs and impact.

Assuming that these statistics are available, it becomes feasible to calculate a variety of indicators.

> *Death and birth rates. The availability of reliable mortality data, undifferentiated for cause, for a unit of analysis permits the calculation of the crude death rate and age-specific death rates based on average population; and infant mortality rate and neonatal mortality rate based on the number of live births in the same year.*
>
> *If any causes can be assigned to deaths with some measure of confidence, then cause-specific death rates may reasonably be calculated based on average population. For the CSA Project, mortality rates associated with diarrhea, malnutrition, measles, and acute respiratory infections are a particular priority.*
>
> *In a similar fashion, maternal death ratios can be calculated if the number of deaths from puerperal causes is known; this is based on number of live births in the same year.*
>
> *If the number of fetal deaths during year can be estimated, it becomes possible to calculate indicators such as fetal death rate, fetal death ratio, and perinatal mortality rate.*
>
> *These mortality rates plus the crude birth rate comprise the major public health indices commonly in use.*

In addition to these mortality-based indices, a significant array of quality of life indicators have been developed covering quality of life issues relevant to public health. The most important category of these are the morbidity-associated indicators of health status.

> *Morbidity. The most important group of quality-of-life indicators falling directly under the category of health status is morbidity. This will be expressed mainly as incidence of chronic and long-term illness and as prevalence of acute illness.*
>
> *For children targeted by the CSA Project, current priority morbidity indicators are: incidence of polio; prevalence of diphtheria, pertussis, tetanus, or measles; prevalence of diarrhea; incidence of hospitalization for diarrhea-associated dehydration; prevalence of acute respiratory infections; incidence of acute 2nd and 3rd degree malnutrition (wt/age); incidence of chronic malnutrition (ht/age); incidence of complicated pregnancies and deliveries; incidence of low birthweight babies; number of pregnancies with intervals of less than 2 years;*

Additional morbidity indicators for children (<5 years and school-age) frequently mentioned are: incidence of injury; incidence of malaria and tuberculosis;

The primary indicators of health status are joined by a number of health-relevant indicators reflecting other aspects of quality-of-life.

Socioeconomic status. This includes the following factors:

Education -- literacy rate for men and women by key age group, proportion of children 12-14 who are literate, proportion of children actively attending school -- by age

Employment and Income -- proportion of households living below poverty level; proportion of households with no permanently employed full-time worker

Lifestyles -- rate of smoking, rate of drug abuse, rate of alcohol abuse

Women's Status -- number/proportion of female householder households

Environmental factors. Factors of interest include proportion/number of households: in substandard housing, with access to clean drinking water, with potable water systems, with adequate latrines, with sewage facilities, and with clean environment free of garbage and human feces.

This set of indicators for needs and impact assessment covers a large proportion of those aspects of major interest to a primary health care program and of the CSA Project, in particular. The next step in using them will be to provide each with an operational definition and decide on the population characteristics by which data will be stratified.

Assessment of Service Coverage and Utilization

The health services of highest priority for the MOH are: immunization; care of childhood diseases such as respiratory infections, diarrhea (with oral rehydration therapy), and malaria; family planning; growth monitoring and nutrition; maternal health; and control of tuberculosis.

The primary health care program of the MOH includes provision of curative care, preventive care, and promotion/education in the community. MOH management at the health center, UDES and central MOH levels need to know the extent to which these services are actually being provided and then whether they are utilized effectively and by what proportion of the population.

The following indicators are those that are being most commonly used or recommended for monitoring and assessing coverage by each of the key programs:

Diarrhea program

proportion of children (< 5 years) with diarrhea in past two weeks ...
- *... treated with ORS*
- *... whose mothers know of ORS*
- *... whose mothers know how to administer ORS*
- *... receiving ORT of any kind*
- *... given appropriate dietary management*
- *... treated in public health facility*

proportion of mothers of children (<5 years) ...
- *... who know of ORS*
- *... who know how to administer ORS*
- *... who know about hygienic habits that prevent diarrhea*
- *... who teach hygienic habits to their children*
- *... who can easily acquire ORS*
- *... who remember mass media messages about diarrhea/ORT*

number of oral rehydration units per unit of population

number full-time equivalent diarrhea program health workers per unit of population

number and proportion of households visited by health worker in past year

proportion of public hospitals using ...
- *... rooming-in*
- *... promotion of breastfeeding*

number of rural water systems constructed

number of rural latrines constructed

Immunization program

proportion of children with complete or partial vaccination coverage for dpt, polio, measles, bcg ... by age group

proportion of mothers who remember mass media campaigns about immunizations

proportion of mothers of newborns in past year who received tetanus toxoid

number and proportion of households visited by health worker in past year

proportion of women in fertile age (in high risk areas) who are immunized against tetanus

rate/number of doses of tetanus toxoid given to pregnant women

Growth/Nutrition program

proportion/number of children enrolled in growth monitoring

proportion of children with up-to-date growth charts

proportion of children weighed during past year -- by number of times and age group

proportion/number receiving suppl. food among ...
- *... children ≤ 5 years*
- *... school-age children*
- *... pregnant or lactating mothers*

proportion of mothers of newborns who breastfeed exclusively for first three months

number and proportion of households visited by health worker in past year

proportion of children of given age (1-month intervals) who receive any breastmilk

proportion of children of given age (1-month intervals) receiving food other than breastmilk -- by type of food

proportion of mothers of newborns in past year who know the benefits of breastfeeding

proportion of mothers of newborns in past year who remember mass media messages about breastfeeding

Family planning program

proportion of women in fertile age ...
- *... using FP -- by type, by age, by status*

... who have ever used a contraception method -- by type
... who know of FP methods and where to obtain them
... who report hearing media campaigns for FP

proportion/number of women in fertile age using MOH family planning services -- by type, by age, by status

proportion of men who know of FP methods and where to obtain them

total fertility rate

mean age of last child

mean length of last birth interval

number and proportion of households visited by health worker in past year

proportion of mothers who remember mass media messages about family planning

Acute respiratory illness program

proportion/number of cases of ari in past two weeks ...
... attended in health service facilities
... whose mothers know the alarms signs for ari
... whose mothers recognize early ari
... whose mothers know how to treat early ari at home
... whose mothers know home treatment for mild ari
... whose mothers correctly treated early ari at home
... whose mothers used adequate treatment for mild ari

proportion of mothers who ...
... know the alarm signs for ari
... recognize early ari
... know how to treat early ari at home
... know home treatment for mild ari
... remember mass media messages about ari

Maternal health program

proportion of mothers of newborns in past year who received ...
... prenatal care visits at least 4 times before birth
... services assuring a safe delivery
... postpartum care at least once within 6 weeks after birth

number of centralized cervical cancer detection centers per unit of population

Assessment of Service Quality

Examples of health services programs or interventions that have not produced the desired impact are depressingly common. There are a number of reasons why an intervention that was demonstrably efficacious in its development might appear to be ineffective in real-world practice.

The first is that there is a failure of theory: that the intervention does not work in the way it was imagined to in its pilot studies. Arguments over program failures heard in international conferences often seem based on this assumption but there is very little evidence in the literature for failure of theory as the most common cause of failure "in fact".

The second reason for the appearance of failure is that the indicators used to measure the impact are too insensitive or wrongly directed to measure what impact there is. This may be a result of misdirected theory or too much noise from other factors affecting the impact being measured.

The third possible reason for apparent failure is that the intervention was simply not carried out as designed: that the quality and quantity of service provided was not adequate to produce the desired outcomes. This is the most likely cause of most of the program failures that have actually been studied and suggests that there is a need for monitoring quality of services that has been, perhaps, underappreciated in the past.

Monitoring quality of service effort along with quantity (coverage) will provide MOH management at all levels with information it needs to exert the most effective and efficient control over primary health care services delivered by the MOH. As part of the local management strategy of the UDES, such monitoring needs to include support services and structural characteristics of peripheral health facilities as well as direct services provided to the user population.

The PRISM Group, as a result of its PRICOR II Cono Sur project, has developed, a detailed and organized set of indicators of structural characteristics and performance quality that can be applied to any of the primary health care programs of the MOH. These will provide a complete assessment of the individual health worker's knowledge, ability, and typical performance in a given program.

Resource Allocation and Utilization

The most common management functions at both the peripheral service facility and UDES level are those relating to resource allocation and the monitoring of their use. This is also the key activity through which strategic decisions (based on assessments of need, impact, coverage, and service quality) are converted into effective and efficient operations management.

There are four major groups of resources that these managers have under their control and each has a number of relatively clear indicators.

Personnel.

personnel deployment by category of worker, area and population size, and projected workload

duty assignments based on individual performance, leave schedules, cyclic activity

training planned based on KAP of individual workers

Drugs/supplies.

level of essential drug use by population size and frequency of diagnosis (drug use by diagnosis is indicator of quality of prescribing practice)

need to shift supplies (active or by attrition) based on over- and under-supply for demand

bulk supply orders based on realistic projections of overall use patterns

Facilities/equipment/transportation.

assign vehicles to meet priority needs for transportation

check transport usage to ensure accountability (mileage, fuel consumption, use-to-purpose) ... need to establish benchmarks

repair and maintenance protocols for each major piece of equipment or type of equipment

check inventory routinely to ensure accountability

Finances.

budget for fuel for vehicles, ravel/daily allowances, office supplies, building and vehicle maintenance ... as function of output

The above lists of indicators are ultimately meant to be complete and parsimonious. At present, there are gaps and many indicators that need to be fully operationalized. This would be the responsibility of the working groups formed under the project. Nevertheless, the existence of this detailed set of indicators will facilitate the ultimate development of standardized indicators for the MOH information system.

Given a set of indicators as output from the HIS, the other important aspect is the HIS process model: how data concerning population and service system characteristics are collected and processed to produce the indicators just mentioned. This is the next step in the description of the model to be implemented.

THE GENERAL MODEL

The general HIS model, which is provided under The PRISM Group's brand name of "PRIMIS Information Systems", contains three distinct data collection subsystems based on origin of data:

transactions (routine service encounter statistics),

community surveillance (active epidemiologic surveillance in the community),

and quality control (monitoring).

These subsystems are shown in the figure on the following page, which presents an overview of the HIS model.

Transactions

The transactions subsystem utilizes data that originate from service encounters between health workers and community members who utilize the MOH. It is immaterial whether these encounters occur in an outpatient clinic, health center, or during a home visit; or whether they involve a health auxiliary, nurse, or physician.

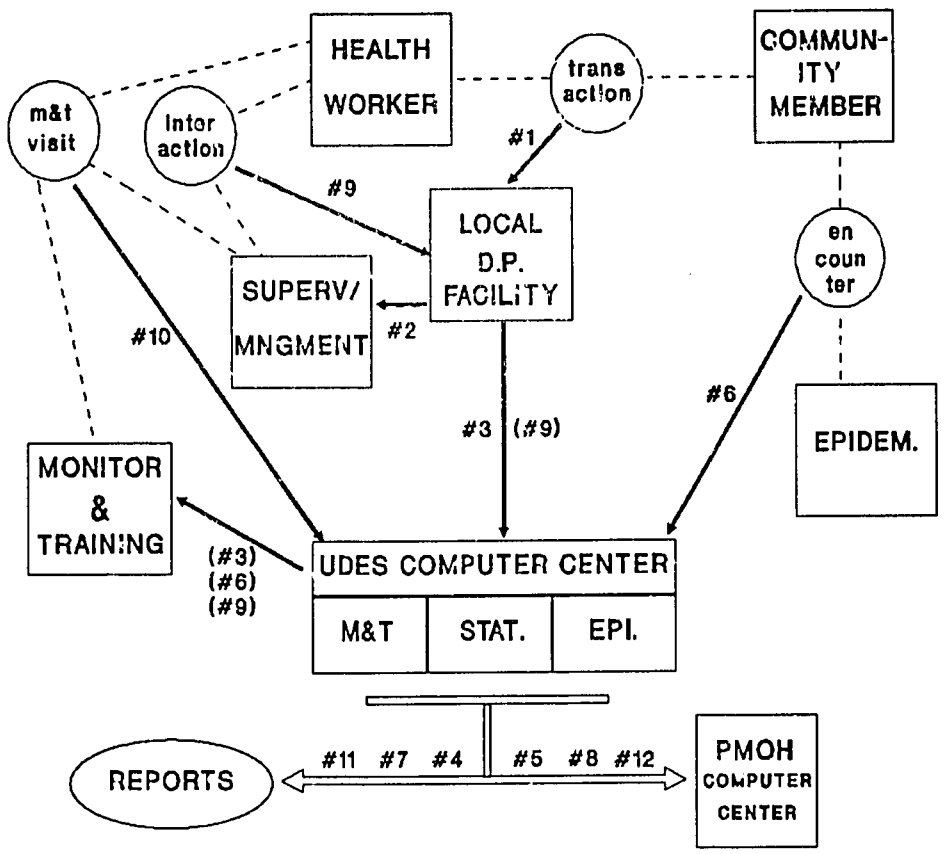

Figure 2. Basic design for PMOH Health Information System

(For simplicity in the following discussion, local facility involved is referred to as "the health center" and the individual as "the health worker", keeping in mind that no such limitation is intended.)

These routine encounter data are sent to the health center's data processing facility #1 where they are tabulated to provide "feedback" (process control data)#2 to the health worker's immediate supervisor and truncated to send#3 to the UDES Computer Center.

Data arriving at the UDES Computer Center are processed to produce reports#4 to send to appropriate recipients (e.g., UDES management, MOH program heads, health center directors) and truncated datasets#5 to send to the central MOH Computer Center, where they will be incorporated into a national database and subjected to further processing.

(To keep the figure uncluttered, report recipients are shown as a single block and the MOH as a final destination only.)

Community surveillance

The community surveillance subsystem is focused on indicators of needs and impact, and is carried out under the direction of the UDES Office of Epidemiology. Epidemiologic data (morbidity, mortality, KAP, etc.) collected directly from community members by one of several possible methods are sent#6 to the UDES Computer Center. There they are processed to produce reports#7 which are sent to appropriate recipients and datasets which are sent to the central MOH Computer Center#8.

Monitoring

The monitoring subsystem focuses on indicators of the quality and quantity (coverage) of service delivery in order to support a system of quality control. The term "monitoring" is used in a management-oriented sense consistent with the concepts of monitoring and evaluation developed by Casley and Kumar (D.J. Casley and K. Kumar, In Project Monitoring and Evaluation in Agriculture, JHU Press, 1987). To quote Murphy and Marchant (J. Murphy and T.J. Marchant, Monitoring and Evaluation in Extension Agencies, World Bank Tech Paper No. 79, 1988):

"...monitoring is defined as an internal function, an integral part of good management which is required whether outside funding is involved or not. Management-oriented monitoring is a continuous, analytical process through which the agency director and technical managers receive frequent updates on three key questions:

Are development activities (infrastructure, services, training, etc.) being implemented as planned, on schedule and within budget?

Are these activities leading to expected results (utilization of infrastructure and services by beneficiaries, increased capabilities through training, changes in behavior patters, etc.?

What is causing delays or unexpected results? Is there anything happening which would lead management to revise/modify the original plans?"

Experience with the PRICOR study in the Cono Sur suggests that monitoring should be closely linked to training as the most effective way to manage quality. In the model, therefore, both activities are placed within a single, hypothetical UDES office called the "Office for Monitoring and Training". This terminology parallels that currently used by the World Bank in talking about their system for agricultural extension, the Training and Visit Extension system (D. Benoir and M. Baxter, Training and Visit Extension, The World Bank, 1984).

Monitoring in this model comprises two distinct data collection efforts. The first, a local management activity, is the result of routine supervisory encounters between the immediate supervisor and the health worker and is based on simple checklists developed for each program to be controlled. The sets of checklists include user guides on their application and interpretation. The results of these encounters are sent to the UDES Computer Center on a monthly basis#9.

The second effort is the Monitoring and Training Visit (MTV) which is carried out cyclically, quarterly for example, by the UDES M&T Office at each health center. The monitoring will include a review of information sent via routes #3, #6 and #9 since the last MTV, and an on-site collection of more detailed and far-ranging indicators of performance at the health center.

The M&T Team will use a set of standardized questionnaires, checklists and interviews to assess basic knowledge, observe capability to perform required service tasks, examine supervisory awareness, gather perceptions by the workers themselves of performance quality, and rate community outcomes. The model is flexible with regard to outcome data, however, and these could be collected in collaboration with community surveillance instead.

A key characteristic of MTV is that pronounced strengths and weaknesses identified during these visits will be included immediately in targeted training done while the M&T Team is still on-site.

The data generated by MTV will be sent to the UDES Computer Center#10 to produce reports which are sent to appropriate recipients#11 and datasets which are sent to the central MOH Computer Center#12.

The UDES Computer Center is, thus, a dynamic center of information processing in this model. It receives data flow from the three subsystems and is responsible for storing these data in accessible

This general model is compatible with a variety of different approaches to data collection and handling that might be ultimately selected by the MOH.

OPERATIONAL APPROACHES TO THE HIS

Transactions

Routine service encounters are currently recorded within the MOH using over 150 forms in a hierarchy of clinical histories, daily registers, monthly and 3-monthly summaries, and forms for special needs -- with new forms produced independently and ad hoc by each central program office as well as by certain UDES offices.

These are supposed to be processed according to formulas developed by the issuing authority to produce the indicators needed to monitor health problems, including reportable transmissible infections, and to evaluate service delivery and coverage by individual programs, professional and staff effort, etc.

The MOH has long recognized that this is an untenable situation and has passed through several attempts to revise and reduce the paper-workload via forms consolidation. Such a major effort was made in 1984-86 with the technical assistance of Westinghouse's Health Systems Group under a USAID Contract.

A design effort was mounted as part of the discussions on IPSS/MOH integration in 1986 to produce a simple, unified daily register that would serve both institutions. This has not been implemented. Finally, the Vice-Minister's office this year sponsored a working group that produced modified designs for the Ministry's basic reports forms. Implementation of these modifications has been suspended pending the initiation of the USAID HIS/MIS project.

Data collection should have a consistent, unified methodology for all service providers that changes little over time and can be easily learned. The HIS reports generator needs to be able to produce routine reports speedily, to modify existing report structures with little cost or effort as program needs change, and to handle demands for special reports in a timely manner: all without requiring changes in the data collection forms and protocols.

The routine statistics dataset should be "log-based" not "tabulation-based. The desired modular structure can be achieved for routine statistics if data collection in the new HIS is based on workers "logging" their daily activities in a standardized manner for subsequent tabulation by microcomputers.

A preliminary version of such a log-based system was developed in a pilot study done by the MOH in Lima's Cono Sur in 1987. That version was one of several pilot implementations of key parts of the HIS which PRISM has been able to do in collaboration with personnel from the MOH.

The development of the primary data collection instrument, or daily register of encounters, should take into consideration the integrated form developed by the informatics team of the MOH/IPSS integration working group. This form currently contains basic patient identifier data, age, sex, residence, diagnoses and treatment. The latter two are to be given a coding structure along the lines developed in the Cono Sur pilot project.

Codes will be based on the International Classification of Diseases for diagnostic and clinical treatment codes and on consensus criteria to be developed within the MOH for treatment and activity codes falling in the areas of preventive services and promotion/education.

With this unified format, the primary source of raw statistical data for all primary health care activities will be a standardized, single line-record of encounters between MOH service providers (physicians, nurses, auxiliaries, etc.) and system users at peripheral service facilities (health centers, posts and outpatient clinics) and in the community.

Certain health services of less immediate priority and less need to be reported as encounter data may be tabulated in a summary encounter format. For example, the over 5 million dental services, of which almost half were simple extractions, make up almost 1/6th of the total encounters. It makes sense for simplified summaries by age group and/or sex to be used for these data rather than add to the data burden by reporting them separately.

This approach also makes it worthwhile to produce a standard health center-based family coding scheme, much like that used in almost all health centers already. Maintaining family clinical history records as a single pack makes a lot of sense. The computerized data management system can help assure that each family in a health center's record system has a unique identifier.

With such an ID, virtually all family members can be uniquely identified as individuals by the combination of it and the age and sex data which is included in each visit record. A special identifier may need to be introduced in the case of same-sex twins or siblings born within the same year. The database manager at the UDES level would back-calculate each age to a birthyear which would be compared for best fit with a directory of existing members of the family whose code was entered. The matching algorithm can be tested against actual record checks during the first year of operation to fine-tune it to maximum accuracy.

This system will not, of course, track people between health centers, nor is it intended to do so. It does give the information system access to the linked

records of a single person visiting a given health center over time. This would be a major advance in the MOH's ability.

These encounter records form the raw dataset from which a limited set of indicators can be extracted manually by the data processing personnel at the health center level. The original dataset, or a truncated version, will be sent as paper records to the UDES Computer Center for entry. It will be desirable to institute a weekly shipment of data to ensure a smooth flow and keep data processing current at both the health center and UDES level.

Process control of key primary health care services at the health center level can be achieved with a kit consisting of clear graphs and charts of progress/effort and a manual on how to produce and maintain them. The goal is to give the statistics personnel at the health center level an immediate return on his/her data processing effort in the form of a continuously updated, easy-to-understand profile of how well the unit is doing in meeting its targets.

An alternative to keyboard data entry via the use of optical scanners could speed data entry by a factor of as much as 5-fold. The PRISM Group has developed an OCR software package (HandEntry) capable of interpreting hand-printed letters that could serve as the basis for such a system. This will be pilot tested as part of the HIS development process.

The database management system (DBMS) under which routine HIS datasets will be processed will be fully relational and programmed in C for maximum speed and portability, and inaccessible for primary manipulation (e.g., altering existing data from "outside" the user interface as is possible with dBase III).

The HIS software to be provided at this level will include a full set of utilities for automatically generating a core series of reports and truncated datasets to be sent to appropriate individuals and offices. This output will be designed in collaboration with Working Groups in the MOH in the first year of the project.

The set of indicators detailed earlier will serve as a draft source for this effort. By the time this stage is reached, managers will be able to "mix and match" from among a set of standardized indicators to create reports tailored to their needs but which introduce no change whatsoever to the data collection or database process. This is the ultimate modularity necessary to maintain a decentralized system of this complexity.

Sophisticated users will be able to use an SQL-based relational query (i.e., a query language that closely resembles English syntax) to generate a wide variety of ad hoc reports as needed. A course in SQL will be included in the specialty courses to be produced during the latter half of the project.

A second, simpler but completely flexible ad hoc user interface to this and the other HIS databases will be provided via the Q&A database package and the Quattro spreadsheet. The DBMS will permit the export from any file, or set of

files within it, an intermediate dataset consisting of selected fields chosen from a menu. Individual records will be screened and included/excluded based on criteria set by the user.

Once exported to a flat-file format in ASCII, this dataset can be imported to Q&A to generate reports integrated with wordprocessing text or to Quattro to produce a wide variety of graphics. Further, either package can be used to manipulate the data and carry out "if-then" types of analysis and simple statistical tests that involve transformations of the dataset. Since this is an isolated dataset, it can be manipulated as desired without danger of damaging the integrity of the original database.

Community Surveillance

Community surveillance includes active epidemiologic surveillance and, perhaps, some aspects of market surveillance, such as consumer preference for and satisfaction with services provided by the local MOH service unit.

As indicated above, routine active epidemiologic surveillance will focus on health status indicators (birth rate, mortality - by age group, life expectancy, fertility rate, and population increase), on the frequency of the most common health problems in the community being served, and on the actual coverage by individual programs based on community sampling.

Data collection can be handled by a variety of different mechanisms (sentinel sites, sample-based surveys, collection of existing data such as birth registries or census data, routine household visits at all health centers). The ultimate design for this feature of the HIS, both collection and reporting, will be produced by the newly created Technical Directorate of Epidemiology in the MOH with the collaboration of the FETP advisors and PHC experts from leading universities.

The HIS team will participate actively in this process to ensure maximum coordination in the development of this subsystem. The indicator set being standardized will be of value to this process and serve as an important bridge between the routine transaction statistics and epidemiologic surveillance.

In this context, attempts to link laboratory data coming to the UDES computer center with patient encounter records arriving separately from the health centers should be avoided. There will be no standardized coding system in place during the implementation of the HIS or, probably, for long after. Non-matches in such a system will create an intolerable workload on the UDES Computer Center staff for no benefit.

Laboratory data can, of course, be analyzed separately as non-linked data and compared to presumptive diagnoses or other indicators generated from the routine encounter records. The primary linkage of laboratory and encounter data should occur, as it now does, by returning results to the initiating site and

entering a confirmed, or rejected, diagnosis for the patient in question. If a reasonable patient identifier system has been put in place as suggested above, the confirmed diagnosis can be easily linked to earlier encounter data via this mechanism.

With respect to the basic data collection process, many public sector health experts favor the suggestion that the MOH move to implement routine household visits by paid health auxiliaries as part of its development of a decentralized primary health care system.

Household visits on a 3-4 times per year basis to all members of the catchment community appears to be a feasible prospect. It means giving one health auxiliary the responsibility for approximately 1000 households, thus creating a demand for, very roughly, 5,000 workers, or about 7% of the current MOH workforce. This would be a truly radical innovation and politically difficult to implement, but it appears to be a proposal that the MOH is going to consider seriously in the coming year.
This would create a substantial routine dataflow that must be taken into account by the planned system for data entry so that it doesn't become a bottleneck. Otherwise, this approach would not differ greatly from any of the others in its demands on the DBMS.

The output from community surveillance will be produced in the same manner as routine statistics, using a similar set of DBMS utilities for routine reports and truncated databases, and the SQL and Q&A/Quattro interfaces for flexible access to the data.

No statistical package was called for in the RFP and the limited statistical capabilities of Quattro will serve for a significant amount of simple modelling that is likely to be done by the actual UDES epidemiology personnel. In addition, it is likely that UDES will have, though the FETP program, a basic statistical package such as Epistat. The ASCII export capacity of the DBMS will permit UDES personnel to generate intermediate flat files in ASCII format that can be imported by Epistat or virtually all other statistical packages.

Monitoring

Routine quality control (QC) monitoring is a relatively new addition to the management tools suggested for public sector health service delivery systems in developing countries but has a long history in private sector enterprises and in the health sector of most industrialized nations. Forms, formulas and protocols designed for QC monitoring in the MOH are currently under development and testing in Lima in an on-going USAID-supported project and will be available for implementation in mid-1989. This will be the basis for the MTV system that was described above.

The MTV approach has been incorporated in a pilot version as part of a planned National Evaluation of Peripheral Services beginning in June, 1989. This evaluation, to be done by The PRISM Group in collaboration with the MOH with funding from USAID on a buy-in to the PRICOR project, will collect performance and basic outcome indicators for all CSA Programs in 64 health centers of the MOH and 24 "policlinicos" of the IPSS during the latter half of this year.

A major feature of the MTV is the use of simulation exercises as a combination monitoring and training tool. This is a new concept in the area of PHC monitoring and evaluation. The indicators and general instruments that have been developed for the MTV have been mentioned.

One factor that may not be apparent is the efficiency with which MTV can be carried out. Experience with the assessment of the MOH 1988 immunization program and the diarrhea program in 1989 in 14 health centers suggests that a team of four (as projected for the national evaluation) can complete performance assessment of a complete health center in just a bit over a 1 1/2 days.

The HIS supposes that the MTV team will make quarterly one-week visits to a given health center. This will allow more than enough time for the monitoring of each key program at least twice a year with targeted training carried out on-site during the same visits.

From the relatively complex instruments developed for the MTV, it is possible to extract a simplified set of indicators for routine use by the immediate supervisors in each program to monitor activities of their personnel on a weekly basis or oftener.

The quality control system calls for routine monitoring data to be sent from health centers to the UDES computer center on a monthly basis and for results of MTV to arrive on a quarterly basis. The data entry load from this is not particularly heavy and can be done by personnel from the M&T Office.

The reporting structure is to a great extent predicated on the instruments that have been developed for MTV. The final format will be developed in collaboration with MOH management personnel for whom much of this intervention has been developed.

Implementation of the monitoring system can be done without a mass training effort. It involves only training a core M&T staff at the UDES and the program directors and/or supervisors from the health centers.

BIBLIOGRAPHY

Benoir, D. and M. Baxter. 1984. Training and visit extension. The World Bank, Washington, D.C.

Berrien, F.K. 1983. A general systems approach to organizations. In Dunnette, M.D. (ed.). Handbook of industrial and organizational psychology. Wiley-Interscience, John Wiley & Sons, New York.

Campbell, J.P. 1977. On the nature of organizational effectiveness. In Goodman, P.S., J. M. Pennings et al. (eds.). New perspectives on organizational effectiveness. Jossey-Bass, San Francisco.

Casley, D.J. and K. Kumar. 1987. Project monitoring and evaluation in agriculture. Johns Hopkins U. Press, Baltimore MD

Cronbach, L.J. 1984. Essentials of psychological testing. 4th ed. Harper & Row, New York.

Donabedian, A. 1966. Evaluating the quality of medical care. Milbank Memorial Fund Quarterly 44: 166-203.

Donabedian, A. 1980. Exploration in quality assessment and monitoring. Volume 1: The Definition of quality and approaches to its assessment. Health Administration Press, Ann Arbor, MI.

Donabedian, A. 1980. Exploration in quality assessment and monitoring. Volume 2: The criteria and standards of quality. Health Administration Press, Ann Arbor, MI.

Donabedian, A. 1980. Exploration in quality assessment and monitoring. Volume 3: The methods and findings of quality assessment and monitoring - An illustrated analysis. Health Administration Press, Ann Arbor, MI.

Gerbert, B. and W.A. Hargreaves. 1986. Measuring physician behavior. Medical Care 24:838-847.

Guilford, J.P. 1954. Psychometric methods. 2nd ed. McGraw-Hill, New York.

Hall, J.A., D.L. Roter, and N.R. Katz. 1987. Task versus socioemotional behaviors in physicians. Medical care 25: 399-412.

Hopkins, K.D. and J.C. Stanley. 1981. Educational and Psychological Measurement and Evaluation. Prentice-Hall, Englewood Cliffs, NJ.

Kline, P. 1986. A handbook of test construction: Introduction to psychometric design. Methuen & Co, London and New York.

Lazarsfeld, P.F. and H. Menzel. 1969. On the relation between individual and collective properties. In Etzioni, A. (ed.). A sociological reader in complex organizations. Holt, Rinehart and Winston, New York.

Lissitz, R.W. and S.B. Green. 1975. Effect of the number of scale points on reliability: A Monte Carlo approach. J. Appl. Psychol. 60(1): 10-13.

Murphy, J. and T.J. Marchant. 1988. Monitoring and evaluation in extension agencies. World Bank Technical Paper No. 79. Washington, D.C.

Nunnally, J.C. 1978. Psychometric theory. McGraw-Hill, New York.

Payne, R. and D.S. Pugh. 1983. Organizational structure and climate. In Dunnette, M.D. (ed.). Handbook of industrial and organizational psychology. Wiley-Interscience, John Wiley & Sons, New York.

Roter, D.L. and J.A. Hall. 1987. Physicians' interviewing styles and medical information obtained from patients. J. Gen. Int. Med. 2: 325-329.

Roter, D.L., J.A. Hall and N.R. Katz. 1987. Relations between physicians' behaviors and analogue patients's satisfaction, recall, and impressions. Medical Care 25:437-451.

Shortell, S.M. and A.D. Kaluzny. 1988. Health Care Management: A text in organization theory and behavior. 2nd ed. John Wiley & Sons, New York.

Van de Ven, A.H. and D.L. Ferry. 1980. Measuring and assessing organizations. John Wiley & Sons, New York.

Werner, D. 1980. *Donde no hay doctor.* La Fundacion Hesperian, Palo Alto, CA.

Williamson, J.W. 1978. Assessing and improving health care outcomes. Ballinger, Cambridge, MA.

Zschock, D.K. (ed.). 1988. Health care in Peru: Resources and policy. Westview Press, Boulder, CO.