Compendium of Indicators for Evaluating Reproductive Health Programs

MEASURE *Evaluation* Manual Series, No. 6

Jane T. Bertrand
and
Gabriela Escudero

August 2002

Volume 1

The manual series is made possible by support from USAID under the terms of Cooperative Agreement HRN-A-00-97-00018-00. The opinions expressed are those of the authors, and do not necessarily reflect the views of USAID.

Other Titles in the Manual Series

NO. 1 Evaluando Proyectos de Prevención de VIH/SIDA: Un Manual con Enfoque en las Organizaciones No Gubernamentales. Bertrand, Jane T. and Solís, Manuel. July 2000.

NO. 2 Quick Investigation of Quality (QIQ): A User's Guide for Monitoring Quality of Care. MEASURE *Evaluation*. February 2001.

NO. 3 Sampling Manual for Facility Surveys for Population, Maternal Health, Child Health and STD

NO. 4 Measuring Maternal Mortality from a Census: Guidelines for Potential Users

NO. 5 A Trainer's Guide to the Fundamentals of Monitoring and Evaluation for Population, Health, and Nutrition Programs

Preface

This *Compendium* reflects the tremendous strides made by many individuals and organizations over the past decade in evaluating reproductive health programs in developing countries. Over a hundred colleagues have contributed to this volume in various ways: attending meetings, preparing text, reviewing drafts of specific sections, and answering requests for information. Appendix A presents a full list of the many contributors to this publication to whom we owe a debt of gratitude.

USAID provided the support for this *Compendium* under the MEASURE *Evaluation* Project. We wish to thank Norma Wilson and Sara Pacqué-Margolis, technical advisors to Measure *Evaluation*, for their assistance throughout this process. Ties Boerma, Director of Measure *Evaluation*, sustained this activity with enthusiastic support, sound technical advice (especially on STI/HIV/AIDS), and good humor.

The authors are especially grateful to a number of colleagues who drafted specific sections of text, listed below in the order in which their sections appear.

Women's Status	Sunita Kishor
Gender Boxes	Karin Ringheim Nancy Yinger Jennifer Wheeler
Policy	Karen Foreit
Management	Anne Young Packham Alison Ellis
Training	Alfredo Fort Sue Brechin Rick Sullivan
Commodities and Logistics	Dana Gelfeld Aronovich Steve Kinzett Tim Williams
Behavior Change Communication	Maria Elena Figueroa D. Lawrence Kincaid
Operations Research	M. Celeste Marin
Quality of Care: SPA, QIQ	Nancy Fronczak Tara Sullivan
Quality Assurance	Diana Silimperi Bruno Bouchet Stephane Legros
Integration of Services	Ian Askew
Safe Motherhood	Mandy Rose
Newborn Health	Mandy Rose
Breastfeeding	Nadra Franklin
Adolescent Reproductive Health Programs	Robert J. Magnani Kristen Nelson
Postabortion Care	Janie Benson
Violence Against Women	Suzanna Stout Banwell Mary Ellsberg Victoria Frye
Reproductive Health in Emergency Situations	Therese McGinn Susan Purdin
Performance Improvement	Alfredo Fort

The authors circulated early drafts of each section to persons with expertise in the subject area, and most sections underwent several rounds of revision. We then sent the final draft to a group of external reviewers, each of whom covered at least five of the different sections in the volume. We would like to express our thanks to this group of external reviewers: Delia Barcelona, Joseph Foumbi, Shireen Jeejoobhey, Cheikh Mbacke, Leo Morris, Elaine Murphy, Luis Rosero-Bixby, Judith Seltzer, Iqbal Shah, John Stanback, John Stover, and Victoria Ward.

Finally, we thank Susan Nicolosi and Marcia Wolters for typing large amounts of text for this volume; Jennifer Wheeler, Shannon Salentine, and Martha Silva for research assistance; Jennie Phillips for editing the final draft; and Brandon Howard, Melody Moore, and Wayne Hoover for final formatting and layout.

This C*ompendium* is made possible by the support from USAID under the terms of Cooperative Agreement HRN-A-00-97-00018-00. The opinions expressed are those of the authors, and do not necessarily reflect the views of USAID.

Table of Contents

Acronyms

ACC/ SCN	Administrative Committee on Coordination/ Subcommittee on Nutrition
AED	Academy for Educational Development
AIDS	Auto Immune Deficiency Syndrome
AMRO	Regional Office of the World Health Organization
ANC	Antenatal Care
AOL	America on Line
APH	Antepartum Hemorrhage
API	AIDS Program Effort Index
AR	Abortion Rate
ARH	Adolescent Reproductive Health
ARI	Acute Respiratory Infections
ASFR	Age Specific Fertility Rate
ATTI	Admission to Treatment Time Interval
BABIES	Birth Weight, Age at death, Boxes, Intervention, Evaluation System
BCC	Behavior Change Communication
BCG	Bacille Calmette-Guerin
BEMFAM	Sociedade Civil Bem-Estar Familiar no Brasil
BEOC	Basic Essential Obstetric Care
BFHI	Baby Friendly Hospitals Initiative
BMI	Body Mass Index
BSS	Behavioral Surveillance Survey
BUCEN	Bureau of Census
BWSMR	Birth Weight Specific Mortality Rate
CAFS	Center for African Family Studies
CAs	Cooperating Agencies
CBD	Community Based Distribution
CBR	Crude Birth Rate
CDC	Center for Disease Control and Prevention
CED	Chronic Energy Deficiency
CEOC	Comprehensive Essential Obstetric Care
CF	Complementary Feeding
CFR	Case Fatality Rate
CHANGE	Center for Health and Gender Equity
CHR	Committee on Human Research
COCs	Combined Oral Contraceptives
CPI	Client-Provider Interactions
CPR	Contraceptive Prevalence Rate
CPTs	Contraceptive Procurement Table
CRLP	Center for Reproductive Law and Policy
CSI	Care and Support Indicator
CSO	Civil Society Organization
CYP	Couple Years Protection
DFID	British Department for International Development
DHS	Demographic Health Survey
DTFR	Desired Total Fertility Rate
EBR	Exclusive Breastfeeding Rate
END	Early Neonatal Death
ENMR	Early Neonatal Mortality Rate
EOC	Essential Obstetric Care

EPI	Expanded Program on Immunization
FDA	Food and Drug Administration
FEFO	First-Expiry/ First-Out
FGC	Female Genital Cutting
FGM	Female Genital Mutilation
FHI	Family Health International
FHI BSS	Family Health International Behavioral Surveillance Survey
FP	Family Planning
FFP/RH	Family Planning/Reproductive Health
FPMD	Family Planning Management Development
GPA	Global Programme on AIDS
GPS	Geographic Positioning System
HARI	Helping Individuals Achieve their Reproductive Intentions
HCP	Health Care Provider
HIS	Health Information Systems
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
HRM	Human Resource Management
ICD	International Classification of Disease
ICPD	International Conference on Population and Development
ICRW	International Center for Research on Women
ICT	Information and Communication Technology
IDD	Iodine Deficiency Disorder
IDP	Internally Displaced Persons
IEC	Information Education Communication
IGSS	Guatemalan Institute of Social Security
INACG	International Nutritional Anemia Consultative Group
IOM/ NAS	Institute of Medicine/ National Academy of Sciences
IOS	International Organization for Standardization
IP	Internet Protocol
IPPF	International Planned Parenthood Federation
IPPF/WHR	International Planned Parenthood Federation / Western Hemisphere Region
IPV	Intimate Partner Violence
IRB	Internal Review Board
ISP	Internet Service Providers
IU	International Units
IUD	Intra-uterine Device
IUGR	Intrauterine Growth Retardation
IVACG	International Vitamin A Consultative Group
JHU	Johns Hopkins University
JHU/CCP	Johns Hopkins University Center for Communication Programs
JHPIEGO	Johns Hopkins Program for International Education in Gynecology and Obstetrics
JSI	John Snow Incorporated
KPC	Knowledge, Practices, and Coverage
LAM	Lactational Amenorrhea Method
LAMM	Latin American Maternal Mortality
LAR	LAM Acceptor Rate
LBW	Low Birth Weight
LIAT	Logistics Information Assessment Tool
LMIS	Logistics Management Information System

LNMR	Late Neonatal Mortality Rate
LSAT	Logistics System Assessment Tool
LUR	Lactational Amenorrhea Method Use Rate
M&E	Monitoring and Evaluation
MASFR	Marital Age-Specific Fertility Rate
MCH	Maternal Child Health
MEASURE	Monitoring and Evaluation to ASsess and Use REsults
MICS	Multiple Indicator Cluster Survey
MIS	Management Information System
MMR	Maternal Mortality Ratio
MNPI	Maternal Neonatal Program Index
MOH	Ministry of Health
MSH	Management Sciences for Health
MOST	Management and Organizational Sustainability Tool
MTCT	Mother to Child Transmission
MUAC	Mid Upper Arm Circumference
MVA	Manual Vacuum Aspiration
NCPD	National Council for Population and Development Kenya
NEHK	New Emergency Health Kit
NFP	Natural Family Planning
NGO	Non-Governmental Organization
NHANES	National Health and Nutrition Examination Survey
NHS	National Health Service
NIDI	Netherlands Interdisciplinary Demographic Institute
NMR	Neonatal Mortality Rate
NT	Neonatal Tetanus
NTMR	Neonatal Tetanus Mortality Rate
OB-GYN	Obstetrics and Gynecology
OCHA	UN Office for the Coordination of Humanitarian Affairs
OD	Organization Development
OR	Operations Research
ORC	Opinion Research Corporation
PAC	Postabortion Care
PAHO	Pan American Health Organization
PAI	Population Action International
PAPFAM	Pan Arab Family Health Surveys
PCS/JHU/CCP	Population Communication Services
PASCA	Programa Acción SIDA de Centro América
PES	Policy Environment Score
PIN	Prevention Indicators
PI	Performance Improvement
PIT	Presumptive Intermittent Treatment
PL 480	Public Law 480
PMP	Performance Monitoring Plans
PMR	Perinatal Mortality Rate
PNA	Performance Needs Assessment
PNC	Postnatal Care
PNMR	Perinatal Mortality Rate
PPC	Postpartum Care
PRB	Population Reference Bureau
PSI	Population Services International
PVO	Private Voluntary Organization

QA	Quality Assurance
QAP	Quality Assurance Project
QAP/URC	Quality Assurance Project University Research Corporation
QC	Quality of Care
QIQ	Quick Investigation of Quality
RAMOS	Reproductive Age Mortality Surveys
RBP	Retinol Binding Protein
RH	Reproductive Health
RHS	Reproductive Health Survey
RPR	Rapid Plasma Reagin
SA	Situational Analysis
SAM	Service Availability Module
SDP	Service Delivery Point
SEICUS	Sexual Information and Education Council of the United States
SPA	Service Provision Assessment
SRH	Sexual-Reproductive Health
STI	Sexually Transmitted Infections
SV	Sexual Violence
TA	Technical Assistance
TAR	Total Abortion Rate
TB	Tuberculosis
TBA	Traditional Birth Attendants
TFR	Total Fertility Rate
TIS	Training Information System
TMFR	Total Marital Fertility Rate
TT	Tetanus Toxoid
UNAIDS	Joint United Nations Program on HIV/AIDS
UNFPA	United Nations Population Fund
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
UTFR	Unwanted Total Fertility Rate
VAS	Visual and Analogue Scale
VAW	Violence Against Women
VCT	Voluntary Counseling and Testing
VDRL	Venereal Disease Reference Laboratory
VE	Vaccine Efficacy
VSC	Voluntary Surgical Contraception
WFS	World Fertility Survey
WHO	World Health Organization
WHO/MI	World Health Organization Micronutrient
WHR	World Health Report
WRA	Women of Reproductive Age
WTFR	Wanted Total Fertility Rate

Summary List of Indicators

This *Compendium* includes the following indicators, organized in two sections: those that cross-cut programmatic areas (described in Part II of this volume) and those that are specific to the different programmatic areas that comprise reproductive health (Part III of this volume). Each indicator is described in detail on the pages indicated below.

Part I
Overview of the *Compendium*

A. Rationale for the *Compendium*

1. Objectives of the *Compendium*
2. Intended Audience
3. Organization of the *Compendium*
4. Selection of Indicators
5. Use of the *Compendium*

B. The Use of Indicators in Program Evaluation

1. Defining Program Evaluation
2. The Value of a Conceptual Framework
3. Program-based versus Population-based Measures
4. Sources of Data
5. Input, Process, Output, and Outcome
6. Causality
7. Quantitative versus Qualitative Indicators
8. The Results Framework
9. Levels of Reporting for USAID Cooperating Agencies

Part I

Overview of the Compendium

A. Rationale

Two events in the 1990s prompted the development of the present *Compendium*. The first, the International Conference on Population and Development (ICPD), held in Cairo, Egypt in 1994, shifted the paradigm conceptualizing women's health. In contrast to the earlier focus on family planning (FP) and (in some countries) on demographic objectives, the new paradigm emphasizes quality of life, gender equity, and human rights. The Cairo Conference caused the international community to reconsider the definition of reproductive health (RH) and the services that programs should offer. As a result, the field of reproductive health has broadened to include multiple areas in addition to family planning: safe motherhood, STI/HIV/AIDS, women's nutrition, breastfeeding, postabortion care (PAC), reproductive health services for adolescents, violence against women (VAW), female genital cutting (FGC), and related topics. In the wake of the Cairo Conference, many governments and non-governmental organizations (NGOs) have risen to meet the challenge of offering such a broad range of services, often with inadequate funding to appropriately address the issues. The goals from ICPD require monitoring of reproductive health outcomes in terms of clearly defined indicators, as well as evidence-based results to demonstrate progress.

A second important trend has been increased emphasis on accountability from two sources: country programs and international donor agencies. Many governments, weary of funding programs that lead nowhere, embrace the concept of program evaluation for the purpose of learning how to strengthen programs in an effort to improve the lives and livelihoods of their populations. With regard to donor agencies, the United States Agency for International Development (USAID), the British Department for International Development (DFID), the United Nations Population Fund (UNFPA), the United Nations Children's Fund (UNICEF), and the World Health Organization (WHO) are among those donors who have focused particular attention on program evaluation, both for the purpose of improving programs and justifying the funds expended on these programs. In addition, private organizations, ranging from the large foundations to specialized NGOs, have actively sought more systematic methods to track progress and to measure results of their efforts.

This interest in program evaluation triggered a proliferation of initiatives to develop indicators on all aspects of reproductive health. The current volume follows in the tradition started by the *Handbook of Indicators for Family Planning Program Evaluation* (Bertrand, Magnani, and Knowles, 1994) and a series of manuals on *Indicators for Reproductive Health Program Evaluation* (Bertrand and Tsui, 1995). The demand for these publications has been high, with over 4,500 copies of these two publications distributed to date. These volumes include many indicators that were tested and used extensively in the field (e.g., total fertility rate, contraceptive prevalence). They also include indicators that were experimental in nature: the "best guess" of experts working on these topics. Over the past five years, program managers and evaluators have experimented with these indicators, especially in new areas of reproductive health (e.g., safe motherhood, postabortion care, and adolescent reproductive health programs). The current volume reaps the benefit of this experimentation, in that many of the chapters herein were drafted or revised by evaluation specialists who have used these indicators in their daily work.[1] As such, the *Compendium of Indicators* reflects far greater field-based experience than did the previous volumes.

Given the diversification and specialization within the field of reproductive health, few people can claim expertise across the full range of reproductive health topics. Whereas these different programs share the common goal of improving the quality of women's and men's lives through improving their reproductive health

[1] This statement does not apply to Parts III.L and III.M, which are emerging areas for reproductive health programs.

status, the actual measures of progress toward that goal differ from one type of program to the next.

1. Objectives of the *Compendium*

The general objective of this *Compendium* is to encourage program evaluation and to improve the quality of work in this area. To this end, the *Compendium* provides a comprehensive listing of the most widely used indicators for evaluating reproductive health programs in developing countries. Moreover, the indicators are organized according to a revised version of the conceptual framework originally developed under The EVALUATION Project. This framework maps the pathways through which programs achieve results (see Figure I.1), and it constitutes a logical framework for developing an evaluation plan with appropriate indicators. The original framework, created for family planning programs, is readily adaptable to other areas of reproductive health. Many sections of the *Compendium* contain more detailed frameworks that explain the pathways for program effects specific to the topic area in question.

Whereas some past evaluation efforts have treated the operations of reproductive health programs as a "black box," this framework specifies how those who design the program expect it to work to achieve results at both the program and population level. Moreover, the framework draws attention to the different aspects of programs (operational areas, access to services, quality of care) that must be working satisfactorily to achieve the desired end result.

The specific objectives of this *Compendium* are:

- To compile in a single publication a menu of reproductive health indicators judged most useful in evaluating reproductive health programs at both the program level and population level;
- To define these indicators in an effort to enhance the consistent use of terms across programs, countries, and donor agencies; and
- To promote evaluation of programs by making indicators readily available to evaluators.

2. Intended Audience

Several different audiences should find this *Compendium* pertinent to their own work, including:

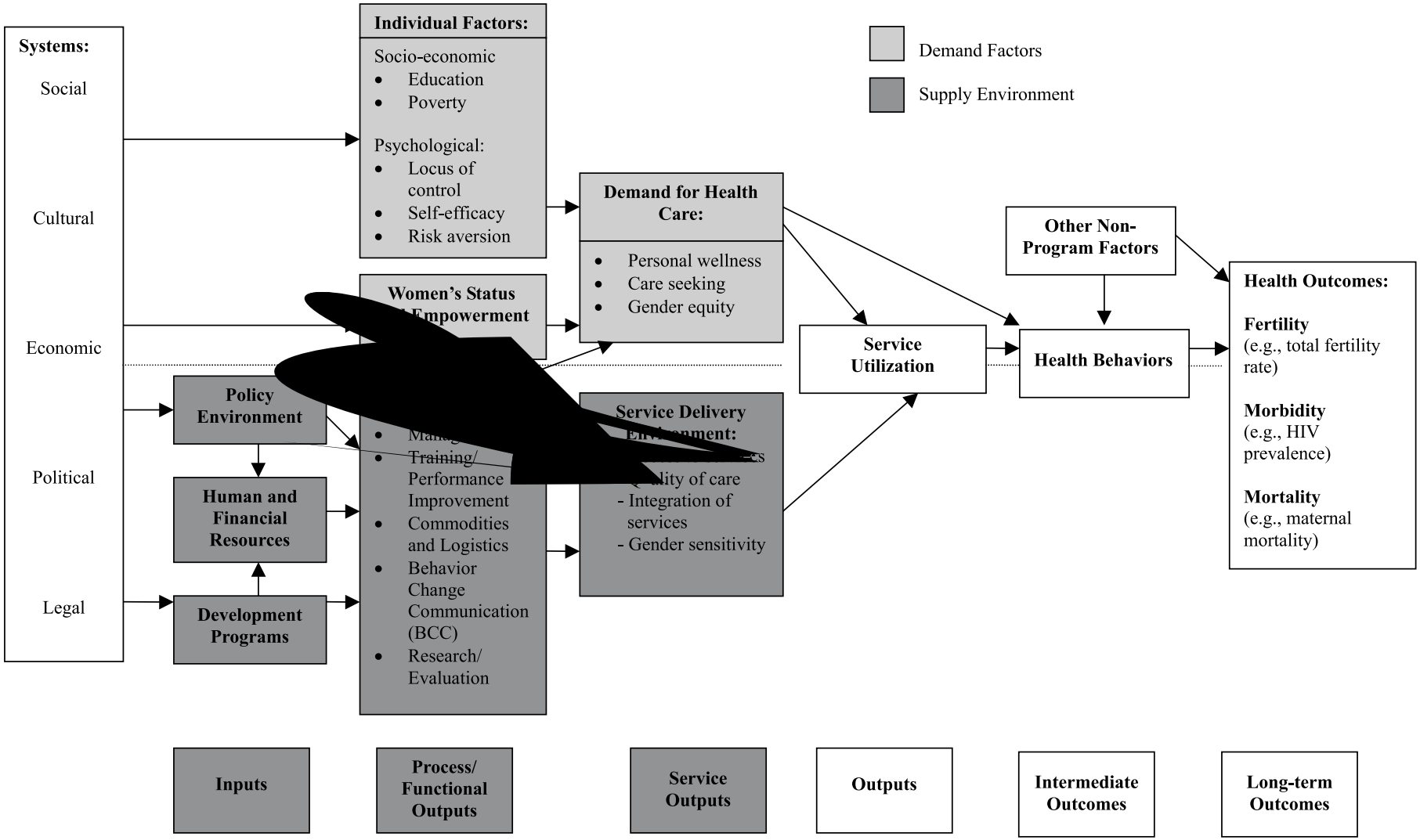
- Administrators, managers, and health workers in reproductive health programs worldwide;
- Staff in international family planning/ reproductive health (FP/RH) agencies responsible for designing and evaluating collaborative projects with host country institutions;
- In-country evaluation specialists responsible for monitoring performance and for evaluating the effectiveness of RH programs in specific settings;
- Private foundations and other donors supporting RH programs;
- Social science researchers (e.g., demographers, sociologists, economists), epidemiologists, and other health professionals; and
- Students and instructors.

3. Organization of this *Compendium*

This *Compendium* is organized in three parts. Part I provides an overview of the document, as well as basic concepts in program evaluation. The rest of the volume presents a menu of indicators (including the definition, data requirements, data sources, purpose and issues) consistent with the conceptual framework presented in Figure I.1, for different types of reproductive health programs.

Part II presents indicators that crosscut the different programmatic areas of reproductive health. It begins with **women's status and empowerment**, which has a pervasive influence on all aspects of reproductive health. In addition to the indicators described in this section, the role of gender is highlighted throughout the *Compendium* in a series of boxes that provide interpretations of other indicators from a gender perspective. The topic of gender appears again toward the end of Part II, section II.H.4, which presents indicators to evaluate institutions on gender equity in their managerial structures and on gender sensitivity in their service delivery systems. Finally, two of the programmatic areas in Part III of the *Compendium* are highly relevant to gender: male involvement and violence against women.

Figure I.1. Conceptual Framework: Achieving Results in Reproductive Health Programs



The rest of Part II deals with the supply environment for reproductive health. The "supply environment" refers to the multiple elements that influence the provision of reproductive health services, including:

- Resources and infrastructure available for development programs;
- The policy environment - political will, legal, and regulatory policies;
- The functional (or operational) areas that support service delivery; and
- The service delivery environment, characterized by access to services and quality of care.

Part II.B outlines indicators for evaluating initiatives to influence the **policy environment**. The subsequent five sections deal with the functional areas of RH programming (**management, training, commodities and logistics, behavior change communication, and research/evaluation**[2]) which collectively shape and influence the supply environment. One can directly evaluate the service delivery environment in terms of **access to services and quality of care,** as well as two sub-elements of quality: **integration of services and gender equity/sensitivity.**

Part III presents indicators specific to the major programmatic areas within reproductive health. It begins with the three areas that have received the greatest attention from the international reproductive health community: **family planning, STI/HIV/AIDS, and safe motherhood**. The latter is followed by a related but relatively new area for program evaluation: **newborn health**. **Women's nutrition** and **breastfeeding** follow. In the wake of the Cairo Conference, other areas have also entered the RH arena: **adolescent reproductive health programs, postabortion care, male involvement, violence against women, female genital cutting,** and **reproductive health in emergency situations**.

We have excluded topic areas that may be important but to date have not been the focus of major nation-wide programs, except in isolated cases or on a small scale (e.g., infertility, health sexuality, and cervical cancer screening).

4. Selection of Indicators

Persons with limited evaluation experience often are bewildered by the process of selecting indicators. Appendix B presents the following tips on the steps for selecting indicators:

- Clarifying the results statements (i.e., objectives to be obtained);
- Developing a list of possible indicators;
- Assessing each possible indicator; and
- Selecting the "best" indicators.

We used the following criteria to select indicators included in this *Compendium:*

- Validity (the indicator measures what it purports to measure);
- Importance (the measure captures something that "makes a difference" in program effectiveness);
- Usefulness (the results point to areas for improvement); and
- Feasibility (data can be obtained with reasonable and affordable effort).

To the extent possible, we selected indicators that had been field-tested, including those measured in Demographic and Health Surveys (DHS) and Reproductive Health Surveys (RHS).[3] In certain emerging areas (e.g., violence against women, female genital cutting), we also included indicators that represent recommendations of those working in these areas, to stimulate dialogue on evaluation in these areas and to provide possible indicators for testing as these new programs become more common.

Two topics, notable by their absence, are technical capacity and sustainability. Despite the importance of these topics to the field of reproductive health, the authors were not able to locate a set of indicators that had been tested and were in use at the field level to monitor technical capacity and sustainability (although the indicators for policy, training, and management do address some elements of capacity building).[4] A group working under the auspices of the MEASURE *Evaluation* Project on indicators for capacity building concluded that such indicators would need to be context

[2] *The Handbook of Indicators for Family Planning Program Evaluation* (1994) included indicators for monitoring and evaluation (M&E). However, to the authors' knowledge, such indicators are not in use in the field. By contrast, the development of indicators for a specialized area of applied research – operations research (OR) – has advanced markedly in recent years. Thus, in this *Compendium,* we have opted to include indicators for OR, rather than for M&E.

specific. This group is developing a set of guidelines that will assist organizations in framing the question and in identifying appropriate means of evaluating capacity building within their organizations (Brown, Lafond, and Macintyre, 2001).

This manual does not address the specialized evaluation of programs that relate to costs (e.g., cost analysis, cost-effectiveness analysis). Evaluators may use a number of the indicators (especially output measures, such as couple-years of protection, number of pregnant women tested and counseled for HIV, number of visits to male-oriented services, among others) to establish the cost per unit of output for a given service. However, the methodology for conducting cost analysis and cost-effectiveness analysis is sufficiently technical that we opt to refer interested readers to other specialized references (Janowitz and Bratt, 1994; Kumaranayake et al., 1998). We encourage evaluators to apply cost analysis and cost-effectiveness analysis in their own programs, despite the fact that we do not address the topic in this manual because of the specialized methodology required.

The wording of many of the indicators in the *Compendium* does not specify the sex of the person; rather, we refer to "audiences, participants, trainers, providers, adolescents, and others" without specifying if these are men or women, for the sake of simplification. At the same time, the descriptions in the text of the indicators often implies that the client is a woman, which is often but not necessarily the case. We have retained the words "she" and "her" to reflect the reality that despite efforts to include men in RH initiatives, the vast majority of clients for RH services in developing countries are women. However, when applying these indicators, the users of the *Compendium* are encouraged to specify the sex of the individuals in question to the extent possible.

5. Use of the *Compendium*

This *Compendium* serves as a reference document for use throughout the international reproductive health community. Although funded by USAID, it applies to reproductive health programs sponsored by a variety of funding agencies, governments, or NGOs worldwide. Specifically, the *Compendium* provides a menu of indicators to be used selectively as part of the evaluation of national programs, regional programs, and country projects. We stress that no program or project would ever attempt to use *all* indicators outlined in this *Compendium*. In fact, for routine monitoring purposes, program managers and evaluators should select a few relevant indicators that are both important to program objectives and easy to collect and interpret. If organizations need more feedback, then they can conduct special studies to evaluate the programs' performance in areas of particular interest to staff. In this case, organizations should stagger these special studies to minimize their research burden.

The indicator descriptions presented herein are designed to promote standardization of definitions and concepts among the international reproductive health community. Whereas standardization is useful, organizations should adapt indicators to their specific circumstances. This approach not only ensures that the indicators are relevant to the organization in question, but promotes ownership of the monitoring and evaluation process. However, organizations that choose to adapt (modify) indicators should clearly state the new definitions and methods of measurement.

Overwhelmed by the sheer volume of indicators in this *Compendium*, many readers will ask: **"What are the 10-15 key indicators that are essential for evaluating reproductive health programs worldwide?"** Indeed, the WHO grappled with this same question and arrived at a consensus list of 17 outcome indicators intended to track progress in country RH programs worldwide following the ICPD in Cairo in 1994 (see Part III.A). These indicators are useful to governments and international donor agencies to measure "the big picture."

However, the vast majority of program managers will find the global list impractical for evaluating their specific interventions for several reasons. First, the global indicators require population-based surveys, which are beyond the scope of organizations working in a specific region or NGOs working with specific sub-groups within the population (e.g., truck drivers and sex workers for HIV prevention). Second, many organizations want evaluation to cover not only the final results achieved, but also the progress made in specific

[3] These surveys are discussed in greater detail in Section I.B.3, below.

[4] Several reviewers cited work in progress or suggested indicators for sustainability, but given the focus of this *Compendium* on indicators in use at the field level, we have not included those under development.

functional areas, such as training, commodities and logistics, or behavior change communication (BCC).

The specific indicators appropriate for use in a given evaluation depend directly on its purpose. For example, the Director of Commodities and Logistics for a national family planning program may track all eight of the quantitative commodities and logistics indicators as well as the logistics systems assessment indicator (described in Part II.E). On the other hand, the Director of the National Family Planning Program will more likely focus on results achieved in terms of contraceptive prevalence, quality of care, and other more global indicators reflecting the collective achievements of functional areas.

This *Compendium* aims to improve program evaluation for the purpose of strengthening service delivery in national RH programs (as well as in regional programs or in specific projects) of a given country. Its purpose is not to generate a report card that pits one country against another for future funding consideration. Although this C*ompendium* encourages the consistent use of definitions and terms across countries or programs, socio-economic and cultural context in which programs operate differs greatly by country or region. For example, it is far easier to achieve a certain level of coverage for a media campaign in a small country with a monolingual population, a well-developed communications system, and a favorable policy environment than in a country where these conditions are not present.

Whereas cross-national comparisons on certain variables are inevitable and may, in certain cases, be useful, the results obtained for these indicators for RH evaluation are subject to misuse if evaluators and program managers do not interpret them in a country-specific context. In most cases, comparing indicators for a given program over time will be more productive than using the indicators for cross-national comparisons.

B. The Use of Indicators in Program Evaluation

1. Defining Program Evaluation

In this *Compendium* we use the term "program evaluation" to encompass (1) routine monitoring and (2) the different forms of evaluation: process, results, and impact. The term "evaluation," as used in this manual, may refer to any aspect of program evaluation.

Routine monitoring and process evaluation measure how well a program is working (Adamchak et al., 2000). By contrast, the evaluation of results and impact relate to the desired change (as defined by the objectives of the program). More specifically:

Monitoring is the routine tracking of a program's activities by measuring on a regular, ongoing basis whether planned activities are being carried out. Results reveal whether program activities are being implemented according to plan and indicate the extent to which a program's services are being used (Adamchak et al., 2000).

Process evaluation measures how well program activities are performed. This information is usually collected on a routine basis, such as through staff reports, but it may also be collected periodically in a larger-scale process evaluation effort (e.g., special studies) that may include use of focus groups or other qualitative methods. Process evaluation is used to measure the quality of program implementation and to assess coverage; it may also measure the extent to which program services are being used by the intended audience (Adamchak et al, 2000).

Evaluation of results measures the extent to which change occurs consistent with the program objectives. Many evaluations focus on change in the knowledge, attitudes, and behaviors of (1) clients/participants in the program or (2) members of the intended audience in the population at large (e.g., women of reproductive age). However, "results" may refer to changes in policies related to reproductive health, management procedures, logistics systems, quality of care in health facilities, and other aspects of the supply environment.

Impact assessment measures not only the change that has occurred, but also the extent to which this change is attributable to the program intervention. This type of evaluation is still fairly limited in international reproductive health programs. Generally, it involves (1) small, clinic-based experiments to determine the effectiveness of a given strategy or (2) multivariate multilevel regression analyses of large-scale surveys, based on DHS-type data, to determine the relative importance of different factors including the program intervention.

2. The Value of a Conceptual Framework

The conceptual framework illustrated in Figure I.1 is adapted from a similar model developed for family planning under The EVALUATION Project. This framework illustrates the pathways by which reproductive health programs achieve their objectives. The column on the far left defines the context in which the program operates: the social, cultural, economic, political, and legal systems in a given society, including that society's reproductive health programs. The top left-hand side of the figure, lightly shaded, outlines the role of **demand** in the effectiveness of a given program. Countries in which the population actively wants the services ("high demand") based on societal norms and preferences will have a far easier time achieving results than those in which the population is indifferent or outwardly negative toward the program.

The lower left-hand side of the conceptual framework lists factors in the **supply environment,** shaded in a darker tone. Countries with strong social and economic development programs provide a more conducive environment in which to promote reproductive health than those without systems to support such efforts. Strong political support ("political will") for a program also facilitates implementation, as illustrated by the family planning program in Bangladesh. Whereas donor agencies and program managers once treated policy as a contextual variable that would influence program implementation, today they actively design interventions (e.g., advocacy) with the aim of shaping the policy environment.

The supply environment also comprises the functional areas that support service delivery and the service delivery environment itself. The functional or operational areas of a program provide the structure for carrying out interventions, including management, training, logistics, BCC, and research/evaluation. Indeed, USAID and other donor agencies fund entire programs that strengthen the operations of agencies in developing countries in these areas. These functional areas contribute directly to the service delivery environment: that is, the services available to a prospective client in a given country. Measures of the service delivery environment focus on access to services and quality of care as well as sub-elements of quality: integration of services and gender equity/sensitivity.

These two sets of factors – supply and demand – jointly determine the level of service utilization in a given country. Although service utilization is not essential to the practice of certain behaviors (e.g., exclusive breastfeeding), it generally plays a key role in helping a client adopt healthy behaviors, through information and counseling, provision of supplies (e.g., condoms for AIDS prevention), or clinical procedures (e.g., insertion of an Intra-uterine device - IUD).

The box labeled "health behaviors" represents the objective of most RH programs: that is, the behaviors that members of the intended audience are encouraged to adopt. Examples include use of contraception for family planning, use of condoms or decrease in number of sexual partners for AIDS prevention, delivery with a skilled birth attendant, and breastfeeding. It is important to recognize that non-program factors may also play a role at this level in influencing both health behaviors and outcomes. For example, women are more susceptible to contracting HIV from an infected partner than men are. Fertility is determined not only by contraceptive use, but also age at marriage, extent of induced abortion, postpartum infecundability, and pathological sterility. The entire chain of causal events, outlined in Figure I.1, leading to specific health behaviors directly affects the ultimate objective of reproductive health programs: improved health outcomes in terms of fertility, mortality, and morbidity.

3. Program-based versus Population-based Measures

For the purposes of program evaluation, it is important to distinguish between program-based and population-based data. Program-based data consist of information available from program sources (e.g., administrative records, client records, service statistics) or information that can be obtained from on-site collection (e.g., observation, client-provider interaction, client exit interviews, mystery client surveys), although routine health information systems are the primary source of program-based data. Also, a follow-up study of clients who attended a clinic constitutes program-based data, in that the information on the clients comes from program records. Although some program-based data correspond to a limited network of clinics providing a specialized service, "program-based" also can refer to programs that are national in scope.

Program-based information is very important for understanding the performance of programs and the type of output they achieve (e.g., number of visits per month to a clinic, number of tetanus shots administered to

Box I.1. Main Sources of Data Used in Program Evaluation

Program-based data:

Service statistics (also known as routine health service statistics or health service data):

- Admission records (by diagnosis)
- Program intake interviews
- Client records
- Registries or databases that track the number of visits, clients, commodities distributed
- Clinic registries and registries for specific services (antenatal, emergency room, operating theater, maternity)
- Laboratory results
- Data from disease surveillance in clinics
- Data from sentinel surveillance in clinics
- Records on cause of death

Facility-based surveys:

- Facility audit of equipment and supplies
- Client exit interviews, client satisfaction
- Observation of client-provider interaction, of trainers
- Logistics site visits
- Service record review

Local surveys:

- Behavioral sentinel surveys (BSS) related to STI/HIV
- Prospective community studies

Surveys and other assessments of providers and other personnel:

- Performance ratings (often based on checklists)
- Pre- and post-test scores for training events
- Competency tests (used in conjunction with training courses)
- Course evaluations (e.g., training)
- Assessments by an external evaluator
- Observation of client-provider interaction
- Interviews with the staff, personnel surveys
- Self-assessment tools
- Mystery (simulated) clients

Program records (other than service statistics):

- Personnel rosters
- Financial data
- Job descriptions
- Official minutes of meetings
- Lists of participants in an event (e.g., training)
- Logistics data: stock levels, forecasts of products (needed for a given program), commodity procurement plans, product lists, planned shipment schedules
- Supervisory records, staff records
- Project documents (e.g., final reports)
- Service delivery protocols, medical and nursing standards
- Access logs generated by a web server

Population-based data:

Census (to determine population of the catchment area)

Vital registration:
- Registration of births
- Registration of deaths (by cause)

Surveys with a nationally representative sample:
- Demographic and Health Surveys (DHS), coordinated by Macro International (including both the core questionnaire and optional modules)
- Reproductive Health Surveys (RHS), by the CDC
- Multiple indicator cluster surveys (MICS), supported by UNICEF
- Rider survey (e.g., BUCEN)
- National surveys conducted independently of these two survey programs (the DHS and RHS)

Surveys with a representative sample of the population of interest:
- Baseline/follow-up surveys for evaluation purposes
- Behavioral surveillance surveys (BSS), used in connection with AIDS
- Opinion polls

Government documents and other official records:
- Official policies, plans, guidelines
- Laws, regulations, special statutes, court decisions
- Health codes, penal codes
- National expenditure budgets, accounts, invoices
- Official minutes of meetings, agendas, attendance lists
- Membership lists of networks or coalitions
- Voting records
- Police records (e.g., for gender-based violence)

Special studies:[5]

- Program Effort Index (e.g., for family planning, AIDS, and maternal and neonatal care)
- Time motion studies
- Media scan
- Online user feedback surveys
- Self-assessment tools
- Content analysis

Qualitative methods (generally applied to clients or other non-representative populations)
- Focus groups
- In-depth interviews
- Pile-sorts
- Ethnographic observation

[5] This category is not mutually exclusive with the previous ones. For example, any of the surveys with clients, providers, government officials, or others could be considered a special study. However, the items listed under this heading do not fit in the other categories on the list.

pregnant women). However, program-based data do not reflect the extent of coverage of these programs (unless one estimates a denominator for the catchment area that converts these program statistics into an estimation of a rate). Moreover, data from program participants are potentially biased (do not reflect the situation of the general population), because of selectivity; that is, persons who opt to participate in programs are often different from the population at large. NGOs tend to evaluate using program data alone, because they do not aspire to national coverage of the population at-large even in the defined area in which they work.

In contrast, governmental programs designed to have national coverage are evaluated in terms of their effect on the general public. The term "population-based" can refer to a smaller geographic region (e.g., the catchment area for a demonstration project, such as a district), provided the data are drawn from a representative sample of the population. The primary sources of population-based data for reproductive health activities are the DHS conducted by Macro International, and the RHS conducted by the Centers for Disease Control and Prevention (CDC). In addition, data are available in selected countries from national-level surveys funded by the host government or by other international agencies (e.g., the India Fertility Survey). The DHS and RHS surveys are particularly useful for measuring demand factors. The DHS service provision assessment (SPA) modules measure factors in the supply environment, but are not as widely applied as are the household surveys. (For further discussion of facility-based surveys, see Part II.H.)

4. Sources of Data

The indicator descriptions below contain a wide variety of data sources. In some cases, different terms are used in different program areas for similar concepts; moreover, the terms can vary from one country to another. For example, FP programs generally use the term "service statistics" to describe the data they keep on numbers of clients and on volume of contraceptives distributed. By contrast, those in maternal health programs may use the terms facility records, clinic registries, health services data, or related terms.

To assist readers in understanding the large number of terms used to describe sources of data, we summarize the major sources of data, as well as the specific categories that fall within each source, in Box I.1. The terms management information system (MIS) or health information system (HIS) refer to multiple types of information, generally some subset of the data listed under "program-based data" in Box I.1.

5. Input, Process, Output, and Outcome

In the earlier manuals produced by The EVALUATION Project, we classified the input, process, and output measures as **program-based**, in contrast to outcome which was classified as **population-based**. This approach is useful in evaluating national programs, such as a national family planning program. However, it is somewhat less useful (especially the term "outcome") for the evaluation of specific functional areas (such as training, logistics, and BCC).

For example, one objective of training programs is usually enhanced quality of care in the service delivery environment. Although the collective efforts of training will contribute to outcomes at the national-level (e.g., increased contraceptive prevalence, increased number of women delivering with a skilled attendant), the most direct and measurable effect of training is improved service quality. In this sense, the desired outcome for a series of training events is quality of care in a specific network of facilities. These results are not population-based, yet they represent the appropriate endpoint in measuring and evaluating training programs. Thus, in Box I.2 we define inputs, process, outputs, and outcomes with the caveat that the "desired outcomes" for a functional area (management, training, logistics, and BCC) may be appropriately measured at the program level.

The 1994 *Handbook of Indicators for Family Planning Program Evaluation* used a slightly different terminology. It classified intermediate outcomes as "effects" (e.g., changes in behaviors or practices) and long-term outcome as "impact" (e.g., changes in fertility or mortality). However, the term "impact" generates a tremendous amount of confusion, since to many evaluation specialists "impact analysis" implies a demonstration of causality. For this reason, we encourage the use of the terms "intermediate" and "long-term outcome," which do not imply causality and (especially in the case of the long-term outcomes) are generally subject to social and economic factors beyond the control of the program.

Box I.2. Definitions of Input, Process, Output, and Outcomes

Inputs are human and financial resources, physical facilities, equipment, and operational policies that enable program activities to be implemented.

Process refers to the multiple activities that are carried out to achieve the objectives of the program. It includes both what is done and how well it is done.

Although a high level of input is generally reflected in satisfactory program implementation, it is theoretically possible to have a high level of input but a poorly delivered program (for example, if a high-level administrator opposed to family planning were successful in blocking service delivery in facilities under his/her control). Conversely, there are countless real-life examples around the world where program staff, with highly inadequate resources, strive, nonetheless, to do the best work they can under the circumstances.

Output refers to the results of these efforts at the program level. Although program managers at the field level are interested in national trends that show the fruits of their efforts (e.g., contraceptive prevalence for family planning, prevalence of breastfeeding for breastfeeding promotion), they tend to limit the evaluation of their own activities to program-based measures, especially measures of output. Two types of output, shown in separate boxes in Figure I.1, are service output (that measures the adequacy of the service delivery system) and service utilization (that measures the extent to which clients use the services).

Outcome generally refers to results of programs measurable at the population level.[6] The evaluation of outcome measures the effect that the program has had on the general population in the given catchment area (such as all women of reproductive age in a given country). It is important to distinguish between two kinds of outcome: intermediate and long-term. Intermediate outcomes tend to refer to specific behaviors or practices on the part of the intended audience – such as contraceptive use, breastfeeding, condom use, consumption of micronutrient supplements – that will affect the desired long-term outcome (of reducing mortality, morbidity, or fertility). The long-term outcome refers to the anticipated results of the program (a change in morbidity, mortality, or fertility). However, the long-term outcome is almost always subject to the influence of non-program factors, including socio-economic conditions and the status of women in a given country.

[6] Note the exception discussed in the text above that "desired outcomes" for functional areas are not necessarily measured at the population level.

Whereas evaluators often report the findings from a program evaluation for the population as a whole, they should disaggregate the results by regions or subgroups, where appropriate. A program that achieves results by providing services to privileged urban residents has achieved less than one that reaches the rural poor. Similarly, where data are available and the n is sufficiently large, evaluators may wish to disaggregate data by ethnic, geographic, and other relevant socio-demographic factors. The reduction of inequalities among regions and subgroups may be a programmatic objective above and beyond the improvement of the average among the general population. Evaluating this type of objective requires disaggregation and measurement by subgroups.

Reproductive health interventions generally address deeply entrenched practices that will only change over an extended period of time. The unprecedented drop in fertility rates worldwide in the late 20th century occurred with record speed, yet it took 20-30 years to accomplish. From a program perspective, it is often impractical to report annually on long-term outcome indicators, even if those outcomes "really matter" (e.g., changes in mortality and fertility). Instead, program evaluation tends to focus on intermediate outcomes (also called effects) that are more directly linked to program effort and expected to change in a shorter period of time. In terms of family planning, the most widely reported intermediate outcome is contraceptive prevalence (i.e., the percentage of women of reproductive age currently using a contraceptive method). HIV prevention programs often evaluate "use of condoms at last sexual intercourse" among members of the at-risk population. In the case of programs to eliminate female genital cutting, an intermediate outcome is the percentage of the population favorable to the continuation of this practice.

6. Causality

Most programs have the objective of achieving some type of change at the population level (among the general public) or at the program level (among clients or participants in the program). The purpose of many program evaluations is to measure whether that change occurs. Yet relatively few evaluations go as far as establishing cause-and-effect between the program and the change (e.g., impact assessment that allows for attribution).

The indicators per se do not define whether an evaluation measures impact. Rather, the study design determines whether one can establish impact (causality). We illustrate this point with two examples. First, suppose that Country X has had a nationwide program to promote breastfeeding over the past five years. DHS surveys conducted in Year two and Year five demonstrate a clear increase in the prevalence of exclusive breastfeeding. Whereas we assume that the program contributed to this increase, we cannot definitely attribute this change to our intervention. Perhaps another NGO may have promoted breastfeeding in the same areas, without the knowledge of the Ministry of Health (MOH). Such confounding factors may be negative in nature, yet positively affect the outcome (e.g., an economic downturn in the country making it too expensive for mothers to purchase milk). This example illustrates a situation in which the program achieved its objective (increase breastfeeding), but one cannot rule out factors other than the program ("confounding factors") that might be responsible for the change. Tracking change in a given population (in the absence of a control group) is also known as monitoring of results or "trend monitoring." Many policy makers are entirely satisfied with this type of evaluation, especially if the results show the desired change in the outcome variable; they are not concerned about confounding factors that might also explain the change. Indeed, many would conclude that if the desired change occurs, then the "program has had impact." By contrast, evaluation specialists recognize that simple tracking of change does not demonstrate causality.

In the second case, suppose the government is interested in evaluating the impact of its breastfeeding initiative. To this end, it randomly allocates the 20 provinces in the country to a treatment and a comparison area (i.e., 10 provinces each). A baseline survey conducted in both areas indicates that the two have similar socio-economic characteristics as well as breastfeeding practices. The government then implements an intensive breastfeeding promotion in the experimental provinces, but not in the comparison area (although it plans to implement the same activities after it concludes the evaluation in the latter, provided the intervention is effective). Three years post-implementation, the government conducts a follow-up study in both areas. The results indicate that exclusive breastfeeding is significantly higher in the experimental than comparison provinces. The results of this quasi-experimental study allow the evaluators to measure what happened in the absence of the

promotion, and thus to attribute the increase in breastfeeding to the intervention. In short, this second case fulfills the requirements for an impact assessment.[7]

Note: in these two examples, the indicators could have been the same. What differed – and allowed for impact assessment in the second case – was the study design. (For a more detailed discussion, see Bertrand, Magnani, and Rutenberg, 1996, Chapter IV.)

Many of the indicators in this manual are listed in terms of percentages, because they involve dichotomous variables with a "yes/no" answer (e.g., use of contraception, use of a condom at last sexual intercourse, use of exclusive breastfeeding, presence of a skilled attendant at delivery). However, some percentages in this *Compendium* are calculated by dichotomizing a continuous variable. For example, the **Percent of Pregnant Women Who Gain at least 1.0 Kg per Month in the Last Two Trimesters of Pregnancy** is derived from the continuous variable "weight gain among pregnant women in the last two trimesters of pregnancy." The use of simple percentages (when the underlying variable is continuous) is justifiable for purposes of tracking change over time. However, the dichotomization of a continuous variable represents a loss of information, and evaluators attempting to determine causality and control for confounding factors will need to use such variables in their continuous form.

7. Quantitative versus Qualitative Indicators

Whereas quantitative research has dominated the field of health and social science research in the past, qualitative research gained wide acceptance during the 1990s, to the point that the latter has become an integral part of many social science research projects. Similarly, qualitative research is widely used in program evaluation, especially in relation to process evaluation (e.g., to measure client satisfaction or participant reactions to the program). Focus groups, in-depth interviews, observation, and interviews with key informants constitute the most commonly used methodologies. Although some researchers have attempted to quantify the results of qualitative techniques, the general trend in the field is to capture the main ideas of respondents through narrative text rather than through percentages and other statistics.

The authors strongly support the use of qualitative methods in program evaluation as a complement to quantitative techniques and quantifiable indicators. Qualitative research is particularly useful in four areas:

- Conducting needs assessments (to learn more about the local situation before designing the program);[8]
- Understanding the local terminology for a given subject (prior to finalizing quantitative data collection instruments);
- Evaluating process (documenting the dynamics of how a program works, as well as its strengths and weaknesses); and
- Developing a clearer understanding of the results obtained from a quantitative instrument (e.g., the attitudes, beliefs, and values that underlie a given finding).

By contrast, quantification is essential for measuring results and impact.

In sum, qualitative research plays an increasingly important role in program evaluation, and we encourage evaluators to use both quantitative and qualitative data to best understand program dynamics. However, the use of qualitative research techniques is beyond the scope of this *Compendium*, which focuses instead on quantifiable indicators used to monitor program activities, to measure results, and to demonstrate impact.

8. The Results Framework

During the mid-1990s, USAID introduced a new approach to monitor its programs throughout the agency, known as Performance Monitoring Plans (PMP). Central to PMP is the results framework – a planning, communications and management tool. The results framework includes the strategic objective and all intermediate results, whether funded by USAID or its partners, necessary to achieve the objective. The framework also conveys the development hypothesis implicit in the strat-

[7] This hypothetical example illustrates the elements needed to demonstrate causality, but it does not accurately reflect the practical realities and difficulties of field experiments. For example, top-level administrators often oppose withholding potential benefits from any groups in their administrative area. Another problem is that comparison areas are rarely "similar" on all relevant socio-demographic variables as well as on the behaviors under study.

[8] Needs assessment is a well-recognized component of the design of programs, though some would question whether it forms part of program *evaluation,* since it occurs before the program is developed. It is part of formative research.

egy and the cause and effect linkages between the intermediate results and the objective. It includes any critical assumptions of the development hypothesis that must hold to achieve the relevant objective. Typically the framework appears in graphic form supplemented by a narrative (PriceWaterhouseCoopers, 2001). Evaluators can apply this conceptual framework to country-level programs (indeed, every USAID Mission worldwide has such a results framework), as well as to the programs of specific USAID-funded projects (a number of which are illustrated in Part II of this *Compendium)*. Other donor agencies have similar tools for program monitoring. Although they are not described in detail here, they also require that evaluators select appropriate indicators to track the progress of the program.

The results framework has a number of advantages. First, it has shifted the focus from a simplistic "bean counting" of activities toward the achievement of results. The number of posters one produces or the number of nurses one trains no longer matters. The question is instead, "what result did the program achieve?" A second advantage of the results framework is that its widespread application throughout USAID has greatly enhanced the understanding of program monitoring and evaluation for all levels of program personnel. Whereas evaluation used to be passed off to a lone evaluator in a back office, the results framework is now an integral part of the design of the program. Program managers are accountable for achieving the results they outline in their results frameworks, and program personnel understand that these indicators represent the criteria on which their efforts will be evaluated.

The results framework does have limitations. First, the framework describes the ways that program interventions will contribute to the ultimate results through cause-and-effect linkages (e.g., lower level results that contribute to the achievement of the strategic objective). In reality, however, the framework rarely traces all possible influences, such that factors other than the program can influence the results observed in ways that the results framework does not depict (i.e., the attribution problem). Second, the desired results are not always easy to measure in quantifiable terms. For example, the Central American HIV/AIDS Prevention Project in Central America (Programa Acción SIDA de Centro America – PASCA) was designed to change the policy environment for HIV/AIDS programs and to strengthen NGO capacity to implement prevention programs in the region of Central America. The need for indicators to measure the policy environment gave birth to the now widely used AIDS Program Effort Index (API), but this measure was totally untested at the time it was first used. Similarly, the measures for assessing institutional capacity building represent the "best guesses" of the expert group assembled to develop them, yet they lack the methodological rigor associated with more common indicators, such as prevalence of contraceptive use or breastfeeding.

A third limitation is not inherent in the framework itself but rather in its application. If program managers are fearful of failing to reach the behavioral objectives of their programs in terms of changing behavior at the population level (e.g., outcomes), they may revert to output measures that are within their manageable interest. For RH areas where output and outcome measures are closely linked (e.g., immunization), this shortcoming is less evident. In contrast, in areas where the link between outputs and outcomes is more tenuous (e.g., women's nutrition), this limitation is more important.

9. Levels of Reporting for USAID Cooperating Agencies

The indicators in this *Compendium* are presented for use at the national level (e.g., prevalence of breastfeeding, the percentage of facilities that offer postabortion care, the percentage of pregnant women who are anemic). However, evaluators should adapt these indicators to suit the needs of the organization using them. For example, if a project works in a single district or set of districts, this geographical area becomes the population of interest for the project. Similarly, if an adolescent project only works in the capital city, youth of a certain age (possibly defined by socio-economic or geographic variables) in this city become the intended audience.

The USAID Cooperating Agencies or CAs (organizations supported by USAID funding that give technical assistance worldwide) have to satisfy two different needs for indicators. First, they work with host-country counterparts in establishing the best indicators to use for evaluating the program or project at the country level or, alternatively, the regional district or city level, as described above. Second, they may need to report the aggregated results of their work in different countries to the donor (e.g., USAID).

The following example illustrates how evaluators can adapt the "same" indicators to assess a program at the host country level and at the global level. A training organization working in collaboration with host country counterparts may adopt the following indicators from Part II.D to measure national capacity for training:

- Capacity to maintain a functional MIS on the training program;
- Existence of a training strategy based on needs assessment to improve quality of service delivery; and
- Systematic evaluation of the organization's training program to improve effectiveness.

This same training organization may then use these indicators to report to USAID on **the number of countries** that achieved these results (of the total in which they worked).

This *Compendium* contains over 300 indicators. Many are in widespread use throughout the world to evaluate RH programs. Others are in an experimentation phase, as reproductive health expands into new areas (e.g., male involvement, female genital cutting, violence against women). The remainder of the *Compendium* consists of a "menu" of indicators to measure background factors of the service delivery environment (in Part II) and outputs/outcomes of 12 different programmatic areas of reproductive health (in Part III).

Part II
Indicators that Crosscut Programmatic Areas

A. Women's Status and Empowerment

B. The Policy Environment

C. Management

D. Training

E. Commodities and Logistics

F. Behavior Change Communication (BCC)

G. Operations Research (OR)

H. The Service Delivery Environment

Part II

Indicators that Crosscut Programmatic Areas

Part II of the *Compendium* consists of indicators that crosscut programmatic areas, that is, that are applicable to the different aspects of reproductive health treated individually in Part III. These crosscutting indicators fall into three categories:

- Background factors;
- Functional (or operational) areas; and
- Service delivery environment.

Background factors describe the context in which reproductive health programs operate. **Women's status and empowerment** is one such factor, in that it influences many aspects of health-seeking behavior: access to the resources to seek health services, autonomy to visit the services, ability to negotiate with one's partner, and likelihood of seeking protection from gender-based violence, to mention a few. The levels of status and empowerment contribute to the level of treatment women receive in the service delivery environment; this treatment in turn affects their decision to continue practicing desirable health behaviors. The **policy environment**, another important contextual factor, can facilitate or hinder the development of reproductive health programs. For example, many attribute the success of selected countries (e.g., Uganda, Thailand, Senegal) in stemming the AIDS epidemic to the support of the national leadership that supports AIDS prevention activities.

Functional areas refer to the operations supporting the actual delivery of services in a given country. These areas often constitute the program divisions within a Ministry of Health or NGO: **management, training, commodities and logistics,** and **behavior change communication.** To this list we add **operations research**, which provides organizations with the means to experiment with alternative strategies, and thus to improve service delivery and program performance.

These functional areas collectively define the service delivery environment – in terms of the quantity and quality of RH services available to the client who seeks them in a given community. In the present day context, one measures "good RH services" in terms of **access, quality of care, integration of services,** and **gender sensitivity**. One process allowing organizations to progress in these areas is **performance improvement.**

Part II of the *Compendium* presents indicators to measure each of these different areas. As the conceptual framework in Figure I.1 shows, these factors form an important part of the causal chain that determines whether a given population will seek services, will adopt specific health behaviors, and will achieve positive health status.

Part II. A Women's Status and Empowerment

- Percent of women who have completed at least four years of schooling
- Percent of women who have weekly exposure to mass media
- Percent of women who earn cash
- Percent of women who mainly decide how their earnings should be used
- Percent of women allowed to go alone to a health center
- Participation of women in household decision-making index

Part II.A

Women's Status and Empowerment

Whereas the sex of an individual is a biological phenomenon, gender is a social construct, defined by societal norms that attribute different roles and values to men and women. Moreover, these sex-specific roles, rights, and obligations are not just different, they also tend to be unequal (Kishor, 1999). We begin the *Compendium* with this section on women's status and empowerment, because these factors influence all subsequent elements in the causal chain that ultimately determines health status.

Unequal gender relations – existing to varying degrees in most if not all countries of the world – feed directly into the status accorded to women in society. Women's status is a term that describes women's situation in both absolute terms and in terms relative to men. The focus of women's status measurement has typically been on women's access to, and utilization of, information and resources (e.g., access to education, access to cash employment). Women's empowerment is a related term that focuses on attention to women's degree of control over their own lives and environments and over the lives of those in their care, such as their children. "Autonomy" is a related concept that also reflects women's control over their lives and environment, as well as status. Although several indicators of women's status and empowerment are available from the DHS core questionnaire or from the women's status module (described in "A Framework for Understanding the Role of Gender and Women's Status in Health and Population Outcomes," Kishor, 1999), we present six that researchers and evaluators have used to date (e.g., field tested). Status and empowerment are intended to reflect the extent to which egalitarian gender relations are achieved.

Gender equity is an end in itself. A recent World Bank publication cites gender equality as a core development issue and objective. Research from economics, law, demography, sociology, and other disciplines demonstrates widespread gender gaps in access and control of resources, economic opportunities, power, and political voice. These gender inequalities not only impose costs on the health and well-being of men, women, and children, but also diminish a country's prospects for development. In addition to these personal costs, societies that discriminate by gender pay a high price in terms of their ability to govern effectively, to reduce poverty, and to pursue economic progress (World Bank, 2001).

Many groups have challenged the societal structures granting men greater power than women, and they fight for a greater balance between the two sexes in all aspects of daily life. Because of the pervasiveness of gender differentials in societies worldwide, the task of shifting this balance of power is mammoth. Nonetheless, some programs are taking on this challenge to redress gender equities, such as those programs focused on attracting and retaining girls in primary and secondary school.

In this section, we address gender as it relates to reproductive health outcomes. A major theme of the 1994 International Conference on Population and Development (ICPD) in Cairo was that gender equity is the single greatest catalyst to fertility decline. As women gain greater control over all aspects of their lives, they will be motivated and able to control their own fertility, presumably at levels lower than the current ones in most developing countries. By the same logic, greater gender equity will allow women to break down the obstacles to receiving treatment for life-threatening complications of childbirth (e.g., lack of access to family resources, requirement of husband's consent to seek treatment), which in turn will decrease maternal mortality and morbidity.

A parallel line of reasoning is that gender equity directly influences health outcomes, in terms of both the supply and demand for services (illustrated in Figure I.1). This perspective does not diminish the value of gender equity as an end in itself. However, it treats gender in relation to factors that ultimately determine health-seeking behavior, service utilization, and desirable health practices.

For example, gender inequity diminishes the likelihood that women will seek health services or perform healthy behaviors because they:

- Lack knowledge of healthy practices and sources of service;
- Have limited access to resources, including nutrition and health care; and
- Lack control over decision-making as it relates to number of children, protection from STIs, and related topics.

The low status of women also affects **the supply environment** (e.g., service delivery). The lack of available emergency obstetrical care in many parts of the world has been linked to the low value placed on women's lives (Rosenfield and Figdor, 2001). Several studies have demonstrated that women of lower status receive treatment inferior to that of their higher status counterparts, even from the same set of providers at the same facility (e.g., Schuler and Hossain, 1998). Women from ethnic minority groups often experience additional barriers when they seek services, especially if they do not speak the predominant language of that country.

Methodological Challenges of Evaluating Women's Status and Empowerment

- **Health professionals generally acknowledge the role of women's status and empowerment in health outcomes, but they may consider such equity "beyond their manageable control."**

Given the wide-ranging nature of gender differentials in society, many program managers feel they have neither the mandate nor the means to directly change this deeply entrenched set of values. Those who have attempted to develop programs to influence power relations in sexual relationships have frequently faced the viewpoint that gender relations are a component of "culture," which is seen as nebulous, static, and impermeable to intervention (Helzner, 1996; Clark, 1998).

- **The traditional public health approach differs significantly from the women's rights/gender empowerment perspective.**

Applying a human rights framework to RH programs means, among other things, focusing as much on the process as the outcome, incorporating efforts to address the gender and power dimensions of reproductive and sexual decision-making into every level of program implementation, and building a sense of entitlement among the seekers and the providers of services (Jacobson, 2000). When faced with competing demands for a very limited cache of resources, many program managers are simply not ready to "take on" a human rights-based approach to gender. Instead, they prefer to focus on providing conditions that will circumvent or negate the ill effects of gender inequities (e.g., providing contraceptive methods that women can take without their husbands' knowledge). Were gender equity a more central part of international RH programming, managers might feel a greater urgency to develop indicators to track progress in this area.

- **Some individuals perceive gender as an amorphous concept that does not lend itself to measurement.**

Concepts central to gender inequity – such as value systems, decision-making, and control of resources – seem abstract; they elude measurement. Indeed, the groups assembled to develop indicators of women's status and empowerment in various contexts have experienced difficulty in deciding which elements to measure. One group took the approach of identifying ways that gender inequities hinder the success of RH programs. Whereas this strategy proved logical and feasible, the group disliked defining women's status as a barrier; rather, they preferred to seek positive means to integrate gender into reproductive health interventions (Yinger et al., 2001).

Blanc (2001) cites the lack of useful and practical measures of power relations in her comprehensive review of the balance of power in sexual relationships. Although certain measures have been linked to specific outcomes in some settings, such relationships may not hold in other settings (e.g., whether they would be cross-culturally valid). As Blanc observes, power relations themselves are rarely measured; thus, if a desirable outcome occurs, assigning it to a change in power relations may be impossible.

- **To measure gender inequity, one must have comparisons, not just a single number.**

Documenting gender inequities (for example, in access to resources) with a single number (e.g., percent, mean/median, is difficult, because gender is relational. Gender inequality reflects that women have lower access to

power and resources than men. Higher status women (e.g., service providers) may also discriminate against lower status women clients on the basis of gender, and thereby limit their access to services. Exceptions include selected indicators related to discrimination in the workplace: percent of women in management positions or societal outcomes of gender discrimination (e.g., female infanticide, violence against women). Thus, documenting disparities in gender equity often requires disaggregation and comparisons of data by sex (e.g., males versus females). Improvement in female education (e.g., percent of girls enrolled in school) over time is generally accepted as a sign of progress contributing to the empowerment of women.

- **Programs generally focus on women, the "disadvantaged gender," with relatively little attention to men's needs.**

With respect to gender differentials in power and in access to resources, men generally appear to have the advantage. Yet with regard to contraceptive and RH services, programs have largely ignored men. This approach not only has excluded men from active participation in maintaining reproductive and child health, but also has clearly put the burden for doing so on women. The HIV/AIDS epidemic has made evident that exclusion of men is untenable, in large part because such exclusion ultimately contributes to the disadvantage and disempowerment of women.

Treatment of Gender in this *Compendium*

Because gender plays such a complex and pervasive role in all aspects of human life, it is challenging to define an exhaustive set of "gender indicators" (for reasons described below). Rather, we treat the issue of gender in four different ways in this *Compendium.*

1. Population-based Indicators of Women's Status and Empowerment

In this section we describe selected population-based indicators of women's status and empowerment that are available from the DHS and other large-scale national surveys. Evaluators can use these indicators in one of three ways. First, the simple tracking of the indicator can document progress (e.g., increasing levels of female education). Second, indicators can serve to document gender differentials in male/female comparisons. And third, indicators can serve to demonstrate the effect of women's status on other behaviors or phenomena (e.g., women with less education have lower decision-making power). For a useful presentation of these relationships from DHS data, see Kishor (1999).

The position of "women's status and empowerment" on the conceptual framework in Figure I.1 illustrates the extensive influence of gender on all subsequent aspects of RH programs. In short, women's status influences demand for both RH services and the supply environment, and through these factors, reproductive health status.

2. Gender in Managerial Structures and in Service Delivery

Gender affects two aspects of the supply environment for reproductive health services in measurable ways: the managerial structure and the service delivery system. For example, men generally make higher salaries than do women for comparable work; women often hold the lowest paying jobs in an organization. In terms of service delivery, women of lower status may get treatment inferior to that of their higher status counterparts, even in the same facility. Providers may inadvertently reinforce gender stereotypes through messages they communicate to clients. Although indicators of these phenomena are relatively new, we include a checklist of factors to assess institutions on gender equity and to evaluate service delivery environments on gender sensitivity. (See Part II.H.4).

3. A "Gender Interpretation" of Other Indicators

Because of the far-reaching influence of gender in all aspects of service delivery, we present a series of boxes in the text of Parts II and III of the *Compendium* that explain how evaluators can interpret other indicators from a gender perspective. Often by disaggregating data by sex (e.g., males versus females), one can identify the effects of gender inequity in a system. The boxes do not present indicators of women's status and empowerment per se, but rather the boxes indicate how to analyze the results from a gender perspective.

4. Indicators for Gender-Relevant Programs

Part III of the *Compendium* presents indicators for specific areas of reproductive health; we present two sections particularly relevant to gender. The first is male involvement. In the wake of the Cairo Conference, programs worldwide have attempted to incorporate men into RH programming to a greater extent, both to support health-seeking behaviors in their partners and to participate directly by adopting practices that foster improved reproductive health. The second topic is violence against women, which constitutes one of the most harmful physical expressions of gender inequity. Intervention programs on this subject are still very new. According to researchers/evaluators in this area, the data on these indicators can be perplexing and counterintuitive (e.g., the reported rates of violent acts may increase at the onset of programs designed to combat violence against women, precisely because more women are willing to admit to abuse). Moreover, the use of such data could damage the very women the programs seek to help, if interviewers do not handle the data collection carefully. Violence screening and services heighten the need to protect clients' confidentiality and to maintain procedures if confidentiality is violated.

Categories of Indicators of Women's Status and Empowerment

The World Bank report entitled *Engendering Development: Through Gender Equality in Rights, Resources, and Voice* (2001) cites three important forms of gender disparity, along with indicators for measurement:

- Social, economic, and legal rights, measured by:
    - Political and legal equality;
    - Social and economic equality; and
    - Equality in marriage and in divorce proceedings.

- Access to productive resources, measured by:
    - Education – primary and secondary enrollment rates;
    - Health – life expectancy at birth, the burden of disease, gender-related violence, HIV/AIDS;
    - Productive Assets – land ownership, access to information, technology, and financial resources; and
    - Employment and earnings – labor force participation, occupation representation ratio, and relative earnings.

- Voice to influence decisions in their communities and at a national level, measured by:
    - Participation in politics; and
    - Representation in elected office.

The indicators that follow (selected in part because they are available through the DHS) focus primarily on access to productive resources and social equality.

Indicator

Percent of Women Who Have Completed at Least Four Years of Schooling

Definition
This indicator measures the percent of women ages 15-49 who have completed at least a primary level of education. For different countries, primary education may vary from four years to eight to ten years.

This indicator is calculated as:

$$\frac{\text{\# of women ages 15-49 who completed four years of schooling}}{\text{Total \# of women ages 15-49}} \times 100$$

Data Requirements
Information on the number of women ages 15-49 who completed primary school and information on the total number of women ages 15-49 surveyed

Data Source(s)
Population-based survey such as the DHS or RHS

Purpose and Issues
Educational attainment of populations vary greatly among countries. For example, in India, where educational attainment of women is very low, even a measure, such as the percentage of women who are literate can suffice. On the other hand, in countries such as Kazakhstan or Colombia where primary education is almost universal, a more appropriate measure may be the percentage of women who have completed high school.

This indicator is preferable to the percentage of women who have completed primary school, since the number of years required to complete primary schooling varies by country (e.g., five years in Egypt, eight years in Kenya). One must also be wary of calculating completion rates for 15-19 year olds, as many are still in school and, in sub-Saharan Africa, many are still in primary school because of the late age of entry into school and because of having to repeat the grade.

Further analysis of this indicator by age group can provide some indication of a changing climate for female education. Specifically, one would expect women 25-29 to have completed more schooling than older age cohorts have.

Educational attainment is important to gender equity for multiple reasons. It provides women with greater self-confidence and with power of logic to operate in an increasingly complex world. It gives them the cognitive skills and training necessary for participation in the workforce. It exposes them to non-traditional ways of thinking and provides alternative modes for behavior. Though education will not guarantee gender equity, it is an essential step toward it.

Indicator

Percent of Women Who Have Weekly Exposure to Mass Media

Definition
This indicator measures the total number of women aged 15-49 who report exposure to either radio, television, newspapers, or magazines at least once a week. The indicator is measured for television and radio. Evaluators may add questions regarding newspapers and magazines in addition to (though usually not instead of) television and/or radio.

This indicator is calculated as:

$$\frac{\text{\# of women ages 15-49 reporting exposure to radio or television at least once a week}}{\text{Total \# of women aged 15-49}} \times 100$$

Data Requirements
Information on the number of women 15-49 reporting exposure to radio or television at least once a week and the total number of women 15-49 surveyed

The DHS categorizes whether women have been exposed to radio, television, newspapers, or magazines as "almost everyday, at least once a week, less than once a week or not at all."

Data Source(s)
Population-based survey such as the DHS or RHS (on selected surveys)

Purpose and Issues
The mass media are one of the most important sources of information and exposure to new ideas, alternative role models, and non-kin-based power structures. The media play an even greater role in countries where women have low or no education, restricted freedom of movement, low levels of employment outside the home, or employment on the family farm. For men, too, the media are likely to be important, but perhaps less so than for women because men tend to have more alternative sources of information than do the women (i.e., they are more likely to be employed, educated, and able to move freely outside the home). Media exposure can be seen as a source of "empowerment" for women just as education is. In health and family planning research, women's exposure even to a single source of media, especially if it is television, is a powerful predictor of attitudes, beliefs, and actions, even controlling for education (Westoff and Bankole, 1997).

Indicator

Percent of Women Who Earn Cash

Definition
This indicator measures the percent of women aged 15-49 who work either at home or outside the home and earn cash. No minimum quantity is specified.

This indicator is calculated as:

$$\frac{\text{\# of women ages 15-49 earning cash}}{\text{Total \# of women aged 15-49}} \times 100$$

Data Requirements
Information on the number of women 15-49 employed earning cash and information on the total number of women surveyed

Note: Evaluators should include women who earn cash for crops in this indicator.

Data Source(s)
Population-based survey such as the DHS

Purpose and Issues
This indicator may reflect several types of empowering effects: exposure to networks other than kin networks, information from sources outside the family, and direct access to resources in the form of cash earnings. This indicator is preferred to the alternative indicator, the percentage of women who are currently employed, because research tends to find that women who are employed but do not earn cash are more "disempowered" than are women who do not work at all. Women who work without earning cash are likely to work on the family farm or to be in dependent situations, which deprive them any outside sources of "empowerment."

Indicator

Percent of Women Who Mainly Decide How Their Income Will Be Used

Definition
This indicator measures the percent of women aged 15-49 who mainly decide the use of the income they earn.

"Mainly" refers to the fact that she may receive some input from her husband, brother, or parent but that she maintains a degree of control over her own earnings.

The indicator is calculated as:

$$\frac{\text{\# of women ages 15-49 who mainly decide how their income will be used}}{\text{Total \# of women aged 15-49}} \times 100$$

Data Requirements
Information on the number of women who mainly decide the use of their income and the total number of women who work for cash surveyed

Data Source(s)
Population-based survey such as the DHS or RHS (on selected surveys)

Purpose and Issues
For women, having some money – however little – that they control is important for their real and perceived financial autonomy and rights. Sources of income and the ability to control how that income can be used makes women less dependent on others for financial support and thus increases flexibility in life options.

An alternative indicator is the percentage of women who participate (alone or with their husbands/someone else) in the decision about how their earnings should be used, although this may be less meaningful in measuring financial autonomy and rights.

Indicator

Percent of Women Allowed to Go Alone to a Health Center

Definition
This indicator measures the percent of women aged 15-49 who are permitted to visit the health center on their own (i.e., without anyone accompanying them).

This indicator is calculated as:

$$\frac{\text{\# of women ages 15-49 allowed to go to the health center alone}}{\text{Total \# of women aged 15-49}} \times 100$$

Data Requirements
Information on the number of women 15-49 who are allowed to go to the health center alone and information on the total number of women 15-49 surveyed

Data Source(s)
Population-based survey such as the DHS Women's Status Module

Purpose and Issues
A woman's ability to move about outside her home is a critical aspect of her empowerment. This indicator may not be appropriate for all countries, since the lack of freedom of movement is not a universal problem. However, this indicator may be appropriate for South Asian and Muslim countries where women's freedom of movement may be restricted.

This indicator specifies "health center" as a particularly relevant destination given the subject of this *Compendium* (regarding reproductive health). However, a more detailed measure – reflecting a number of common destinations combined into an index of mobility, based on the number of total places a woman may go – may be useful. Such places may include the local market, a community center or other nearby meeting place, friends' homes in the neighborhood, church/shrine/mosque/temple, or simply anywhere outside the house or compound. For every place a woman may go unescorted, evaluators assign one point. The total score would reflect the total number of places a woman may go alone.

Being able to go alone to a health center may not indicate all that much about a woman's autonomy (i.e., she may be able to go alone and still not be autonomous), but not being able to go places independently is highly indicative of no autonomy.

Indicator

Participation of Women in Household Decision-making Index

Definition
This indicator measures women's participation in the following five decisions:

- Determining own health care;
- Making large household purchases;
- Making daily household purchases;
- Visiting family or relatives; and
- Deciding what to prepare for daily meals.

A woman participates in a given decision when she alone or jointly with someone else makes the decision. The index is defined as the number of decisions a woman participates in. It is calculated by giving a score of 1 to each decision a woman participates (and 0 otherwise) in alone or jointly with someone else and then taking the sum. The index value will thus range from 0 (participates in none of the five decisions) to 5 (participates in all five decisions).

Data Requirements
Responses to the question "Who in your family usually has the final say on the following decisions: determining your own health care, making large household purchases, making household purchases for daily needs, visiting family and relatives, deciding what to prepare for daily meals?"

Data Source(s)
Population-based survey such as the DHS or RHS (on selected surveys)

Purpose and Issues
While relatively new, this index is perhaps the most direct measure of women's empowerment since it examines women's participation in at least five crucial decision-making processes. This type of information is now part of the core DHS questionnaire, and certain countries may have additional questions about other important decisions. Evaluators should frame the index in terms of women's participation (alone or jointly) in each of the five major decisions.

The higher the index score, the greater the indication of gender equity in decision-making. Aggregated individual index scores provide a measure of gender equity within regions or countries, or over time.

Part II.B
The Policy Environment

- Existence of policies, plans, guidelines that promote access to and/or quality of RH services
- Removal of barriers to RH policy development and/or service delivery
- Number of policy incentives to increase private sector participation in RH service delivery
- Resources available for RH
- Number of new financing mechanisms identified and tested
- Political and popular support for RH
- Participation in RH policymaking
- Number of NGO networks or coalitions working for RH
- Adequacy of the policy planning process

Part II.B

The Policy Environment

> Promoting policy reforms is an important tool in … overall development assistance… Experience in every sector has shown that gains realized from providing training, institutional capacity building, and direct resources are either enhanced or hindered by the policies, regulations, and administrative practices in that sector. When policies and regulations foster equitable opportunities and sustainable economic growth, they play an important part in creating an environment where development can flourish (USAID, 2000).

Policy and *policy environment* are often used interchangeably. The *policy environment* is both extrinsic and intrinsic to RH program operations: it forms part of the socio-political context in which programs must operate, and it influences the scope of program actions, the resources allocated, and the organizational structure of the program itself. We define a *supportive (*or *enabling) policy environment* as one in which (a) laws and executive orders mandate provision of products and services without imposing undue restrictions on providers or eligibility requirements on clients; (b) public and private resources are adequate to ensure full population coverage; (c) government and civil society leaders openly speak in favor of RH services and healthy practices; and (d) the policy formulation process is characterized by good planning principles and broad participation.

The indicators chosen for this chapter explicitly recognize both the broader socio-political context and the policy issues more narrowly focused on RH.

The policy environment is not static, but constantly changes in response to changes in the political and economic arenas, to changes in availability and costs of RH technologies, and to emerging public health issues. Thus, any policy evaluation must address the *processes* by which policies are formulated and revised as well as the policies themselves. Giving voice to groups previously under-represented in the policy process, such as women or rural citizens, may produce pressure for a different array of public services; an adequate policy environment for family planning could be rendered obsolete in the face of a burgeoning HIV/AIDS epidemic; lower prices of medications may provoke debate about offering anti-retroviral therapies, once out of reach because of cost. The chapter also includes an indicator of the policy planning process (also referred to as *policy formulation, policy reform*, or *policy change*).

Defining the boundaries or limits of policy is a challenge. Policy includes formal governmental declarations, laws, and statutes. Policy also covers operational regulations, guidelines, norms, and standards (Cross, Jewell, and Hardee, 2001). It is arguable whether *practice*, such as spending resources or following established norms, should fall within the boundary of policy. Practice may be considered an outcome of the policy environment rather than a component of the environment itself. This "boundary" issue is not unique to policy, but is manifest in almost every program process. For example, training programs seek to improve program performance; yet this *Compendium* considers training indicators separately from performance improvement indicators.

In addition to these definitional issues, program evaluators may want to consider other issues when they adapt the indicators to a specific application. Such issues are briefly summarized below.

Methodological Challenges of Evaluating Policy

- **Policy is difficult to quantify.**

With few exceptions (such as size of health budgets), policy indicators are inherently qualitative. That is not to say they are not objectively verifiable. Most

indicators use a nominal scale (e.g., presence/absence of a policy), and some indicators may be ordinal (e.g., higher or lower checklist ratings). Even when interval or ratio measures are theoretically possible (e.g., percentage of parliamentarians or of the general public in favor of a particular policy), policy program evaluation budgets are seldom large enough to include them on an on-going basis.[1]

- **Policies operate at different levels.**

Within the same country, policies can be enacted at different levels of the program and by different processes. The legislative branch of government and/or the executive branch enact most health policies, the legislative branch through acts of parliament or other laws, and the executive branch through presidential and ministerial decrees, departmental guidelines or norms. Similarly, this *Compendium* does not pre-judge the relative worth of executive vs. legislative policies – a national population law is considered the same regardless of whether it was passed by parliament or issued by the president's cabinet. Decentralization adds a further layer of complexity, when sub-national regions are granted policy and budgetary authority. As a rule of thumb, this *Compendium* defines a "policy" as any guideline or ruling that affects more than one service delivery post. For example, instituting user fees at a single clinic is not considered a policy, whereas using the results of a pilot test to set user fees throughout the entire system (be it a Ministry of Health or a family planning association) is considered a policy.

- **Policy change is usually incremental.**

A given policy is complete when it receives official approval at the highest level at which it was intended (e.g., a legislative action signed into law by the president, program protocols published by program director). This is often a multi-year process; adopting an all-or-none criterion (approved vs. not yet approved) may mask significant improvement in the policy environment. Those involved in field applications may find it useful to include progress toward approval (e.g., drafted, discussed in committee, submitted for approval, approved, revised, and updated) as part of the indicator's definition. In addition to whether a particular policy exists or has been recently adopted, evaluation of the policy environment should consider factors that improve the probability of its implementation, such as political and popular support, and sufficient resource allocations.

- **Several factors affect implementation of policy.**

Supportive policies improve programs and change reproductive health only to the extent that the policies are implemented. Most policy assessments include at least the *content* of the policy or policies (whether it guarantees access to a service, permits a variety of providers). A host of other factors within the policy environment influence policy implementation. These factors include the *actors* involved in the policy reform, the processes used to carry out the reform, and the *context* within which the policy was developed (Walt and Gilson, 1994). Political and popular support, participation, and the planning process itself should be included as policy indicators, because they affect both the likelihood of implementation and the process of policy formulation. Evaluating implementation would be a chapter in itself, because the indicators depend on the specific policy in question. For this reason, the *Compendium* does not specify indicators of implementation or compliance with the content of the policy. However, several indicators include a brief section on "implementation issues," which discusses how implementation evaluation might be addressed.

- **In some circumstances, policy may include unwritten rules of conduct.**

In many cases, unwritten practices govern provider behavior more than published policies or norms do. Do these unwritten rules qualify as "policy?" The answer to this question lies in part in the intervention chosen to remedy a situation. If the remedy chosen is to develop or modify a formal policy, then the unwritten rule of conduct should be considered a policy issue. If the remedy chosen is a different, non-policy programmatic action, then the unwritten rule should not be considered a policy issue. For example, if physicians refuse to offer hormonal contraceptives because they lack experience with modern low-dose formulations and the chosen remedy is contraceptive update training, then the unwritten rule of not refusing hormonals would *not* be a policy issue. However, if the remedy is to draft new norms stipulating that all clients, regardless of age or parity, should be offered hormonals, then the unwritten rule *would* qualify as a policy issue. Stigma and discrimination surrounding HIV/AIDS are another good

[1] See "Monitoring the Policy Reform Process," (USAID, 2000) for a discussion of quantitative, composite, and descriptive approaches to measuring policy.

case in point. By definition, discrimination appears in written legal documents, and as such, is a policy issue. Stigma, on the other hand, is an extra-legal concept dealing with attitudes and behaviors of individuals. If the response to stigma is to conduct public education campaigns on the rights of HIV-positive individuals, it is probably not a policy issue. If, however, the response to stigma is to pass new legislation stating that HIV status does not affect the rights of individuals to public goods and services, it would be a policy issue. In other words, whether or not unwritten rules of conduct qualify as "policy" depends on the nature of the *response* to those rules and not the rules in and of themselves.

In this section on policy indicators, we consider four broad aspects of the policy environment: (a) **formal policies**, (b) **resources and finance**, (c) **support for reproductive health and participation in the policy process**, and (d) **policy formulation**. Taken together, they span the dimensions of a *supportive policy environment*.

The indicators in this section are ordered according to the following four categories:

- Formal policies;
- Resource and finance;
- Support for RH and participation in the policy process; and
- The planning process.

Indicator

Existence of Policies, Plans, Guidelines that Promote Access to and/or Quality of RH Services

Definition

"Policies, plans and guidelines" includes broad health and population policies and laws. They also include programmatic and organizational documents whose objective is to regulate the kinds of services to be delivered, to whom, and under what conditions. They appear in constitutional provisions; legislation; implementing rules and regulations; executive orders; ministerial level decrees, and other measures of a regulatory nature (including related regulations and enforcement mechanisms); official goals and plan programs; statements and other formally documented government directives; standards; guidelines; and decrees (The EVALUATION Project, 1998).

Most developing countries now have some national RH policies or laws in place (although few have a stand-alone RH policy). Experience has shown, however, that macro-level policies, laws, councils, and programs do not guarantee RH service availability and quality. Therefore, we strongly recommend that any policy review include operational policies.

"Promote access" refers to mechanisms that encourage provision of RH services, and increase the number of service delivery points (SDPs) offering services and/or types of services and methods available.

"Promote quality" refers to mechanisms that encourage quality RH services such as technical competence of providers and responsiveness to client needs.

When evaluators measure both access and quality, they should construct separate indicators for each to maintain uni-dimensionality of each.

Not all individual policies will be complete. For example, a national development statement may cite reproductive health as a development issue, without detailing the steps necessary to improve RH. In such cases of incomplete policies, evaluators should consider the aggregate of all policies relating to RH, rather than examining individual documents. To measure changes over time, the indicator should consider only those policies developed or modified during a specific reference period, such as the last calendar year.

Data Requirements

Evidence of policies, plans and/or guidelines Supporting documentation should include the policy/plan/guideline itself, where or by whom it was issued or published, and an explanation of how the policy/plan/guideline promotes access to or quality of RH services. For example: Is support given to a full range of RH dimensions, or for only a single program element? Are all populations – women, men, youth – covered? Is accountability discussed?

At times, evaluators may wish to measure progress towards supportive policies. In this case, they can construct separate indicators for each stage of development (e.g., in draft, submitted for approval, approved), or can devise an ordinal rating scale to track progress from draft to final approval.

Data Source(s)

Actual policy/plan/guideline document with evidence of approval (or submission for approval). A content analysis of the documents should include level (e.g., national, provincial), topic area addressed (e.g., access, quality, FP, HIV/AIDS), and, where applicable, crosscutting issues (e.g., gender, human rights, youth).

Purpose and Issues

The purpose of this indicator is to measure the degree of explicit support for access to and/or quality of RH services on the part of government and other bodies, including service delivery institutions. However, presence or absence of policies alone is of limited usefulness. We encourage evaluators also to include the indicator, **Adequacy of the Policy Planning Process** (the last indicator in this section on the Policy Environment).

An important limitation of this indicator is related to the collection and analysis of all the relevant policy

documents. Evaluators may face difficulty defining "RH policy" within each country for a number of reasons:

- Following ICPD, the scope of RH significantly broadened to include safe motherhood and breastfeeding, aspects of sexual health such as female genital cutting, adolescents, HIV/AIDS, as well as family planning and population growth;
- Because of the extensive scope of policy they must explore, evaluators should carefully identify all RH-related policies. National development plans, poverty reduction plans, and other economic policies may include RH;
- Policy may be enacted by different processes in different countries and thus make cross-country comparisons difficult;
- Determining whether the policy/plan/guideline "promotes access and/or quality" is a qualitative assessment. Refer to Parts II.H.1 and II.H.2 of this *Compendium* for indicators of program access and quality; and
- Implementation of or adherence to policy directives is a separate issue.

In assessing implementation, evaluators must determine whether the health and population policies and laws include an implementation plan that designates institutional roles and responsibilities, time frames and activity plans, budgets, and monitoring and evaluation plans. When assessing implementation of broad policies and laws, evaluators should first consider whether or not an approved implementation plan exists (which will be an indicator in its own right), and then whether the activities have been conducted according to plan. Operational policies, on the other hand, are more specific; evaluation of their implementation should focus on whether their provisions are being carried out in practice.

Gender Implications of this Indicator

A gender perspective on plans and policies examines their content and their implementation strategies.

1. The Contents of the Plans
- Are the contents and language in line with major international agreements, such as ICPD, that focus on sexual and reproductive rights, not on demographic targets?
- Is the language gender sensitive (e.g., using "women" and "men" rather than "couple," which is gender insensitive; couples may or may not have common reproductive health goals or the barriers a "couple" faces may depend on whether the male member or the female member of the couple is seeking care)?

2. Implementation Strategies
- Do the strategies explicitly account for gender specific ways for women and men to access the care they need?
- Do the strategies exclude all elements of coercion or any such elements that act to disempower individual men or women (e.g., policies that give husbands control over the reproductive health of women)?

Indicator

Removal of Barriers to RH Policy Development and/or Service Delivery

Definition

This indicator constitutes a subset of the indicator, **Existence of Policies, Plans, Guidelines that Promote Access to and/or Quality of RH Services**. Whereas the previous indicator includes the broad rubric of policies, laws, and program documents that encourage provision and quality of RH services, this indicator focuses on modifying existing policies to eliminate obstacles to service access and quality.

This indicator is especially pertinent to legal and regulatory reform in RH. Policy barriers may affect participants in the policy process, service providers, and/or potential clients. They may affect both the public and private sectors (such as restrictions on particular contraceptive methods or eligibility requirements for RH services) or may affect primarily the private sector. Kenney (1993) distinguishes five categories of regulatory barriers:

- Regulations that constrain contraceptive options;
- Tax and import policies;
- Advertising and promotion regulations;
- Other regulations affecting the commercial sector; and
- Regulations affecting non-profit organizations.

Added to these are restrictions on access to training and exclusions from policy formulation meetings and other arenas in which policies are made.

Data Requirements

Old and new policy documents, showing evidence of restrictions in the old policy that do not appear in the new policy.

Data Source(s)

Legal and regulatory reviews; actual policy documents with evidence of government approval, submissions for approval

Purpose and Issues

The purpose of this indicator is to measure the extent to which national governments expand participation in developing policy and in providing RH services and facilitate increased access to RH services for all sectors of the population. Removing client eligibility requirements – such as marital status, minimum age, or parity for receiving FP methods or RH care – empowers women and youth to demand the services and products they want. Private sector participation in policy development may ensure that RH programs address the needs of all different groups in a population (e.g., women, men, commercial sex workers, men who have sex with men). The private sector can also be an important provider of RH services, especially in countries where government programs are either overburdened by demand or are unable to reach certain population groups.

This indicator can be quantified in several ways. As a baseline measure, it may be expressed as the number and type of policy barriers that significantly hinder private sector participation. To measure change over time in a country application, the evaluator should count and qualify the policy barriers identified at baseline, which were subsequently removed. Evaluators can measure change through naming and counting those identified policy barriers that do not appear in the new policy. Evaluators should link clearly the barriers identified at baseline, the policy interventions carried out, and the barriers identified at follow-up.

Because policy barriers by their nature tend to be very specific, evaluators can readily assess whether the new policy removes them. For example, if the barrier removed is import duties on contraceptives, evaluators can interview commercial distributors to determine if they no longer pay duties. Similarly, if the barrier is one that constrains contraceptive options – such as requiring parental consent to provide services to unmarried youth under age 18 – evaluators can interview providers to assess their awareness of the barrier removal and can interview youth to assess their ability to obtain services.

Gender Implications of this Indicator

A gender perspective on policy barriers examines the question:

- Do the plans recognize the common and different barriers women and men face in access to health care?

Indicator

Number of Policy Incentives to Increase Private Sector Participation in RH Service Delivery

Definition
This indicator constitutes a subset of the indicator, **Existence of Policies, Plans, Guidelines that Promote Access to and/or Quality of RH Services**. It focuses attention on the private sector.

Policy incentives refer to any course of action that facilitates private sector participation in providing RH services. Such incentives may include tax breaks for private sector organizations that provide RH services or for individuals who contribute to NGOs or mission hospitals providing RH, tariff relief, and public vouchers.

Data Requirements
Evidence of policies enacted that provide incentives

Data Source(s)
Actual policy documents with evidence of government approval, or submission for approval

Purpose and Issues
Governments can *hinder* private sector participation through the policy barriers described in the preceding indicator. The reverse does not hold – governments cannot *mandate* private providers to offer RH services.

The purpose of this indicator is to measure the extent to which governments facilitate the private sector's involvement in providing RH services. It may also indicate the relative importance governments place on the role the private sector can play in providing RH services.

Evaluators have limited experience in applying this indicator in developing countries. Tariff relief that exempts contraceptives from import duties is the most widely-practiced policy incentive to private sector service delivery. In South Korea, the family planning program at one time provided vouchers to reimburse private sector physicians for performing voluntary sterilizations and IUD insertions. Indonesia is testing a similar voucher system with private midwives, and Nicaragua has tested special vouchers for sex workers. Tax codes may offer deductions for charitable contributions to NGOs.

Policy incentives attempt to increase private sector participation. Evaluators must assess not only the presence of incentives (e.g., are vouchers available), but also their effectiveness (e.g., whether private practitioners are serving more clients than they did before receiving incentives).

Indicator

Resources Available for RH

Definition

"Resources available for RH" programs includes money, human resources, physical infrastructure, and material support. Resources may be expressed in monetary forms, such as local currency budgets or dollar conversions; other units, such as number of staff or staff time assigned to RH, are also possible. If used within a single country, the indicator can be expressed in terms of total resources. If cross-country comparisons are intended, the indicator should be expressed over a common denominator, such as resources per capita or RH resources as a percentage of the total budget.

Program administrators mobilize resources through four main sources: direct government (central or local) financing, donor financing (including bilateral, multilateral, and private foundations), user fees, and third-party payment mechanisms such as health insurance. In the face of declining government and donor funding for RH, new (alternative) financing mechanisms such as user fees and health insurance take on added importance. See **Number of New Financing Mechanisms Identified and Tested during a Reference Period**, (the next indicator in this section).

Data Requirements

Evidence of allocations to or expenditures on RH, by source of funding

Data Source(s)

National expenditure budget documents with evidence of approval; national accounts; invoices, and other evidence of expenditures; personnel or staff assignment rosters; time and motion studies

Other sources of information on national funding include the surveys commissioned by UNFPA and the Netherlands Interdisciplinary Demographic Institute (NIDI: www.nidi.nl/resflows/index.html), the UNAIDS/ Harvard University study on national expenditures on HIV/AIDS, and individual country studies of national expenditures and efforts to develop national health accounts.

Purpose and Issues

This indicator measures the commitment of resources by either a government, an NGO, or the private sector to the RH program. Evaluators must carefully define this indicator before they apply it to a country. First, they must define the realm of expenditures – does the evaluator refer to only public sector resources, or also to private expenditures on RH? Money paid out of pocket by individuals for their own care should not be included in this policy indicator, but expenditures made on their behalf by employers or insurers may be considered.

Second, evaluators must decide how to treat the source of public funds. For example, they may exclude donor grants but may include loan funds.

Third, evaluators may track separately capital expenditures (for new or renovated facilities, equipment) and recurrent expenditures for program operations (salaries, supplies, maintenance). Capital budgets may fluctuate widely from year to year, rising to cover construction of new facilities and falling when construction is complete. Thus, a decreased capital budget may not demonstrate or indicate a worsened policy environment. On the other hand, recurrent budgets should show at least maintenance or preferably steady increases over time, to cover growing populations and expanded and/or higher quality services.

Particularly in countries that provide FP/RH services along with other maternal child health (MCH) or primary health care services, evaluators may have trouble identifying and linking the line item in the budget of the appropriate ministry/organization to FP/RH. Moreover, when personnel provide other health services in addition to FP/RH, evaluators may have difficulty determining the proportion of time devoted to FP/RH.

In such cases, evaluators have the following options. First, the most commonly used, though least reliable approach, is to interview supervisors and health workers, asking them to estimate the percentage of their time

spent providing FP/RH services. This percentage can then serve as a basis for allocating labor and other joint costs.

Second, evaluators can conduct a time-use survey of a sample of facilities, using either the technique of patient-flow analysis or direct observation of health workers at specified intervals (i.e., work sampling). Bratt et al. (1999) showed that, compared to direct observation, neither self-reports nor patient-flow analysis reliably estimates allocation of staff time.

Third, another commonly used indicator of government resource commitment to RH is the share of the national budget allocated to family planning and reproductive health. The main problem with this alternative is that RH programs are often financed by several levels of government (e.g., national, state, local). Another problem is that such an indicator is sensitive to variations in the size of the national budget due to political, ideological, or national security considerations.

Fourth, some evaluators convert total expenditures to a per capita measure. This conversion permits cross-country comparisons and at the country level may complement, rather than replace, the total resources indicator.

Fifth, as a precursor to **Resources Available for RH**, evaluators may track, on an interim basis, newly enacted plans or policies (either at the government, organizational, or programmatic level) that attempt to increase resources for RH services. Examples include new, separate budget line items for RH services in national and local MOH budgets, or a directive that insurance plans must cover RH services. Planning to increase resources for RH services may signal an increased recognition of the importance of such services.

Finally, government can enhance resource adequacy by spending existing resources more efficiently.

An important question regarding implementation is whether funds or other resources allocated are actually expended to provide RH services. Many governments fall short of implementing their published budgets. When assessing implementation, evaluators must confirm that the resources allocated to RH programs actually flow to the operational units in the field providing the services. In practice, most evaluations will not be able to follow the money trail down to the operational level.

Indicator

Number of New Financing Mechanisms Identified and Tested

Definition
This indicator complements the indicator **Resources Available for RH**.

This indicator measures the "financing mechanism" – any process that raises funds for RH service provision. Examples of these mechanisms may include: fee for services, sliding fee scales, subsidized services through donor financing, and third-party payment mechanisms such as health insurance.

"Identified and tested" refers to actions that assess the feasibility and appropriateness of certain funding mechanisms for providing RH services. To meet this indicator, a country or program must both identify and test a new financing mechanism.

Data Requirements
Information on type of financing mechanisms identified and/or tested

Data Source
Documents and meeting minutes; pilot tests; study results

Purpose and Issues
Funds for reproductive health services can be mobilized through four main sources: direct government (central or local) financing, donor financing, user fees, and third-party payment mechanisms such as health insurance. This indicator highlights the importance of financial resource mobilization as an essential component of a national plan or policy. Its purpose is to measure the extent to which governments and local NGOs initiate and experiment with different strategies aimed at increasing access to RH services.

Not all new financing mechanisms are necessarily good. Adding a new mechanism like fee for service can be good if it increases available resources for RH, or bad if it suppresses demand. Often economic barriers, such as high fees for services or high transportation costs, restrict access to health services. On the other hand, charging nominal fees for certain reproductive health services may increase demand for such services, because people may associate better quality of services or a greater need for those services with having to pay for them.

In terms of implementation, evaluators will need to distinguish between the testing of a new mechanism and the mechanism's success at increasing revenues without unduly depressing demand. Organizational willingness to test a variety of financing mechanisms signals a positive policy environment, even if the organization ultimately adopts only one or two of the mechanisms.

Indicator

Political and Popular Support for RH

Definition
"Political support" refers to the positions taken by government leaders on RH, both in public meetings and in closed policy deliberations. Support may be manifest in public speeches, voting records, or behind-the-scenes lobbying. Popular support refers to the positions taken by members of the civil society and is measured by civic involvement (e.g., NGOs, media, religious and community organization leaders).

Data Requirements
To distinguish this indicator from the previous indicator, **Existence of Policies, Plans, Guidelines**, evaluators should not gauge political support for RH from official documents such as national development plans. To construct this indicator, an evaluator requires prior definition of the reference group (e.g., which government or civic leaders, stratum of the public at large) and the means to assess the group's opinions. Evidence of oral or written statements or public forums should include dates, position/responsibilities of the person(s) involved, intended audience, and media employed.

Data Sources(s)
To construct this indicator, evaluators may draw upon a variety of data sources, ranging from voting records, quantitative opinion polls of defined leadership groups (e.g., parliamentarians) or of the general public, to key informant interviews. Media scans that archive texts (or audio or video tapes) of official speeches, newspaper articles, government communiqués, official documents, or other public expressions may be available but are difficult to implement and interpret. Evaluators should avoid anecdotal evidence or non-systematic clipping services, especially if they intend to measure change over time.

Purpose and Issues
The purpose of this indicator is to measure the degree of explicit political and popular support for RH services on the part of government, civic leaders, and society at large.

This indicator is basically qualitative, and as such, many consider it fairly "soft," both difficult to quantify and to interpret. Even if one assiduously follows all official speeches and documents of numerous high level decision-makers, assessing such statements is difficult. For example, should reading a prepared speech on World Population Day or at the opening session of a donor-sponsored event be taken as support for RH? Moreover, a single statement by the president of the country may carry more weight than 100 statements by lower level officials.

The impact of political and civil society support is greater if statements continue over a long period of time and if successive leaders make much the same commitment to RH. If, on the other hand, successive leaders vacillate between strong support and weak or no support, the policy impact of such statements may be minimal.

Opinion polls or key informant interviews are the preferred data sources, providing that leaders are willing to be interviewed. These may include parliamentarians or public or opinion leaders previously singled out for an advocacy campaign. Evaluators can measure popular support with opinion polls, readily conducted in countries with "omnibus" marketing surveys.

Experience has shown that media scans of public statements are difficult to maintain and to interpret.

Despite the inherent difficulties in data collection and interpretation, this indicator is one of the few available markers of progress in advocacy for policy change. Opinions of public officials may constitute the earliest signals of impending change in the government's position on RH, whereas civil society support may become increasingly important as governments open the policy process to wider participation. Once governments enact favorable policies or budget resources to RH, continued political and popular support is vital to ensure program implementation.

Evaluators should use this indicator in conjunction with **Existence of Policies, Plans, Guidelines that Promote Access to and/or Quality of RH Services** and **Resources Available for RH**. Support manifest in such concrete actions as constituting a new organizational unit or program to oversee HIV/AIDS or funding a leadership position can be legitimately considered a new policy (such as the one creating the HIV/AIDS body) and increased resources to RH (such as the budget allocated to the HIV/AIDS unit and the official's salary).

Evaluators have used this indicator as a component of several global program assessment measures, such as the Family Planning Effort measure, the POLICY I Project Policy Environment Score (PES), and the UNAIDS AIDS Program Effort Index (API) [Stover, Schwartlander, and Roehnstrom, 2000; Ross and Stover, 2001]. For example, the Lapham/Mauldin /Ross Family Planning Effort measure includes *favorable statements by leaders*: "Whether the head of government speaks publicly and favorably about family planning at least once or twice a year, and whether other high officials also do so." Similarly, the API (presented in Part III.C) includes eight items, rated from 0 to 5, under the dimension of "Political Support." Critics have cited some analysis and interpretation issues, such as treatment of inter-rater differences and measurement of change over time (e.g., comparison of ratings made at different points in time vs. retrospective ratings).

Indicator

Participation in RH Policymaking

Definition
This qualitative indicator measures the *number and breadth* of different governmental, nongovernmental, commercial sector, religious and/or community organizations that participate in the policymaking process, and the *nature of their participation*. Mechanisms may include public hearings, multisectoral boards or consultative committees, and appointment of civil society representatives to official decision-making bodies (USAID, 1998).

Data Requirements
Evidence of individuals and agencies involved in the policy process with information on the nature of their involvement

Data Source(s)
Meeting agendas and attendance lists; focus group discussions; and/or key informant interviews

Purpose and Issues
This new indicator for RH reflects experience over the last decade – from ICPD to decentralization to mobilizing national responses to HIV/AIDS – which amply demonstrates that RH policy transcends central decision makers and even the health sector itself. Responding to these broader needs requires that governments "open up" the policy development process to stakeholders traditionally excluded from decision-making, including them as active participants rather than as passive recipients of decisions made on their behalf.

This indicator explicitly recognizes the links between RH policy and larger issues of democratic governance. From the perspective of governance, widespread participation in the policymaking process is seen as a good in and of itself. However, participation *per se* does not guarantee that resulting policies are better than those enacted through a closed process, although this is a testable hypothesis.

For participation to occur, public institutions must be open to wide involvement in all phases of the policy process, including formulation, implementation, and oversight. For this involvement to occur, mechanisms must exist for the exchange of information and views on the key issues (USAID, 1998).

The purpose of this indicator is to measure the degree to which different organizations (public, private, community, religious, among others) are involved in the RH policymaking process. This indicator relies on the assumption that the greater the number and the more varied the type of organizations involved, and the greater the opportunity for their substantive input, the more that policy will reflect the population's needs.

A limitation of this indicator is its multidimensional quality – including numbers of different actors, breadth of organizational representation, and degree of involvement. Evaluators may have particular difficulty ascertaining the level of "involvement" by different actors. Data collection should solicit information on degree of participation in the process of formulating this policy, involvement in work or discussions leading up to drafting the document**,** and input before the draft document was prepared.

Given the breadth of this indicator, most applications will concentrate on a single factor, such as the number of institutional participants or number of different sectors participating. Evaluators can ask participants to rate their involvement (e.g., on a scale from actively engaged in problem definition and policy formulation, to simply being invited to a policy dissemination seminar), or the degree to which they felt that their opinions were requested and taken into account. Evaluators can design a composite descriptive measure combining all three dimensions (number of actors, breadth of representation, degree of involvement), and can then track both the component profile and the composite score.

Gender Implications of this Indicator

A gender perspective on participation examines the process for developing the plans:

- What was the percentage of women and men who helped to write the initial draft? What organizations or strata of society did they represent?
- Did the intended beneficiaries, including women's organizations, review the plans?

Indicator

Number of NGO Networks or Coalitions Working for RH

Definition
This indicator is based on the premise that there is greater power in numbers. In other words, the more organizations that come together and speak with a joint voice for RH, the more effectively they can present their message. The greater their institutional stability, the more likely that they will be heard and will be effective advocates for RH issues. Evaluators may use this indicator in conjunction with **Political and Popular Support for RH**.

In this section, "network" and "coalition" are used interchangeably to refer to groups of organizations and individuals working together to achieve changes in a policy, law, or programs for a particular issue (POLICY, 1999).

The indicator assesses the status of NGO networks or coalitions that work in support of RH. Three parameters are included – formation, expansion, and strengthening. Depending on baseline conditions, any or all of these may constitute separate indicators.

- Number of NGO networks or coalitions that meet regularly and work in support of RH;
- Number of member organizations and/or individuals belonging to NGO networks or coalitions; and
- Sustainability of NGO networks or coalitions working in support of RH.

Data Requirements
Evidence of network status and functioning, based on pre-determined criteria (number of members, activities, degree of sustainability)

Data Source(s)
Membership lists of networks or coalitions; management and/or financial information systems; meeting minutes; external assessments of sustainability

Purpose and Issues
Democratic governance implies popular participation, including participation by disadvantaged social groups, in both public policymaking and its implementation. By promoting and protecting civil rights, civil society organizations (CSOs) ensure that citizens have the means to express their preferences, engage in dialogue with policy makers, and affect public policy decisions. After governments establish policies, CSOs perform as watchdogs of state performance by demanding accountability in the allocation and management of public resources.

For CSOs to intervene effectively in the policy process, they must gain or strengthen their own advocacy skills. Such skills run the gamut from simply collecting information on the subject at hand to such other tasks as obtaining or allocating human and fiscal resources to advocacy functions, building coalitions and networks, acting to influence policy, and monitoring implementation once a policy decision has been made.

Indicator

Adequacy of the Policy Planning Process

Definition
This indicator measures the process through which policies, plans, guidelines or programs were formulated, developed, or reformed, independent of the documents themselves. It uses a rating scale that brings together a number of parameters into a checklist and is adaptable to individual country situations (POLICY, 2001).

Evaluators assess the adequacy of the policy planning process based on three criteria:

1. Who participates in the planning process; does it
   - Materially involve representative(s) from multiple sectors – public, commercial, and NGO sectors; donors; community or grassroots leaders; special interest groups (e.g., youth, women, and human rights organizations).

2. Do policy makers use empirical information in the planning process; specifically, do they
   - Include information-based needs assessment;
   - Identify and prioritize problems;
   - Consider alternative strategies for addressing identified problems; and
   - Formulate strategies for implementation.

3. Evaluators can readily assess whether the new policy removes the barriers. Does the policy document specify the operational aspects of the program, in that it
   - Includes development of a detailed action plan (inclusive of roles and responsibilities);
   - Includes assessment of resource needs and availability (financial, human, materials);
   - Is medium to long-term (at least one year); and
   - Establishes monitoring and evaluation procedures.

Data Requirements
Results of the Planning Checklist questionnaire or similar qualitative assessment

Data Source(s)
Rating forms or planning checklists (such as the one shown in Table II.B.1). Evaluators collect information from key informant interviews; if the organization involved kept written records, evaluators may review minutes or proceedings.

Purpose and Issues
Planning includes coordinating all aspects of RH policy or program development. This indicator encompasses policy processes and decision-making in both public and private sectors. Broad buy-in and ownership may be as important for program success as the technical bases upon which decisions are made. Therefore, the first set of criteria to ensure adequate planning involves participation in the planning process. The second set of criteria judges the extent to which policy makers use empirical information to understand the RH needs of the population and trade-offs among potential interventions. The third set of criteria addresses operational aspects of the program, from staffing and activity plans to monitoring and evaluation.

The illustrative checklist identifies basic tenets of good planning as presented in the literature on strategic planning. Evaluators can use the checklist to establish a priori criteria to fit the particular aspects of planning that the project is trying to improve.

This indicator is still "experimental" in that evaluators have not routinely used it in policy work. However, to the extent that assistance includes improvement of the planning process as well as the content of the actual policies and plans, it provides an interim measure of progress prior to adoption of the final plan.

Gender Implications of this Indicator:

A gender perspective on the planning process examines the process for developing the plans.

- What was the percentage of women and men who helped to write the initial draft? What organizations and/or strata of society did they represent?
- Did the intended beneficiaries, including women's organizations, review it?

Table II.B.1 Planning Checklist

Persons involved in the planning process – relevant staff, external advisors (consultants), and other relevant stakeholders – use this type of checklist as a baseline assessment to identify current deficiencies with the planning process. In the process of the assessment, this group may identify and may add additional areas that will "improve planning." At the conclusion of the planning process, or in conjunction with the completion of a plan document, this same group of staff, advisors, and other relevant stakeholders will again complete the checklist. A comparison of the checklist results with the baseline assessment will identify specific areas of improvement. To claim "improved planning" results, the country manager/director will write up a description of how planning had improved and submit it along with the baseline assessment, the completed checklist, and a copy of the plan document produced.

Country Name: Date:

Describe nature of plan being developed:

The planning process:

- ❑ Materially involved representative(s) from [number] of the following: public, commercial, and NGO sectors; donors; community/grassroots leaders; special interest groups (youth, women, and human rights organizations)
- ❑ Included an information-based needs assessment
- ❑ Identified and prioritized problems
- ❑ Considered alternative strategies for addressing identified problems
- ❑ Formulated strategies for implementation
- ❑ Included development of a detailed action plan (inclusive of roles and responsibilities)
- ❑ Included assessment of resource needs and availability

  Financial: (specify) _____

  Human: (specify) _____

  Materials: (specify) _____
- ❑ Included medium- to long-term objectives
- ❑ Established monitoring and evaluation procedures

  Name of person:

Part II. C
Management

- Availability of a clear, strategic mission statement
- Responsiveness of strategy
- Capacity to reach annual objectives
- Availability of logical and explicit organizational structure
- Number/proportion of employees whose performance has been reviewed according to performance management standards
- Percent of key positions filled
- Staff turnover rate
- Availability and use of a coherent planning system
- Quality of strategic and operational plans
- Number/proportion of organization/program units systematically using information to monitor performance
- Number/proportion of reporting units submitting a completed routine MIS report on time
- Percent of data elements reported accurately in MIS reports
- Institutionalization of a system of quality assurance
- Availability of budget(s) linked to operational plan(s)
- Effectiveness of financial management systems
- Percent of annual revenue generated from diverse sources

Part II.C

MANAGEMENT

Good leadership and management are essential to organizational development, performance, and sustainability. An organization succeeds because of *what it does* (a shared commitment to accomplish something useful and important) and *how it does it* (the way it functions, decides, evaluates, adapts, and delegates). Good leadership and management are likewise essential to RH programs and to the achievement of national RH goals. Effective and efficient organizational performance is critical in the face of current public health challenges (e.g., AIDS, tuberculosis), health-sector reform (e.g., decentralization, integration, financing), and the needs and desires of an ever-expanding number of clients of reproductive health services from traditionally underserved groups (e.g., adolescents, rural and indigenous populations, men). The objective of management and leadership is to foster and sustain individual employee and organizational performance as well as overall program performance in delivering RH services.[1]

Effective management and leadership consist of a series of essential functions, underlying dimensions, or elements. The functions of managing and leading are carried out *simultaneously* on a daily basis by *multiple* individuals at *multiple* levels of an organization, not just by senior staff.[2]

In many instances, the "boundaries" between the management and leadership functional areas and other functional areas of RH service delivery discussed in Part II of this *Compendium* overlap, because the management or leadership function spans each of the other areas. In other words, an organization's or program's success (or lack thereof) in performance in all of its areas of focus is at least in part attributable to a strength or weakness in its management and leadership. For example, performance in BCC or commodities and logistics management is, in many respects, determined by managers' recognizing the importance of the system and allocating resources to its implementation.

The factors determining how an organization does its work or accomplishes its objectives include: the effectiveness and functioning of individuals at all levels of the organization, the management systems supporting their work, the organizational culture, and the adequacy of human and financial resources.

Ideally, one should evaluate overall management and leadership capacity by taking a long-range view of the organization and the way it develops over time. As organizations grow, they evolve along a management development continuum characterized by four distinct developmental stages.[3] At the first stage, an organization begins to develop a particular management component. By the fourth stage, an organization operates extremely effectively with regard to the management component.[4] Organizations (and the programs they support) pass through these stages at different rates, and evolve to the point that they have a clear mission, strong management structures and systems, and skilled leaders, managers, and staff who can effectively use these structures and systems.

Management systems or capabilities within a single organization will often be in different stages of development because some management systems will receive more attention than will others as the

[1] Hence, this section of the *Compendium* focuses upon organizations that implement programs, rather than upon programs themselves. Thus "organization/program" is often used to signify that the indicator applies to either organizations or programs.

[2] MSH has recently developed a framework describing the essential functions of leading and managing at all levels of an organization (2001).

[3] In the Organization Development (OD) literature, the four stages are referred to as: 1) emergent/initial; 2) launch; 3) consolidation; and 4) mature.

[4] The materials provided in this section draw primarily from Management Sciences for Health and the work of its past Family Planning Management Training and Family Planning Management Development Projects (1985-1990, 1990-1995, 1995-2000, respectively) as well as the current Management and Leadership Program (2000-2005).

organization develops. For example, donors and the organization itself often focus on ensuring that a sound financial management system and practices are in place before they focus on developing an organization's human resource management system. In such a case, the financial management system may be in a more advanced stage of development than is the human resource management system.

Organizational performance – which refers to what an organization does and how it does it – always includes some element of customer satisfaction. One evaluates what an organization does in relation to the goals and objectives it has established, such as expanding access to services to reach a diverse segment of the population, improving quality, improving client satisfaction, increasing or diversifying sources of revenue, reducing costs, or influencing national RH policy. Evaluators should therefore define measures or indicators in relation to the specific long- and short-range objectives set by the organization, many of which are presented in other sections of the *Compendium*.

There are two distinct approaches to measuring management within RH programs. The first approach is to use a standard set of criteria based on national or international norms and standards. Some organizations in developing countries seek certification, for example, from the International Organization for Standardization (ISO), whereas others seek accreditation based on national standards. The second approach is to develop indicators for each management area and component in collaboration with the organization as part of an exercise to review and strengthen its management systems. This approach involves an assessment to determine the baseline stage of development of the organization. In this way, program managers and evaluators can tailor the standards against which organizational performance is being measured to the level of the organization's development as well as the specific context in which the organization functions and offers RH services. The indicators in this chapter of the *Compendium* are derived from actual experience in designing assessments of RH programs in developing countries rather than from rigid standards that are applied to health care organizations.[5]

Methodological Challenges of Evaluating Management and Leadership

- **Self-assessment of performance generates ownership, but it may lack objectivity.**

Organizations may assess their own performance in a participatory fashion, involving staff from all levels, including board members, with the assistance of an external facilitator or a highly skilled internal facilitator. Self-assessment is generally less costly and less time-consuming than external assessments. More importantly, it allows for organization-wide ownership of the findings and commitment to action, because it builds upon existing strengths and addresses major weaknesses acknowledged by the staff.

However, self-assessment can be subjective, unless the internal evaluator obtains hard data to verify the findings as part of the assessment process. Moreover, self-assessment requires careful and thorough consensus building among staff involved. External assessment, on the other hand, may be more objective, because it relies heavily on the review of organizational documents in addition to interviews with staff at all levels of the organizations. An external assessment, however, is likely to be time and resource intensive. Also, unless the assessment is carefully designed and implemented, the organization may not accept the findings or act on them.

Throughout this section of the *Compendium*, the term "evaluator" should be interpreted broadly to include both organizational staff responsible for internal assessments (also referred to as facilitators) as well as persons external to the organization.

- **Evidence of the causal link between management and outcomes remains elusive.**

Researchers have found it difficult to find conclusive evidence of a **causal link** between the indicators of program management/leadership and actual performance in developing country settings. For example, research from developing countries has yet to

[5] See, for example, Baldridge, *Health Care Criteria for Performance Excellence,* 2001.

provide conclusive evidence that human resource interventions (e.g., leadership training) result in greater achievement of organizational objectives. The many intervening variables between human development and end results (e.g., shifts in internal and external conditions, market shifts) make these links difficult to validate. Links between management interventions and organizational outcomes are not clear-cut in research in the public sector, particularly in health care. James Buchan, a researcher in health care human resources from the Queen Margaret University College in the U.K. states, "Evidence based on human resources is mainly U.S. based, but few studies give details of evaluation of quality/outcomes and/or costs…[Studies are] mainly descriptive, weak on methodology and not transferable or generalizable" (Buchan, 2001).

Organization of this Section

This section is organized in a format slightly different from that of others in the *Compendium*. Specifically, it presents key areas of management and leadership. The indicators describe attributes of a well-functioning system, be it in public, private, or NGO sector organizations. The management indicators generally measure progress in the use of a system, the effectiveness of individuals using the system, or the performance of the system. However, some indicators are expressed as the presence or absence of a key component of a system or function itself.

For example, an organization's revenue generation system may be functioning perfectly well when measured against the stated components of a good revenue generation system, yet the organization may still struggle with sustainability if the policy or economic environment is particularly unfavorable for generating new sources of income. Likewise, an organization may have a very strong human resources management system but have high turnover due to factors in the environment that are completely beyond its control.

The indicators included in this section focus on four broad management elements found in all organizations: mission, strategy, structure, and systems (MSH, 1999a).

- **Mission**

The mission is a statement of purpose that frames the values guiding the organization or program and that provides consistency and meaning to its actions. The mission statement answers the question, *"What does the organization do and why?"*

- **Strategy**

Organizational strategies are the approaches that organizations and programs select to define the activities enabling the fulfillment of the mission. Strategy answers the question, *"How will the organization get to where it wants to go?"* Organizational strategies should help prioritize and focus the organization's work, should comply with its mission, and should respond to the demands of the clients served and to the organization's potential market (clients it has the desire to reach).

- **Structure**

Structure addresses roles and responsibilities, lines of authority, and distribution of responsibility, in alignment with mission and strategy. The structure answers the question, *"What is the framework and decision-making structure within which the organization operates?"* A well-defined structure encourages individual and team/group initiative and provides staff clarity in terms of the decision-making authority.

- **Systems**

Systems are the interdependent management areas within an organization that allow it to do its work. Organizational systems answer the question, *"How does the organization carry out its activities?"* Key systems for health organizations and programs include human resource management, organizational planning, information systems, quality assurance, financial management, management of revenues, and management of supplies. The key systems for health organizations and programs are outlined in Box II.C.1.

Box II.C.1 Key Systems for Health Organizations and Programs

- **Human Resource Management**

Human resource management (HRM) is the integrated use of systems, policies, and management practices to recruit, maintain, and develop employees in order to meet the desired goals of the organization. Effective human resource management supports employees in carrying out meaningful and satisfying work as well as help an organization to improve its level of performance and impact (FPMD, 1999).[6]

- **Planning**

Planning is a systematic process to review, modify, and align key elements of the organization's mission, strategy, structure, systems, and program activities in light of changing internal or external conditions. Planning is an essential component in ensuring sustainability. Planning systems cover both annual and long-term planning that further the organization's mission, strategy, and goals/objectives.[7]

- **Management Information Systems**

A management information system is a set of components and procedures organized with the objective of generating information that will improve health care management decisions at all levels of the health system or organization. The MIS represents a key source of data for calculating indicators on service utilization, coverage, and overall performance; a comprehensive MIS includes distinct subsystems for management areas, such as human resources, logistics management, disease surveillance, finances, and workplan monitoring.

- **Quality Assurance**

A quality assurance (QA) system monitors and improves service effectiveness and client satisfaction. Because QA emphasizes a process of constant improvement in operations, it requires long-term organizational commitment and teamwork (FPMD, 1993). Many elements of a high-functioning quality assurance system are embedded in other management systems.

- **Financial Management**

The financial management system collects, records, and reports data on an organization's financial situation. It provides information that helps finance, program, and senior managers to make decisions about allocating resources.

- **Management of Revenues**

Management of revenues means planning for revenue generation and diversification of funding sources through constituency building among clients and donors (current and potential). Organizations tend to have greater success in this area if they have multiple funding sources (e.g., donors, government, third party payments, local community, sale of services/products).

- **Management of Supplies (Logistics)**

Without a functioning supply management system, an organization cannot deliver quality reproductive health services to its clients. Part II.E of this *Compendium* on Commodities and Logistics provides a comprehensive list of indicators for logistics management. Readers are also referred to Chapter 36 of *Managing Drug Supply* (MSH, 1999b) for additional indicators of quality of clinical care, such as average number of drugs per patient, percentage of patients receiving antibiotics, percentage of providers prescribing by generic name.

[6] Some indicators listed under HRM also pertain to Performance Improvement (PI) in this *Compendium* and thus may be found there (see Appendix C).

[7] Capacity in financial planning is addressed under the indicator **Effectiveness of Financial Management Systems.**

Indicator

Availability of a Clear, Strategic Mission Statement

Definition
The existence of a mission statement, which is a written expression of purpose – the overall reason an organization exists.

A mission statement is clear and strategic if it:

- Defines the program or organization's purpose;
- Defines the program or organization's intended clients;
- Is used to guide strategic planning; and
- Is reviewed periodically to assure that it provides a vision for the future and continues to reflect fundamental values and goals.

Data Requirements
Evidence of a written mission statement; other information on the organization's mission

Data Source(s)
Organizational/program documents including plans, staff orientation materials, policy manuals and statements, and marketing materials

Purpose and Issues
This indicator provides a summary measure of the existence of a clear sense of direction in the form of a written mission statement. Having a clear mission statement that is well disseminated within an organization facilitates the setting of strategy and setting of priorities based on the needs of current and potential clients. Although programs and organizations without a formal mission statement may perform in a highly effective manner, a mission statement serves the important function of helping to keep staff focused on the accomplishment of long-term objectives.

Although the basic mission of an organization may remain the same for an extended period, the language of the mission statement must provide a clear focus and a priority for organizational strategies and activities.

This indicator is quantifiable on a scale of 0 to 5, as indicated below.

0	No written mission statement
1	Mission statement exists but fails to conform to the above criteria
2	Mission statement exists and conforms with only one of the above criteria
3	Mission statement exists and conforms with 2 of the above criteria
4	Mission statement exists and conforms with 3 of the above criteria
5	Mission statement exists and conforms with all 4 of the above criteria

Indicator

Responsiveness of Strategy

Definition
The fit between an organization's strategy and its mission and stakeholder needs

A strategy is the long-term plan according to which an organization aims to reach its goals and objectives through a series of activities, inputs, and results. "Responsive" means that the strategy fits the organization's mission and the needs of clients (existing and potential), within its policy and service-delivery environment.

Data Requirements
Evidence of a written strategy; evidence of the quality of the strategy (coherence with mission and results of client-based needs assessment); evidence of analysis of market conditions and needs

Data Source(s)
Review of strategic plan or strategy documents, as well as needs assessment and competitor analysis; interviews with key staff (e.g., managers)

Note: the box to the right indicates the criteria to be used in calculating a score on this indicator.

Purpose and Issues
An organizational strategy should prioritize and focus the work, should comply with the mission, and should respond to the demands of the clients served and the organization's market. A clear strategy continuously and clearly defines the logic leading from activity to outcomes.

Moreover, the strategy must include evidence of linkage and responsiveness to its clients, community, funders, and potential new markets. The following 4-point scale has proven useful in measuring this indicator in reproductive health organizations and programs.[8]

Descriptor	Score
Organizational strategies are formulated with little concern for the perspectives of clients and the demands of the market (the wider community).	1
Client and community perspectives are discussed in formulating organizational strategies, but there is no systematic assessment of these factors (e.g., no market studies, no client interviews). There is no mechanism for involving community/clients in formulating strategies. There is no analysis of competing services.	2
Client needs and desires have been assessed, and markets for expanded and targeted services and products with the community have been defined. These single assessments are used repeatedly over time to guide the development of strategies. Community/clients are only sporadically involved in formulating organizational strategies. Analysis of competing services is carried out sporadically.	3
The needs and desires of clients and the demands of the community are frequently reassessed to identify changes over time and to provide the basis for developing organizational strategies. Clients and community are systematically involved in formulating organizational strategies. A mechanism is in place for regularly analyzing competing services.	4

[8] See Management and Organizational Sustainability Tool (MOST), MSH 1999a.

Indicator

Capacity to Reach Annual Objectives

Definition
Programs or organizations set annual planning objectives at the beginning of the annual planning period and document them in an annual operational plan. Annual objectives should explicitly link to the broader strategic objectives and be achievable within the annual framework. Annual objectives can include quantifiable outputs (e.g., number of HIV clients counseled) to higher-level outcomes involving behavioral or procedural change (e.g., new ways of tapping into funding sources or improved monitoring techniques).

The facilitator (evaluator) assigns a score from 0-3 for each objective, as follows.

Objective	Score
There are no stated objectives	0
Objectives partially met	1
Objectives met	2
Objectives exceeded	3

This indicator is calculated as the mean score across all three objectives.

Data Requirements
List of all annual planning objectives; individual scores for each objective

Data Source(s)
Annual planning document; organizational strategy or other multi-annual planning document; semi-annual and/or annual progress reports (e.g., to stakeholders); assessment by external evaluator or internal facilitator

Purpose and Issues
This indicator provides a "results-oriented" measure of program or organizational planning performance. The indicator is based on the premise that the overarching measure of good management and leadership is whether or not the organization achieves its stated objectives.

The recommended measure for this indicator can only provide a general picture of overall functioning of the organization; the evaluator will require further in-depth analysis of the reasons for failing to obtain stated objectives.

Indicator

Availability of Logical and Explicit Organizational Structure

Definition
The structure of an organization refers to the staffing and decision-making framework that assigns personnel according to their authority and level of responsibility. The structure is clearly articulated when it provides clear lines of authority and accountability, distribution of responsibilities, and lines of communication.

Data Requirements
Evidence that relationships, supervision, roles, and responsibilities have been formally defined; description of how decisions are made in the organization

Data Source(s)
Organizational chart; written job descriptions; policy manual(s); interviews with staff at all organizational levels

Purpose and Issues
This indicator measures whether an organization has a clearly defined structure – roles, responsibilities, and authority – both "on paper" and in practice. Evaluators can find evidence by reviewing whether the organization has:

- An organizational chart (organigram) illustrating authority and communication lines;
- A policy manual clearly defining roles and responsibilities for staff (and board members) as well as the formal system of delegation;
- Job descriptions detailing responsibilities and supervisory lines; and
- A written, defined process to review the structure periodically to ensure consistency with the organization's current strategies.

Measuring these aspects requires developing a simple scale from 0 to 4. A program without any of the four documents listed above receives the lowest score (0), while one with all elements receives the highest score (4).

Organizational decision-making processes may or may not conform to those embodied in the formally defined organizational structure. To make this determination, one can assess the appropriateness of staff for the positions they fill and the extent to which they actually make the decisions called for in the "formal" document (e.g., by interviewing staff to determine how one or more recent important decisions were made).

Assessment for the indicator will tend to be more subjective than for the previous one, which involved simply measuring the existence of documents. However, evaluators can apply the same type of scale as in the previous indicator (ranging from no conformity to full conformity).

The validity of the indicator is based on the assumption that a clear definition of roles, responsibilities, and decision-making in an organization promotes strategic and operational decision–making that optimizes the use of available resources. The organizational structure must be (at least partially) amenable to change as directed by the organization's management/leadership.

Gender Implications of this Indicator

A logical and explicit organizational structure with clear lines of authority and accountability is an important indicator of management capacity. In many organizations, women are plentiful at lower levels of responsibility, but do not reach the top management positions. Many factors contribute to women's ability to rise to management positions in an organization, including their lower levels of school enrollment and literacy. Gender discrimination in the organization can be said to exist when women have the necessary educational qualifications and experience to compete for a particular job but are denied access on the basis of sex. Job descriptions and qualifications can also be written to exclude women by requiring skills or experience that women have no means to gain access to. Logical and explicit structure should include commitment to hiring and promotion on the basis of appropriately identified qualifications.

Indicator

Number/Proportion of Employees Whose Performance Has Been Reviewed According to Performance Management Standards

Definition
The organization's adherence to standards when it supervises and reviews employee performance

Performance review, an assessment of the employee's performance by the supervisor and employee, is ideally based on jointly established work plans, performance objectives, and results related to expectations. The review is the cornerstone of the supervision process and an important element in overall performance management. Performance management is defined as the systems, policies, and procedures used by an organization to define and monitor the work that employees do and to ensure that the tasks and priorities of employees are consistent with the strategy of the organization. Performance reviews address the need of all staff for clear expectations of their work.

Data Requirements
Description of supervision and performance review standards; evidence that supervision and performance reviews adhere to standards

Data Source(s)
Personnel files

Purpose and Issues
Supervision and assessment of employee performance are critical functions of human resource management. Measuring adherence to standards assumes that the organization has written standards defining the requirements for a complete and high-quality review, as well as defining its frequency.

To measure the indicator, an evaluator (external to the organization or the Director of Human Resources for an organization) first reviews written human resources policy documents and establishes the standards for conducting supervision, the frequency of formal performance reviews, and the requirement for written documentation of the review. He/she then reviews a sample of personnel files (or all files if resources are available or the number of employees is small) and checks the documentation to see if supervision/performance review adheres to standards of timeliness, completeness, and accessibility. The evaluator chooses a time period to review depending on the frequency that reviews are supposed to take place (e.g., in the last 12 months).

Standards for completion vary greatly by organization/program. For example, some organizations may require signatures by both the staff member and his/her supervisor in the instance of a joint performance review. In other instances, where a supervision checklist appears in "audit" form, only the supervisor may be required to sign the completed document. Some organizations have the additional requirement of documentation of salary changes/promotion; others require a written performance plan for the next period of review (e.g., 12 months).

The indicator assumes a well-documented supervision/performance review system is in place.

Indicator

Percent of Key Positions Filled

Definition
The effectiveness of the organization's human resource management system to fill key positions

The definition of "key" varies considerably by organization. For health service delivery organizations, clinical standards and organizational norms will dictate the definitions of key staff at the clinic level. The total number of required key staff is often determined as a ratio of physicians, nurses, auxiliary nurses, midwives, and other clinical staff to the catchment population.[9] At the headquarters level of an organization, "key" may include department or unit heads, and critical technical support staff.

Data Requirements
Accurate count of key positions in the organization that are actually filled at the period under review and total number of key positions available in the organization

Data Source(s)
Personnel management information systems; personnel records; organization's organigram; facility survey; accurate payroll system; personnel policy manual

Purpose and Issues
The purpose of the indicator is to measure the effectiveness of the organization's human resource management system to fill key positions and thus to ensure organizational capacity to perform and achieve its objectives. The indicator measures actual performance against the ideal scenario in which all key positions are filled. Most reproductive health organizations spend the majority of their resources on staff; an incomplete complement of key staff can compromise quality of care.

Calculation of this indicator requires an organization to have a clearly defined and agreed-upon definition of "filled." Payroll systems may fail to provide accurate information if staff remain on a payroll (with or without pay) when they are on extended or permanent leave; hence the position is "filled," but no one is performing the functions of that staff member.

Because of inadequacies of personnel information systems, some organizations (especially public sector) rely on an annual count of employees as part of their management information system reports to provide the data for this indicator.

Where payroll and annual reporting systems are inaccurate, including questions about human resources on periodic facility assessments is common. A frequently used indicator is:

- Percent of facilities that have the full complement of staff on duty on the day of a site visit

An important limitation of this indicator is that it does not measure the **competency** of the staff filling key positions. Assessing competency requires a more time-consuming, qualitative analysis of job descriptions, qualifications of staff in positions, and performance reviews.

[9] In some countries, in large health organizations, more complex algorithms that include staff time available, type of patient load (inpatient, outpatient, deliveries, community visits, among others) and administration time are used to determine appropriate staffing levels.

Indicator

Staff Turnover Rate

Definition
The rate at which staff are leaving an organization or program as a proportion of the total staff employed during a reference period (e.g., 12 months).

The indicator is calculated as:

$$\frac{\text{\# of staff who vacated their positions}}{\text{\# of staff employed by the organization or program}} \times 100$$

Data Requirements
Accurate, up-to-date counts of staff who have left positions and of total number employed at the midpoint of the reference period (e.g., 12 months)

Data Source(s)
Human resources information systems; personnel records; organization's payroll system (if accurate); "head count" survey (in the absence of routine personnel information system)

Purpose and Issues
Staff turnover is an important way to measure both the effectiveness of the human resources management system and the overall management of an organization or program. It provides a complementary measure to the previous indicator on key positions filled. If turnover is high, the organization/program must incur additional costs of hiring new staff; these costs include interviewing, checking references, and start-up training, among others. Because human resources often consume greater than 70 percent of reproductive health program budgets, retention of qualified staff, or lack thereof, can have a very large impact on productivity and performance.

Whereas this indicator can raise a "red flag" (signal possible personnel problems), human resource managers may lack the authority to solve the root causes of the problem (e.g., supervision, pay scales, promotion). Further understanding of the causes for turnover requires more in-depth analysis. Some organizations require exit interviews of all employees before departure; examination of these records should indicate if turnover relates to job satisfaction, pay issues, retirement, or other factors that the organization or program can address.

Generally, annual analysis is sufficient for this indicator, although managers may want to examine this indicator more frequently in the case of a perceived increase in attrition. Managers will also want to review it over longer periods of time to facilitate long-term planning for hiring and staff development.

Indicator

Availability and Use of a Coherent Planning System

Definition
The availability and utilization of a systematic process for planning

Planning systems include systematic procedures for short-, medium-, and long-term plans. Their objective is to ensure support for and achievement of the organization's mission, goals, and strategies. A coherent planning system supports the efficient and effective development, implementation, and monitoring of plans.

This indicator measures both the availability and utilization of six key elements during the last planning cycle, as follows:

Available	Utilized
Templates/formats for all planners to follow	Templates used for most recent plan
Schedules for developing, monitoring, and updating the plan	Schedules followed reasonably well for most recent plan
Manuals describing the planning process	Manuals referred to during preparation of most recent plan
A mechanism to assure that activities in the plan are linked with **budgets**	Activities in most recent plan are budgeted
A process for **monitoring** progress	Systematic monitoring performed by managers for most recent plan
A system for generating progress **reports**	Progress reports produced on regular basis for most recent plan

Evaluators can rank an organization on a 12-point scale, calculated by assigning one point for "available" and one point for "utilized" to each of the six items in the table presented here.

Data Requirements
Evidence of a documented planning and budgeting process; evidence of existing strategic and operational plans; evidence of monitoring of plans and schedules

Data Source(s)
Organizational documents; interviews with key staff members

Purpose and Issues
Institutions that perform effective strategic planning can better understand and respond to changes in conditions affecting the organization, and they can more effectively apply available resources to client needs and generate demand for services (USAID, 1999). Managers should refer to such plans when they make management decisions and should monitor and adjust plans continuously to adapt to changing internal and external conditions. Hence, a coherent planning system requires not only the products (plans), but also the process (implementation) grounded in the strategic and budgetary realities of the organization.

Through observing organizational/programmatic documents and through interviewing key informants, an evaluator can consider whether key elements of a planning system exist and whether they have been used during the last planning cycle.

One limitation of this indicator is that it does not assess the quality of the plans that are actually produced by the planning system. However, the next indicator, **Quality of Strategic and Operational Plans**, addresses this issue.

Indicator

Quality of Strategic and Operational Plans

Definition
A strategic plan, a written document stating the key elements of strategy and goals of an organization or program, prioritizes and focuses for the medium- to longer-term (three to five years). An operational plan, a document generally written annually, guides the activities of an organization or program in the immediate future.

Data Requirements
Evidence of existence of strategic and operational plans; assessment of the plans

Data Source(s)
Strategic plans; operational plans; budgets

Purpose and Issues
A quality strategic plan contains:

- Reference to the mission statement of the program or organization;
- A listing of key strategies for the near term (e.g., next one to three years) and a prioritization of said list;
- A vision for the long term (e.g., how the strategic plan will enhance long-term goals of the organization beyond the next five years); and
- A human resource plan defining staffing and training needs required for fulfillment of the strategic plan.

A strategic plan should include a human resource plan to ensure that personnel needs (in terms of required skills) are considered in the institution's planning. Strategic planning that includes human resource needs will help the institution avoid the possibility of management failure due to institutional loss of memory when personnel leave.

When assessing the quality of an organization's strategy, the evaluator must first examine each major strategy and assure consistency with the organization's mission. He/she may use a comparison matrix or checklist. One axis of the matrix can contain the major points of the mission statement purpose(s) and intended clients; the other can contain the major strategic directions. The evaluator can then calculate what proportion of the boxes have a check.

A quality operational plan contains:

- Detailed activities and tasks;
- A definition of personnel and other resources required to accomplish the activities;
- A link to a detailed budget of activities (see the indicator **Effectiveness of Financial Management Systems**); and
- A planned activity to review and adjust the plan to adapt it to changing conditions.

When assessing the quality of an operational plan, an evaluator examines each major activity area and defines consistency with the organization or program strategy. A comparison matrix or checklist as described above can be used.

An evaluator (or an internal management team) can simply use checkmarks or can assign a numerical value to the level of consistency and then can calculate a numeric score. An example of a scale is:

0	**No consistency**
1	**Partial consistency**
2	**Total consistency**

Indicator

Number/Proportion of Organization/Program Units Systematically Using Information to Monitor Performance

Definition
The capacity of managers to use routine information systems to monitor performance

An effective management information system (MIS) processes raw data and produces information that allows decision-makers to understand how well the organization or program is performing. This indicator can apply to information systems designed for producing information to monitor the overall or "bottom-line" performance of the organization/program or the performance of individual management units such as finance or human resources.

Organization/program units in the context of reproductive health refer to either service delivery points, branch offices, or management units/departments within headquarters.

Data Requirements
Written (or computer generated) evidence of use of data (e.g., indicator charts, graphs); information system reports; planning objectives; number of units that show evidence of use of data; and total number of units

Data Source(s)
Assessment of the MIS, conducted during routine supervision or by an external evaluator; interviews with key staff.

Purpose and Issues
When calculating this indicator, an evaluator examines the institutional documents specified above for the presence of the following three attributes of systematic monitoring:

- Data are converted into information on specified indicators of performance;
- Indicators are used to compare performance to objectives or standards established in operational and/or strategic plans; and
- Trend/time analysis of information is available in a way that is accessible to management (e.g., tables, graphs).

Applying this indicator is common in reproductive health programs to measure the use of routine service indicators. This set should include at least one indicator on each of the basic RH services, such as family planning, antenatal and postnatal care, safe-delivery services, STI/HIV treatment and prevention, as well as other programmatic indicators.

The indicator can also apply to a review of other information systems including disease surveillance systems, personnel or human resources, commodities and logistics, finance, and facilities and equipment.

This indicator assumes that the information available to the evaluator is both timely and accurate. The following two indicators in this section address these questions: **Proportion of Reporting Units Submitting a Completed Routine MIS Report on Time** and **Percent of Data Elements Reported Accurately in MIS Reports**. This indicator does not measure the effectiveness of decision-making based on information that is monitored. To do so would require a more complex, in-depth measurement process that involves interviews with key informants and intensive document review.

Indicator

Number/Proportion of Reporting Units Submitting a Completed Routine MIS Report on Time

Definition

The extent to which an organization makes information accessible in a timely fashion

Management Information System (MIS) refers to the mechanisms and procedures for the collection and use of routine data.

"On time" means the report is received within a specified time from the end of the reporting period. Common reporting periods are "within 7 days after the start of a new month," "within 14 days of the start of a new quarter," or other set period.

Data Requirements

Records indicating exact dates when reports were sent from a collection point and received at a unit or office where they are aggregated

Data Source(s)

Log books; stamped and dated reports;
computerized MIS databases

Purpose and Issues

This indicator provides a measure of the extent to which an organization makes information accessible in a timely fashion. The existence of complete and timely information is a pre-condition for measuring the previous indicator regarding the use of information for monitoring. The qualifier "on time" in the definition of the indicator highlights the need to monitor activities and outcomes at a pace consistent with the timeframe that leaders and managers set for planning, monitoring, evaluating performance, and decision-making.

However, deadlines for submitting routine reports must be reasonable given the particular challenges to delivery. Experience in several countries has shown that pressure to submit monthly reports on time leads workers to omit some data. Thus, data are not only incomplete but also cannot be compared to data from institutions submitting figures for a full month.

Evaluators can apply this indicator to other types of reports critical to the management of reproductive health services. Such reports may be quarterly or annual performance reports (which draw on data from multiple sources), drug stock-out reports, notifiable disease or event reports, or periodic population-based reports. To be defined as "on time," a report must be received well before information is required for decision-making.

The indicator does not measure the extent to which upper level management processes the reported data in a timely manner. A complementary indicator is:

- Percentage of reports processed (aggregated manually or entered into a database) according to deadline

Indicator

Percent of Data Elements Reported Accurately in MIS Reports

Definition
The accuracy of reporting and aggregation

A "data element" is a single datum input into an MIS. This term can refer to a single cell on a routine MIS report or in a data entry screen for automated MIS.

Data Requirements
Primary data from health units/programs and aggregated data at all levels

Data Source(s)
Registers, patient records, and/or tally sheets; monthly/ quarterly reports from MIS

Purpose and Issues
This indicator is used in health programs to measure accuracy of reporting and aggregation. Inaccuracies can occur when data are recorded, tallied, transposed onto reporting forms, and aggregated.

Evaluators can select a number of sample data elements and can determine the extent of agreement between (1) data recorded in service registers, patient files, tally sheets or patient files, and (2) data reported on MIS forms.

Item	#Recorded	#Reported	Consistent?
Pill cycles	36	32	No
IUDs	5	5	Yes
Injectables	18	18	Yes
% accuracy		**2/3 or 66%**	

To measure accuracy of aggregation, organizations frequently compare a sample of data elements from the raw data reported from multiple service delivery points (SDPs) to the aggregated total reported to higher levels.

For example:

	SDP 1	SDP 2	SDP 3	Actual Total	Reported Total	Consis -tent?
Pill cycles	32	84	14	130	130	yes
IUDs	5	10	7	22	20	No
Inject -ables	18	21	23	62	49	No
% accuracy:				**1/3 or 33%**		

Program or facility managers may inflate or under-report data for a variety of reasons. A more sophisticated indicator would therefore measure the **relative difference** between recorded and reported data to determine if the problem is over- or under-reporting of data. This approach is more time-consuming and tedious to calculate but better measures reporting accuracy.

If an evaluator wants to examine issues of data quality but lacks the time to do a detailed analysis of accuracy as required in this indicator, he/she may use a more basic measurement of an MIS that looks at **completeness** of data:

- Percentage of health facilities sending reports with no missing data.

Field testing of this indicator shows that it provides a proxy indicator for the overall functioning of information systems. However, more in-depth analysis is necessary to explore whether the **recorded** data reflect reality. This assessment demands direct observation of those staff recording data, a technique that may be too costly to warrant routine use.

Indicator

Institutionalization of a System of Quality Assurance

Definition
Quality refers to offering a service or product in a way that consistently meets the clients' needs. Quality Assurance (QA) is a generic term describing a number of management approaches (Continuous Quality Improvement [CQI], Total Quality Management [TQM]), all of which recognize that many organizational problems result from systems and processes, as well as from a lack of clear performance expectations, rather than negligence on the part of individuals. QA, as it applies to the management of reproductive health programs, generally involves the encouragement of staff members at all levels to analyze systems and processes, to use information to identify the nature and size of each problem, and to design and implement activities to improve services and client satisfaction. (For more detail on QA, see Part II.H.2c)

Data Requirements
Evidence of the availability of quality standards and protocols; budget allocation for QA activities; performance/provider reviews of adherence to standards; client satisfaction feedback on quality; staff feedback on involvement in quality initiatives

Data Source(s)
Organizational documents including service delivery guidelines; interviews with managers, supervisors, and other staff at all levels; budget; staff performance reviews; training curricula; client satisfaction surveys; suggestion boxes

Purpose and Issues
This indicator measures organizational commitment to QA; more detailed indicators can be found at the site listed below.[10]

From a management perspective, the following six items are essential to developing a composite score of commitment to QA:

- **Evidence of integration of quality assurance into the organization's mission and strategy**

By measuring this item, an evaluator will first understand if an environment or organizational culture of quality improvement exists.

- **Evidence of integration of quality assurance into the organization's plans and budget**

Not all QA activities will require a separate line item budget; some of them are combined with other activities and are done only at marginal cost. Assessment of budget allocation can be difficult unless the budget is highly detailed/annotated.

- **Evidence of the availability of quality standards or protocols**

This indicator is easy to measure but requires identification of protocols for service delivery (clinical protocols, counseling) as well as for management (e.g., supervision, storage of supplies, infection prevention, MIS reporting) for each major type of reproductive health service.

- **Performance/provider review of adherence to standards**

As with the previous item, this aspect requires measurement according to service delivery and management standards for each reproductive health area. It requires direct observation of staff (such as by a supervisor or mystery client); measurement can therefore be time consuming depending on the volume of services at a given facility.

[10] Gutierrez, Maria. 2001. http://erc.msh.org/mainpage.cfm?file=2.2.6.htm&module=quality&language=English

- **Mechanisms for obtaining client satisfaction/ feedback on quality**

Measuring client satisfaction is the principal means of knowing whether QA initiatives are reaping any benefits. Among the numerous methodologies for measurement, the most common is the client exit interview. One of the weaknesses of exit interviews is that clients sometimes forget details of a visit or do not know what practices are acceptable. Clients may fear impact of negative responses on availability of and access to services.

- **Mechanisms for collecting provider perspectives on quality**

Measurement in this area can reveal gaps between client and provider understandings of quality. It can also help managers understand the extent to which providers feel encouraged or rewarded for taking initiative to address quality.

Indicator

Availability of Budget(s) Linked to Operational Plan(s)

Definition
The effectiveness of an organization or a program to match available resources with planned activities

A budget is a document that projects the costs, and in many cases, the revenues of a defined activity, program, project, or organization. It is also a financial plan that quantifies programmatic goals and objectives by guiding the allocation of financial and human resources (MSH, 1999c).

Data Requirements
Evidence that activities in the plan have been costed and that resources have been allocated to individual or sets of activities

Data Source(s)
Budgets; chart of accounts; operational plan(s)

Purpose and Issues
This indicator measures the effectiveness of an organization or a program to match available resources with planned activities. Budgets may cover single activities (sub-budgets) or whole programs; a well-constructed budget allows for the "rolling up" of several sub-budgets into a total operating budget.

Measurement of this indicator assumes a certain level of clarity of the budget. For this reason, those preparing budgets must provide information on the units of cost, that is, the number of participants, the specific number and type of materials, the number of days of per-diem, and the number of persons to receive per-diem, for example. This level of detail makes it easier to assign a value or cost to each unit. It also makes it easier to modify the budget because numbers in the plan shift over time. When budgets are broken down into unit costs, tracking changes in costs over time is easier. It is also easier to review quotes that come in from vendors, because historical data are available from prior budgets.

Indicator

Effectiveness of Financial Management Systems

Definition
Financial management refers to managing an organization or program's resources to meet goals and objectives as effectively as possible by using those resources to carry out planned activities. A financial management system is composed of a series of tools and processes that permit the control, conservation, allocation and investment of an organization's or program's resources.

The scoring system for this indicator is as follows:

Descriptor	Score
Expenditures are tracked by budget-line items (e.g., inputs, salaries, utilities, materials) and are recorded as they occur. However, financial reports cannot be generated effectively.	1
Expenditures are not only tracked by inputs, but are also linked to services and materials purchased, and to the activities they support. Financial reports exist but are not used to analyze costs.	2
The financial system produces income/revenue data and case flow analyses; costs are allocated by cost centers (e.g., products/outputs, service units, sets of services). Financial reports, which compare actual expenditures to budget, are sometimes used to analyze costs.	3
High-quality financial reports are linked to budgets and consistently used for management decisions, including allocation of resources.	4

Data Requirements
Information on planned and budgeted activities; expenditure information; evidence that financial reports are used for decision-making

Data Source(s)
Document review of budgets, case flow statements, income statements, balance sheets, and interviews with managers

Purpose and Issues
Managers must understand their current financial situation (liquidity) and their long-term position (solvency) if they are to lead towards effective performance. Evaluators can use a suggested scoring system to measure this indicator. The system combines elements of how expenditures are recorded and tracked and of how financial information is used to make decisions. An underlying assumption is that reports are accurate and timely. These essential functions enable programs/organizations to understand their current and long-term liabilities.[11]

Because measurement of this indicator requires performing a valid document review, evaluators must fully understand locally accepted accounting principles and reporting requirements. Ascertaining how management makes decisions will require not only review of documents such as operational plans, but also interviews with key decision-makers in the program or organization (MSH, 1999a).

[11] Embedded in levels 2 through 4 is the assumption that reports are accurate and timely.

Indicator

Percent of Annual Revenue Generated from Diverse Sources

Definition
The ability to reduce dependency on single-source funding

Revenue are monies or the equivalent received from sales, services, fees, donations, and grants. In the case of grants, only the portion actually spent is considered revenue; the balance may have to be returned to the donor.

Typically, the sources of funds available to reproductive health programs and organizations include: national governmental entities (such as Ministry of Finance/ Ministry of Health), local governmental entities (municipalities or districts); international donors or foundations; sales of services (through charging fees, contracting with other agencies and or insurance schemes)[12]; sales of supplies such as contraceptives; and donations from local donors, corporations or individuals.

Data Requirements
Total resources (funding or in-kind donations) generated or received by the organization/program from each source

Data Source(s)
Income statements; revenue reports; audit reports or other financial records

Purpose and Issues
This indicator measures the ability of managers to use financial management systems to make decisions that will reduce dependency on single sources of funding. Diversity of funding sources spreads the organization's or program's risk of over-dependency on a single source of revenue, and allows the organization/program greater flexibility in determining future directions. Many NGOs find diversification critical to their sustainability over the long term. In addition, with decentralization and the subsequent requirement for local governments/health departments to generate some of their funds, public sector revenue diversification and cost sharing are now relatively common in developing country settings.

This indicator assumes that organizations/programs have basic financial management procedures in place to provide revenue information. For newly decentralized public sector entities, this indicator may be difficult to measure because systems for aggregating information on the revenues generated from service delivery fees, for example, are nascent.

An important consideration in measuring this indicator relates to how funding is earmarked. An organization or program may have diversified sources of funding, but the donors may demand that money be used only for very specific programmatic activities, rather than for general operating or developmental costs that are crucial for sustainability. In addition, the earmarks may not align with the current strategic directions of the organization or program, and this situation may necessitate additional unforeseen expenditures. Hence, evaluators should use this indicator in conjunction with other indicators that permit a more in-depth financial analysis.

[12] These represent the principal forms of direct cost recovery although other mechanisms may exist.

Part II.D
Training

- Number of trainees by type of personnel and topic of training
- Number/percent of trainees who have mastered relevant knowledge
- Number/percent of trainees competent to provide specific services upon completion of training
- Number/percent of trainees deployed to an appropriate service delivery point and job assignment
- Number/percent of trained providers who perform to established guidelines/standards
- Number/percent of training events that achieve learning objectives
- Organization has the capacity to maintain a functional information system on its training program
- Number of faculty and trainers who demonstrate the use of professional core training competencies on the job
- Organization has a systematic process for follow-up and support of trainees after the training event
- Existence of training strategy based on needs assessment to improve quality of service delivery
- The organization systematically evaluates its training program to improve effectiveness
- Demonstrated capacity to carry out training on a sustained basis
- Adaptability of the organization/system to changing needs in a training environment

Part II.D

Training

On the surface, one might consider training an "easy" area to evaluate, thanks to the pre- and post-tests often used in connection with training activities. Although such instruments continue to serve a useful function, they by no means capture the full range of training effects.

In recent years, the state-of-the-art for the evaluation of training has changed substantially. First, organizations are no longer content to evaluate based on the number of training events, number of participants, improved scores on post-test instruments, or other process indicators. Instead, competency-based training has become the standard in organizations worldwide.

Second, whereas in the past, when training was viewed as an isolated set of activities – often the panacea for whatever was ailing a service delivery system – today training programs are expected to address a broader range of issues, including contextual factors that go far beyond the traditional limits of training. For example, factors that affect a person's ability to perform satisfactorily include (1) having clear job expectations, (2) receiving clear and timely feedback on performance, (3) having access to the information, tools, and other resources needed for the job, (4) receiving encouragement or motivation to perform well, (5) having appropriate knowledge and skills, and (6) having organizational support for improved behaviors (e.g., more positive client-provider interactions are more likely to occur and be sustained if the facility or system is overtly client-oriented). Programs have moved beyond conventional training to a process known as "Performance Improvement (PI)." The rationale for PI and the role that indicators play in this process are summarized in Appendix C of this *Compendium*.

Third, where possible, evaluators attempt to measure the effects of training on the service delivery environment itself (i.e., improved access, enhanced quality). The indicators presented in this section stop short of measuring these effects, in large part because most organizations do not have the financial or human resources needed to conduct studies to demonstrate this relationship empirically. Methodologies exist to link the quantity and quality of training with its effects on the service environment (e.g., access to services and quality of care, for which indicators are presented in section II.H of this manual). However, this type of "linkage" cannot be accomplished without a special study based on an experimental or quasi-experimental design or multivariate longitudinal analysis to demonstrate that the facilities receiving training are superior on one or more specific measures than those that did not receive the training.

Kim et al. (1992) present one such study in Nigeria, in which the clients of providers who received training in counseling techniques had a higher rate of follow-up visits than did those whose providers had not received the training. A second example is a study by Dietrich, Guilkey, and Mancini (1998) that linked facility and household data from the DHS in Tanzania; they found that the presence of at least two trained providers in a facility resulted in 11-23 more family planning revisits (i.e., of new and continuing users) compared to facilities without trained providers. Unless program managers and donors are willing to commit funds to such special studies, they basically operate on the **assumption** that good training results in improved performance and enhanced quality of care in the service delivery environment.

No universally accepted word exists in English to describe the person that attends a training event. We have used "trainee" in this section, but recognize the existence of other terms, such as "participant" (which implies more active involvement), "learner" (which reflects the absorption of new knowledge and skills), or "student" (especially in a pre-service education institution). Readers are encouraged to use the term most widely accepted in their local work environment or most appropriate for the activity in question.

A large portion of the personnel to be trained in the context of RH programs will work in a clinical setting,

such as a family planning clinic, STI treatment center, or obstetrical care ward. However, a growing proportion of persons to be trained will work in a non-clinical setting; such groups include community health workers, teachers, peer educators, journalists, women's groups, and others. Whereas this section focuses on training for service delivery, many of the concepts can be adapted to other types of program implementation.

Methodological Challenges of Evaluating Training Programs

Specific methodological challenges of evaluating training programs include the following:

- **"Training" takes many different forms and levels of intensity.**

A given training program may address learning objectives that require as little as a couple of hours to achieve, or it may last a month or more. Moreover, "training" may constitute an isolated activity (which has generally been the case in the past), or it may be one part of an ongoing and integrated program to deal with multiple problems in the service delivery environment. As such, the evaluator must clarify the **type of training** event that is being evaluated and the intended objectives.

- **Training is designed to have multiplier effects, but the evaluation of training rarely captures such effects.**

Training occurs in numerous forms (i.e., types of training) and levels of intensity. Some training programs are set up explicitly to have multiplier effects, such as "cascade training" (e.g., one level of program personnel is trained at a central location). These trainers then begin training groups of providers. With time, the original trainers train more trainers. There is a "cascade" from the initial group of trainers, but all training is based on specific training standards and materials. Other training programs may not be designed in cascade format, but in fact they may produce a spin-off effect when the trained person returns to the service delivery setting. For example, a trained provider may share content and skills with co-workers, either formally through structured on-the-job training or informally by discussing new content or sharing training materials. Because a trained provider may be immediately promoted to another level of care or to an administrative position, the evaluator may have trouble ascertaining the added/amplified effects on that level. In theory, one could conduct a special study to capture the effects of the training at different levels of the system, but such a study would be complex and expensive. In practice, the multiplier effects of training tend to get overlooked in the evaluation process. However, if such effects were overt objectives (and adequate human and financial resources were available), evaluators could measure them.

- **The training – however well executed – may be of little value to the program if organizations select inappropriate participants.**

Traditional group-based training is often considered a "perk." It allows an individual to obtain new (and generally marketable) skills, often in an enjoyable environment away from the pressures or routine of the workplace, with the added benefit of cash payment to cover living expenses (in the case of traditional, off-site training). As a result, the demand to attend a given training course may outstrip the number of slots available. Moreover, officials in high positions may use training opportunities as a means of repaying favors, whether or not the person selected is the most appropriate for the task. For example, on occasion organizations send administrative staff to courses intended for clinical practitioners, and thus waste opportunities. Seniority as well as politics also plays a role in selecting participants for training. Whereas one hopes that this practice is on the decline, it represents a problem in evaluating the effects of the training on the service delivery environment. Training organizations have identified several means of addressing the problem. Some have developed ways to encourage appropriate attendees for training while ensuring that the administrator-level staff members (who are sometimes sent to a training course to enlist their support) are actually involved in the training process in a different way. Alternatively, many organizations are developing other approaches to training, such as distance learning, self-directed learning, peer learning, and on-the-job training.

- **The guidelines and standards against which to evaluate performance may differ by country.**

A number of the indicators refer to guidelines or standards, against which service provider practices are to be evaluated. Some international standards do exist, such as WHO's Medical Eligibility Criteria or the Checklist

for Clients Who Want to Initiate Combined Oral Contraceptives (COCs) in Community-based Programs (Stang, Schwingl and Rivera, 2000; Technical Guidance/ Competence Working Group, 1997 & 2001). However, most governments prefer to establish their own standards and guidelines (or to adapt international ones to their own situation). The benefit of country-specific standards is their relevance to the local context; it is unrealistic to think that a very poor developing country will be able to provide the same quality of care as a country that has "graduated" from donor funding in a given area. Commitment from key constituencies tends to be greater if the standards are developed with local input. However, the existence of local guidelines and standards results in non-comparable results across countries. Since the major purpose of program evaluation is the improvement of service delivery program implementation in a given country setting, the difference of standards across countries should not be considered a major limitation.

- **Ideally, evaluators should assess training in terms of changes in the service delivery or program environment, but doing so requires technical and financial resources.**

Training programs are generally designed to improve performance in a service delivery or program setting (e.g., increasing access, improving quality, making program personnel more gender-sensitive, or improving integration of different services). This *Compendium* includes indicators to measure these elements of the service delivery environment. (See Part II.H.) However, evaluating the extent to which a training intervention achieves changes on these dimensions requires an experimental or quasi-experimental design or multivariate longitudinal analysis. Many training organizations recognize the effectiveness of training, but they lack the financial or technical resources to conduct such evaluations. (The topic of study designs is beyond the scope of this manual, although interested readers are referred to chapter IV of Bertrand, Magnani, and Rutenberg, 1996.)

Although training programs are often asked to "justify" their work through concrete examples of their effectiveness, few program administrators or donor agency representatives are willing to fund evaluations of training effectiveness. This problem is by no means unique to training, but has hindered the advancement of evaluation in this area.

- **Those who attempt experimental or quasi-experimental designs run into problems of "clustering" and intra-class correlation in evaluating training.**

Evaluators often use the individual as the unit of analysis, but individuals from the same service delivery point or those taught by the same trainers using a classroom or group approach are more likely to perform in a similar manner (have less variance) than are those from different locations or those taught by other trainers. This clustering has important ramifications not only for the analysis of the data, but also for evaluators' sample size calculations. Evaluators should consult a statistician or expert in sampling to discuss the best strategy for addressing this problem in the design of their evaluation.

Two Levels of Evaluation

In this section on training indicators, we distinguish two levels of effects: **individual** and **organizational**. Whereas the evaluation of training has tended to focus on the individual service provider in the past, there is a growing trend to evaluate training programs in terms of their effects on the service delivery system (e.g., of the Ministry of Health in a given country). The first five indicators in this system correspond to the individual level, whereas the final eight indicators relate to the system as a whole (i.e., the organizational level).

Indicator

Number of Trainees by Type of Personnel and Topic of Training

Definition
"Trainee"[1] refers to any type of participant in a training event, regardless of its duration. "Type" refers to the different categories of participants (e.g., physicians, nurses, social workers) or the subject matter covered (e.g., IUD insertion, universal precautions for HIV/AIDS prevention, use of a partograph during delivery, peer education techniques, skills to develop a radio script).

Data Requirements
Number of persons (based on an actual list of names for potential verification purposes), their professional positions, and topic of training

Data Source(s)
Records, usually kept by the training division, which are used both for administrative purposes during the training (e.g., distributing per diem) and for monitoring trainees at a later date

Purpose and Issues
This indicator serves as a crude measure of activity; evaluators can use it (1) for determining whether a program/project meets its target and/or (2) for tracking progress from one year to the next. However, the "unit of measurement" is not strictly speaking uniform, in that one trainee may have attended a course for one day, whereas another may have participated in a course for three months.

Evaluators can further improve the measure in several ways:

- Expressing the number of trainees by type of training;
- Expressing the number of trainees as the percent of the number scheduled for training in a given year;
- Expressing the number of trainees as a percent of the estimated number needed to be trained to fill a national program mandate, determined through a systematic needs assessment prior to the initiation of training activities; and
- Expressing the number successfully completing the course as a percent of the total that enroll in the course.

[1] Alternative terms: participant, student, or learner.

Gender Implications of this Indicator

A gender perspective on training assesses the following questions:

1. How are the curricula developed?
   - Are women and men involved in determining what subjects will be covered?
   - Are women and men involved in drafting the contents?

2. What is the content of the curricula?
   - Is the language gender-sensitive?
   - Are the contents in line with ICPD principles on sexual and reproductive rights?
   - Do the contents cover gender-based differences in access to and use of health services?

3. Who carries out the training?
   - What proportion of the trainers are men? Women?
   - What are the roles of male versus female trainers?

4. What training methodologies are used?
   - Are both men and women encouraged to speak-up during training? How often do men speak up? How often do women?
   - What is done to help women participants overcome their shyness/intimidation in groups also containing men? This question is especially important in societies where mixed groups are uncommon.

5. Who receives the training?
   - What proportion of the trainees are men? Women?
   - Are the training sessions held at times and places convenient for both women and men participants, given gender-based constraints, such as restricted mobility, lack of access to money for transport, and household/childcare responsibilities?
   - Is there an equity plan to ensure that all employees get access to training?
   - Is this plan enforced?

Indicator

Number/Percent of Trainees Who Have Mastered Relevant Knowledge

Definition
Evaluators must define "mastery" in terms specific to a given context. "Mastery" conventionally relates to acquisition of knowledge. ("Competency" involves both knowledge and skills; see next indicator **Number/ Percent of Trainees Competent to Provide Specific Services upon Completion of Training**.)

This indicator is calculated as:

$$\frac{\text{\# of trainees that have mastered knowledge}}{\text{Total \# of trainees tested}} \times 100$$

Data Requirements
Listing of individuals; scoring criteria to define "mastery;" evidence of mastery of knowledge (e.g., scores on tests)

Data Source(s)
Administrative records (training files); written tests (e.g., pre-and post-tests of accurate, up-to-date knowledge)

Purpose and Issues
This indicator, commonly used to evaluate training, measures the trainees' ability to retain key information in the short term (during and at the end of training). Low post-test scores reflect inadequacies in the course and/or the inability of trainees to absorb the information. Every training organization that has developed or uses training manuals has identified the knowledge that a category of trainees should acquire on a specific subject. Pre-and post-tests measure this knowledge.

The test results indicate whether the trainee understands certain key points, even though the number and definition of key points will differ by context. The items included in the test should be those most relevant to a particular training exercise, which relate to program performance. If the same questions appear on subsequent tests, this indicator can monitor trends over time within a program and can determine knowledge retention as part of formal training evaluations.

This indicator has two limitations. First, tests lack standardized items. Some training organizations have a list of questions they encourage host country organizations to adopt for testing purposes on a given topic, but some countries opt to design their own questions. This lack of standardization makes it difficult to compare the results from this indicator across countries and even across programs within a given country. Second, the concept of "mastery" is not consistent across settings. For example, in some countries, a passing grade may be 60 percent, whereas in others the required score for passing may be 100 percent. Improved knowledge is only one indication of training effectiveness; by itself, it does not necessarily ensure improved performance.

Despite these limitations, training organizations routinely use this indicator to control the quality of training conducted in connection with their activities.

Indicator

Number/Percent of Trainees Competent to Provide Specific Services Upon Completion of Training

Definition

"Competence" refers to trainee's ability to deliver a service according to a set standard, which may differ according to the training context. Thus, the evaluator must know the standard of the context. Training organizations use "competence" to refer to the acquisition of skills (although performing a skill often requires knowledge). "Upon completion of training" refers to the final assessment given as part of the training event.

This indicator is calculated as:

$$\frac{\text{\# of trainees delivering services according to set standards}}{\text{Total \# of trainees tested}} \times 100$$

Data Requirements

Listing of trainees; pre-established operational definitions of criteria determining competency; assessment of each trainee against established standards for a number of service delivery or programmatic tasks, conducted by an expert observer

Data Source(s)

Competency tests (often in the form of a checklist administered by the trainers and/or external expert observer)

Purpose and Issues

This indicator measures the technical competence of participants who have completed training in a specific skill set. The indicator reflects both the adequacy of the training and the ability of trainees to absorb the information.

Several training organizations working in reproductive health have made considerable efforts to standardize the items on the checklist for given program areas (e.g., family planning) as well as the interpretation of each item on the list (e.g., what constitutes satisfactory performance on that item).

However, at the field level evaluators may use inconsistent criteria to define competency. Some programs may expect a 100 percent grade before they judge the trainee competent in a battery of skills, whereas another organization may judge competency at the 50 percent grade level. In some cases, local standards for the delivery of FP services may not exist, in which case evaluators can use international standards.

Assessing competency generally is more complex than the simple testing of knowledge. Whereas measuring knowledge is easier than measuring competency (i.e., the correct performance of skills), the latter is more likely to define the quality of care that providers give.

Indicator

Number/Percent of Trainees Deployed to an Appropriate Service Delivery Point and Job Assignment

Definition
"Trainees" refer to individuals who participated in a specific training course or event. "Deployed to an appropriate service delivery point" refers to a facility that routinely provides the type of service for which they are trained (e.g., counseling and testing for HIV). "Job assignment" refers to the fact that they are assigned a task at that facility that allows them to perform the skills they obtained during training.

This indicator is calculated as:

$$\frac{\text{\# of trainees in positions where their training is applied in service delivery}}{\text{Total \# of trainees}} \times 100$$

Data Requirements
Listing of trainees at the course or event; place of work and job description of each trainee "X months" (e.g., six months) post-training

Data Source(s)
Program records of trainees; listing of job postings and job titles for employees within a given organization (e.g., Ministry of Health, NGO network of clinics)

Alternatively, a follow-up survey of trainees who had participated in a particular course or event

Purpose and Issues
This indicator measures the extent to which the organization is taking full advantage of the training it provides to its personnel. Ideally, 100 percent of trained personnel will apply their skills to service delivery at some other selected interval post-training (e.g., six months). This indicator provides a quantitative measure of the efficiency of training because it monitors the extent to which organizations assign trained employees to appropriate positions in the appropriate facilities that tap the service delivery skills learned in training.

The limitation of this indicator is its failure to shed light on the reasons for "departures" from service – if a far lower percentage are deployed to appropriate positions than expected. In such a case, the organization in question should separate the "place to which deployed" and "job assignment" to further understand the dynamics at hand.

Ideally, this indicator will accompany the next one measuring the "**Number/Percent of Trained Providers who Perform to Established Guidelines/Standards**." Trained providers must not only work in appropriate facilities, they must also perform the appropriate tasks in the right places; one wants them to be doing the right things as well.

Indicator

Number/Percent of Trained Providers Who Perform to Established Guidelines/Standards

Definition
"Trained providers" refers to individuals who have participated in one or more training events.

"Guidelines/standards" refer to the written criteria adopted by the organization to outline the processes/or implementing of specific procedures.

This indicator is calculated as:

$$\frac{\text{\# of trained providers carrying out specific procedures according to established guidelines/standards}}{\text{Total \# of trained providers evaluated}} \times 100$$

Data Requirements
Listing of trainees; specification of the skill and established standards for the skill; assessment of skills level of trained providers conducted by an expert observer

Data Source(s)
National guidelines/standards for service delivery; and checklists and notes of an expert observer

Written tests can determine knowledge/stated practice of performance to standard.

Purpose and Issues
This indicator measures the retention of skills acquired during training and the application of such skills to the job at hand; it also identifies possible candidates for retraining, or alternatively, for promotion. It measures both the adequacy of the training to impart these skills and the ability of the trainees to assimilate and to retain the information and skills over time.

This indicator goes beyond the previous one to ensure that providers can do their work (a variety of skills/services) according to the standard of the workplace. It measures performance in a work routine or a work day rather than just the skill learned in training.

If a trained provider fails to retain the skills acquired, it is important to explore the reasons. Possible explanations may include a lack of continued practice due to low client load, too much time lapsed since the training, or lack of reinforcement on the job. Conversely, a provider may improve his/her competency by continuously performing the task during the months following the training. In fact, this indicator reflects less the quality of the training than the subsequent work environment of the training (e.g., type and frequency of supervision, demand for the skills).

Evaluators can apply this indicator at a specific interval post-training (e.g., 6 months, 12 months) among those who attended the training course or event. Alternatively, evaluators may apply it to all service providers in the system to capture both the coverage of training and the quality of the instruction (i.e., number/percent of providers who perform to established guidelines/standards).

Indicator

Number/Percent of Training Events That Achieve Learning Objectives

Definition
"Objectives" are outlined in the training curriculum or syllabus.

This indicator is calculated as:

$$\frac{\text{\# of courses that achieve outlined objectives}}{\text{Total \# of courses evaluated}} \times 100$$

Data Requirements
[If assessed by participants] Response to the question "in your opinion, did the course meet the objectives outlined in the first session?"

[If assessed by independent observer with expertise in the content area] Review of the course content and observation of trainees' acquisition of knowledge and skills

Data Source(s)
Evaluation of the training event by trainees upon its completion; or notes of independent course observer

Purpose and Issues
The purpose of this indicator is to determine whether the content of the training provides trainees with the knowledge and skills outlined in the course objectives. Evaluations by trainers/participants are widely used in training sessions for service personnel. Observation by an independent observer with expertise on the topic is more common in training of trainer courses.

Evaluations are subject to a courtesy bias, especially if participants doubt the confidentiality of the exercise or if they have developed a positive interpersonal relationship with the trainers over the course of the event. Those administering the evaluation can best reduce this bias if they stress that the answers will remain confidential and that the trainees should not put their names on the evaluation forms.

Indicator

Organization Has the Capacity to Maintain a Functional Information System on Its Training Program

Definition
An organization's ability to use its information system to track its training activity

"Organization" refers to the Ministry of Health, non-governmental organization, or other institutions responsible for training at the national/regional/institutional level. "Capacity" refers to the personnel, software, and other mechanisms required for an information system. "Information system" refers to a database with information (preferably computerized) that allows easy retrieval of key information.

Data Requirements
Evidence of the existence of a functioning system and its use for training-related decision-making

Data Source(s)
Assessment by an external expert

Purpose and Issues
One measure of institutionalization of training capacity is the ability within the local system to document the numerous national/regional/institutional level indicators of the training activity. These include number of trainees, characteristics of the trainers and of the trainees, content of the courses, number of events/methods used, number of contact hours, standards of competence used for different categories, percent achieving those standards, and cost of the training.

In the past, training programs tended to track their "performance" by reporting the volume of activity performed: number and type of people trained, number of courses conducted, number of contact hours achieved, and so forth. This type of "bean counting" may serve certain purposes for local institutions, but the more sophisticated training environment, places less emphasis on these measures of activity and greater emphasis on results achieved.

A training information system (TIS) is designed:

- To track who was trained in what (for the purpose of identifying gaps that remain);
- To assess how organizations selected trainers and which trainees these organizations should select;
- To link with other data sources to measure the effect of training on service delivery; and
- To maximize the cost-effectiveness of the training.

The criteria used by one training organization as benchmarks of progress on establishing a TIS are as follows:

- Criteria are developed/revised to select appropriate participants for specific RH clinical training;
- A TIS has been established at the national/regional/institutional level to document the number of RH professionals trained, in terms of method and cadre (out of the number eligible for or needing training);
- The TIS links training statistics with service delivery information to enable service delivery gaps to be identified; and
- A mechanism exists for monitoring whether adequate numbers of providers are available/being trained for RH service provision, in terms of method and cadre to meet ongoing/changing service delivery needs.

A good TIS allows an institution to: avoid redundancy; match training plans with needs; replace lost capacity due to high turn-over with new personnel; and improve training inputs (e.g., better trainers, improved curricula, best training practices applied).

Indicator

Number of Faculty and Trainers Who Demonstrate the Use of Professional Core Training Competencies on the Job

Definition
"Faculty and trainers" are those persons knowledgeable in the subject area, designated to improve knowledge and skills through the training activities of a given organization. "Use of professional core training competencies" is context-specific. "On the job" indicates that this assessment takes place in an actual work context (when conducting training or providing services).

Data Requirements
A checklist of competencies that the faculty or trainer should demonstrate

Data Source(s)
Observation by an external expert of faculty and trainers performing actual training activities

Purpose and Issues
One important measure of institutional capacity for training is the ability of staff to conduct training activities using state-of-the-art techniques. These include using participatory learning activities, demonstrating and having trainees practice using relevant job aids, summarizing key messages, and using encouragement rather than negative criticism. These contemporary adult learning techniques contrast sharply with the "classroom lecture format" that has characterized training in the past and is far less effective in achieving training objectives among adult learners, especially those with lower educational levels.

The evaluation must take place in an actual training setting, not in a simulated environment. Such a measure will provide the most accurate assessment of the individual's performance in front of a group of trainees, and it will allow for feedback in improving performance in the future.

Evaluators should share the scores from these assessments and discuss with the persons evaluated, so that the faculty and staff can use this feedback to improve their training techniques and thus the quality of training. The organization achieves little if it documents the quality of training without providing feedback to those involved.

Indicator

Organization Has a Systematic Process for Follow-up and Support of Trainees after the Training Event

Definition
The systematic process for "follow-up" refers to the established mechanism that allows the training organization to locate and to communicate with the trainee at specified periods post-training (e.g., six months, one year). "Support of trainees after training" refers to mechanisms that allow the training organization to respond to questions, doubts, or problems that the trained providers experience in the service delivery environment. (Note: this process is part of the continuum of a transfer-of-training process that provides support before, during, and after training.) Refresher training is one mechanism for supporting trainees long after the training event.

Data Requirements
Lists of persons trained; evidence of attempts to contact each individual post training, including the percentage actually reached, and the result of the contact

Data Source(s)
Program records provided by the staff in charge of this activity, to be reviewed by an external evaluator

Purpose and Issues
The new norms for quality training require that organizations follow-up the persons trained in their system, in contrast to the "train and release strategy" used in the past. For example, the USAID-funded training programs stress "Performance Improvement" (described in greater detail in Appendix C). This emphasis requires the training organization to assess gaps in the service delivery environment that hinder or prevent trained service providers from effectively performing their duties. In this spirit, the current indicator reflects the extent to which a training organization remains in contact with its trainees and attempts to identify and to address problems these employees face in the post-training period when they return to the service delivery environment.

Some organizations may prefer to develop a parallel or similar indicator, number of training programs linked to other performance support systems. A performance support system not only ensures such transfer of skills to the job, but also increases the potential for enhanced performance because it enables the provider's work environment to support this transfer of skills. In the context of performance improvement, this link between training and subsequent performance support is essential to insuring a positive experience for the clients in the system. However, relatively little work has been conducted to date in measuring and evaluating this type of linkage. Thus, this indicator of number of training programs linked to other performance support systems is presented as an indicator under development and in need of further testing.

Indicator

Existence of Training Strategy Based on Needs Assessment to Improve Quality of Service Delivery

Definition
"Based on needs assessment" refers to use of a systematic collection of information from multiple relevant sources that indicates the areas in which more service providers require training and the type of service providers who should receive training.

The "needs assessment" describes the existing service delivery system and identifies the gaps between desired and actual performance of providers. It examines the components described below under the training strategy and may specifically focus on one or a limited set of services. Alternatively, it can be (though rarely is) an overarching assessment of the health services system.

This indicator does not specifically measure the effectiveness of the strategy at improving quality, but it relates to the objectives of training programs, which are performance improvement and enhanced quality of care (discussed in greater detail in section Appendix C in this *Compendium*).

Data Requirements
Evidence of a needs assessment conducted and used in developing the strategy; information from those involved in developing the strategy

Data Source(s)
Program records; interviews with persons responsible for the strategy

Purpose and Issues
A detailed training strategy is essential for effective training. Although a training strategy does not guarantee an effective result, the lack of a training strategy suggests ad hoc efforts with little attention to priorities or the felt needs within the system.

The training strategy shows an integrated approach to improving RH service delivery in standardizing and implementing both pre-service education and in-service training, supported by national guidelines/standards. It builds on national RH service delivery needs identified (from government documents and plans) and describes the role of the comprehensive RH training and education system in the context of the sector. In addition to describing the various institutions, organizations and personnel, it includes the sector components of:

- Licensure/certification of providers;
- Provider deployment/job assignment;
- Provider supervision;
- Qualification of trainers/ trainer development; and
- Participant selection criteria.

The training strategy may also include components of a pre-service/in-service reproductive health training program:

- RH curricular component/course schedule;
- Staff/faculty (classroom instruction, clinical practice);
- Training materials;
- Clinical training sites;
- Quality monitoring system; and
- Training information systems.

For a training strategy to be effective, it must have local commitment. Ideally, the leading staff from the training organization will play a key role in developing the training strategy, either alone or in collaboration with external consultants. Without this local input, the training strategy will garner little support from the upper levels at the local organization in question. Rather, they will likely dismiss the strategy as irrelevant, erroneous or externally imposed.

Indicator

The Organization Systematically Evaluates Its Training Program to Improve Effectiveness

Definition
To systematically evaluate its training program an organization routinely applies indicators such as the first five in this section to its training activities. This evaluation requires systematic data collection, analysis, and reporting of the results to those involved in the training.

Data Requirements
A list of all training events; a list of the indicators and instruments used to evaluate them; and a copy of the results

Data Source(s)
Program records; occasional special studies

Purpose and Issues
As training organizations attempt to develop a "culture of evaluation" to improve their programs, this indicator documents the evolution of the trend. It provides concrete evidence that training organizations (or units) are attempting to obtain systematic feedback and to discuss it with those involved in training efforts.

Training evaluation should form part of the training strategy; the institution should have an evaluator on staff or a regular consultant. Training evaluations should systematically examine the capacity of the trainers, their training materials, tools and methods, and the actual evaluation methodology (e.g., whether checklists measure intended areas, whether they need updating, how to adapt tests to different audiences/learners).

Examining the job performance after training – level three evaluation of Kirkpatrick's training evaluation framework (1998) – should take place every two to three years of a regular training program, if possible. In the interim, training trainers to function as evaluators (working with line supervisors), and adapting training tools (knowledge tests, skills checklists) with which to monitor/observe can document trends in changes in performance.

The evaluation of training can take various forms, ranging from the simplest to the most sophisticated. At the very least, training programs will monitor increased learning using pre- and post-knowledge tests. However, few training organizations consider tests an adequate evaluation of the course, and most prefer (where funds permit) to track the skill level of trained providers, both upon completion of the course and at a period X months later (e.g., 6 months, 12 months).

These evaluation methods refer to the individual trainee. In contrast, many of the indicators in this section refer to the organizational capacity of the system to design and implement effective training. Yet to truly evaluate the effectiveness of training, one must link the training activity to improvements in the service delivery environment. The linkage requires a special study using a quasi-experimental design, in which one contrasts a group of clinics whose providers are trained with a group of clinics whose providers have yet to be trained. This type of operations research study is relatively rare because of the resources required and the burden placed on service delivery to maintain "everything else constant." However, those wishing to definitively demonstrate the link between training and improvement in the service delivery environment will need to undertake such studies. Other techniques involve using multivariate analysis, combining data from facility-based and household surveys (e.g., Dietrich, Guilkey, and Mancini, 1998). Short of that, one simply works on the assumption that improving the competency of individual trainers and increasing the number of locations in which they operate will improve quality and access to service delivery.

Results from a recent study in Indonesia (Kim et al., 2000) on reinforcement via self-assessments and support groups of providers indicate that providers lost skills and knowledge acquired through training within six months, except those who performed self-assessment exercises, who actually improved.

Indicator

Demonstrated Capacity to Carry Out Training on a Sustained Basis

Definition
The nature of the "training" depends on the service delivery areas of interest to the organization, but in this case will relate to the different aspects of reproductive health. "On a sustained basis" refers to the demonstrated ability to maintain this activity over a period of time (e.g., 3-5 years) with decreasing external support.

Data Requirements
Evidence of the implementation and monitoring of a long-term strategy; annual training work plans developed in the country/organizational context to meet identified needs; evidence of review, evaluations, and updating

Budget review with percent of funds for training from internal revenues; alternatively, the institution demonstrates capacity to design and obtain funding for training projects, including evaluation. Review of human resources and equipment; list of training activities completed/replicated in the last three years and projected (long-range, strategic) plans

Data Source(s)
Assessment by an external evaluator with training expertise

Purpose and Issues
This indicator measures the ability to continuously provide quality training on a sustained basis with minimal external input – the ideal of most training organizations.

This indicator is more difficult to evaluate than are others in the section because of the subjective nature of the "capacity" and the lack of standard operational definitions for "sustained basis." An alternative indicator that is more concrete and possibly more practical is the number of training sites and centers performing to quality standards on a regular basis with adequate resources, where resources again refer to funding, sufficient staff and trainers, and internal organizational systems. The limitation with this alternative indicator is that one may encounter a single training site that was satisfactorily fulfilling the needs of a country in the area of training, whereas one may encounter multiple organizations (in a large country) that still have many shortcomings in terms of training. In this case, the number does not equate with "adequacy."

Indicator

Adaptability of the Organization/System to Changing Needs in a Training Environment

Definition
"Changing needs in the training environment" are identifiable changes that require the organization to adjust its training procedures. Examples include introduction of a new contraceptive method (e.g., NORPLANT), the growing demand for counseling and testing services in HIV programs, new techniques for cervical cancer prevention, screening for violence against women.

Data Requirements
A list of changes in the service delivery environment requiring adaptations in training over a certain period; evidence of the organization's willingness and ability to respond to those needs

Data Source(s)
Evidence from program records or other sources of regular, periodic meetings to assess needed changes (e.g., at least once every six months); and/or data collected through a special study

Purpose and Issues
For training organizations to be effective, they must be able to respond to changes in the service delivery environment and in their operations. Evaluators may have difficulty charting an organization's progress in this area, precisely because no objective list of changing conditions in the service delivery environment exists. Moreover, in any given list of changes, some items may be relatively trivial compared to others that have wide-ranging public health implications. Thus, both types of changes cannot receive equal weight.

This indicator is particularly appropriate in the context of an overall assessment of an organization in terms of its training capacity, conducted by an external evaluator with expertise in the training area. The assessment requires an understanding of the local delivery context and cannot take the form of a simple "checklist" or summation of points.

Examples of an organization's adaptability to changing needs in a changing environment include:

- Trainers from the Zimbabwe National Family Planning Council developing a postabortion care training package in 1995; and
- The Center for African Studies' (CAFS) developing and successfully soliciting donor support for a program for AIDS home care in 1999.

Part II.E Commodities and Logistics

- Accuracy of logistics data for inventory management
- Percent difference between consumption forecasts and actual consumption
- Existence of a multi-year procurement plan for each product offered
- Commitment by all stakeholders to carry out established procurement plans for each product
- Percent of facilities that maintain acceptable storage conditions
- Percent difference between the quantity of products ordered and the quantity actually received
- Percent of facilities that experience a stockout at any point during a given time period
- Percent of facilities whose stock levels ensure near-term product availability
- Logistics System Assessment Tool – Qualitative Indicators

Part II.E

Commodities and Logistics

The system for obtaining adequate quantities of contraceptives and other reproductive health supplies and for delivering them to service delivery points (SDPs) constitutes a critical element of family planning and reproductive health operations. Without the products that clients need and without the logistics systems to provide them, no program can expect to meaningfully improve the reproductive health of the people it serves. In short:

No product? No program

As shown in Figure II.E.1, commodities are among the key inputs to any reproductive health program, and logistics systems are among the key processes that enable program success. The figure shows the main mechanisms and sub-components of logistics processes, including logistics management, policies, human and organizational capacity, and financial resource mobilization. It also shows how logistics processes and functional outputs relate to the overall reproductive health conceptual framework shown in Figure II.E.1. These processes and outputs result in product availability to clients – the main direct result of effective logistics systems. Logistics may involve family planning products exclusively (in categorical programs), or an expanded range of reproductive or other health products, as is the case with increasing frequency in many countries. When a program mobilizes human, technical, and financial resources – with a minimum of external assistance – so that the program consistently ensures product availability, access to services, and quality of care consistently in a way that meets clients' needs, the program achieves contraceptive/commodity security. To the extent that logistics systems improve product availability and contribute to commodity security, they also contribute to increased use of reproductive health services and ultimately to improved health outcomes.

Ensuring product availability requires attention to six rights: the *right goods*, in the *right quantities*, in the *right condition*, delivered to the *right place*, at the *right time*, for the *right cost*. As shown in Figure II.E.2, the logistics system is often depicted as a cycle with components of product selection (the right goods), forecasting and procurement (the right quantities, cost), inventory management and distribution (right place, time, and cost), and provision to customers (right place, time, and cost). Information for decision-making is central to the cycle, and quality assurance and monitoring take place throughout. Meeting the needs of end users is the ultimate goal of reproductive health logistics systems, and attention to all six rights is essential to that effort.

Monitoring and evaluating logistics system performance can help managers, donors, and other stakeholders better understand this essential program component and identify ways to improve it. When using the indicators in this section, evaluators should consider the challenges present in several features of logistics systems. Some issues apply to logistics systems in general, and some are unique to integrated reproductive health systems in today's changing health care environment.

Considerations relevant to the evaluation of *any* logistics system include the following.

Methodological Challenges of Evaluating Commodities and Logistics

- **As with many other program components, the causal relation between logistics system improvements and health outcomes is complex and largely indirect.**

Many factors besides commodities and logistics contribute to long-term health outcomes. Although proving the magnitude of the contribution made by effective logistics systems is rarely feasible, it is highly plausible that better systems and increased product availability enable increased use and improved health. But it is beyond the scope of most evaluations to confirm this scientifically.

Figure II.E.1. Conceptual Framework for Logistics, Commodity Security, and Reproductive Health Outcomes

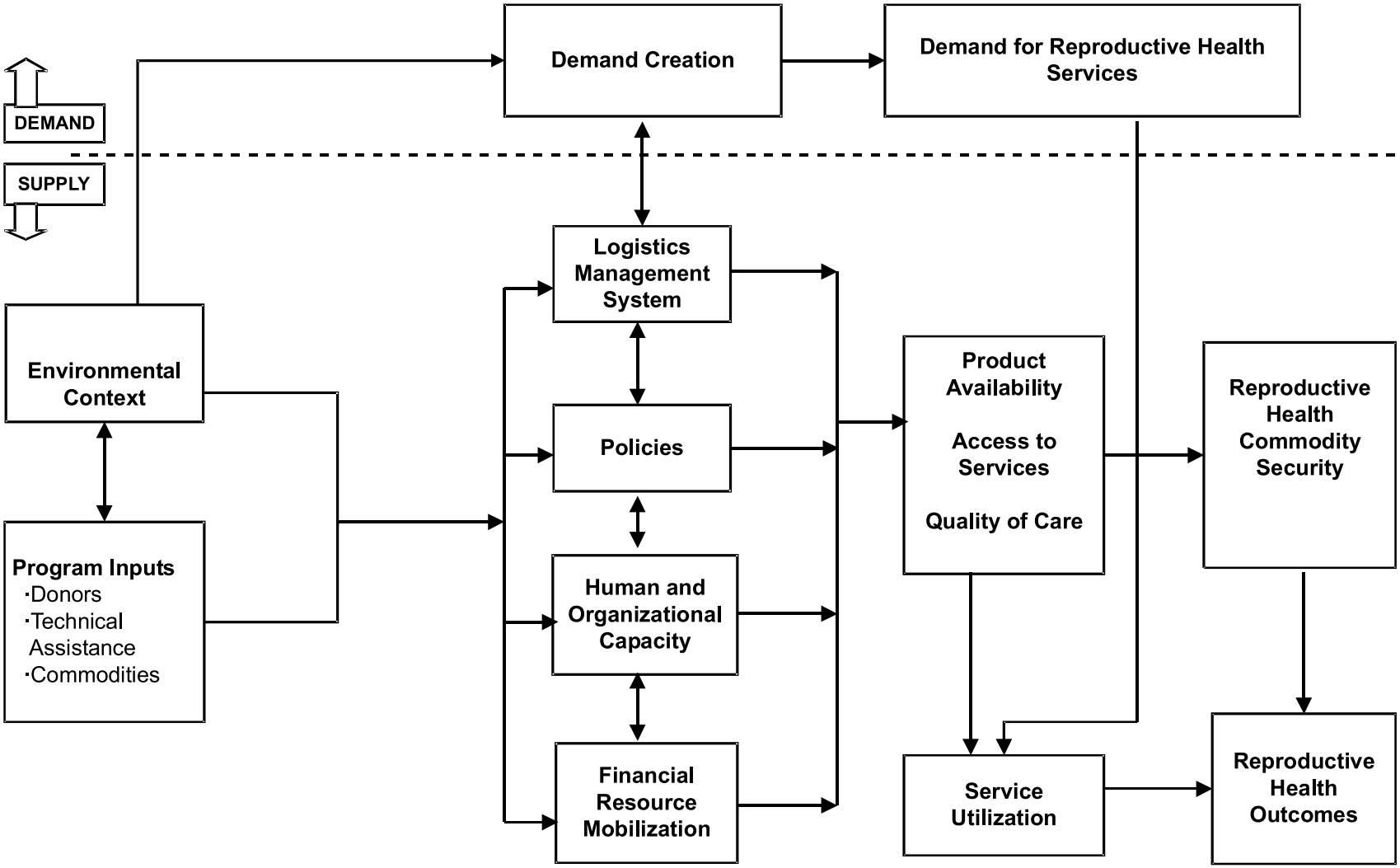
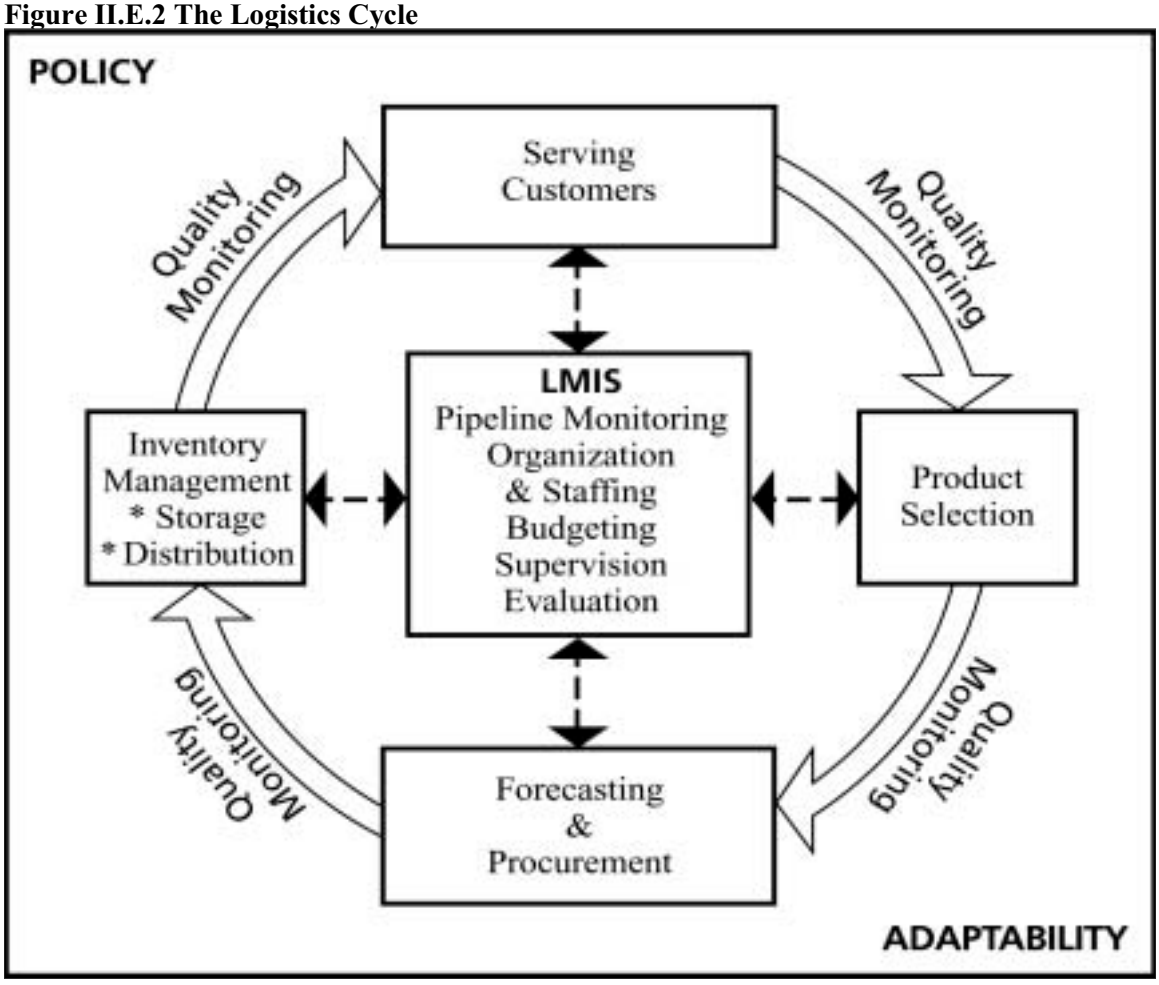


Figure II.E.2 The Logistics Cycle



Source: FPLM/ John Snow Inc., 2000.

- **Logistics indicators (especially stockout frequency and adequate stock levels) are interrelated and should be used together; interpreted separately, they can result in misleading conclusions.**

If evaluators apply the stockout indicator alone, for example, it may not reveal whether products are actually available *to clients*. If reducing stockouts is a strong programmatic priority, service providers may hoard or may ration products to avoid running out. This practice may indeed minimize stockouts, but the result to the client is still the same – no product. When evaluators apply the stockout indicator with adequate stock levels, however, they minimize counterproductive results. The stock status indicator will reveal whether a product is overstocked or under-stocked at any given time and site, an indirect indication of whether rationing is occurring and a direct measure of whether products are actually available when clients need them.

- **Evaluators should interpret logistics indicators in relation to other reproductive health indicators.**

Ensuring that supplies arrive at their intended destinations is not the only objective of a logistics system. Product needs change depending on programmatic interventions, and logistics managers need to constantly communicate with program managers to ensure that supplies go where they can be dispensed appropriately. For example, logistics must be coordinated with training. Clearly, contraceptives such as IUDs, injectables, and Norplant should go only to sites where trained providers dispense them, and where at least potential demand for the product exists. As family planning programs are integrated with a broader range of reproductive health services, these issues become important for an ever-increasing number of products, including STI and HIV test kits, STI drugs and anti-retrovirals, vaccines, and others. Unless logistics and programmatic activities are well coordinated, programs run the risk of expired prod-

ucts, stockouts, inadequate service provision, improper use of products, and ultimately, worsened health outcomes.

- **The measurement of some indicators requires specialized logistics knowledge and on-site evaluation.**

Complete logistics assessments usually require site visits by a logistics system expert. In a typical assessment, M&E specialists may be responsible for the design of tools, sampling strategy, data entry, and analysis, while the actual application of the tools is carried out by logistics experts who may or may not have M&E experience. For example, though in theory storage conditions can be self-monitored, with storage indicators captured through supervision and MIS systems, in practice, site visits by a logistics expert may be the only way to get accurate information. Moreover, a complete logistics assessment (using all the suggested indicators) requires visits to different kinds of sites at different levels, including host-country organization offices, central and district warehouses, and service delivery points, with different information collected at each. This requirement makes monitoring and evaluation of logistics systems potentially more resource intensive than the monitoring and evaluation of other program components.

If these considerations did not make the evaluation of logistics programs difficult enough, recent changes in service delivery strategies further complicate the process. Until recently (and even in many cases today), family planning logistics systems have served categorical (or vertical) programs. In such cases, those logistics systems had to ensure that a small number of contraceptive products reached their intended supply points. Monitoring and evaluation could focus on the effectiveness of those systems at achieving this relatively straightforward goal.

Two major changes in recent years, however, have led donors, cooperating agencies, host-country governments, and logistics managers to rethink the way they manage family planning logistics systems. First, the 1994 United Nations International Conference on Population and Development (ICPD) mandated a broader, integrated approach to family planning within a reproductive health context, and a focus on client rights as opposed to national demographic objectives. Second, health sector reform efforts have accelerated in developing countries around the world. This acceleration, like ICPD, has led to integration, but more to enhance efficiency and economy than to further individual rights. Health sector reform has also led to decentralization and privatization; both of which have created challenges and opportunities for logistics systems as well as for other components of national reproductive health programs. All these changes have occurred in a period of shifting donor commitment that requires greater coordination among all stakeholders, greater ability of logistics systems to adapt to differing donor procedures, and greater emphasis on demonstrating measurable results from donor inputs.

These changes have created new methodological challenges for evaluators of logistics systems in integrated reproductive health programs.

- **With the broadening of the family planning mandate to embrace a reproductive health perspective, logistics systems must manage an increased number of products.**

The number of products alone is not necessarily the problem — automated logistics systems can easily manage thousands of products. But larger systems depend on computers, and most developing country programs do not yet have the resources or capacity for automation. To put the challenge in perspective, categorical family planning logistics systems typically manage fewer than 20 - often fewer than 10 - distinct products, some with multiple brands. When family planning is integrated with programs such as HIV/AIDS, maternal and child health, integrated management of childhood illness, and malaria, the number of products expands proportionately - and the complexity of the system expands exponentially. As just one example, HIV/AIDS programs use products such as test kits, Nevirapine (for reduction of mother-to-child transmission of HIV), antiretroviral drugs, STI drugs, drugs for the treatment of opportunistic infections, home-based care kits, and all the materials needed to provide those products. The increased number and variety of products requires M&E systems to capture, analyze, and manage far more data than they ever have before. Integration may also result in a merging of logistics and health management information systems, a process that can result in the loss of essential logistics data. In practice, the best way to prevent such loss is to maintain and to manage logistics MIS separately from health MIS. If separation is impossible and the two systems merge, evaluators and logisticians should set up reporting systems to ensure that logistics information is easily and continually accessible to those who need it.

- **Beyond sheer numbers, additional products also pose special challenges in logistics forecasting, procurement, storage, and distribution.**

Because the demand for most other products is not as well known as the demand for contraceptives, forecasting needs accurately is difficult. Many products have short shelf lives and thus require more precise procurement planning and inventory management. Because logistics systems integration and decentralization complicate storage and transport, systems that were adequate in a categorical family planning program may fail as the program integrates. Thus, evaluators must measure and interpret logistics indicators in ways that result in appropriate conclusions.

For example, HIV/AIDS condoms, while functionally equivalent to family planning condoms, have a number of unique characteristics that require new ways to manage distribution and new evaluation strategies. In addition to the increased quantity needed (which affects storage and transport decisions), HIV/AIDS condoms are typically dispensed through different outlets from traditional family planning sites. Sites for HIV condoms potentially include everything from bars and brothels to markets, schools, work places, truck stops, barber shops/ beauty salons, and many more. From a monitoring and evaluation perspective, such sites may be far less prepared than family planning sites are to record and report data necessary to measure indicators.

- **Many additional products now being managed are in chronic short supply, with distribution systems based on budgetary constraints or rationing policies rather than on need.**

Family planning products have typically enjoyed strong support from donors, so that supplies are adequate to meet expected demand. Family planning logistics systems, therefore, operate on the assumption that contraceptives are "full supply" products, meaning that in a well-functioning system they should always be available. This assumption allows managers to set maximum and minimum desired inventory levels, and to try to maintain the amount of each product within that range. When a full-supply product such as a contraceptive method is within its planned "max-min" levels, it is said to be "stocked according to plan," and its stock levels are said to be satisfactory.

These measurements do not apply, however, to many essential drugs and other products purchased by national health ministries or provided by donors with insufficient budgets. Such products are purchased too infrequently, or in insufficient quantities, to prevent stockouts between procurements. Max-min inventory control systems by definition can not apply to non-full supply products, and both logisticians and evaluators must treat the desired minimum stock levels less rigidly.

Evaluators must address many factors to fully evaluate logistics system performance, and many of these factors require qualitative assessment, and expert working groups in the past have suggested a two-pronged approach using both quantitative and qualitative assessments. In response, the USAID-funded DELIVER project has developed separate tools to carry out each type of assessment: the Logistics Indicator Assessment Tool (LIAT) and the Logistics System Assessment Tool (LSAT).[1] The LIAT gathers a relatively small number of quantitative indicators to measure key output results that demonstrate whether or not the logistics system is performing well. The LSAT, on the other hand, contains quantifiable sections allowing for monitoring of changes over time, but it serves mainly as a qualitative diagnostic instrument that describes the overall system and helps identify underlying reasons for each system's strengths and weaknesses.

This section first presents the eight quantitative indicators, followed by a description of the Logistics System Assessment Tool. Ideally, evaluators should collect all the indicators as a package, providing a comprehensive picture of the characteristics of a logistics system and its performance, but in practice, not all programs can carry out such a complete assessment. In such cases, measuring any of the indicators individually is still worthwhile, with previously mentioned caveats in mind. The choice of indicators to measure will depend on program objectives, available resources, or other factors. When choosing, evaluators should recognize that product availability (stockout frequency = zero) is the most vital logistics result from the client's perspective, so in that sense, it may be the "most important" indicator. Since stockout data are usually collected through facility surveys, however, evaluators may simultaneously collect data on stock data quality, storage conditions, local forecast accuracy, order fill rates, and stock sta-

[1] Both the LIAT and LSAT questionnaires are provided in Appendix D and E, respectively. These tools are also available on-line at www.deliver.jsi.com

tus. The LIAT can be used to collect such data at the facility level, the application of which can be used to measure five of the core indicators described in this section. Data on national level forecast accuracy, procurement planning, and stakeholder commitment, on the other hand, are obtained centrally and may therefore be less costly to collect. Finally, if one wants to understand the strengths and weaknesses of a logistics system, especially from a qualitative perspective, the LSAT provides the most useful information. In sum, the indicators will be most useful when measured as a package, but if available resources do not allow for this strategy, measuring any of them individually should provide useful information.

Despite the challenges described above, the logistics indicators proposed in this chapter are similar to those used in the past by categorical family planning programs. Logistics systems —whether categorical or partially or fully integrated, large or small, health or non-health — all share fundamental characteristics. All depend on quality data to forecast needs and to order products in the right quantities. Products must be stored and transported effectively, with a minimum of wastage. Inventories should be managed in a way that minimizes the likelihood of wastage or stockouts and that maximizes product availability. Because the indicators suggested here measure all of these elements of a logistics system, they allow managers to better understand its strengths and weaknesses, and to implement interventions that ensure products are available to customers. Thus, the indicators ultimately contribute to contraceptive/reproductive health commodity security and improvements in reproductive health of the communities the program serves.

Indicator

Accuracy of Logistics Data for Inventory Management

Definition
For each method/brand/product of interest, evaluators measure accuracy of logistics data by comparing (1) the physical stock to the recorded count of that stock, and (2) the recorded stock count to the count in reports produced for logistics management information system (LMIS). The discrepancy is expressed as a percentage, as follows:

(1) Accuracy in keeping stock records (see Appendix D, Table 33A, Column 7)

$$\frac{\text{Stock record count} - \text{physical stock count}}{\text{Physical stock count}} \times 100$$

(2) Accuracy in transferring information to the LMIS form (see Appendix D, Table 33b, Column 4)

$$\frac{\text{LMIS report count} - \text{stock record count}}{\text{Stock record count}} \times 100$$

Physical stock, stock record, and LMIS report counts refer to the amount of each product that is shown as undamaged, unexpired, and available for use in a service delivery facility or warehouse. Physical stock count is determined by actually counting the stock in the store. Stock record count is recorded on forms that specific facilities use to track stock balances, transactions, and adjustments over time. LMIS report count is recorded in periodic reports that summarize stock balances, transactions, and adjustments over a period of time, and that are transmitted from one level of the system to another.

Evaluators can report each measure of discrepancy (or agreement) by facility or in the aggregate, and should report for each product of interest. It may also be useful to use these measures to calculate the percent of facilities that keep accurate stock records and produce accurate reports (defined as reports showing that discrepancies for all products fall within a margin of error agreed to by the program).

Data Requirements
Physical counts of total number of products in the facility; recorded inventory, which can be retrieved from the stock ledger or from stock cards; and LMIS reports.

Data Source(s)
Facility survey/logistics site visits to all facilities or to a representative sample

Purpose and Issues
This indicator measures the accuracy of data on product stock levels at various points in the logistics system. Since the supply chain relies completely on stock data to forecast, procure, and deliver the right quantities of product to storage and service delivery sites, this indicator is essential. It highlights the importance of data quality down to the lowest level of the system. The first part of the indicator provides information on whether facilities are accurately tracking their inventories, while the second part tells whether this information is accurately transferred to LMIS reporting forms. The first part uses information on stock levels on the day of the site visit, while the second part compares the most recent available LMIS report to the inventory record balance closest to that date. Since the latter measure requires reviewing historical stock records, evaluators may have difficulty collecting these data. This indicator may also check for leakage in the system, track timeliness in updating stock records, and determine the extent to which programs complete and submit LMIS reports.

Ideally, a program should have no discrepancies between the physical inventories and the two sources of stock level data, but in practice, evaluators should expect some errors. Acceptable levels of error will depend on conditions in each country. In general, discrepancies of over ten percent should cause concern and likely warrant efforts to improve data quality.

Related indicators:

- Percent of facilities that keep accurate logistics data for inventory management
- Percent of facilities that completed and submitted an LMIS report for the most recent reporting period

Indicator

Percent Difference between Consumption Forecasts and Actual Consumption

Definition
For all products that the program has committed to supplying, the percentage difference between forecasts previously made for a year (or other appropriate time period) and the actual consumption or issues data for that period. Evaluators should calculate the indicator for each product for which a forecast is made. If evaluators desire a mean forecast accuracy figure for all products, they should base it on the absolute values of the discrepancies calculated for each product.

This indicator is most commonly measured annually at the central level, but it can also be applied more frequently at lower levels of the system as a measure of facilities' capacity to determine their own order quantities. In either case, the basic formula is the same

This indicator is calculated as:

$$\frac{\text{Forecast consumption} - \text{Actual consumption}}{\text{Actual consumption}} \times 100$$

(See Appendix D, Table 34)

Data Requirements
List of products that the program has committed to supplying; forecasts or order requests by product for the period of interest; and actual consumption or issues data by product for the period of interest

Data Source(s)
Logistics data from LMIS reports, plus (at the national level) key informant interviews; records reviews; demographic surveys; and/or service statistics

National level forecasts and the list of products should come from government or other sources — e.g., Contraceptive Procurement Tables (CPTs) for USAID-supplied contraceptives, recommended orders to donors for essential drugs, or a government forecast of Vitamin A tablets. At lower levels of the system, the "forecasts" would be represented by order requests to the next higher level. Evaluators may obtain consumption data from LMIS reports at any level and at the national level may estimate consumption from demographic surveys or service statistics. They can obtain CPTs or national level forecasts by product through the local USAID Mission, from the USAID-funded DELIVER Project, or from host-country program managers for contraceptive products that USAID supplies. At lower levels of the system, LMIS forms can be used to obtain data on order quantities requested and consumption/issues for each facility in the most recent order period.

Purpose and Issues
At all levels of the system, accurate forecasting helps countries and organizations order the right amount of each commodity, thereby reducing the likelihood of wastage or shortages and increasing the likelihood of meeting client needs with available products. A forecast made using past consumption data and sound forecasting methodologies should approximate actual consumption within a margin of error appropriate for each product. Host-country stakeholders should agree on the allowable margin of error, and evaluators should interpret results in light of real world conditions that may have been impossible to foresee. Forecasts are subject to uncertainty for many reasons, so some errors must always be accepted, particularly at the national level when the forecast period is long. Documenting the reasons for particularly wide discrepancies (including assumptions used in preparing the forecast) helps put the results in perspective and may provide insights for improving future forecasts.

This indicator also indirectly measures data quality, since an accurate forecast can only result if the data used are of good quality.

Related indicators:

- Mean level of forecast accuracy/discrepancy for a range of facilities and/or products; and
- Percent of facilities with forecasts within 5 percent of actual consumption, by product.

Indicator

Existence of a Multi-Year Procurement Plan for Each Product Offered

Definition
For each product procured by a program, a multi-year procurement plan prospectively ensures that the product will be in stock throughout the period, and, for full-supply products, does not result in stock quantities exceeding established maximum levels. This is a yes/no indicator for each product. Ideally, procurement plans should cover three or more years, particularly for donor-supplied products, which may have very long lead times. However, given the practical realities of synchronizing procurement plans with donor budget cycles, evaluators should score a plan that meets the above criteria for at least two years as satisfactory. The indicator is scored affirmatively if columns 7 and 8 in Table II.E.1 show positive stock levels throughout the period of the plan and if stock levels for full-supply products do not exceed the established maximum level.

Data Requirements
List of products that the program has committed to supplying; procurement plan for program-wide product requirements; program-wide stock levels for each product; established maximum levels for each full-supply product; and existence of a forecast.

Data Source(s)
Key informant interviews and records review

The procurement plan should be available at the level where procurement decisions are made. Evaluators can use Table II.E.1 to determine the indicator.

Purposes and Issues
This indicator measures whether the program has adequately planned the procurement and shipping schedule for the products that it has committed to supplying and for which it has forecasted needs. Given budgetary realities, program managers may not have plans for many non-full-supply products (or the plan may be inadequate to avoid stockouts or expiration), but the indicator should measure all products nevertheless. The indicator is useful for showing policy makers and donors where shortfalls or oversupply may occur, and for advocating changes to avoid such supply imbalances. It can also be used to check whether procurement schedules make sense given factors such as shelf life of products and storage capacity relative to patterns of use.

The indicator shows only whether the plan is consistent with historical use and anticipated future consumption patterns at a particular point in time. It does not measure whether the plan is actually carried out. To determine whether products are in fact procured/delivered as planned, one must monitor over time to see whether shipments of goods actually arrived at the expected times and in the right quantity. Program staff can manually monitor small numbers of products, but may require an automated tracking system (e.g., PipeLine, available from www.deliver.jsi.com) in programs managing many products.

The indicator may measure a national program, or individual programs or levels where procurement plans are prepared.

Related indicators:

- Existence of a procurement plan for each product (whether or not it maintains adequate stock);
- Percent of shipments that arrive on time; and
- Percent of shipments that arrive in the right quantity.

Table II.E.1 Instrument for Measuring Adequacy of Procurement Planning

Established Maximum Stock Level (if applicable) = _____ Months

Month (1)	Beginning Balance (2)	Quantity Received (3)	Supplier (4)	Estimated Consumption (5)	Stock Adjustments (6)	Ending Balance (7)	Stock in Months (8)
01-Jan							
01-Feb							
01-Mar							
01-Apr							
01-May							
01-Jun							
01-Jul							
01-Aug							
01-Sep							
01-Oct							
01-Nov							
01-Dec							
02-Jan							
02-Feb							
02-Mar							
02-Apr							
02-May							
02-Jun							
02-Jul							
02-Aug							
02-Sep							
02-Oct							
02-Nov							
02-Dec							

Guidelines for Completion

- In Column 6, "Stock Adjustments," includes the projected amount of lost, expired, and damaged product for each month.

- Column 7, "Ending Balance," is calculated by adding columns 2 and 3, subtracting column 5, and adding or subtracting column 6, depending on the direction of the adjustment.

- Column 8, "Stock in Months," is calculated by dividing the ending stock balance (column 7) by the projected average monthly consumption (normally calculated as the average of the previous three, six, or twelve months of consumption from column 5). This calculation is best accomplished with PipeLine software (available on the Internet at deliver.jsi.com) or a computerized spreadsheet.[2]

[2] Table II.E. 1 is adapted from John Snow, Inc.'s (JSI) PipeLine software, which is available on the Internet at www.deliver.jsi.com.

Indicator

Commitment by All Stakeholders to Carry Out Established Procurement Plans for Each Product

Definition
For each product procured by a program, all stakeholders – donors, lenders, and program managers – have committed to carrying out the established procurement plan. Commitment from donors should ideally be in writing. Further, all commitments should include a planned shipment schedule consistent with the plan, and a confirmation of the budget allocation for the product. All of these conditions should be met before evaluators score the indicator as "yes."

This is a yes/no indicator for each product (though it could potentially be scaled, if so desired). Evaluators may wish to qualitatively assess the strength of commitment to determine whether shipments are likely to occur as promised and as scheduled. Evaluators can assess this commitment through key informant interviews.

If no procurement plan exists for a given product, this indicator does not apply.

Data Requirements
List of products the program has committed to supplying; procurement plans for program-wide product requirements, by product; and acceptable evidence (or lack thereof) of commitment to procurement, including shipment schedules and documented planned budgetary allocations.

Data Source(s)
Key informant interviews and records review of procurement planning documents at the level(s) where forecasting and quantification exercises take place

The product list, planned shipment schedules, and information about budget line item for products should also be available at the central level for each program.

Purpose and Issues
This indicator measures whether key stakeholders are committed to the procurement plan. While logistics planners may not be able to control timing of actual product deliveries, they should at least secure commitments from appropriate stakeholders and should follow up with donors to increase the likelihood that shipments will arrive as planned. Gaining such commitment is an essential component of reproductive health product security.

Gender Implications of this Indicator

Reproductive health security and contraceptive supply can be jeopardized at many points along the logistical supply system. Lack of commitment to maintain supplies of commodities that not only prevent unwanted pregnancies but also may save lives or be essential to maintain health are gender and human rights issues for which those responsible for keeping these commitments are accountable. Numerous "situation analyses" of service delivery sites conducted by the Population Council's operations research program found alarmingly high levels of stockouts of condoms in countries with high HIV prevalence. Other studies have shown that many service providers do not present condoms as a contraceptive option to clients because these providers prefer to promote long-term or more highly efficacious methods. For women who need protection against disease as well as against pregnancy, an uninterrupted supply of condoms is essential. Increasingly, as supply systems must include drugs such as antiretrovirals to prevent mother-to-child transmission and to treat AIDS, lives depend on the commitments made by each party responsible for maintaining supply — from donors and program mangers to supply and stocking personnel.

Indicator

Percent of Facilities That Maintain Acceptable Storage Conditions

Definition
Evaluators should report this indicator for each condition (see Appendix D, Table 35).

This indicator is calculated as:

$$\frac{\text{\# of storage facilities meeting each acceptable storage condition}}{\text{Total \# of facilities reviewed}} \times 100$$

Data Requirements
Checklist of acceptable storage conditions; and data collected for each condition for all facilities or for a representative sample of facilities by an observer knowledgeable about storage requirements

Data Source(s)
Facility survey/logistics site visits

Purpose and Issues
This indicator measures the conditions of storage facilities compared to a list of conditions required to protect the integrity of products. Evaluators can apply the indicator at each level of the logistics system to identify facilities that need improvement.

Evaluators should use the first part of the checklist (see Appendix D, table 35, items 1-13) to assess all storage facilities (including small storage spaces at the SDP level), while they should apply the second part (items 14-18) to larger facilities as appropriate.

Related indicator:

- Percent of facilities meeting all (or a desired percent) of the storage conditions

Indicator

Percent Difference between the Quantity of Products Ordered and the Quantity Actually Received

Definition

For each product that the program is committed to supplying, the percentage difference between the quantity ordered the last time an order was placed and the amount actually received. The indicator should be calculated separately for each product for which orders are placed between levels of the logistics system, or for which national procurement orders are placed. It can be calculated at the individual facility level, for different levels of the system, or for the country as a whole. In most cases, results will be aggregated across many facilities and orders, and in such cases, the indicator should be calculated using the sum of the absolute values of the order discrepancies for each product, divided by the sum of all order quantities.

$$\frac{\text{Sum}|\text{ quantity received} - \text{quantity ordered}|}{\text{Sum quantity ordered}} \times 100$$

Data Requirements

List of products that the program has committed to supplying or a predetermined subset of this list; amount of products requested the last time an order was placed for all facilities or a representative sample of facilities, or the amount of products ordered for the most recent national procurement; and the amount of products actually received in response to the last order or procurement

Data Source(s)

Facility survey/logistics site visits, or procurement records for national procurement orders

Purposes and Issues

This indicator measures the efficiency of a supply chain in ensuring that products reach their destinations in the quantities requested and on schedule. It can be calculated for the supply chain as a whole or for any level or facility that receives supplies based on an order to a higher level. The information can reveal which products are frequently under- or over-supplied, which types of deliveries are most apt to be on time or delayed, and which suppliers or distributors are most/least reliable. Managers and evaluators can use this information to take corrective actions and improve supply chain efficiency.

Caution should be exercised when interpreting the indicator for non-"full supply" products. Many of those products are rationed due to limited resources, so it is to be expected that they would experience greater order discrepancies than full supply products. Such discrepancies, however, may be more due to lack of funds to procure supplies than to inefficiencies of the supply chain.

Related indicators:

- Percentage of all orders that are completely filled and on schedule;
- Average duration of time between the date an order was placed and when it was received;
- Percentage of facilities that received their last order completely filled; and
- Percentage of facilities that received their last order according to schedule.

Indicator

Percent of Facilities That Experience a Stockout at Any Point During a Given Time Period

Definition

This indicator measures the percent of facilities (service delivery points, warehouses) that experienced a stockout – of a method/brand/ product expected to be provided or issued by that site – at any time during a specified period (e.g., the past 6 or 12 months). Evaluators should collect the indicator at all (or a sample of) facilities that distribute or issue products, should calculate the indicator separately for each product, and should aggregate it to calculate the percentage of facilities that experienced a stockout of each product at any time during the period. Evaluators may use Table 32, column 1, in Appendix D, to tabulate data required to measure the indicator.

This indicator is calculated as:

$$\frac{\text{\# of facilities assessed that experienced a stockout of a (}\textit{method/brand/product)}}{\text{Total \# of facilities assessed that distribute or issue (}\textit{method/brand/product)}} \times 100$$

Data Requirements

Information on stock levels of all products of interest for the past 6 (or 12) months at all levels of the system

Data Source(s)

A facility survey/logistics site visit – at all facilities or a representative sample – is usually necessary. In some countries/programs, evaluators may use logistics management information systems or supervisory records, depending on the quality of the information.

Purpose and Issues

This indicator measures product availability (or lack thereof) over a period of time, and serves as a proxy indicator of the ability of a program to meet clients' needs with a full range of products and services. Evaluators should use this indicator in conjunction with the stock status indicator and interpret it with caution, because facilities can avoid stockouts by rationing supplies. Other related indicators (see below) may shed additional light on overall product availability. For example, duration of stockouts may help differentiate between products stocked out for a short period of time (e.g., 1-2 days) versus those stocked out for extended periods. Evaluators may assess reasons for stockouts to help program managers address the underlying causes of this logistics system failure.

If national policy dictates that different brands of the same product cannot be used interchangeably, then evaluators should monitor brands separately. If the policy allows substitution of equivalent brands, and if providers make such substitutions in practice, then evaluators can monitor different brands as a single product.

Using data for a 12-month period allows evaluators to consider seasonal variations in product use, but they may have difficulty obtaining the historical data. Calculating this indicator using data for 6 months is less cumbersome because it requires reviewing fewer reports. If evaluators rely on fewer than 12 months of data, they should investigate seasonality issues.

Related indicators:

- Mean duration of stockouts;
- Percentage of facilities stocked out of any product on day of visit;
- Percent of facilities fully stocked (all products) on the day of visit;
- Mean number of methods stocked out/in stock on day of visit;
- Percentage of products stocked out/not stocked out at any time during past 6 (or 12) months; and
- Mean number of times each method was stocked out in the past 6 (or 12) months.

Indicator

Percent of Facilities Whose Stock Levels Ensure Near-Term Product Availability

Definition
This indicator measures the percent of facilities with stock levels greater than zero and below the established maximum level for each full-supply method/brand/product of interest at a point in time (e.g., the day of visit). Where stock levels are greater than zero but below the established minimum level, evaluators must find an outstanding order for replacement stock, made at or before the time stock levels reached minimum.

This indicator is calculated as:

$$\frac{\text{\# of facilities that have stock levels above zero but below the established maximum level for the product}}{\text{Total \# of facilities reviewed}} \times 100$$

(See Appendix D, Table 31, column 12)

Evaluators can report the indicator at the facility level or aggregate it for a sample of facilities or for the entire program. At any level, evaluators should calculate and report the indicator separately for each product of interest so that each product receives a unique measure. If so desired, evaluators can further aggregate to construct additional indicators, such as the percent of facilities with *all* full-supply products adequately stocked. (See related indicators below.) Averaging all products for an "average" stock level adequacy is not recommended, because oversupply in one product can cancel out undersupply in another, and thus falsely imply that average stock levels were adequate.

Data Requirements
Stock levels of all products of interest at a point in time (e.g., the day of the visit); maximum and minimum stock levels established by the program; historical consumption or issues data for each product at each facility; and records of recent orders (for products below minimum levels)

Data Source(s)
A facility survey/logistics site visit – to all facilities or to a representative sample – is frequently necessary to assess stock levels. Evaluators may collect stock data by taking a physical inventory or by reviewing the stock ledger or stock cards. In some countries/ programs, the LMIS or supervisory/ staff records may provide usable stock-level data. The LMIS should also provide maximum and minimum stock levels along with consumption data by product. Service statistics or similar records may provide the needed data on consumption or issues if the LMIS does not.

Purpose and Issues
This indicator provides an overall measure of whether stock levels of products are adequate at a point in time. It helps reveal overstock situations that could lead to product expiration and wastage, and low stock levels that could result in stockouts or rationing. In applying this indicator, evaluators must carefully evaluate facilities where stock quantities are below established minimum levels. To do so, the evaluator should determine whether a new order was placed when stock levels reached minimum. If such an order is outstanding, then the evaluator may consider stock status adequate, because the order will likely arrive before the facility stocks out. If not, the stock status is inadequate.

Evaluators should apply the indicator only to products the program has committed to keeping in full supply, because stock status at a point in time for non-full-supply products may reflect only the length of time since the last shipment arrived rather than measuring whether inventory management procedures are effective. Ideally, evaluators will measure stock status over a period of time (see related indicator, below), but this approach is usually possible only where the LMIS is automated.

Related indicators:

- Percent of time during a given period that each product of interest is adequately stocked (this indicator requires an automated LMIS system or extensive review of historical stock ledgers); and
- Percent of facilities with all full-supply products adequately stocked for near-term availability.

Indicator

Logistics System Assessment Tool (LSAT) – Qualitative Indicators

Definition
The Logistics System Assessment Tool (LSAT) allows a comprehensive system-level assessment of the performance of a logistics system for any health program managing any health commodity. Evaluators can use the tool with the other logistics indicators in this section to completely assess a program's ability to continually provide health commodities at service delivery points.

Data Requirements
System-level scores for each component and each subcomponent; and detailed information about logistics system processes

Data Source(s)
Evaluators should complete assessments by consensus among program managers or among others with knowledge in logistics management and system operations and performance.

Purpose and Issues
If the programs to be assessed are separate and vertical, the evaluator should complete this tool separately for each program. The instrument is organized according to the components of the logistics cycle shown in the introduction to this section: logistics management information system, product selection, forecasting, obtaining supplies/procurement, inventory control, warehousing and storage, transport and distribution, product use, finance, and organization/staffing. Each section contains a series of objective and quantifiable yes/no questions, as well as open-ended qualitative questions that explore strengths and weaknesses of the logistics system in detail.

The combination of yes/no and qualitative questions allows evaluators to use the tool for both monitoring and diagnostic purposes. Evaluators can average and score quantitatively the yes/no questions in each section to assess progress and improvements in a given logistics system over time, whereas the qualitative ones can help evaluators more clearly understand the system's strengths and weaknesses. Qualitative questions should also reveal the causes of areas of weakness and potential ways to improve them.

The main issue to consider when using the LSAT (from a monitoring perspective) is its potential subjectivity. If evaluators use a consensus exercise to answer each question, experts may disagree about what the response should be. In general, most yes/no questions are as objective as possible, so evaluators should easily find evidence to support one side or the other. In some situations, a condition may be met at some levels of the system or in some regions of the country, but not in all. In general, most questions require that the condition be met at all levels (central down to SDP) to receive a "yes" answer, so again, most scoring should be unambiguous. Finally, the question of reliability of scores over time may be an issue, given that different facilitators or participants may be involved (though evaluators are advised to maintain group consistency to the greatest extent possible). As with the other examples, this potential problem is minimal because the yes/no questions are as clear and objective as possible, but subjectivity can never be completely eliminated. Evaluators should always consider this limitation when they interpret the LSAT's results.

The following summary is a partial listing of the information collected in each section of the instrument (see Appendix E for the LSAT questionnaire).

Organizational Context: Yes/no questions ask about the existence of a logistics management unit and its responsibilities. Qualitative questions ask about relations among key stakeholders and how that affects logistics system performance.

Logistics Management Information System (LMIS): Yes/no questions assess the types of information collected through the LMIS, the purposes for which the information is used, and the extent to which the LMIS is automated. Qualitative questions seek to understand the flow of information at various levels of the system,

whether the information collected is used by program managers, how useful it is in practice, and whether reports are accurate and timely.

Product Selection: Yes/no questions ask about the existence of a national essential drug list, the existence of a National Drug Policy document, and the basic characteristics of each. Qualitative questions inquire about the contents of the documents and probe to determine to what extent they are disseminated and applied at various levels of the system.

Forecasting: Yes/no questions are used to determine what information is used in preparing forecasts, and how frequently they are programmed. Qualitative questions examine the quality of forecasts, the impact of forecasts on budgeting and planning, and the capacity of in-country staff to carry out forecasts without external assistance.

Obtaining Supplies/Procurement: Yes/no questions assess whether procurement plans take into account certain information items. Qualitative questions probe the methods for coordinating procurement planning in the country, and whether, in general, the program procures the right amounts of the right goods.

Inventory Control Procedures: Yes/no questions provide information on the use of state-of-the-art inventory control practices (e.g., first expiration – first out (FEFO); established maximum and minimum stock levels), and whether stockouts occurred at any level during the past 12 months. Qualitative questions look at how well the staff applies the procedures in practice. They also identify which products are most frequently stocked out and why, how the staff handles stockouts and oversupplies, and the effects of stockouts on the program.

Warehousing and Storage: Yes/no questions seek to determine the existence of guidelines for proper storage of all products, for assurance of product quality, and for disposal of hazardous waste and damaged/expired products. Qualitative questions collect more in-depth information on areas where staff could improve storage conditions, and how the staff assures product quality.

Transport and Distribution: Yes/no questions check for a budgetary line item for various components. Qualitative questions assess whether transport is sufficient and whether the distribution system is effective.

Product Use: Yes/no questions query the existence of standard treatment guidelines and procedures for monitoring prescribing practices. Qualitative questions probe to determine the extent to which standard treatment guidelines and universal safety precautions are actually implemented.

Finance: Yes/no questions assess whether the national budget covers logistics, whether a cost-recovery system is used, and what logistics expenses are covered by donors. Qualitative questions examine whether budget allocations for logistics are sufficient to ensure product availability, to examine the source of funds for the logistics budget, and to examine whether donor activities and resources are adequately coordinated.

Organizational Support for Logistics System: Yes/no questions assess the existence of job descriptions, and characteristics of communication channels, information use, decision-making, feedback, supervision, and training. Qualitative questions focus on how various organization and staffing procedures are carried out and how effective they are in practice.

Part II.F
Behavior Change Communication (BCC)

- Percent of audience who recall hearing or seeing a specific message
- Percent of audience that know of a product, practice, or service
- Percent of audience with a specific attitude (toward the product, practice, or service)
- Percent of audience who believe that spouse, friends, relatives, and community approve (or disapprove) of the practice
- Percent of audience that perceive risk in a given behavior
- Percent of audience who experience a strong emotional response (to the communication)
- Percent of audience that are confident that they could adopt the behavior
- Percent of non-users who intend to adopt a certain practice in the future
- Percent of audience who have encouraged (discouraged) friends and relatives to adopt the specific practice
- Number of exposures to websites and other internet-based resources
- Use of the products on a given website

Part II.F

Behavior Change Communication (BCC)

A common thread running through all reproductive health programs is behavior change. Behavior Change Communication (BCC) programs are designed to bring about behaviors that will improve health status and related long-term outcomes. Previously known as Information-Education-Communication or IEC, the change in name implies a switch from materials production to strategically designed programs that influence behavior.

BCC programs include a wide range of interventions that fall into three broad categories:

- Mass media (radio, television, billboards, print material, the Internet);
- Interpersonal communication (client-provider interaction, group presentations); and
- Community mobilization**.**

The indicators in this section – consistent with those in the rest of the *Compendium* – focus on results achieved rather than on activities conducted (i.e., process). Any of these three types of communication – mass media, interpersonal communication, or community mobilization – can generate the results measured by these indicators, including changes in knowledge, attitudes, intentions, and behavior. However, to date evaluators have applied these indicators primarily in relation to communication programs with a mass media component, presumably because the population-based surveys needed to collect such data at the population level are relatively expensive and are only appropriate where the communication program is far reaching.

Much of the work to date on evaluating interpersonal communication has focused on process rather than on results (although notable exceptions exist; Kim et al., 1992). This work is extremely important for understanding the dynamics of these techniques and for identifying ways to improve them in field applications. Illustrative indicators that evaluate client-provider interaction (CPI) appear in the section on evaluating quality of care for family planning (Part II.H.2). However, it is beyond the scope of this manual to cover process indicators for the vast array of innovative interpersonal strategies now in use: worksite BCC interventions, group presentations in villages and neighborhoods, school-based RH/sexuality education, out-of-school programs, outreach to communities and bars, peer education, and related strategies. We refer interested readers to Kim and Lettenmaier (1995), Adamchak et al. (2000), Bertrand and Kincaid (1996).

With regard to community mobilization, Bertrand and Kincaid (1996) have published a list of process indicators, which measure activities conducted but not results achieved. The dynamics of community mobilization projects present a number of methodological challenges to evaluators. For example, the objectives of the intervention may not be fully developed at the time of the baseline survey. The intervention itself may vary markedly in the different communities that implement it. The critical analytic approach used in evaluation may be antithetical to those efforts to develop trust with the community that characterize community mobilization programs. Whereas the objectives of community mobilization include, but are not limited to, changing health behaviors, they also include developing managerial skills, strengthening organizational infrastructure, encouraging participation from different subgroups, and so forth. In this *Compendium,* we give preference to indicators that have been tested in the field. Given that the evaluation of community development initiatives is still young, we omit separate indicators for this type of communication.

The final two indicators in this section relate to the Internet. Although this medium has not been developed to its full potential as a means of reaching relevant subgroups within the general public with reproductive health messages, especially in developing countries, it represents a promising channel for the future as access to the Internet becomes more common. Far from measuring the effects of the Internet on behavior, we

limit ourselves to measuring website usage and reasons for use of this medium.

Methodological Challenges of Evaluating BCC

Evaluating BCC programs has a number of methodological challenges including but not limited to the following.

- **For programs with a mass media component, evaluators cannot identify an appropriate (or any) control group and thus they cannot rule out confounding factors (the attribution problem).**

In field-based program evaluation, many evaluators have had difficulty establishing a control group, often for administrative (e.g., lack of time, lack of money, insufficient interest in evaluation, reluctance to withhold a desirable intervention from part of the intended audience for the sake of research) rather than for technical reasons. However, in the case of programs with a mass media component, it is often virtually impossible to establish a control group (with random allocation of subjects) or even a comparison group (a population with similar socio-demographic characteristics) that is **not exposed** to the communication intervention in question. Without a control or comparison group, one cannot definitively answer the question: "what would have happened in the **absence** of the intervention?" Thus, even if the evaluation shows the desired increase in the outcome variables, one cannot unequivocally attribute this effect to the communication intervention. Other confounding factors (e.g., a communication activity of another organization, political instability, natural disaster, exposure to mass media in general) may have caused or contributed to the observed change (or lack thereof).

An alternative approach for evaluating the effectiveness of BCC interventions consists of conducting a baseline and a follow-up survey among the intended audience that measures:

- A series of ideational variables that represent sequential steps toward behavior change[1] (see the conceptual framework in Figure II.F.1);
- Socio-demographic characteristics of respondents; and
- Data on the intensity of exposure to the specific messages of the communication program (e.g., based on recall of specific messages, number of channels, or some combination of each).

This third type of data (intensity of exposure) allows the evaluator to test "dose-response" as a possible determinant of each ideational variable, after controlling for socio-demographic factors (e.g., education) known to influence health behavior. Although this description oversimplifies the design and statistical analysis involved, it does explain the choice of indicators included in this section of the *Compendium*.

Kincaid outlines four conditions that must be present for evaluators to infer causal attribution: (1) observation of a change or difference in the population of interest, (2) correlation between exposure to the intervention and the intended outcome, (3) evidence that exposure to the intervention occurred before change in the outcome (i.e., time-order), and (4) control or removal of confounding factors (or spurious effects). Whereas these conditions do not entirely overcome the problem related to a lack of a control group, Kincaid argues that the combination of theory-based evaluation and paths of influence observable in the data can lead to convincing results on the effects of communication programs, even those with a mass media component. For further discussion of these issues, see Bertrand and Kincaid (1996).

The different functional areas of FP/RH programs all labor under some pressure to demonstrate their effectiveness, yet this pressure has been particularly strong for BCC programs. One reason is that these programs claim to change behavior at the population level, and in doing so, they invite a "prove it" response. Another is that the range of confounding factors is potentially greater for BCC than for other functional areas. For example, the national logistics system for contraceptives and drugs will probably not spontaneously improve without specific program interventions (e.g., training, equipment). Similarly, service providers in clinics are unlikely to stumble upon instructions for how to insert and remove NORPLANT; rather, they can only learn such techniques through a specific training event. By contrast, the general public will likely learn about HIV transmission through non-programmatic actions (for example, from extensive media coverage in 1991 in the United States when basketball player Magic Johnson

[1] These ideational variables often correspond closely to the objectives of a communication program: increasing knowledge, improving attitudes, creating "intention," and increasing the actual practice in question.

announced publicly that he was sero-positive for HIV). Thus, of the different functional areas, BCC programs have the burden of providing the most rigorous evidence that observed changes in behavior at the population level result from their interventions and not from confounding factors in the environment.

- **A baseline survey was conducted but lacked key indicators.**

In some cases, the group conducting the baseline survey is different from the group performing the evaluation. As such, the former may not think to include certain indicators a subsequent evaluator may want or need. Even when one group performs both tasks, events may occur that were unforeseen at the baseline. In a study in the department of Quiché in Guatemala, a baseline study was conducted for the explicit purpose of evaluating a multi-intervention program. However, those constructing the survey did not know that the program personnel were designing a video on reproductive health problems, with a storyline developed around Mayan families, narrated in the local Mayan language. The program directors subjectively assessed this communication vehicle to have been one of the most effective means of "reaching" this community (traditionally known for being very closed to outsiders) with messages about birth spacing and reproductive health. However, the survey lacked a question to test this premise, and thus the effectiveness of the video was purely speculative.

This problem can also arise if the organization designing the research then passes the responsibility for collecting the data to a market research organization. The market research organization – not understanding the importance of specific questions – may eliminate them from the questionnaire (to shorten the interview time, to improve the flow of questions, or for other reasons that seem logical to them). The result, however, is the absence of key information at the time of analysis. This problem relates to a larger issue – finding and working closely with a reliable market research agency able to assure quality control in the data collection and correct statistical analysis in the reporting of results.

- **The interval between communication program and the follow-up evaluation was too short or too long.**

There is no "established" or "correct" interval between the launch of a communication program and the follow-up evaluation. Advertising firms with multi-million dollar budgets in developed countries would expect to see the effects reflected in sales data in a matter of weeks. However, change in culturally entrenched reproductive health behaviors can take months if not years to achieve. Thus, the interval between launch and evaluation is often dictated by administrative decisions rather than by technical reasons.

As a result, the interval may be too short or too long. "Too short" tends to occur when a given project must be completed within a donor-specified period of time (e.g., the end of project). Often a project that would have had an ample period for the intervention to run finds itself short on time if delays occur early in the project cycle, and thus decrease the time for actual program implementation. In such a case, a program may fail to show results, simply because too little time has elapsed for such effects to take hold.

"Too long" occurs when the program implementation occurs on schedule, but the follow-up evaluation is delayed, often for administrative reasons (e.g., funds not yet available, personnel deployed to another activity) or for factors beyond the control of the program (e.g., the monsoon season). When the interval between the communication intervention and the follow-up evaluation increases dramatically, there is greater opportunity for confounding factors to intercede.

- **The organization lacks the time or money to do an adequate evaluation.**

This problem is by no means unique to BCC; rather, it applies equally well to almost every section of this *Compendium*. In terms of time, the need to conduct the baseline survey before the communication launch may create an unacceptable delay if the program is under a tight schedule. In terms of money, program managers

often feel that the funds would be better spent on doing **more** of the program, even if it means doing **less** of the evaluation; or they fail to budget for evaluation. With the increasing pressures toward accountability, this sentiment may be on the decline but is still evident in many organizations.

- **Obtaining consent forms for human subjects often proves difficult, time-consuming, and expensive.**

Some organizations, if not most, require that all research proposals involving human subjects undergo a thorough review process. A Committee on Human Research (CHR) generally conducts this review and approval process. It usually requires the submission of the proposal, all research protocols and survey instruments, consent forms, local Internal Review Board (IRB) approval, and official translations of Oral Consent Forms into local language.

In most developing countries with communication programs, language and cultural barriers contribute to misunderstandings about informed consent. First of all, many people cannot read; and therefore they may be unwilling or unable to sign any document. Second, even if they are able to read and sign, they may be suspicious about signing any document that looks official for fear that it may be used for other purposes. Furthermore, in many settings, the perception exists that the need to sign something indicates a, perhaps dangerous, hidden agenda.

Obtaining a local IRB approval is another problem for organizations. Complying with CHR requirements and others such as submission of the "final" questionnaire and translations into local languages is difficult, time consuming, and expensive. Frequently, obtaining the approval from a CHR can take up to six months with an average time of three to four months. This delay poses serious time problems to programs.

The Conceptual Framework: How BCC Works

Figure II.F.1 presents a model of strategic communication and behavior change.

In this conceptual framework, communication is treated as an outside factor that affects the other variables in the model. Communication designed to improve skills is identified as instruction. Communication for removing environmental constraints is identified as advocacy, and communication designed to change ideational factors is identified as promotion. The model specifies how and why communication affects intention and behavior: indirectly through its effects on skills, ideation, and environmental constraints.

"Promotion" is central to this section, because it leads to ideational change (that is, a change in the way individuals or populations perceive given practices or behaviors). Promotion is designed to have cognitive, emotional, and social effects, which in turn influence a person's intent to practice a certain behavior and to follow through in doing so. The actual **behavior** is the "desired result" in almost all BCC programs, whatever the specific area or topic. Evaluators often label this behavior the "intermediate outcome" (if measured at the population level).

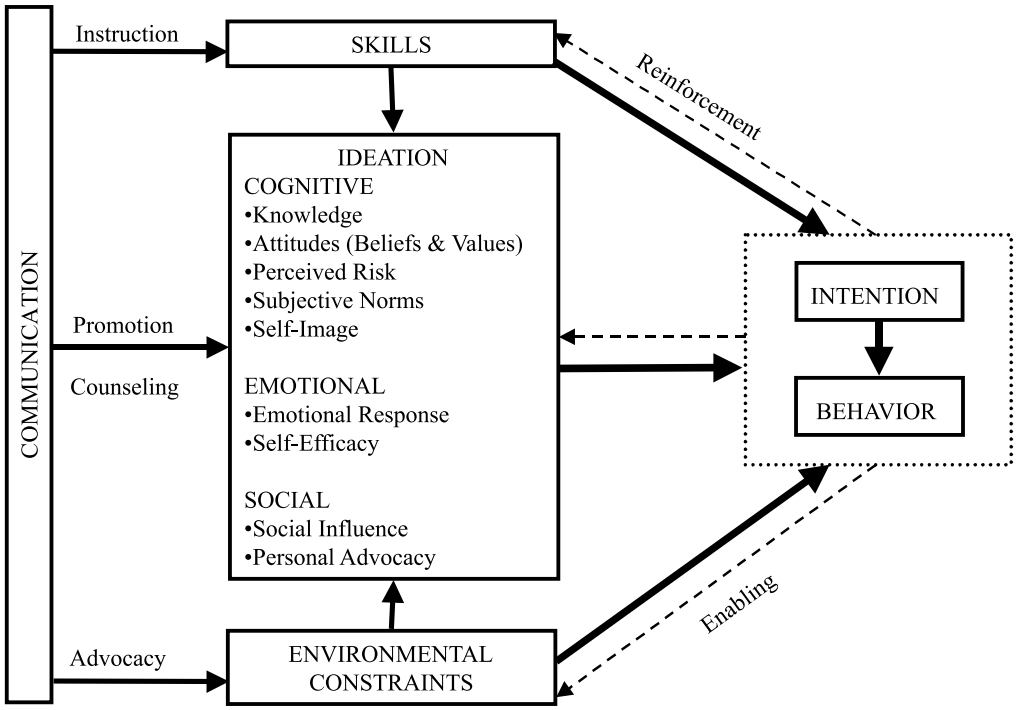
In addition to obtaining data on the actual behavior, evaluators should collect data on all ideational variables that may be relevant to the behavior of interest. Communication is designed to affect ideational variables in order to change behavior. In a pre-post evaluation design, evaluators can compare baseline measures of these variables with post-intervention data. Also, they can assess program effects on the ideational variables by comparing the level of each variable among those exposed and unexposed to the communication program. Evaluators can then use results on the relationship between ideational variables and program exposure to track changes over time and to refine and/or reinforce the communication messages. Research has shown that ideational variables operate as "proximate determinants" and that communication can influence contraceptive use not only directly, but also indirectly through ideation (Kincaid, 2000; Babalola et al., 2001).

Even if one can convince individuals that certain courses of action are desirable, environmental constraints to behavior change often exist – in the form of politically-based barriers (e.g., opposition to abortion even where it is legal), resource limitations, legal constraints (e.g., provision of services to adolescents), and other factors. Advocacy becomes a powerful tool to confront these constraints at the macro level and to minimize barriers to positive behavior at the individual level. The indicators in the following section (except for the last two, which relate to the Internet) focus on the effects of promotion and on the steps toward behavior change.

The conceptual framework in Figure II.F.1 forms the basis for the indicators in this section. Any of the concepts from the framework could translate into the indicators; we have selected those most commonly used in evaluating BCC programs to present in the *Compendium*.

The section on BCC ends with two indicators to monitor communication on the Internet. Although this has not been a primary channel for behavior change communication to date, especially in the developing world, it holds great potential for the future. In the past few years, the Internet has emerged as a source of information for millions of people in developed countries and for those with access to this medium in developing countries. Originally used by researchers and program managers, the Internet is fast becoming a source of reliable, confidential information on reproductive health topics, especially for computer-savvy adolescents.

Figure II.F.1 A Model of Strategic Communication and Behavior Change



Source: Adapted from Kincaid, 2000.

Indicator

Percent of Audience Who Recall Hearing or Seeing a Specific Message

Definition
"Audience" is defined as the intended population for the program (e.g., pregnant women for antenatal care, youth in a specific age range for an adolescent program).

"Recall" may include spontaneous mention and/or aided recall. (The distinction between the two types of recall is explained in the next indicator, **Percent of Audience that Know of a Product, Practice, or Service**.)

"Specific message" refers to a communication with some identifiable aspect (e.g., logo, jingle, character) that the respondent could not name unless he/she had been exposed to the communication.

This indicator is calculated as:

$$\frac{\text{\# of audience members who recall a specific message}}{\text{Total \# of audience members}} \times 100$$

Data Requirements
Self-report from surveys or other measurement tools; sources of information (e.g., how did the user hear about the service?)

Data Source(s)
National, regional, or local sample surveys with members (preferably a representative sample) of the intended audience

Purpose and Issues
"Reaching" the audience is an important first step to increasing levels of knowledge of the products, practices, or services in question. Reaching a large audience is one of the strengths of mass media communication, and recall of specific messages measures the reach of a given communication campaign or message.

This indicator on reach does not in itself indicate if the intended population learned anything from this exposure, or if they approved/disapproved of its content. However, this indicator is important in establishing exposure to the campaign, and if desired, establishing a "dose-response effect" (i.e., an effect that increases with number of exposures to the message(s)).

Evaluators can measure this indicator from a battery of questions to assess the range of communication channels through which the respondent has heard/seen/read a message. The instrument (questionnaire) should ask about all communication approaches used in the campaign or program, as shown in Box II.F.1.

The evaluator may also include a channel not used in the communication campaign to gauge the extent of courtesy bias inherent in the responses. Courtesy bias occurs when the respondent tries to give the socially correct answer or one he/she feels will please the interviewer, rather than the true response. This check is particularly useful where the communication program is on an unusual topic (e.g., vasectomy) or in an environment with relatively few communication channels.

Box II.F.1 Example of Measuring Spontaneous and Aided Recall of Messages

1a. In the past __ (e.g., two months) have you seen, heard, or read anything about breastfeeding in the mass media or anywhere else? Please name all the channels or places where you saw or heard a message on breastfeeding. *(Note: the interviewer circles "spontaneous recall" for each.)*

1b. (For each channel not mentioned, the interviewer asks): Did you see a message on breastfeeding on any of the following channels in the past two months? *(Note: the interviewer circles "aided recall" for any channel mentioned.)*

Mass Media:	Spontaneous	Aided	None
On the radio	(2)	1	0
On the television	(2)	1	0
In a magazine	2	(1)	0
On a poster	2	1	(0)
On a billboard/sign	2	(1)	0
In a pamphlet or brochure	2	1	(0)
Total number of sources of breastfeeding information (add up the numbers of times "1"or "2" is circled):			**4**
To create an index of exposure, one sums the scores from both columns (giving double the weight for spontaneous recall as compared to aided recall).			**6**

A second, more detailed approach is to ask about specific messages. If the questions are closed ended (requiring a response to pre-established categories or a "yes/no" to specific items as is the case in Box II.F.1), then respondents will more likely give a socially acceptable rather than a true response, to avoid looking ignorant or to please the interviewer. For this reason, it is advisable, where possible, to ask the respondent to give some defining characteristic of the message (e.g., a logo, name of a character) to verify "exposure" to the message in question.

The evaluator can sum the responses from this battery of items to construct an index of each respondent's level of recall in order to estimate a dose-response effect. Since BCC campaigns may have several messages, the evaluator can weigh specific messages more heavily for unaided versus aided recall, and then sum them to arrive at a continuous variable measuring level of recall. (For more information on how to create such an index, see Kincaid, Coleman, and Rimon, 1995.)

Gender Implications of this Indicator

If audiences are disaggregated by sex, the percent of men and women able to recall a specific message will likely differ. Behavior change communication specialists planning to reach an intended audience through the media should explore the extent to which gender differences affect access to the media. In many countries, lower levels of literacy among women than men limit the effectiveness of reaching women through print media including newspapers, brochures, posters, and billboards. Women in some societies have less access to television outside the home. Where televisions are only available in public or community places, audiences are more likely to be men with more freedom to go out at night. In societies which highly resist women's movements outside the home, women primarily obtain their information about the world from their male family members; reaching these women with health messages requires other approaches. Women increasingly have access to radio within homes, and radio has been demonstrated as a viable way to reach women with health messages in numerous settings.

Indicator

Percent of Audience That Know of a Product, Practice, or Service

Definition
"Audience" is defined as the intended population for the program, e.g., pregnant women for prenatal care, youth in a specific age range for an adolescent program.

"Know" refers to the percent that can spontaneously name (or alternatively, can recognize when the name is mentioned) a particular practice, product, or service.

This indicator is calculated as:

$$\frac{\text{\# of audience members who know of a product, practice, or service}}{\text{Total \# of audience members}} \times 100$$

A related indicator is the percent of audience who know about a source of supply for a specific product and is calculated as:

$$\frac{\text{\# of audience members who know about a source of supply for a specific product}}{\text{Total \# of audience members}} \times 100$$

Data Requirements
Self-report from surveys or from other measurement tools

Data Source(s)
National, regional, or local sample surveys with members (preferably a representative sample) of the intended audience

Purpose and Issues
Evaluators may measure knowledge of a product, practice, or service in two ways: spontaneous and aided recall, as shown on the previous indicator. For example, the design of the DHS questionnaire facilitates the process of determining which contraceptive methods a respondent "knows." The interviewer first asks the respondent to name any methods that he/she knows for preventing a pregnancy (unaided or spontaneous recall); the interviewer then asks if he/she has ever heard of the methods not mentioned (aided or prompted recall).

Evaluators can combine the unaided and aided knowledge to calculate a percent of the audience that "knows" about (can remember) the product or practice. For example, what percent of the audience has heard of the social marketing brand of condom? When "knowledge" relates to a series of items, such as the different contraceptive methods, then one can measure either (1) the percent that has heard of at least one ___ (method), or (2) the mean number (of methods) known.

One critique of this indicator is that respondent's may claim to have heard of a given product or service simply to avoid appearing ignorant. Interviewers can confirm the respondent's actual knowledge by asking follow-up questions about the characteristics of the method or practice.

Knowing a **source** of the service or product is another important and measurable aspect of knowledge that applies to almost all areas of reproductive health. Survey questions are often phrased: "where would you go if you wanted or needed to get _____ (e.g., a Pap smear, counseling for violence against women, antenatal care). To determine the percent of respondents who "know" a facility, the evaluator should ensure that the responses given correspond to actual facilities (and should not simply "accept" as correct any plausible sounding location).

Indicator

Percent of Audience with a Specific Attitude (Toward the Product, Practice, or Service)

Definition
"Attitude" is defined as a person's favorable or unfavorable *assessment* of a behavior or related construct (such as a specific product or source of service). The assessment is expressed by statements from the audience that relate the behavior with a positive or negative *value* held by the audience.

Data Requirements
Evaluators measure attitude by asking audience members how strongly they agree or disagree with these statements, usually in terms of the following five-point (Likert-type) scale:

Score	Response
1	Strongly Disagree
2	Disagree
3	Not sure
4	Agree
5	Strongly Agree

The statements must all correspond to the same behavior, product, or issue. Evaluators calculate the overall attitude score as the average of the scores from each of the statements used. The higher the average score, the more positive is the audience's attitude towards the behavior. Note: evaluators must reverse scores from negatively worded statements before they compute the average. For example, on a questionnaire about stigma toward HIV/AIDS, a respondent "strongly disagrees" (score=1) with a statement that HIV-positive persons should be quarantined from the rest of society. If most of the items on the scale are worded such that a "5" represents supportive attitudes toward persons with HIV, then the evaluators should convert the "1" on this question about quarantining HIV-positive persons to a "5" before the average is computed.

If the evaluator uses the five-point scale, he/she can also combine the "agree" and "strongly agree" responses to obtain the percentage of the audience with a positive attitude. Examples of attitudinal statements with the underlying values underlined include:

Attitude towards family planning:

- Practicing family planning helps a woman regain her strength before having her next baby;
- Practicing family planning eliminates the fear of getting pregnant; and
- Family planning encourages a wife to become promiscuous.

Attitude towards condom use:

- Using condoms reduces sexual pleasure;
- Using condoms helps prevent HIV/AIDS infection; and
- Using condoms is a sign of infidelity.

Respondents express their values in terms of the expected outcome of the behavior, expected benefit or harm, or positive and negative attributes of the behavior or product.

Data Source(s)
Quantitative: national, regional, or local sample surveys with members – preferably a representative sample – of the intended audience

Qualitative: focus groups, in-depth interviews, pile-sorts, ethnographic observation, knowledgeable informant panels

Purpose and Issues
Attitudes influence all types of social behavior. People generally act in ways consistent with their attitudes. Attitudes may more strongly influence intention than does behavior per se. For example, a person's negative attitude towards smoking may create a strong intention (desire) to stop smoking, but the person may continue smoking because of other factors, such as prevailing social norms and addiction. Communication programs often address the specific beliefs and values that encourage or discourage a particular practice or behavior. In some societies, for example, men believe they have the right to many sexual partners. Mass media

programs can begin to alter that behavior if they portray it as socially unacceptable, harmful to one's family, and threatening to one's own health because it increases the risk of HIV/AIDS infection.

Attitude change (and reinforcement) is one of the ways communication programs ***indirectly*** influence health behavior. Audience members who change or strengthen their attitude because of exposure to the messages of a communication program are more likely to engage in the desired behavior (Fishbein and Ajzen, 1975).

Indicator

Percent of Audience Who Believe That Spouse, Friends, Relatives, and Community Approve (or Disapprove) of the Practice

Definition

"Audience" is defined as the intended population for the program. "Practice" (or behavior) refers to the desired result that the program is trying to achieve among members of the intended population. Practice is the intermediate outcome when measured at the population level. Examples include: contraceptive use, delivery in the presence of a skilled birth attendant, exclusive breastfeeding, use of a condom at last sex with a casual partner.

The practice in question will determine the persons cited in the indicator (e.g., spouse, friends, relatives, and community). For example, the spouse would play a large role in a family planning decision, whereas the views of the extended family or community would highly influence the decision regarding female genital cutting.

This indicator is calculated as:

$$\frac{\text{\# who believe the practice is approved (or disapproved) by spouse, friends, relatives, and/or community}}{\text{Total \# in intended population}} \times 100$$

Data Requirements

Self-report from surveys

Data Source(s)

National, regional, or local sample surveys with members, preferably a representative sample, of the intended audience

Purpose and Issues

This indicator measures the extent to which an individual's social network approves or disapproves of the practice or behavior in question. If he/she perceives approval to be low or negative, then the social cost of adopting the behavior is high. Conversely, if he/she perceives approval to be high, then the social cost is low, and the social environment can facilitate behavior change and maintenance. The implication, supported by theory, is that the decision to adopt a particular behavior is more than an individual decision. Mass media is a useful tool for changing the ways in which audience members perceive their social environment and the extent to which they believe their social network supports a given practice.

A closely related indicator is the percent of intended audience who reports talking about the practice/behavior to one's spouse, friends, and relatives as a result of exposure to a specific campaign. Given the importance of social influence, this indicator measures the extent to which a given communication message or campaign has triggered discussion – either positive or negative – on the topic. Such communication is a first step to reinforcing or changing community norms. Again, the persons of interest in this question may vary with the type of RH intervention.

Indicator

Percent of Audience That Perceive Risk in a Given Behavior

Definition
"Risk perception" is defined as one's beliefs about the likelihood of experiencing negative or harmful consequences. This definition comprises two distinct dimensions: (a) susceptibility to a threat, and (b) severity of that threat. Examples of risk perceptions relevant to HIV/AIDS prevention include:

$$\frac{\text{\# who perceive risk}}{\text{Total \# in intended population}} \times 100$$

and

$$\frac{\text{\# who perceive severity of risk}}{\text{Total \# in intended population}} \times 100$$

Data Requirements
Self-report from quantitative instruments (surveys) or qualitative data collection methods

Data Source(s)
Quantitative: national, regional, or local sample surveys with members – preferably a representative sample – of the intended audience

Qualitative: focus groups, in-depth interviews, pile-sorts, and ethnographic observation (although the latter generally does not yield a precise numerical result)

Purpose and Issues
Risk is the likelihood of a specific event occurring multiplied by the magnitude of consequences associated with that event (Douglas, 1985). Perceived risk parallels this definition as it is (a) one's perceptions of susceptibility or vulnerability to a threat (e.g., What are the chances that a specific event will occur? What are my chances of getting HIV/AIDS if I am abstinent? What are my chances of contracting HIV/AIDS if I use a condom every time I have sex? What are my chances of getting HIV/AIDS if my partner and I are mutually faithful?), and (b) one's perceptions of the severity of that threat (e.g., How serious are the negative consequences associated with the event? What negative consequences are associated with contracting HIV/AIDS? How serious or bad are these consequences?). Presumably the individual thinks of the social, economic, and spiritual consequences, in addition to physical consequences such as sickness or death. For example, "how harmful would it be to my family if I got HIV/AIDS?"

Risk perceptions, commonly referred to as "perceived threat," act as the *motivation* promoting behavior change (although perceived positive benefits can also motivate change). Research has shown that individuals can have all of the knowledge and skills needed, have positive beliefs, attitudes, and intentions toward a specific behavior, yet they avoid engaging in the recommended behavior. They need a trigger to motivate action. Much research has shown that perceived threat is a powerful trigger to action (Witte, 1992 and 1998).

Evaluators can expect desirable behavioral responses when people have strong risk perceptions coupled with strong beliefs of self-efficacy toward the recommended response. Evaluators may expect undesirable behavioral responses when people have strong risk perceptions but they doubt their ability to do a recommended response (e.g., negotiating condom use), and/or they doubt the recommended response will work to avert the threat (e.g., strong rumors circulate in some countries that condoms are contaminated with HIV and actually transmit the infection). Therefore, evaluators must measure perceptions of efficacy when they assess perceptions of risk, so that program staff can devise the best communication messages.

Indicator

Percent of Audience Who Experience a Strong Emotional Response (to the Communication)

Definition
"Emotion" is a physiological response that is subjectively experienced as strong feelings (such as fear, love, hate, anger, disgust, grief, joy, or surprise). Two distinct aspects comprise emotion: the physical/bodily response and the subjectively experienced feeling. Communication experts recognize that the subjective emotional response influences the impact of a message on the receiver.

This indicator is calculated as:

$$\frac{\text{\# who experience strong emotional response}}{\text{Total \# in intended population}} \times 100$$

Data Requirements
Self-report of audience members regarding their subjective response to a specific communication

Data Source(s)
National, regional, or local sample surveys of members of the intended audience

Purpose and Issues
Communication experts recognize that the subjective emotional response of members of the audience toward specific communication greatly affects the potential impact of a given message. Historically, those designing radio and television spots have explicitly crafted their messages to elicit this type of marked emotional response. Market researchers have attempted to measure this type of emotional response as part of routine communication pretesting. However, to date, relatively few evaluators have experimented with measuring this dimension in field-based surveys. In contrast to other indicators in this section of the *Compendium* that are widely used at the field level, this indicator is included for its potential use in evaluating communication programs.

Before evaluating emotional response to a message on a survey of the intended audience, one must first establish that the respondent has seen or heard the communication in question. Evaluators often ask the question in two parts (spontaneous and aided recall):

- What messages do you remember seeing on television in the past _ months (e.g., three months) on the subject of _____ (e.g., AIDS prevention)? (Note: the number of months should correspond to the time period that the spot was actually broadcast); and
- (If the respondent does not mention the message in question, the interviewer asks:) Did you see the message on TV with the young father whose wife died of AIDS?

If the response is affirmative, the interviewer then asks a question to measure emotional reaction to that message. Depending on the message in question, the interviewer can ask the respondent to agree or disagree with a specific statement about the message, such as:

- I became very *sad* when I saw what happens to the orphans of parents who die from AIDS;
- The man who said he would not wear condoms made me very *angry*;
- The woman who asked her husband to wear condoms made me feel *disgusted*; and
- I really *liked* the man who advocated practicing responsible sexual behavior.

Respondents can answer these four questions using a simple "agree/disagree," or using an expanded five-point Likert-type scale: strongly disagree (1), disagree (2), not sure (3), agree (4), strongly agree (5).

Emotion influences the impact of health communication in two ways. First, certain emotions (especially fear) can act as a barrier to change. For example, as long as women are too afraid to ask their partners to use condoms, they are not likely to change their behavior. Conversely, if couples' fear of getting pregnant becomes

more salient (and somewhat greater), these couples will become more receptive to messages about family planning, as well as more motivated to change their behavior. Second, some level of emotional response may be necessary before people will seriously listen to messages and reexamine their own beliefs, counter arguments, and self-rationalizations for maintaining or for changing their behavior. Third, the emotional content of the message may cause it to stand out from the multitude of other nondescript messages that compete for the viewer's attention. As such, the emotional content of the message contributes to its effectiveness.

One limitation of this indicator relates to the measurement of emotion several days or weeks after the respondent sees or hears the message. Certain emotion-charged events are very memorable and remain ingrained in one's mind for some time thereafter. Others produce a fleeting emotional response that may be superseded by other emotions prompted by more recent events. In short, this indicator is subject to some degree of recall bias. However, further testing of the indicator in the context of the evaluation of national communication programs will indicate its value as a pathway in analyzing communication effects.

Indicator

Percent of Audience That Are Confident That They Could Adopt the Behavior

Definition
This indicator measures perceived self-efficacy – the conviction that one can successfully accomplish the behavior required to produce a particular outcome. In other words, it is the confidence a person feels about performing a behavior, for example:

- Using a condom consistently for non-marital sex;
- Abstaining from sex for six months;
- Using a contraceptive method continuously for one year;
- Visiting a family planning clinic for services if they need to; and
- Talking about safer sex with one's partner(s).

This indicator is calculated as:

$$\frac{\text{\# who are confident that they can adopt the behavior}}{\text{Total \# in intended population}} \times 100$$

Data Requirements
Self-report from surveys

Data Source
National, regional, and local sample surveys with members (preferably a representative sample) of the intended audience

Purpose and Issues
Key behavior change theories and models recognize the importance of perceived self-efficacy in the adoption and sustained practice of a behavior. Empirical evidence also abounds to the effect that self-efficacy is a key determinant of behavior change. For example, Bandura (1986) contends that people's judgments about their capabilities (perceived self-efficacy) to organize and to execute courses of action is the most influential mediator of human behavior and plays a powerful role in determining the choices people make.

Perceived self-efficacy helps explain the reasons people's behavior differs widely even when they have access to similar resource bases, such as knowledge and skills. The more confident a person feels about taking a course of action, the more likely he or she will take that course. A woman with little confidence in her capability to insist that her partner use a condom is not likely to adopt that behavior. As self-efficacy increases, the amount of effort a person invests in a behavior also increases. Self-efficacy also affects the level of performance of the behavior, perseverance at a task in the face of challenges, the degree of anxiety or confidence a person brings to the task at hand, and the coping responses exhibited.

To measure self-efficacy, evaluators must distinguish between having the skills necessary to accomplish a particular task and having the perceived self-efficacy to accomplish the task. For example, having the knowledge and ability to use a condom is a measure of skills, not of self-efficacy. The appropriate measure of self-efficacy in this case would be the level of confidence a person has in his/her capability to successfully negotiate the use of a condom. Similarly, knowing how to say "no" to sex is an indicator of sexual negotiation skills, while the belief in one's capability to refuse sex under specific circumstances measures perceived self-efficacy. Obviously, a link exists between the two constructs. Developing self-efficacy about a behavior starts with some level of knowledge about that behavior. Exposure to different situations providing a person the opportunity to practice the new behavior increases perceived self-efficacy. However, a person can know something but not act on that knowledge because he or she lacks the confidence to carry out the required action.

Questions about perceived self-efficacy should be precise and refer to specific circumstances. For example, perceived self-efficacy at negotiating safer sex may depend on the particular context. Therefore, a question that is not context-specific may be a poor measure of self-efficacy. (For specific applications to Adolescent Reproductive Health Programs, see Part III.H: **Percent**

of Adolescents Who Are Confident That They Could Refuse Sex if They Didn't Want It and **Percent of Adolescents Who Are Confident They Could Get Their Partner(s) to Use Contraceptives/Condoms if Desired.**)

Gender Implications of this Indicator

In sub-Saharan Africa, where young girls are four to five times more likely to contract HIV/AIDS than are young men of the same age, HIV/AIDS has increasingly been linked to gender inequality (UNAIDS, 2001). Young adolescent girls may be least able of all demographic groups to successfully negotiate safe sex with a partner. Although self-efficacy may increase with age, in many contexts where condom use is associated with illicit sex, even women who are well educated and/employed and otherwise feel empowered to make decisions may face obstacles when asking a regular partner to use a condom. Interventions that attempt to increase self-efficacy among girls and women may have to engage men as well in order to achieve not only confidence that the behavior could be adopted, but successful adoption of the behavior.

Indicator

Percent of Non-Users Who Intend to Adopt a Certain Practice in the Future

Definition
This indicator measures the intention of non-users to adopt a behavior. "Non-users" are those individuals in the intended audience who do not (yet) practice the behavior in question.

"Intend" is operationally defined as the percent of non-users who answer affirmatively to the question, "Do you intend to ___ (practice a specific health behavior) in the future." Evaluators should define the period, for example, "in the next 12 months."

"Behavior" refers to the desired result the program is trying to achieve among members of the population in question.

This indicator is calculated as:

$$\frac{\text{\# who intend to adopt the behavior in the next 12 months}}{\text{Total \# in intended population}} \times 100$$

Data Requirements
Self-report from surveys. As with attitude, if researchers use a five-point Likert scale, they must decide whether to combine "strong intent" with "some intent" to arrive at the total percentage intending to adopt the desired behavior.

Data Source(s)
National, regional, or local sample surveys with members (preferably a representative sample) of the intended audience

Purpose and Issues[2]
The decision to practice a given behavior (e.g., deliver at a hospital, breastfeed one's baby) can take the form of a statement of intent to act at some point in the future. The value of measuring intention can vary depending on the stage of the program or on the audience in question. For example, if the behavior is still new and infrequently practiced, and the media are trying to create awareness, increase knowledge, and change attitudes, intention to use is an important step to behavior change. In programs where the audience is not yet in a position to act (e.g., adolescents being encouraged to delay initiation of sex), intention to use contraception when appropriate is a good outcome indicator for the short term because long-term behavior change is not available. Another useful indicator for adolescents is intention to delay sexual debut.

Intention is relevant among members of the intended audience that do **not** yet practice the behavior. Evaluators may also measure degree of intention or of probability of adopting the practice by asking members of the audience to respond in terms of a five-point word scale of likelihood (definitely, probably, unsure, probably not, and definitely not).

An alternative approach is to combine two types of intent to arrive at a single desired behavior (e.g., safer sex). Communication programs directed at youth might measure the percent of the audience that report high intention to abstain from sex before marriage or that definitely intend to use a condom every time they have sex.

Mass media generally focus on creating awareness of a benefit or sparking a latent need: (Do you want your baby to be healthier?) Evaluators can measure the ability to create a felt need by the reported intention ("I intend to breastfeed my baby exclusively for six months because it will make him/her healthier"). Based on theories of persuasion, the messages designed for RH BCC interventions usually attempt to link practice to new benefits/values, to increase the importance of those

[2] Although "intention" is a behavioral response that could be classified as an intermediate outcome (and thus listed in Part III of the *Compendium* under specific types of programs), it tends to be used in connection with the evaluation of BCC programs, as an important step between non-use and use, which behavioral studies often overlook.

values, and/or to strengthen the belief that such benefits will indeed result from the practice in question. Evaluators use focus group discussions or in-depth interviews with audience members to identify the most important attitudes related to intention and practice of the behavior.

Intention usually mediates the relationship between attitudes and behavior. Not all individuals who have intentions can act on them immediately. In the case of HIV/AIDS prevention programs, some members of the population may wish to get tested for HIV, but may not have access to a testing facility. Communication programs can affect the attitudes and the intention to change behavior (get tested), without affecting behavior. Intention, as a step in the behavior change process, is "necessary but not sufficient" to predict with certainty behavior change.

Indicator

Percent of Audience Who Have Encouraged (Discouraged) Friends or Relatives to Adopt the Specific Practice

Definition
"Encouraged" refers to speaking positively about the practice; "discouraged" to speaking negatively of it. "Friends and relatives" are self-defined by the respondent. "Adopt the specific practice" refers to the behavior that the communication campaign is attempting to change.

This indicator is calculated as:

$$\frac{\text{\# who have encouraged (or discouraged) adoption of a specific practice}}{\text{Total \# in intended population}} \times 100$$

Data Requirements
Self-report from surveys

Data Source(s)
National, regional, or local sample surveys with members (preferably a representative sample) of the intended audience

Purpose and Issues
Behavior change cannot be expected to continue unless the expected benefits accrue. The positive outcomes of the behavior (e.g., breastfeeding) reinforce that use and increase the probability of continuation. Furthermore, continuation is higher among individuals who not only receive benefits but also tell others about them. Personal advocacy reinforces one's own behavior, while, at the same time, creating a social environment that reinforces others' behavior or influences others to change their behavior (Kincaid, 2000; Kincaid et al., 1999). One can, however, successfully adopt and sustain a behavior without publicly advocating it to others.

Mass media programs can be designed to remind current users that benefits or changes in their lives relate to the practice in question. Programs can also encourage satisfied users to talk to their friends or to have their friends come with them to the clinic. This indicator helps assess the degree to which a behavior has become a new social or community norm, as active encouragement/discouragement by a broad range of family and community members goes beyond perceived changes in social norms and becomes active social support for the practice. The goal of these interventions is to confirm individual behavior change through advocacy of use to others and to establish the behavior in question as an established social norm everyone can follow.

Indicator

Number of Exposures to Websites and Other Internet-Based Resources[3]

Definition
This indicator measures exposure to the Internet. "Hits" are all exchanges between a client (defined not as an individual user but rather a software package such as a browser, e.g., Netscape or Internet Explorer) and the server that supports a given website.

"Website" refers to a collection of electronically linked documents, located on a computer connected to the Internet, which is identifiable by a World Wide Web domain name (e.g., www.___).

"Internet-based resources" refer to information repositories such as databases, bibliographies, and reference materials that individuals can access through the World Wide Web.

Data Requirements
Statistics from log files generated by any web log analysis program, such as WebTrends®, Analog, or Webalizer. See Box II.F.2 for specific indicators available.

Data Source(s)
Access logs generated by the web server itself (e.g., Apache, IIS), and in isolated cases, lists of registered users for controlled websites

Purpose and Issues
Over the past five years, the Internet has become a medium of choice for diffusing communication and for reaching new audiences. With respect to international reproductive health, the primary users are health organizations wishing to publicize key information about their organizations (e.g., mission statement, key activities, contact points, publications, staff roster) and health professionals seeking information on specific topics. The Internet has truly revolutionized information dissemination, by providing easy access to immeasurable stores of information to individuals worldwide. Although barriers to use still exist in many parts of the world (e.g., limited familiarity with use of computers, technical difficulties in obtaining and maintaining the connection, the prohibitive cost to maintain service), the Internet potentially equalizes access to information.

In the United States and other developed countries, the Internet services not only health professionals but also consumers in search of goods and services. With respect to reproductive health, teens can log onto www.teenwire.com, a youth-friendly website created by the Planned Parenthood Federation of America to provide information on a wide range of reproductive health topics and services. In this case, the website becomes an important tool for BCC, because it provides the intended audience with an alternative source of information (pamphlets, booklets, hotlines, informational talks). Although Internet usage is much more limited in developing countries, websites intended to reach consumers in the general public with RH messages are emerging. For example, several websites have been developed in Spanish to reach adolescents with RH information and advice, including:

Sexsalud.com (developed in Peru):
http://www.terra.com.pe/sexalud/scripts/index.cgi

@dolescencia (developed in Mexico):
http://www.puni.uanl.mx/@dolescencia/

Dame Tono (developed by Johns Hopkins University Center for Communication Programs):
http://www.jhmi.edu/~maguilar/menu.html

Corporations have also contributed to the dissemination of RH information for young adults via the Internet. For example, Schering has supported the developments of web sites in Africa and Asia:

http://www.femalelife.co.za
http://www.asia.femalelife.com

[3] A large part of the text for these two indicators was drawn from a presentation by Theresa Norton and Nathalie Likhite.

Web site usage analysis (known as log file analysis) represents a promising methodology for capturing the flow of information from websites to consumers. The quantification is by no means precise, for reasons outlined below, but the statistics provide a crude measure on website usage.

The data captured in a log file vary by type of server used and by the log file format(s) it supports. Data from log file analysis answer the questions

- **Who:** the address of the computer requesting the file;
- **What:** the URL (Universal Resource Locator, or "name") of the file requested; and
- **When:** the date and time of the request.

Other potentially useful information includes:

- The referring URL (the website from which the user linked to this site); and
- The browser and operating system (e.g., PC, Macintosh) used by the requesting computer.

The log file is **not** able to capture data on:

- Number of unique users. A "user" in the log corresponds to an Internet Protocol (IP) address, a number consisting of four groups of digits arbitrarily assigned to a computer or network of computers. The IP address does not necessarily correspond in a one-to-one ratio with an individual person, because the address can represent a "spider" (e.g., an automated search engine, such as Google or AltaVista), a cache (a proxy server or an Internet service provider, such as America On Line (AOL), or a network of computers;
- Individuals' identities (e.g., an email address);
- Qualitative data on reasons for visiting the site, reactions to site contents, actual use of the files viewed, and related information (see next indicator, **Use of the Products on a Given Website**); and
- Files not viewed: log files have no record of files in which no activity occurred so that "least used pages" will not reflect **unused** pages.

The reason the number of requests ("hits") to a website does not equal the number of users relates to "caching," a procedure whereby the system automatically stores a downloaded page in a large-scale memory bank for a period of time to reduce response time. Thus, a frequently requested document may be drawn directly from the cache, and the server has no record of its being viewed. Caching reduces the actual quantity of use recorded by the server, but by an unknown amount.

Another limitation of log analysis relates to the user's geographical location. Log files do not provide a sound basis for tracking geographical distribution of users for two reasons. First, an IP address is a unique number attached to a machine or to a network of machines rather than an address that identifies people. A single IP address can, for example, represent all employees at a given site or organization. Thus, one can not equate the number of unique IP addresses with a number of unique users.

Second, Web log analysis packages tend to base their geographical statistics on the location where a domain name was registered (which may be different from the location of the PC user.) For example, the log file captures individuals from across North America accessing a site through America Online as located in the state of Virginia (headquarters to AOL). Moreover, the system often mixes geographic location ("mx" for Mexico, "cn" for China) and organization type of user (.edu, .org, .com) in the structure of the domain name. Some countries, such as Australia, have mandated the use of both an organizational suffix and a country suffix (e.g., "au" for Australia). However, this convention is not universal; where no country suffix exists, .com, .org, .net could refer to commercial enterprises, organizations, and networks from any country. Thus, log analyzers presenting geographic distributions and organization type breakdowns as separate tables can be misleading. On the other hand, development organizations trying to reach audiences from developing countries can demonstrate evidence of usage of a website by users in specific countries, even if the figure underestimates the actual number of hits from that country.

The information generated by log analysis is useful not only to program evaluators, but also to the administrators of the websites (e.g., in detecting problems, such as "failed hits" on certain pages).

One can also capture usage data on other electronic resources, such as CD-ROM resources and computer-learning centers. For example, INPARRES in Lima,

Peru, developed and tested an interactive multimedia CD-ROM entitled "Isabel: Your Electronic Counselor" to deliver FP, RH, and sexuality information. Adolescents and young people visiting the clinic can access the program – consisting of videos, animated cartoons, text, and audio – using touch screens. An anonymous database tracks user sex, age, educational level, and area of residence.[4]

Despite the limitations of log analysis of website usage, data from this source provide a crude measure of information flow from program websites to the public. Users of this information should not interpret this figure to precisely measure users, but rather to evaluate trends over time and to identify the relative draw to different parts (topics) of a website. Haigh and Megarity (1998) summarize the imprecision in log file analysis by concluding that it is "perhaps best viewed as an art disguised as a science."

Box II.F.2 Useful Indicators of Website Usage for Program Evaluation

Total number of hits: number of requests a browser makes for files on the web server; it is a count of all requests (hits) including HTML pages,[5] images, style sheets, and scripts, among others.

Total number of page hits: A count of successful requests for a specific page (i.e., unit of text, undefined in length, related to a central concept or idea; pages are files that generally end in .htm, .html, or .asp.).

Total number of hits to the home page: Number of times the home page of a website is visited. The home page of a site is generally the first page encountered by most users; it usually contains methods of navigating the rest of the site, provides links to popular areas, and often tries to communicate the purpose of the website and sponsoring agency to users.

Number of hits that were successful/unsuccessful: "Successful" refers to files that were returned to the browser; "unsuccessful" refers to requests that could not be returned or sent. This information is particularly useful to the creator of the website for trouble-shooting purposes.

Weekly/monthly averages: Evaluators can compute each of the indicators above for a specific reference period (e.g., a week, month, quarter, year) for purposes of tracking website usage over time.

Average number of minutes that page is displayed in user's browser: This indicator quantifies the amount of time that visitors to the website spend on different topics (i.e., pages).

[4] http://www.jhuccp.org/isabel/index.stm

[5] An HTML page contains all the text as well as the following information about each graphic image (picture, chart, button, among others): the name of the image file, its location within the page, and its height and width.

Indicator

Use of the Products on a Given Website

Definition
This indicator measures use of website products. "Use" refers to what the person accessing the website or other Internet-based resource does with the information. What was his/her motivation in visiting or subscribing to the resource (e.g., website, e-mail discussion list), and how did he/she use the information found through this resource?

Data Requirements
Self-report of users of a specific Internet-based resource, including websites, web-based forums, and e-mail discussion lists (listserves)

Data Source(s)
Online user feedback surveys and in-person or telephone interviews

Purpose and Issues
Those who provide information to an audience through a website on the Internet or who host an e-mail discussion list (sometimes called a listserve) wish to know the use made of this information: why did the individual access the resource? What did he/she want to learn? How did he/she use the information? Was it used in a classroom presentation? Was it used in a presentation at a professional meeting, to prepare a grant proposal, to keep professionally up-to-date, or to satisfy personal interest only?

E-mail and web-based surveys are useful for gathering this type of information. They tend to work best with closed-ended questions (e.g., a set choice of responses, such as yes/no or a Likert scale of 1-5). A second source, content analysis of information from electronic forums, summarizes the topics discussed, the extent of information exchanged, and the tone of the exchanges.

Several caveats warrant mention in connection with user feedback surveys. First, they may not provide representative results; respondents self-select in answering the questionnaires. Second, users often need some type of reward to motivate them to complete feedback forms and surveys. Third, users may negatively react to attempts to collect data from them. They may resent the delay of a login procedure or may fear for their privacy.

Evaluators should know that the general principles of evaluation that exist when they use print-based surveys (delivered via mail or in-person) remain true and are even more important when evaluators use web-based and email surveys. Informed consent, the tone and language used to invite subscribers/users to complete questionnaires, the privacy/confidentiality of their responses, and the need to inform them of the results of the survey are all very important. In addition, respect of "netiquette" – rules to encourage politeness, civility, and enhanced understanding among users/subscribers – is key in the evaluation process.

For a more in-depth understanding of the end-user applications and the impact of the products, the evaluator may conduct in-depth interviews with selected users to gather responses on the users' experience of the product and the ways they are using the knowledge conveyed (e.g., to strengthen capacity, to improve performance). However, this type of data collection can be costly and time-consuming; also, the interviews are subject to courtesy and recall bias, and persons chosen as the respondents may be suspicious as to why they were selected. Evaluators may avoid these suspicions if they select all members of a group to be interviewed (e.g., all users of a resource within a certain time period). On the positive side, in-person interviews can sometimes elicit better qualitative responses than can online surveys in cultures valuing face-to-face contact.

Gender Implications of this Indicator

Although evidence from the United States indicates that women and girls are as likely as men and boys, if not more so, to use the Internet , the limited information available from developing countries indicates that girls are disadvantaged in terms of access to computers and to web-based technologies. Where such resources are limited, gender power imbalances will very likely favor male use of the Internet over female. Those promoting web-based products – such as interactive sex education programs intended to reach both girls and boys through school or public library computers, cybercafes, and other mechanisms – may need to consider "girls only" hours or sites so that girls will have access to the technology and information available.

Part II.G Operations Research

14 Process indicators

11 Impact indicators

6 Contextual and other factors

Part II.G

Operations Research

What is Operations Research?

Operations Research (OR) is an approach used to improve service delivery or to strengthen other aspects of programs. Although OR can include diagnostic or evaluative studies, the most common use of OR is the intervention study, consisting of five steps:

1. Identifying problems related to service delivery;
2. Identifying possible strategies to address these problems;
3. Testing these strategies under quasi-experimental conditions;
4. Disseminating the findings to program managers and policymakers; and
5. Using the information to improve service delivery programs (Fisher et al., 1991).

This approach is particularly useful in testing new and potentially controversial strategies to service delivery. The implementing organization can experiment with the new approach on a limited scale, without having to adopt it throughout the organization. If the strategy (intervention) proves ineffective or creates unwanted political backlash, then the organization can decide to discontinue it and pursue alternative approaches, at relatively little political cost. If the intervention proves effective and acceptable to the population in question, then the organization can use these results to justify the adoption/expansion of the intervention within the organization. Moreover, the results of a successful OR project may prompt other organizations to adopt the same intervention in their own programs.

The evaluation of OR should address both process and impact. Until recently, there has been relatively little evaluation of operations research, in part because most OR projects are designed to evaluate an intervention. Should one then "evaluate an evaluation?" To the extent evaluation of OR projects occurred, it tended to measure outputs (e.g., how many studies were conducted, how many reports were distributed). With a few notable exceptions (Solo et al., 1998), there was little systematic assessment of impact: the extent to which the OR study resulted in changes in service delivery procedures or policy.

In 1992-93, an OR Working Group, convened under The EVALUATION Project, proposed a set of indicators to evaluate OR studies (later published in Bertrand and Brown, 1997). This work paved the way for the development of a more complete set of indicators under the FRONTIERS Program; evaluators tested these indicators in various countries between 1999 – 2001. The new set of indicators, presented here, measures both how well a study is carried out ("process") and the extent to which a study results in changes in service delivery procedures or policy (referred to in this section as "impact"). In addition, the set includes indicators of context, which describe factors that facilitate or hinder the conduct of OR and the utilization of results; they are useful in explaining what has (or has not) happened, but – in contrast to the indicators of process or impact – they are not scored.

To those interested in more systematically tracking "what happens" as a result of OR studies, this list of indicators should prove beneficial. For others, the exercise may seem too academic and the list of indicators too extensive. Whereas we present the full list of indicators developed to evaluate OR projects, we encourage users to select a subset of these indicators most relevant to their own needs (e.g., relevance, technical soundness, credibility of the findings, and demonstration of change).

Methodological Challenges of Evaluating Operations Research

- **"Impact" is generally defined as change attributable to the project, but OR is generally only one of many influences in decision-making.**

Operations research has been a great catalyst in the field of family planning, and it is now playing an increas-

ingly large role in the area of HIV/AIDS, safe motherhood, postabortion care, prevention of FGC, and related areas of reproductive health. Notwithstanding, an OR study alone rarely results in a major change in service delivery or policy, and demonstrating cause and effect is virtually impossible when evaluating the impact of an OR study on the service delivery environment.[1] Other agencies are often involved in the provision of technical assistance essential to the successful implementation of the intervention. Moreover, a shifting political climate (positive or negative) can influence the utilization of OR results. A more realistic goal is to attempt to demonstrate plausible attribution. To meet the conditions of plausible attribution, the change in service delivery or policy must:

- Be instigated by persons familiar with the OR results;
- Take place after the OR study; and
- Be consistent with the results and recommendations of the OR study.

- **Decision-making is a complex and not necessarily rational process.**

Studies on the role of research findings in decision-making have shown that many other competing factors influence decision-making (Trostle, Bronfman, and Langer, 1999; Anderson et al., 1999; Iskandar and Indrawati, 1996). Program managers and other key decision-makers will only consider implementing recommendations from research they consider to be of high quality, conducted by reputable researchers, consistent with organizational values and needs as well as with social and political context, and able to provide an adequate solution to a recognized problem with available resources. Other less concrete factors such as personal relationships with researchers (Trostle, Bronfman and Langer, 1999) or job security also affect decisions. Context is not easy to measure; yet evaluators must consider it because of its important role in the translation of research recommendations into program and policy change.

- **The term "policy change" covers a large range of actions that differ substantially in their potential impact.**

Policy includes formal government declarations, laws, and statutes which those working in policy refer to as "Policy with a capital P." In addition, policy can refer to the regulations, guidelines, norms, and standards of a given organization (which some label as "policy with a small p"). Within the same country, policies can be enacted at different levels of the program and by different processes. Decisions that are voted into law by parliamentarians or other elected officials or made by the executive branch (e.g., such as the conditions under which induced abortion is legal in a given country) will subsequently affect the entire nation. The policies of a given organization may or may not have widespread ramifications, depending on the size and importance of that organization. For example, a decision by the MOH to introduce adolescent reproductive health services nation-wide should have substantial consequences for the service delivery environment. By contrast, a change in operational policies at a small NGO would have much less potential impact. Because operations research may be undertaken with organizations of varying sizes, from national ministries of health to small, local NGOs, the methodological challenge for evaluating the effects of OR projects on policy is to establish a working definition of the type of policy that will be considered relevant in making this judgment. One possible criterion for defining a "policy change" is that the change in regulations, guidelines, norms, or standards be implemented system-wide within the organization conducting the research (e.g., throughout all the service delivery points operated by the organization).

[1] To clarify, it is possible – with an experimental design – to demonstrate that an intervention caused a certain change in behavior among the intended audience. However, the difficulty lies in determining that a specific OR project was uniquely responsible for a change in service delivery procedures or in policy. Often other events are taking place at the same time that influence the decision-making processes. For example, in Guatemala, different organizations conducted OR studies on the acceptability of DepoProvera distributed through different channels (e.g., community-based distribution and clinics). Subsequently, the MOH adopted Depo as part of its contraceptive method mix. Whereas the studies most likely contributed to the inclusion of Depo in the MOH method mix, one can not rule out the possibility that the MOH might have decided to provide Depo to its clients for other reasons as well (e.g., the approval of Depo by the Food Drug Administration (FDA) in the United States, reports of its popularity in neighboring countries, the willingness of international donor agencies to supply the method). For this reason, we state above that "demonstrating cause and effect when evaluating an OR study is virtually impossible."

- **Usually evaluators cannot measure impact until two to three years after the intervention is completed; however, in the course of this delay, other factors may intervene.**

While there is no golden rule for how long to wait to evaluate the impact of an OR study, at least two or three years are usually needed to allow adequate time for an organization to adopt and institutionalize changes based on the research. An alternative is to wait for all OR studies in a given program to end, and then evaluate them as a group. However, a time lapse of much more than three years may allow too many other changes to take place that might further complicate an evaluation. Due to high staff turnover, evaluators may find it difficult or impossible to contact and interview important informants. Many contextual changes may occur in this time and may further complicate the question of attribution.

- **The responses of key informants are by definition subjective.**

The indicators presented in this section rely on three primary data sources: key informant interviews, project documents, and site visits to observe innovations adopted as a result of an OR project. While key informants attempt to be objective, by definition their answers come from their own perspective. To minimize the bias of subjectivity, the evaluator should interview several individuals regarding the given study to increase the credibility to the information. Where disagreement occurs, evaluators may seek more information from other sources, but ultimately must use their best judgment, because they lack a systematic way to "weight" the opinions of two key informants.

- **The checklist of indicators does not adequately measure or reflect the importance of the dissemination of results.**

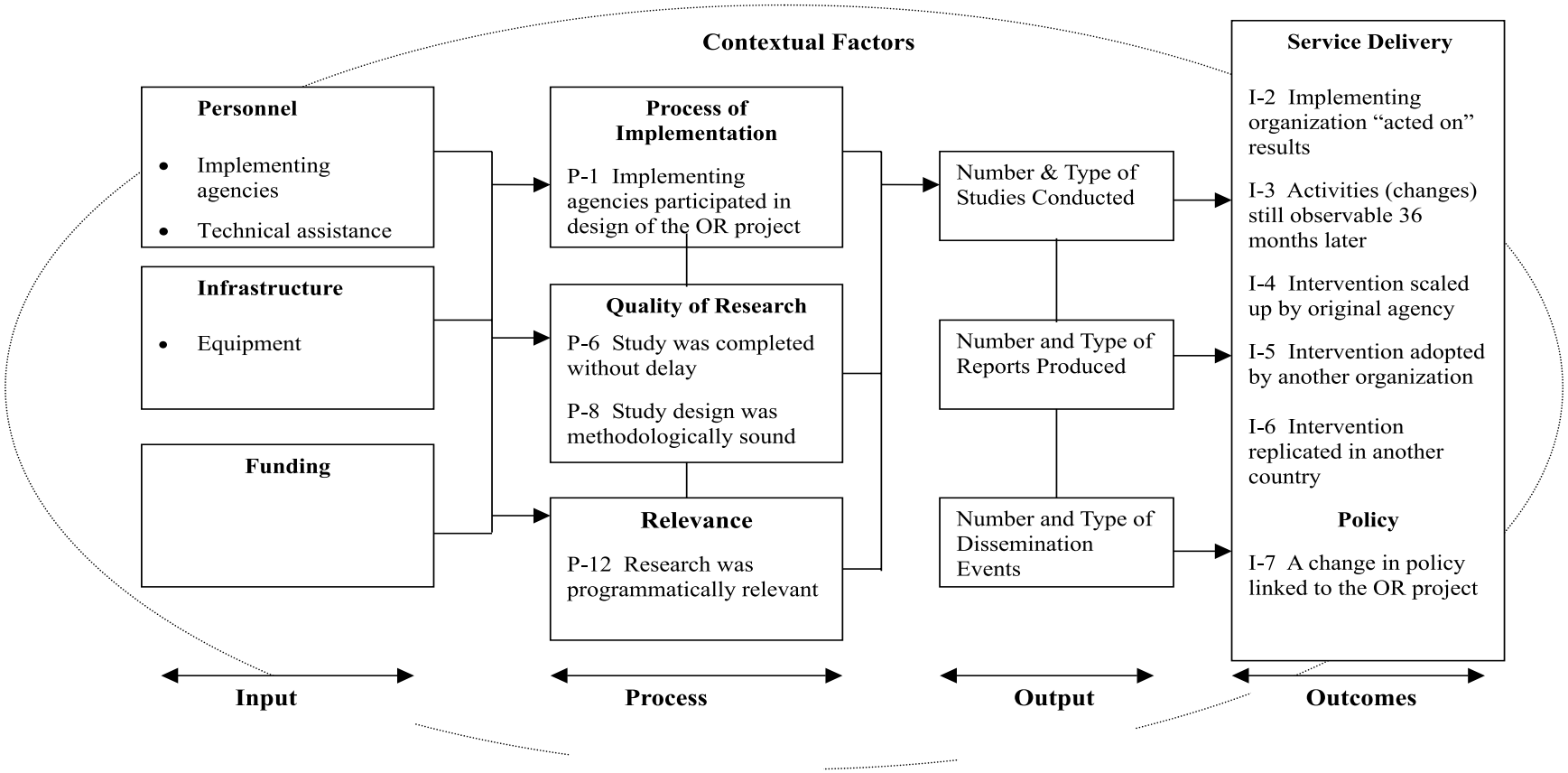
Operations research is conducted with a purpose: to use the results to improve programs. Thus, a necessary (though not sufficient) condition is that appropriate audiences learn about the results and use them in designing their own programs. For example, in 1997, a small Guatemalan NGO tested the use of a necklace to help Mayan couples correctly practice the rhythm method. Use failure rates were low, and most couples (who had wanted to use rhythm) found the method very satisfactory. The researchers widely disseminated the results of this study to other groups working with both Mayans and non-Mayan (or ladino) populations, with the result that the Ministry of Health has subsequently included this method in the range of contraceptives it provides. In contrast to this example of successful dissemination of OR results, many studies in the past have not reached the decision-makers in a position to use the results, with the result that they had minimal impact on the service delivery system.

In short, dissemination is a crucial part of the OR process. Without effective dissemination, OR cannot influence service delivery or policy as it was designed to. However, finding the **right** indicators to measure dissemination has proven elusive for several reasons. First, no standard format for dissemination applies to all situations. Although the end-of-project dissemination seminar is now fairly standard, there are multiple other channels, including face-to-face presentations to high-level officials, technical assistance to the implementing agency, and presentations at international conferences (that lend credibility to the results in the local context). Second, the number of persons they must reach to ensure effective dissemination is not fixed. In some cases, reaching a single individual in a key decision-making role may be sufficient to launch an idea within an organization. In others, reaching a large number of persons through multiple channels over time is necessary before the message penetrates. A recent OR evaluation in Guatemala tested three additional dissemination questions, but none emerged as entirely satisfactory. Thus, we wish to emphasize the importance of dissemination and underscore the challenge that faces evaluators in improving the indicators to measure it.

The Operations Research Conceptual Framework

The conceptual framework used to develop the OR indicators in this section is presented in Figure II.G.1. The framework illustrates that events occurring as part of the process of conducting the OR can affect the utilization of results. In addition, factors beyond the control of the OR project (e.g., staff turnover, changing economic conditions in the country) also determine the extent to which OR findings will be translated to changes in service delivery or policy. Figure II.G.1 includes illustrative indicators of process and impact, on which the evaluation methodology is based. The full list of indicators appears in Box II.G.1

Figure II.G.1 Conceptual Framework for the Impact of Operations Research (with illustrative indicators)



Indicator

Performance of Operations Research and Impact of Results

Definition
This set of 25 indicators serves to collectively evaluate OR studies in terms of (a) the process of conducting the study, and (b) its impact (i.e., utilizing the study results to change service delivery or to influence policy).

An additional six indicators measuring context and other factors are listed in Box II.G.1, but they provide background only; evaluators do not score them in measuring performance of the OR team.

Data Requirements
Assessment of an external evaluator based on available information. Evaluators score each of the 25 items on a scale of 1 (not at all) to 3 (very much so). If indicator I-1 is negative (the intervention was not effective), then I-3, I-4, and I-5 are non-applicable. (See example of the data collection form in Appendix F.)

Data Source(s)
Project documents, in particular the final report of the project

Interviews with key informants, including researchers (especially the Lead Investigator), program managers, and other providers in the service delivery organizations who stand to benefit from the OR, donor agency staff, policymakers and other key decision-makers.

Purpose and Issues
The set of OR indicators allows the evaluator to arrive at a set of numerical scores supported by qualitative justifications for the scores for each OR study under review. The 3-point scale for each item distinguishes among those studies that performed well (3), those that performed satisfactorily but with notable problems (2), and those that did not perform satisfactorily on the relevant indicator (1).

Although other formats are possible, the OR indicators developed under the USAID-funded FRONTIERS Program use a grid format, which doubles as a data collection tool and a reporting format. (See Appendix F) Evaluators can use the blank grid as an interview guide to ensure consistency among key informants, evaluators, and projects. Evaluators can present results for each project using the same grid format. In addition, they can summarize the numerical scores of multiple projects in a table so that one can easily compare performance of studies overall (comparing columns) or can compare specific indicators across studies (comparing rows). This enables evaluators to identify areas of consistent strength and those requiring improvement. (For an example, see Bertrand and Marin, 2001.)

The OR indicators fall into three categories: process, impact, and context/other. Process indicators relate to the conduct of the study; evaluators can assess them immediately upon completion of the study. By contrast, they should assess impact indicators three years later, although impact occurring sooner certainly "counts." In addition, the instrument contains six contextual and other indicators that provide insight into the process but are not included in the scoring because they do not reflect "performance."

While each indicator has a score, the set of indicators does not lend itself to a summary score. Indicators measure different aspects of process and impact that are not necessarily of equal importance, nor is there sufficient experience in evaluating OR to reliably weight them. Rather, evaluators can compare scores individually or in groups of indicators that measure similar aspects, such as participation of the implementing agency at various stages of the study or conduct of subsequent research.

Box II.G.1 Indicators for Evaluating OR Projects

Process indicators (note: each indicator in this set begins with a "P")

P-1 The implementing/collaborating organization(s) actively participated in the design of the OR project

The design of the OR project is the formulation of the study, which includes identifying the problem, establishing the objectives, designing the intervention, and selecting a research methodology. "Active" participation involves contributing original ideas to the work, not simply attending meetings.

P-2 The implementing/collaborating organization(s) actively participated in the implementation of the OR study

"Active participation" indicates that the organization was involved in decision-making and played a technical role in the implementation of the study, for example hiring new staff, conducting training, or analyzing and interpreting results.

P-3 The implementing/collaborating organization(s) participated in developing programmatic recommendations

This indicator asks whether these organizations participated, as well as how, for example, collaboration in report preparation, through formal meetings, and in working groups at dissemination conferences.

P-4 The study accomplished its research objectives

Each study is designed with one or more objectives. This indicator determines whether the study achieved each of its objectives.

P-5 The intervention was implemented as planned (or with some modifications)

Changes between the proposal and implementation of the intervention frequently occur and often are for the better. This indicator seeks to determine whether the organization carried out all of the activities specified in the intervention, allowing for some change in response to local realities. If not, the reviewer should identify any changes between the design and actual realization of these activities.

This indicator is not intended to penalize an organization for making modifications. Rather, it ascertains that the organization made some meaningful change in service delivery (that there was "something to evaluate"). An intervention study fails to show any change in the desired outcome for two plausible reasons: (1) the organization never implemented the intervention or implemented it so weakly that the study hardly constituted a fair test of its potential effectiveness, or (2) the organization fully implemented the intervention but it failed to show the expected results. This indicator attempts to eliminate the first possibility by determining that the intervention was in fact implemented.

P-6 The researcher(s) completed the study without delays (or other adjustments to the timeline) that would compromise the validity of the research design

Study activities are often delayed. This indicator seeks to identify delays that affected the timing of the intervention or that could have reduced the effectiveness of certain activities (e.g., a delay in training resulted

in diluting the effects of the activity; the period between intervention and final data collection had to be cut short, and thus the desired change had insufficient time to take place).

P-7 Key personnel remained constant over the life of the OR project

"Key personnel" are any personnel with a decision-making role in the design or implementation of the subproject. Such personnel include the Principal Investigator, the study coordinator, and counterparts in the collaborating agencies, including key service personnel or government officials actively participating in implementation.

P-8 The study design was methodologically sound (free of flaws that could have affected the final results)

Evaluators should assess this item based on the methodology section of the report and (if appropriate) on discussions with the researchers. Generally, the external evaluator (not a staff member of any of the participating organizations) makes an "informed decision" on this point; key informants may have less knowledge or experience to make this judgment.

P-9 The research design was feasible in the local context

"Feasible" here means "reasonable" or "manageable," a design that could be repeated without unduly draining financial or human resources. "Local context" includes not only program-related factors but also sociocultural or political factors, among others.

P-10 The implementing/collaborating organization(s) judged the OR technical assistance to be useful and provided in a collegial manner

To qualify for a full score, both elements must be positive. If, for example, the advice was technically sound, but counterparts reacted negatively to the manner in which the OR team provided assistance (e.g., in an offensive or condescending way, "imposed upon them"), then the study should receive a lower score on this indicator.

P-11 Stakeholders judge results of the OR study to be credible/valid in the local context

This indicator refers to the judgment of stakeholders (policymakers, researchers, donors, program managers). Utilization of results would be likely limited if stakeholders seriously questioned the validity of the results.

P-12 Research was programmatically relevant

The perceptions of the same stakeholders listed above determine relevance. Relevant research addresses a priority problem of the program, whether a national program of the MOH or a more local program of an NGO.

P-13 Results were disseminated to key audiences, including policymakers, program managers, service providers, and donors

All studies involve dissemination of results. This indicator seeks to determine whether the dissemination strategies used effectively reached the target audience. "Key audiences" are those in a position to act on the results (e.g., policymakers, key decision-makers or service providers in implementing/collaborating agencies, donor agency staff). In addition, dissemination efforts may reach other interested parties (e.g., students

at the local university, members of the international RH community), but the indicator refers only to those in a position to act upon the results.

P-14 Results are readily available in written form

This indicator verifies the existence of a document on the key findings of the study that is well presented (of professional quality) and is locally available in sufficient quantity. This document may appear in a variety of media (e.g. website, CD-ROM) in addition to print. Ideally, results should be available in various formats appropriate to the intended audience: final reports and journal articles for donors and the academic RH community, summaries or research briefs for decisionmakers and program managers.

Impact indicators (note: each indicator in this set begins with an "I")

I-1 The results indicate that the intervention was effective (i.e., that it improved service delivery in the areas identified by the study)

OR studies generally either test one or more interventions or they evaluate changes resulting from interventions already implemented. If all studies found the intervention under study to be effective, then research would be unnecessary. This indicator asks whether the intervention tested successfully improved front-line service delivery (e.g., increase in utilization of services, improved quality of services). Negative results can also be instructive, but they would not influence service delivery except to discontinue an ineffective strategy (see I-2).

I-2 The implementing/collaborating organization(s) "acted on" the results

"Acting on the results" consists of implementing the actual services of the intervention or the activities to support those services (e.g., training courses, development of service delivery guidelines, changes in allocation of personnel, production and testing of IEC materials, supervision, monitoring) if the intervention was effective, or not implementing or discontinuing these services and activities if the intervention was ineffective.

I-3 (If the intervention was effective and continued after the study) The activities tested under the intervention were still observable 36 months post-implementation

"Activities tested under the intervention" are those specific items mentioned in connection with the previous indicator. Where only some of the original activities are observable, the study should receive only a partial score on this indicator. In the case of an improvement that has lasted fewer than 36 months, this indicator does not apply.

I-4 (If the intervention was effective and continued after the study) The original implementing/collaborating organization scaled up the intervention in the same country

Most OR studies are conducted in a specific geographical area. "Scaling up" refers to implementing the intervention activities in additional geographical areas. It can but does not necessarily refer to expansion to the national level.

I-5 (If the intervention was effective and continued after the study) Another organization within the same country adopted the intervention

An organization that did not participate in the OR study "adopts" the intervention by implementing its primary components (see I-1).

I-6 Another country replicated the intervention

Some evidence must exist that links the original intervention to the activities carried out in the other country (e.g., program managers from other countries visited the project site and subsequently adopted similar strategies).

I-7 A change in policy can be linked to the OR project

This indicator measures legislation or other official changes that potentially affect service delivery, for example, authorization for the sale of the pill by non-medical personnel. For further discussion of policy changes, see the introduction to this section.

I-8 The implementing/collaborating organization conducted subsequent OR studies

"Subsequent OR studies" refers specifically to research activities that test interventions. OR studies do NOT include actions or program activities, such as training and materials production (described in I-2) or research for other purposes (e.g., the DHS, epidemiological research).

I-9 The implementing/collaborating organization conducted subsequent OR studies without external technical assistance (TA)

This indicator is included to reflect whether the organization has sufficient capacity to conduct these types of activities as a result of the previous OR experience and has the opportunity to do so.

I-10 The original donor funded new program activities based on the results of the OR study

New program activities are those activities tested in the intervention that the donor had not already funded.

I-11 Other donors provided new or expanded funding based on results of the OR studies

"Other donors" are those donor agencies that did not contribute financial support to the original OR project but subsequently funded the initiation or expansion of program activities – specifically, service delivery or support activities, including training, production of IEC materials, construction or renovations of facilities, and purchase of supplies and equipment. The indicator does not include funding of additional research only.

Context and other factors (note: each item begins with a "C")

C-1 Other factors (not mentioned above) that facilitated the conduct of the research project

Situations that helped complete the activity might be: strong research capacity in the counterpart organization, powerful local person or donor intent on getting answers, good relationship between researchers and program staff, or others.

C-2 Other factors (not mentioned above) that facilitated the utilization of results

This indicator refers to situations that encourage the translation of the results into programmatic actions at the field level. For example, dissemination of results coincides with the planning cycle for a new program initiative or strategy; the intervention is a good match with the organization; and a committed individual continues to provide TA or to promote the intervention beyond the project end.

C-3 Other factors (not mentioned above) that hindered the conduct of the research project

Events beyond the control of the researchers and their collaborators in local organizations may impede a study's implementation. Such factors range from contraceptive stockouts, inter-organizational or interpersonal conflicts, or financial difficulties to political changes, civil unrest, or natural disasters.

C-4 Other factors (not mentioned above) that hindered the utilization of results

Situations that limit the incorporation of research results into policy or programs may also be beyond the control of researchers and their collaborators. Some examples are the intended population opposes the intervention (e.g., believes that contraceptives cause sterility); local authorities for political reasons veto a proposed initiative or change in service delivery; or the health system is restructured.

C-5 The donor used the data from the OR study for a specific purpose

Does any evidence exist that the donor used the data for its own purposes? Donors may use results for resource allocation, funding decisions, or miscellaneous strategic planning, among other possibilities.

C-6 The study included an assessment of costs of the intervention

Evaluators should mention any data collected on the cost of the intervention, primarily for the purpose of cost-effectiveness analysis. This indicator serves informational purposes only, since all OR studies do not necessarily need an assessment of cost.

Part II.H
The Service Delivery Environment

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

Part II.H

The Service Delivery Environment

Rationale for Evaluating the Service Delivery Environment

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

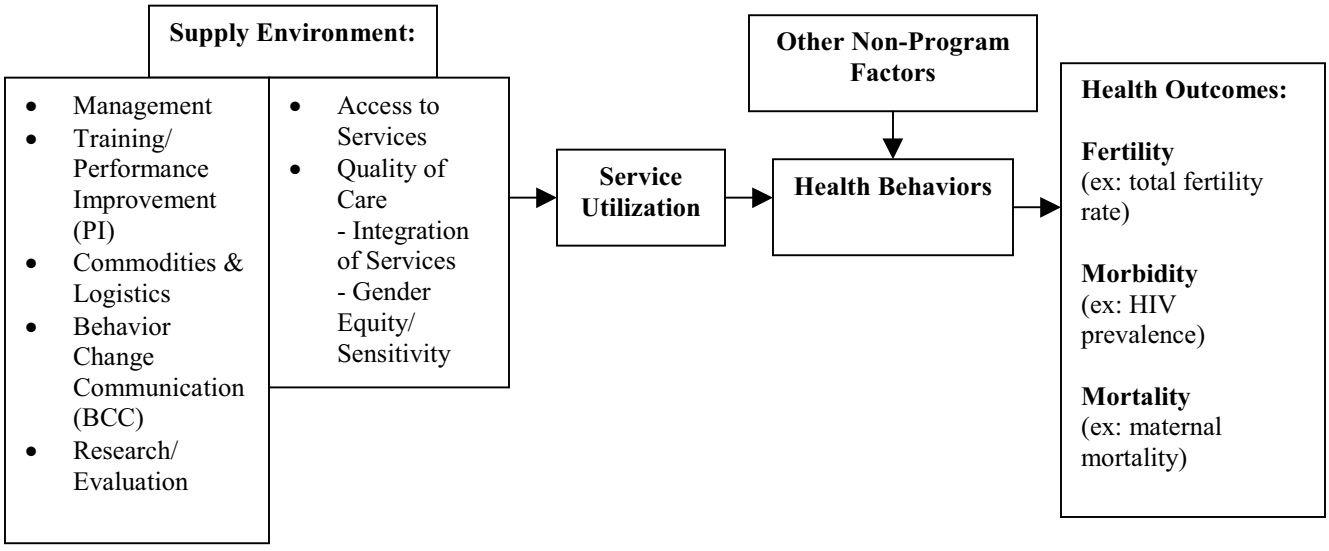
Program evaluation in the early days tended to focus either on counting activities performed (e.g., the number of persons trained, brochures distributed) or results obtained in terms of service delivery (e.g., number of clients, number of clinic visits, number of supplies distributed), rates of use (e.g.,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

This situation has changed markedly in the past decade as evaluators have increasingly focused on the two defining characteristics of the service delivery environment: access to services and quality of care.[1] Of the two, quality of care has been evaluated in far more detail than access , despite widespread recognition of the importance of the latter.

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

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

Figure II.H.1 Service Delivery Environment



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

- The Service Provision Assessment (SPA): a comprehensive set of instruments containing a standardized list of indicators to cover multiple areas of RH: family planning, STI/HIV/AIDS, safe motherhood, and child survival;

- The Quick Investigation of Quality (QIQ): a "short list" of 25 indicators (a subset of the SPA) that measures family planning only; and

- Quality Assurance (QA): an approach that uses the indicators to measure improvement on the issues specific to a given clinic or to a set of clinics as part of a systematic process.

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

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

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

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

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

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

The Situation Analysis (SA) (Miller et al., 1997), also a facility-based survey, did capture the interest and attention of the international reproductive health community. Developed by the Population Council in connection with its operations research project in Africa, Situation Analysis was designed to measure the functioning and quality of care in family planning facilities. Situation analysis was conducted in approximately 21 African countries and in 16 developing countries from other regions of the world between 1989 and 2000.[2]

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

Methodological Challenges of Evaluating the Service Delivery Environment

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

If the purpose of the evaluation is to compare quality of maternal care among countries, then standards need to be the same. Some maternal care standards (such as the use of a partograph as an early warning of insufficient uterine action and of cephalopelvic disproportion) are universally accepted. However, standards may differ by country (e.g., the number of recommended prenatal care visits). When standards are inconsistent, and evaluators lack scientific basis to validate some and to reject others, they can still compare by modifying the indicator (e.g., "percent of women who completed three prenatal care visits on time," to "percent of women who completed the recommended number of prenatal care visits").

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

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

The indicators in this section are organized as follows:

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

Part II.H.1
Access

- Average time/distance to the nearest reproductive health facility offering a specific service
- Number of service facilities offering a specific RH service per 500,000 people
- Cost of one month's supply of contraceptives as a percent of monthly wages
- Percent of facilities with non-medical restrictive eligibility criteria
- Percent of non-use related to psycho-social barriers

Part II.H.1

Access

A primary strategy of health programs is to increase access to services. Different approaches to increasing access include establishing additional facilities, training more health workers, increasing outreach activities, and so forth. Despite the widely acknowledged importance of access as a key feature of the supply environment, this factor is not routinely assessed in RH program evaluation.

Much of the previous research in this area has focused on one aspect or dimension of accessibility: geographic (or physical) access. In this context, access (or accessibility) refers to the degree of difficulty in reaching or obtaining reproductive health services. A variety of measures pertaining to the distance to supply and to service points, the time required to reach these points, and the density of service/supply points within a specified geographic area have been proposed in the literature (Chavoyan, Hermalin, and Knodel, 1984; Hermalin and Entwisle, 1985 and 1988; Tsui et al., 1992; Tsui and Ochoa, 1992). In the case of family planning, the evidence to date tends to confirm the relevance of geographic proximity to contraceptive services as an important determinant of contraceptive use; however, the strength of the relationship between proximity and use in empirical studies has not been as strong as might be anticipated (Tsui and Ochoa, 1992; Boulier, 1985).

Some researchers have distinguished between the terms "availability" (to describe whether a particular method or service is provided) and "accessibility" (to denote a continuum of effort required to obtain services) [Bertrand et al., 1995]. However, the terms are often used interchangeably, and in this *Compendium* we use the word access to reflect the degree of difficulty (or ease) in accessing services.

Access to services is not merely an issue of physical distance, but one that involves other dimensions as well (Chavoyan, Hermalin, and Knodel, 1984; Foreit et al., 1978). Foreit et al. suggested the following as relevant dimensions or elements of accessibility (the authors used the term "availability" in the original text): geographic or physical, economic, administrative, and cognitive. The indicators that follow measure access along multiple dimensions: geographical/ physical (with two separate indicators), economic, administrative, and psychosocial.

Indicator

Average Time/Distance to the Nearest Reproductive Health Facility Offering a Specific Service

Element: Geographical/physical access

Definition
The time (measured in minutes) or the distance (measured in kilometers or miles) from a respondent's place of residence to the nearest service delivery site offering a specific type of reproductive health service (e.g., antenatal care, voluntary counseling and testing, male sterilization)

Data Requirements
Information on the location of the respondent in relation to the service delivery point in question

To obtain aerial distances ("as the crow flies"), one may use a geographic positioning system (GPS). Alternatively, one can map the routes between a given community and an individual service delivery point and can (preferably) obtain measures of travel time and actual distance; these measures are superior to the self-reports of respondents or key informants.

Data Source(s)
Data from facility-based surveys analyzed in relation to data from household surveys (e.g., in the context of a DHS survey)

Purpose and Issues
Ideally, the researcher will determine the distance between the home of an average citizen in country X and the nearest facility providing a specific reproductive health service. In the past, researchers often relied on self-report of survey respondents or of community informants, both of which tended to be highly unreliable. In recent years, researchers have attempted to link the DHS household surveys with surveys of the facilities in the surrounding area in selected countries. In linking the data from the household and facility-based surveys, researchers and evaluators are, for the first time, able to accurately measure distance between these communities and service delivery points (Akin et al., 1998; Seiber and Bertrand, 2001).

There are, however, several caveats to measuring access using this linking technique. First, many DHS household surveys do not include a facility-based survey, or the facility-based survey is not linked to the household survey. Second, the human and financial resources needed to carry out a DHS with both the household and the facility-based components are considerable. Thus, in the best of cases, the linked surveys are conducted only once every 3-5 years. Third, this linking of the two surveys allows for a much more precise measurement of the time and distance between the household of the average respondent and the nearest service delivery point. However, research has shown that clients often elect to use services at some more distant point to preserve their privacy; to obtain a range of services (e.g., specific contraceptive method, or special lab procedures) not available at a facility closer to their home; or to obtain higher quality services (e.g., better client-provider communication).

To date, evaluators have studied physical access as a determinant of service utilization and use, but program managers have not routinely used it for the day-to-day monitoring of program performance, because of the time and expense associated with the above-mentioned linking procedure.

Gender Implications of this Indicator

Distance to the nearest reproductive health facility represents, on one hand, the commitment and resources of government to provide universal access to health care. It is an important variable to consider in terms of women's ability to obtain obstetrical services – maternal and child health care and family planning – particularly in areas where transportation is difficult. Distance to needed obstetrical services and lack of transport to reach a facility offering such services are key variables contributing to maternal deaths. Advocates for safe motherhood argue that these preventable deaths indicate that policy makers undervalue women's lives. Advocates employ human rights conventions to hold governments accountable for providing appropriate and accessible health services (Rosenfield, 2001). On the other hand, when women travel outside of their communities to obtain care at a remote service site, they may do so because some service facilities fail to adequately observe privacy and confidentiality, and women fear the consequences. For example, women who feel they must obtain contraceptives covertly because of perceived disapproval on the part of the husband or extended family may fear retribution if confidentiality is violated. Many small local service outlets have no potential for offering women visual privacy as they wait in line for services. However, even these facilities can observe policies and procedures to protect the woman's confidentiality in reproductive health choices and services obtained.

Indicator

Number of Service Facilities Offering a Specific RH Service per 500,000 People

Element: Geographical/physical access

Definition
The density of service facilities in a given population

Data Requirements
Information on the total number of facilities offering a specific service (e.g., contraceptives, postabortion care, voluntary counseling and testing, micronutrient supplementation) and the total population (or relevant subgroup) in the catchment area

Note: evaluators may limit the denominator to an estimate of the relevant sub-group for the service (e.g., all women of reproductive age for family planning, all pregnant women needing micronutrient supplementation, all adults 15 to 65 for HIV counseling and testing). Because of the difficulties associated with estimating the exact number of persons in need of such a service (e.g., postabortion care), evaluators may opt to use the total population in the denominator.

Data Source(s)
Program records on the service delivery infrastructure; census data on size of population in the catchment area

Purpose and Issues
This indicator gives a broad sense of the density of service delivery points for specific types of reproductive health interventions. It can be useful for advocacy purposes in creating awareness of the deficiencies in the service delivery environment for particular services.

Caveats for this indicator include the following. First, although this indicator gives a ratio of service delivery points per population, it does not reflect the geographical distribution of such points. In the case where service delivery facilities cluster in urban areas, this indicator may yield a more favorable estimate of access to services than individuals in rural areas actually experience. Second, it is easier to collect information on the availability of some services than of others. For example, many countries have fairly accurate lists of family planning services through government or NGO facilities. However, they may not track the number of pharmacies that carry contraception and other reproductive health products, and thus may underestimate the access of the population to these commodities. In more controversial subject areas, such as postabortion care, facilities may provide services but not publicize them widely, and thus may create undercounts on this indicator for those services. A third caveat is that services may exist "on paper" but not at the actual field site.

One potential use of this indicator is to help governments track progress in terms of improving the service delivery environment for the population. However crude this measure is (and how little it reflects the situation of a specific individual in that society), it does represent progress for a government to increase the number of reproductive health facilities per 500,000 in the population (assuming quality remains constant or improves). Moreover, where data on the health service environment are fairly reliable, evaluators may collect this information at relatively little cost to the user.

Indicator

Cost of One Month's Supply of Contraceptives as a Percent of Monthly Wages

Element: Economic access

Definition
"Costs" refer to out-of-pocket expenses for contraceptive supplies and services

This indicator is calculated as:

$$\frac{\text{Cost of one month's supply of contraceptives}}{\text{One month's wages}} \times 100$$

Data Requirements
Information on monthly expenditures on contraceptive supplies and services and estimated monthly income

Data Source(s)
Information from population-based surveys on service and supply costs; fees may also be available from facility records.

Purpose and Issues
This indicator provides a measure of the relative economic burden represented by monthly service and supply costs of contraceptive use. This measure applies specifically to family planning, but evaluators can adapt it to other areas of reproductive health by substituting the cost of the product in question for contraceptives in this definition.

Service and supply costs exceeding one percent of monthly wages for a significant proportion of clients may constitute an economic barrier to contraceptive use (Ross et al., 1992).

The illustrative indicator for this element was chosen from among several alternatives in large part because the data required for its computation are the most likely among the alternatives to be available in a reasonably large number of developing country settings. However, evaluators should recognize that the indicator suffers from several important limitations.

One limitation is that the indicator ignores other costs of contraceptive use that may be just as, or perhaps more, important barriers to contraceptive use than direct service or supply costs are. For example, family planning clients may also incur out-of-pocket expenses for transportation to and from the facility and (possibly) for child care, as well as opportunity costs of time spent traveling to and from the SDP and waiting for service or supplies once clients reach the facility. Thus, a more valid measure of the costs of family planning services would also include these costs in the computation of the indicator.

Another issue concerns the stream of income that evaluators should consider in computing the indicator. Since not all income (gross income) is likely to be available for use in paying for contraceptive services, a more appropriate specification of the indicator will limit the denominator of the measure to monthly disposable income. Furthermore, since men and women do not have equal access to household financial resources in many societies, a further refinement may be to limit the denominator of the measure to income or wages controlled by the client (especially female clients).

Evaluators should recognize, however, that these refinements add to the data requirements for computing the indicator. In many countries, the required information may be available only from special studies. For most practical purposes, the simpler indicator should suffice to guide program management decisions regarding the affordability of contraceptive services. In programs where cost recovery and sustainability are priority management issues, however, the added costs of gathering data required for the more refined measures may be justified.

Indicator

Percent of Facilities with Non-Medical Restrictive Eligibility Criteria

Element: Administrative access

Definition
Eligibility criteria of a non-medical nature, mandated by the service facility or organization, which limit access to RH services for specific categories of individuals

This indicator is calculated as:

$$\frac{\text{\# of facilities with non-medical restrictive eligibility criteria}}{\text{Total \# of medical facilities}} \times 100$$

Data Requirements
Eligibility criteria for services offered by the program

Data Source(s)
Program documents outlining policies and regulations; interviews with staff

Purpose and Issues
This indicator identifies the existence of barriers to services in the form of unnecessary formal program policies, regulations, and procedures; such restrictions, mandated at the policy/program level, exceed those justified on medical grounds. Typical restrictive criteria for different reproductive health services include the following:

Family Planning:

- Age, parity, marital status, and/or spousal consent;
- Requirements for blood tests or pelvic examinations prior to the distribution of oral contraceptives;
- Requirements for multiple visits to receive certain contraceptive methods (e.g., IUDs);
- Requirements of direct physician involvement in the provision of services; and
- A required waiting period of several days between counseling for voluntary surgical contraception and the actual procedure.

STIs:

- Age, marital status, spousal consent;
- Partner notification; and
- Waiting period for HIV test results.

Maternal Health:

- Age, parity, marital status, spousal consent; and
- Requirements for direct physician involvement in provision of services (when other appropriately trained clinical staff [e.g., midwives] could provide the same service).

Adolescent Reproductive Health Services:

- Age, marital status; and
- Parental notification or consent.

Postabortion care:

- Parental or spousal consent prior to treatment;
- Inappropriate judicial requirements (e.g., rape may be a legal indication for elective abortion, and women are not permitted to obtain the service without "proof" of the rape, which might be a court order);
- Provision of uterine evacuation for incomplete abortion only in an operative theatre;
- Provision of uterine evacuation for incomplete abortion only under general anesthesia;
- Provision of uterine evacuation for incomplete abortion only by a physician;
- Inappropriate clinical criteria for use of manual vacuum aspiration for treatment of incomplete abortion; and
- Inappropriate criteria for contraceptive method provision following treatment of incomplete abortion.

Administrative barriers to access occur less frequently for STI/HIV services, which tend to be provided to those who seek treatment.

Indicator

Percent of Non-Use Related to Psycho-Social Barriers

Element: Psycho-social access

Definition
The proportion of women who want to use an RH service who avoid use because of barriers of a psycho-social nature; for example, fear (of negative social stigma, embarrassment, discomfort) or social restrictions (e.g., against women traveling alone to seek services)

This indicator focuses on factors that deter a woman or man who **wants** to use a certain service or practice from doing so. For example, adolescents wishing to obtain information and services from a local clinic might be deterred from doing so by fear of being seen at the clinic or being judged by friends to have low moral standards. A woman wanting to be tested for HIV might fear some type of violent reaction from her husband should he find out. Note: this indicator does not apply to the case where the individual is not even interested in the service.

This indicator is calculated as:

$$\frac{\text{\# who report non-use due to psycho-social barriers}}{\text{Total \# of clients seeking RH services}} \times 100$$

Data Requirements
Information on reasons for non-use of services or RH practices among individuals interested in but avoiding a particular service or practice

Data Source(s)
Population-based surveys; alternatively, focus group discussions (although they do not yield a quantitative result)

Purpose and Issues
This indicator provides a measure of the extent to which access to available RH services is limited by barriers of psychological, attitudinal, or social origin.

Given the context-specific nature of factors falling under this heading, the reasons for non-use of services will likely vary from setting to setting. Thus, the exact numerical figure associated with a particular barrier or factor may be less important than the rank ordering of problems. In view of this, data from focus groups (that do not provide results in quantitative terms such as percentages or ratios) may be more valuable in identifying barriers of this type than data derived from structured interviews are.

Part II.H.2
Quality of Care

A. Service Provision Assessment (SPA)

- Percent of facilities prepared to provide the essential service
- Percent of facilities with systems that support quality service delivery (assessed separately for each service)
- Percent of facilities where _ percent of clients receive the service that meets the expected standards

B. The Quick Investigation of Quality (QIQ)

- Quick Investigation of Quality (QIQ)

C. The Quality Assurance Approach (Applied to Maternal Health)

- Percent of newborns receiving immediate care according to MOH guidelines
- Facility perinatal mortality rate
- Percent of deliveries in which a partograph is correctly used
- Percent of deliveries in which a perinatal clinical record was properly completed
- Percent of mothers examined every 30 minutes during the first two hours after delivery

Part II.H.2

Quality of Care

Quality of care (QC) has been a central focus of the international family planning programs for the past decade. Consistent with the major theme of the 1994 ICPD Conference for more client-focused services, many governments and NGOs worldwide designed and implemented initiatives to improve QC within their service delivery environment. The paradigm behind much of this work is the Bruce/Jain framework, which defines the six elements of quality of care: choice of methods, information given to clients, technical competence, interpersonal relations, follow-up and continuity mechanisms, and the appropriate constellation of services (Bruce, 1990).

With the increased interest in quality that developed during the 1990s, many organizations searched for means to **evaluate** quality for the purposes of (1) underscoring to staff the importance of quality, (2) identifying problems they needed to address, and (3) measuring the effectiveness of interventions designed to improve quality.

In this *Compendium*, we present three approaches to measuring quality of care: the Service Provision Assessment (SPA), the Quick Investigation of Quality (QIQ), and the Quality Assurance (QA) Approach.

A. Service Provision Assessment (SPA)

The most comprehensive tool for evaluating quality of care is the Service Provision Assessment (SPA), a national survey of a representative sample of public and/or private facilities that provide maternal, child, and reproductive health service. In addition to quality, it also measures the general functioning of a network of clinical facilities, and it provides an inventory of available equipment and supplies. The SPA provides a means of assessing strengths and weaknesses in the service delivery environment, which (1) may explain the impact (or lack thereof) of the services on health behaviors in the catchment area, and (2) may guide policy makers and program administrators in prioritizing resources for better health outcomes.

The SPA uses four different data collection methods. The first is an ***inventory*** of resources and support services, which provides information on the "preparedness" of a facility to provide each of the priority services at an accepted standard of quality. As part of the inventory (also known as a facility audit), interviewers ask staff about their qualifications, training, perceptions of the service delivery environment, and related issues.

The second is a ***provider interview***, during which interviewers ask health service providers for information on their qualifications (training, experience, continued education), supervision they have received, and perceptions of the service delivery environment.

The third is ***observation*** of services provided. The observation assesses the extent to which service providers adhere to service delivery standards.

The fourth is ***exit interviews*** with clients who have received services. The exit interview assesses the client's understanding and perceptions of the consultation/examination, as well as recall of instructions regarding treatment or preventive behaviors. Recall of key messages increases the likelihood that the client will successfully follow treatment or will perform the preventive behaviors that optimize healthy outcomes.

The SPA not only measures quality of care but also overall functioning of the facility, as reflected by the set of questions it addresses:

1) To what extent are the surveyed facilities prepared to provide the priority services? (availability of resources)

2) To what extent does the service delivery process follow generally accepted standards? (care process)

3) To what extent do support systems for maintaining or improving the existing services exist, and how well are they functioning? (support services)

4) What are the issues the clients and service providers consider relevant to their satisfaction with the service delivery environment?

The SPA provides the following information on all four types of health services: family planning, STIs, maternal health, and child health:

- Preparedness to provide good quality services;
- Adherence to standards for provision of services; and
- Client understanding of the consultation.

Other data, specific to these topic areas, are as follows:

Sexually Transmitted Infections (STIs)

- Preparedness to offer both basic and higher level diagnosis and treatment of suspected STIs; and
- Preparedness to diagnose and to treat HIV/AIDS-infected persons, including specific program components related to HIV/AIDS prevention, treatment of opportunistic infections, palliative treatment, and family and client support services.

Maternal Health

- Preparedness to provide good quality basic and higher level antenatal care; and
- Preparedness to provide basic and higher level delivery services.

Child Health

- Preparedness to provide good quality immunization services; and

- Preparedness to provide good quality basic diagnosis and outpatient treatment of the seriously ill child.

For each of these health services, the SPA covers the following specific components:

(1) Staff: What is the qualification of staff who provide the service? Have the service providers received periodic continuing education on relevant topics, and how recently has training occurred? Have the service providers received a minimal level of supervision?

(2) Process: Do protocols and standards of practice for each service meet generally accepted quality standards for basic as well as advanced level services at referral facilities? Do providers adhere to the standards of practice for service delivery? The process assessed includes procedures followed, components of physical examinations, as well as the information exchanged between the provider and client (history, symptoms, advice). The SPA assesses if the process during service delivery meets the standards; it does not assess if providers correctly diagnose the problems.

(3) Facility resources, equipment, and supplies: What specific equipment and supplies are available for meeting various levels of service delivery? What are the basic systems and infrastructure that may impact utilization and capacity to provide standard level services? Are the elements required to provide the services meeting the minimum standard, present, functioning, and in the appropriate location for use during service provision? Are there systems for maintaining adequate availability of supplies (inventories; appropriate storage, equipment maintenance and repair/ replacement systems), and is there evidence of their effectiveness?

(4) Systems for evaluating and monitoring services: Are routine information systems up-to-date and able to provide basic client and service provision data? Are there systems for monitoring community coverage if community coverage is expected of the facility?

(5) Facility management: Does the facility have basic management systems in place, and do they include community representation? Does the facility participate in any financing mechanism that impacts the cost to the community or client?

(6) Client understanding: What information regarding the consultation, instructions, or follow up can the client recall?

(7) Service provision environment: Does the facility collect very basic information about the problems staff think should be addressed to improve their working situation and services? Does the facility collect data revealing the opinion of clients regarding issues related to satisfaction with their consultation and the service delivery environment?

The SPA yields data from four different instruments for four areas of reproductive and child health. (In fact, the sheer volume of data generated in this type of survey led to the creation of the QIQ, an instrument less comprehensive in scope, based on 25 indicators and focused exclusively on family planning.) Although evaluators provide the most complete picture of quality and service availability when they assess the four services areas together, a number of factors (e.g., limited human and financial resources, local interest in a particular service area) may dictate a limited scope of the SPA. The SPA has been developed so that evaluators can use each of the modules separately. Appendix G presents a full list of the indicators available from the SPA for each of the four areas: family planning, STI, maternal health, and child health. In addition, several composite indicators – presented on the following pages – assess the service delivery environment across these service areas.

The SPA module measures the service delivery environment. First, it identifies strengths and weaknesses of a set of clinical facilities at a given point in time, and if repeated, the data can demonstrate changes over time (as in Tanzania in the 1990s). Second, if a program is not achieving its desired outcome, the SPA data may reveal service-related reasons for this shortcoming. Third, the SPA data play an important role in a relatively new approach to evaluating program effects. Researchers link facility-based data from the SPA to household-level data from the DHS to demonstrate that changes (improvements) in the service delivery environment improve outcomes at the population level.

In this volume, we present three composite indicators to capture the functioning of subsystems a clinic needs

to achieve an objective. For example, to ensure proper sterilization of equipment, several conditions must be present: sterilization equipment is available, source of heat is available, provider can use the equipment, and items are properly stored after sterilization, among others. Whereas each individual indicator provides useful information, evaluators must combine specific items into composite indicators to evaluate the functioning of a given subsystem. The three composite indicators on the following pages represent Macro International's proposed approach to converting the findings on individual items from the SPA into composite scores. Whereas the composite indicators are still in the development stage (in terms of weighting of items, summary scores across items, and related details), most of the individual items that comprise the summary scores have been tested in field applications.

Indicator

Percent of Facilities Prepared to Provide the Essential Services

Element: Service availability (family planning, STIs, maternal health, child health)

Definition
This composite indicator combines several indicators for preparedness to provide a given service at a minimum standard. It measures the percentage of facilities with each of the following:

- All essential equipment present, functioning, and located in the service delivery area or in reasonable proximity for utilization;
- All essential medications and supplies present; and
- At least one set of staff members assigned to the facility who have either professional or in-service training that qualifies them to provide the service following standard procedures.

Data Requirements (Service specific)
Inventory of functional status of equipment, and location in relation to the service delivery area; inventory of medications and supplies; inventory of all assigned staff and their basic qualification; and information on qualification and continued training related to the service provided, for staff providing the service on the day of the visit

Evaluators will use core definitions for "essential items" (based on generally accepted standards for practice) to compare preparedness to provide the service across various countries. However, evaluators may also adjust the essential items to reflect individual country service standards.

Data Source(s)
Interviews with staff and with persons in charge of each service; facility inventory and physical verification of equipment and supplies

Purpose and Issues
This indicator provides information on the preparedness of a facility to offer a specific service with a minimum standard of quality. It can be used to identify gaps between planned service standards and actual resources on-site, which are required to provide the service to the given standard. These gaps will most often reflect problems with the support systems, such as the commodities and logistics systems, staff allocation, or staff training.

One important aspect of this indicator is that it assesses the resource availability at the delivery site. Evaluators interpret as "not available" any essential items absent from the service delivery area (e.g., a blood pressure gauge sitting in another service delivery area, or supplies locked in a storage closet). This approach more realistically assesses a facility's capability for meeting service standards.

Preparedness is an important measure, because a facility that fails to meet the indicator standard is incapable of providing the service to the established standard.

Evaluators may analyze facility service availability by type of facility, by geographic area, or by sector (e.g., government, private for profit, private non-profit).

The SPA provides an assessment at a single point in time. However, one can evaluate changes over time by repeating the SPA at a given facility or at a set of facilities at periodic intervals (e.g., 3-5 years). With a representative sample, the picture of preparedness should accurately reflect the overall situation at a given level of analysis (e.g., region or nation), although it may not provide an accurate picture of an individual facility.

Several caveats warrant mention. First, the SPA provides data on the extent to which a given facility or set of facilities is "prepared" to provide service to a given standard. It does not measure the actual delivery of the service (e.g., whether the service meets the standard of quality during a given client visit). Similarly, the SPA does not measure whether service providers followed the right process and made the right decisions regarding the course of action for specific clients.

Second, the SPA does not assess preparedness in relation to the potential demand for the services based on normal client load. If the client load is greater than the facility can handle, the overload can increase waiting times, cause staff to neglect tests or other processes for which equipment is in short supply, or cause the clinic to have inadequate supplies to meet client demand. Thus, service quality declines.

Third, preparedness provides only one measure of access; this indicator fails to capture other aspects of access, such as hours the service is provided, and geographic, cultural or financial constraints.

Indicator

Percent of Facilities with Systems That Support Quality Service Delivery (Assessed Separately for Each Service)

Element: System support for quality

Definition
This composite indicator combines several indicators for support systems required for quality service delivery. It measures the percentage of facilities with each of the following:

- Resources and systems adequate for preventing transmission of infection;
- Service-specific written protocols for service delivery;
- Visual aids for educating clients about the service;
- Service providers with in-service training on a related topic within the prior 12 months;
- Service providers supervised while providing the service within the prior 6 months;
- Information systems providing basic information on clients and services provided;
- Use of individual client records/charts; and
- Systems for monitoring service coverage (where relevant).

Data Requirements (Service specific)
Equipment and standard procedures meeting requirements for infection control; relevant infection control supplies in service delivery areas; client information registers/forms and copies of protocols; and information on in-service training and supervision for service delivery staff

Evaluators may use core definitions for "minimum requirement" (based on generally accepted standards) for each of the aspects of quality measured so they can compare system support for quality services among various countries. However, evaluators may augment the essential items and details within each item with country-specific standards.

Data Source(s)
Interviews with staff and with persons in charge of each service; facility inventory and physical verification of equipment and supplies

Purpose and Issues
This indicator provides information on whether a facility has established systems to support and maintain minimum standards of quality. The indicator also assesses how the system functions (e.g., whether staff training is recent; whether staff know the appropriate procedures for disinfection/sterilization). It can identify gaps between planned systems and actual practices. These gaps will most often reflect problems with the support systems, such as the commodities and logistics systems, management and supervision system for the facility, or staff training.

Evaluators can analyze each component of the support system by type of facility, by geographic area, or by sector system (e.g., government, private for profit, private non-profit). Each of these factors may uniquely influence specific items (e.g., availability of items, access for supervision, establishment of acceptable standards) required to support quality of care.

Although this indicator provides information on the existence of systems that should support quality of services, the fact that a facility meets the indicator criteria does not mean that the facility functions to standard and achieves good quality results. For example, although staff may report following appropriate procedures for sterilizing equipment, the SPA does not observe the actual sterilization process. Similarly, although the SPA assesses the staff's level of training, it does not evaluate actual performance and expertise on the job. The assessment of supervision likewise stems from facility reports, not from observation of practice.

Gender Implications of this Indicator

There are many overlaps between services that are gender sensitive and those that provide high levels of quality of care. The elements of a quality program should include gender equitable treatment by providers, respect for the client's dignity and attention to the client's rights to privacy, confidentiality and free and informed choice. If the elements of a gender-equitable program are included in the parameters that measure quality of care, the percentage of facilities that support quality service delivery will also reflect those that give explicit attention to gender.

Indicator

Percent of Facilities Where __ Percent of Clients Receive the Service That Meets the Expected Standards

Element: Adherence to service provision standards

Definition
This composite indicator combines several service delivery indicators for providing good quality client consultation and examination. The indicator measures the percentage of facilities in which providers use standard procedures in their interaction with clients, such as:

- Eliciting essential history and client information;
- Conducting essential physical examination and monitoring;
- Providing treatment (if relevant) or intervention that followed standards, based on the client assessment; and
- Providing essential information to the client during counseling.

Data Requirements
Content of client history elicited during consultation; specific examination procedures conducted; treatment or intervention prescribed/provided; and content of counseling provided to the client

Core definitions exist for "standard processes" providers must follow during service provision (processes based on generally accepted standards for practice); evaluators will use these definitions to compare availability of services between various countries. However, evaluators may augment the essential items with country-specific items that meet their service provision standards.

Data Source(s)
Observation of client-provider interactions

Purpose and Issues
This indicator provides information on the extent to which service providers adhere to specific standards for service delivery. Evaluators should interpret it in relation to the previous two indicators (which measure essential resources and apply to systems supporting the provision of quality of care). If the facility fails to meet the minimum standard for resources and systems to support the provision of quality services, then providers will have great difficulty in adhering to service provision standards. By contrast, if the facility does meet these minimum standards, but the provider fails to adhere to service provision guidelines, this result signals a need to review the adequacy of supervision, training/skills maintenance, and management/leadership.

Evaluators can analyze service delivery procedures followed by type of facility, by geographic area, or by sector system (e.g., government, private for profit, private non-profit). This indicator measures whether the service provision process follows accepted standards. The indicator does not, however, evaluate the accuracy with which providers interpret the information and the appropriateness of subsequent actions.

The major difficulty in measuring this indicator relates to the biases inherent in observing client-provider interactions, as well as issues related to sampling (discussed in the introduction to this section).

B. The Quick Investigation of Quality (QIQ)

In contrast to the SPA, which spans multiple areas of reproductive health, the QIQ was designed explicitly for family planning.

Because quality of care is a complex, multi-faceted issue, evaluators could use literally hundreds of indicators to measure it, and the resulting volume of data would be overwhelming. For practicality, the QIQ was designed as a "short list" of 25 indicators relevant to client behavior and outcomes. These key indicators – while by no means comprehensive – serve as "markers" for a wide range of behaviors or conditions; facilities that perform well on these should perform well on the larger set. For a complete description of the QIQ, see Sullivan and Bertrand (2000). Table II.H.2.1 lists the 25 indicators; the numbering (I-1 to I-25) is consistent with previous publications on the QIQ.

Evaluators can measure each indicator by one (or more) of three methods of data collection, as shown in Table II.H.2.1. The three methods are the same as for the SPA:

- **Facility audit** with selected questions to the program manager;
- **Observation** of client-provider interactions and selected clinical procedures; and
- **Exit interviews** with clients departing from the facility (and previously observed).

The facility audit measures the **readiness** of a facility to deliver quality services. Because evaluators observe client-provider interaction, they can directly assess the actual level of **quality given.** The exit interview provides feedback from clients on their perceptions of the **quality received.** The basic premise is that improved quality of care leads to service utilization, contraceptive adoption, and contraceptive continuation. The items included in the QIQ were drawn largely from the Situation Analysis (Miller et al., 1997) and were later reformulated to be consistent with the SPA (i.e., most of the items contained in the QIQ are included in the SPA instrument for family planning).

Table II.H.2.1 below presents the short list of QC indicators and identifies the type of data collection instrument appropriate for each indicator. These indicators measure five of the six elements of the Bruce/Jain framework (all but "an appropriate constellation of services").

Indicator

Quick Investigation of Quality (QIQ)

Definition
The QIQ is a set of 25 indicators that collectively measure quality of care in family planning programs

Data Requirements
The ratings or assessments of an external evaluator (in the case of the facility audit, observation guide, and mystery client) and self-report on the client exit interview

Data Source(s)
Facility audit; client exit interview; observation guide; and/or mystery client. Note: the first three draw heavily on similar instruments used in Situation Analysis (Miller et al., 1997).

Purpose and Issues
The QIQ is a low-cost, practical instrument that assesses quality of care in family planning programs. Although specific to FP, it has been adapted to related RH topics in several instances (Sullivan and Bertrand, 2000).

The instruments have been tested to date in five countries (Ecuador, Turkey, Uganda, Zimbabwe, and Morocco). The field experience indicates that data collection for obtaining these indicators is practical and that the results are relatively consistent across instruments (i.e., observation and client exit interview [Bessinger and Bertrand, 2001]).

Ideally, one would like to develop a summary score for the full set of indicators. Although it will continue to be valuable to report the strengths and weaknesses of specific facilities or a network of facilities, a summary score will facilitate comparisons among facilities in a network or in a given facility over time. However, to justify the development of a summary score, all elements must "move in the same direction." An in-depth analysis of the data from the field tests in Ecuador and Zimbabwe indicated that most but not all of the indicators were positively correlated; as a result, no summary score was calculated (Sullivan, Rice, and Bertrand, 2001).

Table II.H.2.1 Short list of QIQ Indicators

Indicator Number	Indicator	Client Exit Interview	Observation	Facility Audit
	PROVIDER			
I-1	Demonstrates good counseling skills (composite)	✓	✓	
I-2	Assures client of confidentiality		✓	
I-3	Asks client about reproductive intentions (more children? when?)	✓	✓	
I-4	Discusses with client the method she prefers	✓	✓	
I-5	Mentions HIV/AIDS (initiates or responds)	✓	✓	
I-6	Discusses dual protection (method use)	✓	✓	
I-7	Treats client with respect/courtesy	✓	✓	
I-8	Tailors key information to the particular needs of the specific client	✓		
I-9	Gives accurate information on the method accepted (instructions for use, side effects, complications)	✓	✓	
I-10	Gives instructions on when to return	✓	✓	
I-11	Follows infection control procedures outlined in guidelines		✓	
I-12	Recognizes/identifies contraindication consistent with guidelines		✓	
I-13	Performs clinical procedures according to guidelines		✓	
	STAFF (other than provider)			
I-14	Treats clients with dignity and respect	✓		
	CLIENT			
I-15	Participates actively in discussion and selection of method (is "empowered")	✓	✓	
I-16	Receives her method of choice	✓	✓	
I-17	Believes the provider will keep her information confidential	✓		
	FACILITY			
I-18	Has all (approved) methods available; no stockouts			✓
I-19	Has basic items needed to deliver methods available through SDP (sterilizing equipment, gloves, blood pressure cuff, specula, adequate lighting, water)			✓
I-20	Offers privacy for pelvic exam/IUD insertion (no one can see)	✓	✓	✓
I-21	Has mechanisms to make programmatic changes based on client feedback			✓
I-22	Has received a supervisory visit in past __ months			✓
I-23	Stores adequately contraceptives and medicines (away from water, heat, direct sunlight) on premises			✓
I-24	Has state-of-the-art clinical guidelines			✓
I-25	Has waiting time acceptable to clients	✓		✓

Box II.H.2.1 Definition of 25 QIQ Indicators

I-1: Provider demonstrates good counseling skills (composite of I-2 to I-9)
Good counseling skills are defined in terms of *how* the provider communicates with the client and *what* information the session covers. A provider who demonstrates good counseling skill receives a "yes" score on indicators I-2 to I-9.

I-2: Provider assures client of confidentiality
"Confidentiality" is assurance that the provider will keep private the information the client shares during the session. That is, information such as client records are secure (e.g., in a filing cabinet with a lock) and the provider shares only that information relevant to the case in an appropriate manner (e.g., within the clinic setting, where other clients at the clinic cannot overhear information).

I-3: Provider asks client about family planning intentions (more children? when?)
The provider inquires about family planning intentions: *desire* for more children and *timing* of future births. *Desire* refers to whether or not a woman would like a/another child, and *timing* relates to when she would like to have a/another child.

I-4: Provider discusses with client the method she prefers
"Method preference" is the contraceptive method the client wishes to use when she enters the clinic. The client may prefer a temporary method such as the Pill, the injectable, or the condom if she wishes to space births. Alternatively, she may prefer a long-term method such as the IUD or implant, or a permanent method such as female sterilization if she wishes to limit her births.

I-5: Provider mentions HIV/AIDS (initiates or responds)
HIV (Human Immunodeficiency Virus) is the precursor to the disease AIDS (Acquired Immunodeficiency Syndrome). A provider mentions HIV/AIDS if s/he discusses how the disease is spread and/or how to protect oneself (condom, abstinence, fidelity to an uninfected partner).

I-6: Provider discusses protection (dual method use)
"Dual protection" refers to the use of condoms (to prevent STIs and HIV/AIDS) alone or in addition to another contraceptive method to prevent unwanted pregnancy. Although condoms can prevent both disease and pregnancy, they have not been the "contraceptive of preference" for most couples. (This trend is changing in some countries in response to HIV/AIDS.) Thus, some individuals select or continue to use a highly effective contraceptive for pregnancy prevention, in addition to condoms.

I-7: Provider treats client with respect/courtesy
The manifestation of "courtesy and respect" may differ from culture to culture. In general, courtesy involves greeting a person in a culturally appropriate manner (e.g., exchanging greetings, shaking hands), communicating with words or gestures that the person is welcome, addressing him/her in a positive, non-degrading manner, and so forth.

I-8: Provider tailors key information to the particular needs of the specific client
The provider "tailors" key information to the client's needs by assessing the background of the client (through reviewing the client record and/or through questioning the client about herself) and by adapting the information discussed in the visit accordingly. For example, the provider may ask the client about her reproductive intentions and may only discuss temporary methods if the client reports that she wishes to have more children in the future.

I-9: Provider gives accurate information on the method accepted (instructions for use, side effects, complications)

"Accurate" information is factually correct. In the context of the provision of contraceptives, a provider should discuss current standards on how to use the method, side effects, and any complications that may arise as a result of using the contraceptive method.

I-10: Provider gives instructions on when to return

The provider informs the clients when they should return for their next visit.

I-11: Provider follows infection control procedures outlined in the guidelines

"Infection control" relates to maintaining aseptic conditions in the clinic environment. Some examples are washing hands, wearing sterile gloves, and sterilizing equipment after each use. "Guidelines" are the standards or protocols specified by the program or by national or international standards (or a combination of all three).

I-12: Provider recognizes/identifies contraindications consistent with guidelines

Contraindications are those client attributes that make unsafe her use of a particular contraceptive method. For example, if a client has an STI/HIV, a provider should not advise her to use the IUD because it may lead to pelvic inflammatory disease.

I-13: Provider performs clinical procedures according to guidelines

Clinical procedures are specific actions (other than infection control procedures above) that the program recommends. For example, a common guideline is that providers explain certain aspects of the physical examination to the client.

I-14: Staff treats client with dignity and respect

The "staff" refers to the clinic personnel (other than the provider) whom the client interacts with at the clinic. The staff includes the personnel who admit patients and other auxiliary staff, such as health educators, social workers, and the receptionist. A client receives respect if the staff greet her in a friendly manner and treat her with culturally appropriate courtesy.

I-15: Client participates actively in discussion and selection of method (is "empowered")

Clients "actively participate" in the discussion when they ask questions and volunteer personal information. "Empowerment" refers to a sense of personal agency in a given situation. In the context of health care, a client must feel the confidence to ask for the care she needs. Clients demonstrate empowerment when they ask questions and feel comfortable to select the method they feel is best for them (rather than feeling they must accept the provider's recommendation).

I-16: Client receives her method of choice

The client's preferred method refers to the method she wants when she first walks into the clinic. A client receives her method of choice if she actually receives that method the day of her visit, or if she is referred for or prescribed that method. (Note: This indicator is difficult to track in cases where a client must return for follow-up care or when she is referred elsewhere or prescribed a method.) Note: she may not receive her method of choice if it is medically contraindicated.

I-17: Client believes the provider will keep her information confidential

Confidentiality is the extent to which personal information is kept private within and outside the clinic system. Through her visits to the clinic, a client may come to believe that the clinic will or will not protect her personal information. Information is treated confidentially when the providers use the client names only when necessary, secure client records, and discuss client cases only as necessary in the clinic setting and not

in the broader community. (Note: the field tests of these instruments indicates that provider assurance of confidentiality [I-2] may fail to translate into the client belief that information will remain confidential.)

I-18: Facility has all (approved) methods available; no stockouts

Approved methods are those a clinic is licensed or authorized to carry. For example, some clinics may lack clinical staff trained to administer methods requiring substantial clinical training to deliver (e.g., male or female sterilization, IUD insertion). A "stockout" refers to the inability to give a contraceptive to a client because it is unavailable at the clinic on the day the client visits. Stockouts can vary by method, and by duration and by frequency. For example, a clinic may reliably provide the pill, but have a stockout of condoms for one week every month.

I-19: Facility has basic items needed to deliver methods available through the service delivery point (sterilizing equipment, gloves, blood pressure cuff, specula, adequate lighting, water)

"Basic items," items essential to safely deliver contraceptive methods, are available if at least one is stocked and usable. Necessary equipment varies for each contraceptive method. A clinic should have the minimum equipment required for each method. (Note: one must distinguish between a basic list of instruments and a comprehensive list of instruments. Users of this tool have found it more practical, when they monitor QC, to limit this question to a "short list" of essential items, rather than to include an exhaustive list typically used for supervision.)

I-20: Facility offers privacy for pelvic exam/IUD insertion (no one can see)

Visual privacy refers to an area secluded from others for a physical exam. A private area is a room with a door or a curtained-off area.

I-21: Facility has mechanisms to make programmatic changes based on client feedback

Mechanisms refer to the means available to make needed changes in the program. Some mechanisms for obtaining client feedback include: suggestion boxes, informal interviews with clients, and community meetings with clients from the catchment area. Staff/management must then review the suggestions and decide to act (or not).

I-22: Facility has received a supervisory visit in the past __ months

A supervisor visit – a periodic visit from the program manager or supervisor – ensures that facilities comply with current standards and guidelines. The frequency of these visits may vary according to program needs (e.g., a minimum of once every six months).

I-23: Facility adequately stores contraceptives and medicines (away from water, heat, direct sunlight) on the premises

"Adequate storage" refers to an area set aside to maintain the integrity of contraceptives and of medicines. It is operationally defined as an area away from water, heat, and direct sunlight, and off the ground.

I-24: Facility has state-of-the-art clinical guidelines

"State of the art" means up-to-date information reflecting the current and international guidelines regarding clinical protocols. State-of-the-art guidelines ensure that recommended practices and procedures are based on the most up-to-date research.

I-25: Facility has waiting time acceptable to clients

"Waiting time" refers to the amount of time the clients must wait from the moment they enter the clinic until they see the primary provider. "Acceptable" is based on self-report (i.e., was the "wait time" too long or satisfactory). An acceptable waiting time may differ from country to country; clients come to "accept" what they view as inevitable. Others may define an acceptable wait time as fewer than 30 minutes.

C. Quality Assurance Approach (Applied to Maternal Health)

The Quality Assurance Approach

The Quality Assurance (QA) approach to addressing quality of care issues incorporates three core quality assurance functions: **defining quality, measuring quality,** and **improving quality** (QAP/URC, 2001a.) The QA triangle (see Figure II.H.2.1) effectively illustrates the synergy between these three QA functions. Each core function actually represents a constellation of activities, as explained below. The triangle shape indicates that rather than a unique sequence of steps that initiate QA activities, all core functions need to take place in a balanced manner for a QA strategy to be effective. The greatest impact on quality of care results only when all three functions are implemented in a coordinated fashion. In this *Compendium*, we focus on issues relevant to the measurement of quality.

Figure II.H.2.1 Core Components of a QA Approach

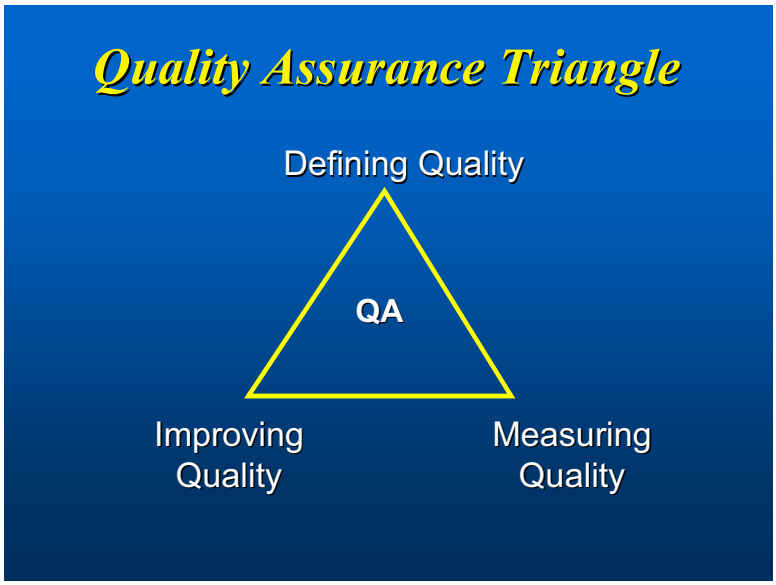
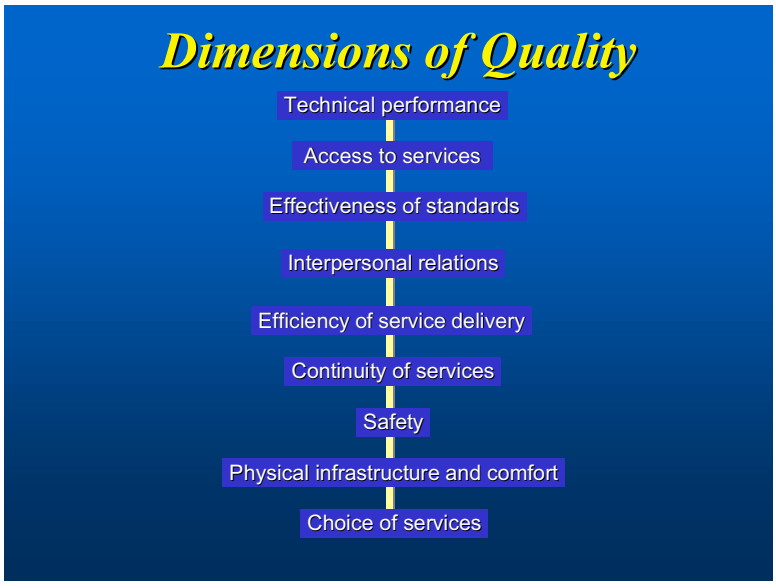


Figure II.H.2.2 Dimensions of Quality



Defining Quality means developing expectations or standards of quality. Standards can be developed for inputs, processes, or outcomes (expected outputs, results or impact on health status); they can be clinical or administrative. Standards state the expected level of performance for an individual, a facility, or an entire health care system. A good standard is reliable, realistic, valid, clear, and measurable. Standards of quality can be developed for each of the nine dimensions of quality shown in Figure II.H.2.2, which cover widely recognized attributes of quality of care. Clinical care standards should be based on the best scientific evidence available. The definition of quality standards also includes stakeholder perception and expectations of quality (including client and community).

If standards do not exist, practitioners must design or adapt them from international standards. Although standards are context specific, universally accepted standards are often a good starting point for developing local standards. Sometimes, even when national standards exist, they must be refined or made operational for local use.

Improving Quality uses quality improvement methods (problem solving, process re/design or re-engineering) to close the gap between the current and the expected level of quality (defined by the standards). This core function applies quality management tools and principles to: 1) identify/determine what one wants to improve; 2) analyze the system of care/problem; 3) develop a hypothesis on which changes (solutions) might improve quality; 4) test/implement the changes to see if they really yield improvement; and 5) based on the results of testing, decide whether to abandon, modify, or implement the solutions (QAP/URC, 2001b).

Measuring Quality consists of quantifying the current level of performance or compliance with expected standards. This process requires identifying indicators of performance, collecting data, and analyzing information. Measuring quality is inextricably linked with defining quality because the indicators for measuring quality are related to the specific definition or standard of quality under study. When standards define quality, measuring quality requires assessing the level of compliance with standards. Hence, measuring quality is easier with a clear definition or standard, because the indicators are directly derived from the expression of the standards. Likewise, measuring quality leads directly to identifying areas for improvement or enhancement – the first step in quality improvement. A few key points in the measurement deserve highlighting; the details appear in different publications (QAP/URC, 2000a and 2000b):

- If one starts the QA approach with measuring quality, the scope of measurement should be limited to what the system is able/willing to improve (i.e., a quality improvement objective must be defined).
- The QA team must be realistic about what data the team can readily collect at the facility level or across the system. A simple performance monitoring system with a limited number of indicators related to the improvement goal is usually very effective.
- Measurement strategies, such as special surveys, self-assessment, audits, and supervision visits, must be carefully designed so that those stakeholders ultimately controlling the quality of care (usually the providers) take full ownership of the quality improvement process. The team in charge of making improvements should fully participate in defining standards, identifying indicators, and developing a measurement strategy.

In this section, we focus on measuring the quality of neonatal and maternal health services, specifically essential obstetrical care, at a facility or operational level. Hence, the illustrative indicators are based on existing Neonatal and Essential and Emergency Obstetrical Care (EOC) standards relating to newborns and safe motherhood. However, even when well-defined national standards exist, defining new standards (and indicators) specific to the needs of the facility and community served may be necessary. Thus, the indicators used to measure quality will vary in each setting, based on the particular standards used and the level of the system (facility, district, regional, or national) on which measurement focuses.

The illustrative standards and indicators presented in this section are drawn from the Latin American Maternal Mortality (LAMM) Initiative. This initiative, implemented in collaboration with the MOH in each participating country and PAHO, aims at reducing maternal mortality by increasing the use, quality, and availability of maternal health services in eleven countries. In three of these countries – Bolivia, Ecuador, and Honduras – QAP and NGO subcontractors provided techni-

cal assistance in measuring quality, using the QA approach. The QA approach improved the ability of health care providers and facilities to detect and manage neonatal care and obstetrical complications at the facility and district levels. Detailed results of this work can be found in the LAMM/Briefer by Askov, Legros, and Camacho, 2001.

Organizational Levels of Care

As explained below, when developing indicators to monitor the quality of maternal care, one needs to consider various organizational levels of the health care system (operational/facility, intermediate, and strategic). The indicators will vary based on the level and on the specific standards selected for measuring quality (QAP/URC, 2001c). Whenever possible, the QA team should select indicators that use available data collected through the routine health information system. A brief description of organizational levels follows.

- **Operational level** - refers to the facility where processes that directly affect users and generate the monitoring data are carried out.
- **Intermediate level** - involves decision-makers who influence quality and the delivery of care, such as heads of departments, sections, or services.
- **Strategic level** - includes top management and political decision-makers at local, provincial, and national levels.

Table II.H.2.2 describes the differences between these levels.

A complete quality monitoring system for neonatal and essential obstetric care will include indicators for the processes, outputs, and outcomes of the different levels of services. They should reflect levels of compliance with the best clinical evidenced-based standards available. The frequency of measuring indicators varies; whereas some require monthly or quarterly measurements (compliance with process of care standards), others will require less frequent measurements (outcome measures such as disease-specific maternal mortality rates). When the number of cases are small, one can pool the data of several facilities in a region. Plotting those data over time allows evaluators to follow trends or patterns of performance and improvement (the product of which is a "run chart").

Although the illustrative indicators that follow have particular use at the facility or service level, one can also use some at national program levels.

Table II.H.2.2 Organizational Levels

Level	Monitoring Objective	Process	People Involved
Operational level	Track service- delivery processes, results, and the availability of inputs	An emphasis on individual processes or services at the facility level	Management and improvement teams that examine the quality of processes at the facility level
Intermediate level	Compare processes carried out at the operational level	An emphasis on the results and products of the combined processes or services	Heads of departments, programs, and services at the district level
Strategic level	Evaluate the overall system results, compare results with the objectives, act on differences, assure the quality of the entire system	An emphasis on the overall results of the system and strategy to better meet objectives	Top organizational management at the national, regional, or local government levels

Source: QAP/URC, 2001c.

Indicator

Percent of Newborns Receiving Immediate Care According to MOH guidelines

Definition
The percent of newborns receiving immediate care according to MOH guidelines

The clinical standard for the immediate care of newborns includes a physical examination consisting (at a minimum) of the following:

- Vital signs;
- Heart rate;
- Respiratory rate;
- Temperature;
- General appearance (e.g., color, malformations, activity level, weight, length, head circumference);
- Gestational age;
- Apgar score; and
- Other interventions (e.g., ophthalmic prophylaxis, vitamin K administration, and completion of the neonatal clinical history).

This indicator is calculated as:

$$\frac{\text{\# of newborns delivered at the facility who receive immediate care according to the MOH clinical guidelines}}{\text{\# of newborns delivered at the facility}} \times 100$$

The indicator is calculated for a specific reference period (e.g., per quarter or per year, depending of the size of the facility).

Data Requirements
Number of newborns delivered at the facility during the reference period who receive immediate care according to MOH guidelines; the total number of newborns delivered at the facility during the reference period

Data Source(s)
Review of medical records; direct observation of providers

Purpose and Issues
Immediate care of the newborn helps identify whether the baby has a normal condition; has a condition or problem requiring urgent, rapid treatment; or has a condition/malformation or other problem that requires transfer to another facility or service.

Providing immediate care to newborns helps reduce neonatal complications, sequelae, and deaths.

Physicians and nurses at all levels can learn to systematically perform immediate care for newborns.

For the newborn record to be a reliable data source, staff must fill the record out consistently and accurately. Ideally, the recording form will specify the standards, will facilitate accurate charting, and will stimulate appropriate actions.

The immediate care guidelines can also include specific protocols for identification and treatment of asphyxia, hypothermia, or hypoglycemia.

A low score on this indicator should trigger some intervention(s) to address skills and knowledge of providers, the organization of care, or both. The district health team or facility supervisor can provide support and encouragement to facility staff to ensure compliance with this important standard. Quality improvement teams can also help the facility determine causes for observed difficulties in compliance with the standard, as revealed by the indicator.

Indicator

Facility Perinatal Mortality Rate

Definition
Perinatal mortality refers to stillbirths (babies born dead after 22 weeks gestation) and early neonatal death (death within the first seven completed days of life). This indicator, the facility perinatal mortality, measures perinatal mortality among deliveries at a facility or network of facilities.

The indicator is calculated for a given period as:

$$\frac{\text{\# of perinatal deaths}}{\text{\# of births}} \times 100$$

Data Requirements
Number of perinatal deaths and number of births during a given period (e.g., 3 months, 12 months) at a given facility or a set of facilities

Data Source(s)
Review of medical records at the facility

Purpose and Issues
The perinatal mortality rate is a key health status indicator that reflects the overall quality of maternal and neonatal care. It depends on the socio-economic status of the community, access to health care, and the quality of that health care.

Several factors influence the level of perinatal mortality, including: effectiveness of treatment for medical conditions and complications during pregnancy, such as diabetes and high blood pressure; availability of special or neonatal care for very sick or premature infants; and availability of facilities for prenatal diagnosis of congenital malformations. The quality of obstetric and pediatric care available and the public health policy also affects the perinatal mortality rate.

Note: The **Facility Perinatal Mortality Rate** reflects quality of care at a given facility or network of clinics. For its application at the population level, see Part III.E.

Indicator

Percent of Deliveries in Which a Partograph is Correctly Used

Definition
The percent of deliveries correctly monitored with a partograph

A partograph is a simple chart that clinical staff can use to monitor labor and identify when it is not progressing satisfactorily.

Correct use is defined as: (1) starting the monitoring process only after the woman begins labor and (2) measuring the essential parameters, such as cervical dilation, descent of fetal head, and uterine contractions.

The indicator is calculated for a specific reference period as:

$$\frac{\text{\# of deliveries correctly monitored with a partograph}}{\text{Total \# of deliveries}} \times 100$$

Data Requirements
Number of deliveries monitored with a partograph; evidence of correct use of the partograph; number of deliveries at the facility during the reference period (e.g., 3 months, 12 months)

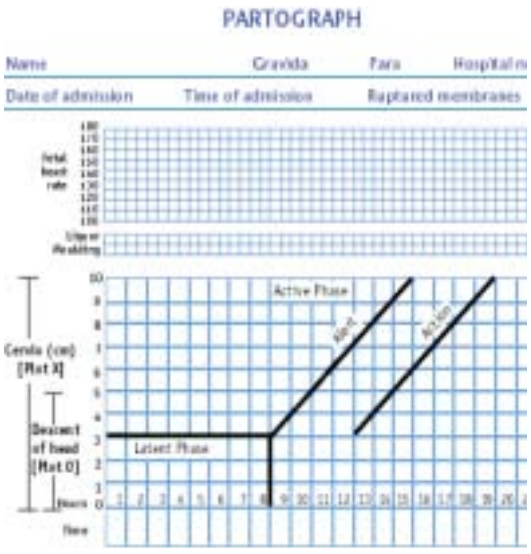
Data Source(s)
Review of medical records; direct observation by supervisor or external evaluator (regarding correct use); and review of the partograph

Purpose and Issues
The partograph displays the dynamic of labor during the first stage of delivery. It records fetal condition, labor progress, and maternal condition, and it provides a visual display of the progress of labor and immediately alerts the care provider to abnormal developments. In this way, the partograph acts as an "early warning system" that detects insufficient uterine action and/or cephalopelvic disproportion leading to obstructed labor (WHO, 1991a).

If properly used, the partograph helps reduce prolonged labor and its sequelae through earlier referral. Midwives, physicians, and nurses at all levels can learn to use and interpret partographs correctly, and thus can reduce cases of prolonged labor, maternal morbidity, and perinatal mortality (Schwarcz, Díaz, and Nieto, 1990).

A low score on this indicator may reveal a need for additional interventions, such as on-the-job training or refresher tutorials for staff.

Figure II.H.2.3 Partograph



Indicator

Percent of Deliveries in which a Perinatal Clinical Record Was Properly Completed

Definition
The attentiveness of staff to maintaining a complete record on the mother and baby from the prenatal period through 22 weeks post-partum

The perinatal record is one or more forms containing information for both the mother and the neonate at each contact – during pregnancy, delivery, and the neonatal period.

This indicator is calculated for a specific reference period as:

$$\frac{\text{\# of deliveries at the facility with a properly completed perinatal clinical record}}{\text{\# of deliveries at the facility}} \times 100$$

The reference period for this indicator is determined locally, but is generally from 3 to 12 months.

Data Requirements
Number of deliveries with a perinatal clinical record completed; number of deliveries at the facility during the reference period

Data Source(s)
Review of medical records; direct observation of providers

Purpose and Issues
The importance of a complete standardized perinatal clinical record is paramount to the quality of maternal care, because it reminds health providers of the standards of care. In Latin America, most Ministries of Health have adapted the model of a "simplified perinatal clinical record" developed by the Centro Latinoamericano de Perinatologia. The standard clinical record has sections for identification data, obstetrical history, pregnancy data, as well as delivery, newborn, and postpartum information. The clinical record has sections in yellow that represent some important factors that can increase perinatal risk.

The format and content of the perinatal clinical record may vary by country, or even within a country if the MOH has not introduced a standardized format. Whatever the format and content agreed upon at the facility, providers should use and know how to complete it accurately for each pregnancy and delivery. This indicator creates awareness among program administrators of the need for a standardized perinatal clinical record or for improvements to an existing one.

This indicator measures the attentiveness of staff to maintaining a complete record on the mother and baby from the prenatal period through 22 weeks post-partum. Whereas staff have a responsibility to maintain records on all women in their clinical facilities, they cannot be held totally responsible for women who do not return to the facility for postpartum care. This indicator primarily measures staff compliance with record-keeping, an important function that improves continuity of care.

Indicator

Percent of Mothers Examined Every 30 Minutes during the First Two Hours after Delivery

Definition
The number of newly-delivered mothers receiving standardized checks every 30 minutes after delivery for the first two hours

"Examined" refers to a series of standardized checks: vital signs, bleeding, and uterine status.

The indicator is calculated for a given reference period as:

$$\frac{\text{\# of mothers examined every 30 minutes during the two hours after delivery}}{\text{\# of deliveries}} \times 100$$

Data Requirements
Number of newly-delivered mothers receiving standardized checks every 30 minutes after delivery for the first two hours; number of women delivering at the facility during the reference period. The checks must be timely (every 30 minutes) to be considered valid.

Data Source(s)
Review of medical records; direct observation from supervisor or external evaluator.

Purpose and Issues
An important proportion of maternal deaths occur after delivery. The most important single cause of these maternal deaths is hemorrhage, most commonly in the immediate postpartum period (WHO, 1999a). Hence, routine checking for vital signs (especially blood pressure) and for vaginal bleeding and uterine status during the first two hours after delivery is an important standard of quality care that will help in the early detection of a potential life-threatening complication.

If a specific post-partum record is available and designed according to the standard of care, then it reminds providers to comply with the standard. If a specific form is unavailable, focusing attention on this standard may trigger the development of a specific job-aid.

Part II.H.3 Integration of Services

- Percent of clients who receive more than one reproductive health service during a given visit

Part II.H.3

Integration of Services

Part III of this *Compendium* presents indicators for multiple areas of reproductive health. The presentation of these different sections may suggest that reproductive health programs consist of a series of vertical interventions. However, as a result of the Cairo Conference and health reform initiatives, we have witnessed increasing integration across different areas of reproductive health in programs at the field level in developing countries. The concept of integration directly relates to "constellation of services" in the Bruce/Jain QC framework (Bruce, 1990).

Despite the wave of enthusiasm for integration, surprisingly little has been done to evaluate the extent to which integration has actually occurred. Rather, program managers and evaluators tend to assess program results based on data for each separate area.

Certain areas within reproductive health naturally link to others. For example, the promotion of breastfeeding goes hand-in-hand with the promotion of the lactational amenorrhea method (LAM). Adolescent programs often promote condoms as the method of choice for preventing both pregnancy and sexually transmitted infections. Antenatal visits provide an opportunity for counseling women on nutrition and micronutrient supplementation. Family planning clinics, STI treatment facilities, and antenatal care offer opportunities for screening women for possible domestic violence.

Not all services "integrate" as easily as one may expect. In the early days of the AIDS epidemic, many assumed that family planning services could readily expand to include condom promotion to prevent STIs. After all, both services involved populations of sexually active adults, and the two shared a common "solution:" the condom. Although efforts toward integration of family planning and STI services continue, the field now recognizes the challenges of combining these services. (Family planning programs tend to serve married women, whereas HIV prevention activities focus on sex workers, truck drivers, and other high-risk groups. The condom is the only viable method of HIV prevention for sexually active adults, yet it is generally less acceptable among adults married or in union than are other methods for contraceptive purposes.)

One approach to evaluating integration consists of observing client-provider interactions (CPI) to determine the extent to which service providers discuss more than one RH service during a single visit. In a related vein, one can determine the extent to which clients in one service (e.g., family planning) receive counseling or services for another (e.g., STI/HIV prevention). Alternatively, one can assess the "readiness to provide integrated services" through facility audits of equipment and supplies or training of personnel in multiple areas.

The International Planned Parenthood Federation/Western Hemisphere Region developed a self-assessment module on "Integrating STI/HIV/AIDS Services into Sexual and Reproductive Health Programs" (IPPF/WHR, 2000a). Although the module does not describe indicators of integration as such, evaluators can use its list of "minimum standards of integration" (module 2, page 3) to develop indicators. This resource contains a questionnaire to collect data on this topic.

In most sections of this *Compendium*, we present indicators tested in actual field programs and proven to yield useful information. Relatively few programs systematically track and report on integration. However, given the importance of integration to the evolving field of reproductive health, we have opted to present one basic indicator for integration and to list a series of illustrative indicators in Box II.H.3.1 The indicators in Box II.H.3.1 involve the integration of family planning and STIs (though one could look at this same set of indicators among new antenatal care clients as well).

Indicator

Percent of Clients Who Receive More than One Reproductive Health Service during a Given Visit

Definition
The clients receiving multiple services during a single visit

"RH service" refers to the areas outlined in Part III of this *Compendium*: family planning, STI/HIV/AIDS, safe motherhood, newborn care, adolescent RH services, postabortion care, breastfeeding, nutrition, FGC, and violence against women.

"A given visit" generally equates with attendance at the clinic on a specific day.

This indicator is calculated as:

$$\frac{\text{\# of clients who receive more than one RH service in a single visit}}{\text{Total \# clients}} \times 100$$

Data Requirements
Responses on a checklist (observation); data on services received (client records); and/or client exit interviews

Data Source(s)
Observation of provider-client interaction; client records (if reliable); client exit interviews

Purpose and Issues
As a crude proxy for integration of services in a RH service delivery facility, one may track the percentage of clients for a given type of service (e.g., family planning) who receive a second type of service during the same visit (e.g., STI counseling and/or treatment). The advantage to this approach is that it communicates to staff the importance of providing integrated services.

Several caveats warrant mention. Not all clients want or need multiple services on a given day. The evaluators should avoid inadvertently punishing a service delivery facility for providing the clients with the single service they desire. Second, evaluators may easily obtain the data for this indicator through an observation checklist or exit interviews, but such data collection requires a special study. The routine data collection at public health facilities may not allow for reliable tracking of this indicator, or it may create an additional recording burden for clinic staff.

The evaluation community will likely develop and refine indicators of integration, given the value placed on integration in reproductive health programs and the potential benefits to clients.

Gender Implications of this Indicator

Vertical programs that compartmentalize such services as family planning, STI diagnosis and treatment, antenatal care, nutrition services, among others, fail to recognize the holistic and interrelated health needs of women. The integration of health services has been a key goal of both the international health reform movement and the action agenda of the International Conference on Population and Development to improve gender equity and reproductive health. To the extent that health services can meet the multiple needs for women within a single site and/or a single visit, the system is responding to a woman's need for improved access to health care and the range of health needs of herself and her children. Integration of services also, in principle, recognizes the value of a woman's time by allowing "one-stop shopping" for a variety of health needs. Whether services are organized in such a manner to actually save her time should also be examined, as well as the overall quality and range of health services provided.

Box II.H.3.1 Illustrative Indicators of Integration:
Family Planning and STI Services

- Percent of new FP clients[1] with whom the provider discusses STIs and/or HIV/AIDS (any aspect of the topic or specific subtopics such as prevention, transmission, symptoms, health-seeking behavior);
- Percent of new FP clients with whom the provider discusses the dual benefits of the condom (to prevent pregnancy and infection);
- Percent of new FP clients orally screened for STIs (broken down to include behavioral risk assessment questions, symptom questions, and questions on the client's and client's partner's history of STIs);
- Percent of new FP clients receiving a physical exam (for STI detection or other purposes);
- Percent of new FP clients diagnosed with an STI syndrome;
- Percent of new FP clients diagnosed with an STI syndrome managed correctly according to a recommended algorithm or referred to an STI clinic, if appropriate;
- Percent of new FP clients receiving treatment on-site (versus those referred elsewhere); and
- Percent of FP clients diagnosed with an STI who are asked to bring in partner for STI diagnosis and counseling.

[1] Some may argue that these integration indicators should apply not only to new FP clients but also to all FP clients. We opted to limit the indicators to **new** FP clients, given that new clients generally participate in a counseling session or in other types of information exchange, whereas this counseling is often absent in resupply visits. In countries where both new and returning clients participate in information/counseling sessions, evaluators may drop the limitation and may include all FP clients.

Part II.H.4
Gender Equity/ Sensitivity

- Gender equity in the organizational context
- Gender sensitivity in the service delivery environment

Part II.H.4

Gender Equity/Sensitivity

Reproductive health programs operate within the cultural context of a given society, including its manifestations of gender inequity. While FP/RH programs do not redress these imbalances at a macro-level, they can promote gender equity in areas within their manageable control. Indeed, reducing gender-related obstacles to improved reproductive health can work synergistically with other development activities to enhance gender equity (Yinger et al., 2001).

This section of the *Compendium* focusing on gender issues in the organizational context, does not try to capture the deep-seated gender inequities existing in most countries worldwide. (See Part II.A.) Nor does the section address the consequences of gender discrimination in the form of injurious social outcomes, such as violence against women (see Part III.K), sex-selective abortion, or female infanticide. Rather, the two sets of gender-related indicators in this section address the following questions:

(1) Is the organization free of gender bias in its managerial structure? Does the organization actively foster gender equity in its routine operations?

(2) Is the service delivery environment free of gender bias toward clients? Does the clinic promote gender equity in the way it offers services?

The two sets of indicators (for gender equity in the organizational context and gender sensitivity in the service delivery environment) draw on two primary sources: the report by the Interagency Gender Working Group (Yinger et al, 2001) and the *Manual to Evaluate Quality of Care from a Gender Perspective* (IPPF/WHR, 2000b). In the Yinger et al. (2001) report, we adapted several items from the list of "gender-related obstacles to achieving RH objectives," identified in the appendix of the report. In the IPPF/WHR manual, we selected and/or adapted a number of the indicators from Appendix 8. In short, the list below represents an effort to combine good ideas from two credible sources into a practical menu of indicators that assess gender-equity in an organizational context. As such, this "instrument" has not been tested in this form, although individual items have been used at the field level. We included these two sets of gender-related measures to encourage the further testing and development of indicators in this area.

Organizations may use the gender-related indicators in three ways. First, they can track these indicators as an ongoing part of monitoring their services. Second, they can set up an external evaluation of the organization based on these indicators, to be conducted by a person familiar with reproductive health programs as well as gender issues. Third, they can use the indicators as a self-assessment tool for a special study to systematically examine their own record on gender equity and sensitivity in the workplace.

Indicator

Gender Equity in the Organizational Context

Definition
"Gender equity" is the equally fair treatment of women and men. To ensure fairness, some societies adopt measures to compensate for historical and social disadvantages that prevent women and men from otherwise operating on a "level playing field." Gender-equity strategies eventually attain gender equality. Equity is the means; equality is the result (Interagency Gender Working Group, 2000).

Data Requirements
Scores from items selected from the menu of indicators in Box II.H.4.1

Data Source
External assessment by an individual familiar with organizational behavior; gender issues; and reproductive health programs. Alternatively, a self-assessment by senior management, based on international standards, adapted for the local setting.

Purpose and Issues
This set of indicators is presented as a menu from which evaluators may select those most applicable to a given work setting. (Evaluators may expand this set to include other items of interest to the organization in question.) IPPF/WHR (2000b) recommends that the evaluation team include (among others) a locally hired gender specialist. The question of gender equity is sensitive, and the process of evaluating gender equity can become highly politicized. For this reason, the organization must select an evaluator perceived to be objective and to have excellent credentials. The evaluation must take place in a climate of impartiality if the results are to carry weight.

Box II.H.4.1 Menu of Indicators:
Gender Equity in the Organizational Context

- Percent of managerial positions held by women;
- Average salary of men versus women in comparable managerial positions;
- Representation of women's health advocates on Board of Directors;
- Participation of women in the conceptualization and design of projects;
- Explicit organizational policy statement that prohibits gender discrimination in hiring, promotion, and retention policies, salaries, and benefits;
- Similarity of supervision procedures for male and female staff (of equal rank);
- Percent of personnel (including supervisors of service programs, receptionists) who receive training in gender sensitivity;
- Elimination of overt gender bias in organization's standards and guidelines;
- Existence of written policies or guidelines to prohibit sexual harassment of staff;
- Organizational commitment (demonstrated by explicit interventions) to:
    - Women's participation (in project activities);
    - Human rights (lobbying for specific causes);
    - Empowerment (e.g., attempts to change community norms regarding women's mobility);
    - Equity (e.g., micro credit systems);
- Disaggregation of program data by sex (where appropriate);
- Equal distribution of opportunities for training and career development between men and women; and
- Equal protection for men and women in organizational policies regarding clients' rights to privacy, informed consent, confidentiality, and delivery of high-quality services.

Indicator

Gender Sensitivity in the Service Delivery Environment

Definition
"Gender sensitivity" is the way service providers treat male or female clients in service delivery facilities and thus affects client willingness to seek services, continue to use services, and carry out the health behaviors advocated by the services. This indicator also measures aspects of the services themselves (e.g., in the case of family planning, whether a range of male as well as female methods is offered).

Data Requirements:
Scores from items selected from the menu of indicators in Box II.H.4.2

Data Source(s)
External assessment by an individual familiar with organizational behavior, gender issues, and reproductive health programs. Alternatively, a self-assessment by senior management, based on international standards adapted to the local setting.

Purpose and Issues
This set of indicators is presented as a menu from which evaluators may select those most applicable to a given service delivery environment. Evaluators may expand this list to include other items of interest in the local context.

For a service delivery facility to demonstrate gender-sensitivity, it must adhere to the principles of informed choice, voluntarism and a target-free approach, which might otherwise not be the case given the low status of women in the locality. A gender-sensitive approach has much in common with a quality of care approach. A program cannot be gender-sensitive if both male and female clients fail to receive complete information and to participate fully in decisions regarding their care and treatment. Many women want opportunities to involve their partners in counseling and in decisions concerning contraceptive use and reproductive and child health. Similarly, many men wish to participate in RH counseling as well as in decisions regarding reproductive and child health, but have felt excluded from this arena.

Box II.H.4.2 Menu of Indicators
Gender-Sensitive Service Delivery Context

- Availability of services to adolescents, single women, widows, homosexuals;
- Absence of requirements that clients have permission of husband or mother-in-law (for married women) or parents (for adolescents);
- Availability of condoms both to women and men;
- Percent of providers in the health facility who are female;
- Availability of a full range of services whatever the sex of the provider (e.g., male doctors provide IUDs for female clients);
- Percent of physicians who are women;
- Availability of female physicians for women who prefer them;
- Non-stigmatizing attitudes toward clients (e.g., unmarried female clients with STIs, homosexuals, sex workers, postabortion care clients, adolescents);
- Number of referrals to other programs that empower women (e.g., related to literacy, income generation, micro-credit, domestic violence);
- Percent of personnel (including supervisors of service programs) who receive training in gender sensitivity;
- Use of gender-sensitive protocols for counseling (e.g., non-discriminating language, two-way communication, equal attention to women in counseling sessions for couples);
- Percent of facilities that, with the permission of the female client, encourage men to visit/attend (to accompany partner, obtain information, or obtain services);
- Equal treatment (e.g., waiting time, courtesy, privacy, information given) for male and female clients;
- Avoidance of gender stereotyping in BCC materials;
- Percent of facilities that are "male-friendly:"
    - Hours convenient to men;
    - Staff receptive to men in clinic; and
    - Materials (posters, pamphlets) directed to men visible and available;
- Percent of service providers trained to detect, discuss, and refer clients to services that handle violence against women (in FP);
- Providers describe female and male sterilization as equally desirable, when appropriate (FP only); and
- Services focused on health outcomes for both the child AND mother (safe motherhood services).