1999

Reproductive health indicators: Moving forward

Kelly Blanchard
Population Council

Batya Elul
Population Council

Saumya RamaRao
Population Council

Recommended Citation

Reproductive Health Indicators:

Moving Forward

Reproductive Health Indicators:

Moving Forward

Kelly Blanchard

Batya Elul

Saumya RamaRao

Authors are listed in alphabetical order

The Population Council is an international, nonprofit, nongovernmental institution that seeks to improve the wellbeing and reproductive health of current and future generations around the world and to help achieve a humane, equitable, and sustainable balance between people and resources. The Council conducts biomedical, social science, and public health research and helps build research capacities in developing countries.

The Council's Robert H. Ebert Program on Critical Issues in Reproductive Health, established in 1988, responds to an awareness that many important reproductive health problems—and the ways women experience them—have been neglected by policy-makers, program planners, and practitioners. Currently the Program focuses on several areas that merit special attention: improving the quality of services in reproductive health programs, managing unwanted pregnancy and preventing the consequences of unsafe abortion, devising new approaches to postpartum care to meet the health needs of the mother and child, and designing programs that address sexually transmitted diseases, including AIDS, within the larger context of women's reproductive health.

Not everything that can be counted counts, and
not everything that counts can be counted.

Albert Einstein

Acknowledgments

Preface

In recent years, a large number of reproductive health indicators have been put forward by various organizations. This paper is designed to help readers, be they researchers, program managers, policymakers, or advocates, sift through and evaluate which potential indicators might be useful in a particular programmatic context. It was initially prepared as a background manuscript for a Population Council meeting on reproductive health indicators held in Cairo in March 1998. The manuscript has since been modified to incorporate ideas and comments expressed at that meeting, as well as to include input from other Population Council colleagues.

Section 1. Introduction

The 1994 International Conference on Population and Development (ICPD) in Cairo led to an explosion of interest in reproductive health. Perhaps the most important outcome of this meeting was a shift in focus among the population and international health establishment from population control targets to women's health issues and women's reproductive rights. This shift has created a demand for tools with which to measure reproductive health. Indeed, over the last ten years, indicators for this purpose have been borrowed from other fields, created anew, and refined in a number of important ways.

Valid and reliable indicators are particularly important as many international organizations continue to invest in reproductive health programs throughout the world. Indicators are needed for identifying problems that require attention, as well as for highlighting specific populations or subpopulations that have the greatest need for interventions and services. Once programs and services have been implemented, indicators can provide valuable information on what works, what does not, and whether programs and services are having an impact on the problems they are designed to address.

A variety of institutions and organizations have attempted to develop reproductive health indicators to meet these needs. The World Health Organization (WHO), for example, has developed a variety of indicators to monitor progress toward their efforts to ensure health for all by the year 2000. WHO has also collaborated with the United Nations Children's Fund (UNICEF) on developing indicators, including a few reproductive health indicators, to assess progress toward the specific targets set at the 1990 World Summit for Children. Additionally, the United States Agency for International Development (USAID) has been engaged, through the EVALUATION Project, in the development of indicators, largely for monitoring reproductive health programs.

These efforts, however, have been plagued by several shortcomings. First, they have focused almost exclusively on monitoring programs at the national level to the exclusion of other uses, such as evaluating women's health status. Second, no standard definitions for many common reproductive health problems have been established across communities, regions, and organizations, an omission leading to confusion and questions about the interpretation of available data. Third, the data necessary for calculating many of the reproductive health indicators developed by these organizations are not often readily available or, where available, are frequently of dubious quality. Finally, many indicators have been adapted from other areas of public health without due consideration of their applicability to reproductive health.

Perhaps the most significant advance in the development of reproductive health indicators has been the implementation of a number of in-depth, community-based studies of reproductive health. These studies have focused attention on women's health issues and were instrumental in the shift of interest to this subject at the ICPD conference. These studies took a new approach to measuring women's reproductive health, using both qualitative and quantitative data-collection strategies and amassing information not only on morbidity, but also on women's experience of ill health.

Community-based studies, as they have been conducted, though, are not without methodological problems. For example, to date, they have been conducted in only a few discrete settings and have not been replicated in a wider range of countries and populations, using a standardized protocol. Additionally, the innovative indicators used in these studies need to be validated and their technical robustness must be demonstrated.

This paper sets out to lay the groundwork for a discussion of the means by which to develop and refine reproductive health indicators. It does not provide an exhaustive review of reproductive health indicators or of the technical aspects of their development and calculation, but rather provides the foundation for a discussion of future research needs in this area.

Presented below is a conceptual framework for guiding research on reproductive health indicators followed by a review of the potential purposes for which such indicators can be used. The next section includes a brief overview of how indicators are derived and discusses technical criteria that can be used to assess them. The fifth section describes the data sources typically used to calculate reproductive health indicators, their advantages and their disadvantages. Section six illustrates the concepts that have been outlined by reviewing the particular qualities of a few sample indicators. The final section of the paper offers a brief catalogue of other agencies that have made significant contributions to the development of reproductive health indicators and the work that they have undertaken.

Section 2. A Conceptual Framework

The development of useful indicators is best guided by a theoretical framework that postulates the relationship between different variables of interest. Such a framework requires a detailed understanding of the variety of factors that might affect an outcome of interest, typically a disease state or health condition. The input-output-outcome-impact framework shown in Figure 1 is designed for identifying indicators for programmatic purposes. This

Figure 1: Levels of evaluation			
Process Evaluation		**Impact Evaluation**	
INPUT	**OUTPUT**	**OUTCOME**	**IMPACT**
(Program design and plan)	(Implementation)	Change in:	Change in:
• People	• Information	• Awareness	• Morbidity
• Goods	• Skills	• Behavior	• Mortality
• Services	• Services	• Health practices	• Disability
• Money			• "Health"

framework suggests a relationship between the typical activities of health improvement programs, *inputs* and *processes*, and relates them to particular *outcomes* and *impacts*, which

are usually defined as changes in behavior and health status, respectively. Although outcomes are usually the point of greatest interest—changes in health status are typically the ultimate goal of any program—they are often difficult to measure, and the link to program activities per se is difficult to establish. Changes in health outcomes can be related to a variety of factors, many of which are unrelated to the program in question. Thus, input and process indicators are often used as surrogates to capture the activities of a program that presumably have the desired influence on the condition of interest, or on overall health status. Typically, although inputs are supply-side oriented, outcomes and impacts are measured at the level of the individual who uses health services or in the community at large.

Fortney (1995) has described a potential framework for understanding the components of reproductive morbidity. This framework separates reproductive morbidity (a subset of women's morbidity) into obstetric (maternal), gynecologic, and contraceptive categories. Such distinctions are important in order to assure that rates of "reproductive" morbidity are comparable between studies. These categories can help researchers to clarify what specific type of morbidity they are interested in, and how it might be related to other, more general, or more specific, types of morbidity. This framework can also help focus measurement efforts on those areas where no or little reliable information exists.

This framework does not provide information on determinants of reproductive health outcomes of interest, however. Such a framework would be helpful for identifying different points along causal pathways that could be targeted for indicator development. These types of frameworks have been developed in other areas, for example child and infant mortality and morbidity (Mosley-Chen framework). Different frameworks could be developed for different components of reproductive health, including maternal mortality/morbidity; reproductive tract and sexually transmitted infections (RTIs/STIs) including HIV; and violence against women. The development of such frameworks requires further research on the determinants and covariates of specific reproductive morbidities and conditions. Some of this information can be gleaned from the in-depth community-based studies that have been conducted to date, but more studies are needed in different areas and populations of women. A better understanding of the determinants of reproductive morbidity would help to focus the discussion of reproductive health indicators.

Conceptual frameworks are important for developing indicators because they take into consideration many determinants of broad health outcomes. Understanding the determinants or components of health outcomes and their effect on the outcome of interest is critical for choosing valid indicators. Although we may often rely on input or process indicators because measuring changes in outcome is difficult or impossible, a link between the determinant or component we are addressing and the outcome of interest must be established, and other determinants or components that may affect our ability to effect change in that outcome must be recognized.

Section 3. Purposes of Reproductive Health Indicators

The purpose for which an indicator will be used is critical in selecting one that is appropriate, and one that will lead to accurate conclusions. Because each measure of an aspect of

reproductive health requires a different level of precision or refinement, a single indicator can rarely serve a variety of purposes. Several possible uses of reproductive health indicators are listed below by the level of precision required for each, in order of increasing precision.

Advocacy. Indicators are often used by advocates, including women's health advocates, advocates for better health of underserved populations, and health professionals who play the role of advocate to governments, donors, or individuals. In general, less refined indicators may be useful for advocacy purposes. Statistics such as the number of maternal deaths per year resulting from unsafe abortion or the average number of women who die in childbirth per day are often sufficient to raise awareness about or engage interest in a problem. For advocacy purposes, magnitude tends to take precedence over precision. Advocates usually do not have the resources to undertake research in order to generate the indicators they require, and such research is not the focus of their work. Typically, therefore they use readily available indicators, such as those that are calculated annually and published in accessible reports.

Needs Assessment/Feasibility Studies. Reproductive health indicators may also be used to assess needs or to demonstrate the feasibility of a particular program. Although many reproductive health indicators used for this purpose may also play a role in advocacy by highlighting a specific problem that has not received sufficient attention, these indicators are intended to guide a project or intervention design in a specific community. As such, they must be more precise than those used for advocacy alone. For example, the indicator must reflect the current reproductive health status of the target population. Yet, because indicators used to assess needs are usually focused on establishing baseline figures, rather than on evaluating the effectiveness of an intervention, they may be less precise than those used for monitoring or evaluating the impact of a program.

Group, Regional, and International Comparisons. Indicators used to compare reproductive health information across groups, regions, or countries require greater precision than do those employed to assess program needs. When reproductive health measures are used to compare information across groups, regions, or countries rather than to determine the health status of a single population, comparability of the populations must be considered in terms of such factors as age structure and access to health services. In addition, the definition of the condition being measured by the indicator must be the same for the populations being compared. In order to make a valid comparison, for example, to state that more women in country A than in country B suffer from pre-eclampsia, the national rates of pre-eclampsia must not differ because pre-eclampsia is defined more liberally or frequently in country A than it is in country B.

> **Box 1: Get out your dictionaries.** Ensuring standardized case definitions has proved to be a particular problem. Definitions of reproductive morbidities vary by country, by region, and also by provider. For example, when Bulut et al. (1995) reanalyzed the prevalence and distribution of pelvic inflammatory disease (PID) that were reported in their study employing the criteria used by researchers in three other studies, the estimated prevalence ranged from 3.7 percent to 30.4 percent.

Monitoring Existing Programs. An extensive literature is available on indicators that can be used to monitor programs (Bertrand et al*.,* 1996; Rashad et al. 1994; Rossi and Freeman 1993; Smith and Morrow 1991.) One of the main issues in choosing an indicator is ensuring that it reflects the activities of the program itself, rather than extraneous trends in the health service or catchment population. The input-process-outcome-impact framework shown in Figure 1 can help identify appropriate indicators for monitoring programs. Although policymakers or program managers may be more interested in the outcome of a program or intervention (see below), input and process indicators are most often used for routine monitoring because they are directly related to program activities. "Input" refers to the resources provided by a program (for example, drugs supplied to a clinic for treatment of STIs or contraceptive methods on hand), whereas "process" denotes the measure of programmatic activities undertaken (for example, number of clients counseled about HIV prevention or number of condoms distributed). Because program monitoring often occurs at routine intervals in a program cycle, both input and process indicators should be available disaggregated by specified time periods. For this reason, data collection for these indicators should be incorporated into the program or intervention design insofar as is possible.

> **Box 2: The multipurpose indicator.** A single indicator best serves a single (or limited number of) function(s). Indicators developed for a specific purpose may not have characteristics critical for other purposes. For example, the number of maternal deaths per year by country may be used by advocates to highlight the differential risks of maternal death around the world and to stimulate funding for improvements in maternal health services, but maternal deaths occur too rarely to measure program impact. Similarly, the inconsistency of definitions of maternal death, and the differences in age structures of populations, may make international or interregional comparisons meaningless.

Assessing the Impact of Interventions. Assessing the impact of interventions or programs has proved more difficult than has monitoring program inputs or activities. Identifying the impact of a program requires that a link be established between program activities and any change in disease status or other desired outcomes. Often, documenting such link in the population under study requires not only lengthy and costly follow-up but also necessitates an understanding of the disease process and a clear idea of where in that process the program might cause a change. For example, a program undertaken to reduce the incidence of HIV might include counseling of clients with STDs about safer sex practices and provision of condoms. Proving, however, that counseling and the provision of condoms results in behavioral change and a reduced risk of acquiring the disease is much more difficult than determining that these activities have taken place. In order to establish such a relationship, the co-factors affecting the population of interest that might also have an impact upon the spread of HIV must be accounted for. Many of these co-factors are difficult to quantify; others may not even be known. Furthermore, identifying aggregate indicators of impact is difficult for programs that provide integrated services or more than one service, or those that do not have disease-specific goals. The goal of the latter sort of program

may be overall improvement in reproductive health for which no clear or operational definition exists, or improvements in a variety of outcomes that may or may not show uniform improvement or decline.

Resource Allocation and Strategic Planning.[a] Indicators of the cost-effectiveness of programs or interventions are often used to plan the combination of services to be offered at various service delivery points, and to develop national or regional strategies of health-care provision. Collecting data with which to calculate these indicators is extremely difficult and time consuming. Indicators such as disability-adjusted life years (DALYs) saved per unit cost or other measures of cost-effectiveness require that all the costs of an intervention be calculated, including the various components of capital costs. Additionally, a number of assumptions must be made regarding depreciation, inflation, and the amount of various capital inputs, including staff salaries, that should be attributed to the intervention or program. Such calculations often involve extensive and expensive costing studies. Moreover, measures of cost-effectiveness may be incomparable across regions because of the variability in costs of imported materials or local labor.

Section 4. Criteria for Assessing Indicators[b]

Typically, more than one indicator can be used to summarize the same phenomenon or condition. For instance, the quality of antenatal services can be assessed, to some degree, by an indicator consisting of "the number of visits made by pregnant women," as well as by the "proportion of pregnant women having their first visit in the first trimester." In order to select the most useful indicator, appropriate selection criteria must be agreed upon. A single indicator is unlikely to fulfill all of the agreed-upon criteria, however. Consequently, the relative importance of various selection criteria must be weighed depending on how the indicator will be used. Some of the most commonly used selection criteria (World Health Organization 1997) are reviewed below. Although the list is not comprehensive, it can be used to spark discussion of other potential criteria.

Scientific Soundness. Indicators should possess proved technical robustness. The scientific soundness of an indicator is comprised of four main components: validity, reliability, specificity, and sensitivity.

---

[a] Although a large literature exists on measuring the cost-effectiveness of programs and services (Barnum 1987; Bobadilla et al. 1994; Murray 1993; Murray et al. 1994; World Health Organization 1988), indicators for this purpose are not discussed here.

[b] As background to this section, we have provided a description of how indicators are derived in Appendix I. While much of the information presented in this section is not specific to indicators or to reproductive health, we have approached it with an eye focused on the development of reproductive health indicators. As much as possible, we offer reproductive health examples.

Validity refers to the extent to which an indicator represents what it is supposed to, or, in other words, to how well it reflects reality. Validity can be evaluated on several grounds: First, the extent to which a measurement corresponds to theoretical concepts concerning the phenomenon under study can be determined. Whether the measure being used to describe an issue or factor is in fact describing that issue or factor can be assessed. For example, a woman's report concerning her use of oral contraceptives is likely to have greater validity than her husband's report. Finally, whether the current measure of the phenomenon produces results that are closely related to other independent measures of the same phenomena can be evaluated. For example, results can be "triangulated" from different data-collection methods. In a groundbreaking community-based study in Giza, Egypt, the authors measured the validity of self-reports of reproductive tract infections and genital prolapse by comparing and measuring the association between self-assessments and medical diagnoses (Zurayk et al. 1995). If the two measures were highly associated, then self-assessment could serve as an inexpensive measure of health that stood the test of validity.

Box 3: Validating women's self-reports of major obstetric complications. In recent years, several Safe Motherhood studies have assessed the usefulness of asking women about signs and symptoms of obstetric complications using survey methods. (Six studies have been conducted to date in Benin, Bolivia, El Salvador, Ghana, Indonesia and the Philippines; unfortunately only two are published: Stewart et al. 1995; Ronsmans et al. 1997.) These studies were undertaken to determine if self-reported symptoms, compared with clinical and/or laboratory results, can predict correctly the occurrence of major obstetric complications. In 1996, a task force established to review the data from these studies concluded that interview data are unlikely to be valid or reliable for identifying the prevalence of obstetric complications (Stewart et al. unpublished). The task force reported that women's retrospective self-reports of complications are not an accurate means of estimating the proportion of women who need medical treatment for obstetric complications. For instance, one of these studies, in which 340 inpatients who had presented at a rural district hospital in Ghana with life-threatening obstetric complications were interviewed, found that 76 percent of the complicated cases and 75 percent of the uncomplicated cases were correctly identified by a self-reported symptoms algorithm, representing substantial false-positive and false-negative rates (Sloan et al. unpublished).

Reliability reflects the extent to which an indicator produces the same results if it is assessed more than once under the same conditions (that is, with the same measurement tool and the same population). Count data, such as the number of maternal deaths in a population, are generally reliable if records are available. On the other hand, when attitudes, beliefs, or perceptions are being measured, the same questions may generate different responses depending on the respondent's mood, state of mind, or interest in participating in the survey, as well as on the relationship between the respondent and the interviewer and where the interview is conducted.

Sensitivity and *specificity* are typically used in the field of epidemiology to measure the validity of screening tests. Sensitivity is the proportion of persons in the screened population who are accurately identified as diseased by the screening test for the disease. Specificity is the proportion of truly non-diseased persons who are accurately identified by the screening test. The definitions of these measures, however, are often interpreted more broadly in the context of reproductive health indicators. For instance, sensitivity is used to express whether an indicator reflects changes in the factor of interest. Similarly, specificity refers to the ability of an indicator to reflect *only* changes in the issue or factor under consideration without being influenced by changes in other factors.

Data Requirements. In order to minimize the burden of data collection on service providers and other health personnel, indicators should be based on data that are routinely collected and available data, if possible. Readily available data sources are often plagued by problems of representativeness and reliability, however (see below). For this reason, data from population-based surveys are often needed to supplement routinely collected data. Using these surveys, however, increases costs and, consequently, may make indicators based on these surveys less accessible to those working in the field.

Representative. Ideally, indicators should be representative of all the issues or populations that they are expected to cover. Generally, data routinely collected by health institutions have limited coverage and, consequently, reflect only a select group. For example, the indicator "prevalence of severe anemia in pregnant women" is not representative because not all pregnant women are screened for anemia. In many cases, however, supplementing routine data can increase the representativeness of a given indicator. For instance, the indicator "number of facilities with functioning comprehensive essential obstetric care per 500,000 population," which appears on the short list of national and global indicators published by WHO (see section 5), would be more representative if, in practice, data were collected from both private and public facilities.

Ethical. Data collection, analysis, and presentation should protect the rights of the individual to confidentiality, freedom of choice in supplying data, and informed consent regarding the nature of the data required.

> **Box 4: Who's asking the question?** Research in various settings suggests that respondents' reports of symptoms vary by who is interviewing them. Indeed, several studies show that women report fewer symptoms to lay interviewers than to clinicians. For example, in a community-based study conducted in Turkey, just over one-third of respondents reported experiencing reproductive morbidity when interviewed at home, whereas more than 80 percent reported the presence of a reproductive malady to the clinician when they were examined at a clinic (Bulut et al., 1995). The authors concluded that the participants' memories were jogged during the home interview, leading them to report more morbidity to the clinician. Other reasons for this finding may include women's expectations of treatment and the presumed confidentiality of the clinic setting. Similar results have been reported in general health surveys as well (Ross et al. 1986).

Understandable. Indicators should also be simple to define and interpret. Consequently, WHO discourages the use of composite indicators.

> **Box 5: Syndromic management of STIs.** In recent years, efforts have increased substantially to identify the best management practices concerning STIs in resource-poor settings. On the programmatic level, many health-care workers have begun to rely on algorithms of symptoms, signs, and risk factors to predict infection without the need for laboratory diagnosis. A detailed review of validation studies in this area, however, found that management strategies using symptoms, signs, and risk factors have poor predictive value; they miss a substantial proportion of women in need of treatment and misclassify many women as requiring treatment when they do not (Haberland et al. 1999). For instance, comparisons of self-reports and medical assessments (a combination of clinical examinations and laboratory tests) of RTIs in Bangladesh found that only 68 percent of symptomatic women also showed evidence of vaginal, cervical, or pelvic infection (Wasserheit et al. 1989). Similarly, a study conducted in Giza, Egypt, that compared women's self-reports of vaginal discharge with medical assessments (clinical and/or laboratory) of RTIs notes a relatively high sensitivity of 79 percent but a low specificity of 26 percent (Zurayk et al. 1995).

Section 5. Data Sources

We devote this section to discussion of various data sources for reproductive health indicators and what can be gleaned from each of them to measure reproductive health.

5.1 Traditional Data Sources

Traditionally, data for calculating indicators have come from four main sources: vital statistics, censuses, service-delivery statistics, and sample surveys. The benefits and drawbacks of using data from each of these sources and the information relevant to reproductive health indicators that can be obtained from them are described below.

Vital Statistics. Vital statistics are data collected routinely on live births, deaths, marriages, and, in some cases, abortions. In theory, vital statistics provide a large amount of information on a continuous basis with complete coverage of a population, although the events recorded may be limited in scope and strictly defined. In most countries today, and particularly in developing countries, however, vital statistics are rarely routinely collected, or are collected only from certain areas and socioeconomic classes. The infrastructure required to aggregate data, as well as the central office for the collection and dissemination of data are often lacking. Moreover, in countries where large proportions of births and deaths occur outside of health-care facilities, these events often go unreported, so that available statistics are representative of select portions of populations.

Even in ideal situations, vital statistics provide only a small amount of the information needed to derive reproductive health indicators because events other than births, deaths, and marriages are usually not recorded. These data, however, can be used to calculate the numerators for the total fertility rate, the maternal mortality "rate," and mortality rates from other causes. Vital statistics can also be used to calculate the infant mortality rate, since, in principle, both births and infant deaths are recorded. Population denominators, however, cannot be calculated from vital statistics (unless the data are complete for a long period of time).

The Census. Reliable data on population size gathered by national censuses have proved valuable in calculations of reproductive health indicators. Specifically, censuses can provide important information for the denominators of many key reproductive health indicators, like the maternal mortality rate and the infant mortality rate. Censuses usually have good coverage and allow a great deal of information to be gathered at one time. Moreover, the information collected in a census can be modified by adding new questions. A variety of drawbacks are inherent in using census data for calculating reproductive health indicators, however. Censuses are expensive to conduct and require significant technical expertise in survey methods. Additionally, although, in theory, censuses provide complete information about a nation's population, in reality, census data often do not include poor and marginalized populations. Because censuses occur infrequently, establishing figures for intercensal periods involves extrapolation, which may lead to inaccurate conclusions. Finally, censuses rely on self-reported information, making much of the data vulnerable to recall bias (see box 6). Many of the conditions pertinent to reproductive health indicators may be impossible or difficult to recall over long periods of time.

Service Statistics. Service statistics are data collected routinely at service-delivery points (SDPs) concerning the conditions or diseases observed among clients. Because service statistics usually include results of diagnostic tests (unlike most other data sources described in this section), they are not subject to the biases of self-reports. Service statistics, however, suffer

> **Box 6: When to ask the question?** The period of recall is an important factor determining responses, regardless of the type of data-collection instrument. A study of postpartum morbidity in Bangladesh offers important insights on this issue (Goodburn and Graham, 1996). In this longitudinal study, women were interviewed about their pregnancy and delivery experience twice during the antenatal period—at pregnancy registration and at seven months' gestation—and four times during the postpartum period–within five days of birth, at two weeks, six weeks and 12 weeks. Concerning a number of events, including whether the respondent had had premature or postterm labor, had had labor lasting greater than 24 hours, had been attended by a traditional birth attendant, had had her placenta extracted, and had experienced moderate or severe postpartum hemorrhage, the proportion of women reporting each event decreased at 12 weeks postpartum compared to that reporting at five days postpartum. In the context of reproductive health, most studies, aside from those assessing pregnancy and delivery-related complications, have tended to use current-status measurements.

from several other inherent drawbacks and biases. First, they are available only for populations who present at health services and thus are not representative of the population at large. Indeed, in the developing world, accumulated evidence from a variety of settings suggests that individuals suffering from various ailments or conditions rarely seek care at SDPs. Rural women, in particular, may be constrained from seeking care for several reasons, including lack of awareness of the ailment (for example, asymptomatic RTIs/STIs), inaccessibility of treatment sources (as a result of cost, or physical or social distance), and lack of sufficient autonomy to seek care. Additionally, conditions not culturally defined as requiring medical assistance will not be included in service statistics unless such conditions are discovered in a patient presenting for another reason. Service statistics also suffer from some of the same problems as do vital statistics in terms of infrastructure and technical requirements. Finally, health-care workers may not be motivated to report service-delivery statistics or may be overburdened with other tasks, supervision of reporting may be inadequate, and service-delivery statistics may be oriented generally toward administrative matters rather than toward monitoring the health of populations.

> **Box 7: How do you ask the question?** The exact format (open-ended or pre-coded) of a survey question and the extent of probing for each reproductive health condition are important determinants of a response. Indeed, results from several studies suggest that, when compared with open-ended questions, symptoms checklists elicit higher levels of positive response. Checklists are believed to structure the interview, making it easier for the interviewer to conduct. They also jog the memory of the respondent, and provide standard definitions of illnesses (Kroeger 1983). In a community-based study conducted in Turkey, for example, 7 percent of women interviewed reported a reproductive health morbidity in response to open-ended questions compared with 65 percent who reported such a condition when it appeared on the symptoms checklist (Bulut et al. 1995). Several other studies have reported similar increases in reporting when a symptoms checklist or additional probing follows open-ended questions. In a study of Rajasthani women, for example, the proportion of women reporting menstrual problems increased from 28 percent to 48 percent, and those reporting prolapse increased from 10 percent to 20 percent upon detailed probing (Grant and Measham 1996).

In theory, however, service-delivery statistics can provide information about a range of conditions that are of interest as reproductive health indicators, including STIs, infertility, RTIs, reproductive cancers, and obstetric conditions, as well about contraceptive-use rates and number of births attended by a trained clinician. Service statistics, however, do not usually provide an accurate population denominator, because most service-delivery points do not have a well-defined catchment population. If reliable census or survey data exist for an area or region, data from these sources can be combined with service statistics to calculate an indicator.

Sample Surveys. Sample surveys, also known as population-based studies, have served as data sources for a variety of indicators that are related to reproductive health. In general, these surveys are well designed and provide accurate and representative data. Perhaps the

most best known sample surveys in the reproductive health field are the Demographic and Health Surveys (DHS). DHS data have been used to calculate indicators such as total fertility, wanted fertility, and infant and child mortality rates. In recent years, the DHS have also tested a variety of reproductive health modules, including some addressing domestic violence. Additionally, research groups and government offices in developing countries have collaborated with the DHS to interview subsets of the initial DHS sample on specific reproductive health issues, such as the sub-survey conducted in the Philippines in 1993 (National Statistical Office and Macro International 1994). This survey assessed the proportion of women who experienced a variety of reproductive health problems and their use of services for these problems.

Like censuses, however, sample surveys are usually dependent upon self-reports of information and are, therefore, prone to recall bias. A number of techniques have been developed to deal with this problem. For instance, many questions used in DHS surveys have been pre-tested extensively and validated. Additionally, sample surveys provide information for only one fixed time point, which restricts their usefulness in detecting trends over time. Repeat surveys can limit this disadvantage, but the cost and technical expertise required to conduct such surveys may be prohibitive. Sample surveys generally rely on multiple-choice questions, with pre-coded responses based on the researcher's a priori knowledge. Although such an instrument design encourages the use of standardized terms and concepts, it also severely limits the scope for dialogue between the interviewer and the respondent (Assogba et al. 1991).

5.2 Alternative data sources

Given the limitations of traditional data sources for the development of reproductive health indicators, as discussed above, we now turn to alternative data sources.

Sentinel Surveillance System. Sentinel surveillance is perhaps best known in the reproductive health field for the collection of data related to HIV/AIDS (Mann and Tarantola 1996). Yet, sentinel surveillance systems have been used for many years to track trends of other conditions, including tuberculosis and acute respiratory infections, and may be relevant to the collection of data for other aspects of reproductive health.

A surveillance system is used to monitor trends of a particular condition over time through the repeated collection of samples from the same population. Sentinel sites may be established to gather information from populations at different degrees of risk or in different geographic areas that are assumed to represent a random sample of the population under study. Consequently, testing at these sites serves as a "sentinel" by providing a warning of what is presumed to be happening in the population at large.

Sentinel surveillance has the advantage of generating data quickly and cost-effectively. Additionally, the location of sentinel sites can be changed to reflect shifts in the spread of disease. For instance, as HIV transmission shifted from particular groups, such as commercial

sex workers and intravenous drug users, to the general heterosexual population, sentinel sites have shifted from STD clinics to ante-natal clinics.

Sentinel surveillance, however, suffers from several methodological shortcomings. First, because sentinel surveillance does not employ random sampling methods the data collected at sentinel sites may not be specific enough for evaluating program effects or for estimating incidence or prevalence of a condition in the population as a whole. In these cases, incidence or prevalence data estimated from sentinel populations would have to be compared with that estimated

Box 8: Quality of care. In recent years, there has been a growing consensus that indicators of quality of care are needed to provide policymakers, program managers, and service providers with guidelines for improving services. To date, work in this area has focused mainly on contraceptive choice, staff competence, and information provided to clients. The quality of client-provider interactions, however, has proved particularly difficult to measure. Indeed, providers who are unresponsive to women's requests ironically may fare well in some quality assessments. For example, a woman who uses oral contraceptive pills and requests a refill but receives a thirty-minute review of all available contraceptive methods from an unresponsive provider would figure in the numerator of the indicator "percentage of clients given information on all available methods," resulting in a misleading assessment of the quality of client-provider interactions.

from cross-sectional population-based studies in order to assess which sentinel populations most closely represent the condition in the general population. Additionally, sentinel surveillance may suffer from "participation" bias depending on the condition in question. As is the case with service statistics, sentinel surveillance conducted at service-delivery points is representative only of the people who seek care at health-care centers. With regard to HIV/AIDS, if testing is voluntary rather than unlinked and anonymous, results may not be valid nor representative because of the stigma attached to those infections. Operationally, sentinel surveillance is impractical for dealing with the full panoply of reproductive health conditions. A sentinel surveillance system can rarely be used to monitor trends of more than one unrelated condition at a time because each problem requires observation of a different sentinel population. For instance, although a sentinel site established at an antenatal clinic may provide data on HIV/AIDS, as well as on several other STDs, such as syphilis, it would be a poor source of data for the prevalence of uterine prolapse in the community at large.

Service-delivery-point-based Studies (SDP studies). Typically, the information available to researchers concerning various aspects of reproductive health has been gathered from hospital- or clinic-based studies. For the most part, information from clinics has been available on STDs, maternal mortality, and abortion whereas information on infertility and gynecological problems has been collected rarely. Much of the current knowledge about the prevalence and distribution of STIs comes from studies of attendees of STD, antenatal or family planning clinics (Brunham and Embree 1992). Similar situations exist for information about abortion and maternal mortality. Clinic-based abortion research has provided information on the profile of abortion seekers (that is, about their age, parity, marital status, contraceptive use), on service-delivery issues such as complications observed and

13

types of procedures used, and on infection-prevention techniques employed, although these data may be thoroughly contaminated where abortion is illegal.

SDP-based research is attractive for several reasons. Foremost is the comparative ease of conducting clinical and laboratory investigations in settings where trained personnel and equipment are available. This can result in substantial cost reductions for conducting research in medical settings. Additionally, as noted above, clinic attendees, especially those showing no symptoms of the condition under study, form an important sentinel group for determining prevalence at the population level. Finally, research focusing on service-delivery improvements is best conducted at SDPs.

Although SDP-based studies can provide valuable information, as has been noted, they are not representative of the population at large because respondents are self-selected for the study. As noted in our discussion of service statistics, for a variety of reasons, many women do not seek care at medical institutions. Therefore, SDP-based studies do not provide insight into the prevalence of ailments and conditions in the general population.

Community-based Studies. Community-based studies are an important innovative data-collection methodology that redress to some degree the inadequacies of SDP-based and sample-survey data-collection methods. Community-based studies have focused attention upon reproductive health issues. For example, such studies conducted in Giza, Egypt (Zurayk et al. 1995) and Gadchiroli, India (Bang et al. 1989) provide information on the extent of gynecological symptoms and morbidity among women in the developing world. Increasing evidence that gynecological problems are not confined to clinic-based populations, but rather appear to be pervasive in the general population, has spurred innovative research in reproductive health. These studies have provided a bank of substantive information, but have also raised a series of methodological issues. The term "community-based study" is used here to denote the in-depth study of particular villages or sites. The sample in community-based studies may be smaller and less representative than are those found in sample surveys or population-based studies. Typically, community-based

---

Box 9: Laboratory testing. Laboratory results are meaningful only for conditions which are amenable for such testing, such as reproductive and urinary tract infections and anemia. Laboratory tests, however, are often difficult to conduct under field conditions. Storage and transport of specimens to referral laboratories are problematic. Furthermore, some laboratory tests are expensive. For many conditions, a variety of laboratory tests – some simple and some more complicated – can be used to detect a particular condition. For example, testing for anemia can be performed by using simple procedures such as color coding of blood or by more detailed serological analysis. Similarly, RTIs can be detected by a range of tests including simple microscopy of wet mounts to cultures. Additionally, the prevalence of infections is an important predictor of detection. In settings with low prevalence, it is easier to miss positive cases than in high prevalence settings. Many "gold standard" tests are beyond field implementation and may not be suitable for community studies.

studies use a triangulation of measurement tools, collecting information from women's self-reports of symptoms, from clinical examinations conducted by trained medical personnel, and from laboratory tests.

Many community-based studies complement quantitative survey methods with qualitative methods in order to provide a rich tapestry of information about reproductive morbidity. For instance, in Giza, Zurayk and her colleagues (1995) employed both a validation study and a series of auxiliary qualitative studies in an effort to better understand the context of women's reproductive health. One such auxiliary study entailed case studies of women diagnosed as having gynecological problems in the initial study who were referred outside of the community for further medical care (Khattab 1992). These case studies highlighted the role of the woman's child-rearing responsibilities, the gender and generational hierarchies within households, and the culturally proscribed norms of tolerance and endurance in the women's recognition of illness and in their consequent patterns of health-seeking behavior.

Community-based studies also highlight how women evaluate their symptoms in terms of the physical and social disturbance they cause and how this evaluation is translated into treatment-seeking behavior. The Giza study showed that physical discomfort such as pain, itching, foul odor, or a feeling of wetness prompted women with a vaginal discharge to seek care. Results from that study also showed that the severity of the condition and the wider-ranging social and personal consequences of the ailment are significant factors in a woman's decision to seek care. Such findings have also been reported in community-based studies conducted in India. For example, one Indian study found that the severity of a discharge or menstrual problems prompted women to seek care (Kanani et al. 1994). A second study conducted in India found that white discharge had implications for overall physical, mental, and sexual health. In particular, white discharge was seen as causing a host of ailments including visual, gastric, and urinary problems, and general aches and pain (Bang and Bang 1994). Moreover, this

> **Box 10: What are we talking about? Disease, illness or sickness?** The use of several research methodologies simultaneously or sequentially ensures that community-based studies are more than a simple measurement of disease prevalence. Indeed, by highlighting the social context in which reproductive morbidity is experienced such studies can provide a holistic view of reproductive health. In this context, Zurayk and Kabakian (1996) stress the need to distinguish between disease, illness and sickness as the latter two provide a more complete description of an individual's experience of disease. Indeed, the concepts of illness and sickness allow for an investigation of cultural explanations and perceptions of health and disease causation. Furthermore, they provide a social meaning to health and ill health. For example, it is frequently cited that often individuals do not recognize their ailment as a medical condition which requires attention or care. An ethnomedical study in Rajasthan, India revealed that women suffering from a white vaginal discharge attributed it to physiological weakness resulting from poverty. Thus, these women did not seek health care as they did not perceive the root cause of their discharge to be medical (Grant and Measham 1996).

same study highlighted the hierarchy of treatment in the informal health sector: Women suffering from a white discharge initially treated themselves with home remedies provided by an elderly, experienced woman. If no improvement followed this treatment, a traditional healer was consulted. If the discharge was perceived to be the result of witchcraft, however, a witchcraft healer was sought. Finally, an allopathic healer was visited only if all else failed and the resources to seek such care were available.

Community-based studies, as conducted in recent years, however, have been shown to have several methodological flaws. Studies aiming to measure levels of reproductive morbidity can be biased by the self-selection of respondents. For instance, women who agree to participate in the study may do so in the hope of receiving advice or treatment for their ailments, whereas those who perceive themselves to be healthy may refuse to participate. The self-selection problem tends to be greater in studies that include a clinical examination than it is in those that use self-reported data on symptoms. Such a situation inflates the variable under study by increasing the number of women whose positive responses are reflected in the numerator and by decreasing the number of women whose responses could potentially form part of the denominator.

Several community-based studies have also reported substantial rates of sample loss for the clinical examination component of the study. For example, seven community-based studies conducted in India (Koenig et al. 1996) all noted significant sample loss as great as 80 percent. Similar results are reported for neighboring Bangladesh where 24 percent of symptomatic women refused clinical examination after being interviewed (Wasserheit et al. 1989). By ways of contrast, a community-based study in Turkey was successful in examining 80 percent of the respondents who had participated in the home interviews and, similarly, in the Giza study, 91 percent of the women interviewed were later examined (Bulut et al. 1995; Zurayk et al. 1995). Although culture-specific reasons may explain the varying levels of participation in the different phases of the studies, differential efforts of health-care providers to build rapport with the community may also account for these differences. These results highlight the need to establish rapport

Box 11: How do you define the condition?
Qualitative health research suggests that in many cultures and languages, a multitude of terms exists to denote the same reproductive health morbidity. Moreover, culture-specific constructs identify the term as either a positive or a negative manifestation. For example, Bang and Bang (1994) reported that in one Indian community, women used 12 different expressions and five major categories to describe white discharge. Respondents may also use terminology that indicates conditions other than the literal meaning. In India, "weakness" is often used to refer to a discharge (Koenig et al. 1996). Clearly, if respondents are unable to understand the question, they will not be able to respond appropriately, and measurements will be affected. Investigators, too, need to understand the terminology and the meanings for the various conditions so that they can identify and classify them properly. A fuller description of the condition usually facilitates a simpler and more accurate classification of the variable. For example, a vaginal discharge can be described in terms of several characteristics such as volume, odor, consistency, and color. Similarly, menstrual problems can be characterized by volume of bleeding, and by days of menstrual flow.

with the community and with key household decisionmakers, including husbands and mothers-in-law, before beginning study-sample recruitment.

Underreporting of morbidities is another important methodological problem with studies that rely on self-reports of symptoms, such as those that are community-based. Two longitudinal studies conducted in India found that more women reported a gynecological morbidity in the latter rounds of a survey than in the initial ones (Koenig et al. 1996). Another study of more than 270 women conducted in northern India reported a greater than seven-fold increase (from 5 percent to 38 percent) in the number of respondents reporting a gynecological problem over the course of a year (Koenig et al. 1996). A study of postpartum women in Bangladesh (Goodburn and Graham 1996) showed that reports of reproductive morbidity increased during the four-week study period. Indeed, validation by physical examination— for those conditions that could be confirmed by a physical examination, such as lower abdominal pain—showed that although an increase occurred in the reporting of the symptom, no corresponding increase was evident from clinical diagnosis.

To date, comparing results from community-based studies has been difficult because standard definitions for many common reproductive health problems have not been established. Definitions of STIs, for instance, have varied widely, leading to confusion and to questions about the interpretations of available data. Indeed, variations in definitions have led to differing estimates of the prevalence of the same STIs in one community (see Box 1). Therefore, although several studies suggest that STIs and RTIs are major causes of morbidity, the scope of this problem remains uncertain.

Although community-based studies have dealt effectively with several shortcomings of traditional data sources, several important measurement issues remain unresolved. Further work is required to develop and validate standards for measuring the reproductive health of women on a community level.

Section 6. Examples of Indicators

In the previous sections, we reviewed the purposes for which indicators may be used, criteria for assessing their usefulness, and possible data sources. In this section, we draw on the concepts summarized earlier and discuss the effectiveness and limitations of a few sample indicators.

The *maternal mortality ratio* measures the annual number of maternal deaths per 100,000 live births. Maternal deaths are rare events even in settings with high maternal mortality, and they require large sample sizes from which to derive reliable estimates. Field-workers for the Indian DHS, for example, interviewed nearly 90,000 women (the largest DHS conducted to date) to derive a maternal mortality ratio of 437 deaths per 100,000 live births for the two-year period preceding the survey (IIPS, 1995). Despite the large sample size, the sampling errors for the maternal mortality estimates are large. In many settings, classification of the cause of the death is extremely difficult, especially where a large proportion of deaths occur outside of institutional settings. Deaths from complications of abortion may be misreported

17

or underreported where such procedures are illegal. Similarly, many deaths of pregnant women resulting from domestic violence are not reported as maternal deaths. For these reasons, the maternal mortality ratio is not an appropriate indicator for assessing effectiveness of safe motherhood programs at the district level. It can be used, however, as an indicator to measure progress at the national level.

The *percentage of deliveries attended by trained personnel* and *the percentage of institutional deliveries* are indicators of the coverage of delivery services. These percentages can also be used to measure access to and use of delivery services and are often used to monitor program performance at both national and subnational levels. The numerators of these indicators are the numbers of deliveries attended by trained personnel and the numbers of deliveries occurring in institutions, respectively. In both instances, the denominator is the total number of deliveries occurring in the reference period. Robust indicators of the use of delivery services can be derived from sample surveys, because these can provide information for both the numerator and the denominator. However, if service-delivery statistics or other institutional records are used to generate data, detailed information on the numerator may be available, but the denominator is usually based on estimates of the projected number of deliveries occurring in the area in the reference period. The representativeness of this indicator depends on the proportion of births that occur at health-care facilities, and the accuracy of the reports of the total number of births in the reference period.

The indicator *reported rate of non-accidental injury to women aged 15-44* has been developed to assess the level of violence against women in an effort to inform policy and to plan effective programs. The data requirements for this indicator are complex. In order to obtain an accurate picture of the level of violence, data are required on all reported cases of violence from several sources, including police stations, health institutions, crisis centers, and women's shelters. Because a considerable proportion of violent episodes are never reported to any institution, the indicator is likely to underestimate the true incidence of violence against women, even if data are collected from a number of sources. Although sample surveys may include questions on violence, levels of non-response or misreporting tend to be high. Similarly, although health-care institutions may offer an estimate of the proportion of clients who are victims of violence, women seeking care at those institutions may over-represent those who are severely injured because those not seeking care or seeking care outside the formal sector will be ignored.

The indicator *percentage of men with urethritis* attempts to assess the extent of potential STIs among men in the general population. Data for this indicator traditionally come from sample surveys: men are either asked about symptoms or provide a semen sample. Self-reporting of symptoms suffers, however, from the problems discussed earlier, and the collection of specimens is time consuming, expensive, and often infeasible. For these reasons, collection of accurate data for the numerator is difficult, resulting in a poor indicator and potentially erroneous conclusions about STI prevalence in the general population. If the purpose of the indicator is not to assess the true prevalence of urethritis, however, but rather to indicate minimum levels of infection for advocacy and planning purposes, the indicator may be sufficient.

Section 7. Other Work in the Field of Reproductive Health Indicators

General activities in the field of reproductive health indicators over the past few years are reviewed below. Most efforts have been aimed at developing national-level, aggregate indicators and little, if any, attention has been allotted to community-level indicators.

WHO   WHO has been one of the forerunners in the development of reproductive health indicators. Following the World Summit for Children in 1990, the organization has collaborated with UNICEF (see below) to generate a list of indicators for assessing progress toward the summit's goals. In recent years, the family and reproductive health unit has also worked specifically to develop reproductive health indicators and has delineated selection criteria for these indicators, as described in section 3. After hosting the Interagency Technical Meeting on reproductive health indicators in early 1997, WHO proposed a list of 13 reproductive health indicators to serve as a useful minimum.

UNICEF   UNICEF has been working for several years to develop indicators to monitor progress toward the specific targets set at the 1990 World Summit for Children. The agency initially produced a list of impact and process indicators for assessing progress toward the Mid-decade Goals and have worked since with WHO to expand the earlier list to evaluate global goals for the year 2000. Only a few of these indicators are related to reproductive health. In 1995, UNICEF also issued the second edition of "Maternal Mortality: Guidelines for Monitoring Progress" (Maine et al. 1995). In this publication, the authors review ten indicators related to safe pregnancy.

UNFPA   At the initiation of advisors from the Country Support Teams, the Technical and Evaluation Division of the United Nations Population Fund (UNFPA) has been working since 1994 to develop a list of indicators for assessing the components of reproductive health identified at the ICPD. They have created a list of 70 indicators, which include both quantitative and qualitative types.

USAID   USAID has funded the EVALUATION Project to coordinate multi-disciplinary subcommittees of the Reproductive Health Indicators Working Group, to review existing indicators and to develop new indicators of reproductive health. The project has focused almost exclusively on indicators for evaluating the output or outcome of programs rather than on assessing needs or evaluating health status. Subcommittees were convened on safe pregnancy, breastfeeding, sexually transmitted diseases and AIDS, adolescence, and maternal nutrition. Each subcommittee was asked to provide a short list of 59 indicators. Interestingly, the subcommittee on family planning did not produce a short list, but rather offered 95 indicators, stating that the choice of key indicators for evaluating family planning programs worldwide depends entirely on the purpose of the evaluation.

Section 8. Conclusion

In this paper, the importance of using appropriate indicators for the purpose required is emphasized. Indicators required for program monitoring and evaluation are necessarily

different from those required for outcome or impact measurement. A similar distinction is necessary for indicators used at different levels—national, subnational, or those used for a specific program.

Although a considerable bank of knowledge concerning reproductive health issues exists that is based on service-delivery-point data, the recent advances in community-based research have added significantly to the measurement of reproductive health status. Community-based studies have highlighted new problems and important areas for work—both in terms of necessary interventions and improved services and in light of the need to develop and validate good reproductive health indicators. This method of measurement should be expanded in order to gather information about different populations of women in a variety of regions and cultures. For this endeavor to be successful, however, a standard protocol for conducting studies must be developed. Second, the indicators employed must be validated for technical soundness to ensure that they will provide comparable data. Once the results of these studies are applied more broadly, and appropriate interventions are identified and implemented, the routine suffering of women resulting from unmeasured reproductive morbidity can be addressed and perhaps alleviated.

The knowledge derived from community studies can form the basis for the design of larger-scale studies that are replicable and sustainable. The possibility of including questions on reproductive health in large-scale surveys has to be explored more rigorously. Questions on reproductive behavior and attitudes toward family planning were once considered to be too personal or sensitive, and thus were likely to elicit responses that were invalid or unreliable. Determining more effective ways of asking questions and identifying which questions to ask has led to the development of better questions that yield both valid and reliable answers that can be included routinely in large-scale surveys. The limitations of such surveys cannot be ignored, however. Only when a variety of data-collection methods are developed, all of which provide valid and reliable information, will the true picture emerge showing the reproductive health status of the world's women that is necessary for program design, evaluation, and monitoring.

Appendix I: How Indicators are Derived

With the present focus in the reproductive health community on health status, the presence or absence of a specific illness or condition among a population is a matter of general interest. In many instances, we rely upon epidemiological measures of disease frequency to summarize the distribution of disease among different groups.

The most basic measure of disease frequency is a simple *count* of cases of the illness or condition of interest. Such figures are important for strategic planning in the health system, and for allocating scarce health-care resources. Counts of cases provide an idea of the number of people who will require a specific health service or treatment. Counts of cases, however, reflect the size and composition of the population from which those cases arose. In order to compare the disease burden of different populations, or of subgroups in a single population, the base population from which cases arise and the time period of data collection must be considered as well.

The most commonly used and simplest type of indicator that includes a measurement of the base population is a *proportion*. A proportion is a ratio in which the numerator is a part of the denominator, and it is usually expressed as a percent. Some examples include the fetal death "rate" (that is, the number of fetal deaths/the number of births where the number of births is defined as live births plus fetal deaths); the percent of women giving birth in a health-care facility (that is, the number of women giving birth in a health-care facility/the number of women giving birth); and the case fatality "rate" (that is, the number of deaths due to a condition/the total number of cases with that condition).

A *ratio* involves a comparison of a numerator and a denominator in which the numerator is not a part of the denominator. For example, in the maternal mortality ratio (that is, the number of maternal deaths/the number of live births), the numerator does not figure in the denominator.

The word "rate" must be used and interpreted with caution. In the public health literature, the term refers loosely to measures that are actually ratios or proportions, rather than true rates. A true *rate* includes a measure of time in the denominator. The denominators of true rates are measures of person-time, for example, person-years or person-months. Denominators of true rates are population measures rather than events, such as births or deaths.

Prevalence is a measure of the number of persons having a specific condition or illness at a designated time. It is defined as the number of existing cases/total population from which those cases arose. All prevalence measures are also proportions, and can be interpreted as the risk of any individual in the population of having the condition of interest at one particular time. In terms of reproductive health indicators, the prevalence of a variety of STIs or conditions such as uterine prolapse might be of interest. Prevalence incorporates a measure of the size of the population from which the cases arose. When comparing the prevalence (or in fact any measure of disease frequency) of certain conditions between groups, however, the definition of a "case" must be consistent for all groups, and the populations must be

21

comparable. Differences in the age structure or other characteristics between populations may lead to incorrect conclusions about the relative importance of a condition in each population.

Incidence is a measurement of the frequency at which new cases of a specific condition or illness arise in a given population over a specified period of time. *Cumulative incidence* (that is, the number of new cases of disease arising during a given period of time/total population at risk) can be interpreted as an estimate of the probability or risk that an individual will develop the condition of interest during a particular period. The period involved can be a number of years or can refer to periods in an individual's life (for example, before age ten). This measure implies that all of the individuals under study were followed for the entire time period delineated; in other words, the cohort from which the information is derived cannot experience any loss to follow-up. In terms of reproductive health indicators, a woman's lifetime risk of developing cervical cancer, or the number of women who present with RTI symptoms over a certain period of time might be of interest.

Box 12: Incidence vs. Prevalence.
Incidence is inherently difficult to measure for illnesses because such a measure requires that a cohort of individuals from the population of interest be followed over a period of time. This procedure is both costly and time consuming. Consequently, prevalence measures are often used instead because they can be determined from the data gathered in cross-sectional surveys. Caution must be used when prevalence measures are interpreted, however, because prevalence is a reflection both of the incidence of disease in the population and of the average duration of the disease or condition. The duration of disease is influenced in turn by other factors such as the availability and effectiveness of treatment and the screening measures employed.

The *incidence rate* is defined as the number of new cases in a period of time/ total person-time of observation. When calculating incidence rates, individuals can be followed for various lengths of time and can contribute differing amounts of time to the denominator. Such calculations assume that all the individuals all have equivalent susceptibility and, consequently, that time contributed from different individuals is interchangeable (for example, following eight individuals for two years would yield the same data as would following 16 individuals for one year). Looking at the incidence rate of a particular condition provides an idea of the frequency with which individuals experience that condition. For example, the incidence of HIV seropositivity indicates the rapidity of the spread of infection in a particular population. Comparing the incidence between different populations may help make clear which factors affect the spread of infection.

In addition to the measures mentioned above, other indicators are derived using demographic methods and, essentially, are averages of a specific condition of interest. Examples include life expectancy at birth, at other ages, or at a certain time (these data can be disaggregated by sex), and the total fertility "rate," which is an estimate of the average number of children a woman would have should she experience over her lifetime the fertility rates that exist currently for women in every age group.

References

Assogba, L., O. Campbell, and A. Hill. 1991. "Advantages and limitations of large-scale health interview surveys for the study of health and its determinants." In *The Health Transition: Methods and Measure.* Eds. John Cleland and A. Hill. Canberra: The Australian National University Press.

Bang, R., A. Bang, M. Baitule, Y. Choudry, S. Sarmukaddam, and O. Tale. 1989. "High prevalence of gynecological diseases in rural Indian women." *The Lancet* 1(8,629): 85–87.

Bang, R. and A. Bang. 1994. "Women's perceptions of white vaginal discharge: Ethnographic data from rural Maharashtra." In *Listening to Women Talk about their Health: Issues and Evidence from India.* Eds. J. Gittelsohn et al. New Delhi: Har-Anand Publications.

Barnum, H. 1987. "Evaluating health days of life gained from health projects." *Social Science and Medicine* 24(10): 833–841.

Bertrand, J., R. Magnani, and N. Rutenberg. 1996. *Evaluating Family Planning Programs.* Chapel Hill: The EVALUATION Project.

Bobadilla, J.L., P. Cowley, P. Musgrove, and H. Saxenian. 1994. "Design, content and financing of an essential national package of health services," *Bulletin of the World Health Organization* 72(4):653-662.

Brunham, R.C. and J.E. Embree. 1992. "Sexually transmitted diseases: Current and future dimensions of the problem in the third world." In *Reproductive Tract Infections: Global Impact and Priorities for Women's Reproductive Health.* Eds. A. Germain et al. New York and London: Plenum Press.

Bulut, A., N. Yolsal, V. Filippi, and W. Graham. 1995. "In search of truth: Comparing alternative sources of information on reproductive tract infection," *Reproductive Health Matters* 6: 31–39.

Fortney, Judith. 1995. "Reproductive Morbidity: A Conceptual Framework." *Family Health International Working Paper* No. WP95–02. Research Triangle Park, NC: Family Health International.

Goodburn, E. and W. Graham. 1996. "Methodological Lessons from a Study of Post-Partum Morbidity in Rural Bangladesh." Paper presented at IUSSP seminar on Innovative Approaches to the Assessment of Reproductive Health. Manila, 24–27 September.

Grant, J. and D. Measham (eds.). 1996. *Reproductive Tract Infection, Lessons Learned from the Field: Where do We Go from Here*? Report of a seminar, Population Council's Robert H. Ebert Program on Critical Issues in Reproductive Health. New York, NY. 6–7 February 1995.

Haberland, N., B. Winikoff, N. Sloan, C. Coggins, and C. Elias. 1999. *Case Finding and Case Management of Chlamydia and Gonorrhea Infections Among Women: What We Do and Do Not Know.* New York: Population Council, Robert H. Ebert Program on Critical Issues in Reproductive Health.

International Institute for Population Sciences (IIPS). 1995. *National Family Health Survey (MCH and Family Planning), India 1992-93.* Bombay: IIPS.

Kanani, S., K. Latha, and M. Shah. 1994. "Application of qualitative methodologies to investigate perceptions of women and health practitioners regarding women's health disorders in Baroda slums." In *Listening to Women Talk about their Health: Issues and Evidence from India.* Eds. J. Gittelsohn, et al. New Delhi: Har-Anand Publications.

Khattab, H.A.S. 1992. *The Silent Endurance: Social Conditions of Women's Reproductive Health in Rural Egypt.* Amman: UNICEF and Population Council.

Koenig, M., S. Jejeebhoy, S. Singh, and S. Sridhar. 1996. "Undertaking community-based research on the prevalence of gynecological morbidity: Lessons from India." Paper presented at the IUSSP seminar on Innovative Approaches to the Assessment of Reproductive Health. Manila, 24–27 September 1996.

Kroeger, A. 1983. "Health interview surveys in developing countries: A review of the methods and results." *International Journal of Epidemiology* 12(4): 465–481.

Maine, D., V.M. Ward, T.M. Wardlow, J. McCarthy, and A.S. Birnbaum. 1995. *Maternal Mortality: Guidelines for Monitoring Progress.* New York: UNICEF.

Mann, J. and D.J.M. Tarantola (eds.). The Global AIDS Policy Coalition. 1996. *AIDS in the World II: Global Dimensions, Social Roots and Responses.* New York: Oxford Press.

Murray, C.J.L. 1993. "Quantifying the Burden of Disease: The Technical Basis for Disability Adjusted Life Years." *Harvard Center for Population and Development Studies Working Paper* 93.03. Cambridge, MA: Harvard University Press.

Murray, C.J.L., J. Kreuser, and W. Whang. 1994. "Cost-effectiveness analysis and policy choices." *Bulletin of the World Health Organization* 72(4): 663–674.

National Statistics Office (NSO) [Philippines] and Macro International. (MI). 1994. *National Safe Motherhood Survey 1993.* Calverton, MD: NSO and MI.

Rashad, H., R. Gray, and T. Boerma (eds.). 1994. *Evaluating the Impact of Health Interventions.* Liège: IUSSP.

Ronsmans, C., A. Endang, S. Cohen, and A. Zazri. 1997. "Women's recall of obstetric complications in South Kalimantan, Indonesia." *Studies in Family Planning* 28(3): 203–213.

Rossi, P.H. and H. Freeman. 1993. *Evaluation: A Systematic Approach*. Newbury Park, CA: Sage Publications.

Sloan, N., P. Arthur, E. Amoaful, S. Adjei, and B. Winikoff. "Validity of self-reported symptoms to identify major obstetric complications." New York: Population Council. Unpublished.

Smith, P. and R.H. Morrow (eds.) 1991. *Methods for Field Trials of Interventions Against Tropical Diseases*: *A Toolbox*. New York: Oxford University Press.

Stewart, M.K. and M. Festin. 1995. "Validation study of women's reporting and recall of major obstetric complications treated at the Philippine General Hospital." *International Journal of Gynecology and Obstetrics* 48: S53–S66.

Stewart, M.K. et al. "Statement from a task force meeting on validation of women's reporting of obstetric complications in national surveys." Unpublished.

Wasserheit, J.N., J.R. Harris, J. Chakraborty, B.A. Kay, and K.J. Mason. 1989. "Reproductive tract infections in a family planning population in rural Bangladesh." *Studies in Family Planning* 20(2): 69–80.

World Health Organization. 1988. *Training Modules for Household Surveys in Health and Nutrition*. Geneva: World Health Organization.

———. 1997. *Interagency Technical Meeting on Reproductive Health Indicators for Global Monitoring: Report of the Meeting*. Final draft report of a technical meeting Geneva, 9–11 April. Geneva: World Health Organization.

Zurayk, H, H. Khattab, N. Younis, O. Kamal, and M. El-Helw. 1995. "Comparing women's reports with medical diagnoses of reproductive morbidity conditions in rural Egypt." *Studies in Family Planning* 26(10): 14–21.

Zurayk, H and T. Kabakian. 1996. "Measurement of reproductive morbidity: The usefulness of perceived/reported morbidity on reproductive tract infections." Paper presented at the IUSSP seminar on Innovative Approaches to the Assessment of Reproductive Health. Manila, 24–27 September.