Measuring reproductive health: review of community-based approaches to assessing morbidity

Ritu Sadana[1]

This article begins by reviewing selected past approaches to estimating the prevalence of a range of morbidities through the use of household or community-based interview surveys in developed and developing countries. Subsequently, it reviews epidemiological studies that have used a range of methods to estimate the prevalence of reproductive morbidities. A detailed review of recent community or hospital based health interview validation studies that compare self-reported, clinical and laboratory measures is presented. Studies from Bangladesh, Bolivia, China, Egypt, India, Indonesia, Nigeria, Philippines and Turkey provide empirical evidence that self-reported morbidity and observed morbidity measure different phenomena and therefore different aspects of reproductive health and illness. Rather than estimating the prevalence of morbidity, interview-based surveys may provide useful information about the disability or burden associated with reproductive health and illness.

Keywords: reproductive medicine; epidemiological studies; health surveys; disease notification, methods; review literature; comparative study; developed countries; developing countries.

Introduction

Awareness of the extent and consequences of reproductive ill health has increased over the past decade (*1–3*). Coalitions of international agencies, women and development movements, feminist movements and an array of nongovernmental organizations have forced attention on the global distribution of women's reproductive illness and its neglect (*4–6*). This rising awareness is in part due to the development of international perspectives on health and inequities and to the premise that improving women's reproductive health is an important intrinsic goal, not simply a means to other objectives.

Despite advancing on the policy agenda, a number of researchers argue that women's health and reproductive health lack clear definitions and rigorous assessment methods. Graham & Campbell conclude that the low priority on reproductive health and the lack of available information are self-reinforcing and constitute a "measurement trap" (*7*). This trap exists because of a narrow conceptualization of women's reproductive health, poor existing data sources, restricted indicators of health that focus solely on measures of disease — particularly mortality — and limited measurement techniques to facilitate community-based data collection.

The findings of the Global Burden of Disease study further confirm that reproductive morbidity and associated disability must be taken account of, beyond mortality. The pattern of disability-adjusted life years (DALYs) lost from reproductive ill health either due to premature mortality or morbidity associated with reproductive conditions is substantially different from that for deaths alone. This is because of the young age of many of those who die from conditions associated with reproductive ill-health and the large component of years lived with disability (YLDs) resulting from many of these conditions (*8, 9*).

Until recently, most efforts to estimate the prevalence of reproductive mortality and morbidity were primarily based on hospital data or poor quality vital statistics that were not representative of the population. Other sources include research focusing on issues other than women's reproductive health, such as prevalence of contraceptive use or child survival or isolated studies covering non-representative samples of women (*10*). Without population-based information on reproductive health and illness, efforts to quantify the extent of premature death and disability attributed to specific conditions or diseases must rely on limited data extrapolated to larger regions of the world (see e.g. *11*). One way forward is to improve research methods for estimating the prevalence of reproductive morbidity in the community, particularly for developing regions. Since the

[1] Epidemiology and Burden of Disease, Global Programme on Evidence for Health Policy, World Health Organization, 1211 Geneva 27, Switzerland (email: sadanar@who.ch).

literature on this topic is vast, this review focuses on the application of interview-based measurement techniques to facilitate community-based data collection.

Self-reported morbidity

Early findings in industrialized countries

Efforts to measure the community-based prevalence of reproductive morbidity using household interview surveys in developing countries are reminiscent of the epidemiological research that began several decades ago to assess morbidity at the population level in industrialized countries (12). The following are some of the merits of the household interview approach to estimating the prevalence of morbidity over approaches that rely on hospital statistics or medical examinations: a greater breadth of the population is covered given higher response rates and lower cost; interpretation of findings is simplified; and generalization to the source population is achieved given the population-based sampling frame and strategy (13). Early research findings raised several important methodological issues concerning the validity (i.e. the degree that the measurement measures what it purports to do) and the reliability (i.e. the degree that repeated measures are consistent) of this approach as a means to estimate the prevalence of morbidity in the population.

Validity has been directly estimated by comparing *self-reported* morbidity with external criteria of *observed or measured* morbidity (e.g., clinical examinations, laboratory diagnosis or cross-checks against medical records) for a wide range of conditions, or indirectly by comparing self-reported patterns with expected patterns in different age groups or other subpopulations. Different estimates and standards of association and agreement have been employed depending upon clinical or population-based study designs, as noted in Table 1. For accurate measurement of the prevalence of observed morbidity at the population level, a combination of both high sensitivity, to detect as many cases as possible, and high specificity, to avoid overestimation of cases, is necessary (14). As far as the utility of data collected to identify individuals for further treatment is concerned, high sensitivity is desirable for conditions that have serious negative consequences and are treatable, whereas high specificity is desirable for conditions that are not easily treatable or curable (15). Most early studies did not adjust for the degree of agreement between self-reported and observed measures or for agreement due to chance alone, for example, as estimated by the weighted kappa statistic (16). Even fewer studies estimated the predictive value of a positive test or predictive value of a negative test that takes into account the actual prevalence of the morbidity within a population (17).

Some studies have investigated whether the sensitivity of the questionnaire may be improved through the use of different interview schedules that employ symptom tracer lists beyond disease labels (18), or by different matching criteria that classify individuals within disease categories based on symptom profiles (19). Reliability has been estimated primarily by checking similarities in data collected from a given individual by re-interviewing, from an individual and a proxy respondent such as a family member, or by investigating differences elicited from lay and medical interviewers. For example, Elinson & Trussell compared self-reported morbidity obtained through household surveys with lay interviewers and observed morbidity from three different US population-based morbidity surveys conducted during the 1950s (20). The sensitivity differs for conditions that are hospitalized, non-hospitalized but which received medical attention, and medically unattended conditions, as well as by age, education and income, and examining physician. Based on results from Baltimore, MD, the sensitivity of self-reported morbidity only reaches 22% for all chronic conditions, and for cervicitis, a mere 8% (21). Various factors were hypothesized to account for this substantial underreporting of morbidity: individuals are not aware of their conditions, the use of proxy respondents, recall problems, the deliberate withholding of information, differences in the conceptualization of what constitutes disease or symptoms, limitations of the interview schedule, and language or communication difficulties.

Nevertheless, Woolsey et al. noted that the differences between self-reported and observed morbidity provide evidence that the "dividing line between a healthy state and a diseased state is not sharp; in fact, this phenomenon has the characteristics of a continuum... [that] can be investigated along more than one scale" (18). Furthermore, Mechanic & Newton concluded that morbidity is likely to be reported when conditions are salient to an individual and where the social and psychological barriers to reporting are absent (22).

Taken together, these early studies and reviews clearly document that self-reported morbidity elicited from health interview surveys is only slightly to moderately associated with observed morbidity for a range of conditions. The high degree of false-negative self-reports prevents valid estimation of the prevalence of morbidity, particularly at the individual level. In addition, psychological and behavioural factors, along with health and medical knowledge, influence self-reporting of morbidity.

Recent findings from industrialized countries

Recent studies on the validity of self-reported morbidity in industrialized countries document a wider range of agreement between self-reported and observed morbidity in aggregate. However, comparisons among studies are complicated by different populations, diseases or conditions, and measures of association. Dealing specifically with women and

Table 1. **Measures of association between self-reported and observed morbidity**

		Observed morbidity		
Sensitivity: $se = a/(a+c)$	Self-reported	yes	no	
Specificity: $sp = d/(b+d)$				
False negative: $f- = c/(a+c)$	+	a	b	a+b
False positive: $f+ = b/(b+d)$	−	c	d	c+d
Positive predictive value: $PV+ = a/(a+b)$		a+c	b+d	n
Negative predictive value: $PV- = d/(c+d)$				
Prevalence: $pr = (a+c)/n$				
Percentage of observed agreement: $P_o = (a+d)/n$				
Percentage of expected agreement: $P_e = [(a+c)(a+b) + (b+d)(c+d)]/n^2$				
Kappa statistic: $P_o - P_e /1 - P_e$				

Adapted from *14–16.*

reproductive health, Colditz et al. assessed the validity of self-reported morbidity of major diseases among a large cohort of female nurses (*23*): depending upon the type of cancer, medical record reviews confirmed 69–99% of women's self-reports of nonfatal cancers. In a related validation study of menopause status, medical record reviews confirmed close to 100 % of self-reported status (*24*). However, it could be argued that nurses within the USA are much more familiar with medical conditions than individuals from the general population and thus more inclined to self-report morbidity accurately (*25*). In a review of close to 30 studies that compared self-reported and observed morbidity, Harlow & Linet found that, for larger studies evaluating a range of chronic illnesses, underreporting was more problematic than overreporting. Six of the studies reviewed examined women's recall of reproductive events and exposures, and the authors concluded that reproductive-related events appear to be more accurately recalled than self-reports of chronic conditions.

In contrast, Oakley et al. compared self-reports of pregnancy-related events with medical records, and concluded that for some conditions, especially events relating to birth, women may be more reliable sources than medical records (*26*). As far as the duration of labour is concerned, hospital records define the start of labour synonymously with that of hospital admission, while women time it from their own experience of physical signs, which on average, start two hours before admission. More recently, Zapka et al. have documented differences among socioeconomic groups in this respect (*27*). The odds of agreement between self-reports of mammography and the information on medical records is 3.6 times greater for women attending private clinics than for those using public clinics. The authors suggest that women who attend public clinics are less likely to undergo mammography and are more likely to be transient and thus have incomplete medical records. They postulate that the information communicated to these women may differ from that in private clinics "perhaps making the event or information less salient to them" (*27*).

Selected findings from developing countries

In developing countries, the practical benefit of obtaining epidemiological data through relatively inexpensive household interview surveys to assess and monitor changes in health status, as well as plan for health services, remains attractive. This is because service-based health data continue to reflect small samples of the overall population and few comprehensive morbidity registries exist. It is not surprising that comprehensive reviews of studies of the prevalence of morbidity using health interview surveys in developing or transitional societies also question the validity, reliability and comparability of the estimates obtained. What is more surprising is that most studies that rely on self-reported morbidity to estimate prevalence do not include validation components or critically assess the findings obtained. Reviews of recent studies provide extensive evidence for the limitations of health interview surveys (*28–32*). Summarized below are some of the methodological, conceptual and procedural limitations that hinder the interpretation and usefulness of findings.

Methodological limitations. Prevalence estimates based on self-reports of morbidity for similar populations and conditions appear to be sensitive to minor differences in methodology. Non-standardized interviewing techniques, the use of open-ended or closed questions, the degree of probing, the inclusion of proxy respondents and variations in the length of recall periods, illustrate some of these differences. These may contribute as much to the variations in estimates of the prevalence of morbidity as does any real difference. The infrequent use of tracer list conditions reduces the standardized reporting of morbidity. The frequent neglect of culture-specific disease classifications and definitions of illness reduce the comprehensiveness of findings. Furthermore, estimating the duration and incidence of illness episodes is especially difficult since questions require individuals to report retrospectively into discrete episodes the symptoms they experienced. There is often, however, limited correspondence between an episode of illness and the recall period utilized. Recall periods of more than 2–4 weeks for closed questions, or a few days for open-ended questions, appear to introduce bias from underreporting and misclassification.

Conceptual limitations. Conceptual limitations relate to self-reported morbidity and its cultural, socioeconomic and psychological correlates. Inadequate knowledge of the study population and its expected patterns of illness may prevent appropriate interpretation of results. The neglect of local perceptions and interpretations of symptoms and signs, such as whether mild conditions are viewed as morbidity or natural events, may lead to inaccurate

comparisons across populations. Ignoring that class and gender selectively shape these differences may lead to inaccurate comparisons within populations. Limited understanding of the morbidity of interest may also make results less useful. For example, the infrequent recognition of the seasonality or epidemicity of many symptoms and conditions is especially problematic for recall periods shorter than 12 months.

Procedural limitations. Procedural limitations include the failure to encourage local populations to participate in the design of questionnaires and sampling approaches, or the neglect of traditional providers and alternative care sources. Failure to communicate and discuss findings with the population under study, health authorities, and media seriously hampers the potential use of research findings as an input to health policy. For example, the frequent failure to disaggregate findings by socioeconomic class, small geographical areas, or distance to different types of health facilities, reduces the usefulness of results for subsequent analyses or decision-making.

Comparison with industrialized countries. Broadly confirming the conclusions of validation studies in industrialized countries, studies in developing countries recognize that self-reported morbidity and observed morbidity measure different phenomena and that their comparison should not be expected to yield similar prevalences of morbidity. In contrast, health interview surveys are valuable because of their potential to estimate the prevalence of conditions that may only be self-reported; identify conditions that escape the attention of health services; investigate individual, social and environmental determinants of the self-report of morbidity and subsequent health seeking actions; and assess the consequences or impact of illness. Surveys that combine interviews and examinations provide a more comprehensive profile of morbidity, yet require greater resources (*33*).

For household interview surveys that include a validation study, differences in the types of self-reported morbidity assessed contribute to the variation in sensitivity and specificity, in comparison with observed morbidity. WHO's *Training modules for household surveys on health and nutrition* classify self-reported morbidity into the following types (*34*): symptoms reported without any interpretation (e.g. headache, backache); illnesses that have been interpreted within the social context and have received a lay diagnosis (e.g. anaemia, rheumatism); symptoms that have been previously diagnosed through a clinical interview or examination and then reported by the individual (e.g. tuberculosis, diabetes); and conditions for which the professional diagnosis has been misunderstood or misreported by the individual (e.g. professional diagnosis is schistosomiasis, while the individual reports anaemia). For example, Kroeger notes that the self-report of a morbidity that requires clinical or diagnostic tests to confirm diagnosis—such as hypertension—may provide an estimate of people's knowledge of disease rather than estimate its true prevalence (*34*).

More recently, Murray & Chen have distinguished these differences in reporting based on three categories of morbidity (*35*): conditions that may be both self-reported and observed (i.e. symptomatic conditions); only self-reported conditions (i.e. pain and suffering); and conditions only observed or measured through professional, clinical or laboratory assessments (i.e, asymptomatic conditions). They also conclude that individuals who are aware of asymptomatic morbidity have more contact with the health services and are more knowledgeable about health problems. Evidence from both within and across population-representative surveys in developed and less developed countries show that higher income groups or countries report greater levels of morbidity than lower income groups. This is the case even though the opposite is usually documented in surveys that include clinical or laboratory examinations or medical record reviews (*36, 37*).

Few studies in developing countries have reported interview-based diagnosis of morbidity through the use of algorithms that combine different categories of self-reported morbidity, as an alternative to the direct matching of self-reported and observed morbidity. Kalter argues that the advantages of validated algorithms over interview-based diagnoses are significant (*38*). Brief algorithms may be as sensitive and specific as longer interview schedules and therefore facilitate rapid community assessment. Additionally, a range of algorithms for the diagnosis of interest may be calculated for use in different situations, such as high sensitivity in order to treat all suspect cases, or high specificity to minimize misclassification if over-treatment is a concern.

Self-reported reproductive morbidity
Selected findings from developing countries

Recent efforts to measure reproductive morbidity in developing countries stem from a broader effort to complement mortality indicators or hospital-based studies with measures of acute and chronic morbidity. Signifying one step towards the identification of nonfatal conditions amenable for interview-based diagnosis, several frameworks and taxonomies focus on specifying operational indicators of reproductive morbidity (*39, 40*).

A WHO working group defined reproductive morbidity "as any morbidity or dysfunction of the reproductive tract, or any morbidity which is a consequence of reproductive behaviour including pregnancy, abortion, childbirth or sexual behaviour [and] may include those of a psychological nature" (*41*). Three categories of reproductive morbidity and its subcategories were distinguished: obstetric morbidity (i.e. direct, indirect and psychological maternal morbidity); gynaecological morbidity (i.e. direct,

indirect and psychological morbidity of the reproductive system, including sexually transmitted diseases); and contraceptive morbidity (i.e. local and systemic morbidity caused by modern or traditional fertility regulation).

Demographic and epidemiological surveys. The application of demographic and epidemiological survey techniques to measure the prevalence of reproductive morbidity in developing countries is discussed elsewhere (*10, 42*). Hill et al. provide a comprehensive review of many recent studies and outline three approaches taken (*43*). Briefly, since the 1970s nationally representative sample surveys that primarily addressed fertility and contraceptive use represent the first approach (e.g. World Fertility Surveys (WFS), Demographic and Health Surveys (DHS), Centers for Disease Control and Prevention (CDC) Reproductive Health Surveys, PAPCHILD and the Gulf State Surveys). The sample sizes and topics covered within most of these surveys preclude estimation of adult mortality or morbidity, since they largely reflect international priorities focusing on child health. In an extensive review of demographic data collected in less developed countries, Cleland notes that neither the WFS nor DHS effort was encouraged by its sponsors to devote substantial resources to field experiments (*44*), including alternative approaches to estimating female reproductive morbidity. Secondary analyses of data collected through these large-scale surveys, however, offer an indirect assessment of reproductive health at the population level, for example, sterility (*45*).

A second approach is the more recent inclusion of specific modules on reproductive morbidity within these large-scale household interview surveys. Smaller scale validations studies, such as the qualitative (*46*) and case–control (*47*) studies nested within the Philippines Safe Motherhood Survey or the case–control (*48*) study nested within the maternal morbidity study in Menoufeya Governorate, Egypt, illustrate this efficient approach.

A third approach is population- or community-based household surveys primarily dedicated to estimating the prevalence of reproductive morbidity. These studies rely on self-reported morbidity, observed morbidity, or some combination of these. Earlier community-based studies include the WHO Family Formation Pattern Studies that investigated household formation patterns and maternal and child health outcomes. Although these studies represent a multinational collaborative effort that included gynaecological examinations, variation in reporting and clinical methodologies reduce confidence in the prevalence estimates and prevent comparisons of self-reported and observed morbidity within and across studies (*49, 50*).

Other recent cross-sectional and prospective household interview surveys continue to rely solely on the self-report of reproductive morbidity as a means of establishing prevalence. The limitations of this type of study design require critical review. For example, data collected from 3600 women in south India indicate that those from urban areas self-report a greater number of symptoms associated with less well-defined morbidity (i.e. milder conditions such as menstrual problems and anaemia) than women from rural areas. Urban and rural women, however, are equally likely to report symptoms associated with more distinct morbidity (i.e. potentially more serious conditions such as lower reproductive tract infections and acute pelvic inflammatory disease) (*51*). Also, the higher the women's education level, the higher the reporting of morbidity during antenatal and natal periods (*52*). These results confirm previous findings on the correlates of self-reported morbidity. Despite these limitations, Bhatia & Cleland argue that cross-sectional, interview-based retrospective studies remain the most feasible option for the study of maternal morbidity in developing countries (*53*).

Some studies use a combination of self-reported and observed morbidity but fail to compare findings in terms of sensitivity and specificity (*54, 55*). Other studies are not specifically designed to compare self-reported and observed morbidity (*56*); suffer from low participation or case identification rates (*57*); or collect observed morbidity data limited to symptomatic women (*58*) or to women who self-report as having one or more chronic morbidities (*48*).

A growing number of studies include qualitative investigations of women's perceptions and descriptions of reproductive conditions. Most of these efforts are an attempt to improve the conceptual and methodological limitations of household interview surveys as well as the interpretation of results. Among the applied ethnographic and anthropological methods commonly utilized are the following: informal, open-ended interviews; illness narratives, observations and/or participant observations; sorting and ranking of key concepts; and focus group discussions (*59, 60*). Such investigations often document how women describe in their own words their experience with illness, signs and symptoms, and probable cause or consequences of illness. This is particularly useful for studies that include women who are least likely to be familiar with biomedical disease categories. For example, a medical diagnosis of acute pelvic inflammatory disease may be self-reported as "severe pain in the womb, vaginal discharge, and fever"; pre-eclampsia as "ankle swelling"; or prolapse as "feeling of a mass or swelling coming out of the vagina or leaking urine when coughing or sneezing." By design, local descriptions vary by study population. In-depth investigations may also overcome procedural limitations and increase overall participation rates by enhancing rapport with the community.

Validity and reliability of self-reported reproductive morbidity

Concerted efforts to overcome the methodological inconsistencies discussed above have sparked con-

siderable research efforts to develop valid and reliable community-based health interview surveys. The objective of these surveys is to measure the prevalence of reproductive morbidity, notwithstanding the recognition that clinical and laboratory data provide better measures of prevalence. Researchers justify this effort to investigate the feasibility of questionnaires for community diagnosis of reproductive morbidity because of the serious complications of many of the conditions and their cumulative impact on women's health. Also, the high cost of clinical examinations, the unavailability of reliable diagnostic tests appropriate for field conditions, and the high refusal rates to participate in gynaecological examinations contribute to the reliance on interview-based investigations.

The results of earlier research on self-reported morbidity indicate that many attributes of reproductive morbidity make these efforts particularly challenging. For example, many conditions are asymptomatic or lack distinct symptoms (1); are stigmatized and thus likely to be misreported (61); or are so prevalent that their symptoms are considered the norm and thus are not reported as morbidity (62). In addition to stigma, a "culture of silence" prevails that on one hand reflects women's reluctance to reveal private problems to strangers and on the other reflects women's inferior status within the family hierarchy of power (63). Furthermore, local knowledge of reproductive physiology, events and illness shape women's interpretation of signs and symptoms of reproductive conditions differently from those for non-reproductive conditions (64).

Comparison of self-reports with clinical examinations, laboratory diagnosis or medical records. The prevalence of reproductive tract infections, prolapse and many other conditions, as determined by clinical and laboratory measurements, is greater than expected or than that previously known in the community. However, for most reproductive conditions it is not surprising that women's self-reported morbidity is minimally related to clinical and laboratory diagnosis. The sensitivity and specificity of interview-based methods to estimate prevalence compared with the observed or measured morbidity are listed for a range of reproductive morbidities in Table 2, along with the study design and sample. As expected from previous findings, prevalence estimates based on self-reported morbidity are generally more specific than sensitive. Furthermore, results from studies conducted in Bangladesh, Egypt, the Philippines and Turkey document that different approaches to asking questions influence estimates of validity and that different interviewers and interview conditions influence the reliability of prevalence estimates. Results from seven studies in different countries illustrate several of the key methodological issues and are discussed in greater detail below.

Giza, Egypt. Zurayk et al. compared approximately 500 women's self-report of symptoms and signs with clinical and laboratory diagnosis for a range

of reproductive tract infections and prolapse in Giza (65). To compare women's reports of symptoms of discharge with the diagnosis of reproductive tract infections from laboratory examinations, the authors used the following cut-off levels identifying the self-report of morbidity: a report of the presence of discharge; a report of at least one feature of the discharge considered medically suspicious; and a report that the discharge is "unusual" for the woman. Sensitivity and specificity varied considerably for each cut-off level: 79% and 26%, 66% and 40%, and 14% and 88%, respectively. The positive predictive value and overall agreement was approximately 50% at each level. Conversely, women's reporting of discharge is to a greater degree substantiated by physicians' observations, with sensitivity and specificity reaching 91% and 61%, respectively, and a positive predictive value of 86%. Together, these findings indicate that although women are able to report signs and symptoms, laboratory diagnoses are required to confirm the prevalence of reproductive tract infections.

Cobancesme, Turkey. A study of ca. 700 women sampled from the registry of a family planning facility in Cobancesme (66–68) estimated the validity and reliability of women's self-reports compared separately with clinical interviews, examinations, and laboratory-diagnosed morbidity for five groups of conditions: reproductive tract infections, urinary tract infections, pelvic relaxation, anaemia and menstrual disorders (see Table 2). The self-reported prevalences based on symptom check-lists for these conditions was up to three times greater than those for the observed conditions, except for self-reported pelvic relaxation and the diagnosis of prolapse, where the reverse was documented. Estimates of the prevalence of all condition groups differed between clinical interviews and laboratory diagnoses. For example, clinical interviews yielded approximately twice as many reproductive tract infections as laboratory diagnoses. Women were more likely to underreport conditions to the lay interviewer and overreport conditions to physician interviewers (however, all lay interviews preceded physician interviews). Analysis of reported symptoms by contraceptive use indicated that women who used IUDs were significantly more likely than the users of other contraceptive methods to report menstrual disorders, but not the other morbidities under study. Corrected for chance, the reliability of agreement between lay and physician interviews was highest for pelvic relaxation (66%) and lowest for menstrual disorders (40%) and upper reproductive tract infections (37%). The authors concluded that the reliability of the household questionnaire depends partially on the skills and abilities of interviewers, since differences among lay interviewers were statistically significant for conditions that have fewer distinct symptoms and signs, such as anaemia and menstrual disorders.

Analyses of the Giza and Cobancesme data demonstrate the low sensitivity of interview-based

Table 2. **Studies investigating the validity and reliability of the self-report of reproductive morbidity**
(a more detailed version of Table 2 is available on the Web version of the *Bulletin* at: http://www.who.int/bulletin))

Author/ place of study/ study design	Sample size	Results	Comments
Wasserheit (*58*), 1989 Comilla, Bangladesh Cross-sectional community-based survey	2129 interviewed 472 symptomatic women clinically examined (74% participation rate of symptomatic women)	– 68% of those clinically examined were diagnosed with vaginal, cervical or pelvic infection – Sensitivity of self report of reproductive tract infections: 0.68	– Excluded asymptomatic women – Low participation rate may bias results
Bang et al. (*57*), 1989 Maharashtra, India Cross-sectional community based survey	654 interviewed and clinically examined (59% participation rate of interviewed women)	– 55% of all women report at least one gynaecological or sexual disease while 92% observed – 43% of asymptomatic women diagnosed with morbidity – Overall sensitivity of self-reported symptoms with diagnosis: 0.59	– Low participation may bias results
Younis et al. (*80*), 1993 Zurayk et al. (*78*), 1995 Giza, Egypt Cross-sectional community based survey	509 interviewed and clinically examined (91% participation rate of community sample)	– 52% of all women diagnosed with at least one RTI *Validity:* *Discharge* (*se*: 0.79 *sp*: 0.26) *Prolapse* (*se*: 0.91 *sp*: 0.61)	– High participation rate reduces selection bias affecting results
EFCS (*48*), 1995 (Egyptian Fertility Care Society) Menoufeya Governorate, Egypt Nested case–control validation study	Community-based survey: 4522 women interviewed – 1155 report 1 or more chronic morbidities (25.5% of all women) Case–control study: – 89 cases (report chronic condition & not received treatment) – 57 controls (67 % participation rate of eligible cases)	– 92.8% of all women examined had been circumcised in the past *Validity:* *Uterine prolapse* (*se*: 0.28 *sp*: 0.86) *hemorrhoids* (*se*: 0.49 *sp*: 0.84) *recto-vaginal fistula* (*se*: 0.00 *sp*: 0.98) *stress incontinence* (*se*: 0.27 *sp*: 0.85) *dyspareunia* (*se*: 0.08 *sp*: 0.96)	– Selection of cases limited, low participation rates for both cases and controls; incomplete sampling – Sensitivity of questions addressing chronic conditions found to be relatively low, whereas specificity much higher – Limited generalizability of case–control findings to community survey
Stewart & Festin (*47*), 1995 Philippines Case–control hospital based study	230 women interviewed and medical records reviewed (38% participation rate of cases and controls identified)	*Dystocia n = 48* (*se*: 0.69 *sp*: 0.97) *Haemorrhage n = 53* (*se*: 0.70 *sp*: 0.78) *Sepsis n = 9* (*se*: 0.89 *sp*: 0.83) *Eclampsia n = 16* (*se*: 0.44 *sp*: 0.96)	– Low participation rates and high loss to follow up – Limited to hospital deliveries – Small *n* for each condition – Various combinations of questions to obtain highest *se/sp*
Brabin (*56*) ,1995 River State, Nigeria Cross-sectional survey	Community-based survey: – 410 adolescents (93% participation rate for examinations) – 458 women (88% participation rate for examinations)	*Adolescents <20 years.* – self reported vaginal discharge 82% – 19% diagnosed with STD; 40% with any RTI *Women* – Self reported discharge 63% – 17% diagnosed with STD; 23% with any RTI	– Self reported and observed morbidity only reported in aggregate – Significant reporting of vaginal discharge, itching and lower abdominal pain that is not diagnosed as a medically recognized infection
Goodburn et al. (*71*), 1995 Manikganj District, Bangladesh (one of BRAC population surveillance districts) Prospective community-based study	1403 followed from initial registration to 12 weeks post partum (68% participation) – of these, 701 complete all 6 interviews	– 33% report serious complications during follow-up period *Ankle oedema* (*se*: 0.50 *sp*: 0.97) PV+: 0.68 *sepsis reported at 2 weeks* (*se*: 0.33 *sp*: 0.89) *sepsis reported at 6 weeks* (*se*: 0.28 *sp*: 0.97)	– Simple clinical exams, no laboratory diagnosis – Different recall periods – Authors conclude that over time, women reported more morbidities potentially due to increased awareness and participation in longitudinal study

Table 2. **Studies investigating the validity and reliability of the self-report of reproductive morbidity (continued)**

Author/ place of study/ study design	Sample size	Results	Comments
Bulut et al. (*66*), 1995, 1996 Ronsmans et al. (*67*), 1996 Filippi et al. (*68*), 1997 Cobancesme, Turkey Cross-sectional clinic-based survey	867 interviewed 696 clinically examined (80% participation rate for examinations of women originally sampled and interviewed)	*Validity:* *Menstrual disorders* (*se*: 0.45 *sp*: 0.93) *pelvic relaxation* (*se*: 0.17 *sp*: 0.96) *RTI* (*se*: 0 .49 *sp*: 0.81) *UTI* (*se*: 0.13 *sp*: 0.81) *anaemia* (*se*: 0.58 *sp*: 0.42) *reliability:* Kappa = 40–66%	– Low participation rate may bias results – Differences between lay and clinical interviews – Differences between prompted and unprompted self-reports
Ronsmans et al. (*74*), 1997 South Kalimantan, Indonesia Case–control hospital-based study	284 recruited within hospitals, cases 204 interviewed, controls (72 % participation rate)	*Dystocia n = 72* (*se*: 0 .44 *sp*: 0.92) *haemorrhage n = 65* (*se*: 0.75 *sp*: 0.99) *eclampsia n = 12* (*se*: 0.59 *sp*: 0.91) *pre-eclampsia n = 31* (*se*: 0.42 *sp*: 0.97)	– Different participant characteristics of those interviewed at home and at the hospital – Not generalizable to home-based interviews given biases of hospital based validity study – Also compared specificity to other obstetric complications – Various combinations of questions to obtain highest *se/sp*
Koenig et al. (*79*), 1998 Cross-sectional community-based survey *Combined data from seven Indian studies conducted between 1985-1994* Two urban sites: Baroda and Bombay Five rural sites: Maharashtra, West Bengal, Karnataka, Gujarat and Rajasthan	3174 women clinically examined across all seven sites (reflecting 19%-86% of women interviewed in each of the sites, i.e., participation in clinical examination low particularly for rural sites)	*Self-reported%* Menstrual problems 33–65 Excessive discharge 13–57 L. abdominal pain 9–21 Lower backache* 5–39 Dyspareunia* 1–7 ⩾ 1 more conditions 55–84 *Clinically observed%* Vaginitis 4–62 Cervicitis 8–48 Cervical erosion 2–46 PID 1–24 Prolapse <1– 7 ⩾ 1 more conditions 26–74 *excludes Rajasthan site	– Methods varied considerably across sites, including sampling strategies and clinical definitions – Low participation may bias results, especially in some rural areas – Data aggregated across sites and with wide ranges, sensitivity and specificity not reported – Authors focus on research lessons learned – Rapport with community developed over months
Seoane et al. (*75*), 1998 La Paz and Cochabamba, Bolivia Case–control hospital-based study	1027 women admitted to hospital reviewed to meet case and control eligibility criteria Cases (*n* = 257) Controls (*n* = 428) Non-eligible (*n* = 428)	For best single question with highest sensitivity: *Malpresentation n = 106* (*se*: 0.96 *sp*: 0.70) *labour disorders n = 106* (*se*: 0.21 *sp*: 0.90) *haemorrhage n = 34* (*se*: 0.79 *sp*: 0.68) *eclampsia n = 22* (*se*: 0.50 *sp*: 0.99) sepsis *n* = 3 (*se*: –* *sp*: 0.99) * not calculated since only 3 septic cases	– Cases and controls differed by an index of socioeconomic class and parity – All interviews conducted on last day of discharge: not generalizable to potential community based interviews – *se* and *sp* reported for single questions rather than combinations – One of the few studies that gives confidence intervals for point estimates of *se/sp*
Kaufman et al. (*76*), 1999 Chengjiang and Luliang, Yunnan Province, China Population-based cross-sectional survey	1189 – 2020 women clinically examined and tested for up to five reproductive tract infections – 57% of women interviewed were screened by laboratory tests for all five infections – 85–90% of eligible women aged 15–49 years from 14 selected villages were located and interviewed	For best single question with highest sensitivity (self-reports vs. gold standard laboratory tests) *Trichomonas (n = 2020)* *se*: 0.58 *sp*: 0.65 *PV*+: 0.20 *candidiasis (n = 2020)* *se*: 0.36 *sp*: 0.63 *PV*+: 0.20 *bacterial vaginosis (n = 1189)* *se*: 0.42 *sp*: 0.59 *PV*+: 0.15 *gonorrhoea (n = 2020)* *se*: 0.29 *sp*: 0.93 *PV*+: 0.01 *chlamydia (n = 1643)* *se*: 0.42 *sp*: 0.62 *PV*+: 0.06	– Low participation rates for laboratory tests – 8 of the 14 villages without laboratory results for bacterial vaginosis and chlamydial infections and thus may bias both prevalence and validity estimates – Algorithms based on self-reports not published

methods and caution against the use of algorithms (see *69*) that combine symptoms reported by women and/or clinical signs observed by physicians for the detection of sexually transmitted diseases, similar to findings from Nigeria and Zaire (*56, 70*).

Manila, Philippines. Stewart & Festin (*47*) used a retrospective case–control design to compare hospitalized women's recall of severe obstetric complications and symptoms with data abstracted from medical records at an urban, public hospital in Manila. Cases were selected from medical records documenting women who experienced at least one of the following conditions during the previous 4 years: haemorrhage, dystocia due to obstructed labour, eclampsia or sepsis. Controls included women who were admitted for delivery during the same time period and did not experience any of the four conditions under investigation. Using algorithms that combined responses to several questions rather than responses from one question, the authors were able to increase sensitivity and specificity associated with the self-report of haemorrhage and dystocia (see Table 2). Although the combination of low numbers of cases and a low follow-up rate (38%) may bias the estimates considerably, the findings suggest that self-reported data from large-scale household surveys underestimate severe obstetric morbidity.

Manikgarnj, Bangladesh. Within a prospective cross-sectional community-based study investigating the distribution of obstetric morbidity in Manikgarnj District in rural Bangladesh, investigators enrolled approximately 2100 pregnant women (*71*). Starting at 28 weeks of pregnancy until 12 weeks postpartum, six household interviews and simple physical examinations were scheduled at regular intervals. Close to 1000 women completed interviews within 48 hours after delivery and 1400 women completed interviews at 12 weeks postpartum. Only 700 women completed all six interviews and examinations. The sensitivity of self-reported infection compared with clinical diagnosis of sepsis was significantly lower (see Table 2) than that achieved within the Philippines study discussed above. For signs of pre-eclampsia, self-reported ankle oedema at the antenatal interview yielded both low sensitivity (50%) and positive predictive value (68%). As expected, the reliability (assessed by the kappa statistic) of self-reported morbidity over two time periods was higher for well-defined symptoms (e.g., swelling, 76%) and lower for less distinctive signs (e.g., vaginal discharge, 44% or pain, 20%). The authors noted that additional factors contribute to discrepancies between multiple self-reports, including the recording of symptoms that begin between interviews or differences in the degree of probing that may heighten women's sensitivity to previously existing symptoms. Over the course of this longitudinal study, evidence emerged of a spurious increase in the prevalence of some self-reported morbidities, i.e., a Hawthorne effect (*72*). Qualitative (*73*) and quantitative findings highlight that self-

reports of other postpartum morbidity, in particular perineal tears and infection as well as problems associated with breasts and breastfeeding, were also high and deserve more attention given the impact of these problems on daily life.

South Kalimantan, Indonesia. A recent case–control study in South Kalimantan, Indonesia, investigated whether women's reports of their experience of childbirth accurately represents the magnitude of obstetric morbidity diagnosed within a hospital (*74*). Women who experienced severe obstetric morbidities ($n = 169$ with dystocia, haemorrhage or hypertensive diseases of pregnancy – eclampsia or pre-eclampsia) were recruited from three hospitals as were a sub-group of women ($n = 115$) from the lowest socioeconomic class who had had spontaneous vaginal deliveries with no complications. Of those recruited, 72% were subsequently interviewed either at discharge or roughly 2–12 months later at home. The authors first tested the sensitivity and specificity of single questions and combinations of questions for comparison with medical diagnosis. Only self-reported haemorrhage reached the target specificity of >95%, while maintaining a sensitivity of >50% (see Table 2). In general, there was poor agreement between self-reported morbidity and medical record review, although the degree of agreement varied with the type of complication. Interviews with women tend to overestimate the prevalence of medically diagnosed problems, such as those associated with excessive vaginal bleeding or dysfunctional labour. However, self-reports of eclampsia agreed to a greater extent with reviews of medical records.

Bolivia. A similar cross-sectional hospital-based study in La Paz and Cochabamba, Bolivia, compared women's self-reports of obstetrical complications (i.e. malpresentation, labour disorder, haemorrhage, eclampsia and sepsis) with hospital medical records and clinical examinations (*75*). Of 1027 women giving birth at either hospital included in the study, cases ($n = 257$) were identified from the results of medical examinations, partograph findings, and the outcome of the mother's labour and delivery. Controls ($n = 428$) were selected from women who had a normal labour, delivery, and immediate postpartum period, excluding those with caesarean sections or extensive tears. Cases were more likely to have a lower index of socioeconomic status and be primiparae. However, there were no significant differences between cases and controls in terms of estimates of the validity of the questions they were asked. The lowest sensitivities were for questions related to labour disorders and the highest for postpartum haemorrhage and malpresentation. All questions relating to labour disorders and eclampsia had high specificity values (see Table 2). However, positive predictive value estimates were generally low for individual questions: these ranged from 14% for excessive bleeding as a sign of postpartum hemorrhage, to almost 65% for any seizures as a sign of eclampsia. The authors note that

since the prevalences of obstetrical complications in the community are usually much lower than those found in hospital settings, the community-based positive predictive value estimates would be even lower than those estimated within this hospital-based study.

Yunnan, China. With the goal of improving the diagnosis and management of reproductive tract infections locally, a study in the rural province of Yunnan, China, evaluated the accuracy of women's self-reported symptoms, clinical diagnoses using algorithms, and low-technology microscopy and biochemical tests with "gold standard" diagnostic tests (*76*). Five infections were assessed: trichmonas, candidiasis, bacterial vaginosis, gonorrhoea and chlamydial infections. Approximately 85% of all eligible women were interviewed. Of these women, 57% underwent laboratory tests for all five infections (*n* = 1153) rendering the prevalence estimates from this study subject to bias. Nevertheless, the sensitivity, specificity and positive predictive value of the different approaches are shown in Table 2. If only self-reported symptoms are relied on, 42–100% of true cases would be missed. Alternatively, if self-reported symptoms alone were used to diagnose and treat the infection, 79–100% of women would be incorrectly diagnosed as having the condition and thus be treated inappropriately. Kaufman et al. note that although clinical examinations provide greater diagnostic accuracy than self-reports, positive predictive value is roughly equivalent to the overall prevalence of each infection. This indicates that the prevalence of clinical signs of disease is high in the non-infected population. The high prevalence of multiple infections (e.g., > 9%) further complicates diagnosis based on symptom reporting or clinical examinations. Depending upon the condition, field-based methods (i.e. wet mounts, Gram staining, pH of discharge, and potassium hydroxide staining) provide different degrees of accuracy. For trichomonas (using both the wet-mount and Gram staining tests) and candidiasis (using both pH and wet-mount tests), sensitivity reached approximately 85%, and specificity and positive predictive value, 100%. No combination of the field tests provided similar accuracy for the other three infections (not shown in Table 2).

Different designs, sampling strategies, variation in participation or follow-up rates, and differences in the population prevalence of the range of reproductive morbidities investigated have considerable impact on the estimation of validity and potential bias. Ronsmans has critically assessed how selection bias affects the estimates of conventional measures of agreement specifically in the context of community or hospital-based reproductive morbidity validation surveys (*77*). Her analysis shows that when a specific morbidity's prevalence is low (e.g. =5%), a survey tool with a specificity and sensitivity of >50% will always overestimate the prevalence of disease, unless the specificity approaches 100%. In addition, the sensitivity and specificity of a set of questions

depend upon the prevalence of reported symptoms in the population studied. Sensitivity is biased upwards and specificity downwards if the study population has a higher proportion of symptomatic women than the general population, whereas the opposite is true if the study population has a lower proportion of symptomatic women. Ronsmans concludes that without knowing the actual prevalence of a morbidity, it is impossible to determine how well a questionnaire will predict the prevalence of disease in unselected populations.

Conclusion

Such validation studies, primarily focusing on a range of gynaecological and obstetric morbidity among women in developing countries, provide empirical evidence that self-reported morbidity and observed morbidity measure different phenomena and therefore different aspects of reproductive health and illness. Differences across study sites, ignoring methodological variations, also underscore that women's self-reports may vary with the medical and sociocultural context, as may clinical and to a lesser extent, laboratory observations. Yet it is not surprising that these studies collectively confirm that woman may either under- or overreport morbidities and that different sets of questions and probes influence the sensitivity and specificity of survey tools. Differences in sampling, participation bias, and the actual population prevalence of disease also influence the estimated validity, generalizability, and comparability of results. These conclusions are somewhat disappointing considering the need to improve epidemiological estimations of reproductive morbidity in the community.

Echoing the findings from earlier studies in industrialized countries, researchers nevertheless consistently argue to place greater value on women's self-reports of the *experience* of reproductive health and illness. Consequently, household interview surveys may be better suited to estimating the impact and context of reproductive morbidity. Zurayk et al. note that self-reports of symptoms and signs provide insight on the "feeling of ill-health in the community" as well as the salience of conditions to women, as illustrated by the "discomfort or interference with their daily routines, or with their feeling of dignity" (*78*). Based on the results from several validation studies in India, Koenig et al. conclude that "little is known about how such morbidity impacts on women's ability to fulfil their various roles — economic, domestic, marital and sexual — or their mental health and well-being" (*79*). Rather than obtaining valid estimates of the prevalence of morbidity, interview-based surveys may provide useful information concerning the disability or burden associated with reproductive morbidities. Methodological studies to develop and estimate the validity of indicators of physical, mental and socio-

economic consequences of reproductive illness warrant further attention and review. ∎

Résumé

Evaluer la santé génésique : examen des approches communautaires visant à évaluer la morbidité

Au cours des dix dernières années, on a mieux pris conscience de l'importance et des conséquences d'une mauvaise santé génésique. Malgré les progrès enregistrés dans le programme d'action, un certain nombre de chercheurs estiment que la santé de la femme et la santé génésique n'ont pas des définitions claires et des méthodes d'évaluation rigoureuses. Une façon d'aller de l'avant consiste à mettre au point et à tester des outils destinés aux enquêtes par entretiens et permettant d'estimer la prévalence de la morbidité génésique.

Dans la première partie de cet article, on passe en revue certaines approches utilisées dans le passé pour estimer la prévalence d'une série de pathologies par le biais d'enquêtes par entretiens menées à domicile ou dans les communautés, dans les pays développés et dans les pays en développement. Pour un large éventail de pathologies, la validité des résultats a été directement estimée en comparant la morbidité *auto-évaluée par les patientes* (par ex. au cours d'entretiens) aux critères extérieurs de la morbidité *observée ou mesurée* (par ex. à l'occasion d'examens cliniques, de diagnostics de laboratoire ou de vérifications croisées avec les dossiers médicaux). La plupart des études qui se basent sur la morbidité signalée par les patientes pour établir la prévalence ne comportent aucun élément de validation, ni n'évaluent de façon critique les résultats obtenus. Diverses limites méthodologiques et conceptuelles empêchent de bien interpréter et d'utiliser ces résultats. Les estimations de la prévalence basées sur l'auto-évaluation de la morbidité pour des populations et dans des conditions analogues semblent être sensibles à des différences mineures de méthodologie. Des techniques d'entretien non normalisées, le recours à des questions ouvertes ou fermées, l'importance du sondage et les variations observées dans la durée des périodes couvertes illustrent certaines de ces différences. Les questions d'ordre conceptuel ont trait au phénomène de l'auto-évaluation de la morbidité et de ses composantes culturelles, socio-économiques et psychologiques. Les questions méthodologiques ont trait à la nature participative de l'évaluation et au courant de communication qui s'est établi.

Dans la deuxième partie, on analyse les efforts déployés récemment pour mesurer la morbidité génésique dans les pays en développement. Des efforts concertés visant à surmonter toutes sortes de limites méthodologiques ont suscité d'importantes recherches ayant pour objectif d'élaborer des enquêtes sanitaires en communauté menées au moyen d'entretiens, qui soient pertinentes et fiables. L'objectif de ces enquêtes est de mesurer la prévalence de la morbidité génésique, même si l'on sait que les données cliniques et de laboratoire donnent une meilleure mesure de cette prévalence. On présente ici un examen détaillé des études de validation des enquêtes sanitaires récentes effectuées dans des communautés ou en milieu hospitalier au Bangladesh, en Bolivie, en Chine, en Egypte, en Inde, en Indonésie, au Nigeria, aux Philippines et en Turquie, comparant les résultats de l'auto-évaluation à ceux des examens cliniques et de laboratoire. Les pathologies évaluées sont notamment les suivantes : infections vaginales, cervicales ou pelviennes, prolapsus, fistule recto-vaginale, hémorragie, septicémie, éclampsie, troubles du cycle et anémie. Dans toutes ces études, pour la plupart des affections génésiques, l'autoévaluation des femmes a peu de liens avec le diagnostic clinique et de laboratoire. Comme pouvaient le laisser penser des résultats antérieurs, les estimations de la prévalence basées sur l'auto-évaluation de la morbidité sont généralement plus spécifiques que sensibles. En outre, les résultats des études effectuées au Bangladesh, en Egypte, aux Philippines et en Turquie indiquent que si l'on pose les questions selon des approches différentes, la validité estimée des résultats ne va pas être la même et que des enquêteurs et des conditions d'enquête différents vont influer sur la fiabilité des estimations de la prévalence. Les résultats de sept études illustrent plusieurs des principaux problèmes méthodologiques qui se posent et y sont analysés plus en détail. Il n'est pas surprenant que ces études confirment collectivement qu'une femme va sous-notifier ou surnotifier des pathologies par rapport aux mesures observées, et que différentes séries de questions et de sondages vont influer sur la sensibilité et la spécificité des outils d'enquêtes. Des différences d'échantillonnage, des biais de participation et la prévalence réelle de la maladie dans la population influent également sur la validité, la comparabilité et la généralisation éventuelle des résultats. Ces conclusions sont quelque peu décevantes vu la nécessité d'améliorer la comparabilité des estimations épidémiologiques de la morbidité génésique dans la communauté.

Reprenant à leur compte les résultats des études antérieures effectuées dans les régions industrialisées, les chercheurs persistent néanmoins à accorder une plus

grande valeur à l'*expérience* qu'ont les femmes de la santé ou de la morbidité génésique. Plutôt que d'offrir des estimations fiables de la prévalence de la morbidité, les enquêtes par entretiens peuvent fournir des informations utiles sur les incapacités ou le fardeau

associés à la morbidité génésique. Il est donc justifié de faire le point sur les travaux méthodologiques visant à élaborer des indicateurs permettant de mesurer les conséquences physiques, mentales et socio-économiques de la mauvaise santé génésique.

Resumen

Medición de la salud reproductiva: examen de los métodos de evaluación de la morbilidad basados en la comunidad

La sensibilización respecto a las dimensiones y consecuencias de la mala salud reproductiva ha aumentado durante el último decenio. A pesar de los progresos del programa de acción, varios investigadores sostienen que, en lo tocante a la salud de la mujer y la salud reproductiva, faltan definiciones claras y métodos de evaluación rigurosos. Una manera de intentar corregir esa situación consiste en diseñar y poner a prueba instrumentos de sondeo basados en entrevistas que permitan estimar la prevalencia de la morbilidad reproductiva.

En la primera parte de este artículo se examinan determinados enfoques empleados en el pasado para calcular la prevalencia de diversas enfermedades mediante encuestas basadas en entrevistas domiciliarias o comunitarias en los países desarrollados y en los países en desarrollo. Se ha estimado directamente la validez de esos métodos comparando la morbilidad *autonotificada* (p. ej., mediante entrevistas) con la morbilidad *observada o medida* mediante criterios externos (p. ej., exámenes clínicos, diagnósticos de laboratorios o cotejo con los archivos clínicos) en el caso de una amplia gama de enfermedades. En la mayoría de los estudios que se basan en la autonotificación de la morbilidad para establecer la prevalencia no se emplean componentes de validación ni se evalúan críticamente los resultados obtenidos. La interpretación y la utilidad de los resultados tropiezan con diversas limitaciones metodológicas, conceptuales y procedimentales. Las estimaciones de la prevalencia basadas en la autonotificación de la morbilidad, considerando poblaciones y enfermedades similares, parecen ser sensibles a pequeñas variaciones de la metodología. Las técnicas de entrevista no normalizadas, el uso de preguntas abiertas o cerradas, el grado de detalle y la distinta duración de los periodos de rememoración son algunas de esas diferencias. Las cuestiones conceptuales guardan relación con la morbilidad autonotificada y con los factores culturales, socioeconómicos y psicológicos con ella relacionados, mientras que las cuestiones procedimentales guardan relación con la naturaleza participativa de la evaluación y el flujo de la comunicación.

En la segunda parte se analizan actividades llevadas a cabo recientemente para medir la morbilidad reproductiva en los países en desarrollo. Los esfuerzos concertados desplegados para superar diversas limitaciones metodológicas han propiciado numerosas investigaciones orientadas a elaborar encuestas de entrevistas sanitarias de base comunitaria válidas y fiables. El objetivo de estas encuestas es medir la prevalencia de la morbilidad reproductiva, aun admi-

tiendo que los datos clínicos y de laboratorio proporcionan mejores estimaciones de esa variable. Se presenta una revisión detallada de estudios recientes de validación de entrevistas sanitarias de base comunitaria u hospitalaria, llevados a cabo en Bangladesh, Bolivia, China, Egipto, Filipinas, la India, Indonesia, Nigeria y Turquía, en los cuales se comparan las estimaciones resultantes de la autonotificación, del examen clínico y de las pruebas de laboratorio. La morbilidad evaluada incluye, entre otras dolencias, diversas infecciones vaginales, cervicouterinas o pélvicas, el prolapso, la fístula rectovaginal, la hemorragia, la septicemia, la eclampsia, los trastornos menstruales y la anemia. En el conjunto de los estudios, para la mayoría de los problemas de salud reproductiva, la morbilidad autonotificada por las mujeres estaba sólo mínimamente relacionada con la deducible a partir del diagnóstico clínico y de laboratorio. Según cabía prever a juzgar por los resultados de investigaciones anteriores, las estimaciones de la prevalencia basadas en la autonotificación de la morbilidad son por lo general más específicas que sensibles. Además, los resultados de estudios realizados en Egipto, Turquía, Bangladesh y Filipinas muestran que la manera de formular las preguntas influye en las estimaciones de la validez, y que los diferentes entrevistadores y las diferentes condiciones de entrevista influyen en la fiabilidad de las estimaciones de la prevalencia. Los resultados de siete estudios ilustran varias de las cuestiones metodológicas clave y se tratan más detalladamente. No es de extrañar que, considerados globalmente, estos estudios confirmen que las mujeres pueden subnotificar o sobrenotificar las enfermedades en comparación con la morbilidad observada, y que los diferentes conjuntos de preguntas y modalidades de encuesta afectan a la sensibilidad y la especificidad de los instrumentos de sondeo. Las diferencias de muestreo, el sesgo de participación y la prevalencia real de las enfermedades en la población también influyen en las estimaciones de la validez, las posibilidades de generalización y la comparabilidad de los resultados. Estas conclusiones son algo decepcionantes si tenemos en cuenta la necesidad de mejorar y hacer comparables las estimaciones epidemiológicas de la morbilidad reproductiva en las comunidades.

Coincidiendo con los resultados de estudios anteriores llevados a cabo en regiones industrializadas, los investigadores abogan sin embargo de forma sistemática por dar mayor peso a la *experiencia* autonotificada por las mujeres en relación con su salud reproductiva y con las enfermedades que la socavan. Más que aportar estimaciones válidas de la prevalencia

de la morbilidad, las encuestas basadas en entrevistas pueden proporcionar información de utilidad sobre la discapacidad o la carga asociadas a los problemas de salud reproductiva. Es preciso revisar el trabajo metodológico a fin de desarrollar indicadores de las consecuencias físicas, mentales y socioeconómicas de los problemas de salud reproductiva.

References

1. **Wasserheit JN.** The significance and scope of reproductive tract infections among third world women. *International Journal of Gynaecology and Obstetrics,* 1989, suppl. 3: 145–168.

2. **Walsh JA et al.** Maternal and perinatal health. In: *Disease control priorities in developing countries.* Oxford, Oxford University Press, 1993: 363–390.

3. **Liskin LS.** Maternal morbidity in developing countries: a review and comments *International Journal of Gynaecology and Obstetrics,* 1992, **37** (2): 77–87.

4. **Fathalla MF.** Inequity in reproductive health: the challenge to obstetricians and gynaecologists. *European Journal of Obstetrics, Gynecology and Reproductive Biology,* 1992, **44**: 3–8.

5. **Germain A, Nowrojee S, Pyne H.** Setting a new agenda: sexual and reproductive health and rights. In: Sen G, Germain A, Chen LC, eds. *Population polices reconsidered: health, empowerment and rights.* Boston, MA, Harvard University Press, 1994: 27–46 (Harvard Series on Population and International Health).

6. *United Nations International Conference on Population and Development. Report of the International Conference on Population and Development, Cairo, 5–13 September 1994.* New York , United Nations, 1995 (1994A/conf.171/13).

7. **Graham WH, Campbell OMR.** Maternal health and the measurement trap. *Social Science and Medicine,* 1992, **35** (8): 967–977.

8. **AbouZahr C.** Maternal mortality overview. In: Murray CJL, Lopez AD, eds. *Health dimensions of sex and reproduction.* Cambridge, MA, Harvard University Press, 1998 (Global Burden of Disease and Injury Series, Vol.3).

9. **Murray CJL, Lopez AD.** Quantifying the health risks of sex and reproduction: implications of alternative definitions. In: Murray CJL, Lopez AD, eds. *Health dimensions of sex and reproduction.* Cambridge, MA, Harvard University Press, 1998 (Global Burden of Disease and Injury Series, Vol.3).

10. **Filippi VG, Graham WJ, Campbell OMR.** *Utilizing survey data on maternity care in developing countries: an illustrative study.* London, Maternal and Child Epidemiology Unit, London School of Hygiene and Tropical Medicine, 1990 (Publication No. 3).

11. **World Bank**. *World development report 1993: investing in health.* New York, Oxford University Press, 1993.

12. **Downes J, Collins SA.** A study of illness among families in the Eastern Health District of Baltimore. *Milbank Memorial Fund Quarterly,* 1940, **18**: 5–26.

13. **Breslow L.** Uses and limitations of the California Health Survey for Studying the Epidemiology of Chronic Disease. *American Journal of Public Health,* 1957, **47**: 168–176.

14. **Galen RS.** New math in the lab: predictive value theory, combination testing, evaluating published data. *Diagnostic Medicine,* 1979, **4**: 31–50.

15. **Fleiss JL.** *Statistical methods for rates and proportions,* 2nd edit. New York, John Wiley & Sons, 1981.

16. **Cohen J.** Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychology Bulletin,* 1968, **70**: 213–220.

17. **Vecchio TJ.** Predictive value of a single diagnostic test in unselected populations. *The New England Journal of Medicine,* 1966, **274**: 1171–1173.

18. **Woolsey TD, Lawrence PS, Balamuth E.** An evaluation of chronic disease prevalence data from the health interview survey. *American Journal of Public Health,* 1962, **52** (10): 1631–1637.

19. **Rubin T, Rosenbaum J, Cobb S.** The use of interview data for the detection of associations in field studies. *Journal of Chronic Diseases,* 1956, **4**: 253–267.

20. **Elinson J, Trussell RE.** Some factors relating to degree of correspondence for diagnostic information as obtained by household interviews and clinical examinations. *American Journal of Public Health,* 1957, **47**: 311–321.

21. **Krueger DE.** Measurement of prevalence of chronic disease by household interviews and clinical evaluations. *American Journal of Public Health,* 1957, **47**: 953–960.

22. **Mechanic D, Newton M.** Some problems in the analysis of morbidity data. *Journal of Chronic Diseases,* 1965, **18**: 569–580.

23. **Colditz GA et al.** Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *American Journal of Epidemiology,* 1986, **123** (5): 894–900.

24. **Colditz GA et al.** Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *American Journal of Epidemiology,* 1987, **126** (2): 319–325.

25. **Harlow DE, Linet MS.** Agreement between questionnaire data and medical records. *American Journal of Epidemiology,* 1989, **129** (2): 233–248.

26. **Oakley A, Rajan L, Robertson P.** A comparison of different sources of information about pregnancy and childbirth. *Journal of Biosocial Science,* 1990, **22**: 477–487.

27. **Zapka JG et al.** Mammography use among socio-demographically diverse women: the accuracy of self-report. *American Journal of Public Health,* 1996, **86** (7): 1016–1021.

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

29. **Ross DA, Vaughan JP.** Health interview surveys in developing countries: a methodological review. *Studies in Family Planning,* 1986, **17** (2): 78–94.

30. **Temaeus I et al.** Health surveys in developing countries: the objectives and design of an international programme. *Social Science and Medicine,* 1988, **27** (4): 359–368.

31. **Murray CJL et al.** Adult morbidity: limited data and methodological uncertainty. In: Feachem RG et al., eds. *The health of adults in the developing world.* Oxford University Press, New York, 1992: 113–160.

32. **Gray RH.** Interview based diagnosis of morbidity and causes of death. In : Boerma JT, Stohrer A, eds. *Measurement of maternal and child mortality, morbidity, and health care: interdisciplinary approaches.* Liege, Ordina Editions, 1994.

33. **Fisher G, Pappas G, Limb M.** Prospects, problems and prerequisites for national health examination surveys in developing countries. *Social Science and Medicine,* 1996, **42** (12): 1639–1650.

34. *Training modules for household surveys on health and nutrition.* Geneva, World Health Organization, 1988 (unpublished document WHO/HST/ESM/88.1).

35. **Murray CJL, Chen LC.** Understanding Morbidity Change. *Population and Development Review,* 1992, **18** (3): 481–503.

36. **Mackenback JP, Looman CWN, van der Meer JBW.** Differences in the misreporting of chronic conditions, by level of education: the effect on inequalities in prevalence rates. *American Journal of Public Health,* 1996, **86** (5): 706–711.

37. **Murray CJL.** Epidemiology and morbidity transitions in India. In: DasGupta M, Chen LC, Krishnan TN, eds. *Health, poverty and development in India.* Delhi, Oxford University Press, 1996: 122–147.

38. **Kalter HD.** The validation of interviews for estimating morbidity. *Health Policy and Planning,* 1992, **7** (1): 30-39.

39. **Graham WJ et al.** *Asking questions about women's reproductive health in community-based surveys: guidelines on scope and content.* London, London School of Hygiene and Tropical Medicine, 1995 (Maternal and Child Epidemiology Unit Publication No. 6).

40. **Fortney JA** *Reproductive morbidity: a conceptual framework.* Research Triangle Park, Durham, NC, Family Health International, 1995 (Working Papers No. WP95-02).

41. *Measuring reproductive morbidity. Report of a technical working group, 30 August to 1 September 1989.* Geneva, World Health Organization, 1990 (unpublished document WHO/MCH/90.4).

42. **Fortney JA.** Reproductive epidemiological research in developing countries. *Annals of Epidemiology,* 1990, **1** (2): 187–194.

43. **Hill A, Bhatti S, Wittgenstein F.** *Assessing the reproductive health of women in developing countries: current knowledge and future opportunities.* Boston, MA, Department of Population and International Health, Harvard School of Public Health, 1995 (background paper prepared for UNFPA global framework for reproductive health).

44. **Cleland J.** Demographic data collection in less developed countries: 1946–1996. *Population Studies,* 1996, **50** (3): 433–450.

45. **Larsen U.** Sterility in sub-Saharan Africa. *Population Studies,* 1994, **48**: 459–474.

46. **Jacobson N.** *Pregnancy in Cagayan de Oro, Philippines: a qualitative study in conjunction with the Safe Motherhood Survey.* Arlington, VA, John Snow Inc.; and Macro International Inc., Calverton, MD, John Snow Inc.,1992 (Report submitted to the MotherCare project).

47. **Stewart MK, Festin M.** Validation study of women's reporting and recall of major obstetric complications treated at the Philippine General Hospital. *International Journal of Gynaecology and Obstetrics,* 1995, **48** (suppl): S53–S66.

48. *The Egyptian Fertility Care Society: study of the prevalence and perception of maternal morbidity in Menoufeya Governorate, Egypt. Final Report.* Cairo, Mohandessin, 1995.

49. **Omran AR, Standley CC.** *Family formation patterns and health: an international collaborative study in India, Iran, Lebanon, Philippines and Turkey.* Geneva, World Health Organization, 1976.

50. **Omran AR, Standley CC.** *Family formation patterns and health — further studies: an international collaborative study in Columbia, Egypt, Pakistan and the Syrian Arab Republic.* World Health Organization, Geneva, 1981.

51. **Bhatia JC, Cleland J.** Self-reported symptoms of gynecological morbidity and their treatment in South India. *Studies in Family Planning*, 1995, **26** (4): 203–216.

52. **Bhatia JC.** Levels and determinants of maternal morbidity: results from a community-based study in Southern India. *International Journal of Gynaecology and Obstetrics*, 1995, **50** (suppl 2): S153–S163.

53. **Bhatia JC, Cleland J.** Obstetric morbidity in South India: results from a community survey. *Social Science and Medicine*, 1996, **43** (10):1507–1516.

54. **Datta KK et al.** Morbidity pattern amongst rural pregnant women in Alwar, Rajasthan — a cohort study. *Health and Population: Perspectives and Issues,* 1980, **3** (4): 282–292.

55. **Prual A et al.** Severe obstetric morbidity of the third trimester, delivery and early puerperium in Niamey (Niger). *Revue Afrique de Santé,* 1998, **2** (1):10–19.

56. **Brabin L et al.** Reproductive tract infections and abortions among adolescent girls in rural Nigeria. *Lancet,* 1995, **345**: 300–304.

57. **Bang RA et al.** High prevalence of gynecological diseases in rural Indian women. *Lancet,* 1989, **1**: 89–89.

58. **Wasserheit JN et al.** Reproductive tract infections in a family planning population in rural Bangladesh. *Studies in Family Planning,* 1989, **20** (2): 69–80.

59. **Gittelsohn J et al.** *Rapid assessment procedures (RAP): ethnographic methods to investigate women's health.* Boston, MA, International Nutrition Foundation, 1998.

60. **Campbell O et al.** *Social science methods for research on reproductive health.* Geneva, World Health Organization, 1999 (unpublished document WHO/RHR/HRP/SOC/99.1).

61. **Barreto TV et al.** Investigating induced abortion in developing countries: methods and problems. *Studies in Family Planning,* 1991, **23** (3): 159–170.

62. **Dixon-Meuller R.** The sexuality connection in reproductive health. *Studies in Family Planning,* 1993, **24** (5): 269–282.

63. **Khattab H.** *The silent endurance: social conditions of women's reproductive health in rural Egypt.* Cairo, The Population Council, 1992.

64. **Sadana R, Snow R.** Balancing effectiveness, side-effects and work: women's perceptions and experiences with modern contraceptive technologies in Cambodia *Social Science and Medicine,* 1999, **49** (3): 343–358.

65. **Zurayk H et al.** Comparing women's reports with medical diagnosis of reproductive morbidity conditions in rural Egypt. *Studies in Family Planning,* 1995, **26** (1): 14–21.

66. **Bulut A et al.** In search of truth: comparing alternative sources of information on reproductive tract infection. *Reproductive Health Matters,* 1995, **6**: 31–39.

67. **Ronsmans C et al.** Clinical algorithms for the screening of chlamydia trachomatis in Turkish women. *Genitourinary Medicine,* 1996, **72**: 182–186.

68. **Filippi V et al.** Asking questions about women's reproductive health: validity and reliability of survey findings from Istanbul. *Tropical Medicine and International Health,* 1997, **2** (1): 47–56.

69. *Report of a WHO Study Group: Management of patients with sexually transmitted diseases.* Geneva , World Health Organization, 1991 (WHO Technical Report Series, No. 810).

70. **Vuylsteke B et al.** Clinical algorithms for the screening of women for gonococcal and chlamydial infection: evaluation of pregnant women and prostitutes in Zaire. *Clinical Infectious Diseases,* 1993, **17**: 82–88.

71. **Goodburn EA et al.** *An investigation into the nature and determinants of maternal morbidity related to delivery and the puerperium in rural Bangladesh.* Dhaka, Bangladesh Rural Advancement Committee, 1994.

72. **Goodburn EA, Graham WJ.** Methodological lessons from a study of postpartum morbidity in rural Bangladesh. Paper presented at: *IUSSP Seminar on Innovative Approaches to the Assessment of Reproductive Health, Manila, Philippines, 24–27 September 1996.*

73. **Goodburn EA, Gazi R, Chowdhury M.** Beliefs and practices regarding delivery and postpartum maternal morbidity in rural Bangladesh. *Studies in Family Planning,* 1995, **26**: 22–29.

74. **Ronsmans C et al.** Women's recall of obstetric complications in South, Kalimantan, Indonesia. *Studies in Family Planning,* 1997, **28** (3): 203–214.

75. **Seoane G, Castrillo M, O'Rourke K.** A validation study of maternal self reports of obstetrical complications: implications for health surveys. *International Journal of Gynaecology and Obstetrics,* 1998, **62**: 229–236.

76. **Kaufman J et al.** A study of field-based methods for diagnosing reproductive tract infections in Rural Yunnan Province, China. *Studies in Family Planning*, 1999, **39** (2): 112–119.

77. **Ronsmans C.** Studies validating women's reports of reproductive ill health: how useful are they? Paper presented at: *IUSSP Seminar on Innovative Approaches to the Assessment of Reproductive Health, Manila, Philippines, 24–27 September 1996.*

78. **Zurayk H et al.** Concepts and measures of reproductive morbidity. *Health Transition Review,* 1993, **3** (1):17–40.

79. **Koeing M et al.** Investigating women's gynaecological morbidity in India: not just another KAP survey. *Reproductive Health Matters,* 1998, **6** (11): 84–96.

80. **Younis N et al.** A community study of gynecological and related morbidities in rural Egypt. *Studies in Family Planning*, 1993, **24** (3): 175–186.