PN-AAE-
872-04

WORKSHOP ON HEALTH IMPACT EVALUATION

"A REVIEW OF METHODOLOGICAL APPROACHES TO EVALUATING HEALTH PROGRAMS"

IIP OCCASIONAL PAPER, NO. 1

Report of a workshop held at
The Johns Hopkins University
School of Hygiene and Public Health
October 9 - 10, 1986.

Report prepared by
Dr. Ronald H. Gray, M.B.B.S., M.F.C.M., M.SC.
Workshop Coordinator
Institute for International Programs
The Johns Hopkins University
School of Hygiene and Public Health
615 N. Wolfe St.
Baltimore, Maryland 21205

301/955-7983

Funded by
the United States Agency for International Development (USAID)
under grant number DAN-5057-G-SS-5088

CONTENTS

Introduction	Page 1
Measurement of Mortality	Page 2
Measurement of Nutritional Status	Page 5
Measurement of Morbidity	Page 7
Measurement of Coverage and Utilization	Page 13
Integration of Demographic and Epidemiologic Studies	Page 15
Sampling Strategies	Page 16
Anthropological and Observational Studies	Page 19
Cost Effectiveness Analysis	Page 20
Executive Summary and Recommendations for Further Research	Page 20
Workshop Agenda and List of Speakers	Appendix A
Key References	Appendix B
Tables and Figures	Appendix C

The Institute for International Programs (IIP) was founded in 1985 at The Johns Hopkins University School of Hygiene and Public Health to coordinate interdepartmental and interdisciplinary research on health projects in developing countries. The Institute is involved in activities in many fields of international health, with a major focus on child survival programs in developing countries. Faculty members associated with the Institute are organized into working groups. This publication is a product of the Evaluation Working Group, and is the first in a series of occasional papers to be published by the Institute for International Programs.

Introduction

The Institute for International Programs at the Johns Hopkins University, School of Hygiene and Public Health sponsored a two-day workshop to review methodological approaches to evaluation of the impact of health programs on mortality and morbidity among infants and children in developing countries. The agenda and list of speakers and are given in Appendix 1. Also enclosed is a list of key references suggested by the speakers for each topic on the agenda.

There was consensus that all health programs require provider-based (tier 1)* information on inputs and outputs to monitor activities and facilitate management decisions. These data can usually be derived from administrative records. In many projects it is desirable to evaluate programs at a population level (tier 2) by measuring knowledge of and attitudes toward health activities, as well as coverage and utilization of services. These tier 2 data are usually derived from population based sample surveys which may be single purpose enquiries such as immunization coverage (EPI), use of oral rehydration therapy (ORT) and contraceptive prevalence surveys (CPS), or broad-based health interview surveys (HIS). Emphasis was also given to the importance of disease surveillance for monitoring of immunization programs. In some cases it may also be possible to assess the efficacy of certain health measures in the community. For example, vaccine efficacy can be estimated from surveillance studies if information is available on immunization status of cases, the efficacy of nutritional programs can be monitored by institution or population based anthropometry, and contraceptive use-effectiveness can be measured from retrospective surveys. Although tier 2 evaluation allows monitoring of a program in the community, for most health interventions it is difficult to translate measures of coverage or efficacy directly into estimates of the program impact on morbidity and mortality (tier 3). Impact evaluation, particularly

* The United States Agency for International Development (USAID) has suggested three levels of evaluation, designated as tiers 1, 2, and 3 depending on data sources and outcome measures. Tier 1 implies process evaluation based on service statistics, tier 2 implies measurement of population coverage and use-effectiveness based on survey data, and tier 3 implies measurement of program impact derived from special studies.

estimates of mortality change, requires special studies which can only be conducted under limited circumstances. Such studies entail measurement of the level and trends of mortality and/or morbidity in relation to program activities. Thus tier 3 impact evaluation incorporates, but is not a substitute for routine program evaluation at tier 1 and 2 levels.

There was extensive discussion of the value and rationale for health impact evaluation. It was agreed that impact evaluation is only appropriate if the results are of importance to policy decisions regarding program strategy, allocation of resources or prediction of project outcome, and in some cases for accountability and justification of expenditures. However, impact evaluation is only feasible if declines of mortality or morbidity are large enough to be measurable, and this is contingent on the level of program achievement, the type of intervention and relevant outcome measure, as well as the sample sizes required for study and the duration of observation. These factors must be weighed in relation to the costs of the research.

The Workshop considered the methodology and problems of health impact evaluation in terms of mortality and morbidity measurement, including special applications of surveillance studies and case-control studies, as well as measurement of coverage and utilization of services, including observational studies, and cost-effectiveness assessment.

Measurement of Mortality

Dr. Kenneth Hill gave an overview of direct and indirect methods for estimating mortality levels and trends, primarily from survey data. A summary of the information that can be obtained, and the advantages or disadvantages of each approach is shown in Table 1. The importance of repeated surveys was emphasized both for the measurement of trends in mortality, and as a means to rectify deficiencies in data quality. The "truncated maternity history" approach collecting information on dates of each birth and the survival or date of death for deceased children over a period of five years prior to interview was considered the most reliable and economic method, but great care needs to be given to questionnaire design and interviewer training to ensure consistency over time and to minimize errors. Reporting errors, particularly the

omission of deaths of very young infants, and misstatements of age or timing of events were considered to be the major problems with this method. Sampling errors were also of concern, since with rare events large numbers of respondents are required to provide precision, but logistic/cost constraints often entail sample sizes below those desirable for statistical purposes. It may not, therefore, be feasible to show statistically significant declines in mortality under most circumstances, and impact evaluation may be confined to demonstrating plausible trends rather than testing the hypothesis that a program significantly reduced mortality.

Other approaches such as questions on the date of birth and survival of the most recent child or births and deaths of children within the past 12 months generally provide poor data due to omission and mistiming of events. A promising new method is to ask women at time of delivery about the survival of their last born and next to last born children. This allows estimation of the probability of dying up to approximate ages 2 and 5, respectively. However, the data provide average mortality estimates over a variable period preceding interview which limits the ability to detect recent trends, and since interviews are generally conducted in institutional settings where mothers come for obstetric care, self selection, which may change over time, can undermine external validity and introduce bias. Furthermore, changing prevalence of family planning or declines in fertility can selectively affect the population of mothers and distort mortality trends measured by this method.

Dr. Douglas Ewbank discussed the use of sample registration systems, particularly for evaluation of overall mortality trends in response to health activities in defined populations. The use of multiround surveys was considered more cost-effective than continuous household surveillance, but ideally it is desirable to have more than one survey round per year so as to capture information on seasonal variation and to obtain complete reports of early infant deaths. It is often difficult to obtain baseline mortality data prior to implementation of health activities, and there is frequently need for initial retrospective maternity history studies to establish pre-project mortality levels. Costs can be reduced and supervision enhanced by the

use of health workers to collect data, but care must be taken to avoid biased reporting and excessive demands on personnel time. Some studies have used "control" or "comparison" areas in a quasi-experimental design to monitor non-program trends. However, problems of comparability can arise if control area population characteristics or data collection differs from that of the experimental study areas. In all follow-up studies careful training and supervision is required to maintain data quality and consistency. An advantage of population sample registration areas is the ability to better link individual level mortality data to socioeconomic characteristics, health utilization or morbidity data for research purposes, but such intensive investigation may undermine the 'representativeness" of the study population relative to the general population and limit generalization of findings.

Ms. Susan Zimicki reviewed the collection of data on causes of death. Existing vital records are seldom adequate for this purpose and since most deaths occur without medical attendance, information on diseases preceding death must usually be obtained from retrospective interviews of family members (often termed "verbal autopsy"). Open-ended questions addressed to relatives do not provide reliable information on most diseases because culturally specific classificatory systems often do not coincide with recognized medical diagnoses, and secondary interpretation of responses by interviewers is often biased. Studies have been conducted using symptom-prompted questions with structured questionnaires which record the duration and severity of the symptoms or signs prior to death. An example of a record form is given in Figure 1. The timing of symptoms is important to ensure that relatives do not report unrelated illnesses that occurred long before the terminal events. The medical diagnosis can be based on computer algorithms or review by a group of experts. Also, multiple causes of death and their temporal sequence can be ascertained (e.g., malnutrition or wasting prior to the terminal illness). These methods have been field tested in Bangladesh and Senegal with promising results, particularly in the diagnosis of tetanus, measles, diarrhea and dysentery. Interview time was generally 20 minutes, and recall appeared to be good up to 10-12 months after death. Preliminary anthropologic review is needed to identify and pretest culturally specific questions, and further research

should focus on validation studies to determine the sensitivity and specificity of symptom/sign complexes for diagnosis of specific diseases. Such validation could be conducted by retrospective interviews of mothers whose children died in medical institutions where the diagnosis is known.

Cause of death data can be of value in defining health priorities and for estimating the potential mortality impact of specific intervention strategies. However, it is important to recognize that the majority of childhood deaths (excluding those in the neonatal period) result from the cumulative burden of recurrent or chronic illness and malnutrition. Thus, the measurement of morbidity and nutritional status among survivors is an integral aspect of impact evaluation.

Measurement of Nutritional Status

Drs. Frederick Trowbridge and Michele Foreman reviewed methods for measurement of nutritional status in cross-sectional or longitudinal surveys and nutritional surveillance studies. Clinical diagnosis of malnutrition is seldom feasible in the field with the exception of pathognomic conditions such as xerophthalmia. Also, biochemical measures are too complex and expensive for most purposes, and dietary surveys seldom provide sufficient detail on the quantity or nutritional value of foodstuffs consumed. Thus, most studies depend on anthropometry using measurements of weight, height or arm circumference which have the advantage of being quantitative indices related to general nutritional status, and are cheap and easy to collect under field circumstances. Weight-for-age is the most commonly used indicator of nutritional status in children, but does not distinguish between acute and chronic malnutrition; height or length are more difficult to measure but best reflect past nutrition (e.g., stunting); and the weight/height ratio can overcome errors in age estimation and provide a sensitive indicator of recent illness or nutritional deficits. However, the interpretation of this ratio may be complicated by population variation in body build observed in different racial or ethnic groups. The mid-upper arm circumference which measures fat, muscle and bone, is the simplest and cheapest measure, and has been widely used in relief operations as a robust but nonspecific measure of general nutritional

status between ages 1 to 4 years. Skinfold thickness, although a good indicator of fat stores, requires considerable expertise and can only reliably be used in special studies.

The validity of these anthropometric indices can best be estimated from their ability to predict subsequent morbidity and mortality. From present data, weight- or height-for-age appears to perform best in this regard. The sensitivity and specificity of arm circumference appears to be similar to these more common measures, but needs further empirical testing. Arm circumference should at present be considered as an additional measure of nutritional deprivation to be used along with weight and height, and may serve as a useful measure by itself in relief operations where more complete assessment is not feasible. Although it is not clear whether anthropometric indices are true predictors of mortality in their own right, or whether they reflect underlying socioeconomic or educational factors, the prevalence of moderate or severe malnutrition may be useful as a proxy measure for mortality risk in circumstances where clear relationships between nutritional status and mortality risk have been established. This relationship will, however, vary from one population to another, and the only advantage of using anthropometry in this context is the smaller sample size requirement for survey purposes. The classification of these continuous measures such as weight and height into categories or grades of nutritional status requires definition of cut-off points based on comparison with a standard or reference growth distribution using either percentiles or Z scores. The choice of an appropriate reference population and cut-off levels can be problematic since even the widely used WHO reference population based on the Fels Research Institute and NCHS data has artifactual discontinuities at age 24 months resulting from differences in initial data collection methods for the under two's (longitudinal) and the over two year olds (cross-sectional).

Cross-sectional anthropometric surveys are relatively simple and can provide high quality data, but the costs are not insubstantial. Longitudinal followup provides the highest quality data at considerable cost and complexity, however the results may be affected by selective attrition and regression toward the mean over time. A compromise approach is the use of surveillance data based on long-term monitoring

of available service statistics. However, as with all service based information, varying quality or completeness of measurement and selective service utilization by subgroups of the population can compromise interpretation. Of particular concern in this regard is the non-use of services by the most disadvantaged subgroups who are most at risk. Although surveillance data can be useful in detecting unexpected deterioration in nutritional status, repeat cross-sectional surveys or longitudinal studies are required for measuring the impact of health programs on nutritional status.

Measurement of Morbidity

The measurement of morbidity from selected immunizable diseases was reviewed by Dr. Neal Halsey. Hospital or clinic based data selectively include more severe cases of illness and may not accurately reflect disease patterns in the community, but surveillance of routine institutional records can be of value in monitoring immunization programs and detecting epidemic outbreaks. Similarly, routine reporting of notifiable diseases can be of value, although, as indicated by U.S. experience with measles, the proportion of all cases notified is liable to change as immunization coverage improves.

More precise measurement of morbidity in the community is difficult. Prevalence of prior infection can best be estimated from sero-epidemiologic surveys, but these are expensive and may be impractical under most circumstances. However, for diseases such as measles, interviews using local terminology or structured questions on symptom/sign complexes (e.g., fever, rash, desquamation, cough, conjunctival infection, etc.) have high sensitivity and specificity in most settings, largely because measles is a well-recognized illness which can occur only once in an individual's lifetime. The critical issue is the age at infection since early age of onset is associated with increased severity, complications and case fatality, and with persistent growth deficits among survivors.

The prevalence of paralytic polio can be assessed by lameness surveys of the general population, basing the diagnosis on an asymmetric flaccid paralysis with no sensory loss, and a history of acute onset associated with fever and no subsequent progression of paralysis. WHO

lameness surveys in over 100 countries reveal a high impact of the disease, and evidence of declining prevalence among older children may be indicative of selective mortality. Passive surveillance for paralytic polio substantially underestimates the prevalence of the disease, although routine hospital statistics and notification can help monitor immunization programs. Periodic school based lameness surveys are probably the most cost-effective method of monitoring program impact, despite potential bias due to selective or changing school enrollment. Of particular concern is the nonattendance of paralyzed children.

Neonatal tetanus has a high case fatality and is a major cause of neonatal mortality, especially in South Asia. Also, studies indicate that two doses of tetanus toxoid to pregnant women or improved cord hygiene are highly efficacious in preventing the disease. Thus, the measurement of the impact of preventive programs can be based on retrospective maternal interviews regarding deaths during the first month of life, using structured questions on the cessation of crying or suckling, and the onset of spasms or flaccid paralysis in infants who were apparently normal for at least three days after birth.

The frequency of recurrent illnesses such as diarrhea or respiratory infection can be difficult to measure due to the possibility of repeated episodes, the need for a defined recall period, and less precise maternal recollection of common illnesses. Morbidity from diarrhea in community surveys can be ascertained from a history of loose or watery stools with or without blood and mucous over a period of two weeks preceding interview. Also, acute lower respiratory disease can be monitored by a history of fever, cough and respiratory distress, though this is subject to considerable respondent error. Even physician diagnosis of respiratory disease is problematic without supporting radiological evidence.

There is a pressing need to validate the diagnosis of acute immunizable diseases based on respondent recall and to determine the sensitivity and specificity of proxy measures. For most situations of high disease incidence, diagnostic sensitivity is more important than specificity for monitoring intervention programs.

Dr. Gordon Smith discussed disease surveillance using existing health record systems. Most ministries of health collect a large amount of information often based on WHO recommendations or requirements. However, it is recognized that these data are incomplete and of variable quality, and that collection is inconsistent between institutions or over time. To overcome this problem sentinel surveillance of a small number of well-supervised centers focusing on specific diseases and using consistent case definitions may provide a cost-effective alternative. Another approach is local area monitoring of largely urban populations. For example, a reduction of clinic admission rates for severe dehydration can be used as an indirect indicator of the success of an ORT program. Surveillance can also focus on sentinel diseases such as measles both to monitor program performance and to interrupt transmission as part of the control strategy. There is a need for rapid feedback so that the personnel collecting data can see the relevance of the information to their own work. Also, emphasis must be placed on active surveillance, since passive measures are too vulnerable to selectivity of reporting and changing quality over time.

The Workshop considered special approaches to the evaluation of program impact on morbidity and particularly the use of case-control studies. Dr. John Briscoe and Ms. Beverly Young described the use of case-control studies for evaluation of the impact of water and sanitation programs on diarrhea in the Philippines and Malawi. The objective was to field test a rapid and inexpensive method for assessment of environmental improvements. Cases were children presenting to clinics with diarrhea, and controls were children attending the same clinics with diseases thought to be unrelated to water or sanitation (control children had respiratory infections, measles and malaria). Information on water and sanitation exposure was collected by clinic interview and validated by subsequent home visits. The risk of diarrhea associated with improved water and/or sanitation was estimated from the odds ratio after adjustment for confounding. In both country studies, the risk of diarrhea was reduced in the presence of improved water and sanitation, and this was most marked if analysis was restricted to cases with proven enteric pathogens which are likely to be most affected by these environmental conditions. However, the

study revealed several problems with the case-control approach which require further study.

Misclassification of disease or exposure status is always a difficulty, and studies of diarrhea are particularly problematic because of multiple etiologic agents, only some of which might be related to the exposure of interest. Also, the exposure (e.g., good water supplies) may be hard to define or to measure by interview alone, and can vary in quality over time (e.g., system breakdown in response to flooding). This nondifferential misclassification bias may have diluted the observed association between diarrhea and water/sanitation. Selection bias is an inherent problem of institution based studies because disproportionately more severe cases are included, and because of confounding by socioecoromic factors associated with both disease risk and likelihood of health service utilization. Another important selective factor is distance from the clinic. Choice of appropriate clinic based controls and adjustment for socioeconomic characteristics or place of residence can minimize these biases, but may inadvertently cause "overmatching" of cases and controls, thus diluting the apparent impact. For example, socioeconomic status and place of residence may be strongly correlated with adequacy of water and sanitation. Furthermore, with recurrent diseases such as diarrhea, individuals may be included in the study more than once (e.g., a control may later become a case and vice versa); this is called incidence density sampling and raises serious problems of overrepresentation of those children with recurrent or multiple illnesses, thus introducing variable bias into the measured effects. However, including an individual only once (cumulative density sampling) may underestimate the importance of recurrent or multiple illness. It is probably best to analyze data separately for repeated episodes, and for first episodes of illness so as to ascertain the magnitude of these effects. Finally, the experience with these case-control studies indicates the need for substantial sample sizes to measure small effects while allowing adjustment for confounding and estimation of interaction. Further empirical studies are required and several investigations of shigellosis, measles and low birthweight are ongoing.

Applications of epidemiologic methods to studies of vaccine efficacy for immunization program monitoring and impact evaluation were discussed by Dr. Walter Orenstein. The vaccine efficacy equation based on attack rates in the vaccinated and unvaccinated can easily be expressed as a relative risk (VE(%) = (1-RR) x 100 where VE denotes vaccine efficacy and RR denotes relative risk.) Thus efficacy can be estimated from actual attack rates measured in prospective studies, or as the proportions vaccinated among diseased and non-diseased individuals in the total population. Curves showing vaccine efficacy in relation to the proportion of cases vaccinated, and the overall prevalence of vaccination provide a valuable tool for program managers (see Figure 2). In all vaccine efficacy studies care needs to be exercised in disease definition, case ascertainment, and assessment of vaccination status. Disease definitions must be highly specific and sensitive to avoid misclassification, and as discussed earlier, lay or even physician diagnoses may be in error. Also during outbreak investigations, obtaining a history of disease prior to vaccination or an outbreak may be useful in order to avoid confusion between natural immunity and vaccine efficacy. In most circumstances it is best to omit subjects with a prior history of disease. Including such persons in the calculation of efficacy results in a falsely low efficacy with the magnitude of the error dependent upon the differences in vaccine uptake between ill and well prior to the outbreak, and upon differences in preoutbreak exposure to disease among vaccinees and nonvaccinees. In many instances, where characteristics of vaccinees and nonvaccinees are similar, including persons with prior disease in the efficacy calculation, will result in only a small error.

Methods for case ascertainment vary depending on the frequency of the disease in question. Population based surveillance is probably the most accurate, but with rare diseases such as paralytic polio one may need to rely on clinic surveillance. The determination of vaccination status is difficult, and only vaccine cards or clinic records can provide reliable information. An unconfirmed history of vaccination is non-specific and subject to serious error. Problems also arise with multiple dose vaccines if individuals have inadequate vaccination due to partial completion of the recommended schedule, and it is best to treat

such inadequately vaccinated subjects as a separate group in analysis. There is also a need to ensure comparable disease exposure among the vaccinated and unvaccinated children since the socioeconomic or educational characteristics of the parents may influence both the likelihood of contracting the disease and of obtaining vaccination. Minimal socioeconomic data are needed in order to adjust for such potential confounding.

Cohort studies using total populations or random samples can be used to estimate vaccine efficacy in outbreak situations if attack rates are high, but these studies may be expensive and difficult to conduct. Another effective approach to outbreak investigation is measurement of secondary attack rates among vaccinated and unvaccinated children in families, which can be used as mini-cohort and have the advantage of intrinsic adjustment for disease exposure. However, this is contingent on identifying a sufficient number of families.

Coverage surveys using EPI clusters can be used in non-outbreak settings if a disease history is obtained for older children aged 24-35 months. The advantage is that integration with ongoing EPI surveys minimizes costs and problems of personnel training, but the disadvantages are the reliance on uncertain parental diagnosis, recall of past disease, and an increase in sample size requirements. In case-control studies vaccination histories among cases of an immunizable disease are compared to histories of matched controls, and vaccine efficacy estimated from the odds ratio. The advantage is the economy of effort needed to obtain vaccination histories from a small number of controls rather than the whole population. A particular application is to situations where household immunization cards are not available, and vaccination status must be determined from clinic records. However, with diseases such as polio, the large number of subclinical infections may lead to misclassification of disease status and an underestimation of efficacy. Also, with common diseases (over 10% prevalence) the rare disease assumption needed for the odds ratio approximation to the relative risk is invalid, and adjustment is needed to avoid falsely high estimates of efficacy. The case-control approach has been used in several studies but needs more methodologic development.

In summary, measurement of vaccine efficacy is valuable both for program monitoring (to detect vaccine failures) and for assessment of program impact. Epidemiologic approaches hold promise of contributing in a cost-effective and reliable manner to the collection of this information.

Measurement of Coverage and Utilization

Dr. William Taylor reviewed measures of program performance, and described the methods used by the Control of Communicable Childhood Diseases (CCCD) program at CDC. Direct or indirect process measurements of program inputs and outputs can often be obtained from routine data (e.g., statistics on the number of vaccines given, timeliness with respect to age and completeness in terms of immunization schedules), and clinic review can provide information on health facility assessment using the quality of service performance, and availability and continuity of supplies as indices. However, these tier 1 provider-based studies may not reflect services in the community, which require tier 2 population-based surveys of coverage, utilization and efficacy.

There has been extensive experience with health interview surveys in both developed and developing countries which has recently been reviewed by a number of authors (see references). Poor validation or standardization of questionnaires, as well as culturally specific problems in responses to questions on illness, complicate interpretation of survey data on morbidity and health care utilization. Much greater attention needs to be given to questionnaire design, development of a locally appropriate "vocabulary" of illness and grading of perceived severity, and where possible, limitation of questions to the simplest indicators.

There have been numerous cluster sample surveys to monitor health care utilization/coverage, particularly with regard to immunization, ORT use and family planning. However, many of these surveys focused on only one intervention and thus provided limited and potentially selective information, and there is need for integrated surveys that can provide an overview of health activities in the community.

In most situations, health administrators need to know the population that is actually covered by services as well as the level of

service utilization. This requires community level surveys of samples drawn from the whole population. However, in settings where sampling frames are inadequate and there is only very limited health service coverage restricted largely to populations within the catchment area of health facilities, an intermediate level of evaluation of households near to health centers may be more appropriate. Such intermediate level surveys can provide valuable information on program implementation in those populations best served by the services, but will not provide data on coverage.

A variant of this intermediate level approach is the use of Lot quality assurance sampling (LQAS) to determine whether services in a clinic or local area achieve certain predefined "acceptable" or "unacceptable" levels. LQAS can provide a simple quantitative estimate of program performance using very small sample sizes. However, in order to conduct an LQAS of either clinic records (e.g., vaccination cards) or households, it is necessary to have an adequate sampling frame from which to select units for observation. In most developing countries the major expense and difficulty of population based studies is to construct an up-to-date and comprehensive sampling frame, and the size of the final survey population is often not a critical factor because the prevalence of service indicators of interest is usually of reasonable magnitude. Thus in a community study, LQAS may not provide a major advantage over other sampling strategies, since the most difficult task of devising a sampling frame is still necessary, and the small sample sizes merely result in a dichotomous outcome of acceptable versus unacceptable, whereas a moderately large simple random sample could provide precise estimates of levels of coverage/utilization. Modification of LQAS using multiple random starts to select households or children may be a cost-effective alternative, but has not been tried empirically. However, in a clinic setting where sequentially numbered records are available, LQAS can be used to monitor performance over time, and repeated LQAS surveys or surveys in multiple clinics can be combined to provide more precise estimates that could potentially be used to evaluate clinic based programs.

Integration of Demographic and Epidemiologic Studies

Dr. Ron Gray reviewed the similarities and differences between demographic and epidemiologic approaches to health impact evaluation. In general, demographic studies estimate mortality levels in relation to underlying socioeconomic characteristics without adequate attention to proximate determinants of mortality, particularly use of health care; whereas epidemiologic studies tend to focus on intermediate determinants but are often constrained by inadequate data on mortality estimation and underlying socioeconomic factors. There is a need to integrate the two survey approaches and to provide more data on diseases associated with death. However, collection of additional information may result in excessively long or complex questionnaires which create problems of respondent compliance, and require more skilled and better trained interviewers. Also, longer interview times may not be feasible in the large-scale surveys required for mortality estimation. An alternative approach is the use of population based case-control studies which require smaller sample sizes than surveys, and allow more detailed interviews so as to better reconstruct chains of events and define risk factors. Cases and controls can be identified during population surveys and reinterviewed indepth. For example, cases can be defined as any infant or child death in a specified period, or a death from a specific disease (e.g., diarrhea or measles). In addition, population based case-control studies are not subject to the selection bias that frequently arises with institution based studies. Thus, integrated multiround surveys and case-control studies provide an underutilized but economic and feasible approach to the evaluation of health programs.

All studies of health impact require some baseline measurement of health status and mortality, but it may be difficult to decide whether and when to conduct follow-up studies. Large-scale mortality surveys are expensive and demanding, and it would not be appropriate to mount follow-up rounds unless the health program had achieved levels of coverage at which measurable declines in mortality can be expected. Also, the magnitude of the projected decline in mortality will determine the sample size required for follow-up surveys. To determine whether health coverage is sufficient to warrant further mortality surveys, it is necessary to conduct periodic utilization/coverage (tier 2) surveys

to monitor the program. These tier 2 surveys require smaller samples and simpler interviews than mortality impact studies, and can be fielded at six-month or annual intervals.

There is little information to date with which to estimate the expected mortality impact that may be attributed to specific health interventions. For example, we have inadequate estimates of the reduction in diarrhea case-fatality attributable to ORT use in a community setting; in fact, we often do not know the proportion of deaths due to watery diarrhea or the efficacy of ORT use by mothers in the community. Some of this information can be obtained during baseline or tier 2 monitoring studies, but so far it is not possible to accurately predict the mortality decline that might be achieved by specified levels of ORT coverage. Similar problems arise with immunization programs or nonspecific interventions such as family planning, and with integrated programs providing multiple interventions. There clearly is a need for much more work to determine the mortality reduction attributable to specific health activities, similar to demographic estimates of births averted by family planning programs.

Sampling Strategies

Dr. Stan Lemeshow discussed sampling strategies, with emphasis on cluster sampling. Usually clusters are geographic entities defined by political or administrative boundaries (e.g., villages or city blocks), and most health related surveys in developing countries use cluster sampling for reasons of feasibility and economy. Cluster sampling is often the only feasible approach because sampling frames of households or individuals are seldom available, whereas a list of potential clusters is usually available or can be easily compiled. Also, because interviews can be conducted in a restricted location, there are major savings in the costs of listing households within a cluster, as well as travel time, ease of supervision and survey logistics.

The objective is to obtain clusters that are as heterogeneous as possible by using a multistage sampling scheme, preferably with selection probabilities proportional to size. The latter has the advantage of logistic ease and statistical simplicity in analysis, since the samples are self weighted (i.e., the size of clusters does not enter

into estimates of precision). Ideally, once a cluster is selected the second stage sample of enumeration units should be based on random selection or total coverage. However, random selection requires a listing of enumeration units which may not be available, and may be expensive and difficult to construct. As a consequence, many health surveys (particularly of EPI or ORT) use non-random, second stage sampling by applying selection rules, such as starting with a randomly selected initial household and proceeding in a predefined systematic manner (e.g., spiraling) to identify enumeration units by sequentially visiting contiguous households until the desired number of subjects are studied. In larger clusters such as towns or scattered rural populations, a central starting point is identified by a local landmark (e.g., city square, church or school) and interviewers proceed from the central point to the boundary to select a random starting household. A further variant is to subdivide the larger clusters into contiguous zones and randomly select a zone in which the starting household is identified.

In all these approaches, the non-random second stage sampling can introduce problems, because the households are not selected independently of each other and may not represent the totality of households in the cluster. A particular concern is that geographically related households can lead to "pocketing," whereby contiguous units may be similar with respect to the variable of interest (e.g., immunization status or disease incidence), but unrepresentative of the cluster as a whole. Another concern is the potential for bias if the interviewer can determine either the starting point (because local conditions preclude a true random start), or exercise discretion in selection of successive households (e.g., avoiding difficult slum areas). Further bias can occur if there are no arrangements for revisiting unoccupied households.

The EPI Surveys of 30 clusters and 7 children per cluster to estimate immunization coverage, have been evaluated by use of simulation models which allow comparison with simple random sampling. Under most conditions the EPI strategy achieved reasonable results given the required level of precision of 10 percentage points. However, in simulations with a high degree of pocketing, there was considerable bias (generally leading to underestimation), and much greater variability was

observed with the EPI approach than the simple random sample. As a cautionary note, it was emphasized that cluster surveys should not be used to estimate coverage for local areas because of instability in such local estimates.

Variants of the EPI cluster sampling strategy have been used for studies of the incidence of neonatal tetanus and polio, and to estimate ORT treatment rates for diarrhea, but these applications have not been evaluated by comparison with empirical or simulated simple random sampling. Also, the use of cluster surveys to measure community level variables such as access to health services, water or sanitation may be misleading, because these factors tend to be common to an aggregate of households in a locality. Thus, community level variables are more likely to differ between areas than between individual households within an area.

The sample sizes required for cluster surveys are larger than those needed for simple random surveys because of the design effect, which is the ratio of the variance obtained by cluster sampling to the variance obtained by simple random sampling. The sample size required for a cluster survey depends on the estimated size for a simple random sample at a given level of precision, multiplied by the design effect. Thus, the magnitude of the design effect has important implications for logistics, costs and feasibility, and can vary greatly depending on the intracluster distribution of the factor of interest (i.e., pocketing), and its frequency in the population. For example, surveys of diseases such as measles or pertussis with a high incidence and propensity for local epidemic outbreaks, may be associated with large design effects. The estimation of the design effect requires either information from previous surveys in the same population or generalization from other studies.

In summary, cluster sampling provides the only feasible approach for surveys in most settings, but care must be taken to avoid bias both in cluster selection, and particularly in second stage selection of enumeration units. More work needs to be done on application of cluster sampling to measurment of health interventions other than EPI, and to estimation of design effects for these applications.

Anthropological and Observational Studies

The contribution of traditional anthropologic studies and observational studies to evaluation was reviewed by Dr. Kevin O'Reilly. These data complement the provider or survey data described in previous sections, and because such studies are less culture bound and more "assumption free" than structured interviews, they can yield important insights and generate new hypotheses. Also, preliminary anthropologic studies can be of critical importance for developing culturally specific questionnaires. For example, qualitative studies can define cultural categories or perceptions of illness and identify appropriate terminology. This information can be obtained by interviewing individual informants or by focus group discussion, though the latter tends to yield "public information," which may not truly reflect private individual attitudes or beliefs.

Observational studies are also needed because many behaviors do not result from decisions at a conscious level, and because responses to interviews may present normative prescriptions rather than an accurate description of behaviors. The advantages of structured observation are that the information is quantitative, it can be made relevant to the program under evaluation, and it can avoid selective perception by the interviewer or respondent. Using structured reporting forms, an observer in a household can record all relevent activities at 15-minute intervals for varying periods of time over the course of a day or more, and the presence of an observer does not usually distort behaviors after a period of adaptation. Thus, maternal care of a sick child can be carefully monitored and described. Training of observers by role-playing activities and use of videotapes can be accomplished in a short time, and the studies can be conducted rapidly.

Recent work has led to proposals for rapid behavioral or ethnographic assessment using both traditional ethnographic methods and structured observation. Since many child survival interventions such as feeding behaviors or treatment of diarrhea depend on maternal care in the domestic setting, these studies may be of major importance to our understanding of how women utilize health knowledge and technology and how cultural, educational, or socioeconomic factors affect this process.

Cost-Effectiveness Analysis

The role of cost-effectiveness in impact evaluation was reviewed by Dr. Peter Berman. Cost-effectiveness is a technique for comparing alternative means of achieving a desired effect and selecting the most efficient in terms of cost (i.e., minimize the cost-effectiveness ratio, or maximize effect given a fixed cost). It is primarily a managerial decision tool, and differs from impact evaluation in that the latter attempts to measure the absolute magnitude of effects related to specified inputs, whereas cost-effectiveness addresses effects relative to costs.

Cost-effectiveness analysis requires a single outcome measure of effect, and accurate cost estimates for the intervention, as well as clear identification of relevant alternatives. The cost-effectiveness ratio depends on capital and recurrent costs of the health delivery system, service mix and quality, and utilization (case mix, access, acceptability, etc.). Cost-effectiveness can be best applied to comparisons of health technologies and alternative delivery strategies, but cross-national comparisons are problematic, and it may be difficult to estimate marginal costs for selective or targeted interventions focused on high risk subgroups. Also, measures such as cost per death averted are imprecise and may be misleading, and measures such as days of productive life added may be more socially acceptable indices.

The costs of health impact evaluation were discussed, and it was agreed that the cost-effectivenss of these rather expensive studies was a function of their potential value for policy decisions.

Executive Summary and Recommendations for Further Research

Health impact evaluation (tier 3) is the measurement of the effect of health programs on mortality and morbidity. It incorporates measurement of coverage, utilization and efficacy of services at a community level (tier 2), and provider based information on service inputs and outputs (tier 1). The objective of impact evaluation is to provide information for program managers and policy makers, so as to facilitate decisions on priorities and strategies for health care, to predict project outcome, and in certain circumstances to demonstrate that resources invested in health have achieved a desired impact on

health status. Impact evaluation requires special studies of the levels and trends of mortality and morbidity in relation to program activities, and because of cost or logistic difficulties, such studies can only be conducted under limited circumstances. This Workshop, sponsored by the Institute of International Programs (IIP) of the Johns Hopkins University, focused on the methodological problems and new research directions for health impact evaluation.

Mortality measurement is difficult in the absence of a reliable vital statistics or surveillance system, but demographic methods have been developed to measure both levels and trends of childhood mortality with medium to high accuracy. Single or repeated sample surveys in which mothers are asked about births and deaths of children over the five years preceding interview ("truncated maternity history") was considered the most reliable and economic approach. Further research is needed, however, to devise improved questionnaire designs and sampling procedures. Another approach, asking mothers at time of birth about the survival of their two most recent births, appears promising because of its simplicity, but needs more empirical research. The most difficult area of mortality studies is ascertaining the diseases associated with death from interviews with relatives of the deceased person ("verbal autopsy"). Although studies with structured symptom prompted questions appear promising, much further research is needed to validate responses, devise appropriate questionnaires and gain experience in diverse cultural settings. High priority should be given to this issue because information on diseases associated with death is important both for health impact evaluation and for the setting of health priorities.

Data on causes of death reflect, but do not directly measure, the cumulative burden of recurrent and chronic illnesses and malnutrition in survivors. Nutritional surveys or clinic based surveillance using weight and height measurements in children provide indices of nutritional status which are also powerful predictors of subsequent morbidity and mortality. However, this traditional anthropometry presents logistic difficulties in the field, and arm circumference, because of its simplicity, may be an important alternative measure which warrants extensive empirical research to better define its value as an indicator of nutritional status. Measurement of morbidity, particularly

due to immunizable diseases and diarrhea, is of critical importance to tier 2 and tier 3 evaluation. Institution based surveillance, although of value for program monitoring, can be unreliable for impact evaluation because of vulnerability to changing patterns of health care utilization. Population based morbidity surveys using symptom prompted interviews can suffer problems of diagnostic accuracy similar to cause of death investigations, and there is a need to further validate diagnoses based on respondent recall. Experience with survey data on diseases such as diarrhea, measles and polio are however encouraging.

Morbidity surveys are usually combined with information on health care utilization and coverage, particularly immunization, use of ORT and family planning. However, many investigations are limited to one disease or intervention (e.g., EPI or ORT) and research is needed to develop integrated survey methods, particularly with regard to sampling strategies and questionnaire structure.

New research approaches designed to relate health outcomes to program inputs were discussed, particularly the application of case-control methods for studying program impact on acute diseases or cause-specific mortality. Case-control studies have considerable advantages of small sample size requirements and the ability to adjust for the effects of confounding variables. The reduction in risk of disease or death associated with the program can be estimated from the relative risk (i.e., odds ratio). A variety of new epidemiologic approaches to evaluation of immunization programs have been developed such as normograms providing curves of vaccine efficacy in relation to immunization prevalence (Figure 1), and case-control studies of vaccine efficacy and outbreak investigation. All these innovations need further research and field testing before standard methods can be developed for widespread application.

The selection of study respondents for surveys in developing countries is difficult because of inadequate sampling frames and logistic constraints of field work. Cluster sampling is often the only feasible approach and has been widely used. Although clusters are generally selected by random procedures, second stage sampling units of households or individuals within clusters are frequently selected by non-random or quasi-random methods, and concern was raised over

potential selection bias and measurement errors due to "pocketing" of program inputs. Also, sampling strategies devised for one purpose (e.g., EPI measurement of vaccination coverage) are often applied to other health problems (e.g., ORT use) without adequate consideration of the statistical validity of such extrapolation. There is a pressing need for theoretical and empirical research on sampling methods appropriate to impact evaluation in developing countries, particularly with regard to integrated demographic and epidemiologic surveys in which multiple outcomes require samples of different size and often individuals with different characteristics (e.g., age). Also, research is needed on new sampling strategies aimed at limited target groups such as the catchment population of a clinic.

Another approach to program evaluation is Lot Quality Assurance Sampling (LQAS), in which program performance is dichotomized into "acceptable" or "unacceptable" levels and small samples selected at random to determine the prevalence of the indicator measures. LQAS has advantages in clinic settings where sequentially numbered records are available as a sampling frame, and repeated sampling of records can be used to monitor services over time. LQAS requires much smaller samples than random sampling in this context. However, LQAS may not be advantageous for small population based surveys, because of difficulty in devising an adequate sampling frame, and because the dichotomous outcome provides less detailed information than would a moderate sized random sample. There is a need for further research to define the appropriate applications and limitations of LQAS, and to develop simple methodologies.

Qualitative anthropological studies or quantitative observational studies have important roles in evaluation. Anthropological studies can help define cultural categories of illness or perceptions of services which yield important insights and facilitate the development of culturally relevant questionnaires. Also, structured observational studies can be used to describe behaviors of relevance to health (e.g., feeding during illness) which cannot be measured by interview techniques. There is need for research on how ethnographic methods can be better incorporated into demographic or epidemiologic studies.

The importance of cost-effectiveness in evaluation and managerial decision making was also considered. The cost-effectiveness ratio allows measurement of diverse costs relative to clearly defined outcomes, whereas impact evaluation measures the absolute magnitude of changes in outcome in response to specified inputs. Thus the two approaches complement one another.

In conclusion, health impact evaluation can only be conducted in selected circumstances where the technical expertise and logistic facilities make such studies feasible. In most settings, health impact studies cannot be approached with the rigor demanded of experimental research, and what is needed is plausible evidence of improved outcomes in response to program activities, derived from multidisciplinary observational study methods. This workshop identified the strengths and weaknesses of available methodologies and suggested new directions for future research.

Appendix A

WORKSHOP ON HEALTH IMPACT EVALUATION
School of Hygiene & Public Health

Topic	Speakers
Chairman - Dr. Henry Mosley	
Introduction and Overview	W. Henry Mosley
Mortality	
1. Direct and indirect methods for measuring trends in mortality using survey data	Kenneth Hill
2. Use of sample registration data or mortality surveillance for the evaluation of health programs	Douglas Ewbank
3. Information on cause of death	Susan Zimicki
Morbidity & Methods of Measurement	
Chairman - Dr. Ron Gray	
4. Measurement of nutritional status	Rick Trowbridge
5. Nutritional surveys and longitudinal studies	Michele Forman
6. Measurement of morbidity from selected diseases	Neal Halsey
7. Morbidity surveillance including sentinel surveillance	Gordon Smith
Chairman - Dr. Neal Halsey	
8. Cluster surveys	Stan Lemeshow
9. Case-control methods for evaluation of water and sanitation	Beverly Young John Briscoe
10. Epidemiologic methods for studies of vaccine efficacy	Walt. Orenstein
Chairman - Dr. Robert Black	
11. Integration of demographic and epidemiologic methods for mortality and morbidity studies	R. H. Gray
12. Methods for measuring coverage/utilization let quality assurance	Bill Taylor
13. Observational studies	Kevin O'Reilly
14. Cost-effectiveness of impact evaluation	Peter Berman
15. Summary, strategies and priorities for the future	Ronald Gray

Appendix B

KEY REFERENCES

1. Kenneth Hill, "Direct and Indirect Methods for Measuring Trends in Mortality Using Survey Data"

Illustrative Analysis of Maternity Histories:

Somoza, J. (1980) "Illustrative Analysis: Infant and Child Mortality in Columbia." London: Scientific Reports No. 10, World Fertility Survey.

Adjustment of Registered Deaths:

Brass, W. (1975) "Methods for Estimating Fertility and Mortality from Limited or Defective Data." Chapel Hill, NC: Carolina Population Center, Laboratory for Population Statistics.

Bennett, N. and S. Horiuchi (1981) Estimating the completeness of death registration in a closed population. Population Index 47(2).

United Nations (1983) Indirect Techniques for Demographic Estimation, Manual X. New York: United Nations Department of International Economic and Social Affairs; Population Studies No. 81.

Survival of Children, Parents and Spouses:

United Nations (1983) Indirect Techniques for Demographic Estimation, Manual X, Chapters 3 & 4. References to original sources can be found therein.

Intercensal Methods:

Preston, S. (1983) An integrated system for demographic estimation from two censuses. Demography 20(2).

Preston, S. and N. Bennett (1983) A census-based method for estimating adult mortality. Population Studies 37(1).

Survival of the Previous Child:

Brass, W. and S. Macrae (1984) Childhood mortality estimated form reports on previous births given by mothers at the time of a maternity: 1. preceding births technique. Asian and Pacific Census Forum 11(2).

2. Douglas Ewbank, "Use of Sample Registration Data or Mortality Surveillance for the Evaluation of Health Programs"

Berggren, W.L., Ewbank, D.C., and G.G. Berggren (1981) Reduction of mortality in rural Haiti through a primary-health-care program. New Eng J of Med 304:1324-1330.

D'Souza, S. (1984) Small-area intensive studies for understanding mortality and morbidity processes: two models from Bangladesh - the Matlab Project and the Companiganj Health Project. In Data Bases for Mortality Measurement, United Nations, Population Studies No. 84:146-158.

Mata, L.J. (1978) The Children of Santa Maria Cauque: A Prospective Field Study of Health and Growth. Cambridge, MA: The MIT Press, pp. 43-91.

Moriyama, I.M. (1984) Advantages and disadvantages of continuous registration systems for national, subnational and differential mortality analysis. In Data Bases for Mortality Measurement, United Nations, Population Studies No. 84:47-53.

Padmanabha, P. (1984) Use of sample registration systems for studying levels, trends and differentials in mortality: the experience of India. In Data Bases for Mortality Measurement, United Nations, Population Studies No. 84:54-65.

Wyon, J.B. and J.E. Gordon (1971) The Khanna Study. Cambridge, MA: Harvard University Press, pp. 43-91.

3. Susan Zimicki, "Information on Cause of Death"

Garenne and Fontaine (1986) "Assessing Probable Causes of Deaths Using a Standardized Questionnaire." IUSSP Seminar on Comparative Studis of Mortality and Morbidity, Siena, Italy, July 7-12, 1986.

Gemert, Valkenburg, and Muller (1977) The diagnosis of measles under field conditions. Trop Geog Med 29:303-313.

Jackson (1985) Malaria in Liberian children and mothers: biocultural perceptions of illness vs. clinical evidence of disease. Social Science and Medicine 20(12):1281-1287.

Puffer and Griffith (1967) Patterns of Urban Mortality. Washington, DC: PAHO, pp. 15-19 (inquiry and classification methods); pp. 247-249, 260-262 (weighing system).

Puffer and Serano (1973) Patterns of Mortality in Childhood. Washington, DC: PAHO, pp. 73-76 (multiple causes); chap. XVII (changes in assignment of cause); Appendix I (form).

Ross and Vaughan (1986) Health interview surveys in developing countries: a methdological review. Studies in Family Planning 17(2):78-94.

Zimicki (1986) "Old and New Approaches to Assessment of the Cause Structure of Mortality: A Case Study from Bangladesh." IUSSP seminar on Comparative Studies of Mortality and Morbidity, Siena, Italy, July 7-12, 1986.

4. Rick Trowbridge, "Measurement of Nutritional Studies"

Trowbridge F.L. and N. Staehling (1980) Sensitivity and specificity of arm circumference indicators in identifying malnourished children. Am J Clin Nutrition 33:687-696.

Trowbridge F.L. and H.C. Stetler (1980) Nutritional status surveillance in El Salvador. Bulletin of WHO 58(2):327-332.

5. Michele Forman, "Nutritional Surveys and Longitudinal Studies"

Feinstein, A.R. (1978) Clinical biostatistics: the architecture of cross-sectional research (conclusion). Clin Pharmacol Ther 23(4):481-493.

Brink, E.W. et al. (1983) "The Egyptian national nutrition survey, 1978. Bulletin of WHO 61(5):853-860.

Trowbridge F.L. and L. Newton (1979) Seasonal changes in malnutrition and diarrheal disease among preschool children in El Salvador. Am J Trop Med Hyg 28(1):135-141.

Brown, K.H., Black, R.E., and S. Becker (1982) Seasonal changes in nutritional status and the prevalence of malnutirtion in a longitudinal study of young children in rural Bangladesh. Am J Clin Nutr 36:294-302.

6. Neal Halsey, "Measurement of Morbidity from Selected Diseases"

Black, F.L. (1982) Measles in Viral Infections of Humans, edited by A.S. Evans. New York: Plenum, pp. 397-418.

Evans, A.S. (1982) "Surveillance and Seroepidemiology." In Measles in Viral Infections of Humans, edited by A.S. Evans. New York: Plenum, pp. 43-66.

Voorhoeve, A.M., Muller, A.S., Schulpen, T.W.J., et al. (1977) Machakos project studies: agents affecting health of mother and child in a rural area of Kenya: III. the epidemiology of measles. Trop Geogr Med 29:428-440.

Grenfell, B.T. and R.M. Anderson (1985) The estimation of age-related rates of infection from case notifications and serologic data. J Hyg Camb 419-436.

Heymann, D.L., Floyd, V.D., Lichnevski, M. et al. (1983) Estimation of incidence of poliomyelitis by three survey methods in different regions of the United Republic of Cameroon. Bulletin of WHO 61:501-507.

Bernier, R.H. (1984) Some observations on poliomyelitis lomeness surveys. Rev Infect Dis 6(Supple)S:371-375.

Joseph, B., Ravikumar, R., John, M., Natarajan, K., Steinhoff, M.C., and John, T.C. (1983) Bulletin of WHO 61:833-837.

Newell, K.W., Lehmann, A.D., LeBlanc, D.R., and Osorio, N.G. (1966) The use of tetanus toxoid for the prevention of tetanus neonatorum. Bulletin of WHO 35:863-871.

Expanded Programme on Immunization (1982) Prevention of neonatal tetanus. Wkly Epi Rec 18:137.

Black, R.E., Huber, D.H., and Curlin, G.T. (1980) Reduction of neonatal tetanus by mass immunization of nonpregnant women: duration of protection provided by one or two doses of aluminum - absorbed tetanus toxoid. Bulletin of WHO 58:927-930.

World Health Organization Guidelines for a sample survey of diarrheal disease morbidity, mortality, and treatment rates. WHO/CDD/SER/84.6.

7. Gordon Smith, "Morbidity Surveillance Including Sentinel Surveillance"

Schaffner, W., Scott, H.D., Rosenstein, B.J., and Byrne, E.B. (1971) Innovative communicable disease reporting. HSMHA Reports 86(5):431-436.

Thacker, S.B., Choi, K., and Brachman, P.S. (1983) The surveillance of infectious diseases. J Amer Med Assoc 249(9):1181-1185.

Doll, R. (1974) Surveillance and monitoring. Internatl J Epi 3(4):305-314.

Langmuir, A.D. (1963) The surveillance of communicable diseases of national importance. New Eng J Med 268:2-13.

Rutstein, D.D., Mullan, R.J., Frazier, T.M., et al. (1983) Sentinel health events (occupational): A basis for physician recognition and public health surveillance. 73(9):1054-1062.

8. Stanley Lemeshow, "Cluster Surveys"

Lemeshow, S., and Stroh, G. (1986) "Sampling Techniques for Evaluating Health Parameters in Developing Countries." Prepared for the National Academy of Sciences, 1986.

Lemeshow, S., Tserkovyni, A.G., Tulloch, J.L., Dowd, J.E., Lwanga, S.K., and Kenja, J. (1985) A computer simulation of the EPI survey strategy. Internatl J Epi 14(3):473-481.

Lemeshow, S., and Robinson, D. (1985) Surveys to measure programme coverage and impact: A review of the methodology used by the expanded programme on immunization. World Health Statistics Quarterly 38(1):65-75.

Wolff, M.C., and Black R.E. (1986) "Manual for Conducting a Lot Quality Assessment in an Oral Rehydration Clinic." Department of International Health, School of Hygiene and Public Health, The Johns Hopkins University. Prepared for WHO (unpublished).

9. Beverly Young and John Briscoe, "Case-Control Methods for Evaluation of Water and Sanitation"

Baltazar, J., Briscoe, J., Mesola, V., Solon, F.S., Van Derslice, J., and Young, B. Case-control studies of the effect of environmental sanitation on diarrhea morbidity, II: Philippines. (submitted for publication, 1986).

Briscoe, J., Feachem, R.G., and Raharan, M.M. (1985) "Measuring the Impact of Water Supply and Sanitation Facilities on Diarrhoea Morbidity: Prospects for Case-Control Methods." (WHO/CWS/85.3 CDD/OPR/85.1) Geneva: World Health Organization.

Briscoe, J., Feachem, R.G., Rahaman, M.M. (1986) Evaluating Health Impacts: Water Supply, Sanitation and Hygiene Education. Ottawa: IDRC Press.

Briscoe, J., Baltazar, J., and Young, B. Case-control studies of the effect of environmental sanitation on diarrhea morbidity, III: methodological implications of two field studies. (submitted for publication, 1986).

Smith, P. and Day, N.E. (1984) The design of case-control studies: the influence of confounding and interaction effects. Int J Epi 13:356-365.

Young, B., and Briscoe, J. Case-control studies of the effect of environmental sanitation on diarrhea morbidity, I: Malawi. (submitted for publication, 1986).

10. Walter Orenstein, "Epidemiologic Methods for Studies of Vaccine Efficacy"

Orenstein, W.A., Bernier, R.H., Dondero, T.J., Hinman, A.R., Marks, J.S.. Bart. K.J., and Sirotkin, B. (1985) Field evaluation of vaccine efficacy. Bulletin of WHO 63(6):1055-1068.

11. Ronald Gray, "Integration of Demographic and Epidemiologic Methods for Mortality and Morbidity Studies"

Gray, R.H. (1985) "The Integration of Demographic and Epidemiologic Approaches to Studies of Health in Developing Countries." Paper for session 1.15, IUSSP General Conference, Florence, Italy, June 5-12, 1985.

Gray, R.H. (1986) "Epidemiologic Methods and Case-Control Studies of Mortality and Morbidity." Paper for the IUSSP meeting on Comparative Studies of Mortality and Morbidity, Siena, Italy, July 7-12, 1986.

Malison, M.D., et al. (1986) "Estimating Use of ORS, Vaccine Coverage and Childhood Mortality: A New Approach from Experience in Uganda." Centers for Disease Control (unpublished).

Mosley, W.H. and Chen, L.C. (1984) An analytical framework for the study of child survival in developing countries. Population and Development Review 10(Suppl):25-45.

12. Bill Taylor, "Methods for Measuring Coverage/Utilization Lot Quality Assurance"

Kroeger, A. (1983) Health interview surveys in developing countries: a review of the methods and results. Intl J Epi 12:465-481.

Ross, D.A., and Vaughan, J.P. (1986) Health interview surveys in developing countries: a methodological review. Studies in Family Planning 17:78-94.

Kroeger, A. (1985) Response errors and other problems of health interview surveys in developing countries. World Health Statistics Quarterly 38:15-37.

13. Kevin O'Reilly, "Observational Studies"

Patton, M.Q. (1980) Qualitative Evaluation Methods. Sage.

Appendix C

TABLE I

COMPARISON OF METHODS FOR MEASURING MORTALITY TRENDS

Method	Meas	Age-Sex Patterns	Trends	Geog. Detail	Socio-Econ. Detail	Accuracy	Cost: Per Cap	Cost: Overall
a) Direct								
Census/V.R.	Bo	Y	RA	H	M	Var.	L	H
Multi-Round S.	Bo	Y	RA	L	M	M-H	H	M
Surveillance	Bo	Y	RA	L	H	H	H	H
Maternity Hist.	Ch	Y	IP	L	H	M-H	H	M
Retro. Q. on Deaths	Bo	Y	RA	Var.	M	L	L	L
Survival of Most Recent Birth	Ch	Sex Only	RA	Var.	M	L	L	L
b) Indirect								
CEB/CS	Ch	Sex Only	IA	Var.	M	M-H	L	L
Survival of Previous Child	Ch	N	RA	L	L	?M	L	L
Census Age Distributions	Ad	(Y)	RA	L	L	Var.	L	L
Survival of Close Relations	Ad	Sex Only	IA	L	L	L-M	L	L

Key:

Col. 1: Ch = Child Mortality only, Ad = adult mortality only, Bo = both child and adult

Col. 2: Y = yes, N = no, (Y) = yes theoretically, but rarely in practice

Col. 3: RA = through Repeated Applications only

IP = from Internal detail, and Period specific

IA = from Internal detail, but only broad Averages over time

Cols 4-8: H = High, M = Medium, L = Low, Var = Variable

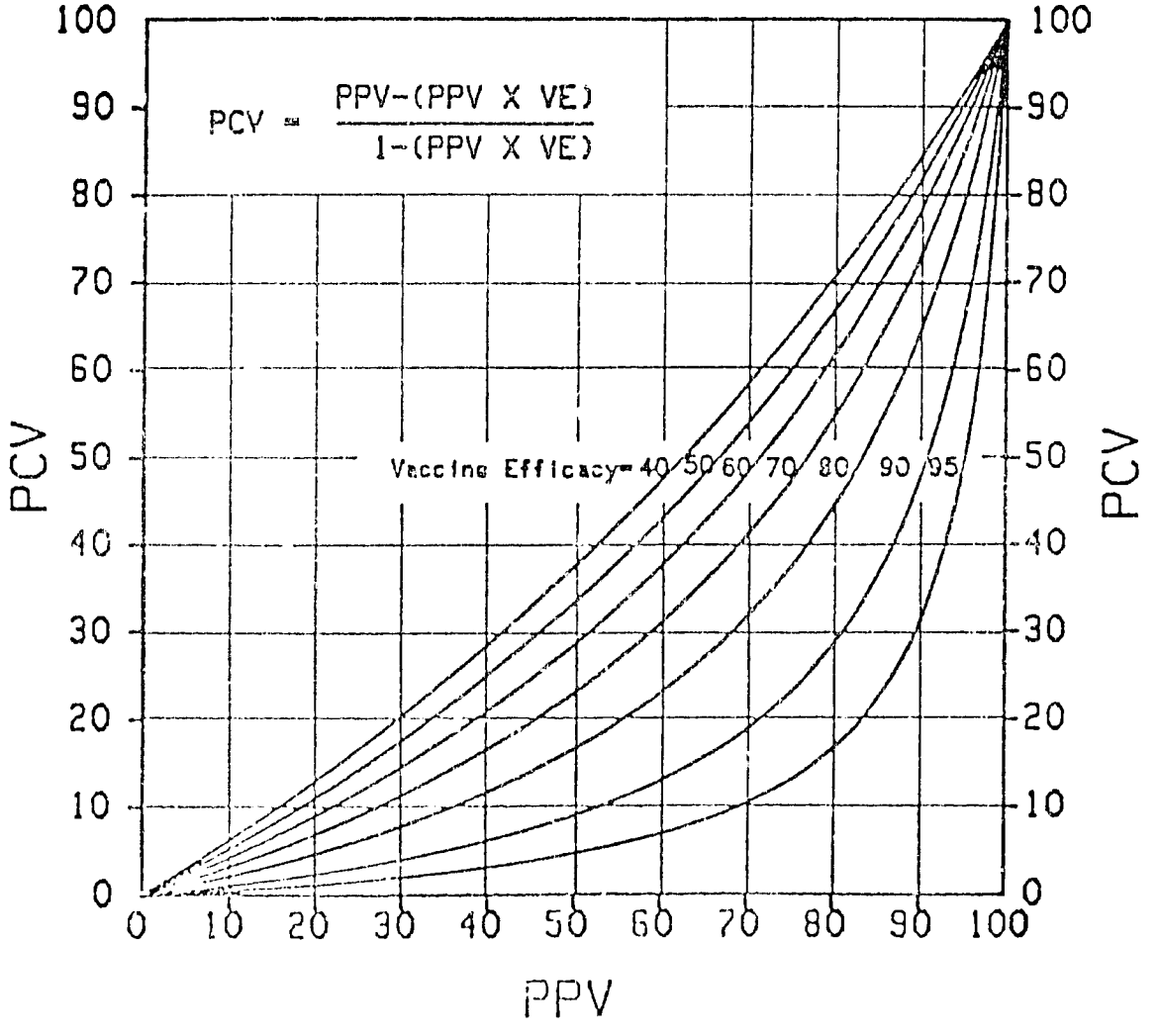
FIGURE 1

Symptom	Present ? No Yes	How long before death did it start?	How long did it last? Until death	How long before death did it stop?	
Fever					Was the fever high or low? Did it occur all the time some of the time. When? ____
Pain					In what part of the body was the pain? _____ Was it severe not so severe?
Rash (such as chickenpox or measles)					Was the rash large small both? Was there water inside? No Yes Did the rash itch? No Yes Where on the body was the rash? _____
Cough					Was this whooping cough? No Yes Did the person vomit after coughing? No Yes Was there blood in what was coughed out? No Yes
Trouble breathing					Did the person's chest hurt? No Yes Was it easier to breathe sitting up? No Yes Was there a sound with breathing? No Yes Was the skin pulled between the ribs? No Yes
Vomiting					Was there blood in the vomit? No Yes Was the vomit black? No Yes
Diarrhoea or dysentery					Was this diarrhoea dysentery both diarrhoea and dysentery? Was there blood in the stool? No Yes Was there mucous in the stool? No Yes How many times a day defecation? _____ Was defecation painful? No Yes Did the eyes become sunken? No Yes

SOURCE: Zimicki, S. "Old and New Approaches to Assessment of the Cause Structure of Mortality: A Case Study from Bangladesh." IUSSP Seminar on Comparative Studies of Mortality and Morbidity, Siena, Italy, July 7-12, 1986.

FIGURE 2

Percentage of Cases Vaccinated (PCV) per Percentage of Population Vaccinated (PPV), for 7 Values of Vaccine Efficacy (VE)



SOURCE: Orenstein, W.A., Bernier, R.H., Dondero, T.J., et al. Field evaluation of vaccine efficacy. Bulletin of WHO 63:1055-1068, 1986.