WATER-SUPPLY AND WASTE-DISPOSAL MANAGEMENT: IMPACT-EVALUATION GUIDELINES

United Nations Centre for Human Settlements (Habitat)

Nairobi 1987

CONTENTS

FOREWORD

The large number of people unserved with water-supply, sanitation and refuse-disposal services in developing countries is partly a consequence of low priorities and limited funds assigned to these sectors in national development plans. While it is believed intuitively that water supply, sanitation and refuse disposal, alone or as part of a comprehensive intervention package, will reduce disease transmission and trigger developmental processes, it is difficult to demonstrate this causal relationship in practice. The inability to pinpoint and quantify most benefits from basic-service interventions is one of the reasons why these essential services have not received the attention they deserve.

Donor agencies, that finance a large proportion of the water- supply and waste-management projects in developing countries, have been recently forced to examine critically the allocation and use of development funds. Consequently, evaluation is increasingly incorporated as a component in the overall project cycle. Most evaluations are, however, confined to assessing the efficiency and effectiveness of the project intervention, and little attention is given to assessing its impact. Not much is, therefore, known about project-impact assessment.

If water supply and waste management are to be accorded high priority in national development strategies and correspondingsectoral investments are to be increased, a wide appreciation of the beneficial impacts of project interventions is essential. This report attempts to elucidate some of the issues involved in impact evaluation of water-supply and waste-management interventions and defines basic guidelines for designing impact-evaluation studies. Impact evaluation in the water-supply and waste-management sectors still remains morean art than a science. It is hoped that the issues identified in this report will stimulate discussion on the subject amongst professionals, increase awareness of the benefits of basic-service provision amongst policymakers and planners, and further understanding of this little-known subject.

I wish to acknowledge the assistance of Dr. M. G. McGarry in the preparation of this report.

Arcot Ramachandran

Dr. Arcot Ramachandran
Under-Secretary-General
Executive Director

INTRODUCTION

Evaluation, as a science, has been developing rapidly over the past two decades. Evaluation in the water-supply, sanitation and refuse-management sector has made slow progress but is now being bolstered by the insistence of most donor agencies that evaluations be incorporated in all projects before funding. Furthermore, many governments are calling for evaluations of projects in these sectors, whether externally funded or not. This is primarily in response to questions being raised as to the effectiveness of governmental initiatives and their long-term success rates which have generally been disappointingly low.

It is surprising, in view of the hundreds of millions of dollars that are spent annually in this sector, that so little is known of the impacts of providing clean water, sanitation services and waste- disposal facilities. All but a few past evaluations have focused on health improvements, but studies have seldom yielded concrete results until recently. Efforts spearheaded by the London School of Tropical Medicine and Hygiene are only now providing quantified assessments of health impacts. However, very little effort has so far been focused on social, economic and development impacts.

The prospective evaluator is typically confronted by a myriad of questions, not the least of which are who will use the results, what is the most cost-effective way of extracting reliable data, and what will be the scope and focus of the evaluation. The evaluator must set out well-defined parameters to maximize the usefulness and reliability of the results. This report tries to clarify some of these issues, to provide a background to the evaluation process and to act as a guide for the evaluator in designing impact-evaluation studies. Each evaluation differs according to local conditions and needs, so that there is no blueprint to follow, and this report does not pretend to provide one.

These guidelines have been written for evaluators, project managers, economists and students working in the water-supply, sanitation and waste-management sector. It first presents the framework into which impact evaluation falls and then focuses on various aspects of impact evaluation itself - principles, constraints, design, planning, the evaluation team, field surveys, data-collection methods and measurement parameters. The vast majority of evaluations that are carried out today are short-term evaluations. Thus, greater attention is given to cross-sectional evaluations (i.e., measurements made at a single, common point in time) carried out over months than to longitudinal evaluations (i.e., similar measurements made at different stages over a period of time) requiring years and substantial financial backing to complete.

How to use these guidelines

This report has been written so that the reader does not have to go through it in its entirety before planning an evaluation. For example, the first chapter, covering the history of evaluation in this sector, is informative but not essential reading. There are several pointers within that chapter, however, that might prove useful, especially on the importance of planning evaluation early in the project cycle, so that meaningful baseline data are collected before the project starts, for analysis on its completion.

The second chapter describes the three different kinds of evaluation - efficiency, effectiveness and impact - and shows how they interrelate. It is essential to know what kind of evaluation is being undertaken and why: the three kinds differ significantly in objectives, methodology and eventual use. Although these guidelines are concerned primarily with impact evaluation, there is often a need to include certain elements of effectiveness evaluation in impact evaluation, in order to understand why impacts occur (or do not, as the case may be). The chapter also describes the various kinds of objectives (goals, purposes and outputs) and how they relate to the three kinds of evaluations. Definition of these levels of objectives is always essential for effective project planning and, naturally, evaluation planning as well.

The principles of evaluation, the various subjects of impact evaluation and preconditions for evaluation are set out in the third chapter. There are many sources of error and bias which are bound to affect the evaluation, if the planner fails to recognize and take steps to avoid them. This chapter also describes various kinds of experimental designs that can be used in setting out an evaluation. Unfortunately, in most projects, there is no pre-project evaluation planning and, therefore, no baseline-data collection. The lack of baseline data prevents the conducting of longitudinal evaluations and confines most evaluation efforts to the cross-sectional design.

In the fourth chapter, a rural water-supply and sanitation project is taken as an example to illustrate how an evaluation is carried out. Careful selection of the evaluation team - the senior professionals as well as the interviewers and support staff - is crucial to the success of evaluation. This chapter describes what to look for in developing survey protocol, training interviewers, planning the fieldwork and providing support for field teams.

The fifth chapter covers the question of what data should be collected in the field and how measurements can best be made. Choosing what parameters to measure largely determines how effectively resources will be used and how efficient the evaluation will be in pinpointing impacts. The choice has to relate to financial constraints, human resources, purpose of the evaluation and identity of end-user. This chapter elucidates most of the important parameters (objectively verifiable indicators) for economic, social, health, community and environmental impacts.

The sixth chapter focuses on how measurements are taken and what are the best sources of information.

I. IMPACT EVALUATION

A. *Historical background*

Development of evaluation methodology

Project evaluation, covering efficiency, effectiveness and impact, has developed since the early 1950s. In those days, formal evaluation methodology emerged at both the country level and from within the United Nations system, but the focus was on efficiency and effectiveness rather than impact. It has been only recently that impact evaluation has taken its place as a tool in assessing project worth.

Impact studies in developing countries in the late 1950s started from a narrowly-focused health base then gradually expanded to include broad socio-economic issues by the early 1970s. In the early days, the concern of international development agencies was the impact of their water-supply and sanitation projects on diarrhoeal diseases. Among the first studies, the most notable was a series of empirical cross-sectional studies sponsored by the World Health Organization (WHO) in Bangladesh, Egypt, Islamic Republic of Iran, Mauritania, Sri Lanka, Sudan and Venezuela: these and others are summarized by Saunders and Warford (1976). Relatively few longitudinal studies on the health impacts of water and sanitation projects have been undertaken in developing countries, mainly owing to the difficulties in experimental control (Warner, 1981).

In the mid-1960s, the implementation of rural-development projects in the newly independent countries of Africa and Asia led to an interest in the overall effects of these projects. Studies with broad interdisciplinary foci were undertaken, with a switch from cross-sectional to longitudinal approaches. Some of the earliest field studies of the broad impacts of rural water-supply projects were carried out in East Africa in the 1960s, including comprehensive cross-sectional investigations of domestic water use in Kenya, Uganda and the United Republic of Tanzania (White and others, 1972). By using questionnaires, analysing excreta and urine samples, performing clinical diagnoses and examining existing records, the evaluation team considered the relationships between improved water supply and disease reduction, and the overall costs of the lack of safe water. They suggested a classification system for water-related diseases by the manner in which they are related to and/or transmitted by water: these diseases were divided into four transmission categories - water-borne, water-washed, water-based and water-related-insect-carried.

This classification was refined by Bradley and Feachem in 1977 and a similar conceptual system for diseases related to excreta was developed by Feachem and others in 1978. In East Africa, the evaluators were able to confirm that water supply alone does not result in substantial benefits. Their approach and methodology greatly influenced subsequent investigators: for example, the disbenefits of inaccessible water which had to be carried over long distances were measured in terms of energy-loss and converted into monetary terms.

Carruthers (1973) was the first person to propose a conceptual model of impact hierarchy, in which improved water supply is considered to bring about "direct effects" (such as increased labour availability) and "second-order benefits" (such as increased crop yields). First-order benefits are still only potential, since they are dependent on certain conditions being met (e.g., user's adoption). In turn, first-order benefits are preconditions for second-order benefits, but the second-order benefits are, in addition, subject to the fulfilment of other conditions as well. Thus, impacts are shown to occur in sequence, in the sense that one level of benefits is dependent on the achievement of a previous level and on the meeting of other conditions. Carruthers' model gave recognition to an integrated approach in impact evaluation and to a holistic view of the project- development cycle, by distinguishing the necessity for complementary inputs. The argument is that water is considered an important but not sufficient condition for development; it requires complementary inputs for success.

In 1975-1976, a broad impact study was carried out by a multidisciplinary team on the rural water-supply programme in Lesotho (Feachem and others, 1978). It consisted of cross-sectional comparisons of water use and detailed investigations of health, community-participation, institutional, political and economic aspects. Unfortunately, the lack of baseline data forced the evaluators to rely on hospital records in the project and control areas. Surprisingly, it was concluded that water supplies, as installed and used in Lesotho, had little impact on health and that diarrhoeas and typhoid were not primarily waterborne. There were no spin-off development activities attributed to the improved water supplies. Feachem's team consolidated Carruthers' argument that there are complex linkages between multiple inputs, complementary preconditions and final impacts, and that water should be considered not as a single input but as an integral component of a development scheme. Their evaluation led to a later comprehensive presentation of evaluation methodology (Cairncross and others, 1980).

Shuval and others (1981) presented a "threshold-saturation theory", linking sanitation, health and socio-economic conditions. The theory states that, for a given socio-economic level, investment results in health improvement but that there is a minimal socio-economic standard below which health benefits are not achieved. Likewise, there is an upper level at which benefits will not increase further. Using adult literacy data (as a socio-economic indicator) and life expectancy at birth (as a health indicator) from 65 countries, the authors found some tentative empirical support for their theory. Later, results from field studies by Magnani and others (1984) in the Philippines supported this theory.

In the field of health-impact evaluation, Briscoe and others (1985) have proposed the "case-control study" design which may overcome many defects in standard designs. In contrast to other designs, the case-control study (also known as "case history" or "retrospective") proceeds from effect to cause, not from cause to effect. For instance, in a project area having both improved and unimproved water supplies, people reporting to clinics having diarrhoea (the "cases") are compared - with respect to the water sources they have used - with those reporting other infections (the "controls"). Using statistical analysis, the relative risk of diarrhoea among people using unimproved sources in comparison with users of improved sources is estimated. The main advantages of the case-control study over other methods are: (a) the required sample size is smaller; (b) the sensitivity and specificity of the disease measure used are substantially higher; (c) only a single round of data is required; and (d) early results are available. Although still having some methodological problems, the case-control study design holds promising potential that warrants development of detailed procedures and field-testing.

Donor agencies have only recently systematically formulated project-evaluation methodologies. For example, the Canadian International Development Agency (CIDA) adopted a Logical Framework Analysis approach in 1980, focusing on three levels of evaluation, project efficiency, effectiveness, and impact. A similar project evaluation model, focusing on system operation, performance and impact, was suggested by the United States Agency for International Development (USAID) (Warner, 1981). These two project-assessment models coincided with a third defined by WHO (1983), in which identical levels of evaluation were defined as system functioning, utilization and impact. The covergence of methodology is illustrated below:

CIDA	USAID	WHO
Evaluation foci	Evaluation levels	Evaluation types
Efficiency	Operation	Functioning
Effectiveness	Performance	Utilization
Impact	Impact	Impact

In summary, impact evaluations of water-supply, sanitation and solid-waste-disposal projects first focused on disease. Gradually, studies of economic values of health improvements, in terms of reduced medical-care costs and increased labour productivity, emerged. Recently, impact-evaluation studies have expanded their coverage to social, economic and institutional aspects within a systematic Logical Framework Analysis.

From the donor's viewpoint, the current situation is best described by CIDA (1977):

"For a long time, it was thought that evaluation was not really necessary as long as the parties involved were reasonably satisfied with what had been achieved.

"Over the years a more critical attitude has emerged. In spite of the growing resources allocated to foreign aid, the gap between the rich and the poor has been widening steadily. Politicians, journalists and the general public want to be better informed about the effectiveness of the aid which is paid out of their tax dollars. Accordingly, the agencies which plan and manage development projects must take a hard look at their operations".

Some examples of impact evaluation

Selected examples of impact evaluations from a variety of developing countries are summarized in this section. Although the review is neither comprehensive nor complete, it does give an overview of how project evaluation has evolved. 1/

Kenya. Carruthers (1973) reviewed the impact and economics of water-supply projects in Kenya. One of the early studies was undertaken around 1965 and applied both longitudinal (before and after) and cross-sectional (with and without interventions) approaches. It indicated that piped water did not bring about significant gains in labour productivity from health improvements, although, besides striking health improvements in children, there were improvements in housing conditions and vector control. However, this 1965 evaluation is beset with technical and sampling problems that make the extensive data difficult to interpret (White and 8 others, 1967).

The Philippines. A six-year evaluation of the Provincial Water Project was carried out for USAID through three rounds of household surveys, to measure impact indicators and explanatory variables. The first surve

1/ For comprehensive reviews, see Esrey and others (1985), Feachem (1984), Hughes (1981), McJunkin (1982), Saunders and Warford (1976) and UNICEF (1983).

was conducted prior to project implementation, the second one to two years after the completion of the project water systems, and the third about five years after completion. The evaluation was quite elaborate, and statistical analyses were extensively applied in data processing and presentation, using sophisticated computer programmes. Given these factors, the results may be considered disappointing, but the lessons gained were valuable:

-The magnitude of economic gains from increased household businesses could not be quantified with the data available.

- There was no conclusive evidence that the project had a substantial health impact.

-The five years allowed for health impacts to appear may not have been sufficient for those impacts to mature.

- Considering the rapid deceleration of economic growth in the country during the evaluation period, it was possible that benefits of the project (e.g., on nutritional status) were offset by deteriorating economic conditions in the project area.

- Some potentially useful morbidity data were lost, owing to poor quality control in data collection and processing during the baseline survey.

When this study was designed (1974-1975), there were no existing USAID guidelines or precedents for water-supply project impact evaluation, and one of the objectives of the evaluation was to establish evaluation methodology for the agency. It was concluded that the evaluation design had methodological problems and should not be replicated in the future. Instead, "smaller-scale, cheaper and more efficient methods" which "favor the measurement of short-term project effects instead of long-term health impacts" were suggested (Magnani and others, 1983).

Thailand. An ambitious Potable Water Project in rural Thailand was implemented during 1966-1972, with 250 systems installed, each with a capacity of 10-50 m^3/hour. The interesting fact about this project is that it was evaluated at different points of time, using different foci - first, efficiency and effectiveness evaluation (Office of the Auditor General for East Asia, 1972); then, effectiveness and impact evaluation (Frankel, 1973); and, finally, impact evaluation (Dworkin and Pillsburg, 1980). The final impact assessment for this project was part of a series to assess the effects of USAID programmes in a number of countries, and these studies were more "management reviews" than impact evaluations, not following any concrete evaluation model. They are "not in the mainstream of impact evaluation nor do they contribute to a general understanding of impact assessment" (Warner, 1981).

United Republic of Tanzania. Impact evaluations of water-supply projects in the United Republic of Tanzania revealed some unexpected results and problems. Some interesting findings (Heijnen and Conyers, 1971; Westman and Hedkvist, 1972) are:

- Confounding variables were numerous; for example, many schemes wereimplemented in settlements where people had already been better off than others in the same area who were not served at all by the project.

- The distance travelled to obtain water was not always reduced with improved water supply; in some cases, it increased during the wet season, because of the switch from traditional sources to improved sources or because more trips were made than before.

- People living less than five minutes' walk from a water tap spent more time collecting water than they did before project completion, owing to increased water consumption.

- In one area, the cattle population showed a rapid increase following the completion of a reservoir, causing overgrazing and soil erosion.

- As the project area had already experienced rapid economic growht, it was impossible to say how much spin-off development was attributable to the project.

- Expected spin-off economic development activities (such as irrigation and fish culture using water from new supplies) did not materialize, owing to lack of technical support for these activities.

The message from these evaluations is clear and sounds a warning to countries implementing their water programmes with unrealistic targets and exaggerated benefit predictions. Warner (1973) later conducted impact studies of water-supply projects implemented in the United Republic of Tanzania during 1968-1970. Longitudinal studies, using household questionnaires, field-testing, observations and records, were carried out on project and control areas. Initially, 30 benefit hypotheses (which are in fact a mixture of goals and purposes) were drawn from defined national development goals. However, as a result of this work, only nine project objectives were recommended for planning water projects: the others were not recommended, since they were either unproven and high-risk or showed little empirical support for the existence of benefits.

In light of the above case histories, it is evident that important concepts, that are well-known today, in fact emerged only recently. Impact evaluation has evolved quite quickly: as recently as 10 to 15 years ago, many impact evaluations still did not follow any systematic methodology. As a consequence, many lessons were learned in a relatively short time: many aspects that are now familiar and taken for granted (for example, the benefit of increased income-generating activities owing to time savings in water collection) were not anticipated until quite recent evaluations.

B. *Problems in impact evaluation*

The previous section has revealed some of the difficulties in impact evaluation; they are numerous and interactive. Those most commonly encountered are presented below.

Lack of baseline data

Inadequate planning frequently results in inadequate or no baseline data being available at the time of evaluation. This is primarily because the project implementing agency and, often, the financial supporting agency do not insist on detailed evaluation of the project in the early planning stages. Unfortunately, the demand for evaluation comes too late in the project, forcing the evaluator to carry out short-term cross-sectional evaluation, with little or no longitudinal data on which to base comparisons over time.

Moving targets

Although the physical appearance of low-income communities in developing countries may not appear to change drastically over the years to an outsider's eyes, fundamental dynamic changes take place there more rapidly than in wealthy communities. Population growth, migration, educational advancement and community development are common uncontrolled changes occurring over the period of impact evaluation, all having an effect on the results. Without establishment of a proper control group for comparison purposes, the impacts of water, sanitation and solid-waste management will invariably be confounded with the effect of these uncontrolled variables.

Over-ambition

With the current emphasis on multidisciplinary approaches and comprehensive analyses, the evaluator often attempts to incorporate too many variables in one study. This leads to excessive and unmanageable data which cannot be adequately analysed or used. It is best to focus on a few key issues and recognize the constraints of limited resources and time.

Inadequate behavioural information

Evaluators tend to focus on the effects of an intervention without understanding the interactions between the intervention and the user-community. For example, improvements in health may be identified and assumed to be a result of latrine installations, without anyone appreciating the fact that most of the people (particularly,the children) are not using the latrines. In this case, the evaluator may be able to say that the effects accrued following an intervention but cannot say why they occurred.

Health-impact evaluation difficulties

In addition to the above there are several methodological problems specifically associated with evaluating health impacts. These were reviewed by Blum and Feachem (1983). Briefly, they are:

(a) In asking about disease episodes such as diarrhoeal episodes, the information gained from the family is likely to be incomplete and unreliable. There may be an unwillingness to disclose personal information to the interviewer; even the mother may not have complete information on her children's diarrhoeal history; and, finally the ability to remember even one's own diarrhoeal patterns is often limited. The maximum recall period over which reliable information on diarrhoea can be obtained is 48 hours.

(b) Disease identification is imprecise, as respondents may describe a wide variety of symptoms. Thus, what the interviewers mean by, for example, diarrhoea may be quite different to the respondent's meaning. Variations in interpretation between respondents are also common.

(c) The effects of water-related and sanitation-related diseases vary according to age group. Diarrhoea, for example, is much more prevalent in young children than in adults. Impact evaluations often fail to distinguish between age groups, and the reasons why an intervention is or is not having an impact becomes difficult to assess.

(d) Disease is influenced by the season, often being more prevalent during certain seasons in tropical climates than during others. Ideally, evaluation would be conducted throughout the year or, at least, during the key seasons.

C. *Resource constraints*

The size and coverage of evaluation are determined by the need for the information and the availability of financial, institutional and human resources. Financial constraints are usually the principal determining factor, but time availability is also adeterminant. In these terms, evaluations can be categorized into three types:

(a) In-depth longitudinal evaluations are normally carried out where resources are the least limiting and there is a need for thorough understanding of specific interventions and impacts. In the past, these studies have been fraught with methodological problems, as discussed above. Much wasted effort and funds have resulted in collection of large quantities of data from which few *definitive conclusions* could be drawn. Reviews of in-depth evaluations are presented in the health-impact literature (Esrey and others, 1985; Blum and Feachem, 1983).

(b) The second level of evaluation has been termed "opportunistic", in that it capitalizes on existing water, sanitation and solid-waste management interventions and reflects the realities of resource constraints. These are tailored to suit the project and may be cross-sectional as well as longitudinal in nature. They are limited in the number of variables they study but, while being limited by resource constraints, they are nonetheless effective in assessing key selected impacts. To be fully effective, however, they must be well planned and implemented, preferably from the project's inception. Opportunistic evaluations are usually carried out by the project itself or by a specialized evaluation team. When the project conducts its own evaluation, external evaluators are normally brought in to assist in the evaluation design and final data analysis, and to ensure that the evaluation remains objective.

(c) The third type of evaluation is short-term and is constrained by lack of planning, resources and time. It is normally carried out by an external funding agency and is primarily concerned with how efficiently the inputs were utilized by the project and how effectively the outputs were achieved.

D. *Timing of evaluation*

Timing over the project cycle

Different types of evaluation are timed at various stages of the project cycle. During the project feasiblity-study stage, a baseline study is carried out, and its data are used in project planning, appraisal and evaluation. At the end of the feasibility study and before approval, the project's merits are assessed by an appraisal.

Monitoring is carried out on a regular basis during project implementation, to check project progress and quality control. Monitoring is primarily concerned with project efficiency. Effectiveness (often termed "ongoing") evaluation can begin early in the project, with the first installation of facilities or services.

On project completion, an end-of-project report is usually prepared, presenting the results of a completion evaluation. The evaluation, at this stage, pays less attention to efficiency and more to effectiveness as well as to some early impacts. Some time after project completion, on maturation of impacts, an ex-post evaluation is undertaken,thistime focusing almost entirely on project impacts. Thus, as the project proceeds, the purposes and foci of evaluation change to suit changing needs and interests.

Timing of evaluation surveys

Several indicators (particularly those related to health) are influenced by climatic and seasonal variations. For example, seasonal influences can be expected to bear on:

(a) Community interest and concern about water supply, which may be high during the dry season but fall as alternative supplies become available in the wet;

(b) Community surveys, which are facilitated by people being available during holidays or when they are not working in the field;

(c) Diarrhoeal morbidity, which is strongly influenced by climatic variation (bacterial diarrhoea morbidity peaks in the warm season in many countries, whereas viral diarrhoeas peak in the cool);

(d) water quality, which is strongly influenced by seasonal variations in rainfall.

II. LOGICALEVALUATION FRAMEWORK

In evaluating water-supply, sanitation and solid-waste projects, there are hundreds of parameters that can be measured, and there are hundreds of factors influencing these measurements. Any evaluation can be extremely complex and confusing, if it is not set out in a logical framework. This chapter deals with project planning and how project evaluation fits into the planning framework. It describes the three types of evaluation (efficiency, effectiveness and impact) and how they determine whether project objectives (results, purposes and goals) have been achieved.

In preparing a project, planners set out its objectives in a Logical Framework Analysis (LFA). This states the expected results, purposes and goals and the means of determining if they have been reached. LFA is now used by several international assistance agencies, such as the German Agency for Technical Co-operation (GTZ), CIDA, USAID and the United Nations system.

LFA is used here as the basis of the Logical Evaluation Framework (LEF). Within this framework, information is presented on why the project is undertaken, what the project is to achieve, and what the expected impacts or goals are. The LEF is described in this chapter, toprovide an overall framework for evaluation methodology and to highlight how impact evaluation fits into and relates to it.

The LEF offers several advantages:

(a) It is an approach which allows everyone to understand all elements of the evaluation within a systematic framework;

(b) It gives the different levels of project objectives and the kinds of evaluation necessary to assess whether they have been reached;

(c) It distinguishes between different types of project evaluation, with clear distinctions between how efficiently the project was managed, how effective it was and whether it had the desired impact;

(d) Being a standard way of setting out the elements of the project, it can be used worldwide to compare evaluations of projects in different countries.

The LEF is shown in figure 1. In the first column, the elements of the project are set out, separated into goals, purposes, outputs and inputs; these are described in the next section. Objectively Verifiable Indicators (OVIs), shown in the second column, are indicators which measure how well (or poorly) the project has performed. They are verifiable, meaning that they can be checked by other evaluators, and they are objective, meaning that they are accurate and not biased.

Project

Project objectives are set out in three levels by the LEF goals, purposes and outputs. The hierarchy is as follows:

(a) The goal of the project is its highest-level objective and includes its impact on the recipient population and environment. The project's goal is the overall reason why the project is implemented. Examples are improvements in health, productivity and environmental quality.

(b) The purposes of the project describe what needs to be achieved by the project to attain its goals. It describes, for example, how the installed facilities should function, and can be utilized and maintained by the community.

(c) The results or outputs of the project are usually its physical installations, such as the number of water-supply standpipes to be installed in the area.

Supporting these objectives are other project elements, namely activities and inputs, which define steps and resources which are necessary to achieve the results or outputs of the project.

Objectively verifiable indicators

The LEF's objectively verifiable indicators (OVIs) are dependent variables which indicate whether or not project results, purposes and goals have been or are being achieved: as the term implies, they must be objective (accurate and without bias) and verifiable. They must constitute objective evidence rather than subjective opinion: that is, in using them, the same conclusions should be drawn, even if different individuals (enthusiast or sceptic) carry out the evaluation. The indicators must relate to the objectives being studied and to the type of evaluation. For example, finding out how many hand-pumps were installed may measure a project's output, but this says nothing about the project's purpose achieved, because the number of hand-pumps says nothing about whether they are being maintained or they are working: neither does it measure achievement of the project's goal, as it does not give any information on the impact that the hand-pumps have, for example, on time saved in collecting water or improvements in public health.

Data collection and sources

Data for efficiency evaluations are usually taken from project reports and then verified in the field. For effectiveness evaluations, most of the information must be obtained from the project team and recipient communities, and from observations of the physical installations. For this purpose, questionnaires, structured discussions and observational techniques are designed and used by the evaluation team. For impact evaluation, most of the information is inferred through observation, structured conversation, household surveys and field and laboratory analyses.

For both effectiveness and impact evaluations, the evaluation team is usually made up of experienced professionals (engineers, epidemiologists, sociologists and statisticians) and a survey team. The survey team can consist of student sociologists and engineers, people from the community and the project team itself. The composition of the survey team will depend a great deal on the individual project evaluation and the funds and human resources available. Detailed consideration is given to impact-evaluation methods, personnel and indicators in chapters V and VI.

Assumptions and preconditions

Inevitably, there are risks of projects failing to meet their objectives. When a project is being planned, risks which are beyond the project's control should not be so high as to endanger project success. For example, it must often be assumed that the government's commitment to the project will not diminish or that there will be no natural catastrophe, such as a hurricane: there is a need to judge whether these risks are worth taking before starting the project. There are five levels of assumptions corresponding to goals, purposes, outputs, inputs and activities: these assumptions are important to the three levels of evaluation, as they are sometimes the reason behind project failure. For example a project may be requried to initiate necessary action to import and supply well-drilling equipment for water-supply development and might assume no undue delays in recruiting qualified project staff to commence project execution and in obtaining government permission and Customs clearance for this purpose. In practice, however, delays incurred at these stages could have a drastic influence on the actual outputs of the project and limit the overall number of wells developed during the life of the project. Likewise, it may be assumed that the provision of sanitation facilities will automatically lead to wide spread utilization by all age groups. In practice, however, failure on the part of children to use the facility may prevent achievement of the project purpose and cause the project to fall short of overall goals, such as a reduction in the prevalence of specific diseases.

Preconditions are also listed in the LEF: these must be met before the project can start. For example,two preconditions for a water-supply project are that all agencies must approve the project, and that adequate financing must be made available. It is often possible to identify specific bottle-necks to the delivery process in any project and attempt to overcome these by setting these out as pre- conditions to project funding or approval. Even so, there is little guarantee that the preconditions would be conformed with, and there is a need to verify how realistic these assumptions and preconditions were when the project was first initiated.

Three kinds of evaluation

Efficiency, effectiveness and impact evaluation are related to the project elements they cover (see figure 1). Although they overlap, they are distinct in what they evaluate and the kinds and uses of information they generate. Efficiency and effectiveness evaluations are carried out at various stages during the project cycle, while impact evaluation is made after the installations are in place. Some impacts do not mature until well after project completion, and this is particularly true of health impacts.

Project-efficiency evaluation

Efficiency evaluation is primarily concerned with how results were achieved, given project inputs and activities. The amount and quality of outputs are compared to the resources and means mobilized. Efficiency evaluation deals with management issues - for example, whether the resources were promptly supplied and well managed. Efficiency can be measured at other levels (purposes and goals), but the central concern is how well the resources were converted into installations and services.

Project-effectiveness evaluation

Effectiveness evaluation looks at the achievement of project purposes. It fills the gap between the physical installations or facilities and their intended impact. For example, an evaluation is made of whether the facilities are functioning and are being well utilized. Effectiveness evalaution is also partly concerned with the achievement of a project's outputs or results and its goals.

Project-impact evaluation

Impact evaluation pertains to the effect which the project has on the recipient population and on the development of the sector and the country as a whole. Impacts will be both positive and negative, foreseen and unforeseen. Impact evaluation focuses on the ultimate effects of the project and seeks to answer such questions as:

- Did the project bring about the intended effects? For each effect: if not, why not? if so, to what degree and why?

Goals

The goal of a project is usually set by central planning bodies and is not normally the responsiblity of project staff. In water- supply, sanitation and solid-waste-disposal projects, some of the goals, for example, can be:

Goals	Examples
Economic improvement	Increased productivity and income-generating activities
Social improvement	Equitable distribution of benefits
Health improvement	Reduction of water-related and sanitation-related diseases
Community improvement	Enhanced community organization and confidence
Environmental improvement	Conservation of water resources and improved environmental quality

Purposes

Project purposes relate to how well the installations or services are functioning and,most often, include achievement of technical standards. Examples are:

Purposes	Examples
Water supply	Water quality and quantity standards
Sanitation	Sanitation-unit performance
Solid-waste disposal	Performance of collection services
Drainage-basin conservation	Water-resource protection and development
Hygiene education	Long-term personal and household hygiene improvements
Institutional development	Demonstratedorganizational capabilities by the community committee or local authority
Financial management	Achievement of revenue recovery and financial viability
Human resources development	Performance of trained staff or community members

Results/outputs

The outputs are the facilities installed, services provided and institutional development achieved by the project. Examples are:

Outputs	Examples
Facilities	Numberofhouseholds served, hand-pumps installed, sanitationunits built
Financial management	Accounting systems installed
Education	Hygiene-education programmes established and operating
Human resources development	Community members and project staff trained
Community development	Community-level organizations (e.g., water committees) formed

Inputs

Inputs are quite straightforward: they consist of everything that is put to use in the project. They are financial resources, human resources (including professionals and semi-skilled and unskilled labour), materials (land, construction materials, physical facilities), engineering (plans and designs) and project management (including financial management and institutional/community development).

Activities

Activities use project inputs to achieve project results or outputs. They fall into various categories or groupings, such as: (a) project preparation and planning; (b) institutional development; (c) construction; (d) operation and maintenance; and (e) project administration.Examples are:

Activity	Examples
Project preparation and planning	Drafting of project document establishing institutional framework for execution; and and defining of work programme
Institutional development	Conducting training courses or in-service training; improving personnel policy and career advancement within agency
Construction	Construction of facilities and auxiliary structures
Operation and maintenance	Controlling chlorine levels in water supply; repairing hand pumps
Project administration	Allocating physical and human resources commensurate with the requirements to meet delivery

- Did it have any effects that were not anticipated, and why?

- From the impacts observed, what additional activities should be carried out as remedial measures or as additional inputs to enhance the project?

- What lessons can be learned, to improve other projects, from the results of the evaluation?

Positive/negative impacts. Impacts can be both positive and negative; the evaluator should look at both benefits and costs of a project. As an example, increased water supply can reduce skin and eye infections. However, it can also result in excess wastewater in and around the town, and pools of stagnant wastewater are breeding places for the Culex mosquito which carries filariasis. Thus, increased water supply can, in some cases, actually worsen public health. Similarly, water, sanitation and solid-waste management projects are usually beneficial to the environment, but there can be negative side (external) effects which must be detected and measured. Although environmental impacts are often relegated to second priority, such side-effects as toxic leachates from refuse landfill sites and groundwater pollution from on-site sanitation can be very important.

Unforeseen impacts. New water supplies may create income-earning opportunities which did not previously exist and, thus, can result in improved nutrition if used for home gardening. Yet if tariff schemes are not carefully designed, they could widen the gap between rich and poor, since the poor may not be able to afford adequate supplies.

Effectiveness and efficiency. Information on project efficiency and effectiveness is important to impact evaluation. It cannot be assumed that facilities were properly installed and are being well used and maintained. Knowing how effective the project has been is essential to understanding why the various impacts occurred. Thus, although impact evaluation is primarily concerned with achievement of goals, it begins with determining whether outputs and purposes were achieved.

III. PRINCIPLES OF IMPACT EVALUATION

A. *Scope of the evaluation*

The scope of the evaluation depends on what the evaluation is for and what resources are available to carry it out. The broadest impact evaluations are those used by project planners, managers and policy-makers. These cover economic, social, health, community, environmental, and, sometimes, political impacts.

There are many factors which determine the size and cost of impact-evaluation exercises. For example, the number of communities and families to be surveyed depends on the expected variations in impacts within the population and the degree of accuracy (level of confidence in the conclusions) required by those using the results. In-depth evaluations are usually carried out by research institutions; owing to budget limitations, these tend to concentrate on one kind of impact. To the researcher they provide a valuable source of information about one area of interest, but to the planner they are not broad enough for policy decisions.

Evaluations can be carried out by project staff as the project progresses. These focus on efficiency and effectiveness and later, once the project is completed, concentrate on its impact. This "internal" kind of evaluation is economical but cannot avoid being subject to bias: using non-project personnel goes a long way to avoiding bias. Impact evaluations are customarily "external" and are carried out by professionals who have had little or no previous link with project personnel. Inevitably, however, outside evaluators need field support which only the implementing agency can provide: most impact evaluations are, therefore, mixed (external and internal) and are, thus, able to keep costs and bias to a minimum.

The sectors

These guidelines are concerned with the evaluation of impactsof water-supply, sanitation and solid-waste management. They are intended for use in evaluating projects in these sectors in low-income urban settlements and rural towns of developing countries.

Water supply includes improvements in water quality and quantity made accessible to the user through wells,hand-pumps, standpipes, yard taps and house connections.

Sanitation refers to improvements in excreta and wastewater collection and disposal. These include on-site facilities, suchas ventilated improved pit latrines and pour-flush toilets, and off-site centralized collection systems, such as small-bore sewers and shallow-sewer systems, followed by treatment by stabilization ponds.

Solid-waste management pertains to collection, transport, off- site disposal and, sometimes, reuse of household, commercial and industrial (non-hazardous) wastes in a safe and hygienic manner. This is carried out by the municipality or community, often with the co- operation of residents in depositing household refuse at designated sites in the project area.

Community development is inherent as a parallel activity in all the above interventions. Specifically, community development arises from the creation of community committees, involvement of the community in all stages of the project, community-level training and education, and confidence gained by the community through successful participation in the project.

The impacts

The impacts considered by the evaluation define its scope.

Economic. Although not specifically targeted for economic development, improved water-supply, sanitation and solid-waste management does lead to economic changes in the beneficiary community. These changes vary from increased crop/livestock production to introduction of other income-generating activities, such as brickmaking, brewing, cloth-dyeing and ice-making. Project outputs, such as water and compost, may be sold to produce a direct income-generating benefit from the project.

Social. Although project goals are seldom specifically social in nature, social impacts are inevitable in water, sanitation and solid- waste management projects. Alleviation of the drudgery of daily water cartage is an important benefit to the family, yet difficult to quantify and measure. Other similar social benefits are common in such projects but are seldom checked, since projects are usually oriented to economically productive sectors. The use of communal taps has an impact on communications amongst women which may even affect the relationships between social groupings and economic strata within the community.

Health. The impacts of water, sanitation and solid-waste management are age-specific. Young children (particularly at the age of weaning) are most susceptible to diarrhoea and related infections: for this reason, most health-impact evaluations focus on diarrhoeal morbidity and nutritional status of children under five. The diarrhoea-causing pathogens of interest to impact evaluations are *Escherichia coli* (ETEC), rotavirus, *Campylobacter jejuni* and *Shigella*. There are several other non-diarrhoeal diseases which can be influenced by water and sanitation interventions, including ascariasis, hookworm, schistosomiasis, filariasis and guinea worm, to mention a few.

Community. Water-supply, sanitation and solid-waste management projects are often the first development projects undertaken by communities. With increased organizational abilities and self- confidence, the community can go on to other developmental activities, such as improving roads or promoting electricity supply or agricultural credit. The initial project acts as a "bridgehead" for subsequent spin-off development. The transfer of technology, whether represented by television, hand-pumps or pour-flush toilets, to a low- income community is bound to have impacts. For example, the project may train individuals in the community in maintenance, book-keeping and building skills. These can have a long-term beneficial impact on low-income communities, many of which have no access to technical education.

Environmental. Water-supply, sanitation and solid-waste-disposal projects have an impact on the environment, and the ultimate decision on whether the ecological consequences are worth risking for the benefits gained is a subjective value-judgement. However, there should, wherever possible, be an attempt to measure the damages in relation to the benefits in such a way that a social choice can be made about the relative worth of a project. Unfortunately, environmental issues are interdisciplinary, interactive, biological and probabilistic, and, because information is always deficient, the outcome has a considerable degree of uncertainty or risk. Compounding the lack of certainty is the conflict between development and conservation, which can never be absolutely reconciled and which, in many cases, leads to inaction and oversight on environmental matters. Given these limitations, it is prudent to analyse most "environmental" impacts in terms of their ultimate consequences and reclassify them under other appropriate impact categories. Therefore, the pollution of groundwaters by nitrates from on-site sanitation or sanitary landfills, for example, is not considered an environmental impact but, since its consequences are health-related, is considered under health impacts.

B. *Prerequisities for evaluation*

Objectivity

Objectivity is the guiding principle of evaluation, but biases introduced by poor selection of the evaluation team, by collection of information from inappropriate sources and by improper analysis of data are very common. Biswas (1981) provides some examples of how bias creeps in:

(a) Obtaining information from the rural elite and missing the people for whom the facilities were intended;

(b) Making contact with men but not women;

(c) Relying on data provided by senior project staff and in reports, without questioning the junior staff in the field;

(d) Questioning only the users of the facilities but not those who are not using them;

(e) Focusing on the technology while neglecting the social aspects;

(f) Extrapolating trends over the entire project area on the basis of a glimpse of one specific point and period in time.

Elimination of confounding variables

Confounding variables are those that are outside the project intervention but cause similar impacts. For example, a study of the impact of a sanitation project on a community's organizational ability could confound the project's impact with that of a road-building project which also required the community to form a project team and committee. Confounding variables must be avoided or, at least, be known and separated from the project-intervention's impacts before conclusions are drawn.

Unfortunately, there are usually many confounding variables which are stumbling blocks to the vast majority of impact evaluations. It is often simply not possible to eliminate or, even, to quantify their effects. Nevertheless, it is essential to know what the confounding variables are, design the evaluation with them in mind and measure them throughout the evaluation.

Baseline data

Project evaluations have a long history of being beset by constraints, owing to lack of baseline data. Evaluations should be planned at the beginning of projects, so that the same kind and quality of data are collected at the beginning (baseline) as at the end (post-project). All too often, the team of evaluators is brought in to evaluate a project, only to find little or no information on what the area was like before the project intervention, and having to rely on guesswork and hearsay. Cross-sectional comparisons between communities with and without facilities are possible and one way of overcoming lack of baseline data: however, selecting communities which are truly comparable is extremely difficult. Baseline information is always desirable and, at most times, essential to impact evaluation: it is needed not only to cover the impact variables themselves but also to cover confounding variables which may change over time (such as educational levels), giving false readings of the impact. Good baseline data are relatively easy to collect at the beginning of a project, if the evaluation is well planned at the start, but very difficult to generate once the project is under way.

Statistical analysis of data

The reliability of conclusions drawn by an evaluation on the significance of observed impacts can only be accurately determined by statistical analysis. Unfortunately, the vast majority of evaluations are performed without statistical analysis, and the evaluator has to rely on "educated guesses" as to how important the impacts were and what they were caused by. For example, conclusions drawn on the impact of improved refuse disposal in only two communities (one with improvements and one without) are highly suspect, because of the small sample size and minimal statistical basis of comparison. Only where the response to the intervention is similar in each population can small sample sizes be used. Where there are big variations in response to a given intervention, large sample sizes are required to achieve an adequate level of significance for the conclusions drawn by the evaluation. In all cases, statistical analysis of data is highly desirable.

C. *Evaluation variables and sampling*

In designing an evaluation there are several considerations which must be taken into account to establish whether the project was the cause of effects being measured. The evaluator must select certain effects to be measured and must attempt to show that there is a high possibility that these effects or impacts resulted from the installations made and services provided by the project.

Independent variables

The independent variable is normally the intervention made by the project. Comparisons are made with populations which have not been affected by the project, and these "with and without" project situations can be considered as the two levels of the independent variable. The "without" project situation can also be considered as a control: thus, the population which has not been affected by the project and which is used as a comparison to the population that has been affected is termed the control group. Independent variables can be both qualitative and quantitative in nature: "with and without" situations are qualitative, whereas a several-levels service (such as the amount of water supplied) is a quantifiable independent variable. Drawing conclusions about an independent variable would be relatively easy, if the population was only affected by the independent variable under study. Unfortunately, in evaluation work, this is never the case: there are numerous "irrelevant" variables which also affect the way a population behaves. Although it is not necessary to know why these variables influence the situation (making them irrelevant to the work), it is necessary to know to what extent they interfere with or confound the effects of the independent variable under study.

Confoundment of effects

In designing an evaluation, it is important to minimize the confounding effect which an irrelevant variable has on the variable of interest. For example, the incidence of disease may be hypothesized to be influenced by the level of water supply. Baseline measurements can be made and compared with post-project disease incidence levels. However, it may be later noted that the communities have made considerable economic gains, resulting in substantial improvements in housing and education in the period since the project began.

Economic status, housing and education are all known to have influence on the incidence of disease, and, in essence, these "irrelevant" variables have confounded effects of the water-supply variable under study. In this situation, it is not possible to separate these confounded effects.

Error variance

The other concern about irrelevant variables is the degree to which they cause "error variance". There will always be some variance in data obtained from individuals within a population, even in seemingly identical circumstances. This variability is attributed to differences in the individuals themselves, such as age, education level and social status. The error variance is that which cannot be attributed to the independent variable under study. Where error variance is large, it is difficult to determine the effects of the independent variable with any reasonable degree of certainty (significance), so every precaution must be taken against increasing error variance, by standardizing the way data are collected, by selecting independent variables carefully and by increasing the size of samples so that any observed effects are statistically significant.

Randomization

To some extent, randomization avoids systematic bias introduced by irrelevant variables, if it is possible to select randomly which individuals are to be subject to which level of the independent variable. Unfortunately, this is rarely the case in evaluating water-supply and waste-disposal projects, because communities are preselected on the basis of other criteria, such as the need for service or their ability to afford and maintain the infrastructure. Thus, it is simply not possible randomly to assign communities or individuals to "with" and "without" project situations. However, it is possible to use randomization to advantage in another way. Randomization is used, for ex-

Table 1 Required sample sizes (after Briscoe and others, 1986)

Frequency of disease in the unserved population (percentage)	Reduction in frequency to be detected					
	10	20	30	33	40	50
0.2	1600 000	380 000	160 000	130 000	85 000	50 000
1	320 000	76 000	32 000	25 000	17 000	10 000
5	62 000	15 000	6 000	5 000	3 200	2 000
10	29 000	7 000	3 000	2 400	1 500	950
25	10 000	2 400	1 000	800	550	330

same point in time. For instance, a selected group of villages A, which implement refuse collection, may be compared to control villages B, without refuse collection (see figure 2). Any differences in the dependent variables measured (OVIs) are attributed to the project intervention-in this case, refuse collection. However, irrelevant variables may confound the effects; the villages A and B may not be similar in all respects apart from the intervention. Likewise, the error variance associated with the dependent variable(s) may be large, and the effect of the intervention "hidden"

Figure 2 Schematic characterization of types of project evaluation

Type of evaluation	Time scale	
	Before project intervention	After project intervention
Cross-sectional	No measurements	Community A (with intervention) Control community B (without intervention)
Longitudinal	Community A (without intervention)	Community A (after intervention)
Combined	Community A (without intervention)	Community A (after intervention)
	Control Community B (without intervention)	Control community B (without intervention)

ample, in identifying which households are to be interviewed in a given community: in this way, systematic bias, which may unknowingly creep in as a result of differences in income level or geographic location within the community, is limited.

The dependent variable (or the objectively verifiable indicator)

The dependent variable is that which is measured to determine the effect of the independent variable. Choosing the right dependent variable or OVI is extremely important to the evaluation: ideally, the OVI should be reliable and sensitive to the independent variable under study. Cost is another consideration: it should be possible to gather information on the OVI at reasonable (if not lowest) cost. All else being equal, the OVI which is easiest to obtain is chosen: however, "all else" is seldom equal, and trade-offs between sensitivity, reliability, cost and ease of measurement are inevitable.

Selecting a sample

It is seldom possible or necessary to collect data from the entire population in a project area. Samples are selected to represent the population at large, so that generalizations (as to the effects of the independent variables) can be made. In the "with" and "without" evaluation, independent-variable sample communities are selected to cover the range of conditions within the project area and ensure, to the greatest extent possible, that the samples are representative of the "with project" population. In selecting sample communities which are "without project", those communities which are similar (in ways affecting dependent variables) to the project communities should be identified.

The sample size is another very important consideration. An unnecessarily large sample should be avoided, but at the same time it must be ensured that the sample is large enough to be able to prove, with reasonable certainty, that the effects resulting from the project within the selected sample can be generalized to apply to the population at large. The decision on how large the sample will be is concerned with the "power". Power is the probability of correctly concluding that differences in the effects of the levels of the independent variable exist. The required power depends on the size of the effect being detected and the expected error variance: it also depends on the acceptable risk of making a mistake in concluding that an effect is taking place.

Reference is made to Briscoe and others (1986) in which sample sizes for health-impact evaluations are discussed. The example is given of an evaluation which is to be carried out to measure the impact on the incidence of diarrhoeal disease in children under five years of age. If the average number of diarrhoeal attacks per year is 2.2, and diarrhoeal-incidence data are based on a 48-hour recall period, it would be expected that the frequency of positive answers to the question, "Has your child had an attack of diarrhoea within the past 48 hours?", would be 1.2 per cent. In this case, the required sample size to detect a 33 percent reduction in diarrhoeal incidence would be larger than 20,000. There would then be a 90 per cent chance of detecting the required reduction level at a 5 per cent significance level. A required sample size of 20,000 is alarmingly large, but it comes from both the kind of experimental design being used ("with" and "without" type) and the very low level of positive responses expected (1.2 per cent). This is common to most health-impact evaluation studies and has plagued researchers in this specific field over the past three decades. Briscoe and others (1986) recommend another kind of experimental design to overcome these difficulties, which is discussed later.

For impacts other than health, the expected positive response level is very much higher than 1.2 per cent. For example, well above 70 per cent of the project's families might enjoy substantial reductions in time spent in collecting water as a result of the project bringing water to within 200 metres of houses through standpipes or hand-pumps. A similar percentage would be expected to be positively affected by installation of latrines. Table 1 illustrates the effect of increased frequency of disease and the reduction in frequency to be detected. Parallels exist between these parameters and, say, frequency of families spending three hours daily collecting water and the required reduction in frequency to be detected. The latter should be well above 50 per cent for a rural water-supply project to be considered successful. As the frequencies of the levels of the dependent variable rise to 25 per cent and above and as the reduction in frequency to be detected reaches 50 per cent, the sample numbers quickly reduce to practical levels.

D. *Experimental design*

The experimental design of the evaluation has a strong bearing on its eventual worth. In particular, it establishes the statistical significance of the evaluation's conclusions. Unfortunately, practical considerations inevitably limit the choice of experimental design to considerations of variables and sample sizes. In theory, there are numerous designs to select from, but, in practice, (owing to the nature of evaluation of projects) the choice is narrowed to one or two.

Cross-sectional design

In cross-sectional evaluations, the "with and without" project sample groups are compared at the

This low-cost simple experimental design can also be termed a one-on-one comparison, as the differences between groups of villages are attributed to the intervention. Single cross-sectional evaluation has the advantage of producing results quickly: however, apart from the most obvious impacts, the conclusions drawn from such an evaluation design are often inconclusive. It is unfortunate that many evaluations are forced into accepting this type of experimental design: lack of reliable baseline data provides little alternative to the evaluator but to opt for it.

Longitudinal evaluations

As the name suggests, longitudinal evaluations are carried out over time, usually beginning with baseline-data collection and concluding with post-project surveys. There are two basic types - experimental and non-experimental - differing in the way they control irrelevant confounding variables. The experimental design controls them by choosing comparison or control groups which are similar to those within the project, except that they are unaffected by the project. Non-experimental designs control these variables by statistical means. This technique is used in the "case-control" method described for health-impact evaluation by Briscoe and others. (1986): the case-control method is used where the incidence of disease is very small.

The experimental design can either be "true" or "quasi" in nature. Where individuals can be randomly assigned to specific levels of the independent variable (for example,treatments), the true experimental design is possible. This is often feasible in controlled environments, such as in the laboratory; however, it is never possible in evaluations of projects in the field. The evaluation is commonly relegated to the quasi-experimental design which is historical or retrospective in nature.

Where data are collected periodically throughout the project, it is termed a time-series evaluation. The steps are:

(a) Prior to project intervention-pre-project baseline-data collection;

(b) Periodically during implementation-ongoing or mid-term evaluation;

(c) Upon completion-terminal or end-of-project evaluation; and

(d) Some time (maybe, several years) after completion-post-project evaluation, when the interventions are expected to have their full development, and the full impacts of the project occur.

In each evaluation, the data are compared with those collected in the baseline study before project implementation and with any previous data. The purpose is not only to compare situations at two points in time but also to observe the trends over time.

Ongoing evaluation assesses the overall performance of the project and is the basis for deciding whether project objectives are being met. It is a management-oriented evaluation, identifying undesirable unexpected side-effects, as the project proceeds. Terminal and post-project evaluations assess the overall achievements and impacts of the project.

Cross-sectional and longitudinal designs are often combined. Here, comparisons are made before and after the intervention (longitudinal) and between samples (cross-sectional) at both points in time. The advantage of having data from the control through time is that it is possible to identify any variation in the dependent variables in the control not due to the project intervention. Thus, the impact of the project on the sample "with" project can be clearly established as being caused by the project intervention.

Table 2 shows how the different types of evaluation relate to the various stages of project development. In some cases, a second post- project evaluation is performed, some time after corrective measures (as suggested by previous evaluations) have been taken. Longitudinal evaluation is considered to be more reliable than cross-sectional evaluation. However, its disadvantages include high costs and difficulty in maintaining long-term interest and continuity in the evaluation process; in many areas in developing countries, the turn- over rate of staff is high, making evaluation difficult. As with the cross-sectional evaluation, the longitudinal evaluation suffers from having limited control over confounding variables.

Table 2 Evaluation in project-development stages

Stages	Project activities	Evaluation activity	Evaluation focus
Identification	Awareness of needs Assignment of planning responsibilities Sector-analysis and project-identification report		
Preparation	Feasibility report Appraisal	Baseline-data collection	
Approval	Investment decision		
Implementation	Construction of facilities Supporting activities Monitoring Project progress report	Ongoing evaluation	Efficiency and effectiveness
Operation	Operation and maintenance of facilities Continuous provision of service Monitoring Inspection and audit End-of-project report	Ongoing evaluation	Efficiency and effectiveness
Evaluation	Evaluation report	Completion evaluation	Effectiveness (and impact)
	Evaluation report	Ex-post evaluation	Impact

Note: Underlined activities are related to evaluation.

IV. OBJECTIVELY VERIFIABLE INDICATORS (OVI)

After consideration of the framework for evaluation and the history, principles and design of impact evaluation, it is necessary to turn to means of measurement - the objectively verifiable indicators (OVIs). These are measurements of dependent variables which reveal whether the project's impact objectives have been achieved. They must accurately measure the impact without - bias or subjective judegement: as far as possible, they must be quantifiable (this is important to the analaysis of large quantities of data collected in the field) and they must be verifiable, that is, they must yield the same conclusions even if different evaluators carry out the evaluation. While there are key indicators that will be used in nearly all evaluations, there is no standard set of indicators nor are there standardized measurement techniques. Thus, questions on community participation that would be asked in one country or region would be phrased in quite a different way in another area, to ensure that what the respodent is answering is what the interviewer meant.

This chapter gives an overview of many kinds of indicators, points out their values, pitfalls and relevance, and makes suggestions as to how they can be incorporated in impact evaluation. The following indicators have been chosen for their relevance to short- term and opportunistic evaluations, and they are intended for use with a minimum of laboratory back-up. Although specialized expertise will always be required in the design and planning of evaluations, the following indicators are recommended for use by field staff, trained in evaluation procedures but without high academic qualifications.

Impact evaluation is still an art not a science, as there is no consensus on research methodology or indicators. Some considerations on OVIs for impact evaluation are worth mentioning here:

(a) Impacts can be positive or negative; the evaluation should include both.

(b) Impacts that are not anticipated during project planning stages frequently occur, and, while focusing on the project's objectives and goals as a basis for evaluation, the evaluator should also strive to reveal any unexpected impacts.

(c) When adopting a set of indicators or parameters, the evaluator should make sure that they are mutually exclusive, in order to avoid 'double counting' (e.g., the OVI "Crop/Livestock production" is used when agricultural products are not sold but, in cases where they are, the OVI would be "Home income-generating activities").

(d) A good OVI should be:

(i) Relevant and valid (it actually measures what it is supposed to measure);

(ii) Sensitive to any changes in the impact being investigated;

(iii) Cost-effective (the results should be worth the resources spent on them);

(iv) Independent (e.g., OVIs for impact evaluation cannot be also used for effectiveness evaluation); and

(v) Timely (i.e., able to provide information within a reasonable time).

OVIs corresponding to common project goals are listed in table 3.

A. *Economic impacts*

Water, sanitation and solid-waste management are the primary purview of women in many developing countries. Accordingly, their lives are far more influenced by these interventions than those of men. Women, as household managers, are the primary waste depositors and water carriers and users, and they are the main decision-makers on how much and what water is used for what purpose, and how household wastes are disposed of. Women have a strong influence over household hygiene in many respects - by the way they bathe and clean their children, enforce hygiene practices (e.g., hand washing before each meal), dispose of the children's excreta, sweep the house, store the water in the house etc. Consequently, women are singled out here for special consideration.

Impacts on women are best measured through interviews by women with women in the household, in the absence of men. The following are impacts in terms of changes caused by the project:

(a) Income, expenditures, savings and management of the project (planning, design, implementation, operation and maintenance);

(b) Creation or strengthening of women's groups/clubs for purposes of their involvement in the project;

(c) Technical skill improvement - the training of women for operating and maintaining the facilities once installed;

(d) Opportunities for income generation - increased availability of water and fertilizer, and ability to sell water and retain earnings.

Table 3. Breakdown of project goals

Economic:	- Time savings in collecting water and time use
	- Crop production through irrigation
	- Home income-generating activity
	- Commercial, manufacturing, industrial, water and waste use
	- Water and waste sales
	- Job creation
	- Resource conservation
	- Realization of tourism potential
Social:	- Drudgery alleviation
	- Creation of recreational opportunities
	- Status-raising
	- Increase in self-esteem
	- Equitable distribution of benefits
Health:	- Reduction in specific measurable diseases
	- Improvements in nutrition
Community	- Community independence, motivation and self-help
	- Organization and management
	- Strengthening of leadership
	- Spin-off developmental activities
Political:	- Community support to government efforts
Environmental:	- Environmental-quality improvements

Economic-impact evaluation is both simple and complex: simple in the sense that many of the economic indicators are readily observed and complex in that their quantified measurement and interpretation can be extremely difficult and complicated. For example, economic benefits of water supply and/or solid-waste disposal may be thoroughly confounded with other variables, such as advances in education and improvements to other infrastructure, such as roads. Wherever possible, evaluation should distinguish between the true economic impacts which occur as a result of project interventions and other economic impacts occurring as a result of other surrounding development events.

Time savings

The most obvious and attractive benefits to users (especially women) in water-supply, sanitation and solid-waste management projects are the time and energy saved in not having to carry water from far away, the convenience of being able to defecate in a facility close to home and the economies of effort in not having to organize their own refuse disposal. Of these three, the time saved in fetching water is the most important and will be discussed in detail. Similar principles apply, however, to improved sanitation and refuse disposal.

Comparisons can be made between water-collection time before and after installation of facilities, but care must be taken to identify the real situation. For example, it may be anticipated that all water will becollected from the new tap, but close observation might reveal that drinking water still comes from the old source, as people prefer its taste. Much can be gained from questioning the women about water collection, but a complete understanding can only be obtained through continuous observation at the water point (care being taken to avoid the observer's presence interfering with water-gathering practices).

All time savings are welcome, as there are several associated benefits:

(a) *Health and nutrition:* less body energy is spent on water collection;

(b) *Project success:* almost without exception, the time savings and convenience offered by a new water system are perceived and appreciated by users more than any other benefits - thus, the greater the savings, the better the chance that the project facilities will be properly operated and maintained and that the users will be willing to pay for the service, making it financially viable;

(c) *Economic:* time saved can be spent on income-generating activities (see below);

(d) *Formal education:* educational opportunities are increased for children who used to collect water instead of going to school;

(e) *Community education:* time is available for women to gather in social groups or to listen to the radio or watch television, facilitating community educational programmes, such as hygiene ducation.

Information on this OVI is most often collected by questionnaire and checked by observation. The results are compared with baseline data gathered during the project preparation stage or with data collected from unaffected control communities. The time and energy spent in water collection in most cases are proportional to the collection distance but may be influenced by other factors, such as topography, climate, queueing and talking at the water site. In some projects, evaluators quantify time savings by calculating the body energy (in terms of calories) spent in water collection. However, accurate assessment of energy saved is unlikely, because many factors are not measurable, making comparisons between villages difficult. For example, differences will usually exist between pathway conditions, topography, microclimates (shaded paths save energy) and water-hauling devices (pushing watercarts requires less energy than carrying water on the head). White and others (1972) give a review of methods of measuring energy expenditure in water collection.

The economic value of time savings may be considered using two approaches, as suggested by Cairncross and others (1980). Briefly, the first approach assigns an implicit value to the time of water collection, to be discounted at a suitable rate of interest over the project's life. Such a calculation enables policy-makers to see time- saving benefits in monetary terms. However, in many cases, this method is not realistic, as unemployment prevails; the actual opportunity cost of the time spent on water collection is close to or equals zero. The second approach is to examine what people do with time saved and place a social and economic value on that for comparison purposes.

Time savings can be negative: in other words, projects can actually increase time taken to collect water, defecate and dispose of refuse properly. Some examples are:

(a) *Increased water consumption.* For a house near a new water source, users might collect more water than before, owing to increased water consumption. In this case, the increased water consumption is interpreted as a positive impact, but the time spent in collecting it is negative.

(b) *Queueing.* This not usually necessary at traditional water sources but may be required at improved sources, such as communal standpipes.

(c) *Seasonal variations.* Often the traditional water source is much further away from the household during the dry season than the wet. As has been observed in United Republic of Tanzania (Heijnen and Conyers, 1971), where new communal water points are located between the dry-season and the wet-season collection points, time is saved during the dry season but not the wet. This demonstrates the need to understand the local situation in detail and to make observations in all relevant seasons.

(d) *New water uses.* Time lost in using more water than before the project cannot be ignored; improved hygiene implies new water uses in the house and longer periods spent in washing of clothes, in personal hygiene and in cleaning of premises.

Crop/livestock production

Increased livestock and crop production, resulting from improved access to water, is likely to provide nutritional benefits. Provided that baseline data are available and accurate observations of water use are possible, changes in crop and livestock patterns are a good impact indicator. Differentiation should be made between the types of crop or livestock and how they benefit the family. Improved sanitation and solid-waste management creates opportunities for resource recovery (compost, fertilizer, biogas and aquaculture). Where such reuse is practised, the benefits of increased food production should be included in the impact evaluation. Changes in agricultural production can be observed through aerial photography.

However, changes at the household level are likely to be most accurately measured by questionnaire and observation. The economic value of increased nutrition resulting from increased agricultural production can then be determined.

Home income-generating activities

There are several activities which can be initiated or enhanced by improved water supply and appropriate resource recovery. Income generation implies products being sold outside the house, e.g., through:

(a) Horticulture (vegetables, flowers);

(b) Livestock production (poultry, dairy cattle, pigs);

(c) Cloth-dyeing;

(d) Brewing.

Commercial water and waste water infrastructure facility uses

The benefits of urban water-supply, wastewater-collection and solid-waste-disposal management are commonly shared by industry and commerce. Improved water supplies will benefit all water-using industries, as may be explicitly stated in the project purposes. Likewise industries that produce both solid and liquid wastes benefit from existing infrastructure that take care of these.

In urban areas, water-supply, sewerage and refuse-disposal infrastructure provision often leads to increased development of industries which require these facilities. The industries in turn provide employment to urban residents and contribute to the economic development of urban centres. While it is possible to recognize this causal relationship, it is often difficult to quantify precisely the extent of industrial activities stimulated exclusively by the availability of services. A myriad of other factors, such as the availability of credit, other services (e.g., electricity) raw materials and local skills, will all have a bearing upon the establishment of industries in any urban area. It is, however, possible to assess whether a given service-pricing policy encourages or discourages industrial activities. The assessment of the impact of pricing policies could serve as a useful guide to amending or revising tariff structures to stimulate or discourage industrial activities. Structured questionnaires and field evaluations of industrial service demands and supplies and the extent to which individual-service facilities are developed by industries themselves serve as indicators of the impact of service-pricing policies.

All these impacts may be anticipated but are difficult to quantify. One approach to use is the price that the beneficiaries are willing to pay for services. This is perhaps the only way impacts can be quantified objectively, although care must be taken to separate out subsidies and other trade-offs.

Sales

All project products and services can, in one way or another, be sold, e.g.:

(a) Water from the project area can be sold to other areas where water is scarce or saline or otherwise undrinkable;

(b) Resources recovered from wastes, such as refuse and sewage, an be sold as fertilizer etc.;

(c) Reclaimed wastewater can be sold for agricultural and industrial purposes;

(d) Energy generated from refuse and sewage treatment can be sold;

(e) Tariffs collected from the community can benefit the community and the agency in paying for operating, maintaining and expanding the water supply.

Quantification of sales is normally made through monetary value of sales made. Although this provides a good basis for before-and-after project comparison, care must be taken to uncover associated impacts, such as the equitable distribution of wealth, institutional and human resources development, and the purposes to which new financial resources are put.

Employment generation

The project will invariably generate employment, whether only in construction jobs or in permanent jobs, such as in the refuse- collection service. Such employment is relatively easy to identify and quantify; more difficult to evaluate are the secondary impacts resulting from raised economic status of those employed (e.g., education and improved housing). However, certain groups may lose their employment as a direct result of a project, including, for example, "sweepers" who may have been earning their living by collecting refuse and excreta: these groups should not be ignored in project design, as the impact of the project on them could be disastrous. These people are the lowest paid and usually lowest in social status; alternative forms of employment are seldom available to them. They can, however, be formally brought into the project and, after training, given alternative employment as treatment-plant operators, refuse collectors and sorters, artisans and building contractors.

Tourism potential

Tourism benefits from reliable clean water supplies both by reduction in costs of individually treated water, which may have been necessary for hotels, and by the fact that tourists place a value on potable water from the tap in their rooms. Certainly, epidemics of water-related diseases have a drastic effect on tourism. The removal of refuse from tourist areas (especially beaches) also enhances tourism.

Service interventions can enhance or reduce tourist potential. The potential use of a dam as a tourist attraction can prove to be of considerable value to the region where it is located: by contrast, the discharge of untreated or inadequately treated wastewaters into watercourses and waterbodies could reduce tourism potential in the concerned areas. The value of such potential is difficult to estimate and can only be approximated through a comparison of the "with" and "without" scenarios, where the relative value of increased tourism is established by comparing areas with the infrastructure services with others that do not possess these.

B. *Social impacts*

Social impacts are inherently intangible in nature and are also complex, being specific to the social setting in which they occur. Their measurement by questionnaire survey, structured conversation and observation is considered in chapter V. There will be many instances where, without lengthy anthropological study, the complete social impact of a project will not be understood, but the social impacts discussed below are those that are considered important enough to be included in relatively short-term impact evaluations attempting to cover project impacts in several representative communities.

Effective and accurate communications are at the heart of good social-impact study. The evaluator must recognize that (in all likelihood) he/she comes from different cultural and educational background from the communities being surveyed. Thus, what is intended to be said by the evaluator may not be what the community understands as being said: likewise, the response may not be understood by evaluator as intended by the community. Personal interpretation of responses must, therefore, be careful: local dialect should be used in communications, and it is desirable, if not essential, that the sociologist in charge of the work has first-hand experience working in communities in the region.

Drudgery alleviation

Drudgery alleviation is closely allied to time savings discussed previously, except that the 'social' side of savings cannot be measured in monetary or energy terms. The drudgery of water collection depends on the distance and conditions of carriage: in some areas, women spend most of the day and almost all of their productive energy carrying water to their families in the dry season.

Measurement of drudgery is difficult, but before-and-after project comparisons are possible. Naturally, responses to questions about feelings of hardship are very subjective and difficult to quantify, but drudgery alleviation is an important impact in many projects and should not be ignored. In most instances, not all of the time saved in collecting water is reallocated to economic uses, but time-saving adds to leisure time, which is in itself beneficial.

Recreational opportunities

Water, sanitation and solid-waste management will have little impact on recreation, apart from making available sites for sports after refuse, wastewater and excreta are removed from the streets. This result is best measured through structured discussion with community leaders and checking by observation.

Social status

Changes in social status as a direct result of the project are very difficult to observe and quantify. Indicators of pride of ownership, such as the decoration of latrines, do not point to changed social status. Changes of social status can occur within a community, owing to inequitable distribution of benefits, and are usually considered as undesirable in water and sanitation projects. Social status is not recommended as an OVI in impact evaluations, unless such impacts are very apparent.

Self-esteem

Changes in self-esteem often result from the introduction of water supply, sanitation and solid-waste disposal interventions. These changes are manifested in terms of changes in attitude which look down upon past practices and can be elicited through questionnaires and structured conversations.

Benefit distribution

Almost without exception, an implicit aim of water, sanitation and solid-waste projects is the equitable distribution of benefits amongst all users/recipients. The study of benefit distribution answers the questions: (a) who gets what and why?; (b) is the distribution equitable?; and (c) are there ways to improve the the distribution? These are not easy questions to answer accurately.

The socio-economic structure of the project area may need to be studied, to identify individuals or groups who have advantages in access to project services, owing to their wealth and social or political positions. Benefit distribution may not be equal even amongst those having the service. Water supply may be available through yard taps to the entire community, yet only those with sizeable plots will be able to use the water for vegetable production.

Once the socio-economic structure and the distribution of access and benefits are determined, the equitability of benefits distribution an be assessed as being consistent with project policy and objectives.

Evaluation of benefit distribution thus involves the three different levels of evaluation, namely: (a) outputs, to evaluate the access; (b) purposes, to evaluate the benefits; and (c) impact, to evaluate the effects of benefit distribution on the population and, if necessary, to determine means of improving benefit distribution.

C. *Health impacts*

The evaluation of health impacts of water-supply, sanitation and solid-waste-management projects is perhaps the most challenging task for the impact evaluator. These are both indirect and direct indicators of health changes.

Indirect hygiene-related indicators

Indirect indicators pertain largely to improvements in hygiene and sanitation in and around the household. They are only indirectly indicative, in that the assumption is made that improvements in them will result in improvements in health. Some examples are:

(a) Personal cleanliness (observations of clothes, hands, fingernails);

(b) Water-quality protection during cartage, storage and use in the house;

(c) Dishwashing practices;

(d) Reuse of dirty water for washing;

(e) Latrine usage and hygienic conditions;

(f) Anal-cleansing practice;

(g) Perceptions of illness and of why people get sick.

These and indicators like them are primarily project-purpose (effectiveness) oriented but need to be included in the impact study, if the evaluation is to reveal why various health impacts occur.

Direct health indicators

Knowledge of the principal water-related and excreta-related diseases in the project area is essential. This can be sought directly from the women of the households, who will provide symptomatic information only, or from local health centres. Unfortunately, there are seldom health centres which have adequate outreach toservethe population, and those that do commonly lack adequate mechanisms for collecting health data, except from patients who attend the clinics. Data that are available tend not to be systematically recorded and are frequently unreliable: for instance, in large projects covering wide areas, different health centres may

have different service-access levels (e.g., owing to each centre's reputation or the local presence of competent traditional healers) and reporting methodology (e.g., some centres include dysentery in the diarrhoeal-disease category while others exclude it). The first problem is how to interpret the existing health records kept at these centres: this can be, in part, solved by observing the way health personnel diagnose patients and comparing it against health records. All health records need to be checked against field observations.

Direct measurement of health impact is commonly made by using the following indicators:

(a) Mortality rates;

(b) Morbidity, incidence rates of diarrhoea and/or dysentery;

(c) Prevalence rates of excretion of enteric pathogens;

(d) Prevalence rates and intensities of intestinal helminthic infections;

(e) Prevalence rates of eye and skin infections.

Mortality rates are only relevant if there is sufficient mortality in the project area caused by water-related and sanitation-related diseases: they are of no use in assessing the impact of solid- waste management. Care must be taken in interpreting existing mortality information, as the cause of death is, as often as not, incorrectly reported, and/or there are other illnesses, such as malnutrition and pneumonia, associated with the death.

The World Health Organization (1981) suggests the use of the "diarrhoea death ratio" as an indicator of diarrhoea prevalence. This is the ratio of deaths caused by diarrhoea in children under five to all deaths in the same under-five age group. Official records for such data are inadequate in many developing countries, so a field-survey technique has been developed for use with this indicator. Experience with this technique has been quite favourable in some cases.

One of the most common indicators is morbidity. Mothers can be asked of their children's episodes of severe attacks of diarrhoea (mild cases are of doubtful public health significance), during regular visits to the household. The data may be unreliable or incomplete, because of reluctance to divulge personal or family informationor of inability to remember accurately episodes of diarrhoea. The optimum frequency of data collection for these purposes is as often as logistically possible, but the maximum return time should be one or two days, although some researchers have found one to two visits per week provide reliable data (Blum and Feachem, 1983). Again, the definition of "diarrhoea" is imprecise and varies from respondent to respondent: a clear definition must be made and tested to ensure that it is common to both interviewer and respondent.

Where there are few constraints on data collection and samples can be regularly taken and accurately analysed for causative pathogens, a number of selected diarrhoeal pathogens should be considered: they should be known to be causes of diarrhoea in specific age groups. The effect of climate/season must also be known. Briscoe and others (1985) suggest that, for a study of children under five years of age over a whole year, those pathogens of emphasis in many countries would be *Escherichia coli* (ETEC), rotavirus, *Campylobacter jejuni and Shigella*. It is noted that the quantificabtion of rotavirus and ETEC requires relatively sophisticated laboratory facilities.

Where it is known that intestinal helminths are a significant health concern, the following can be included as OVIs: *Ascaris lumbricoides, Enterobius vermicularis, Ancylostoma duodenale,* the schistosomes, *Strongyloides stercoralis, Taenia saginata* and *Taenia solium, Trichuris trichuria,* the guinea worm *Dracunculus medinensis* and the filaria *Wuchereria bancrofti.* Their importance to health must be ascertained before their inclusion in the evaluation; again, some of them require moderately sophisticated laboratory equipment and expertise. Even where laboratory facilities are not available, selection of a few organisms which are relevant to the evaluation and easily recognized (at least by simple and low-cost diagnostic methods that can be performed on the spot) should be included in the evaluation programme, if they exist in the project area and are influenced by project interventions. A notable example of these diseases is guinea worm, the prevalence of which can be expected to diminish in endemic areas within a year after water-supply interventions. For this reason, this OVI can be used to distinguish early health impacts of an improved water-supply component from the benefits of other components.

Other indicators that can be considered in resource-constrained evaluations are the prevalence of roundworm and hookworm. These two parasites are strongly related to sanitation and hygiene-education interventions.Their eggs can be readily recognized under a microscope by a technician with little training, and stool examinations for these helminth eggs do not require complex equipment and, hence, are best performed on-site.

A majority of the indicators so far discussed aid in the verification of positive health impacts of water supply, sanitation and solid-waste disposal interventions. Some negative health impacts also result in certain cases. In fact it is not uncommon for service interventions to demonstrate a shift in the predominant disease patterns of a community over a period of time. While some of the negative health impacts directly affect the community benefiting from the intervention, others affect communities not directly served under the intervention. For example, there are three principal diseases which can be exacerbated by water-supply projects. These are schistosomiasis, malaria and filariasis. All depend on the creation of breeding sites, increasing the vectors which transmit the diseases. Increased schistosomiasis (bilharziasis) and malaria result from creation of open reservoir storage behind dams and open channels for water conveyance, and are therefore likely to affect riparian communities. The vectors responsible for these diseases breed in clean waters. Bancroftian filariasis, however, can result from increased surface areas of polluted water which serve as breeding sites for the mosquito vector and is likely to affect the beneficiary community directly. Such sites result from the accumulation of sullage, sewage and in some cases poorly constructed sanitation facilities such as septic tanks, pit latrines and aqua privies. Prevalence and intensity changes in these diseases can form the basis of their OVIs. Laboratory facilities will be required for diagnosis and analysis of blood and stool specimens.

Mosquitoes, flies and rodents are often associated with on-site sanitation and solid-waste disposal interventions. Changes in the prevalence and intensity of diseases that they transmit may therefore result, following service interventions. It is, however, difficult to estimate the variations resulting from an intervention and to establish the economic benefit of the intervention in terms of the reduced cost in controlling their numbers and in treating the diseases they spread.

While most of the health impacts so far discussed relate to biological-disease-transmitting agents, certain interventions could also cause diseases through chemical agents. An example of this is the increased prevalence of methemoglobinemia ("blue body syndrome") that could result from nitrate pollution of groundwater sources used for potable supplies following the provision of on-site sanitation and sanitary-landfill facilities. Prevalence and intensity changes in the disease will once again form the basis of its OVI. Other phenomena such as the increased prevalence of respiratory diseases resulting from air-quality deterioration following the introduction of incinerators to dispose of solid wastes are more difficult to establish and evaluate. As before, it is however, difficult to estimate the increased cost of controlling and treating the diseases caused by chemical agents.

Nutritional status and anthropometry

Nutritional status of children in a certain age range (e.g., four years and below), using standard anthropometric measurement techniques, has potential as a reliable OVI in developing countries. The height and weight of children are measured, individual ages are noted, and then the data are compared against the standards applicable in the country (which may be available in the form of a growth chart). The following indicators have been frequently used United Nations Administrative Committee on Co-ordination Task Force, 1984):

(a) Weight-for-age - the simplest and most common nutritional indicator;

(b) Height-for-age-a specialized measurement of stunting (prolonged or temporary nutritional deprivation, or insufficient food intake); and

(c) Weight-for-height-the accepted measure of wasting (i.e., sufficient food intake but impaired food efficiency, as in cases of helminthiases) when used in conjunction with height-for-age.

The weight-for-height and height-for-age indicators help distinguish between chronic malnutrition (where stunting is present) and acute malnutrition (where wasting is present).

Anthropometry (anthro: human; metry: measurement - the practice of measuring and weighing the proportions or parts of the human body) has been used by USAID to evaluate water-supply projects in a number of countries. A less sophisticated technique than anthropometry is mid-arm measurement using a three-colour tape wrapped around the mid- arm of children. The colour calibration gives a quick indication of the nutritional status of the child measured: green means good, yellow is bad, and red means very poor nutritional conditions. The basis for using children's nutritional status as an OVI is that the quality and quantity of water and sanitation service directly affect the incidence of water-related diseases which in turn have a direct effect on the nutritional status of children. The cause-and-effect chain events are most meaningful in childhood (for details on the interactions between diseaseand nutrition,see the review by Rosenberg and others (1976)). Furthermore, nutritional status serves as an indicator of far- reaching impact. Even so, it is nonetheless extremely difficult to weed out the effect of confounding variables, such as improved economic status, which are likely to have an equal, if not more, pronounced impact on nutritional status. Increased water supply may lead to increased vegetable gardening and livestock raising which change the nutritional diet.

Since the use of children's nutritional status as an OVI for evaluating water-supply and sanitation projects is a relatively new idea, its potential merits for developing-country application include (Magnani and others, 1984; Struba and Isely, 1981):

(a) There is recent evidence that anthropometric indicators are as responsive to the interventions of water-supply and sanitation provision as is diarrhoea;

(b) As anthropometric values are continuous (compared with discrete predetermined ranges of values as in a questionnaire), they are well suited to statistical analyses;

(c) Anthropometry gives quite accurate results and, hence, is considered less prone to bias than other data-gathering techniques;

(d) Anthropometric values are easily understood by non- technicians;

(e) The technique is simple, relatively low-cost, requires a minimum of equipment for fieldwork and does not require skill: large amounts of data can be collected quickly at the local school or any place where children gather.

Drake and others (1980) identified some limitation in using anthropometry as an indicator of nutritional status. However, in general, the advantages of anthropometry far outweigh its limitations.

D. *Community impacts*

Water, sanitation and solid-waste management projects are inherently developmental in nature. To many low-income communities, improved water supplies are first prority; often, these projects are the first requiring community-wide participation. The formation of a committee which has a project focus and the committee's ultimate success in seeing the project through have long-range developmental impacts, sometimes larger than the water-supply project itself.

There are wide variations between communities in terms of their organization and hierarchy. Cross-sectional evaluations which require control (non-intervention) communities to be similar to those within the project area are unlikely to succeed in separating out these differences. Baseline studies must collect data on institutions, organizations, communications and hierarchy within the community prior to project inception. This information is best obtained through structured discussions with leaders, individuals working in the community, such as teachers, and a representative group of householders. This baseline information will be compared withdata collected during the impact evaluation at project completion and several years later, to reveal institutional development within the community and community response to the project's promotion, training and organizational inputs. These are primarily purpose or effectiveness measurements, and they are the background for explanations of why impact occurred. Impact is measured through historical analysis of initiatives taken during the project and of new ventures thereafter. All of the data required are not quantifiable or analysable using coded questionnaires and statistical analysis: nonetheless, development impact is very important and should not be omitted from the evaluation.

Some "spin-off" developmental activities may result from the project. Having succeeded in its first project, a community might gain self-confidence and go on to other initiatives, such as improving roads, community meeting halls, schools and recreational areas. Project committees might also become core groups which can lobby government for electricity, land tenure and political autonomy. The key is that a successful water, sanitation and solid-waste management project does not stop with provision of services alone: it builds organizational ability and confidence within the community. These are basic ingredients for spin-off developmental activities later on. Community impact is, thereby, an integral part of the outcome of the project and its impact evaluation.

One spin-off development which often follows basic-service interventions in low-income squatter settlements, for example, is the eneration of housing improvements which result from the increased security provided to residents. The provision of water-supply, sanitation and refuse-disposal facilities by governmental agencies is often seen by residents as a formal recognition of their plot and house, and thus motivates them in making capital and labour investments: in fact, the provision of basic services is often used as a principal strategy in upgrading low-income areas. Experience in Brazil and Sri Lanka has demonstrated that, contrary to common belief, low-income communities are often willing to pay the full economic cost of essential services as a means of meeting basic needs and increasing their security of tenure. While it is possible to identify such spin- off benefits of basic-service interventions, through the application of questionnaires and structured discussions with the concerned communities, it is difficult to measure, for example, the increased investments released for shelter improvement as a consequence of the increased security that results from the provision of basic services.

In urban areas, water-supply, sewerage and refuse-disposal infrastructure provision often leads to increased development of industries which require these facilities. The industries in turn provide employment to urban residents and contribute to the economic development of urban centres. While it is possible to recognize this causal relationship, it is often difficult to quantify precisely the extent of industrial activities stimulated exclusively by the availability of services. A myriad of other factors, such as the availability of credit, other services (e.g., electricity) raw materials and local skills, will all have a bearing upon the establishment of industries in any urban area. It is, however, possible to assess whether a given service-pricing policy encourages or discourages industrial activities. The assessment of the impact of pricing policies could serve as a useful guide to amending or revising tariff structures to stimulate or discourage industrial activities. Structured questionnaires and field evaluations of industrial service demands and supplies and the extent to which individual-service facilities are developed by industries themselves serve as indicators of the impact of service-pricing policies.

E. *Political impacts*

Political motives often determine priorities in selecting areas to be provided with water-supply, sanitation and refuse-disposal facilities. It is intuitively believed that the introduction of these services to a community will increase political support to the party in power and to other government endeavours. While this is true, in general, it is however, extremely difficult to dissociate the impact of other confounding variables from those directly attributable to the intervention. An attempt to assess the extent of such support resulting from service interventions can be made through questionnaires and structured conversations which monitor political support over a period of time and enquire into some of the fundamental reasons behind the changes. Even having established the magnitude of political support directly attributable to service interventions the assigning of values to such changes in support will nonetheless be subjective and will vary from one area to the next and between cities depending on a myriad of factors such as the perceived need for the intervention, political support generating potential of the best alternative investment (for example in agriculture) etc.

F. *Environmental impacts*

Water-supply, sanitation and solid-waste management project impacts are closely linked to environmental improvements. Two environmental impacts warranting description here include aesthetic quality and desertification.

Aesthetic quality. Any disruption of natural conditions during or as a consequence of an intervention will have a resultant visual impact. Although subjective, such impacts could be assessed through appropriately designed questionnaires and structured discussions. Such assessments could also be supported by detailed laboratory analysis demonstrating changes in ecosystems as a function of the intensity and variety of a range of species. For example, the change in the ecosystem within a water body that results from the discharge of wastewaters into it could be assessed by classifying water quality according to animal and plant life found in it.

Desertification. Defoliation, deforestation and consequent desertification is a serious concern in arid and semi-arid regions where water supplies (wells) attract excessive numbers of nomads and livestock. This results in overgrazing and eventual desertification of the area. It can be measured by serial aerial photography. However, it is seldom a concern of urban and rural towns water supply development in settlement.

V. CONDUCTING THE EVALUATION

A. *Evaluation planning*

The importance of planning the evaluation during the early stages of project planning was stressed in chapter I. The ultimate use of the information collected in the field dictates the scope of the evaluation and its "depth". Evaluating a project for its impact on small business would require a very different LEF from one, say, inquiring into the project's impact on diarrhoeal disease. The project's objectives are clearly established and agreed upon by all concerned agencies long before implementation. Impact evaluation relates to achievement of the project's goal, but, naturally, this goal has to be broken down into various components during the evaluation planning. In this case, there are economic, social and other goals, all broken down into elements, until each element is sufficiently detailed to be separately assessed in the field. Subgoals and sub-subgoals should be clearly agreed upon by all concerned early in project planning, so that there is a clear and common understanding of exactly what is expected of the project, i.e., what goals it will be expected to achieve in the long run.

At the same time, it is important to know what the expected purposes and outputs are. Although this is an impact evaluation, it is essential to have an understanding of how many facilities were installed and if they are being maintained and used properly. Thus, a minimum of efficiency and effectiveness evaluation is carried out in parallel with impact evaluation. This provides a broad understanding of the reasons why the various impacts are (or are not) occurring. Thus, from the outset, details of the evaluation plan are established and agreed upon. This is a far cry from the usual situation where evaluators are called in to evaluate projects long after their completion, without the project's objectives having been properly defined and with no baseline data having been obtained at the beginning of the project.

B. *Baseline data*

Lack of baseline data endangers the validity of impact evaluation. The evaluation team is faced with not knowing the conditions in the communities prior to the project and having to hypothesize or generate data from neighbouring communities outside the project area. By planning the evaluation in detail during the project's early preparation stages and getting baseline data immediately after approval (before implementation), it is ensured that the data collected will be relevant to the eventual post-project evaluation to be carried out several years later. This is because the OVIs and their methods of measurement in the field are clearly established in the evaluation plan. The appropriate timing is illustrated in figure 2. Baseline data can most easily be collected during the social and technical surveys required for the detailed design.

C. *The evaluation team*

Whereas it may, at first, seem logical to use the project team to carry out the entire evaluation, there are several reasons why external evaluators are preferable. Well-qualified professionals from outside the project bring new ideas and expertise to the project, and this is particularly valuable in identifying how the project can be improved. Often differing views on how the project should have been carried out are held by various project staff members, but external professionals are able to comment on the project without jeopardizing their postion and can, thus, maintain a degree of impartiality. Inevitably, the project team has vested interests in seeing that the project gets a favourable evaluation, so the external evaluator is brought in specifically because he/she is impartial. Finally, having been immersed in the project for so long, the project team is likely not "to be able to see the wood for the trees", but the external evaluator is able to stand back from the project and acquire an overview. It is for these reasons that donor agencies commonly require external evaluators to be involved in evaluation, especially when the evaluation is going to be used to support an application for further funding.

External evaluators, however, cannot work alone. It is extremely difficult, for example, to work in low-income communities without guidance from the project team. When appropriate, members of the project team could assist in the evaluation, by providing logistic support, and, in some instances, could actually participate in the evaluation. This is feasible where the efficiency and impartiality of the evaluation are not jeopardized by the presence of the project team in the field. For example, where relationships between the project team and the communities have failed, the presence of the project team in the field with the evaluators could prove disastrous to the evaluation. Yet where relationships are good, the project team can provide invaluable support to the evaluators and, in fact, can form part of the evaluation team: this has the extra advantage of having project personnel familiarized with and confident in the evaluation methodology.

The professionals leading the evaluation should, if at all possible, have prior experience in carrying out evaluations. They should also be selected for their expertise in the particular foci of the evaluation (economic, social, health, environmental etc.). Great care and attention should be paid to selecting the field surveyors (interviewers). They are the evaluation's "front line" and must be depended upon to acquire information accurately and efficiently while maintaining good relations with the community. In many instances, the senior members of the evaluation team are not able to speak the local dialect (with the exception of, perhaps, the sociologist), so the interviewers should, if possible, come from the project area itself and be able to relate closely to the householders they interview. The local university is often a good source of interviewers in the form of graduate students; however, they must have the right orientation and attitudes towards the rural and urban poor, to avoid disputes which can result in conflict or, at least, disrespect beween interviewers and community members. Previous experience with interviewing in low- income areas is highly desirable, but, as a very minimum, the interviewers must have empathy with and understanding of those they interview.

It is necessary to determine how many interviewers are required. The team cannot be so small as to take months in the project area and not so large as to become unwieldy and logistically unmanageable. For example, if a 10 per cent sample of houses is required from 50 villages of 200 houses each, and each interview and transition between houses takes, on the average, one hour, 1,000 interview-hours are required. Thus, the interviewing could be carried out by five interviewers over a five-week period (assuming a 40-hour week), but an allowance of 50 per cent additional time for contingencies, such as bad weather, breakdown of transport, poor communications etc., should be made. In most instances, interviews should be carried out by a male/female team, with the woman interviewer taking the lead in discussions with the women of the household and the male interviewer talking with the men of the household.

A supervisor should be available to the interviewers at all times in the field, not only to check that reliable data are being obtained but also to provide support as necessary. Supervision is commonly the responsibility of the evaluation team's sociologist who also usually designs the survey's instruments and methods: otherwise, an experiencedsupervisor can be brought on to the team specifically for the purpose of supervision. While in the field, the team members should act together to draw on their multidisciplinary resources. Accommodation facilities should be chosen to ensure interaction between the team leader, the sociologist,the engineerand the interviewers, and every advantage should be taken of the fact that the team is isolated (away from its offices) and stays out in the survey area.

D. *Survey-instrument development*

There is a variety of survey techniques which can be used in the field, including interview by questionnaire, structured and unstructured discussion and observation: survey instruments are coded (where possible) to assist data analysis. Before being applied, instruments require field-testing and upgrading. Inevitably, field-testing of the instruments results in important improvements in survey technique, and, in some cases, it can completely alter the way the survey is carried out. Field-testing and instrument-upgrading take several days and may require a second run, if there are alterations in the original format. Surveys should be carried out in representative communities (and not only those close at hand), although this may mean travelling some distance to field-test in communities which, for some reason, differ significantly from others. The field test should cover the main sectors of the community (all income levels, religious groupings etc), the aim being to polish the questionnaire, making it relevant to all households being surveyed in the project area.

Field-testing should incorporate trial runs at data analysis: it is often realized at this point that the quantity of data being collected is unnecessarily large. Trial runs at data analysis through the computer also check the relevance of the coding to the computer programme and pin-point any "bugs" in the programme before the main data are run. Trial data should be analysed in their entirety, to determine just which data are important to the conclusions of the evaluation and which can be omitted. The tendency to collect more data than are useful is far stronger than that to collect too few. The consequence is frequently a surfeit of data which overload the data analysis and dilute the evaluation effort, thereby reducing its overall value.

E. *Interviewer training*

Interviewer training is best carried out during and after the field tests. It offers an informal opportunity for explanations of the underlying principles behind the selected survey technique and for contributions by the interviewers to instrument upgrading. In this way, the interviewers participate in the design of the evaluation and, having contributed to it, naturally deepen their interest in and commitment to the fieldwork ahead. Interviewers are asked to do very repetitive work; the same questions are asked over and over again. How questions are asked is just as important as what is being asked, but "interview fatigue" sets in after only a few houses, unless the interviewer is enthusiastic about the work and is intimately familiar with the interview technique and instruments.

The interviewer should be confident of the support which the supervisor offers in the field; he or she should also be aware that his or her work is being checked. The temptation to fill in questionnaires and invent discussions is great, especially if the interviewer is not properly supervised. In this light, it is dangerous practice (despite the incentive) to pay interviewers on the basis of the number of households interviewed.

F. *Co-ordination and logistic support*

The amount of co-ordination and logistic support required for the evaluation team in the field naturally depends on the size and scope of the evaluation: however, it always seems to be underestimated. When the sample size and number of interviewers are large, it is surprising how much of the team leader's time is taken up in simply co-ordinating who is where, at what time and by what means of transport. Communications within the team are strong when it consists of only two or three and only one vehicle is used; however, when 10 or 15 interviewers are scattered across two or three villages, the job of co-ordination becomes difficult. Then, organization, using timetables and written instructions, becomes essential. For example, an evaluation covering 20 villages might have four teams of interviewers, each requiring transport, supervision, advisory support, food and accommodation. Commonly, there are no telephones, and road access can be very difficult.

The essential things to ensure are; (a) that all members of the team are well informed; (b) that communication between them is good; (c) that members function as a unified team (this being encouraged by staying out in the field and having as much contact with one another as possible); (d) that timetabling is realistic; (e) that punctuality is insisted on; (f) that support in the field, such as accommodation, food, vehicles, provision for emergencies and introductions to the communities, is carefully planned; (g) that there is flexibility to deal with inevitable unforeseen events; and (h) that there is respect for team leadership. Again, the field test of survey instruments is invaluable in identifying the key support requirements for the team in the field.

G. *Fieldwork*

Fieldwork should be entirely standardized through good interviewer training and use of detailed survey instruments. This is very important, in that the data gathered by several interviewers are analysed together, so errors introduced through differences in interviewers and their interviewing techniques must be kept to an absolute minimum. Adequate time should be allowed for the unexpected: it is very likely that, among 5 or 10 interviewers, one will fall ill or otherwise be unable to function and that the whole timetable will have to be rearranged to spread the work among those remaining. It is not uncommon for field surveys to take 50 per cent more time than originally anticipated, and changes in timetable mean increased budgetary requirements.

The supervisor and senior staff should review the documentation being brought in by the interviewers every night. This helps identify inconsistencies, errors and omissions in time for correction. Data analysis is made easy if there are no gaps in the data.

H. *Data analysis*

It is advisable to carry out at least some data analysis in the field. If the data are to be analysed manually, it is necessary to be selective about which data are to be analysed: however, at least one or two key questions should be analysed. This will ensure that interviewers are consistently handing in their data every evening and help identify the general trend of the findings. Knowing that their data are being analysed day-by-day gives added incentive to the interviewers.

Battery-driven portable computers are available and are extremely useful in keeping up with data analysis while the survey is going on. In particular, lap-portables, with 640 Kilobytes memory capacity and discs, are within the price range of most evaluations and have capacity to cope with the data. There are also, on the market, statistical-analysis software packages which suit evaluation-data analysis: otherwise, analyses can be made using data-base-management software. At a very minimum, the data can be set out in a spreadsheet. The point is that computerized data analysis is within the reach of most evaluations and should be carried out while the data are being collected. This avoids all the work being lumped at the end of the fieldwork and allows for any "bugs" in the system or inconsistencies in the data to be finally worked out.

Another key to success is simplicity. It seems that the law of diminishing returns applies to large quantities of data and complex analysis techniques. Evaluations must be kept simple: it is far better to succeed in providing a few well-selected hypotheses than to end up with a myriad of ill-founded conclusions of questionable significance. It is the nature of impact evaluations that the important impacts are already observable and need only be confirmed by reliable data collection in the field: it is wise to be selective in what hypotheses are to be tested and then to concentrate on these few.

VI. DATA-COLLECTION METHODS

A. *Baseline data requirements*

The importance of baseline data cannot be overemphasized. Evaluation is made simple by knowing the situation and conditions in the area before the project was undertaken: if collected properly, baseline data allows accurate comparison of pre-project and post-project conditions. Naturally, the kind of baseline data needed depends on the scope of the eventual evaluation: it is for this reason that the evaluation should be planned early in the project cycle, i.e., during the initial feasibility study. In fact, baseline data should be taken using the same means of measurement as are used in the evaluation, so that the pre-project and post-project data can be directly compared.

At a very minimum, baseline data should include background information on the project area before project implementation, such as population characteristics (household size, numbers and growth rate), current practices of water use, defecation and solid-waste disposal, education levels, income distribution and employment characteristics. This information should normally be collected as part of the project feasibility study and design surveys: the information requirements for both project design and project evaluation are similar.

Every attempt should be made to plan the evaluation during the feasibility stage of project planning, so that the requirements for baseline data are detailed prior to their collection in the field during the project's detailed design phase. Where this has not been possible, the baseline data will have to be as broadly scoped as possible: this means collecting information on a wide range of economic, social, health, community and environmental parameters, all of which may not necessarily be used in the final evaluation. The team conducting surveys for the detailed design is in the best position to collect baseline data. As this team will, in most cases, carry out the project, its collecting baseline data will be useful, if it is eventually directly involved in the evaluation. The survey team will, of course, need to be trained, in that the kind of information required in baseline data-collection goes well beyond the purely technical information normally used in engineering design of installations.

B. *Data-collection methods*

There are several ways in which data can be collected in the field. The most common are observation, structured conversation, questionnaires and sampling for laboratory analysis. These methods and their purposes and pitfalls are described below. First, however, the evaluator should look to secondary data available through existing reports and documents.

Secondary data

Once the evaluation has been approved and resources set in place to carry it out, the first task is to collect all secondary data which may be relevant. This will include the project's feasibility study and detailed design documents, progress reports on the project, reports from ministries which have been peripherally involved in the project (the ministry responsible for economic planning, for example), local government and, of course, the evaluation plan and baseline data. It is often surprising how much information can be gleaned by casting a wide net for secondary data: it takes time, however, to gather them.

Often, some of the reports are considered to be confidential, and official permission is required before they can be released to the evaluation team: this takes time, so collection of secondary data should begin as early as possible.

Data collection by observation

There are two ways by which information is collected by observation: (a) observation in the natural setting; and (b) household observation. The purpose behind observing behaviour in the natural setting is not to cause interference which would alter behaviour. An example is observing water-collection practices at a standpipe. The observer positions himself/herself at a distance over an extended period of time, so that his/her presence is accepted and causes no changes in the usual water-collection practices of the people.

The data collected by such observation is very useful, but there are difficulties associated with observation in natural settings. The first is, of course, that the observer must spend a great deal of time in observing several situations; e.g., water-collection habits change during the day, week and season. Also information on the people drawing water is often unavailable; the only data that can be collected without disturbing the sitation are those related to water- collection habits. In other words, a disproportionate amount of the evaluation team's resources are spent on collecting relatively few data.

Also, boredom sets in quickly: passive observation is usually considered dull work, and the temptation to "cheat" is strong. Another problem concerns reporting: there is a tendency to report "what catches the eye", and the result is a series of disjointed notes which are of little use to the person trying to collate and analyse the data from several observational points. The two ways by which boredom can be somewhat alleviated and reporting systematized are: (a) time-sampling; and (b) checklisting. In time-sampling, the observer collects information only during predetermined time periods (on the hour for 15 minutes, for example): this allows the observer to take breaks and carry out other work in between. Checklisting involves the use of a carefully formulated checklist to focus the attention of the observer on the required behavioural information: care must be taken, however, to include all relevant items on the checklist and to train the observer well in what he/she must look for and record.

Household observations are opportunistic in that they are taken by the trained surveyor of specific items while visiting the household. Here, the surveyor may be interviewing the householder while, at the same time, taking mental notes of hygiene, water and sanitation conditions and practices around the home. Later, when conditions permit, the surveyor records the observations on a coded observational form. This is an efficient and accurate way of obtaining information about household conditions and practices. First, however, the observer must be invited into the house and the required situations: in most societies, this will only be possible if the observers are women. Secondly, the observers have to be well trained to know what they are looking for. Finally, predicting all situations for observing is not an easy task, especially in communities where houses vary considerably. Household observation may be an efficient method of collecting useful data, but the pitfalls are many: great care has to be taken in designing and field-testing observational methods.

Structured conversation

Structured conversation is used to elicit information from individuals within a community (such as a teacher, health worker or religious leader) without the use of a questionnaire. Questions are open-ended but are structured so that all selected topics are covered in a consistent fashion. No reporting forms are used during the interview; the discussion is kept as open and natural as possible. Structured conversation is usually carried out by the supervisor and other senior members of the evaluation team, as experience and skill are essential in acquiring reliable information.

Household surveys

Household surveys are used to obtain information about the families and their opinions. Normally a minimum of 10 per cent of families is surveyed, and questionnaires are used to standardize the format and method of interview, so that interfamilial responses are comparable. The questionnaires are coded, so that responses can be easily analysed, especially by computer. The questionnaire-based survey differs from the observational survey in that it interrupts normal routine; the interviewee must agree to participate and he or she must be willing to provide answers without fear of repercussions.

Surveys can be used to cover a far broader scope of subjects than observations, but there are several limitations and pitfalls associated with questionnaire-based surveys and the way they elicit information. The person being interviewed is asked to report his or her reaction or opinion on being questioned (this is termed self- reporting), but there is always the question of how reliable and accurate the responses are in self-reporting. There are several reasons for inaccuracy: the respondent may be unable to recall, owing to failure in memory, or may for one reason or another be unwilling to provide the correct answers to the questions. Biases may creep in as a result of the way the questions are asked: the "loaded" question is an example. Questions about personal hygiene are difficult to formulate without being loaded: responses to the question "do you wash your hands after defecation?" are bound to attract responses which are biased towards what the respondent thinks "should" be the answer.

Some subject areas are sensitive and should be handled through discussion and not questionnaires: money, for example, is always a difficult subject. An interviewer asking questions about income or possessions is likely to come under suspicion; the respondent will naturally want to know how the information he/she provides is going to be used. Likewise, questions about religion and politics are likely not to elicit accurate responses, especially within communities which are traditionally against government and associated institutions.

Questionnaires do not permit many open-ended questions or a naturally flowing discussion. They typically limit the number and variety of options which the respondent has in answering: the respondent may only be allowed to answer "yes" or "no", but it is also possible to provide the respondent with a rating scale in response to a question (excellent, good, bad, very bad). Even where a rating scale is provided it is an observed fact that respondents give higher ratings than in a naturally flowing discussion. Nonetheless, it is simply not possible to use open-ended questions in a questionnaire: analysis of open-ended responses from many households is too cumbersome and time-consuming. Questions that touch on vested interests will also not attract accurate responses: a surveyor from a water authority asking questions about felt needs within a community should not be surprised to find water supply high on the list.

The questionnaire-based household survey is useful as a method of acquiring specific information from large numbers of households, but it has drawbacks associated with the need to rely on standardized questions and responses and the inherent potential for biased data.

Field/laboratory analyses

Analysis of samples from the field offers one of the best ways to achieve accuracy and replicability which are the key characteristics of the objectively verifiable indicator. Most such analyses, however, are more applicable to effectiveness than to impact evaluations. Sample analysis does allow quantified measurement of variables: however, it is still important to ensure that other extraneous variables do not interfere with the comparability of results. For example, water- quality measurement by coliforms in the source water should take seasonal, sampling-location and climatic factors into account. Similarly, analysis of the quality of drinking water should account for contamination of the water beyond the tap, in the storage container in the house and from drinking utensils. Merely sampling the old well and comparing water from it with the water from the new tap will not give an accurate assessment of water quality as it is actually drunk.

Techniques used in analysing the samples being compared should be similar. Standard methods for chemical and microbiological analyses areavailable, but the calibre of the analyst can change over time. For example,many projects include institutional development, of which upgrading the analytical laboratory and its personnel forms a part. The baseline data in such cases may be suspect.

C. *Data sources*

The best source for secondary data is the implementing agency itself. It offers past reports (e.g.,feasibility-study and progress reports), project files, field notes, design drawings and baseline data. The most useful information comes from the staff of the implementing agency, however. Post-project evaluations are usually carried out some years after the installations have been made, and staff changes commonly limit the accuracy of the information and opinions about the project during construction. It is always best to understand the staff's relationship with the project and to be cognisant of any vested interests. It is best to know the internal politics of the agency before embarking on rounds of discussions with its staff: this will ensure that discussions on sensitive issuesare handled properly and will also enable the evaluator to identify who are the most reliable sources for information.

In this respect,the most difficult task the external evaluator has is to give a fair and accurate assessment of the project in a way that will not create undue antagonism within the agency, which may offset the value which the evaluation offers in improvingfuture projects. Knowing just how far to drive a point home in an official report without putting key agency staff on the defensive requires skill and judgement. The most frequently used approach is to include key agency staff in the evaluation team, so that they feel part of it and respect its conclusions. This, of course, has to be done without jeopardizing the evaluation itself.

The project

The physical infrastructure and the services provided by the project should be reviewed by all impact evaluations. Information is required not only on how many installations were made but on whether they continue to function properly. This pertains to efficiency and effectiveness but is important to impact evaluation just the same. Detailed information on the installations is usually available in records held by the implementation agency, but these should always be checked in the field.

Community leaders

The leaders of the community are important sources of information, but it is necessary to know what kind of leadership and authority they exert. There are several kinds of leaders: formally elected or appointed, informal (political), merchants, landlords and institution-based leaders, such as teachers and religious leaders. It is often best to start with the formal leader recognized by government, but others may have a great deal (sometimes more) to offer the evaluation. The evaluation team should be wary of being "steered" towards one influential group or another.

Health centres

Health centres are useful sources of information but again must be assessed for their own biases and accuracy of data provided. The nurse, health-care worker and sanitarian are good sources, if they are actively working in the project area, but health statistics collected by clinics are frequently deficient, for several reasons. For example, statistics collected through clinic visits by community members complaining of diarrhoea cannot be generalized as being representative of the entire population: under-reporting is common. The clinic naturally attracts the families whose houses are closest, but many families prefer to visit the traditional doctor or midwife.

Misreporting is also common, since the precise cause of complaint may be difficult to diagnose, owing to other concurrent causes of illness and disability. Reference is made to Briscoe and others (1986) for further details on information sources for evaluating health impact.

Women

Water and waste management is considered a " women's issue". It is usually Women who are responsible for obtaining water, protecting it and using it in and around the house. Likewise, wastewater discharge and refuse disposal fall under their purview. From all perspectives, women are very important as sources of information for the evaluation. Most cultures do not welcome outside male interviewers into houses, and women interviewers are required for this purpose. Women as interviewers are often able to ask questions of the women of the household separately from the men. In this way, answers can be obtained without distortion which may occur if male family members are sitting in on the conversation.

Families

Families are the primary source of information for the evaluation. As described earlier, the questionnaire survey is normally the preferred method, coupled with observations taken within the household. It should be ensured that the sample of houses taken is representative of the general population, otherwise, the information will be biased. For example, there is always a great temptation to leave out those households which are unoccupied when the interviewers make their first round: this cannot be called a random sample. Biases occur if the reason for not being in or refusing to answer the door influences the information which would have been given: such would be the case if those not responding objected to the evaluation because of objections to the project itself.

Care must be taken not to oversurvey the community. A particular community may undergo several surveys for various purposes, but, after not getting any significant benefit from these surveys, the community may continue to receive visitors politely but provide inaccurate information. The surveyor must respect the privacy of the family and recognize that its time is just as valuable as his or her own. Neither the community nor the family should be subjected to survey "overdose".

Some questions will pertain to the householder who is commonly the senior male in the family, but most others will require participation by the women. Family members cannot be expected to interrupt their daily routine to answer questions: the survey must adapt to local work and holiday patterns, and approach the families when they are most available for interviewing.

VII. CONCLUSIONS

Increased awareness and measurement of the economic, social, health, community and environmental impacts triggered by the introduction of water-supply, sanitation and solid-waste-disposal facilities could promote investments and the assignment of high priority for these elements in national plans. Impact-evaluation studies, designed with a focus beyond health improvements and including social, economic, community and environmental aspects, defined within a systematic logical evaluation framework (LEF), present an opportunity to determine the efficiency, effectiveness and overall benefits and effects of basic service-intervention projects. Impact evaluation is at a stage where objectively verifiable indicators of impact and methodologies for their quantification can be consolidated and standardized within LEF, to permit measurement of basic-service project impact and promote awareness of the diverse benefits of the provision of basic services.

Careful design of evaluation studies, use of appropriate and statistically representative methods of data-gathering from reliable sources, technically correct data processing and adequately selected and trained evaluation teams are all essential to ensuring unbiased conclusive evaluation results. The evaluations that are most likely to be successful in yielding conclusive results are those that focus on key issues, take account of behavioural changes, are appropriately timed, are disaggregated to take account of age-related basic-service utilization patterns and are designed to minimize the effect of variables which are outside project intervention but which cause similar impacts. Evaluations planned at the beginning of the project permit the collection of high-quality data from project inception to completion. A combination of cross-sectional and longitudinal evaluation designs optimizes the cost and reliability of evaluation conclusions and the time needed for their completion. High-quality baseline data are essential for all impact evaluations and have a notable bearing on the overall quality of the evaluation.

Traditionally, impact assessment has been confined to health aspects, and indicators and methodologies for these are well developed. Indicators of social, community and environmental impacts are little developed: developmental impacts associated with human- settlement upgrading have been particularly neglected in the past. Revelant objectively verifiable independent indicators, which are cost-effective and sensitive to any changes in the impact being investigated, are essential for assessing basic-service intervention impacts. Careful selection and field-testing of data-collection procedures, such as observation, structured conversations and discussions and household surveys, and the adoption of standard procedures for field/laboratory analyses are important for ensuring accuracy and replicability of the measured impact.

REFERENCES AND FURTHER READING

Biswas, A.K. (1984). "Monitoring and evaluation of an irrigation system". International Journal of Water Resources Development, 2(1): 3-25.

Blum, D., and Feachem, R.G. (1983). "Measuring the impact of water supply and sanitation investment on diarrhoeal diseases: problems of methodology". International Journal of Epidemiology, 12(3): 357-365.

Bradley, D.J. (1977). "The Health Aspects of Water Supplies in Tropical Countries". In Water, Wastes and Health in Hot Climates, R. Feachem, M. McGarry and D. Mara (eds.) (Chichester, John Wiley & Sons.)

Briscoe, J., Feachem, R.G., and Rahaman, M.M. (1985). Measuring the Impact of Water Supply and Sanitation Facilities on Diarrhoea Morbidity: Prospects for Case-Control Methods (Geneva, World Health Organization Environmental Health Division and Diarrhoeal Disease Control Programme).

Briscoe, J., Feachem, R.G. and Rahaman, M.M. (1986). Evaluating Health Impact: Water Supply, Sanitation and Hygiene Education (Ottawa,International Development Research Centre).

Cairucross, S., and others (1980). Evaluation for Village Water Supply Planning (The Hague, WHO International Reference Centre for Community Water Supply).

Canadian International Development Agency (1977). Project Development-Perspectives and Methodology (Ottawa).

Carruthers, I.D. (1973). Impact and Economics of Community Water Supply-A Study of Rural Water Investment in Kenya, Agrarian Development Studies Report No.6 (Ashford, Kent, School of Rural Economics and Related Studies, Wye College).

Chambers, R. (1981). "Rural povertyunperceived: problems and remedies". World Development, 9(1):1-19.

Cullivan, D.E., and Careaga, R. (1977). "Environmental impacts of water supply projects in developing countries". In C.G. Gunnerson and J.M. Kalbermatten (eds.) Environmental Impacts of International Civil Engineering Projects and Practices (New York, American Society of Civil Engineers).

Drake, W.S., Miller, R.C., and Humphrey, M. (1980). Analysis of Community-Level Nuturition Programs (Washington, D.C., Office of Nutrition, USAID).

Drucker, D. (1981). "Ask a silly question-get a silly answer: community participation, entry points and the demystification of planning" (unpublished manuscript).

Dworkin, D.M., and Pillsburg, B.L.K. (1980). The Potable Water Project in Rurual Thailand, Project Impact Evaluation No.3 (Washington, D.C., United States Agency for International Development).

Elmendorf, M.L. (1981). Women, Water and the Decade, WASH Technical Report No. 6 (Arlington, Virginia, Water and Sanitation for Health Project).

Elmendorf, M.L., and Isely, R.B., (1981). The Role of Women as Participants and Beneficiaries in Water Supply and Sanitation Programs, WASH Technical Report No. 11 (Arlington, Virginia, Water and Sanitation for Health Project).

Esrey, S.A., Feachem, R.G., and Hughes, J.M. (1985). "Interventions for the control of diarrhoeal diseases among young children: improving water supplies and excreta disposal facilities". Bulletin of WHO, 63(4) : 757-772.

Feachem, R.G. (1984). "Water supplies for low-income communities: resource allocation, planning and design for a crisis situation". In R. Feachem, M. McGarry and D. Mara (eds) Water, Wastes and Health in Hot Climates, (Chichester, John Wiley and Sons).

Feachem, R.G.., (1984). "Interventions for the control of diarrhoel diseases among young children: promotion of personal and domestic hygiene". Bulletin of WHO, 63(3): 467-476.

Feachem, R., and others (1978). Water, Health and Development (London, Tri-Med Books Ltd.).

Frankel, R.J. (1973). Evaluation of Effectiveness of Community Water Supply in Northeast Thailand, Final report submitted to the Ministry of Public Health (Bangkok, Asian Institute of Technology).

Glasgow, M., and Howard, P. (1983). Minimum Evaluation Procedure for Health Education Evaluation in Water and Sanitation Projects (Amsterdam, Royal Tropical Institute).

Gunn, A., and Ballance, R.C. (1983). Maximum Benefits to Health... An Appraisal Methodology for Water Supply and Sanitation Projects (Geneva, World Health Organization).

Heijnen, J.D., and Conyers, D., (1971). "Impact studies of rural water supply". In A. Tschanneri (ed.) Water Supply-Proceedings of the Conference on Rural Water Supply, (University of Dar-es-Salaam, 5-8 April 1971).

Hughes, J.M., (1981). Potential Impacts of Improved Water Supply and Excreta Disposal on Diarrhoeal Disease Morbidity: An Assessment Based on a Review of Published Studies. Consultation Report (Geneva, World Health Organization, Diarrhoel Disease Control Programme).

International Drinking Water Supply and Sanitation Decade. Suggested Guidelines for Incorporating Community Communication and Participation in Project Formulation.

Kawata, K., (1978). "Of typhoid fever and telephone poles: deceptive data on the effect of water supply and privies on health in tropical countries". Progression Water Technology, 11 (1/2): 3-43.

Kirk,Roger E., (1982).Experimental Design: Procedures for the Behavioural Sciences (Belmont, California, Brooks/Cole Publishing Co. Wadsworth Inc.).

MacCormack, C.P., (1983). Minimum Planning and Evaluation Guidelines for Women, Health and Development. (London School of Hygiene and Tropical Medicine).

McJunkin, F.E., (1982). Water and Human Health (Washington, D.C., United States Agency for International Development).

Magnami, R., and others (1984). Evaluation of the Provincial Water Project in the Philippines (Washington, D.C., International Statistical Program Centre, Bureau of the Census).

Myers, J.L., (1979). Fundamentals of Experimental Design (3rd ed.) (Boston, Allyn and Bacon, Inc.).

Office of the Auditor General for East Asia (1972). Audit Report No. 8-493-73-3 (Washington, D.C., United States Agency for International Development).

Perrett, H., (1983). Social Feasibility Analysis in Low-Cost Sanitation Projects (TAG Technical Note No. 5) (Washington, D.C., World Bank).

Perrett, H.E., (1984). Monitoring and Evaluation of Communication Support Activities in Low-Cost Sanitation Projects (TAG Technical Note No.11) (Washington, D.C., World Bank).

Rosenberg, I.H., Solomon, N.W., and Lovin, D.M., (1976). "Interactions of infection and nutrition: some practical concerns". Ecology of Food and Nutrition, 4:203.

Saunders, R.J., and Warford, J.J., (1976). Village Water Supply:Economics and Policy in the Developing World (Baltimore, Johns Hopkins University Press).

Shiffman, M.A., and others (1978). :Field studies on water, sanitation and health education in relation to health status in Central America". Progression Water Technology, 11 (1/2): 143-150.

Shuval, H.I., and others (1981). "Effect of investments in water supply and sanitation on health status: of the shold - sanitation theory", Bulletin of the World Health Organization, vol. 59, pp. 243-248.

Simpson-Hebert, M., (1983). Methods for Gathering Socio-cultural Data for Water Supply and Sanitation Projects (TAG Technical Note No.1) (Washington, D.C., World Bank).

Struba, R.J., and Isely, R.B., (1981). The Choice of Health Status Indicators to Evaluate Water and Sanitation Project in North Cameroon: A synthesis of Available Information, WASH Technical Report No. 5 (Arlington, Virginia, United Nations Administrative Committee on Co-ordination, Water and Sanitation for Health Project).

Task Force on Rural Development (1984). Guiding Principles for the Design and Use of Monitoring and Evaluation in Rural Development Projects and Programs (Rome, International Fund for Agricultural Development).

United Nations Childrens Fund (1983). Abstracts of Evaluation Reports of UNICEF-Supported Water Supply and Environmental Sanitation Programmes. (New York, United Nations Childrens Fund).

United States Agency for International Development (1974). Project Evaluation Guidelines (Washington, D.C.).

Warner, D.B., (1981) Social and Economic Preconditions for Water Supply and Sanitation Programs, WASH Technical Report No.10. (Arlington, Virginia, Water and Sanitation for Health Project).

Westman, B., and Hedkvist, F., (1972). The Impact of the Rural Water Program in Tanzania.

White, G.F., Bradley, D.J., and White, U.A., (1972). Drawers of Water-Domestic Water Use in East Africa (Chicago, University of Chicago Press).

World Health Organization (1981). Manual for the Planning and Evaluation of Diarrhoeal Disease Control Programmes WHO/CDD/SER/81. 5) (Geneva, World Health Organization).

World Health Organization (1983). Minimum Evaluation Procedure (MEP) for Water Supply and Sanitation Projects (Geneva, World Health Organization).

Whyte, A. (1983). Guidelines for Planning Community Participation in Water Supply and Sanitation Project (Geneva, World Health Organization).