

WHO/CHD/97.2
Distr.: GENERAL
Original: English

Evaluating the Impact of National CDD Programmes



**Division of Child Health and Development
World Health Organization
Geneva**

© World Health Organization 1997

This document is not a formal publication of the World Health Organization (WHO), and all rights are reserved by the Organization. The document may, however, be freely reviewed, abstracted, reproduced or translated, in part or in whole, but not for sale or for use in conjunction with commercial purposes.

The views expressed in documents by named authors are solely the responsibility of those authors.

a 52238

WHO/CHD/97.2
Distr.: GENERAL
Original: English

Evaluating the Impact of National CDD Programmes



CONTENTS

Executive Summary	Page 3
Section I. Background and Purpose	Page 4
Section II. Conceptual Frameworks	Page 4
- Evaluation Frameworks	Page 4
- Framework for Assessing Programme Impact on Diarrhoea Morbidity and Mortality	Page 7
Section III. Evaluation Design	Page 8
Section IV. Steps of the Evaluation	Page 11
Step 1. Review programme documentation and the availability of data on coverage and impact	Page 11
Step 2. Decide whether the evaluation is feasible	Page 12
Step 3. Define the initial objectives and hypotheses	Page 14
Step 4. Assemble the required data	Page 16
Step 5. Redefine the evaluation hypotheses	Page 27
Step 6. Analyze the data	Page 27
Step 7. Interpret the results	Page 30
Step 8. Write the final report and disseminate the findings	Page 33
REFERENCES	Page 34

Executive Summary

The implementation of National CDD Programmes in several countries has been accompanied by reports of declining diarrhoea mortality and severe morbidity. These claims have not often been adequately documented. This raises the need for a more rigorous evaluation of programme achievements. This document provides guidance on the main steps involved in carrying out an evaluation of the impact of a NCDDP, based on recent evaluations carried out in Brazil and in the Philippines.

A conceptual framework characterizes programme achievements in terms of adequacy (whether programme objectives have been met) and plausibility (how likely is it that the programme itself can be held responsible for the achievements). The framework also classifies indicators as measuring performance (including services provision, utilization and coverage) or health impact.

The guidelines then lead the evaluation team through a succession of steps. Initially, there is a need for an overview of the existing data regarding performance and impact evaluations, in order to decide whether an evaluation is feasible. Criteria to assist in making this decision are presented. Next, the objectives and working hypotheses should be defined and the data to test them should be gathered. The main data sources on performance indicators will be programme documentation and existing surveys, while for impact data from vital statistics and record reviews are most often used. The importance of the critical evaluation of data quality is highlighted. Detailed guidance is provided on the main sources of bias, accompanied by specific suggestions on how these may be overcome. After screening the available data, objectives and hypotheses may need to be redefined.

The next step is the analysis of the data and their interpretation in light of the hypotheses and conceptual framework. The importance of pooling together data from as many different sources as possible is stressed. Because secondary data are known to be affected by several types of bias, reliance on a single source might lead to major problems in interpretation. While the assessment of adequacy objectives is straightforward, plausibility analyses are more complex. They may entail, for example, trying to rule out external factors as explanations for a declining trend in diarrhoea deaths, or correlating ORT use rates with mortality rates across different geographical units, or examining case-fatality trends in selected facilities before and after implementation of ORT. The plausibility that any decline in morbidity or mortality rates was due to programme activities will be enhanced by showing that different types of analyses point in the expected direction.

Finally, guidance is provided on the preparation and dissemination of the evaluation findings.

Section I. Background and Purpose

National Control of Diarrhoeal Diseases Programmes (NCDDP) have been established in over 100 less developed countries since 1980. Monitoring and evaluation activities are an integral part of these programmes. A systematic, data-based assessment of the achievements and constraints of a NCDDP is a part of the Focused Programme Review process recommended by WHO¹. The FPR guides a team of national and external participants through the identification of achievements and priority issues, the summary of available data and collection of additional information if needed, the identification and in-depth analysis of priority issues and possible solutions, and the development of a work plan to implement solutions and monitor their implementation.

Recently, several countries have reported declining trends in diarrhoea morbidity and mortality. Some of these countries have expressed a need to conduct a more rigorous evaluation of the extent to which their programmatic efforts have led to reductions in diarrhoea severity and mortality. The purpose of this document is to provide guidance in planning such an impact evaluation. The topics addressed include: a conceptual framework for evaluations; the general design of the evaluation; prerequisites for a successful evaluation; a step-by-step discussion of how to proceed with the evaluation, including possible methods and sources of data; major measurement challenges likely to be faced by the evaluation team; and data analysis and interpretation.

The target audience for this document includes independent researchers and NCDDP staff that may be involved in evaluating such programmes.

Section II. Conceptual Frameworks

Prior to the evaluation, it is necessary to decide what type of evaluation will be carried out. This section provides a conceptual framework for evaluation aimed at helping the research team with this decision. It also contains a discussion of factors affecting diarrhoea incidence, severity and case-fatality. Such factors must be taken into account when assessing the impact of the programme.

Evaluation Framework

The main objective of an evaluation is to influence decisions. Evaluation design depends on who the decision maker is and on what types of actions will be taken as a consequence of the findings. In the case of NCDDP evaluations, senior officers at the Ministry of Health may want to be reassured that the programme is moving in the right direction and that it has improved the health of children. NCDDP evaluation results may also have an important role in advocacy, as well as in influencing international organizations and donors regarding their continued support for the programme.

Based on the need to design evaluations aimed at supporting decision-making, Habicht and colleagues² have proposed a framework with two main axes. The first refers to the type of indicators to be assessed, whether these refer to the provision of services, to their utilization, to the coverage

reached in the population, or to the programme's health impact^a. Both provision and utilization may be assessed from health services records, but coverage and impact are population-based indicators, usually requiring either household surveys or vital registration statistics. Provision, utilization and coverage indicators are often described as performance indicators. Table 1 shows the definitions of these terms. A logical order leads from provision to impact. The services must be available, accessible and of adequate quality in order to be used. This utilization will result in a given population coverage. Finally, the achieved coverage may lead to an impact on health. Any important shortcomings in the early stages of this chain will result in failures in the later achievements.

Table 1. Types of indicators for evaluating a National CDD Programme.

INDICATOR	QUESTIONS	COMMENTS
Provision ("Inputs")	Are the services or activities available? Are they accessible? Is their quality adequate?	Firstly, the services must be provided, that is, available, accessible to the target population and of adequate quality.
Utilization ("Outputs")	Are the services being used?	Second, the population must accept the services and make use of them.
Coverage ("Outcome")	Is the target population being reached?	Third, this utilization will result in a given population coverage.
Impact	Were there improvements in health?	Finally, the achieved coverage may lead to an impact on morbidity or mortality.

The distinction between utilization and coverage is subtle but important. Utilization relates to the number of services, activities or supplies distributed. In the CDD context, this may represent, for example, the number of ORS packets distributed or the number of mothers who attended hygiene education sessions. Utilization, as mentioned, may be assessed from health services records. However, it is often impossible to estimate actual population coverage from these data, because a few children may have received a large number of packets, or a few mothers attended several sessions. Simply dividing the number of attenders by the population may thus be misleading. Coverage assessments provide more precise information by estimating through household surveys the actual proportion of the target population who received the services or activities. For example, the proportion of children with diarrhoea who received ORT, or the proportion of mothers who know the basic principles of hygiene education. Coverage is a particularly useful measure, representing

^a In the evaluation literature, these indicators are often referred to as inputs, outputs, outcomes and impact, but Habicht et al propose the use of the terms provision, utilization, coverage and impact as these have more precise meanings.

the interface between services or activities delivery (the managerial process) with the population (the epidemiological picture).

The second axis of the framework refers to the type of inference statement to be made. The simplest statement refers to adequacy, that is, as to whether or not the indicators are moving in the expected direction. However reassuring, it may not be sufficient to show that - for example - diarrhoea mortality is falling because this may be due to factors external to the programme. The second level of inference, therefore, refers to plausibility, that is, to how plausible or likely is it that the programme itself has led to the observed trend. Plausibility analysis requires some type of comparison, either with areas where the programme has not been fully implemented or a before-and-after comparison. It also requires specific attempts to rule out external factors (see below) that might confound the observed association. Finally, the highest level of inference refers to probability analysis. This requires a randomized trial in which the programme is allocated to some geographical areas and not to others, to allow a strict assessment of its effectiveness. This latter type of evaluation design does not apply to the evaluation of existing national programmes where randomized allocation of the intervention did not take place.

The two axes may be combined. Table 2, modified from Habicht et al², shows possible types of NCDDP evaluations that may be designed through their combination. In general, the complexity and costs of evaluations increase from the upper left corner of the table to its bottom right corner. Also, evaluations of provision should precede those of utilization, these should precede those of coverage, and so on. Adequacy evaluations should also precede those of plausibility. The framework therefore proposes a logical sequence for performing evaluations and may contribute to saving time and resources. Throughout this document, reference will be made to this framework.

It is important that the review process should make use of all available information on provision, utilization and coverage - obtained, for example, through household or health facility surveys. The critical assessment of all information is essential for building up the plausibility that the NCDDP has effectively led to a health impact.

Table 2. Examples of possible evaluations of National Control of Diarrhoeal Diseases Programmes

Type of evaluation	Performance			Impact
	Provision	Utilization	Coverage	
Adequacy	Changes in availability of ORS in health centers	Changes in numbers of ORS packets distributed in health centers	Measurement of percentage of all diarrhoea episodes treated with ORT in the population	Measurement of trends in diarrhoea mortality in the programme area
Plausibility	As above, but comparing services with full implementation of the NCDDP with those with partial implementation	As above, but comparing services with full implementation of the NCDDP with those with partial implementation	Comparison of ORT coverage between areas with full NCDDP implementation versus those with partial implementation (dose-response)	Comparison of diarrhoea mortality between areas with full NCDDP implementation versus those with partial implementation (dose-response)
				Assessment of likelihood that observed trends in diarrhoea mortality were in fact due to the NCDDP and not to external influences
Probability	As above, but following the random allocation of the programme to intervention or control areas of the country (not applicable to evaluation of existing NCDDPs)			

Framework for Assessing Programme Impact on Diarrhoea Morbidity and Mortality

A CDD programme does not stand alone in influencing diarrhoea morbidity and mortality. Other programmes in different sectors (health, nutrition, education, etc) as well as external factors may affect it. Table 3 shows some of these factors, that may have to be taken into account in the evaluation. The table is not exhaustive - in many countries, other programme-related and external factors may also have to be considered.

Table 3. Selected CDD Programme-related interventions and external factors that may affect diarrhoea incidence, severity or case-fatality.

CDD PROGRAMME-RELATED INTERVENTIONS	EXTERNAL FACTORS
Factors affecting diarrhoea incidence	
Hygiene education Breastfeeding promotion Promotion of safe weaning practices Measles vaccination	Low socioeconomic status Inadequate water and sanitation Seasonal and climatic factors Poor hygiene practices Lack of breastfeeding Unsafe weaning practices Measles incidence
Factors affecting episode severity	
Oral rehydration therapy	Malnutrition Vitamin A/other micronutrient deficiencies Inadequate household management
Factors affecting the case-fatality of severe episodes	
Oral and parenteral rehydration therapy Dysentery management Persistent diarrhoea management	Malnutrition Inadequate care-seeking behaviour

Section III. Evaluation Design

The design of the evaluation is based on which indicators will be assessed and what type of inference is to be made. This section addresses the main steps involved in a NCDDP evaluation, which are spelled out in detail in the forthcoming sections. The experience from Brazil shows that this evaluation may be carried out in a 12-month period. Figure 1 includes a flowchart with the main steps of the evaluation process.

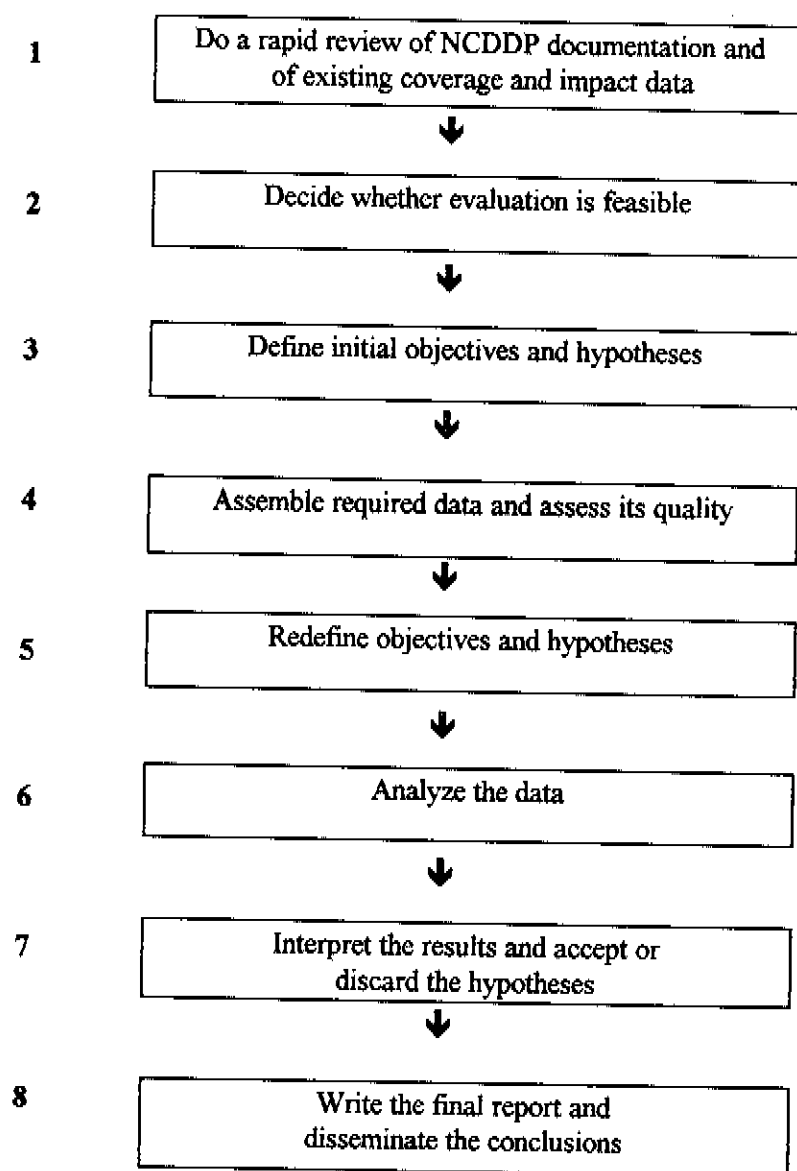


Figure 1. Flowchart with the steps in NCDDP evaluation exercise.

The evaluation effort should involve members from the NCDDP as well as external experts. The team should include at least one epidemiologist with experience in programme evaluation as well as in handling vital statistics and in analytical methodology. Members or consultants with skills in demography and statistics may also be required. Ideally, the principal investigator would be external to the programme, possibly affiliated with a national university or research institute. The team should also include one or more members with good political contacts in the government health and statistics sectors as the collaboration of these bodies is essential. The leading members of the research team - say, two to three investigators - should be appointed early in the process described in Figure 1.

First, evaluations must start with an accurate description of programme activities. Typically, this step will address indicators of provision and utilization of services. The evaluation team should clearly define the intermediate or shorter-term outcomes that they expect to result from programme activities. These can be improvements in the resources available for diarrhoea prevention or case-management in facilities or communities (indicators of provision), increased use of existing services (indicators of utilization) or changes in knowledge or behaviours (indicators of coverage). The re-examination of targets and subtargets from the NCDDP plan should help focus the evaluation effort. These targets may serve as the basis for judging the adequacy of the observed changes by assessing whether the pre-established goals have been met.

As noted in the conceptual framework for evaluation, only after specifying the indicators of services provision, utilization and coverage should the evaluation team consider the health impact of the programme, that is, changes in diarrhoea-related morbidity and mortality.

Regarding impact indicators, most NCDDPs have centered their efforts in case-management activities, particularly the promotion of Oral Rehydration Therapy (ORT). ORT will prevent and treat dehydration among children with acute watery diarrhoea, and therefore is expected to reduce hospital admissions and deaths due to diarrhoea. ORT does not affect diarrhoea incidence nor does it prevent the occurrence of either persistent diarrhoea or dysentery - although it has a useful subsidiary role in the management of these conditions. ORT-focused programmes, therefore, may have an impact on overall rates of diarrhoea deaths and hospital admissions, particularly those due to acute watery episodes, but not on diarrhoea incidence (see Table 3). NCDDPs with a strong emphasis on preventive measures - such as breastfeeding promotion, personal and domestic hygiene, etc - may also affect diarrhoea incidence. These issues must be borne in mind when selecting impact indicators. These indicators are further discussed, with examples, on page 26.

Having reviewed the achievements of the programme and the nature of existing data, the core research team will have to decide whether an evaluation is feasible (Step 2 in Section IV). If this is the case, then the team may be expanded to incorporate additional investigators, and the initial objectives and hypotheses must be drafted (Step 3). At this stage the research team will have to decide whether it will satisfy itself with adequacy statements, that is, limiting its task to the documentation of the expected changes in performance and health impact. In most cases, the answer to this question will be negative, and the team will want to move further into showing the plausibility that these changes have been due to the NCDDP. This has important implications for the design and analysis of the evaluation. Plausibility statements require some sort of comparison (either before-and-after or between different regions with varying intensity of the programme implementation) as well as control of external or confounding variables. There is thus a need for data on time trends in programme-related, impact and confounding variables, as well as on their geographical distribution.

The research team will then set out to obtain the relevant data and assess its quality in detail (Step 4), building upon the early assessment of availability and quality of data that was done in Steps 1-2. After this, it may have to redefine the objectives and hypotheses in light of the existing data

(Step 5). Next, the team will analyze the data (Step 6) and interpret the findings (Step 7), concluding its task by writing the final report and disseminating the conclusions (Step 8).

Section IV. Steps of the Evaluation

Step 1. Review programme documentation and the availability of data on coverage and impact.

This step entails an early review of the programme documentation to describe its implementation and main activities, and a rapid overview of data quality. This will often be done prior to a detailed planning of the evaluation, as this review process will help answer some of the questions posed in Step 2 regarding the feasibility of the evaluation, as well as being essential to the definition of its objectives and hypotheses (Step 3).

The CDD Programme Management Flowchart (Annex A) provides a model of programme implementation. The first step in evaluating the NCDDP is to examine actual programme activities in relation to the flowchart and the national plan. Programme activities should be clearly described and quantified. Much of this information can be drawn from Phase 1 of a Focused Programme Review, if one has been conducted recently. Information from various sources should be compared, and their relative quality assessed, before arriving at an accurate description of programme implementation.

The product of this step should be a detailed programme implementation summary, describing what was done, where, by whom, how and when. Whenever possible, attention should be directed to the quality and intensity of activities (e.g., training) and not only to the fact that an activity occurred. Box 1 lists some of the possible data sources for obtaining the information required in this step.

Box 1. Possible sources of data on programme implementation activities

Key documents	Key informant interviews
NCDDP Plan of Action and policy documents	DTU manager
Programme review reports (e.g., annual reports)	MOH Health Communications and
Country Profile (CDD), Annual Training Summary	Training staff
Survey reports (household, health facility)	Health/vital statistics staff
Consultants' reports	EPI staff, MCH staff
Routine reporting data (e.g., diseases notification)	Donor agencies representatives
Sentinel surveillance	UN agencies staff (WHO, UNICEF)
Supervisory and monitoring reports	
Reports of training activities	
Budget records	
Relevant research studies	

Drawing on the results of the above exercise, the evaluation team should clearly specify the achievements expected from programme activities, including the target populations, the specific changes in knowledge and behaviours that were forecast, and the expected relationship between these variables and diarrhoea morbidity and mortality. These goals will help set the basis for stating the evaluation objectives and hypotheses (Step 3 below). As an example, Box 2 lists some of the targets of the Philippines NCDDP.

Box 2. Goals of the Philippines National CDD Programme for 1987-92	
Provision	<ul style="list-style-type: none"> ● ORS to be available in 100% of health outlets during the 1992 diarrhoea season ● 85% of the population to live within 5 km or 1 hr from an outlet with ORS ● 100% of government health facilities to have staff knowledgeable in ORT and who use it for dehydration ● 100% of ORT corners to have adequate supplies, equipment and IEC materials ● 100% of the ORS requirements to be procured or produced
Coverage	<ul style="list-style-type: none"> ● 35% of mothers to know ORS and use it properly
Impact	<ul style="list-style-type: none"> ● Reduction by 50% of diarrhoea mortality (from 8.5 to 4.3 deaths/1,000 underfives) ● Reduction by 20% of diarrhoea morbidity (from 2.8 to 2.2 episodes/child-year)

An important component of this first step is to verify the availability of population-based data on programme coverage and health impact. The key informants are likely to be aware of major surveys regarding diarrhoea management. Also, staff from the Health Statistics Unit or the National Bureau of Statistics will provide information on the availability and quality of morbidity and mortality data.

Step 2. Decide whether the evaluation is feasible.

Countries interested in conducting this type of evaluation should first consider carefully whether they are likely to be successful in producing useful information. At a minimum, the evaluation team should carefully consider whether the following prerequisites can be met:

- *Availability of specific information on programme activities.*

The quality of the evaluation will rest on whether programme activities have been adequately documented. The team should carefully assess the quality and completeness of information that can be obtained on programme implementation before embarking on an evaluation of this type.

- *Existence of baseline and follow-up data for performance and impact indicators.*

The choice of specific indicators to be studied will be based on programme objectives, targets and activities. Nevertheless, there is a need for information on the quality of case-management at home (for example, the ORT use rate) and in health facilities, as well as on diarrhoea morbidity and mortality. These data should be available for at least two points in time, preferably prior to programme implementation (or at the very early programmatic phase) and at a time sufficiently later to allow changes to have occurred, say 3-5 years later.

- *Availability of information on other programmes or external factors that may have affected the performance or impact indicators under study.*

As discussed, even when improvements in programme performance or impact are shown to be adequate, it is possible that these changes were due to factors other than the NCDDP activities. To increase the plausibility of the attribution of the effect to the programme, data should be available on alternative explanations for improvements in performance (e.g., increases in health worker pay, major training or supervision inputs by other public health programmes; reduction in fees charged for services) or in health impact (e.g., changes in measles immunization rates, water or sanitation coverage, vitamin A supplementation or breastfeeding duration). A thorough knowledge of the country situation as well as of the NCDDP performance and of local risk factors for diarrhoea is required for identifying the relevant confounding variables or external factors.

- *Reason to believe that the level of programme activities justifies expectation of changes in coverage and impact.*

Evaluations of coverage should not be carried out unless there is sufficient reason to believe that programme activities have led to changes in the provision and utilization of services. Likewise, impact evaluations should not be attempted unless important changes in coverage are assumed.

- *A willingness to accept that definitive conclusions on the relationship between programme activities and health status are unlikely to result from the investigation.*

Given that the above prerequisites can be met, the programme should also consider whether the objectives of the evaluation are realistic. If the objective is solely or primarily measuring

the health impact of programme activities, the evaluation team should be aware that the study is unlikely to prove that changes in health status are due to programme activities. The study will have several limitations, among them the evaluation design (in most cases limited to a before-and-after comparison), the absence of an adequate concomitant control group, and the reliance on information of variable quality and completeness. The evaluation will attempt to "build up the case" that changes observed in health status can be plausibly attributed to programme activities, but will be unlikely to establish causality. Such type of formal proof requires a strict probability evaluation, which as seen above is not feasible in the context of NCDDPs. The strength of the plausibility statement will depend on the intensity of programme implementation; on the association in time between implementation, coverage and impact; on the magnitude of the change in coverage and health impact; and on the ability to demonstrate that these changes are unlikely to be due to alternative interventions.

Unless these five prerequisites are met, it may be difficult and frustrating for the evaluation team to successfully accomplish its task.

Step 3. Define the initial objectives and hypotheses.

The evaluation team should clearly state its general and specific objectives, as well as the hypotheses under study. To spell them out, the team should be well acquainted with what data may be available. The general objective may be:

"To assess the effectiveness of the activities of the NCDDP in improving the quality of diarrhoea case-management and prevention and in reducing diarrhoea-related morbidity and mortality."

Specific objectives may address adequacy or plausibility. The examples below are generally based on the Brazil NCDDP Evaluation. Examples of adequacy objectives might include:

"a) To describe annual trends in the numbers of:

ORS packets produced or procured;

health-service attendances of children with diarrhoea^b;

health services staff trained in diarrhoea case-management;

^b Interpretation of data on the number of health attendances of children with diarrhoea is complex. Early in a NCDDP, increased numbers of attendances may be interpreted as a measure of programme success in raising awareness of families that diarrhoea is a potentially serious event and that care must be sought. Later in a programme, appropriate case-management at home may result in a decrease in the number of attendances. No specific hypothesis, therefore will be made regarding the desirability of increased or decreased numbers of attendances.

educational materials produced and distributed;

mass media campaigns.

- b) *To assess trends and levels of ORT coverage (the proportion of diarrhoea episodes in children under five years that were treated with ORT).*
- c) *To assess trends in diarrhoea-related hospital morbidity (such as OPD attendances, admission rates, duration of hospitalization).*
- d) *To assess trends in diarrhoea mortality."*

Examples of plausibility objectives may include, for example:

- "e) *To relate trends in diarrhoea mortality and hospital morbidity to the availability of ORS and to ORT coverage*
- f) *To relate the geographical distribution of diarrhoea mortality to levels of ORT coverage.*
- g) *To attempt to rule out external factors that may account for these observed associations".*

Specific hypotheses should be related to each objective. They may be stated either in terms of general trends ("There has been a substantial improvement...") or, whenever possible, in terms of specific NCDDP goals ("The objective has been reached"). The specific programme expectations - such as those listed in Box 2, Step 1 - should form the basis for evaluation hypotheses. If programme activities have focused in a particular geographical area, for example, the hypothesis might posit that performance improvements or mortality reductions will be greater in this area than in a comparable area for which data are also available. Similarly, the hypothesis might suggest that health centres in areas where supervisors have been trained might have more consistent stocks of ORT than health centres in areas with untrained supervisors.

As an example, the hypotheses in the Brazilian evaluation were:

- "a) *There have been substantial improvements (and/or the NCDDP goals have been met) regarding the number of ORS packets produced or procured; health services staff trained in diarrhoea case-management; educational materials produced and distributed; and mass media campaigns.*
- b) *There has been a major increase in ORT coverage.*
- c) *Diarrhoea-related hospital morbidity has declined significantly.*

- d) *There has been an important reduction in diarrhoea mortality.*
- e) *The decline in diarrhoea mortality and hospital morbidity coincided with the growing availability of ORS and with increased ORT coverage.*
- f) *States or provinces with higher levels of ORT coverage had lower diarrhoea mortality rates.*
- g) *External factors do not seem to account for the latter two associations".*

As a part of hypothesis generation, the evaluation team should clearly specify the time period for the evaluation of each hypothesis and the justification for this period. Factors that may be important in this decision include when a national policy was adopted, when the NCDDP was established or specific activities initiated, or the periods for which data (both for baseline and post-intervention) are available. For example, in Brazil the diarrhoea control activities were started around 1982-83, and at the time of the evaluation mortality data were available for 1980 to 1989; the evaluation exercise was therefore limited to the decade of the 1980s.

Carefully stated, well-defined objectives and hypotheses will help guide data collection and analysis. In turn, the availability of data will help define which objectives may be met and which hypotheses may be tested, generating an interactive feedback cycle which is further discussed in Step 5.

Step 4. Assemble the required data.

Once the objectives and hypotheses have been defined, the evaluation team should identify the key variables that will be needed to test each hypothesis. Performance and impact evaluations will be discussed separately as they often draw on different data sources and pose different measurement challenges. The need for data on variables not directly related to the programme - including quality control, denominator and external variables - will also be discussed. Note that, while examples are provided below, they are not intended as a comprehensive list.

At this stage, it will also be necessary to decide whether additional data collection - as opposed to using existing data - is required. In most cases, collection of new data through field surveys will fall outside the scope and timetable of NCDDP impact evaluations. However, in some instances focused health facility surveys to examine changes in case management or hospital record surveys to document reductions in diarrhoea admissions may be warranted. Both in the Brazilian and Philippine evaluations, it was deemed desirable to carry out rapid retrospective surveys of hospital records which are discussed under the topic "Diarrhoea severity".

A thorough assessment of the quality of the data used in the evaluation is required. Indicator-specific data quality issues are discussed below. In addition, Box 3 gives some guidance on what to look for when assessing the quality of routinely collected data.

Box 3. Checking the quality of routinely collected data

Carefully check the routinely collected information and interview staff involved in its collection and analysis. Whenever possible, outsiders with a critical knowledge of the information system such as University researchers or staff from international organizations (WHO, UNICEF, etc) should also be interviewed. Some of the data characteristics causing concern include:

- Lack of a well-defined population base (catchment area) for the information being provided
- Low coverage of the data collecting system within the catchment area, leading to under-registration of events
- Changes in event definitions during the study period
- Implausible levels of indicators relative to the level of development of the region
- Unusual fluctuations from year to year, that may suggest poor data collection
- Gaps in time series, due to lack of data collection in some periods
- Evidence of increases or decreases in registration levels, leading to spurious trends
- Trends limited to a single area or subpopulation, in the absence of a reasonable explanation

Indicators and data sources for programme performance

The initial programme review described in Step 1 will have already provided information on existing indicators of services provision, utilization and coverage. Data sources for these will typically include National Programme Profiles, household surveys (HHS), health facility surveys (HFS), Demographic and Health Surveys (DHS), supervisory reports, stock records, etc. A list of the information available from HHSs and HFSs conducted using WHO guidelines is provided in Annex B. Whenever appropriate, rather than describing programme achievements in absolute numbers, it is useful to refer them to a denominator: for instance, x% of the governmental health facilities have an ORT corner, and y% of health workers have received training in case-management.

As an example, at an early stage in the Philippines evaluation, desirable key indicators of provision and utilization were listed, along with possible data sources and time frames (Table 4). While data on some indicators were known to be available, others were likely - but not certain - to become accessible to the evaluation team. The former are listed as "essential" indicators in tables 4 to 6 while the latter as described as "additional". The question marks in these tables refer to data sources that were known to be available but which, at that early stage, were not available to the researchers.

Table 4. Possible key indicators of services provision and utilization in the Philippines evaluation.

Essential indicators	Purpose	Source	Years
Number of ORS packets produced/imported	Availability of rehydration salts	NCDDP Private sector?	1980-93
Number and geographical distribution of health centres providing ORT (e.g. with ORT corner) and/or preventive activities	Accessibility of services	NCDDP	1980-93
Number of training courses in diarrhoea case-management and/or prevention	Quality of case-management and preventive activities	NCDDP	1980-93
Number of messages broadcast through the mass media, posters, etc	Intensity of educational activities through the mass media	NCDDP	1980-93
Number of diarrhoea prevention education sessions held	Availability of educational activities in health centres	NCDDP	1980-93
Total NCDDP budget	Funds allocated to the programme	NCDDP	1980-93
Number of ORT packets distributed in health centres and commercial outlets	Utilization of packets	NCDDP Private sector?	1980-93
Number of children underfive attending health facilities due to diarrhoea	Utilization of health services	HIS-DOH* NCDDP Private sector?	1980-93
Number of mothers attending diarrhoea prevention education meetings	Utilization of educational activities	NCDDP	1980-93
Knowledge and practices of health workers regarding case-management and prevention	Quality of case-management and preventive activities	Health facility surveys in selected provinces	1989-91
Proportion of diarrhoea attenders who receive ORT, antibiotics, education on prevention	Quality of case-management and preventive activities	Health facility surveys in selected provinces	1989-91
Knowledge and practices of hospital doctors regarding case-management	Quality of case-management in hospitals	Selected hospital surveys	1990-91
Additional indicators	Purpose	Source	Years
Number of mothers watching or listening to mass media campaigns	Utilization of mass media campaigns	TV/radio audience polls	?

* Health Intelligence Service, Department of Health

National CDD staff will also be able to provide information on existing coverage indicators based on household and demographic surveys. The desired indicators for the Philippines are summarized in Table 5.

Table 5. Possible key indicators of coverage in the Philippines evaluation.

Essential indicators	Purpose	Source	Years
Proportion of children with diarrhoea in the past two weeks who received ORS/RHF/increased liquids/continued feeding	Coverage of correct case-management	NDS/DHS* ORT survey**	1983 1988 1989 1993
Proportion of mothers who are aware of ORS/ORT	Coverage of case-management education	NDS/DHS	1983 1988 1993
Proportion of mothers who have ever used ORT	Coverage of case-management education	NDS/DHS	1983 1988 1993
Additional indicators	Purpose	Source	Years
Proportion of mothers who know the correct preparation and administration of ORS	Coverage of case-management educational activities	Household surveys?	?
Proportion of children with diarrhoea who receive adequate ORT (preparation and volume)	Coverage of correct case-management	Household surveys?	?
Proportion of mothers who know danger signs for care-seeking	Coverage of case-management educational activities	Household surveys?	?
Proportion of mothers who know basic preventive measures	Coverage of preventive educational activities	Household surveys?	?

* National Demographic Survey/Demographic and Health Survey

** Department of Health, 1989 ORT and Feeding Practices Survey

Assessing the quality of management is an essential part of NCDDP evaluations. Table 4 included four quality indicators included in the Philippines evaluation, and several of the coverage indicators listed in Table 5 also refer to management quality. Additional quality indicators may include the proportion of diarrhoea cases at health facilities who are a) correctly assessed, b) correctly rehydrated (orally or IV) and c) correctly managed; proportion of dysentery cases treated with antibiotics; proportion of mothers given correct and sufficient advice; etc. These indicators should be included whenever possible.

Data quality issues that often arise in relation to coverage indicators include:

- **Problem:** Non-representative, geographically delimited or convenience samples have been used.

Possible solutions: Carefully document the sampling procedures used. Unless these have resulted in comparable samples, this problem will limit the comparison of results from different surveys.

- **Problem:** Data for key variables are not comparable. For example, for variables related to ORT use, care must be taken to ensure that the same types of fluid were characterized as ORT in the different surveys.

Possible solutions: Ideally, information should be available separately on use of ORS and of recommended home fluids (RHF), on increased intake of liquids and on continued feeding. If the original computer files are available, existing survey data may have to be reanalyzed to ensure that consistent definitions have been used.

- **Problem:** Sample size is not sufficient for disaggregating data for subnational units. Surveys with adequate sample sizes at the national level may fail to provide precise data at state or province level that would be required for testing ecological correlations. For example, the 1993 Philippines NDS/DHS had a national sample of 8,500 children, of whom about 10% had diarrhoea in the two weeks preceding the interview. This resulted in an average of 65 children with diarrhoea for each of the country's 13 regions, a sample which will give a margin of error of plus or minus 12 percentage points if ORT use is 50%^c. This margin of error is probably too large for meaningful comparisons.

Possible solutions: Group together neighbouring states, provinces or regions so that each analytical unit includes, say, 200 children with diarrhoea. This has its drawbacks, however, since dissimilar administrative units may have to be grouped and the resulting number of units may be too small for statistical analysis.

Indicators and data sources for programme impact

The measurement of diarrhoea morbidity and mortality is challenging for the many reasons that will be discussed below. These indicators will be grouped in three broad categories: diarrhoea incidence, severity and mortality.

^c Assuming a design effect due to cluster sampling of 1.0 (which was the actual finding of the Philippines survey) and a 95% confidence level.

Diarrhoea incidence

As discussed in Section III, one should only expect substantial changes in this indicator as a result of the NCDDP if strong emphasis has been given to preventive activities. In addition, incidence must be measured through population-based surveys and documentation of time trends requires at least two comparable surveys in the same population. Methodological difficulties, with their proposed solutions, include:

- **Problem:** The surveys may have used different definitions of diarrhoea, varying recall periods or inconsistent wordings for the questions. Minor changes in definitions or in the words used may lead to completely different measurements.

Possible solutions: Ensure that the same definitions and wording have been used. If possible, reanalyze survey data with uniform definitions and recall periods.

- **Problem:** The surveys may have been carried out in different seasons. Even if this did not occur, variability in average temperatures or rainfall from one year to another may lead to poor comparability.

Possible solutions: Carefully document the climatic conditions at the time of the surveys. Using surveillance data, assess the magnitude of seasonal effects and adjust for these in the analysis.

Due to the above difficulties, most countries are likely to find that assessment of the impact of the NCDDP on diarrhoea incidence is not warranted.

Diarrhoea severity

There are several ways to assess diarrhoea severity. Possible indicators include changes in the number and proportion of hospitalizations due to diarrhoea; changes in the degree of dehydration or in episode duration seen among in- and outpatients; and hospital case-fatality rates. Sources of data on these indicators include household surveys, routine disease surveillance data, clinic and hospital records.

In most countries, data on hospital admissions are not routinely collected at national level. In both the Brazilian and Philippines evaluations, it was thought worthwhile to review the records of major hospitals to assess a possible trend in diarrhoea admissions, as well as changes in the characteristics of the episodes, namely a reduction in the proportion of acute watery diarrhoea. The methods of the Brazilian hospital study are summarized in Box 4.

Box 4. Hospital records review in Brazil.

Reliable data on trends in hospital admissions by cause were not routinely available in Brazil. In each of the nine state capitals in the Northeast region, the largest public pediatric hospital was visited by a physician or registered nurse who reviewed systematic samples of at least 360 records of children under five, for each year from 1980 to 1990. The annual proportions of admissions due to diarrhoea were calculated for each hospital. In four hospitals the case-notes for earlier years had been discarded so that shorter time series were obtained (starting in 1981, 1982, 1984 and 1985). Linear regression techniques were then used to extrapolate the data to earlier years. When data from all hospitals were pooled together, the proportion of admissions due to diarrhoea fell from 57% of infant hospitalizations in 1980 to 30% in 1990.

A more detailed study covered children admitted in the beginning of the decade (1980 or the earliest available year) and in 1990. The case-notes of systematic samples of children admitted with diarrhoea were reviewed, totalling 2,224 admissions in the nine states. Information was collected on several variables related to characteristics of the diarrhoea episode and its management. In each state, the possibility of changes in the catchment population of the selected hospital was investigated, but this did not appear to have occurred.

Methodological problems with obtaining information on diarrhoea severity from household surveys are the same as those discussed above under diarrhoea incidence. Obtaining information from clinic and hospital records also have their own limitations, including:

- **Problem:** Hospital cases may not be representative of the severe episodes in the community. Distance and socioeconomic factors may influence the likelihood of admission.

Suggestions: Estimate what proportion of severe cases end up in hospital. A rough guess may be provided by the proportion of diarrhoea deaths occurring at home, but this statistic may also be biased since these deaths are less likely to be reported. However, even if hospital cases are a biased sample, they may still be useful for looking at time trends.

- **Problem:** The hospital's catchment population may have changed over time. Other nearby hospitals may have been opened or closed, or there have been changes in the fees being charged or in health insurance coverage. There may have been major population changes in the area due to migration.

Suggestions: Use information from health officials and hospital administrators to ensure that the catchment population has remained stable.

- **Problem:** There may have been changes in diagnostic criteria over time. This may be due to changes in medical practice (for example, stating that malnutrition is the underlying cause of admission in children with both diarrhoea and malnutrition) or to economic/ administrative

reasons (for example, changes in the amount reimbursed by national insurance for an admission due to diarrhoea relative to other causes).

Suggestions: Interview pediatricians to check that diagnostic criteria have been consistent in the time period, and administrators to assess economic/administrative changes.

- Problem: Clinics prescribing "only ORT" for diarrhoea episodes may lead families to shift to private practitioners who are more likely to prescribe drugs.

Suggestions: Check existing survey data to verify the proportion of diarrhoea episodes treated in the private sector (in the Philippines this was about 20%). Interview key informants to check whether such a shift may have happened. The main concern is not that a proportion of the children are being taken to private practitioners, but that this proportion may have changed substantially in the time period under study and therefore affected the interpretation of time trends. This is unlikely to affect hospital admissions, however.

When looking for health facility-based data, the evaluation team should keep its eyes open for the possibility of using existing data that may have been collected for other purposes. In the Philippines, the NCDDP collected information from selected hospitals on case-fatality rates before and after establishment of ORT corners or diarrhoea training units, as Box 5 shows. This type of information may be useful for building up the plausibility of an impact of the programme.

Box 5. Reduced diarrhoea case-fatality in Filipino hospitals

Six hospitals reported case-fatality data before and after establishment of ORT corners or diarrhoea training units. Five of these showed reductions ranging from 50% to 75%, while in the sixth there was a 65% increase. The median case-fatality before the intervention was 3.3% against 1.1% after the intervention. These results are encouraging but a possible limitation is that hospitals with marked reductions may be more likely to report their data than those with no change or a worsening situation. It would be useful to collect data on a random sample of hospitals or according to some pre-defined criteria, such as the largest hospital in each region.

Diarrhoea mortality

Reduction in mortality due to diarrhoea is likely to be the major impact indicator in NCDDP evaluations. Special care is therefore needed to ensure that the data are properly collected, analyzed and interpreted. The major source of such data are vital registration statistics. In some countries, mortality data may be available from sentinel sites or from repeated surveys such as DHS. Such data, however, are seldom broken down by cause of death which means that although they may be useful for examining overall trends and checking the quality of the registration system, they will not provide

information on diarrhoea-specific mortality^d. It is recommended that evaluation teams analyze the data separately for infants and for children aged 1-4 years, to check whether similar trends are present in both age groups.

The main shortcomings of using data from vital statistics include:

- **Problem:** A substantial proportion of the deaths are not registered. Under-registration is common in most less-developed countries, with official death rates often being less than half the actual rates. Registration tends to be particularly low for early deaths (e.g., neonatal), for those occurring in rural areas and for those among the poorest. A study in Brazil showed that diarrhoea deaths were more likely to be under-reported than those due to other causes. The coverage of the death registration system may also have changed during the study period.

Suggestions: Compare official mortality levels with those generated from censuses or representative surveys (e.g., DHS) through indirect methods or birth histories, to gauge the extent of under-registration. Ensure that roughly similar declines are observed from all sources. In addition, examine trends in the total number of deaths at all ages since any major fluctuations may reveal instability of the data collection system. Use proportionate mortality statistics (for example, proportion of infant diarrhoea deaths over all infant deaths; or proportion of infant diarrhoea deaths over deaths due to all causes at all ages) as this will reduce registration biases^e. Compare trends in diarrhoea deaths with those in causes which are less likely to have changed, for example, malformations or other perinatal causes^f. Under-registration per se is not an impediment for using vital statistics to examine trends, so long as the coverage of the registration system remains fairly constant.

- **Problem:** Criteria for the attribution of cause of death may have changed. Improved completion of death certificates and diagnostic procedures, for example, may shift many deaths previously attributed to septicaemia or to ill-defined causes to diarrhoea. Likewise, malnutrition may be more often recognized as the underlying cause of death, encompassing some of the deaths formerly ascribed to diarrhoea. Changes in coding of the underlying cause of death at the central health statistics level may also have occurred.

^d Although some surveys have used verbal autopsy techniques to assess the frequency of causes of death, they have two main shortcomings. First, the specificity of the diagnosis of diarrhoea through verbal autopsy tends to be low. Second, the actual number of deaths due to each cause in these surveys is often small so that estimates tend to have low precision.

^e If diarrhoea deaths are less likely to be registered than those due to other causes, then improvements in registration coverage would tend to increase, not decrease, the proportionate diarrhoea mortality.

^f Note that in some countries neonatal death registration may have improved due to the safe motherhood initiative and the EPI focus on neonatal tetanus. MOH officials may be able to provide information on the intensity of these programmes.

Suggestions: Interview pediatricians and mortality statistics clerks about changes in criteria. Examine trends in deaths due to different causes, including septicaemia, malnutrition and ill-defined causes. Check whether trends in these causes could explain the observed increase or decrease in diarrhoea deaths.

- Problem: Infant mortality rates are calculated using the number of registered live births as the denominator. Since under-registration also affects birth statistics, an additional source of error is introduced.

Suggestions: Use the population under one year of age obtained from censuses or surveys as the denominator for calculating the age-specific mortality rate for infants.

Because all types of health impact measures will have their limitations, it is desirable to include as many indicators as possible in the analyses. Table 6 shows the list of indicators to be collected in the Philippines evaluation.

Table 6. Possible key indicators of impact in the Philippines evaluation.

Essential indicators	Purpose	Source	Years
Trends in infant mortality due to diarrhoea and other major groups of causes (proportionate and with population denominator)	Assess mortality trends in infancy	HIS-DOH*	1975-93
Trends in child (1-4 years) mortality due to diarrhoea and other major groups of causes (proportionate and with population denominator)	Assess mortality trends in 1-4 year-olds	HIS-DOH*	1975-93
Trends in proportions of hospital admissions due to diarrhoea and other major groups of causes among infants in selected hospitals	Assess trends in hospital admissions in infancy	Hospital record reviews	1980-93
Trends in proportions of hospital admissions due to diarrhoea and other major groups of causes among 1-4 year olds in selected hospitals	Assess trends in hospital admissions in 1-4 year-olds	Hospital record reviews	1980-93
Diarrhoea case-fatality in selected facilities before and after implementation of ORT/DTUs	Assess trends in case-fatality	Hospital record reviews	1980-93
Clinical patterns of diarrhoea among hospitalized children (acute watery/persistent/dysentery)	Assess changes in clinical patterns	Hospital record reviews	1980-93

* Health information system - Department of Health

Indicators and data sources for additional variables required

In addition to indicators of programme performance and impact, data are required for providing the denominator for rates, for data quality checks and for assessing changes in external factors.

Denominators

These will include the number of live births per year as well as - for the reasons discussed above - the population under one year of age. The number of births may be obtained from the vital registration statistics while for the number of under-ones it will be necessary to interpolate and/or extrapolate data from censuses and surveys. Proportionate mortality calculations require the number of deaths due to all causes in infancy, childhood and at all ages.

Data quality checks

To estimate the completeness and stability of vital registration, one should gather data on the results of censuses or surveys that have provided indirect or birth history based estimates of mortality levels and trends. A complete picture of official mortality statistics should be obtained, including the absolute number and rates of deaths at all ages, of all infant and childhood deaths and of those due to major groups of causes, specifically including perinatal causes, malformations, septicaemia, malnutrition and ill-defined causes.

External factors

The factors to be investigated should be limited to those hypothesized to affect the specific performance and impact indicators under study. External factors are often country or region-specific. They may include changes in the structure of the health services or payment policies; efforts by other programmes that may increase utilization or affect care-seeking; political and economic factors affecting the motivation or working habits of health workers; etc. Information on these factors should be obtained through interviews with key informants in the health sector.

Other factors that may explain decreases in diarrhoea mortality or morbidity may include changes in socioeconomic (income, employment, education) or environmental (water, sanitation) variables; breastfeeding rates; immunization (particularly measles) coverage; anthropometric and vitamin A status. It is desirable to have comparable information on these variables both at the beginning and at the end of the study period. For variables such as socioeconomic status, the evaluation team should give preference to those most sensitive to the distributional shift at the high-risk end of the scale (e.g., use the proportion of families below the poverty line rather than the mean income level).

These variables will be typically obtained from national censuses and representative surveys such as DHS. Since several of these indicators are population-based, the same methodological problems and possible solutions discussed under "coverage indicators" are applicable.

Step 5. Redefine the evaluation hypotheses

The careful examination of the nature and quality of data available for testing the original hypotheses is likely to show that some of these cannot be addressed, and perhaps suggest additional hypotheses for the evaluation. At this stage, therefore, the initial hypotheses may be redefined.

Step 6. Analyze the data

The data obtained in the previous steps should be summarized, related to each evaluation hypothesis, and an analysis plan should be drafted. An example, abstracted from a programme evaluation in Brazil, is presented in Annex C³.

Two steps of the analysis should be separated, since they present different degrees of complexity. These are the adequacy and plausibility analyses.

Adequacy analyses

These analyses are straightforward. If there are absolute programmatic goals, such as those listed in Box 2 for the Philippines, then the first task is to assess whether or not the goal has been met. For example, the evaluation may show that the goal that 35% of mothers should know ORS and use it properly has been amply met.

Goals expressed as trends - for example, the 50% expected decrease in under-five mortality between 1987 and 1992 in the Philippines - require that the indicator be calculated for the whole time period, or at a minimum for its beginning and end. Special time series techniques, such as moving averages, may be used to smooth data which otherwise show wide fluctuations.

Statistical significance is not relevant to adequacy statements, although in some cases the calculations of confidence intervals for survey-based data may be useful.

Plausibility analyses

Plausibility statements are strengthened by parallel or consistent changes in data from different sources. Even descriptive time trends may be used for this purpose. For example, whenever possible it is desirable to obtain data series starting prior to the implementation of the NCDDP. If diarrhoea mortality is shown to have been stable in the pre-programmatic era and starts to drop after its implementation, this provides reasonable evidence for a real impact. The same applies for the demonstration of marked reductions in hospital case-fatality following implementation of ORT.

The case for plausibility would be strengthened by showing that external factors have not accounted for the mortality or morbidity changes observed in the study period. Box 6 includes a list of possible external (or confounding) factors that may affect mortality and morbidity rates.

Box 6. Possible external factors to be ruled out in a plausibility analysis

When analyzing the possible impact of NCDDPs, the effects of changes in the following external factors should be ruled out whenever possible:

External factor	Expected changes on diarrhoea mortality and morbidity
Socioeconomic factors (GNP per capita, family income, parental schooling, unemployment levels, etc)	Socioeconomic changes are inversely related to diarrhoea levels
Climatic factors (droughts, floods, etc)	Disasters may increase diarrhoea levels
Water supply and sanitation	Improved water and sanitation will tend to reduce diarrhoea
Breastfeeding duration and other feeding practices	Increases in breastfeeding duration will reduce diarrhoea incidence and severity
Prevalence of child malnutrition	Changes in nutritional status may reduce diarrhoea severity ⁸
Food/micronutrient supplementation programmes	Effective food or micronutrient (e.g., vitamin A) supplementation may improve nutrition and reduce diarrhoea severity
Immunization coverages	Immunization against measles will reduce severe diarrhoea cases

In the Brazil evaluation⁴, information on the prevalence of 12 possible confounding factors was obtained for the pre- and post-intervention period. Five of these showed substantial changes, including indicators of poverty, water supply, breastfeeding, measles vaccination and malnutrition. Earlier research in that country had provided relative risks of diarrhoea mortality associated with each of these factors, so that the impact of the observed changes could be modelled mathematically. Other countries may or may not have local information on relative risks for this simulation exercise, but if these are not available data from the international literature may be used.

⁸ Changes in malnutrition prevalences are not completely "external" to a CDD programme because these may result from improved diarrhoea management. A possible strategy is to analyze the data with and without adjustment for changes in malnutrition levels, and to compare both sets of results.

Ecological analyses may further strengthen plausibility. In these, levels of programme performance indicators - such as ORS distribution or ORT use rates - are correlated on a geographical basis to impact measures. In Brazil, this analysis was carried out using the nine states in the Northeast region; in the Philippines, the 13 regions in the country will be used. Regression and correlation techniques are the basic tools for the analysis of ecological studies. Since the number of observations (states, regions, etc) is usually small, nonparametric statistics such as Spearman's rank correlation coefficient should be used. Plausibility will be strengthened by showing that other confounding factors do not seem to be associated with the impact measure to the same degree as the programmatic variables, but available data on these factors may be limited. In most cases, formal control of confounding through multivariate regression analyses is not warranted since the number of observation units is small.

Step 7. Interpret the results

As for data analysis, the interpretation of adequacy results is reasonably straightforward. The original goal will or will not have been met, the observed trend will or will not be in the expected direction and magnitude. Box 7 shows the adequacy statements arising from the Brazilian evaluation. There were important increases in ORS distribution and in the ORT use rate. Impact indicators included major declines both in diarrhoea mortality and hospitalization rates, among infants as well as among children aged 1-4 years. The evidence of similar declines in two separately-collected impact indicators, both of them in two different age groups, is particularly reassuring.

As its name indicates, a plausibility statement is largely based on value judgments of experts in the field, including the evaluators. Statistical tests may and should be used, but their results can be overruled by judgement. For example, plausibility statements may be made even when the statistical power of the evaluation was low and the correlation coefficients were not significant. By pointing in the expected direction, these coefficients provide reassurance that the programme is moving along as desired.

The plausibility of a programme impact must be judged on the basis of the overall weight of the evidence. Box 8 shows some examples of research findings that will strengthen a plausibility statement.

In Brazil, changes in the external factors accounted for about half the decline in diarrhoea mortality, leaving the other half unexplained. In addition, there was a strong inverse ecological correlation between ORT use rates in the nine states and diarrhoea mortality (Box 7). The weight of the evidence here supports the case for a real impact of the programme.

Box 7. Summary of the Brazil evaluation.

Adequacy statements:

Provision/utilization: Increase in ORS distribution

Coverage: Increase in ORT use rate

Impact: Declining diarrhoea mortality and hospitalization rates

Plausibility statements:

Impact: External factors do not seem to account for observed decline in mortality

Inverse ecological association between ORT use and mortality

Box 8. Research findings that would strengthen the plausibility of the impact of a NCDDP

These are some examples of findings that will support a plausibility statement:

- Coincidence of increased programme activities with start of decline in diarrhoea mortality or severe morbidity (or acceleration of their rates of reduction)
- Coincidence of periods of reduced programme activities with increased diarrhoea mortality or severe morbidity (or deceleration of their rates of reduction)
- The observed declines cannot be explained by changes in external factors
- Faster rates of decline (or lower levels of morbidity/mortality) in areas with more intense programmes, if possible after ruling out confounding factors
- Reduction in case-fatality in facilities offering improved management
- Programme acceptors (e.g. infants whose mothers know/use ORT) show lower mortality rates than non-acceptors, if possible after controlling confounding factors
- Mortality rates in non-acceptors are similar to those in pre-programme periods

Note that the last two statements require data from field studies at the individual level that may not be available in many countries.

There are a number of methodological issues which are likely to be faced when interpreting evaluation results. These include:

- The feasibility of initial goals. If a goal has not been achieved, one may ask whether this represents a true failure or if the goal may have been unrealistic. For example, 100% goals are often stated but may be almost impossible to reach; in this case, is 95% a failure? Careful interpretation of the programme achievements in relation to the feasibility of initial goals is required.
- The possible effect of population heterogeneity. If ORT use is high among severe cases then a) a large impact may be observed even when overall ORT use is low, since those most at need will be receiving it anyway; b) further increases may not lead to mortality decline; and c) higher use rates may not be associated on a geographical basis with lower mortality levels. Whenever possible, information on ORT use rates should be stratified by disease severity

indicators, even if only the reported severity (mild/intermediate/severe) as perceived by the child's mother.

- The problem of competing mortality risks. Strong implementation of other health programmes - for example, EPI, ARI or safe motherhood - may lead to decreases in other causes of death and thereby reduce or eliminate decreasing trends in proportionate diarrhoea mortality. Computer simulations may help partly overcome this problem.

Step 8. Write the final report and disseminate the findings

A final report should be written stating in detail the background, objectives and hypotheses of the evaluation. The indicators and their sources should be described, as well as any additional data collection required. The main results should be discussed and interpreted, and the analytical methods characterized.

The final report should be circulated to government and international bodies, including donors. The national scientific community may be reached by publishing papers in medical or public health journals. However, it is just as important to ensure that the evaluation results are returned to the population. For this purpose, the evaluation team should use the press and other mass media, particularly the radio which has a large audience in many less developed countries.

REFERENCES

1. Division of Diarrhoeal and Acute Respiratory Disease Control. Guidelines for Conducting a Focused Programme Review (FPR). Geneva: WHO, July 1992 (draft).
2. Habicht JP, Victora CG, Vaughan JP. A framework for linking evaluation needs to design choices, with special reference to the fields of health and nutrition. UNICEF Evaluation Working Papers 1995 (in press).
3. Victora CG. Recent trends in diarrhoea mortality in Northeastern Brazil. Final Report to WHO/CDR, 1994.
4. Victora CG, Olinto MTA, Nobre LC, Barros FC. The recent fall in diarrhoea mortality in Northeastern Brazil: did ORT play a role? Health Policy and Planning (in press).