

Strengthening Monitoring and Evaluation in the World Bank's Nutrition Portfolio

REPORT AND RECOMMENDATIONS
FROM A WORLD BANK MEETING

May 2003

STRENGTHENING MONITORING AND EVALUATION IN THE WORLD BANK'S NUTRITION PORTFOLIO

Report and Recommendations from a World Bank Meeting

Meera Shekar and Amanda Liddle

May 2003

Strengthening Monitoring and Evaluation in the World Bank's Nutrition Portfolio

Meera Shekar^a and Amanda J. Liddle^b

^aSenior Nutrition Specialist, Health, Nutrition and Population, The Human Development Network, The World Bank, Washington, DC, USA

^bConsultant, Health, Nutrition and Population, The Human Development Network, The World Bank, Washington, DC, USA

Report and Recommendations from a World Bank Meeting,
World Bank, Washington DC, USA. May 21-22, 2003

Abstract: Recent developments at the World Bank towards a greater emphasis on results based management, on achieving the Millennium Development Goals (MDGs), and a shift away from vertical programs towards programmatic lending and poverty reduction strategy credits (PRSCs) provided a backdrop to this discussion meeting on monitoring and evaluation for nutrition. The meeting sought to understand the key issues constraining monitoring and evaluation in Bank-supported projects in countries, and to use this as a basis for development of a larger monitoring and evaluation framework for nutrition investments within the Bank. The issues were explored through a series of 'clinics' with country-level partners and World Bank Task Team Leaders from Bangladesh, Madagascar, Eritrea and Mauritania, to resolve specific questions/issues in these countries. This was supplemented with sessions on key monitoring and evaluation issues through presentations and case studies (i.e., 'learning-by-doing'). The clinics resolved specific issues for each of the country cases, but also provided the context for understanding the Bank's monitoring and evaluation work. Key questions that evolved during the meeting were addressed in the discussions and eight key recommendations were identified for strengthening the Bank's nutrition investments. Among other recommendations, it was agreed that the Bank must continue the process started in this meeting towards developing a clear strategy to guide monitoring and evaluation in future nutrition investments.

Keywords: monitoring and evaluation, nutrition, Millennium Development Goals (MDGs), Bangladesh.

Disclaimer: The findings, interpretations and conclusions expressed in the report are entirely those of the authors, and do not represent the views of the World Bank, its Executive Directors, or the countries they represent.

Correspondence Details: Meera Shekar, The World Bank, Mail Stop G7-701, 1818 H Street N.W., Washington, DC 20433. Tel: 202-473-6029; Fax: 202-522-3489; E-mail: mshekar@worldbank.org

This report, and the presentations, are available on the following website:
<http://www.worldbank.org/nutrition>

TABLE OF CONTENTS

Foreword	vii
Acknowledgements	ix
Executive Summary	xi
INTRODUCTION	1
PART 1: MONITORING AND EVALUATION - CONCEPTS	2
A. Choosing Evaluation Designs to Suit the Evaluation Questions	2
B. Monitoring as a Complement to Evaluation.....	3
C. Benefit Incidence Analyses.....	4
D. Developing In-Country Capacity for Monitoring and evaluation: The Challenge and the Way Forward (An example from the DANIDA-supported Community-Based Nutrition Program in Kenya).....	5
PART 2: MONITORING AND EVALUATION – ISSUES IN DEPTH	6
Clinic 1: The Bangladesh Integrated Nutrition Project (BINP).....	6
Clinic 2: The Community Nutrition Project (NUTRICOM), Mauritania.....	11
Clinic 3: The Use of Monitoring (and Evaluation) Data to Measure Impact and to Plan Future Programs- using the case example: Nutrition II, Madagascar	12
Clinic 4: Sampling Issues - Using the Case Example: Early Childhood Development Project (IECD), Eritrea.....	14
PART 3: THE WAY FORWARD: Discussions AND RECOMMENDATIONS	16
ANNEX 1: BACKGROUND PAPER	19
ANNEX 2: MINUTES FROM NUTRICOM PROJECT,	41
ANNEX 3: CONSULTATION AGENDA	43
ANNEX 4: CONSULTATION PARTICIPANTS	49

FOREWORD

Knowing what works, and why, and using this information to inform decisions and to stimulate greater investment in nutrition, is key to achieving several of the Millennium Development Goals (MDGs). A focus on monitoring and evaluation in our own nutrition work is therefore timely.

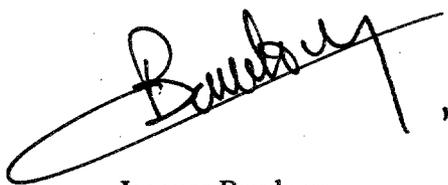
The focus on the Millennium Development Goals (MDGs) and the emphasis on development effectiveness have added to the impetus for strong monitoring and evaluation. Nutrition has a potential role in the achievement of at least four (and possibly six) of the eight MDGs. Potential support for achieving the MDGs hinges on a strong evidence base and reliable evaluation results from existing investments. The emerging shift in the nature of country support from the more traditional project instruments (that finance pre-defined activities within vertical projects) towards broader program or sector-wide support and new lending modalities, such as the Poverty Reduction Strategy Credits, add to the monitoring and evaluation challenge.

Earlier this year, Bank staff within Health, Nutrition and Population (HNP) and other sectors discussed new development strategies with the World Bank President, Mr. Wolfensohn, to redefine core functions and to outline a new vision for HNP. A strong recommendation from these discussions was to enhance HNP capacity to lead monitoring and evaluation activities.

This consultative meeting hosted by the World Bank was therefore timely. An expert panel reviewed specific monitoring and evaluation issues faced by Bank staff and country teams from Bangladesh, Madagascar, Eritrea, and Mauritania. The meeting re-visited the evaluation design and advised on the analysis plan for the Bangladesh Integrated Nutrition Project (BINP). Overall, the BINP evaluation plan was assessed by the expert panel as one among the few large scale nutrition projects world wide that meets the design pre-requisites for a sound evaluation, though some of the objectives are not addressed by the study. The panel also explored more generic issues and options, such as the critical role of monitoring and evaluation within nutrition, evaluation design features, the capacity needed for evaluation and for monitoring, how monitoring and evaluation could be conducted within the context of programmatic and sector-wide lending, and data ownership and dissemination of evaluation results. The meeting identified several key issues for Bank follow-up. We support the meeting's recommendation, among others, that monitoring and evaluation in existing nutrition investments should be strengthened, and that the Bank should consider building in explicit incentives for monitoring and evaluation, both with its own staff and with client countries.

Lessons and materials from the meeting will be incorporated into training modules on monitoring and evaluation for courses on nutrition to be developed jointly by the Human Development Network and the World Bank Institute.

We would like to thank the members of the expert panel on monitoring and evaluation for nutrition for their time and effort in helping the Bank think through these critical issues.



Jacques Baudouy
Director
Health, Nutrition, and Population (HNP)



Charles Griffin
Director
South Asia Sector Human Development (SASHD)

ACKNOWLEDGEMENTS

The meeting was conceptualized and prepared by Meera Shekar with inputs from Milla McLachlan of the Nutrition Cluster, Health, Nutrition and Population unit of the Human Development Network, World Bank. The consultation report was prepared and edited by Meera Shekar and Amanda Liddle, with inputs from Meri Vanharanta. Important contributions were made by the presenters and discussants at the meeting: David Coady (International Food Policy Research Institute), Ed Frongillo (Cornell University), Jean-Pierre Habicht (Cornell University), Nasreen Haq (Action Aid, Bangladesh), Rezaul Karim (The Institute for Nutrition and Food Science, University of Dhaka, Bangladesh), Barkat-e Khuda (International Centre for Diarrhoeal Disease Research, Bangladesh), Sascha Lamstein (Tufts University), Reynaldo Martorell (Emory University), Andrew Tomkins (Institute for Child Health, London, UK), representatives in Mauritania from the Ministry of Health, the NUTRICOM Project, UNFPA, UNICEF and WHO (via video conference), and by the following World Bank staff: Harold Alderman, Ritu Chhabra, Cherif Diallo and the Mauritania country team, Rae Galloway, Michele Gagnolati, Robert Hecht, Kees Kostermans, Milla McLachlan, Tawhid Nawaz, the Nutrition II Madagascar project task team, Christine Lao Pena, and Meera Shekar. Amanda Liddle coordinated and facilitated preparations for the meeting, assisted by Merced Doroteo. The graphics were formatted by Victoriano Arias. The final report was formatted by Gisele Biyoo.

The Nutrition Cluster gratefully acknowledges the collaboration of the World Bank Institute, particularly Abdo Yazbeck and Meri Vanharanta, in the recording of the interactive learning sessions and their potential use in future capacity development sessions.

The Netherlands Ministry of Foreign Affairs generously provided financial support, which is gratefully acknowledged.

The monitoring and evaluation expert panel consisted of:

Ed Frongillo (Cornell University)
Jean-Pierre Habicht (Cornell University)
Nasreen Haq (Action Aid, Bangladesh)
Reynaldo Martorell (Emory University)
Meera Shekar (World Bank)
Andrew Tomkins (Institute for Child Health, London, UK)
Howard Nial White (OED, World Bank)

Lessons and materials from this seminar will form the basis of a training module on monitoring and evaluation for a core course on nutrition to be developed by the Human Development Network and the World Bank Institute.

EXECUTIVE SUMMARY

Recent developments at the World Bank towards a greater emphasis on results based management, on achieving the Millennium Development Goals (MDGs), and a shift away from vertical programs towards programmatic lending and poverty reduction strategy credits (PRSCs) provided a backdrop to this meeting on monitoring and evaluation. The meeting sought to understand the key issues constraining monitoring and evaluation in Bank-supported projects, and to use this as a basis for development of a larger monitoring and evaluation framework for nutrition investments. The issues were explored through a series of clinics with country partners and World Bank Task Team Leaders (TTLs) from Bangladesh, Madagascar, Eritrea and Mauritania, supplemented with sessions on key monitoring and evaluation topics. The clinics resolved specific issues for each of the country cases, and provided the context for understanding the Bank's monitoring and evaluation work. The meeting specifically advised on the analysis plan for the Bangladesh Integrated Nutrition Project (BINP). Overall, the panel assessed the BINP evaluation plan as one among the few large-scale nutrition efforts worldwide that meets the design pre-requisites for a sound evaluation. The panel also noted, however, that the indicators captured in the study did not address some of BINP's objectives, and made several recommendations for strengthening the analyses.

Four questions were addressed. (i) what is the role of monitoring and evaluation in nutrition, and how can the Bank strengthen its commitment to evaluation? (ii) what capacities are needed for evaluation and how can this capacity be developed in client countries? (iii) how should monitoring and evaluation be positioned within the context of programmatic lending and PRSCs? (iv) who owns monitoring and evaluation data, how can reports be disseminated better, and how can data be made more available for follow-up analyses?

The meeting resulted in eight key recommendations. These were as follows:

The World Bank should

- (a) Strengthen monitoring and evaluation in all existing Bank nutrition investments, to contribute towards an evidence base for future investments; program and investments decisions at all levels must be based on the best available evidence. To achieve this, the Bank should consider explicit incentives for TTLs in the Bank, and for client countries.
- (b) Enhance capacities of Bank TTLs and client countries for monitoring and evaluation through training and "learning by doing" in existing operations. Partnerships with the World Bank Institute, and with academia and external institutions, both in the south and in the north, should be considered for this purpose.
- (c) Explore alternative and creative funding mechanisms for evaluation, to ensure that good quality evaluations are conducted effectively and timely.
- (d) Re-examine systems for strengthening data collection so that the interface between supervision reports, monitoring systems and evaluation data is strengthened, and most importantly the use of these data for decisions at all levels is facilitated.
- (e) Invest in a review of appropriate monitoring and evaluation methodologies and tools in the context of programmatic lending and PRSCs.
- (f) Publish and widely disseminate results from evaluations through national and international forums as well as among the communities from whom data were collected. While countries should continue to own their data, every effort must be made to make these data more available for further analyses.
- (g) Create and archive an evaluation database for the World Bank's nutrition investments.
- (h) Continue the process started in this meeting towards developing a clear strategy to guide future work on monitoring and evaluation for nutrition.

INTRODUCTION

In 2002, the World Bank engaged in approximately \$19.5 billion total lending (including \$8.1 billion IDA credits and 11.5 billion in IBRD lending). Roughly 18-20% of IDA credits represent grants. Of the total investments, health and social sectors together account for about 12-13%. Within that overall share, investments in nutrition are somewhere in the range of 10% of the total HNP investments. This suggests that the nutrition portfolio is a relatively small component of both IDA and IBRD investments. In 2002, the total Loan amounts committed to nutrition are approximately \$ 725 million across 38 countries (17 in Africa, 4 in East Asia/Pacific, 4 in Eastern and Central Europe, 8 in Latin America and Caribbean, 3 in Middle East and Northern Africa and 2 in South Asia). About 66 currently active projects claim to include some nutrition related activities, but only about 9 of these have more than 30% of total investments dedicated to direct nutrition interventions. Further, only 34 projects (including the above 9) have invested more than 10% of the total loan amounts in nutrition. Four new projects are expected to roll out this financial year, all with nutrition expenditures at less than 5% of total (and \$5 million cumulatively). Despite these relatively small investments in nutrition, in this era of greater attention to development effectiveness and competing demands for limited country and donor resources, monitoring and evaluation are essential tools for building a strong evidence base for all investments, including nutrition. A strong evidence base is the appropriate springboard for positioning advocacy for greater investments in nutrition.

As recently as 1999, project appraisal documents did not necessarily require key indicators to assess the impact of a project; however, the Bank is now increasingly focusing on change in outcomes (and not only on inputs or process variables), as measures of project success. In order for monitoring and evaluation to be used effectively, it is important to be realistic in setting goals and in selecting indicators for projects. Realism also requires careful consideration of the questions to be addressed by the evaluation, client and Bank capacity to measure impact, and overcoming barriers, such as, the misconception that the cost of monitoring and evaluation is prohibitive.

The World Bank convened a panel of nutrition monitoring and evaluation specialists at the Bank's Washington headquarters on May 21-22, 2003. The overall objectives of the meeting were to:

1. Review monitoring and evaluation issues within the World Bank's nutrition operations;
2. Identify opportunities for strengthening of monitoring and evaluation within the World Bank's nutrition portfolio, including development of in-country capacity for monitoring and evaluation;
3. Provide an interactive learning opportunity for the Bank's Task Team Leaders to strengthen monitoring and evaluation in their operations by presenting and discussing the issues faced in the field with the advisory panel; and
4. Provide inputs for a participatory learning/training monitoring and evaluation module for integration into the core course on nutrition to be developed in partnership between the nutrition team of the Human Development Network and the World Bank Institute.

The meeting first addressed key monitoring and evaluation concepts. It then considered various monitoring and evaluation issues in depth, often through "clinics" tailored to address constraints and issues from the World Bank's existing nutrition portfolio in countries. Most notably, the panel advised in-depth on the analysis plan for the evaluation of the Bangladesh Integrated Nutrition Project (BINP). Other "clinics" included those for Mauritania, Eritrea and Madagascar. The meeting concluded by outlining the future direction of monitoring and evaluation within the Bank's nutrition operations. This report summarizes the meeting and identifies the key recommendations for follow-up.

PART 1: MONITORING AND EVALUATION - CONCEPTS

Note: All presentations are available on the following website:
<http://www.worldbank.org/nutrition>

A. CHOOSING EVALUATION DESIGNS TO SUIT THE EVALUATION QUESTIONS JEAN PIERRE HABICHT, CORNELL UNIVERSITY

Key 'Take Home' Messages:

What is the evaluation for? There are several considerations in designing an evaluation: who is the decision maker? what are the relevant questions? and what kinds of decisions will be made on the basis of the evaluation? If one cannot identify the question that will be answered, and the use for that answer, it is not an evaluation.

Who uses the evaluation? There should be consideration of who will make the decision relative to the evaluation findings, and who will implement the decision. This means that the decision-makers must be involved in the design and implementation of the evaluation, which is often very difficult because those who pay for the evaluation (often the funders), are usually different from those who will use the results. The Bank should consider how to make the link between those who pay for the evaluation, and those who actually should use the findings of the evaluation.

Regarding design: One type of design (i.e., probability, plausibility and adequacy), is not necessarily better than another – each has a different purpose and meaning. However, the best evaluations are those that suit the evaluation question. **Adequacy evaluations** are limited to describing whether or not the expected changes have taken place. Adequacy performance evaluations assess how well program activities have met the expected objectives, whereas adequacy impact evaluations assess whether health or behavior indicators have improved among program recipients. **Plausibility assessments** attempt to control for the influence of the external factors that might have caused the observed effects. This is done by choosing control groups before an evaluation is begun, or during the analysis of the data. **Probability evaluations** aim at ensuring that the observed difference between the program and control areas is not due to confounding, bias, or chance. These may require randomization. Plausibility designs are often regarded as the gold standard. However, this approach is seldom mandatory, or even feasible, for routine evaluations of program effectiveness.

The main objective of an evaluation is to generate (persuasive) results that can affect the decision to be made on issues such as, scaling-up, replicability and best-practice consensus. The specific purpose of the evaluation should determine the appropriate evaluation design.

(For more details on this presentation, please see background paper in Annex 1)

B. MONITORING AS A COMPLEMENT TO EVALUATION **REYNALDO MARTORELL, EMORY UNIVERSITY**

Key 'Take Home' Messages:

Monitoring and evaluation are two distinctly different activities, although data from one can complement the other. Monitoring is the on-going collection and review of information on project implementation and performance. It can be quantitative data from Management Information Systems (MIS) that addresses whether a project is proceeding as planned, or qualitative/investigative data which considers why implementation problems are occurring. Evaluation is an independent evaluation of project implementation and performance (baseline, mid-term, final survey).

Monitoring provides data for participants (but has no control group and only allows for adequacy evaluation). Monitoring results are often more 'favorable' than evaluation results. Evaluation surveys lack the full context necessary to 'tell the full story'. Together monitoring and evaluation data allow for comprehensive evaluations (plausibility).

Monitoring data provide the continuous, evolving context, disaggregated to the needed level, within which to interpret evaluation results, and within which key program management decisions can and need to be situated. It can also inform about the best timing for evaluation surveys (e.g., PRAF Honduras).

Monitoring and evaluation data can often differ. The reasons for 'better' results from monitoring are:

- worse off may be missed and not counted in the target population;
- providers may inflate progress; and
- memory loss can be a problem in recall survey questions.
- training and technical skills of evaluators are often better than those of providers (e.g., digit preference in anthropometry); and
- inconsistent definitions used by monitoring and evaluation and across time.

Monitoring is often made to be too complex and demanding for community level workers, and complex monitoring data has limited use for decision making. In general, low importance is given to monitoring and research units in developing countries, and the technical capacity of project units to analyze data, interpret results and to write reports is often low.

Improvements can be effected by:

- 'lean and mean' MIS, linked effectively to decision making at all levels;
- increased technical capacity and status of monitoring and research units; and
- fostered innovation through operations research.

The comments made after the presentations by Drs. Habicht and Martorell progressed along the following (edited) discussion line:

- While many people do not realize how important timeliness is for evaluation and for monitoring, the Bank does and it needs to contract people who also understand that.
- Constructs about supervision and monitoring are not currently well thought-out. Supervision should not be about policing, but should be an on-going understanding and dialogue. Monitoring data (limited number of key indicators) should be aggregated at appropriate levels, and used for

supervision, and for local diagnosis of problems. It is crucial to consider who will use these data at each level, and what it will be used for.

- While we need to separate monitoring and evaluation as they are very different, they also need to work together. For example, in Nicaragua there is an integral monitoring system (for a growth monitoring program, that is working well because it is used for supervision and identifying problems that need to be solved - a problem solving tool. However, in Bolivia a supervision firm has been hired. It is not working well because the people who need to make the changes are not involved in the monitoring. The overall management system and a sense of an overall learning organization are essential for monitoring to work.
- People often do not trust monitoring data because of the ‘fudge factor’ of those who report the monitoring data. This issue needs to be addressed. Furthermore, there is a need to monitor the monitoring system. Even in a good overall project, there are places where the monitoring is better than in other places (even within one country and one project/program).
- Currently, there is greater emphasis on the role of civil society in development, and community and civil society are users of evaluation information. Therefore, they should have greater access to evaluation findings. We should consider how civil society and the community could be included in future evaluations.

C. BENEFIT INCIDENCE ANALYSES **DAVID COADY, INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE**

Key ‘Take Home’ Messages:

Although benefit incidence analysis is a useful indicator of the overall incidence of public expenditures, it is severely constrained by lack of disaggregated data on the cost side. In addition, for policy purposes information on "marginal benefit incidence", i.e. the benefit incidence of changes in public expenditures, is much more useful. In this respect, there is no substitute for evaluating the benefit incidence of individual program components that make up total public expenditures.

The major points made following the presentation by Dr. Coady progressed along the following (edited) discussion line:

- How do you get the cost data? Is there a template for collecting cost data?
Programs rarely have cost information. We have tried to get detailed list of costs and we have tried to link it to the activities (for PROGRESA, PRAF, and RPS). We are trying to bring the three together. Often it is hard to find who the beneficiaries are.
- In addition to the cost issues outlined in the presentation, benefit incidence analyses present a useful technique for analyzing the data to answer questions such as who participates? Do the poorest participate more or less? Do the richest (or the poorest) income quintiles benefit more from the program? This is a similar methodology, but a slightly different context.

D. Developing In-Country Capacity for Monitoring and evaluation: The Challenge and the Way Forward (An example from the DANIDA-supported Community-Based Nutrition Program in Kenya)

Andrew Tomkins, Institute for Child Health, London, UK

Key 'Take Home' Messages:

The Community Based Nutrition Program in Kenya shows that with the right kind of support, communities are able to plan, initiate, monitor, and sustain community based nutrition projects, including securing their own funding, and integrating monitoring into the community's plans.

In this project, community capacity was built up through a four year visioning process and the project has shown that community capacity for planning, initiating, monitoring and sustaining the community-based interventions can be developed through this process. Issues of scaling-up, replication and sustainability are addressed by three-year 'beacon' projects that aim to move into other communities. In addition, the project has also seen an increasing awareness among satellite communities. The project will monitor the 'beacons' after a three-year period to assess whether capacities developed in these beacon communities are sustained, and whether nutrition improvements can be documented. While DANIDA provided support for the initiation of this project, the project has designed a program for different funding levels and is aiming to replicate this model in other parts of Kenya with support from other donors.

The major points made following the presentation by Dr. Tomkins addressed a larger examination of monitoring and evaluation issues. The (edited) discussion line progressed as follows:

- Empowering communities to trigger resources is essential (and it is important to bring in the lessons learned from other similar projects). It is necessary to have a government that accepts decentralization, has the political willingness to allow the process, and has some resources. This can often not be a problem in rural communities, provided there is a focus on nutrition security (instead of something very general). However, it is more difficult in urban areas.
- Ownership of the data being generated must be considered. It belongs to the communities, but could be used well by others as well.
- Questions:
The Bank is moving away from lending for specific sectoral activities. It is increasingly releasing funds into a pool either through the basket financing for SWAPs, Poverty Reduction Strategy Credits, or Sector Development programs:
 - What adaptations are being made for monitoring and evaluation for these new funding modalities?
 - What is, or will be, evaluated if the Bank makes loans that go into a big pool?
 - Social Funds have largely funded infrastructure in the past - is this a viable model for nutrition projects?

Responses:

- The concept behind sector-wide lending is to support the government's overall budget with rolling funding and goals that can be monitored. However, no program of this type has been evaluated for outcomes as yet. Methodologies for sector-wide evaluations remain to be developed.
- There will always be a mix of program types. The way forward is not clear, but Poverty Reduction Strategy Credits seem to be more and more popular, though some countries continue to have sector-specific investments, and some will continue with stand-alone

projects. Other bi-laterals such as DFID are also moving from specific health sector funding to budget support. This shift increases the urgency for being able to ‘make the case for nutrition’ to Ministries of Health and to Ministries of Finance. However, communities are often not organized or capable enough to do this, and social action funds often do not include nutrition.

- It is vital to make the case for nutrition from the Minister of Finance’s perspective – why would he/she invest in nutrition? This requires presentation of nutrition’s contribution to human development and to productivity. Sound evaluations on specific projects are crucial to provide the evidence needed to support these claims. However, care with indicators must be taken so that PRSP and MDG indicators are measured, so that any impact or achievements can be attributed appropriately.
 - The nutrition community is already taking evaluation seriously, but also needs to take the lending issues seriously.
- Next Steps:
 - In response to the changed lending approach the need for developing new and appropriate evaluation methodologies and approaches is urgent.
 - Continued dialogue on (i) who owns monitoring and evaluation data; and (ii) how we facilitate sharing, but not misuse, of these data, and (iii) conscious efforts for strengthening evaluation capacity and for building incentives for monitoring and evaluation.

PART 2: MONITORING AND EVALUATION – ISSUES IN DEPTH

The Clinics

CLINIC 1: THE BANGLADESH INTEGRATED NUTRITION PROJECT (BINP)

The primary objective of this clinic was to consult with the expert panel on the following issues:

1. Given the design and sampling methods employed for the Baseline, Mid-term and End-line evaluation surveys in Bangladesh, how can the BINP data be best analyzed to produce a credible and quality product?
2. Review and advise on the evaluation design for the follow-on National Nutrition Project (NNP).
3. Consider how nutrition evaluation capacity can be strengthened for the future in Bangladesh?

The first presentation in this clinic (by Rezaul Karim and Sascha Lamstein) laid out the design and preliminary results from the BINP. The second presentation (by Nasreen Haq) laid out the context in Bangladesh within which the BINP evaluation results need to be situated. These presentations were followed by extensive discussions on the strengths and weaknesses of the BINP evaluation design, and the best analytic strategies for these data. A summary of the agreed-upon analysis plan is presented at the end of this section. The last presentation in this clinic (by Barkat-e-Khuda) laid out the design for the baseline study for the follow-on National Nutrition Project (NNP) in Bangladesh.

Copies of each of the presentations are available in Annex 2.

The major points made following the presentations by the Bangladesh team progressed along the following (edited) discussion line:

A Summary of the discussion's on BINP end-line analyses:

- Design and analyses:
 - Overall, the BINP evaluation meets all design pre-requisites for a good evaluation, i.e., it is feasible, has control and treatment groups, pre and post data to account for secular trends, measured indicators, and adequate lag-time. However, there is some slippage of indicators vis-à-vis objectives. The control groups used in the study seem to have been 'contaminated' over the life of the project, thus diluting the size of program impact. It is therefore important to have external survey data and to construct a 'super-control' through opportunities, such as, the HKI nutritional surveillance data. These could help account for secular trends.
 - The timing of the evaluation in the Bangladesh Nutrition Program was correctly postdated as there were no reasons to expect any impact within a short intervention period. It was not necessary to have a probability design, but it is necessary to show plausibility. The report must consider how best to show plausibility.
 - For the evaluation to be meaningful, it is important to answer the following questions: What is the purpose of this evaluation? What decisions will be affected by this evaluation? and How would we have done this evaluation differently if certain issues had been known in advance? The value and use of evaluations must be thought through to be clear on what we hope to gain from them, and how the results will affect change in future programs and their evaluations. The monitoring and evaluation data quality should be such that decisions on program continuation or modification can be reliably made on this basis.
 - In the case of BINP, external validity (generalizability), of the results is not as important as internal validity, since cultural characteristics of each country define what works in the environment in question.
 - There is a need to present the BINP results in several different ways and not simply by presenting statistically significant changes with p-values < 0.05. The p-value does not tell the magnitude of the impact, and, in some cases, some changes that do not reach statistical significance may have important physiological implications. Additionally, due to the ceiling effect (only the underweight will benefit from energy supplementation), following the shift in weight or length distributions can be more informative than comparing baseline and end point values.
 - Sharing evaluation results with the public would increase their value generate important discussions and confidence towards World Bank operations by reducing suspicion. It would also encourage the Bank to analyze the results from different perspectives, which would give a more complete picture of the true affect. One possibility is to present the results in different 'envelopes' or assumptions about, for example, exclusion criteria, income and malnutrition levels.
 - What is the cost-effectiveness of BINP? (what are the effectiveness measures). The study should consider cost-effectiveness against what, and the hidden benefits (i.e., supervision).

- The data on the effectiveness of the intervention is there; it simply needs to be pulled together in a more convincing manner to deliver the message more persuasively.
- Consider what the Bank wants to do with the data? What can the data tell us about adjustments to the follow-on NNP project? Can NNP targets and objectives be reviewed in-line with the findings of this study?
- **Monitoring:**
 - Setting achievable goals is important and also gives incentives for monitoring.
 - The quality of the monitoring data should not be under-estimated, as it can sometimes be more accurate than evaluation data. An in-built monitoring system, consisting of key indicators, would allow us to act in a timely manner rather than having to wait three years for the midterm results. Therefore, it is critical to develop a methodology that includes monitoring in all programs.
 - Monitoring data can be used to comment on the quality of the intervention. This can be used to justify future programs (an example of how collecting crucial evidence pays back in the long run).
- **Sampling:**
 - For future programs, the differences between participating *‘thanas’* could be taken into account in sampling (e.g., to control for the similitude of subjects and controls). The basic differences can be taken into account by adjusting for them, e.g., in a regression model. This will reduce the need to include other variables in the model (as the power of the analysis goes down with an increasing number of variables included).
 - Matching controls (according to baseline values of selected variables), is recommended and reduces the number of controls needed. In some cases, it is essential to be able to randomize the treatment, though this was not necessary for BINP.
 - Every time people are asked to volunteer in the program the question ‘What do we know about the people who did not want to participate?’ is raised. Consider the possibility of a selection bias: are the people willing to participate somehow different from the ones who do not, and thus, more or less prone to show an impact of the program? Correct for bias through analytical techniques if needed.
 - Numbers of non-responders in the evaluation surveys should be reported on to add to the background information.

BINP - Endline Evaluation Analysis Plan of Action

Follow-up actions:

The most immediate use of the BINP evaluation results will be for decisions by the Government of Bangladesh, the World Bank and other donors, re: the need for fine-tuning the design of the existing National Nutrition Project (NNP), as well as a review of the targets and goals for the NNP. Data from this evaluation will also be used for making decisions about future investments in nutrition in Bangladesh - their nature, components, cost-effectiveness, and scope. Accordingly, and to make the analyses feasible, the meeting agreed on the following phased plan for analysis of the BINP data:

PHASE I:

In the first phase, the researchers will focus on:

- Validation of the sampling frame to ensure comparability across the baseline, mid-term and final surveys
- Construction of “super-controls” from secular trend data from HKI, or other such sources in Bangladesh, to add to the strength of the control group comparisons
- Review of seasonality issues related to weight-for age in Bangladesh

Analytic questions for the first phase analyses relate to:

- The impact of BINP – both summative analyses on what impact, as well as, the size of the impact
- To the extent feasible in a preliminary analysis, address issues of attribution of the observed impact to BINP or otherwise
- Who benefited from BINP (e.g., did the poorest benefit more or less? Did younger or older children benefit more? Did younger or older mothers benefit more or less?)
- Did those who we thought would benefit actually benefit?
- Is there a dose-response between those who participate more and those who benefit more?
- Can we say anything (preliminary) about cost effectiveness?

Results from these analyses will inform both the fine-tuning of the NNP, and formative research for future investments in Bangladesh.

PHASE II:

The second phase of the analyses will focus on more detail re the components of the BINP package:

- What components of BINP seem to have contributed to the observed impact more than others?
- What is the relative cost-effectiveness of each of these components?
- What are the implications for scaling-up such components? What components should be scaled up? What components seem to require additional operations research before decisions about scaling-up are made?

PHASE III:

In the third phase, the analyses will focus on bringing together the quantitative results with more qualitative information on key issues such as:

- Feasibility (and lessons) for evaluating and monitoring large scale programs
- Telling the complete story of BINP – the trials and tribulations

While the first phase analyses are ongoing, the second and third phases will require additional funding support, which could be sought from the World Bank South Asia team, the World Bank anchor, and/or bi-laterals and other international donors. The third phase may also require some additional qualitative data collection.

In order to facilitate these analyses, the data must be archived and made available for the analyses. Funds for this will be needed from the World Bank’s South Asia team. Technical support will be provided from the World Bank’s nutrition anchor team.

B Suggestions/recommendations for consideration in the NNP evaluation design:

- The most important recommendation for NNP was to ensure that some of the avoidable mistakes for BINP evaluation are not repeated in the case of NNP. Key among these mistakes is a focus on the design of individual surveys (baseline, midterm and end-line) rather than on the overall evaluation design. In the case of BINP, this led to different sample sizes for each survey, additions/deletions of key data from each survey, and some times the use of different survey tools/questionnaires. The use of multiple contractors for the evaluation added to the challenge of comparability of data across surveys.
- Other recommendations are as follows:
 - Regarding measurement:**
 - a) Children and adults – it is crucial to measure length for age;
 - b) Consider if it is appropriate to measure pregnancy weight in the last trimester and create a “normogram” for validation or correction of birth weight data;
 - c) The NNP survey is currently measuring children up to two years of age. It should measure up to five years of age in order to be able to capture the cumulative impact on growth; and
 - d) The usefulness of 24 hr recall data is limited. However, if it is used, decide what question the data will answer and ask specific sub-questions that will provide the necessary data to answer the main question.

Regarding sample size:

Sample size estimations are currently based on previous NNP targets. They should be revised to be based on the smallest difference in key outcomes (such as height and weight) of physiological significance that the project would aim to bring about.

Process variables:

Add qualitative methods to capture client perceptions about the program.

- Consider further discussion, in Bangladesh, of this panel’s suggestions for reviewing and updating the NNP evaluation plan.

CLINIC 2: THE COMMUNITY NUTRITION PROJECT (NUTRICOM), MAURITANIA
JP Habicht, Kees Kostermans, Meri Vanharanta with Cherif Diallo and the Mauritania country team

NOTE: This clinic was conducted via videoconference between Washington DC and Mauritania. See Annex 2 for the minutes of the videoconference prepared by Cherif Diallo.

Background and Issues:

Mauritania has a US\$4.9 million Learning and Innovations Loan (LIL) from the World Bank. The objective of this LIL is to learn and innovate for nutrition improvement on a small scale, with a view to scaling up at a later time. Implementation of the project has been considerably delayed, and the monitoring and evaluation component has been floundering. This clinic was set up with a large team of nutrition professionals in Mauritania (from the Ministry of Health, UNFPA, UNICEF and WHO), to streamline the monitoring and evaluation efforts for the project.

Monitoring and evaluation of the nutrition intervention is yet to be implemented for most areas. The main issues are i) a multitude of data on the 38 indicators is being collected; ii) data collection is not complete, and data are often not forwarded to higher levels in the system; and iii) data analysis is lacking.

The discussion with Dr. Habicht brought forth the following suggestions:

- 1) Think carefully about what information is needed to follow the progress and then reduce the number of indicators to very few key indicators. Start with just two indicators: the number of children seen and the number of children found malnourished. A small number of indicators would allow the team to start collecting the data on these parameters as soon as possible, and the remaining time would be sufficient to get the surveillance going in all areas. Once the collection and the delivery of the data are on going, additional indicators could be added if feasible. At the end of the program, we will, at least, be able to present the number of children reached, and the percentage of malnourished children attending the clinics at the beginning of the program in comparison with the percentage at the end.
- 2) Do not analyze the wealth of data on the 38 indicator-system for the moment.
- 3) Indicators to consider later are: (a) coverage data of the target population; and (b) data which show individual improvements of children enrolled in the program.
- 4) The delivery of the surveillance data (e.g., once/twice in two months) from the regional centers to the control center must be carefully planned.

General points:

As the national nutrition policy is currently under development in Mauritania, some examples of nutrition policies from other countries could be helpful.

Micro-credits were discussed, and the issue of earmarking government loans to nutrition was pointed out. Follow-up information should also be gathered from malnourished families given micro credits to find out how much is spent on additional food, if any.

It was evident that the Team in Mauritania had the solutions to most issues discussed, so the meeting served as a reinforcement for these ideas to be implemented in practice.

CLINIC 3: THE USE OF MONITORING (AND EVALUATION) DATA TO MEASURE IMPACT AND TO PLAN FUTURE PROGRAMS- USING THE CASE EXAMPLE: NUTRITION II, MADAGASCAR

Ritu Chhabra, Rae Galloway and Reynaldo Martorell

Clinic Summary:

The second Community Nutrition Project in Madagascar targets approximately 1 million children under 3 years, 2.5 million children between 3 and 14 years and 700,000 lactating women. The project has goals to reduce underweight, vitamin A deficiency, iron deficiency, and helminth infections, and improve community awareness and capacity to take action in the targeted populations.

The project has funded and implemented baseline and midterm surveys and an end-of-project survey is planned. It has designed a monitoring system, collecting data on a monthly basis throughout the project's life. A preliminary report from an ongoing review indicates that these data, available electronically, are of good quality. The follow-on project is currently under discussion and will use the lessons and findings from this project in its planning. The clinic extensively discussed the project's monitoring and evaluation status. The technical experts on evaluation assisting the team during the clinic were surprised and impressed with the level of attention given to monitoring and evaluation under the project. Yet, several problems were encountered by the project team in collecting and interpreting data for the project which were extensively discussed and commented upon by the technical experts:

- 1) The team expressed concern that the control areas were limited in number, that they were "contaminated" with non-BINP project inputs from other agencies and NGOs, and that they were selected arbitrarily, rather than randomly or as matched controls. The problems of being able to identify matched (or randomized) control areas in real-life programs was discussed and the technical experts felt that it is often difficult in a program setting to find areas that can act as "true" controls, since there are other parallel economic and development activities occurring in over time. It was recommended that the 1997 MDHS data or the HKI nutrition surveillance data be used as a national comparison, even though the MDHS data are not disaggregated to the district level.
- 2) Another issue of concern related to the target group for the evaluation. In the baseline survey all children under 5 years of age were part of the sample but in the end-line survey only data for the target group (under three) were captured. It was recommended that although the target group for the project is children under 3 years of age, since stunting is a cumulative indicator, it would have been useful to see the impact of the program on prevalence of stunting in children 4-5 years of age. However, the lack of data for older children in the final evaluation limits these analyses.
- 3) Although confounding factors can be dealt with in the statistical analysis, it is wise to take note of potential confounding factors that might be affecting the results.
- 4) On the issue of using monitoring data for evaluation purposes, even if the monitoring data are of good quality, these data are, by definition, limited to participants only. The team should, therefore, carefully consider self-selection and sample selection issues when using monitoring data to supplement or triangulate evaluation results.
- 5) The question of distinguishing between the impact of counselling versus food supplementation was discussed. Under the project there are communities that are receiving World Food Programme food for mothers and children along with counselling and other program inputs such as vitamin A, deworming, etc. while other communities are not receiving food but are receiving

all the other program inputs. It was suggested that there is an opportunity to determine the impact of different program inputs such as food, counselling and food, and counselling by itself:

- ⊖ community-based monitoring could examine whether mothers and health workers reported increase in knowledge and improvement in practices (e.g., understand the concept of gaining weight);
- ⊖ when a child is not gaining weight, it should be ascertained if the quantity, quality or frequency of feeding is an issue. Is the reason for the child not gaining weight because of lack of food in the community? Is the problem of children not gaining weight less of a problem where WFP is providing food? It was recommended that the body mass index (BMI) of the mother be taken to determine if the mother is underweight, as it should not be automatically assumed that the child is the one who needs the intervention.
- ⊖ the role of a multi-media campaign as a communication strategy to counter negative practices and beliefs should be evaluated as well as the effectiveness of counselling.

CLINIC 4: SAMPLING ISSUES - USING THE CASE EXAMPLE: EARLY CHILDHOOD DEVELOPMENT PROJECT (IECD), ERITREA

Ed Frongillo, Christine Lao Pena, Sascha Lamstein and Meera Shekar

The Eritrea Early Childhood Development Project (IECD) project has a component for improving Child and Maternal Nutrition. A sub-component, The Community Based Growth Monitoring and Promotion (GMP) Centers, is being phased-in and implemented slowly. It was agreed with the National Statistics Office (NSO) that of the 30 children listed in each of the GMP registers, a random sample of 10% would be an adequate representative sample from each of the GMP sites selected.

Issues:

- 1) For Monitoring:
 - i) Is it acceptable that the baseline data is collected by the implementers, not an independent agency?
 - ii) What is the correct sample size and sample composition?
 - iii) How should the project proceed, given the low capacity in the 'zobas' (regions)?
- 2) For Evaluation:
 - i) Is it reasonable to use DHS data as a baseline?
 - ii) Weight for age is measured for children and their nutritional status is presented to the community via the GMP. Can this be considered as the alternative baseline information rather than the DHS?
 - iii) There may not be comparable data for end of project evaluation.

Discussion Suggestions:

Two points need to be emphasized:

- i) Often sampling needs are complex (i.e., clustering), and that needs to be taken into account in the planning and analysis; and
- ii) The sample size has to be large enough to have adequate power for the effects of interest, or conversely, our expectations of effects have to be in line with what is feasible and affordable. This is a very real problem in program evaluations.

The government and the Bank should first discuss what information they both need to make decisions. Specifically, they should consider:

- i) What do the decision makers need to be able to move forward and make decisions?
- ii) Is it plausible to attribute any improvements to the project ?
- iii) Does anyone in the government or the Bank need to be convinced about growth monitoring and promotion, and community driven development if the framework for the project is community development and empowerment?

After these issues have been addressed, a feasible evaluation design can be considered.

The group suggested that:

- i) Since evaluation capacity in Eritrea may be limited, evaluation could be done with some external help. The project should check with the government/Ministry of Health as to which neighboring countries are acceptable collaborators.

- ii) Sample size calculations: With a 3:1 ratio for “experimental” vs “control” areas, a Standard Deviation of 1.2, a total sample size of approximately 600 (plus) can detect a 0.3 Z score change. However, this does not take into account any clustering one might decide to do, or any non-response. Therefore, a somewhat larger sample size should be considered.
- iii) Have a baseline and a final impact evaluation with good quality data (no mid-term since the time period is short). The baseline would need to include a household survey on socio-economic characteristics and knowledge (specifically regarding the behavior issues addressed by the project). However, some process evaluation mid-way through the project should also be considered.
- iv) The project should consider a ‘plausibility’ study in one region and an adequacy study in other regions. There is probably no need to go to all 9 regions; alternately one may wish to select one *Zoba* from each geographic region.

Regarding monitoring, the group suggested that the team:

- i) Consider who would obtain this information and how? It is very important to: a) have every piece of monitoring data linked to a trigger point for action; and b) actually use the information regularly (or collect the data as often as it would be used).
- ii) Collect the regular growth monitoring data from about 10% of the children enrolled in the program in each center. Aggregation of the data at regional and national levels could yield useful information.
- iii) Use the monitoring data, complemented with centre-level qualitative data on the quality of the program implementation (training of learners, regularity of growth monitoring and promotion sessions, availability of food supplements), because they could yield useful information.

Regarding evaluation the group suggested that:

- i) DHS data cannot be used as a baseline because of the timing of phasing-in of the project, and comparability of the DHS data collection areas to this project’s data collection areas.
- i) The project measures children under five because of the lag time for effect on growth.
- iii) The study should report on weight gain in grams and length gain in cms, as governments consider these measures give a more convincing argument than Z scores.

PART 3: THE WAY FORWARD: DISCUSSIONS AND RECOMMENDATIONS

The recommendations evolved from the following questions that developed as issues of importance during the meeting:

- 1 What is the role of monitoring and evaluation within nutrition, and within other sectors, and how can the Bank strengthen its commitment to evaluation?
- 2 What capacities are needed for evaluation, and how can this capacity development be addressed in developing countries?
- 3 How should monitoring and evaluation be positioned within the context of programmatic lending and PRSCs?
- 4 Who owns monitoring and evaluation data, how can evaluation reports be disseminated better, and how can data be made more available for follow-up analyses?

Many of the answers to these questions were discussed within the context of the strengths and constraints of how the Bank does business (including its changing lending practices from investment to programmatic), and in nutrition specific contexts.

Bank Context:

- Within the Bank, evaluation research is carried out through two mechanisms a) through the Operations Evaluation Department (OED), which is responsible for ‘feedback’ after project completion; and b) it is built into lending (2% of the borrowing is allocated for monitoring and evaluation). OED does not conduct any ex-ante evaluations. Therefore the challenge of much of the evaluation work lies within lending operations.
- The New Human Development Vision at the Bank emphasizes the role of monitoring and evaluation by focusing on issues such as: evidence based policy decisions, making learning and knowledge the heart of human development, better evaluation and monitoring at the household level, global support for monitoring and evaluation, global emphasis on developing a sound system of household surveys, making projects learning pilots, and helping clients learn from projects. The meeting was encouraged by, and supported, the emphasis on shifting lending portfolios towards investment in knowledge of what works and what can be scaled up through budget support.
- Although it is more difficult to make nutrition a priority within the PRSP approach than within vertical or sectoral programs, the group stressed the need to prioritize nutrition, as it’s potential contributions to poverty reductions are large. Documenting monitoring and evaluation lessons can assist this process.
- The current international focus on the Millennium Development Goals (MDGs) is fully embraced and integrated into the work of the Bank. The meeting extensively discussed nutrition monitoring and evaluation under the MDG umbrella. There was consensus that the MDGs provided the opportunity for re-positioning nutrition within the larger development agenda.

- There are many ways of achieving improvements in nutrition, and clarifications are needed in regard to the ‘nutrition MDGs’ and the role of nutrition interventions in achievement of the ‘other MDGs’.
- The evidence base for nutrition’s potential contribution to the MDGs must be generated from evaluations of large-scale nutrition projects. Since the World Bank is among the few agencies supporting large-scale nutrition projects, the role of evaluation becomes even more critical in the Bank’s work. Building this evidence base will also include, for example, more operations research on which direct nutrition interventions are cost effective and can be effectively scaled-up, and the role of growth monitoring and promotion as a tool and community organization for focusing or targeting nutrition relevant interventions. Additional evidence must also continue to be gathered on whether the inclusion of direct or in-direct nutrition investments (versus excluding these), would yield additional benefits. Data are needed to document whether such payoffs are limited to nutrition outcomes alone, or whether those go beyond nutrition to impact on other development outcomes (health, education, poverty). These are important considerations for monitoring and evaluation in the context of the strong focus on the MDGs and on programmatic lending and Poverty Reduction Strategies.

Constraints for monitoring and evaluation:

The meeting discussed several constraints to monitoring and evaluation in Bank-supported investments:

- Clients are responsible for project implementation, including monitoring and evaluation, but oftentimes there is not much buy-in by clients;
- Despite monitoring and evaluation being a Bank corporate priority, the incentives for Task Team Leaders (TTLs) to support monitoring and evaluation are not strong;
- Good evaluations must be planned for in advance, very early in the life of the project. Timing of evaluations is key. However, this needs a long-term technical and financial commitment to evaluation, which is not always followed through at the Bank. Often-times Bank TTLs change during the life of a project and this may further jeopardize monitoring and evaluation plans;
- It is not clear that the best methodologies to evaluate community-driven development projects, which are increasingly being promoted at the Bank, have been identified. Methodologies to evaluate community-driven development (including growth promotion), should be carefully re-considered and applied.
- Building the evidence base and the case for nutrition investments in the context of PRSPs will require combining nutrition impact analyses with costs and cost-effectiveness (rates of return), to compare with other investments and sectors. However, these are complicated analyses that need to be thought through and implemented carefully.

Recommendations:

The recommendations addressed the questions that evolved during the meeting’s discussions. The following are the eight key recommendations:

1. Monitoring and evaluation in existing Bank nutrition investments must be strengthened, to form an evidence base for future nutrition investments. Programmatic and investment decisions must be based on the best available evidence. To achieve this, the Bank must consider explicit (non-monetary) incentives for Task Team Leaders (and for clients) for including strong monitoring and evaluation in Bank investments.

2. Capacities of TTLs for monitoring and evaluation must be enhanced through a combination of systematic training and learning by doing. Investment in capacity development in client countries is also critical. It is essential to strike the right balance between conducting good quality evaluations, and developing in-country evaluation capacities during the evaluation of existing projects (learning-by-doing). Although in-country capacity analysis should be fostered, the Bank should also use the data, in-house, to distill lessons for its own operations. The Bank should consider partnerships with the World Bank Institute and with external institutions, including academia, both in the South and in the North for this purpose.
3. Currently, funding for evaluation is built in to the lending program. This is often a disincentive for evaluation, since client countries often hesitate to use loan funds for quality evaluation. This issue needs to be more creatively addressed in future investments. It may be appropriate for the Bank to explore separate funding options for evaluation, and to encourage government commitment to release data after the analysis/project is completed.
4. The Bank should re-examine systems for strengthening data collection wherein the interface between supervision reports, monitoring information systems and evaluation data is strengthened and the use of these data for decisions at all levels is facilitated. The challenge is to think about:
 - a) *why* are we collecting the data (i.e., what decision has to be made from the data), and not be fixated on the *process* of data collection;
 - b) what *ought* to happen to those data; and
 - c) what *actually* happens with those data.

The best way we can learn about this is through good examples (e.g., BINP), and use that as a way of not just critiquing the system but figuring out what else is necessary to make the system work.

5. The Bank needs to invest in a review of appropriate monitoring and evaluation methodologies and tools in the context of PRSCs and programmatic lending, including the documentation and identification of the consequences of non-intervention in human capital terms.
6. Results from evaluations must be widely published and disseminated in national and international forums as well as among the communities from whom data were collected. While countries must continue to own their data, every effort must be made to make these data more available for follow-up analyses. This can be difficult to accomplish because governments, and sometimes the Bank, can be fearful that the data will be misinterpreted or misused.
7. The Bank must invest in creating an evaluation database for nutrition. This will require archiving all available data, negotiating agreements with governments for sharing these data, defining data-sharing policies, and identifying mechanisms for institutionalizing this process for future data.
8. It is essential that the Bank develop a clear monitoring and evaluation strategy to guide work on this in the future. While this strategy may focus on nutrition issues in the short-term through a continuation of the discussions started in this meeting, in the longer term, it needs to be considered and accomplished across-sectors. A committee should be formed to bring together all groups working on monitoring and evaluation throughout the Bank. The committee should continue the work of this meeting, to look seriously at Bank-wide monitoring and evaluation, and how it can be achieved across sectors and integrated into operational projects and programs. The committee should address issues such as the Bank's strengths and weaknesses when working with its clients to address monitoring and evaluation. It may also be strategic to involve other development partners in this work.

ANNEX 1: BACKGROUND PAPER

**Habicht JP, Victora CG, Vaughan JP.
Evaluation designs for adequacy, plausibility and probability of public health programme
performance and impact.
International Journal of Epidemiology 1999; 28:10-18**

Reproduced in this report with author and Journal permission

LEADING ARTICLE

Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact

JP Habicht,^a CG Victora^b and JP Vaughan^c

The question of why to evaluate a programme is seldom discussed in the literature. The present paper argues that the answer to this question is essential for choosing an appropriate evaluation design. The discussion is centered on summative evaluations of large-scale programme effectiveness, drawing upon examples from the fields of health and nutrition but the findings may be applicable to other subject areas.

The main objective of an evaluation is to influence decisions. How complex and precise the evaluation must be depends on who the decision maker is and on what types of decisions will be taken as a consequence of the findings. Different decision makers demand not only different types of information but also vary in their requirements of how informative and precise the findings must be. Both complex and simple evaluations, however, should be equally rigorous in relating the design to the decisions. Based on the types of decisions that may be taken, a framework is proposed for deciding upon appropriate evaluation designs. Its first axis concerns the indicators of interest, whether these refer to provision or utilization of services, coverage or impact measures. The second axis refers to the type of inference to be made, whether this is a statement of adequacy, plausibility or probability.

In addition to the above framework, other factors affect the choice of an evaluation design, including the efficacy of the intervention, the field of knowledge, timing and costs. Regarding the latter, decision makers should be made aware that evaluation costs increase rapidly with complexity so that often a compromise must be reached. Examples are given of how to use the two classification axes, as well as these additional factors, for helping decision makers and evaluators translate the need for evaluation—the *why*—into the appropriate design—the *how*.

Keywords Evaluation, public health, diarrhoea, impact

Accepted 1 April 1998

Funding agencies are increasingly requiring quantitative evaluations of the impact of public health programmes, to meet increased demands for accountability. The present paper addresses summative evaluations of established interventions, rather than formative evaluations whose purpose is to fine tune programme implementation.¹ The results of summative

evaluations are to be used to make decisions about the programmes evaluated. Such 'instrumental' use of evaluation results is on the increase.² This is a distinct situation from what was observed in the past, when evaluations had limited 'instrumental' use but affected programmes and policies less directly, through changing perceptions. This difference in the uses of evaluations is important because one is more likely to reach decision makers when the use is 'instrumental', since the evaluators can ascertain what information is necessary for the decision-taking.

It is generally understood that other factors weigh as much or even more than quantitative evaluation results in the final decisions about programmes. However, the inferences from quantitative evaluations should be pertinent to the decisions if

^a Division of Nutritional Sciences, Cornell University, USA.

^b Universidade Federal de Pelotas, Brazil.

^c London School of Hygiene and Tropical Medicine, UK.

Reprint requests to: Prof. CG Victora, Departamento de Medicina Social, Universidade Federal de Pelotas, CP 464, 96001-970 Pelotas, RS, Brazil.

these evaluations are to have any utility, and thus experienced evaluators design their evaluations to address the specific questions of concern to decision makers.^{3,4} Of great importance for evaluation design is the type of inference required by decision makers, an issue which so far has not been addressed in the epidemiological literature.

This paper uses conventional epidemiological designs to discuss the above points, drawing from the authors' experience in the fields of health and nutrition in developing countries. It is particularly addressed to assessing effectiveness, that is, the large-scale achievements of interventions which, under ideal controlled conditions, have a known efficacy.

Why do the evaluation? Who will be influenced?

It is well recognized that formative evaluations must be done with those who have authority for the changes that need to be instituted. Less well understood is that the same kind of participatory research is essential in almost all summative evaluations for the information to be actually appropriately used in decision making.

Based on the findings from a summative evaluation, a decision maker may decide to continue, change, expand or end a project or intervention. The first task for the evaluation planner, therefore, is to define the target audience for the evaluation results, since the responsibilities and expertise of the decision makers will affect what questions should be asked.

Different decision makers not only ask particular questions but also require distinct kinds of inferences from the quantitative data. In other words, the answer to the question on *why* do an evaluation will affect its inferential design. For example, a donor agency may wish to document a statistically significant impact on mortality, while a district health manager may be interested in knowing whether a certain coverage was reached if the cold chain is functional. This does not imply that one kind of evaluation is more 'scientific' than the latter, as both types can and should be equally rigorous, in the sense of providing information that is sufficiently valid and precise for the decisions to be taken. The first type of evaluation provides evidence of effectiveness, being relevant to a decision to expand the programme. The second, on the other hand, assesses the overall adequacy of changes in outcomes, and may support a decision that no changes are required.

The evaluation designer should thus work with the decision makers for planning a study that will satisfy their requirements, that is, which will address the *why*. A conceptual framework is presented below to help *how* to design the evaluation. Note that an evaluation may be aimed at more than one category of decision maker. In this case, the design must take into account their different needs.

Classification axes

The proposed classification is based on two axes. The first refers to the indicators, that is, whether one is evaluating the performance of the intervention delivery or its impact on health or behavioural indicators. The second axis refers to the type of inference to be drawn, including how confident must the decision maker be that any observed effects were in fact due to the intervention.

First axis: What do you want to measure? Indicators of provision, utilization, coverage and impact

A useful way of looking at evaluations of health and nutrition interventions is to ask what is to be evaluated. The answer to this question will determine what will be measured. One may evaluate the provision or utilization of services, coverage or impact. Table 1 presents the outcomes of interest in a logical order leading from provision to impact. The services must be provided so that they are available and accessible to the target population and of adequate quality. Second, the population must accept the services and make use of them. Third, this utilization will result in a given population coverage. Coverage is a particularly useful measure, representing the interface between service delivery (the managerial process) with the population (the epidemiological picture). Finally, the achieved coverage may lead to an impact on behaviour or health. Any important shortcomings at the early stages of this chain will result in failures in the later achievements. For each outcome Table 1 presents a relevant question and an example of an indicator useful in the evaluation of a programme for the control of diarrhoeal diseases aimed at young children with emphasis on the promotion of oral rehydration solution (ORS). In subsequent tables the term *performance* evaluation will be used to encompass evaluations of provision, utilization and coverage, as separate from *impact* evaluations.

The evaluator should choose the indicators on the basis of discussions with the decision makers. The complexity of the evaluation designs and the extent of data collection will also depend on the decision maker's intended use of the results. As discussed, local managers may need summative data on provision and utilization to improve them within a health centre or in a district. On the other hand, national or international agencies may require assessments of coverage or impact to justify further investments in the programme. It also depends on how much one is willing to pay for the evaluation. Provision or utilization may be assessed by visiting services or using routine information systems. Coverage or impact, however, almost always require field data collection with important cost implications.

Second axis: How sure do you want to be? Types of inference: adequacy, plausibility, probability

The second axis refers to the kind of inference (adequacy, plausibility or probability), as well as on how confident decision makers need to be that any observed effects are in fact due to the project or programme. Both performance and impact evaluations may include adequacy, plausibility or probability assessments.

Adequacy assessment: Did the expected changes occur?

Inferences about the adequacy of programme outcomes depend on the comparison of the performance or impact of the project with previously established adequacy criteria. These criteria may be absolute—for example, distributing 10 million packets of ORS to children with diarrhoea or achieving 80% ORT use rate—or may refer to a change—for example, a 20% decline in reported diarrhoeal deaths in the programme area. Even when specific goals have not been established, performance or impact may still be assessed by measuring general time trends, such as an increase in coverage or a reduction in mortality.

Table 1 Example of indicators for evaluating a diarrhoeal diseases control programme

Indicator	Question	Example of indicators
Provision	Are the services available?	• No. of health facilities offering CDD activities per 100 000 population
	Are they accessible?	• Proportion of the population <10 km of a health facility with CDD activities
	Is their quality adequate?	• Proportion of health staff with recent CDD training
Utilization	Are the services being used?	• No. of attendances of under fives with diarrhoea per 1000 children • No. of ORS packets distributed
Coverage	Is the target population being reached?	• Proportion of all under fives with diarrhoea who used ORT
Impact	Were there improvements in disease patterns or health related behaviours?	• Time trends in diarrhoeal deaths and hospital admissions

Table 2 Characteristics of adequacy evaluations

Type of evaluation	Measurements	In whom?	Compared to what?	Inferences
Adequacy			Predefined adequacy criteria	Objectives met
Performance (provision, utilization, coverage)	Programme activities	Implementation workers Programme recipients		Activities being performed as planned in the initial implementation schedule
Cross-sectional	Once		Absolute value	
Longitudinal	Change		Absolute and Incremental value	
Impact	Health and behavioural indicators	Programme recipients or target population		Observed change in health or behaviour is of expected direction and magnitude
Cross-sectional	Once		Absolute value	
Longitudinal	Change		Absolute and Incremental value	

Adequacy assessments require no control groups if results are to be compared with set criteria (e.g. 90% exclusive breast-feeding rate by the age of 4 months). For assessing the adequacy of change over time, at least two measurements will be required, thus increasing the complexity of the design. Nevertheless, adequacy assessments are usually much less expensive than the other two types.

The main characteristics of adequacy evaluations are summarized in Table 2. Adequacy performance evaluations assess how well the programme activities have met the expected objectives. For example, these may include assessments of how many health centres have been opened, how many ORS packets or other drugs are available, how well health workers have been trained, how many children used the services or what coverage has been achieved in the target population. The evaluation may be cross-sectional, carried out on a single occasion, during or at the end of the programme. It may also be longitudinal, requiring baseline data or including repeated measurements for detecting trends.

Adequacy impact evaluations assess whether health or behavioural indicators have improved among programme recipients or among the target population as a whole. Again, the assessment may be cross-sectional or longitudinal. An advantage of adequacy assessments is that they can often use secondary data so that evaluation costs are much reduced.

Adequacy evaluations are limited to describing whether or not the expected changes have taken place. When assessing provision or utilization, one may reasonably ascribe an observed success to the programme being evaluated. For example, improved case management skills among health workers and

increased distribution of ORS may be safely attributed to a Control of Diarrhoeal Diseases (CDD) programme. When measuring coverage or impact, however, it may be difficult to infer that any observed improvements were due to the programme since there is no control group to ensure that these changes would not take place anyway. The observed improvements may have been caused by outside influences such as secular trends in mortality or malnutrition, general socioeconomic improvements, and the presence of other projects in the same area, etc.

Adequacy evaluations may also show a lack of change in the indicators. Under usual conditions, this suggests that the programme has not been effective. However, under special circumstances—such as a general deterioration in socioeconomic situation, a famine or another emergency, or general failure of other services—a lack of change may show that the programme has been effective in providing a safety net for the affected population. This scenario is further discussed in the plausibility section.

Despite their inability to causally link programme activities to observed changes, adequacy evaluations may provide all the reassurance necessary that the expected goals are being met and lead to continued support for the programme. For many decision makers, more complex evaluation designs will not be required, particularly since these would demand additional time, resources and expertise. If the evaluation finds that the programme goals are not achieved, further evaluations may be required to identify the causes for the failure and to guide remedial action. For other types of decisions, adequacy statements must be combined with either plausibility or probability assessments to deliver the necessary inferences.

Plausibility assessment: Did the programme seem to have an effect above and beyond other external influences?

Some decision makers may require a greater degree of confidence that any observed changes were in fact due to the programme. Plausibility appraisals go beyond adequacy assessments by trying to rule out external factors—called hereafter ‘confounding factors’—which might have caused the observed effects. A statement is plausible if it is ‘apparently true or reasonable, winning assent, a plausible explanation’.⁵ Table 3 summarizes the main types of plausibility evaluations.

Plausibility assessments attempt to control for the influence of confounding factors by choosing control groups before an evaluation is begun, or afterwards during the analyses of the data.

There are several alternatives for choosing a control group but the final choice is often dictated by opportunistic criteria, that is, by taking the best advantage of the existing situation. Control groups may include:

(a) *Historical control group*: the same target institutions or population. This approach entails a comparison of change from before to after the programme, accompanied by an attempt to rule out external factors.

(b) *Internal control group*: institutions, geographical areas or individuals that should have received the full intervention but did not, either because they could not or refused to be reached by the programme. Often, reception of a programme is variable. The indicators may then be compared between three or more groups of communities or individuals with different intensities of exposure to the intervention. A dose-response relation between intensity of the intervention and the observed performance or impact allows a stronger plausibility statement than findings from comparison between all and nothing groups. These approaches require comparisons of cross-sectional data collected at the end of the programme cycle.

Another kind of internal impact assessment is the use of the case-control method⁶ to compare previous exposure to the

programme between individuals with and without the disease. An advantage of the ‘case-control’ method is that it can be initiated relatively early after the initiation of the programme and may deliver definitive results earlier.

(c) *External control group*: one or more institutions or geographical areas without the programme. In this case, the comparison may be cross-sectional (intervention versus control at the end of the programme cycle) or longitudinal-control (comparing intervention and control at the beginning and at the end of the cycle).

The use of any of the above control groups results in much more plausible conclusions than if no controls are used. Plausibility is often markedly improved if they are used in combination. For instance, staggered interventions that begin at different times in separate areas allow the combination of historical data with external controls represented by areas where the intervention will start later; that in turn will have historical controls.

The intervention and control groups are supposed to be similar in all relevant characteristics except exposure to the intervention. This is almost never true since one of the comparison groups can be influenced by a confounding factor that does not affect the other group as much. For example, if the (control of diarrhoeal disease) CDD programme is implemented in an area with a better water supply than the control area, a difference in diarrhoeal mortality may be due to improved water and not to the programme. Dealing with confounding requires the measurement of probable confounders and their statistical treatment through matching, standardization, stratification, or other forms of multivariate analysis.⁷

Control of confounding is particularly important when internal comparisons are being made. Individuals who refuse the intervention or those who could not be reached often also differ from recipients in a number of other ways.

Confounding is also critical when using historical controls. This design is similar to an adequacy evaluation, in which a trend is recorded without external comparisons. To characterize

Table 3 Characteristics of plausibility evaluations

Type of evaluation	Measurements	In whom?	Compared to what?	Inferences
Plausibility			‘Opportunistic’ or non-randomized control group	The programme appears to have an effect above and beyond the impact of non-programme influences
Performance (provision, utilization, coverage)	Programme activities	Implementation workers Programme recipients (dichotomous or dose-response)		Intervention group appears to have better performance than control
Cross-sectional	Once		Control group	
Longitudinal	Change		Before-after	
Longitudinal-control	Relative change		Comparing before-after between intervention and control	
Impact	Health and behavioural indicator	Programme recipients or target population (dichotomous or dose-response)		Changes in health or behaviour appear to be more beneficial in intervention than control group
Cross-sectional	Once		Control group	
Longitudinal	Change		Before-after	
Longitudinal-control	Relative change		Comparing before-after between intervention and control	
Case-control	Once	Target population	Comparing exposure to programme in diseased (cases) and non-diseased (controls)	

a plausibility evaluation, however, one must also attempt to exclude other possible causes for the observed trends, for example, by assessing whether a decline in diarrhoeal mortality might have been due to socioeconomic development, to improved water supply and sanitation, to nutritional or other health interventions. This may be accomplished by estimating how much mortality would have decreased as a result of external changes and comparing that with the observed decline.⁸ A special situation is when no important improvement was observed,⁹ but using the above simulation approach one shows that a deterioration was expected. In this case, one may plausibly state that the programme was successful in preventing the situation from getting worse as a result of external hardships.

In many aspects, plausibility assessments are akin to 'natural experiments'. The evaluator will take advantage of the opportune existence of a control group to examine the effect of a programme. As its name indicates, a plausibility statement is largely based on value judgments of experts in the field, including the decision makers and the evaluators.

Plausibility assessments encompass a continuum, ranging from weak to strong statements. At the lower end of the plausibility scale are the simple comparisons with a control group, with an attempt to discuss and rule out possible confounding. At the higher end of the scale, one may have several comparisons and mathematical simulations. To reach the highest level of plausibility, one must formally discard all other likely explanations for the observed improvements. For example, plausibility would become stronger by consecutively showing that: (a) diarrhoeal mortality fell rapidly in areas with the CDD interventions (congruency of expected trend); (b) diarrhoea did not fall in the areas without the CDD interventions (not due to general changes in diarrhoea in the area); (c) changes in other known determinants of mortality could not explain the observed decline (lack of measurable confounding); (d) there was an inverse association between intensity of the intervention in the programme areas and diarrhoeal mortality (congruency of dose-response); (e) mothers with knowledge of ORT had fewer recent child deaths than those without such knowledge (congruency of mediating variables); (f) mortality among non-acceptors in the programme area was similar to that of the control area (congruency of lack of impact in the absence of the intervention); (g) the increase in ORT coverage was compatible with the degree of mortality reduction (congruency of magnitude of effect on mediating variables).

From an academic standpoint, the main shortcoming of plausibility assessments is that one cannot completely rule out all alternative explanations for the observed differences. However, by the time one had demonstrated point 'g' such alternatives are so unlikely as to be negligible. Furthermore, from a more practical, programmatic point of view, even less stringent plausibility statements are often sufficient for deciding about the future of a programme, because the cost to the decision maker of making a mistake is sufficiently low that higher plausibility is not necessary.

Probability assessment: Did the programme have an effect ($P < \alpha\%$)?

Probability evaluations aim at ensuring that there is only a small known probability that the difference between programme and control areas were due to confounding, bias, or to chance. These

evaluations require randomization of treatment and control activities to the comparison groups, being the gold standard of academic efficacy research.

While randomization does not guarantee that all confounding is eliminated (a common erroneous belief) it does ensure that the probability of confounding is measurable, being part of the error associated with the significance level used ($P < \alpha\%$), where P is chosen on the basis of considerations discussed below under 'Magnitude of sample'. The confounding factor does not even have to be known for this procedure to work. Thus randomization assures that the statistical statement of association is directly related to the intervention. This means that the statement of statistical probability of such a 'probability' evaluation relates directly to the causality of the intervention, and is not simply a statement that the comparison groups are different as is the case for all the other designs. We will not further discuss the details of probability evaluations here since these are adequately described in standard textbooks, in particular preventing biases that accompany the intervention from clouding the evaluation.

The main characteristics of probability assessments are listed in Table 4. There are a number of reasons why probability evaluations are often not feasible for assessing programme effectiveness.¹⁰ Firstly, the evaluator must be present at a very early stage of the programme planning cycle to design the randomization. Eligible services, communities or individuals have to be listed and randomized to intervention or control groups. Unfortunately, evaluators are often recruited only well after the programme has been implemented.

It is also necessary to overcome political influences affecting the choice of where to deploy the new intervention. Interventions are usually regarded as desirable and political pressures are put on planners, often resulting in the programme being directed to more influential communities. To ensure the use of random allocation, the evaluator must directly influence the implementation process. Alternatives have been proposed, including the 'stepped wedge design' (or 'experimentally staged introduction'¹¹) in which the intervention is deployed in a randomized sequence but eventually extended to all eligible communities or individuals. This eventual extension, as resources become available, is necessary not just for political but also for ethical reasons. This means that randomized designs are not appropriate for looking at effects with long time lags after the intervention begins.

The stringencies of probability trials may result in situations that are artificially different from the reality to which the results must be extrapolated, in other words, that the assessment lacks external validity. The probability assessment may have a high internal validity in showing that the intervention caused the results. But this gain in internal validity may be useless because the lack of external validity renders the results irrelevant to the decisions that need to be made.

Due to those and to other reasons,^{10,12} there are many limitations to the use of the probabilistic approach in assessing large-scale programmes. If the intervention has proven efficacy in field trials, few experienced decision makers would require measuring the effectiveness of every programme through a probability design. However, key individuals in donor or international agencies, as well as the evaluators themselves, may have been trained to regard probability assessments as the gold

Table 4 Characteristics of probability evaluations

Type of evaluation	Measurements	In whom?	Compared to what?	Inferences
Probability			Randomized control group(s)	The programme has an effect ($P < 0.05$)
Performance (provision, utilization, coverage)	Programme activities	Implementation workers Programme recipients		Intervention group has better performance than control
Longitudinal-control	Relative change		Comparing before-after between intervention and control	
Impact	Health and behavioural indicators	Programme recipients		Changes in health or behaviour are more beneficial in intervention than control group
Longitudinal-control	Relative change	Target population	Comparing before-after between intervention and control	

standard and fail to understand that this approach is seldom mandatory or even feasible for the routine evaluation of programme effectiveness.

In spite of the limitations of probability assessments there are times when these are essential, such as the first vitamin A supplementation trials that proved the lethality of subclinical avitaminosis A.¹³ These early studies had no external validity relative to the implementation of public health interventions even though they were essential to show the need for public health action in populations with subclinical avitaminosis A.

Combining adequacy, plausibility and probability inference objectives

The inference axis has in fact two components that vary together to a large extent. The first component is categorical: adequacy, plausibility and probability evaluations require different designs and result in different inferences, not just in the conclusions to be drawn from statistical tests, but also substantively. The importance of these questions for evaluation design is not discussed in the epidemiological literature.

For instance, a probability inference may deliver a rigorous inference that the intervention caused an impact, without any insight on whether the impact was adequate. Feasibility considerations indicate that some adequacy objectives can be incorporated into the design of plausibility and probability assessments at little added cost. Thus all evaluations should be designed to permit some adequacy inferences.

Logically, there appears to be no advantage of adding plausibility objectives to a probability evaluation. Both are directed at inferring that the intervention had an effect: the first by trying to exclude other explanations for the findings, the second by direct statistical testing. The strength of the inference is greater for probability evaluations, so that there would be no apparent advantage of adding plausibility objectives. However, it turns out that decision makers are not comfortable with a single piece of evidence, no matter how convincing this may be to statisticians or epidemiologists. For example, the exemplary vitamin A probability trials were not believed by many because of lack of congruency.¹⁴⁻¹⁶ This means that some plausibility should be built into probability designs, for example by providing data on confounding variables and, even more

importantly, on mediating variables. In the authors' experience, most decision makers are particularly sensitive to evidence of congruency, both from epidemiological data as well as from qualitative components of the evaluation that should complement the former. This congruency is often so persuasive that it may even outweigh impact results that do not quite reach statistical significance. It is the congruency of many pieces of evidence that ultimately persuades.

The inference axis has a second component, that is closely related to the first, categorical one. This component reflects the strength of inference about the causality of programme effect. The progression leads from a description without a comparison group, to comparison with possibly biased control groups, and finally to a comparison with a probably unbiased control group (through randomized trials). This second component of the inference axis, unlike the first, is well described in the epidemiological literature¹⁷ and is only briefly discussed in the present paper.

Combining the indicators and the inference axes

Each of the four components of the indicators axis (provision, utilization, coverage, impact) may be assessed according to the three types of inference (adequacy, plausibility, probability). An example is given below in Table 5.

Other factors influencing the choice of evaluation design

In addition to what indicators the decision makers wish to measure and to how certain they want to be, other factors may affect the choice of the appropriate type of evaluation. These include the large-scale efficacy of the intervention, the sector of knowledge to which it pertains, and the timing of the evaluation.

Efficacy

In a perfect world, interventions would only be widely applied at population level after their clinical and public health efficacy had been proven. However, this efficacy is often not demonstrated before practical public health interventions are initiated.

Table 5 Examples of possible evaluations of Diarrhoeal Diseases Control Programmes

Type of evaluation	Provision	Utilization	Coverage	Impact
Adequacy	Changes in availability of ORS in health centres	Changes in numbers of ORS packets distributed in health centres	Measurement of percentage of all diarrhoeal episodes treated with ORT in the population	Measurement of trends in diarrhoeal mortality in intervention area
Plausibility	As above, but comparing intervention with control services	As above, but comparing intervention with control services	Comparison of ORT coverage between intervention and control areas (or dose-response)	Comparison of diarrhoeal mortality trends between intervention and control areas (or dose-response)
Probability	As above, but intervention and control services would have been randomized	As above, but intervention and control services would have been randomized	As above, with previous randomization	As above, with previous randomization

The known efficacy of an intervention, therefore, is another important factor affecting the choice of evaluation design. Let us have two examples of evaluations, from the perspective of international and donor agencies. First, the efficacy of measles immunization is well proven. If adequacy evaluations show that the cold chain is operational and that coverage is high, there is little need for evaluating the impact of immunization programmes on disease rates, or even on changes in immunity to measles. The case is rather different, however, relative to using vegetarian foods to improve vitamin A nutrition. Their efficacy has not yet been established. Demonstration of increased ingestion¹⁸ is insufficient to persuade donors of the utility of this approach without measures of vitamin A status and at least a strong plausibility design. In fact this is a case where more probability designs are likely to be necessary to persuade decision makers to implement these interventions.

Sector of the programme

The subject area of the intervention is another important factor. This paper has concentrated on health and nutrition programmes but the approach can be adapted to other areas. As a general rule, more stringent evaluations seem to be demanded in the health field. For example, health impact evaluations often require the demonstration of a mortality reduction, which will only take place if a number of intermediate changes occur successfully. In other fields, a decision maker may be satisfied with, say, improved performance in a test (in education), an increased crop yield (in agriculture), or greater water consumption (in water/sanitation). In addition, in most other fields the effect is measured solely among the programme recipients, while in health and nutrition more stringent criteria require measurement of coverage or of impact on the whole target population.

Besides differences in the kinds of outcomes measured, distinct sectors require very different degrees of certainty before declaring an intervention as efficacious or effective. Some public policy and programme decisions depend entirely on plausibility statements. This is particularly the case in economics.¹⁹ Even within the health sector there are marked differences in judging the efficacy of interventions, whereby nutritional interventions appear to be held to higher standards than other health interventions.²⁰

This variability in standards of certainty required by decision makers in judging the efficacy and effectiveness of interventions

is a major barrier to rational public policy. It is therefore important to specify the levels of certainty that are achievable by the designs used, whether they are adequacy, plausibility or probability designs. Comparisons of expected impact for competing interventions across sectors need take these differences in certainty into account.

Timing and timeliness

The time when the evaluation is planned is fundamental. Probability assessments, as noted, require the evaluator to be present before the programme starts so that communities or individuals may be randomly allocated. All longitudinal methods, including those with a control group, also require baseline information to be collected before the programme, or else reliable secondary information for the pre-programmatic period. In general, evaluations of provision and utilization may be carried out sooner and more frequently, as they help local decision makers improve the interventions more quickly than waiting for longer term results. On the other hand, coverage and particularly impact evaluations are often undertaken later in the programme cycle and are often once-off activities. As a general rule, no less than 3-5 years are required for an intervention to show an impact. Several years or decades may be required for showing an impact on diseases with long incubation periods, such as AIDS, chronic diseases or the generational effects of improved nutrition.

As a general rule, evaluations should be planned when the programme itself is being designed, even if actual data collection is only foreseen at a later phase. Adequacy and plausibility evaluations may be instituted after the programme is under way. However, adequacy evaluations are more meaningful if there are clear and feasible pre-set goals, and plausibility evaluations often require baseline information from the pre-programmatic period.

The evaluation should deliver the answers to the decision makers in time for them to take these results into account in their decisions. Perfect information from an ideal evaluation is useless if it arrives after the decision is already made, an all too frequent situation. Therefore evaluators should determine not only what decisions are going to be made but when those decisions will take place. The design and conduct of the evaluation should then be organized to meet these deadlines, and all evaluation designs should include timeliness as part of their objectives.

Magnitude of sampling

The number of people sampled and the distances between them are major determinants of the costs of the evaluation. The number of areas, and the number of people to be sampled within the areas, is determined by calculations based on the willingness of the decision makers to be given erroneous results. Usual practice in scientific research is to accept as true that a treatment has an effect 5% of the time when in fact there is no effect—an alpha error or significance level of 5%. This is an almost sacrosanct figure among academicians. Usual practice in academic research is to declare no effect 20% of the time when there really is an effect—a beta error of 20% usually referred to as a power of 80%. The lower the setting of the per cent alpha and beta errors the greater will be the sample size.

From the above it is obvious that scientists are willing to not identify a beneficial result four times more often than to be mistaken in declaring such a result when it is absent. Most public health practitioners would be very unhappy with that trade-off for evaluating their programmes, and many would set the opposite trade-off. At any rate the sacrosanct 5% significance limit needs to be questioned before being accepted automatically. For instance accepting 20% for both alpha and beta errors would reduce the sample size by 35–40%, below that acceptable to many scientists. Setting explicit per cent error levels that are appropriate for the decision maker is in fact more scientific than blindly accepting conventional levels. Thus one can set the errors much higher in many programme evaluations than in efficacy trials, if the results are not to be used for scientific inferences for which low alpha errors are necessary.

Costs

Costs are often the major factor affecting the choice of a design. Decision makers are particularly sensitive to this aspect, for often they will be asked to provide the necessary funds from the overall programme budget. A full discussion of evaluation costs is beyond the scope of this paper, but evaluators should discuss with decision makers the budgetary implications of different designs, including the following issues:

- (a) Is a full summative evaluation worth doing?
- (b) Is there a need for collecting new data? If so, at what level?
- (c) Does the design include an intervention-control or a before-and-after comparison?
- (d) How rare is the event to be measured and how small is the difference to be detected?
- (e) How complex will the data analysis be?

Choosing the evaluation design

This section discusses how to combine evaluation designs and also summarizes some of the main points presented above. The classification axes presented above should be used for discussing with decision makers which evaluation design or designs may be used for each programme. Table 6 shows some areas which may typically concern different decision makers in the field of health and nutrition.

Complex evaluations (for example, those with a probability approach or impact assessments) should not be carried out before ensuring, through less costly evaluations, that the process is moving in the expected direction.³ Table 7 shows a heuristic sequence of evaluations with growing complexity, that would

Table 6 Possible areas of concern of different decision makers

Types of evaluation	Provision	Utilization	Coverage	Impact
Adequacy	Health centre manager		District health manager	
	International agencies		International agencies	
Plausibility		International agencies		Donor agencies, scientists
Probability		Donor agencies, scientists		

Note: The shaded areas represent those of greater concern for international (e.g. UN) agencies.

Table 7 Hypothetical example of flow of evaluations from simpler to more complex designs, to allow decisions by local, district and national managers

Axes	Provision	Utilization	Coverage	Impact
Adequacy	1st	2nd	3rd	4th (b)
Plausibility			4th (a)	5th
Probability				

be carried out based on the results of simpler evaluations. It is based on the evaluations listed on Table 5, and contemplates the concerns of local, district and national decision makers.

One would start by ensuring that ORS is available in the health centres, and next check that the population is utilizing this service. The third stage would include a household survey to assess whether the ORT coverage goal has been reached. So far, all evaluations have been adequacy statements. Next, the decision maker could opt for either showing that coverage is higher in the intervention than in the control areas (option 4(a), a plausibility statement that the higher coverage was due to the programme), or perhaps for attempting to show a reduction in diarrhoeal mortality compared to before the programme (option 4(b), an adequacy statement). This logical sequence helps in deciding the actual sequence, which also depends on funding, administrative and political considerations.

In conclusion this paper is designed to foster the development of a logical framework by which health and nutrition programmes can be judged and compared to other public interventions. The major premise is that the objective of an evaluation is to influence decision makers. How complex and precise the evaluation must be depends on who the decision maker is and on what types of decisions will be taken as a consequence of the findings. Both complex and simple evaluations, however, should be equally rigorous, whether they assess the adequacy of an intervention's effects, or assess the plausibility, or the probability that the intervention caused these effects.

In addition to the above framework, other factors affect the choice of an evaluation design, including the efficacy of the intervention, the field of knowledge, timing and costs. Regarding the latter, decision makers should be aware that evaluation costs increase rapidly with complexity so that often a compromise must be reached.

Acknowledgement

This work was supported by the Evaluation and Research Office, UNICEF.

References

- ¹ Mohr LB. *Impact Analysis for Programme Evaluation*. Newbury Park: Sage Publications, 1992.
- ² Rossi PH, Freeman HE. *Evaluation*. Newbury Park: Sage Publications, 1993.
- ³ Mason JB, Habicht JP. Stages in the evaluation of ongoing programmes. In: Sahn DE, Lockwood R, Scrimshaw NS (eds). *Methods for the Evaluation of the Impact of Food and Nutrition Programmes*. Tokyo: United Nations University, 1984, pp.26-45.
- ⁴ Habicht JE, Mason JB, Tabatabai H. Basic concepts for the design of evaluation during programme implementation. In: Sahn DE, Lockwood R, Scrimshaw NS (eds). *Methods for the Evaluation of the Impact of Food and Nutrition Programmes*. Tokyo: United Nations University, 1984, 1-25.
- ⁵ *New Webster's Dictionary and Thesaurus of the English Language*. Danbury, CT: Lexicon Publications, 1992.
- ⁶ Schlesselman JJ. *Case-Control Studies. Design, Conduct, Analysis*. New York: Oxford University Press, 1982.
- ⁷ Rothman KJ. *Modern Epidemiology*. Boston: Little, Brown and Company, 1986.
- ⁸ Victora CG, Olinto MT, Barros FC, Nobre LC. The recent fall in diarrhoea mortality in Northeastern Brazil. Did ORT play a rôle? *Health Pol Plan* 1996;11:132-41.
- ⁹ DeVanzo J, Habicht J-P. Infant mortality decline in Malaysia 1946-1975. The roles of changes in variables and the structure of their relationships. *Demography* 1986;23:143-60.
- ¹⁰ Black N. Why we need observational studies to evaluate the effectiveness of health care. *Br Med J* 1996;312:1215-18.
- ¹¹ Cook TD, Campbell DT. *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin Co, 1979, pp.375-76.
- ¹² Kirkwood BR, Cousens SN, Victora CG, Zoysa I. Issues in the design and interpretation of studies to evaluate the impact of community-based interventions. *Trop Med Int Health* 1997;2:1022-29.
- ¹³ Sommer A, Tarwotjo I, Dunaedi E et al. Impact of vitamin A supplementation on childhood mortality: a randomized controlled community trial. *Lancet* 1986;i:1169-73.
- ¹⁴ Costello AML. Vitamin A supplementation and child mortality (letter). *Lancet* 1986;ii:161.
- ¹⁵ Gray RH. Vitamin A supplementation and child mortality (letter). *Lancet* 1986;ii:162.
- ¹⁶ Martinez H, Shekar M, Latham M. Vitamin A supplementation and child mortality (letter). *Lancet* 1986;2:451.
- ¹⁷ Rothman KJ, Greenland S. *Modern Epidemiology, 2 edn*. Philadelphia: Lippincott Raven, 1998.
- ¹⁸ Smitaslri S. On planning and implementing Vitamin A interventions: Linking scientific knowledge to effective action. In: Garza C, Pelletier D (eds). *Beyond Nutritional Recommendations: Implementing Science for Healthier Populations*. Cornell University Press, Ithaca (in press).
- ¹⁹ Judge GG, Griffiths WE, Hill RC, Lee T-C. *The Theory and Practice of Econometrics*. New York, NY: Wiley, NY 1985.
- ²⁰ Hill K (ed.). *Child Health Priorities for the 1990s*. Baltimore: The Johns Hopkins University Institute for International Programmes, School of Hygiene and Public Health, 1992.

ANNEX 2: PRESENTATIONS

Clinic 1: The BINP End Evaluation:

This presentation by **Rezaul Karim**, (The Institute for Nutrition and Food Science, University of Dhaka, Bangladesh), and **Sascha Lamstein**, Tufts University outlined the BINP evaluation plan and preliminary results.

The Presentation Slides:



**Bangladesh
Integrated
Nutrition Project**

END EVALUATION

Introduction to BINP

- Started in 1995
- Covered 59 of 490 (12%) rural upazilas
- Implemented by a Project Office responsible to the Ministry of Health and Family Welfare
- Taken over by NNP in March 2003

BINP Components

- National Level Nutrition Component
- Community Based Nutrition Component (CBNC)
- Inter-sectoral Nutrition Component

National-Level Nutrition Component Activities

- Program development
- Institute building
- IEC development
- Strengthening existing nutrition activities
- Project management
- Monitoring and evaluation

Inter-Sectoral Nutrition Component: Activities

- Home gardening
- Homestead poultry rearing

Community-Based Nutrition Component: Activities

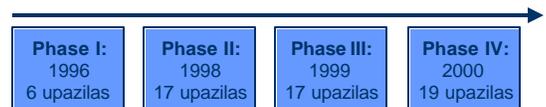
- Target children under 2 and PLWs
- Monthly growth monitoring and promotion
- Supervised supplementary feeding
- Micronutrient supplementation
- Follow-up home visits
- Secondary target of adolescent girls through adolescent forums

Community-Based Nutrition Component: Goals

- Reduction of severe PEM by 50%
- Reduction of moderate PEM by 33%
- Reduction of low birth weight incidence by 50%
- Improvement in maternal weight gain to 7 Kg among 50% of pregnant women
- Reduction of iron deficiency anemia by 33%
- Elimination of vitamin A deficiency disorders by 50%
- Elimination of iodine deficiency disorders by 50%

Community-Based Nutrition Component: Implementation

- Implemented in phases:



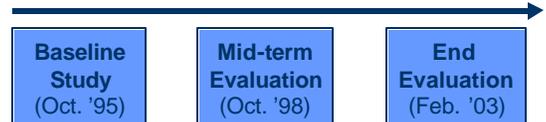
- Implemented by contracted NGOs
- Implemented through Community Nutrition Centers (CNCs) at the local level.

Community-Based Nutrition Component: Implementation

- Each upazila was divided into CNCs.
- Each CNC covered 200-250 households (1,000-1,500 population).
- One Community Nutrition Promoter (CNP) per CNC managed local service delivery.
- One Community Nutrition Organizer (CNO) per supervised 10-12 CNPs.
- NGOs supervised and facilitated upazila-level work.

Community-Based Nutrition Component: Baseline & Evaluation

- Baseline study, mid-term evaluation, and endline evaluation were conducted in the six Phase I upazilas:



- Control upazilas were identified at the baseline.
- The same control upazilas were included in the mid-term and end evaluations.

Community-Based Nutrition Component: Sample Upazilas

Project Upazilas

Gabtali, Bogra
Bhanga, Faridpur
Banaripara, Barisal
Shahrasti, Chandpur
Rajnagar, Molvi Bazar
Mohammadpur, Magura



Control Upazilas

Hajiganj, Chandpur
Sonatala, Bogra

Community-Based Nutrition Component: Baseline & Evaluation

- Comparisons between project and control upazilas were made for each study.
- A six-cell comparison was conducted for the endline evaluation:

TYPE	BASELINE	MD-TERM	END-LINE
Program	Cohort a	Cohort b	Cohort c
Control	Cohort d	Cohort e	Cohort f

Endline Evaluation: Objectives

- Measure project impact in the six first phase upazilas on:
 - nutritional status of children under 2 years and women of reproductive age;
 - pregnancy weight gain;
 - birth weights;
 - receipt and use of iron tablets by PLW and adolescent girls; and
 - coverage of vitamin A to post partum women.
- Make pre-post and project-control comparisons.

Endline Evaluation: Components

- Under two children, mothers, and adolescent girls household survey
- Pregnancy weight gain survey
- Birth weight survey
- Service provider assessment through FGD
- Review of project monitoring system
- Cost effectiveness analysis

Endline Evaluation: Key Indicators

- Demographic characteristics and socio-economic status
- Provision and utilization of services
- Caring practices
- Nutritional status of children under two years of age, women, and adolescents
- Pregnancy weight gain
- Birth weight

Baseline Survey: Sampling

- 10 mouzas were drawn from each upazila using PPS.
- Mouzas were divided into clusters of 40-50 HH each.
- One cluster was selected from each mouza at random.
- All households with under 5 children were interviewed.
- In all 2473 project and 938 control HH were interviewed.
- Anthropometric measurements were taken from children and mothers from 1199 project and 448 control households.

Mid-term Evaluation: Sampling

- Followed similar methodology and approach as the end evaluation.
- In all 4854 project and 1622 control HH having under 2 child were interviewed.
- 1561 birth weight and 1398 pregnancy weight gain measurements were collected.

Endline Evaluation: Sampling

- Sample size quotas per upazila:
 - children under two questionnaire: 600
 - pregnant women questionnaire: 160
 - newborn infants questionnaire: 170
- Primary sampling units (PSUs):
 - CNCs in project upazilas
 - Comparable population (village/para or HA unit) in control upazilas

Endline Evaluation: Sampling

- 17 PSUs were selected from each upazila using PPS.
- Lists of households with children under 2 and pregnant women were obtained from CNCs, HAs or FWAs.
- All HHs with children under 2 in each PSU were interviewed to collect data on children under 2, their mothers, and any adolescent girls (11-19 years) in the household.
- All pregnant women in each PSU were also interviewed.
- If the quotas were not reached after completing interviews in the 17 PSUs, additional PSUs were selected at random to complete the quota.
- All newborn infants born in the upazila during the study period were interviewed with the help of informants.

Endline Evaluation: Preparation

- Three questionnaires were developed for data collection:
 - Children under two (for collection of data on children, mothers and existing adolescent girls)
 - Pregnant women
 - Newborn infants
- The questionnaires were pre-tested.
- The questionnaires and methodology were shared with BINP and NNP.

Endline Evaluation: Team Building

- 71 enumerators and 8 supervisors were initially selected.
- The enumerators and the supervisors received a 4-day residential training.
- The 64 best enumerators and the supervisors were retained for field work.

Endline Evaluation: Data Collection

- One data collection team comprising 8 enumerators and one supervisor was sent to each upazila.
 - 5 enumerators completed the children under two questionnaires.
 - 2 enumerators completed the birth weight questionnaires.
 - 1 enumerator completed the pregnancy weight gain questionnaires.
- Completed questionnaires were sent to the project office periodically.

Endline Evaluation: Realized Sampling

- Questionnaires completed per upazila:
 - Children under two questionnaire: 613 - 633
 - Pregnant women questionnaire: 162 – 199
 - Newborn infants questionnaire: 170 - 171

Endline Evaluation: Data Processing

- **Field quality control:**
 - Completed questionnaires were reviewed nightly by supervisors.
 - Project investigators conducted supervisory visits.
- **Central editing:** Completed questionnaires were reviewed in the project office before data entry.
- **Data entry:**
 - Data were entered using double entry method.
 - Anthropometric data were entered using ANTRHO.
 - All the other data was entered using Fox-Pro 2.6.

Endline Evaluation: Data Analysis Considerations

- Realization of BINP goals and objectives
- Statistical significance between project and control and between baseline, mid-term and endline
- Relationship between nutrition knowledge, behavior change, and nutritional outcome
- Whether there is a true control
- Lack of baseline data for certain key indicators (i.e. pregnancy weight gain, birth weight, adolescent girls nutritional status)

Endline Evaluation: Data Analysis to Date

- Dummy tables were developed to highlight key indicators.
- Initial data analysis was conducted and tables were prepared to show:
 - description of the samples;
 - means and frequency distributions of some process and outcome indicators; and
 - comparative measures between baseline and endline and between midterm and endline.
- A draft report and cost effectiveness analysis were prepared.

Endline Evaluation: Initial Findings

Table 8: BMI of Target Non-Pregnant Women

INDICATOR	BASELINE (1995)		MID-TERM (1998)*		ENDLINE (2003)	
	Project (N=544)	Control (N=216)	Project	Control	Project (N=3590)	Control (N=1199)
Cut-Off (%)						
Less than 18.5	46.3	49.5	67.3	64.9	48.5	46.8
18.5 and Above	53.7	50.5	32.7	35.1	51.5	53.2

* Mid-term data is taken from the Mid-Term Evaluation Report and, therefore, no tests of significance could be run. BMI is for all women of child-bearing age in the household (15-49 years), no distinction of whether they are pregnant. Baseline and Endline data are only for women that are not pregnant.

Endline Evaluation: Initial Findings

Table 9: Pregnancy Practices

INDICATOR	BASELINE (1995)		MID-TERM (1998)		ENDLINE (2003)	
	Project (N=555)	Control (N=222)	Project	Control	Project (N=3729)	Control (N=1238)
Antenatal check-ups during last pregnancy^{cd}						
Yes	12.1	13.5			81.0	55.1
No	87.9	86.5			19.0	44.9
Iron tablets taken during last pregnancy^{cd}						
Yes	16.8	21.6			83.9	41.8
No	83.2	78.4			16.1	58.2
Rest during last pregnancy^{de}						
Less than usual	43.1	53.2			10.0	24.7
Same as usual	43.1	22.1			23.0	29.1
More than usual	13.8	24.8			67.0	46.2

Endline Evaluation: Initial Findings

Table 10: Childcare Practices

INDICATOR	BASELINE (1995)		MID-TERM (1998)		ENDLINE (2003)	
	Project	Control	Project	Control	Project	Control
Fed colostrums^{a,d}	(N=573)	(N=227)			(N=3726)	(N=1235)
Yes	66.5	78.4			95.7	92.6
No	33.5	21.6			4.3	7.4
Initiation of breastfeeding^e	(N=573)	(N=229)			(N=3729)	(N=1238)
Immediately after birth	14.3	12.2			84.0	56.2
Within 24 hours of birth	55.1	61.6			8.5	19.3
More than 24 hours after birth	30.5	25.3			7.2	24.0
Never	-	0.9			0.3	0.5
Initiation of solid foods^e	(N=441)	(N=163)			(N=3729)	(N=1238)
Less than 1 month	-	-			8.0	11.9
1-5 months	29.0	25.8			59.4	55.9
6-10 months	51.0	47.2			11.8	11.2
11-24 months	17.7	23.9			0.1	0.1
Still not					20.6	20.3
Cannot remember	2.3	3.1			0.1	0.6

Endline Evaluation: Initial Findings

Table 14: Nutritional Status of Target Children (6-23 months)

INDICATOR	BASELINE (1995)		MID-TERM (1998)*		ENDLINE (2003)	
	Project	Control	Project	Control	Project	Control
Weight-for-Age Z-Score	(N=442)	(N=169)	(N=3883)	(N=1266)	(N=2411)	(N=800)
Severe, -3.00 and below	30.8	27.8	18.2	19.5	13.0	14.4
Moderate, -2.99 to -2.00	28.5	29.0	37.3	38.6	36.3	36.6
Normal, -1.99 and above	40.7	43.2	44.5	41.9	50.7	49.0
Height-for-Age Z-Score^e	(N=430)	(N=162)	(N=3883)	(N=1266)	(N=2409)	(N=799)
Severe, -3.00 and below	36.0	34.6	21.4	19.8	16.6	23.3
Moderate, -2.99 to -2.00	27.0	32.1	28.2	29.2	30.6	28.8
Normal, -1.99 and above	37.0	33.3	50.4	51.0	52.8	47.9
Weight-for-Height Z-Score^e	(N=429)	(N=162)	(N=3883)	(N=1266)	(N=2403)	(N=800)
Severe, -3.00 and below	11.4	8.0	2.2	2.1	1.0	1.3
Moderate, -2.99 to -2.00	14.7	14.8	13.2	16.4	11.5	8.3
Normal, -1.99 and above	73.9	77.2	84.6	81.5	87.6	90.5

Endline Evaluation: Initial Findings

Table 20: Cost Effectiveness

YEAR	PRESENT (1996-97) COST OF REMOVING ONE CASE INCREMENTAL MALNUTRITION (TK)	
	Severely underweight	All underweight
1999-00	9501	16381
2000-01	8022	13830

Endline Evaluation: Data Analysis Plan

- Further analyze the differences between project and control over baseline, mid-term, and endline.

Indicator	Baseline (1995)		Mid-Term (1998)*		Overall PP Change (B/M)	Endline (2003)		Overall PP Change (B/E)
	Project	Control	Project	Control		Project	Control	

- Prepare tables on the distribution of benefits by SES and other disaggregations.
- Conduct tests of significance and regression analyses where applicable.

Clinic 1 (continued): BINP - Country Context

This presentation by Nasreen Haq, (Action Aid, Bangladesh), focused on putting the BINP evaluation results in perspective.

The Presentation Slides:

Nutrition Context in Bangladesh

- High prevalence of undernutrition in children

	0-23 months	24-59 months
Wt/age<2sd	41-46%	53-61%
Ht/age<2sd	40-45%	50-54%
Wt/ht< 2sd	9-15%	8-14%

Nutritional Surveillance Project in 2001 (HKI/IPHN)

Nutrition Trends

- The NSP has documented trends in nutritional status of children since 1990. 1.8 % decline in stunting since 1990. Prevalence of underweight fell by 1.3%.
- Evaluation of BINP must separate the secular trends from actual impact of program.

Household Food Security

- Households consuming less than 1805 kcal/person/day: 24%
- Households consuming 1805- 2122 kcal/person/day: 23%

Nutritional Surveillance Projects in 2001 (HKI/IPHN)

Undernutrition in Mothers of under 5 children

- Maternal wasting ranged from 35-41% in 2001 as measured by BMI<18.5

Nutritional Surveillance Projects in 2001 (HKI/IPHN)

Seasonality

- Seasonality pattern observed in both children and mothers. Most prominent in children.
- No improvement in wasting pattern since 1990.

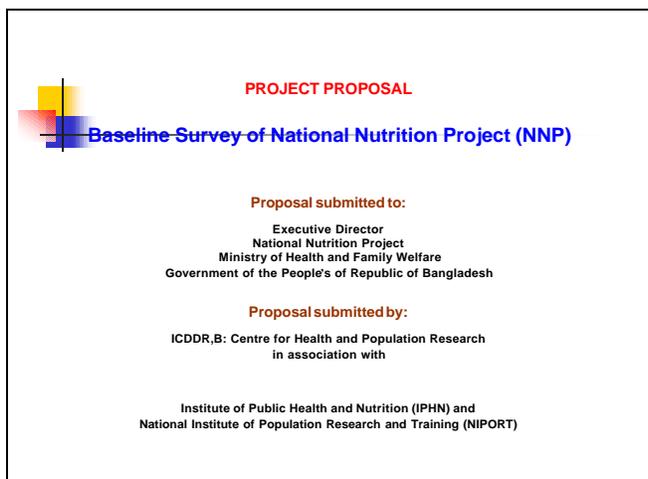
Considerations for Evaluation of BINP

- Seasonality
- Use of z score
- Independent data – use of the NSP
- Indicators on improved coverage of services
- Diversity of implementation
- Assess the interventions for household food security.
- Should not only focus on nutrition status indicator
- Assess the design of the BINP

Clinic 1 (continued): BINP – The National Nutrition Project (NNP) Link

This presentation by Barkat-e Khuda (International Centre for Diarrhoeal Disease Research, Bangladesh) laid out the design for the baseline study for the follow-on National Nutrition Project (NNP).

The Presentation Slides:



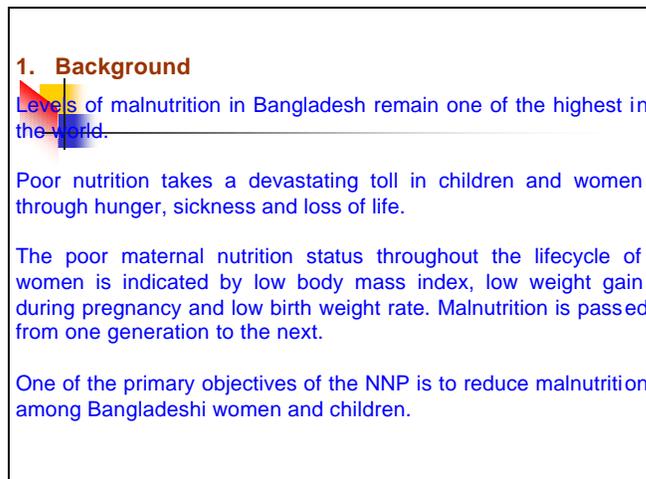
PROJECT PROPOSAL

Baseline Survey of National Nutrition Project (NNP)

Proposal submitted to:
Executive Director
National Nutrition Project
Ministry of Health and Family Welfare
Government of the People's of Republic of Bangladesh

Proposal submitted by:
ICDDR,B: Centre for Health and Population Research
in association with

Institute of Public Health and Nutrition (IPHN) and
National Institute of Population Research and Training (NIPORT)



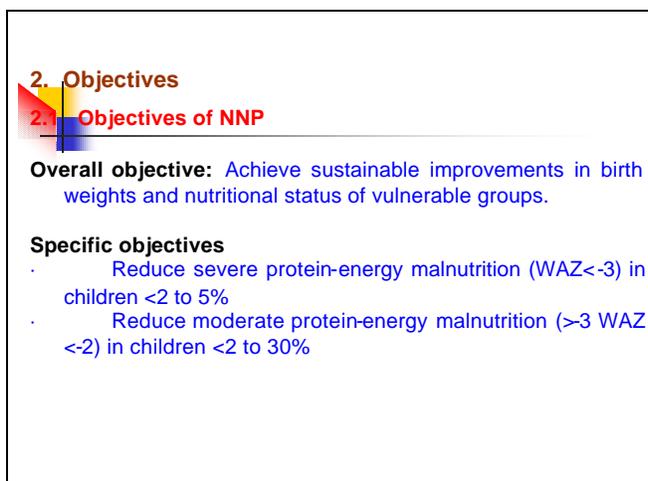
1. Background

Levels of malnutrition in Bangladesh remain one of the highest in the world.

Poor nutrition takes a devastating toll in children and women through hunger, sickness and loss of life.

The poor maternal nutrition status throughout the lifecycle of women is indicated by low body mass index, low weight gain during pregnancy and low birth weight rate. Malnutrition is passed from one generation to the next.

One of the primary objectives of the NNP is to reduce malnutrition among Bangladeshi women and children.



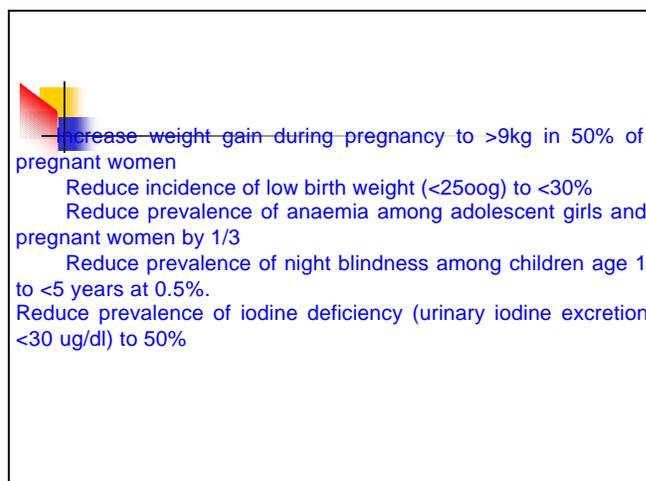
2. Objectives

2. Objectives of NNP

Overall objective: Achieve sustainable improvements in birth weights and nutritional status of vulnerable groups.

Specific objectives

- Reduce severe protein-energy malnutrition (WAZ<-3) in children <2 to 5%
- Reduce moderate protein-energy malnutrition (>3 WAZ <-2) in children <2 to 30%



- Increase weight gain during pregnancy to >9kg in 50% of pregnant women
- Reduce incidence of low birth weight (<2500g) to <30%
- Reduce prevalence of anaemia among adolescent girls and pregnant women by 1/3
- Reduce prevalence of night blindness among children age 1 to <5 years at 0.5%.
- Reduce prevalence of iodine deficiency (urinary iodine excretion <30 ug/dl) to 50%

2.2 Objectives of the NNP Baseline Survey

1. Obtain statistically valid samples from the 44 new NNP Upazilas
2. Obtain valid baseline results on the specific objectives of NNP
3. Obtain valid data from appropriate number of control Upazilas
4. Obtain valid data from appropriate samples from 53 old Upazilas

Have optimum matching variables from End Line Survey of BINP Phase I Upazilas

3 Research Design and Methods

3.1 Variables

(i) Primary variables

1. Weight and age in children <2
 2. Weight and height in pregnant women
 3. Birth weight
 4. Anaemia among adolescent girls and pregnant women
 5. Iodine deficiency in adolescent girls and pregnant women
- Pregnancy weight**

(ii) Secondary variables

1. Socioeconomic data
2. Infant and children
- Ø Feeding, and growth monitoring and promotion (GMP) records
3. Adolescent girls
- Ø Weight, Height, Age, Diet, anaemia, urinary iodine
4. Pregnant women
- Ø Gravida, Pregnancy complications, Pregnancy duration, ANC (visit, services, type and quality), Rest, Dietary practice, Knowledge on food, and plans for breastfeeding

5. Micronutrient

- Ø Use of iodised salt, micronutrient intake in food (24 hour recall), iron tablet intake (pregnant women), and Vitamin A capsule intake at delivery
- Ø Urinary iodine will be estimated for children, adolescent, pregnant and lactating women; and salt will be estimated for adequacy of iodine content
- 6. Stool samples will be tested for test of ova of parasites in selected samples

(iii) Others variables

1. Participation in home gardening
2. Participation in Poultry project of NNP
3. Participation in VGD programme

3.2 Sampling

The primary objective of the Baseline Survey is to provide estimates with acceptable precision for some selected measurable indicators to assess the current situation and measure the impact of the project inputs. The baseline survey, the mid-term survey, and the end-line evaluation make a 'six-cell' study design (Table 1).

Table1: Framework for comparing data of the Baseline Survey with the midterm evaluation and endline evaluation

Type	Baseline	Midterm	End line
Programme	a	b	c
Control	d	e	f

Sample Design

- The major domains are:
 - (i) 44 new upazilas
 - (ii) 53 old BINP upazilas, not currently under INFS endline evaluation
- Independent estimates at division level for 53 old upazilas
- Independent estimates at division level for 44 new upazilas
- Independent estimates at division level for 12 control upazilas

Minimum Sample Size Per Division for 44 new Upazilas and 12 control Upazilas

Estimation of sample size is based on changes in measurable indicators in the project upazilas compared with the control upazilas.

Sample estimates will be self-weighted.

Ratio of sample sizes from the programme and the control upazilas will be 3:1.

Minimum Sample Size for 53 old Upazilas

This survey is also intended to be the NNP baseline for the 53 BINP upazilas, not currently under end evaluation by INFS.

For determination of sample size in 53 old upazilas, the same methodology used for 44 new project upazilas will be used.

The samples will be taken to ensure the phase wise upazilas with proportionate samples among phases II, III, and IV.

3.5 Methodology

Types of data collection

- a) Survey on questionnaire - Demography and socioeconomic aspects, caring practices, and source of water use
- b) Anthropometry
- c) Blood sample for anaemia
- d) Urine sample for iodine status

Procedure

1. For HH survey, pre-tested questionnaire will be used
2. For anthropometry, trained survey team will measure weight, height and MUAC. Children <2 will be measured for length at 2 mm of precision
3. For pregnancy weight gain, a team will collect data at least for 3 months in the last trimester of pregnancy
4. For birth weight, a trained person from the area will collect birth weight data within 24 hours of delivery

For anaemia, finger prick collection of blood sample for estimation of haemoglobin will be done by trained assistants. For iodine status, urine will be collected from all age groups in the field and sent to Dhaka. Household salt samples will be collected to estimate presence of iodine. A proportion of stool samples will be tested for prevalence of parasites. The variables will be matched with those of end line evaluation by INFS, to the extent relevant.

3.7 Supervision and Quality Control

A major responsibility of ICDDR,B would be to maintain the quality of data and its analysis.

External validation will be in place within two weeks of commencement of baseline survey.

An Independent Quality Control Team (IQCT) will be visiting the survey sites to re-interview and re-examine 2% of respondents interviewed over the preceding 15 days. The team will examine onsite interviews and anthropometric measurements of women and newborn babies.



Discrepancies detected will be taken up with the survey agency for corrective measures.

The quality control team will retrain the survey team, if necessary. If gross mistakes are found, survey will be stopped at any point of time.

3.5 Data Analysis

Data analysis

An expert team will develop a data analysis plan. ICDDR,B will take the major responsibility, with IPHN and NIPORT providing necessary assistance.

ANNEX 3: MINUTES FROM NUTRICOM PROJECT, MAURITANIA CLINIC CHERIF DIALLO

SUMMARY OF MINUTES OF VIDEO CONFERENCE (translated from French)

**Secretariat of State for the Status of Women
NUTRICOM Project
Wednesday, May 28, 2003**

Summary of decisions made following the video conference of May 22, 2003:

At 5 p.m. on Thursday, May 22, 2003, the Resident Mission of IDA organized, on behalf of the NUTRICOM project team and a number of departments involved in the nutrition sector, a video conference between Nouakchott and Washington on the topic of the monitoring and evaluation (M & E) of Bank-financed nutrition projects. This ultimately turned out to be a two-way debate with a group of IDA specialists on ways and means of setting up the NUTRICOM project's monitoring and evaluation system. The full list of participants is in Annex 4.

Description of the videoconference:

Mme. Mounina mint Abdellah, Director of the NUTRICOM Project and Messrs. Chérif Diallo, Task Manager, and Sall Aliou, Senior Nutritionist at NUTRICOM, described NUTRICOM's structure, operations and current M & E system. They emphasized the absence of such a system for the project, and the negative impact that this is having on its implementation.

Mr. Mohamed Lémineould Moujtaba, from the United Nations Fund for Population Activities (UNFPA), emphasized, in connection with the external evaluation of NUTRICOM, the lack of a baseline, mid-term survey that would, in the normal course of events, occur between the baseline survey performed at project start-up and the one carried out upon completion, and that would indicate whether the activities initiated are beginning to produce the desired results (or, in other words, whether NUTRICOM is on the right track).

Mr. Sall Aliou provided some clarifications concerning:

- Data-gathering tools used at the Community Nutrition Centers (*Centres de nutrition communautaire*, CNC) or at the Rural Community Nutrition Centers (*Centres de nutrition communautaire rural*, CNCR): growth monitoring sheets, home visit check-off sheets, quick-reference guide cards, supervision sheets, weigh-in records and IEC attendance sheets;
- The circuit through which this data passes, from the source (either a CNC or a CNCR) to the project management unit (PMU);
- The absence of summary monthly reports, which are also not in a standard format; and, finally
- NUTRICOM's lack of an M & E software package that would facilitate the processing of the data gathered and, hence, the decision-making process.

Mr. Ba Oumar dit Foussala, Procurement Specialist, added that, due to the withdrawal of NUTRICOM's NGO partners, each regional team will need to: 1) produce summary reports for each CNCR in their region; 2) produce a regional report (combining the reports of the various NCRs in their region); and then 3) submit these to the PMU, which will then incorporate them into the national report.

Mr. Sall Aliou also drew participants' attention to the array of indicators (38 in all for the nutrition component alone!) required of NUTRICOM by IDA, and to the cumbersome procedures that this entails in terms of constructing and interpreting them (in addition to the fact that not all of them are relevant.)

Mr. Jean-Pierre Habicht suggested that only a small number of relevant indicators be retained to help in decision-making, and that the list of those indicators be sent to Washington (to Messrs. Jean-Pierre Habicht and Kees Kostermans) so that there can be a constructive dialogue between the PMU and the group of experts participating in the video conference. He called attention to the importance of training and retooling for community nutrition agents (*agents de nutrition communautaire*, ANC) for proper data entry and accurate interpretation of the data sheets and records.

Mr. Mohamed Ag Bendeche, IDA consultant, pointed out the importance of documenting the project's failures as well as its successes, in order to ensure its replicability (and so that lessons learned from errors during this pilot phase might be put to good use when the project is expanded to the rest of Mauritania.) He attaches more importance to the validation of the project process than to project impact.

Mr. Youssouf Koïta, from UNICEF, 1) having noted that in Mauritania, the monitoring and evaluation system is always the neglected component during the project design phase; and 2) having tried to determine where such a system might fall short; 3) described his experience with the Expanded Program on Immunization (*Programme Elargi de Vaccination*, PEV), and stressed the need to assign responsibility to the 'decentralized' services at the regional level in order to create an efficient system.

Mr. Mohamed Lémine oul Moujtaba asked how the link between improved nutrition and income-generating activities (IGAs) was to be measured. Messrs. Kees Kostermans and Jean-Pierre Habicht observed that this was a complex issue, but that, in their opinion, that depended on how the information was to be used. Mr. Mohamed Ag Bendeche felt, for his part, that it was an issue of the use that is made of profits generated by such activity. (If, for example, they serve to finance a fund for assistance to seriously malnourished target populations, then the link between IGAs and nutrition is directly and positively demonstrated.)

Mr. Chérif Diallo reminded participants that, in order to ensure the sustainability of the CNCs (once the project closes), NUTRICOM has established some IGAs. Mr. Sidi oul Sidi Bouna, Micro-Project Specialist, gave a brief description of the IGA component and suggested that standard models of monitoring and evaluation reports be submitted to Messrs. Kees Kostermans and Jean-Pierre Habicht. They accepted this suggestion.

The videoconference ended at 7:30 p.m.

Decisions:

As a result of the videoconference, the PMU team and the Task Manager agreed:

- to include, in the terms of reference of the branch office managers, junior nutritionists and agro-economists, a precise description of their obligations (particularly in the area of monitoring and evaluation);
- to select a limited number of simple indicators for monitoring nutrition and IGAs; subsequently (starting in the 3rd or 4th months, for example), one could gradually introduce additional indicators deemed relevant, taking care not to overload the mechanism in place;
- to submit a file (for comment) to the experts who participated in the video conference in Washington, including: 1) a summary of the video conference; 2) selected monitoring and evaluation indicators and their justification; 3) standard format for monitoring and evaluation reports on the activities of micro-projects (Annex 5);
- to propose a model for the organization of monitoring and evaluation (see Annex 3); and, finally
- to bring the regional teams together in Nouakchott to inform them of the monitoring and evaluation system selected and to discuss it with them.

ANNEX 4: AGENDA

Monitoring and evaluation for Nutrition Investments in Bangladesh

Wednesday, May 21, 2003

8.30am – 4.00pm

World Bank, Washington DC

The Nutrition Team - Health, Nutrition and Population, Human Development Network

Objectives:

The primary objectives of this meeting is to consult with an expert panel on the following issues:

- 1 Given the design and sampling methods employed for the Baseline, Mid-term and End-line evaluation surveys in Bangladesh, how can we best analyze the Bangladesh Integrated Nutrition Project (BINP), data to produce a credible and quality product?
- 2 Review and advise on the evaluation design for the follow-on National Nutrition Project (NNP).
- 3 How can nutrition evaluation capacity best be strengthened for the future in Bangladesh?

A secondary objective will be to:

- 1 Provide inputs for a participatory learning/training monitoring and evaluation module for integration into the core course on nutrition being developed by the Human Development Network and the World Bank Institute (WBI).

Approach:

The meeting will primarily be a consultation between the Bangladesh team and the expert panel on how best to strengthen the Bangladesh BINP end-line evaluation and plan for appropriate evaluation and sampling designs for the follow-on National Nutrition Project (NNP). Materials from these consultations will be used as inputs for the monitoring and evaluation module within the core course on nutrition being developed in partnership between the Bank's nutrition team and the WBI.

References and documentation:

Resources and references for further reading will be provided for appropriate sessions and shall also be made available on the Bank's intranet.

Participants:

Participants for this meeting shall include expert panel members, Task Team Leaders from Bangladesh, the BINP Evaluation team, and nutrition thematic group members from the World Bank.

Agenda:

8.30 am: Coffee

9.00 am: Opening/Welcome/Introductions - *Chairs: Tawhid Nawaz/Milla McLachlan*

9.15 am: Bangladesh baseline and mid-term evaluation: progress to-date and analysis plan - *Rezaul Karim, Sascha Lamstein*

- 9.45 am: Contextual issues – *Nasreen Huq, ActionAid Bangladesh*
- 10.00 am: Questions/Clarifications - *All*
- 10.30am: Coffee
- 10.45 am: Brief commentary from panel members
Key design issues - *Jean-Pierre Habicht, Cornell University*
Sampling issues - *Ed Frongillo, Cornell University*
Monitoring issues - *Reynaldo Martorell, Emory University*
- 11.00 am: Open discussions on analysis plan for the BINP data, within the context of the design issues and the sampling frame
- 12.30 pm: Lunch
- 1.30 pm: Follow-up discussions on analysis plan
- 2.30 pm: Evaluation design for the NNP - *Barkat-e-Khuda, ICDDR, Bangladesh*
- 3.00 pm: Coffee
- 3.20 pm: Feedback on NNP design - *All panel members*
- 3.45 pm: Way forward - *Meera Shekar*
- 4.00 pm: Close - *Chair*

Strengthening Monitoring and evaluation in the World Bank's Nutrition Portfolio
Thursday, May 22, 2003
8.30am - 5:30 pm
World Bank, Washington DC

**The Nutrition Team – Health, Nutrition and Population,
Human Development Network**
(Interactive Learning Sessions Prepared in Partnership with the World Bank Institute)

Introduction:

In this era of greater attention to development effectiveness and competing demands for limited country and donor resources, monitoring and evaluation to build a strong evidence base for nutrition investments is becoming increasingly important. At the same time, important questions are being asked as to whether most evaluations address the right questions:

- What are the right evaluation questions?
- What are the best evaluation designs to help answer these questions?
- What additional methods can be used to complement traditional evaluation paradigms?
- How can monitoring data be used to strengthen/complement evaluation results?
- How can the Bank help build institutional capacity for monitoring and evaluation through country operations?
- What are the common monitoring and evaluation issues/constraints in Bank operations in nutrition?
- How can we maximize monitoring and evaluation within the context of common constraints faced in the field?
- How can the World Bank's Task Team Leaders strive to systemically build monitoring and evaluation into operations?

Approach:

The seminar is an interactive learning event addressing issues around monitoring and evaluation in nutrition projects. Task Team Leaders working on nutrition projects will present issues they face on the ground and an advisory panel will discuss specific monitoring and evaluation topics related to these issues. Following these discussions, participants will break out into four clinics focusing on:

- d) The Nutricom Project in Mauritania;
- e) Use of monitoring data for evaluation - case example from Madagascar;
- f) Developing capacity and commitment for monitoring and evaluation; and
- g) Sampling issues - case example from Eritrea

Objectives:

- 1) Review and understand the monitoring and evaluation issues in the World Bank's nutrition operations;
- 2) Provide an interactive learning opportunity for Task Team Leaders to strengthen monitoring and evaluation in their operations;
- 3) Identify opportunities for systemic strengthening of monitoring and evaluation within the World Bank's nutrition portfolio, including development of in-country capacity for monitoring and evaluation through learning by doing;
- 4) Provide inputs for a participatory learning/training monitoring and evaluation module for integration into the core course on nutrition to be developed in partnership between the nutrition team and the World Bank Institute (WBI).

Panel members:

Ed Frongillo (Cornell University)
Jean-Pierre Habicht (Cornell University)
Nasreen Haq (Action Aid, Bangladesh)
Reynaldo Martorell (Emory University)
Meera Shekar (World Bank)
Andrew Tomkins (Institute for Child Health, London, UK)
Howard Nial White (OED, World Bank)

Lessons and materials from this seminar will form a basis for a training module on monitoring and evaluation for a core course on nutrition to be developed by the Human Development Network and WBI.

Agenda:

8.30 am: Coffee

9.00 am: Introductions and Opening remarks - *Robert Hecht, Acting Director HDNHE*

9.10 am: Monitoring and evaluation perspectives from a Task Team Leader - *Kees Kostermans, AFTH2*

9.20 am: Choosing evaluation designs to suit the evaluation questions - *Jean-Pierre Habicht, Cornell University*

9.40 am: Monitoring as a complement to evaluation - *Reynaldo Martorell, Emory University*

10.00 am: Beneficiary incidence analyses - *David Coady, International Food Policy Research Institute*

10.20 am: Developing in-country capacity for monitoring and evaluation for nutrition: the challenge and the way forward - *Andrew Tomkins, Institute of Child Health, UK*

10.40 am: Coffee

Chair: Milla Mc Lachlan, HDNHE

11.00 am: Monitoring and evaluation issues in World Bank Nutrition Operations - *Harold Alderman, DEC*

11.10 am: Discussion

12.00 noon: Way forward - *Meera Shekar, HDNHE*

12.15 Lunch

1.15 – 3.15 pm: Clinics on specific monitoring and evaluation issues.

Parallel Sessions on:

A. Mauritania - Cherif Diallo TTL, WHO, UNICEF and Mauritania country team (Via Video Conference)

Facilitators: Kees Kostermans and Jean-Pierre Habicht (Room J3-044)

B. Use of Monitoring Data for Evaluation: Case example from Madagascar
Facilitators: Rae Galloway, Ritu Chhabra and Reynaldo Martorell (Sunlight Salon, G8)

C. Developing Capacity and Commitment for MONITORING AND EVALUATION
Facilitators: Milla McLachlan and Andrew Tomkins (Room G7-109)

D. Sampling Issues
Facilitators: Meera Shekar and Ed Frongillo (Room G7-043)

3.30pm: Coffee

Chair - Robert Hecht, HDNHE

4.00 pm: Panel discussion: How can the World Bank over-archingly strengthen monitoring and evaluation in it's nutrition operations? - *Panel members*

5.30 pm: Wrap up

ANNEX 5: PARTICIPANTS

Monitoring and evaluation for Investments for Nutrition Investments
World Bank
Wednesday, May 21, 2003
Participants List

Name	Organization	E-Mail
Carla Bertoncino	AFTH1	cbertoncino@worldbank.org
Caryn Bredekamp	SASHD	cbredenkamp@worldbank.org
Sadia Chowdhury	SASHD	schowdhury@worldbank.org
Edward Frongillo	Cornell University	EaF1@cornell.edu
Rae Galloway	HDNHE	rgalloway@worldbank.org
Michele Gragnolati	SASHD	mgranolati@worldbank.org
Jean-Pierre Habicht	Cornell University	Jh48@cornell.edu
Nasreen Haq	Action Aid, Bangladesh	nasreen@actionaid-bd.org
Rezaul Karim	Dhaka University, Bangladesh	Krezaul@udhaka.net
Barkat-e Khuda	ICDDR B, Bangladesh	barkat@icddrb.org
Sascha Lamstein	Tufts University	Sascha.lamstein@tufts.edu
Amanda Liddle	HDNHE	aliddle@worldbank.org
Reynaldo Martorell	Emory University	Rmart77@sph.emory.edu
Milla McLachlan	HDNHE	mmclachlan@worldbank.org
Tawhid Nawaz	HDNHE	tnawaz@worldbank.org
Meera Shekar	HDNHE	mshekar@worldbank.org
Andrew Tomkins	Institute of Child Health, UK	a.tomkins@ish.ucl.ac.uk
Meri Vanharanta	WBIHD	mvanharanta@worldbank.org
Howard White	OEDST	hwhite@worldbank.org

Strengthening Monitoring and evaluation in the World Bank's Nutrition Portfolio
World Bank
Thursday, May 22, 2003
Participants List

In Washington D.C.

Name	Organization	E-Mail
Harold Alderman	AFTHD	halderman@worldbank.org
Carla Bertoncino	AFTH1	cbertoncino@worldbank.org
Caryn Bredenkamp	SASHD	cbredenkamp@worldbank.org
Vanessa Brooks	ICDDR B, Bangladesh	vbrooks@icddrb.org
Ritu Chhabra	HDNHE	rchhabra@worldbank.org
Sadia Chowdhury	SASHD	schowdhury@worldbank.org
David Coady	International Food Policy Research Institute	D.Coady@Cgiar.Org
John Fiedler	LCSHH	jfiedler@worldbank.org
Edward Frongillo	Cornell University	EaF1@cornell.edu
Emanuela Galasso	DECRG	Egalasso@worldbank.org
Rae Galloway	HDNHE	rgalloway@worldbank.org
Michele Gagnolati	SASHD	mgranolati@worldbank.org
Jean-Pierre Habicht	Cornell University	JH48@cornell.edu
Nasreen Haq	Action Aid, Bangladesh	nasreen@actionaid-bd.org
Robert Hecht	HDNHE	rhecht@worldbank.org
Rezaul Karim	Dhaka University, Bangladesh	Krezaul@udhaka.net
Barkat-e Khuda	ICDDR B, Bangladesh	barkat@icddrb.org
Kees Kostermans	AFTH2	Kkostermans@worldbank.org
Sascha Lamstein	Tufts University	Sascha.lamstein@tufts.edu
Amanda Liddle	HDNHE	aliddle@worldbank.org
Tazim Mawji	HDNSP	tmawji@worldbank.org
Reynaldo Martorell	Emory University	Rmart77@sph.emory.edu
Judith McGuire	LCSHH	jmcguire@worldbank.org
Milla McLachlan	HDNHE	mmclachlan@worldbank.org
Elaine Ooi	OEDPK	eeoi@worldbank.org
Christine Pena	AFTH4	cpena@worldbank.org
Siddhartha Prakash	AFTKL	sprakash@worldbank.org
Vinita Ranade	ECAVP	vranade@worldbank.org
Meera Shekar	HDNHE	mshekar@worldbank.org
Andrew Tomkins	Institute of Child Health, UK	a.tomkins@ish.ucl.ac.uk
Meri Vanharanta	WBIHD	mvanharanta@worldbank.org
Howard White	OEDST	hwhite@worldbank.org
Jeffrey Yau	University of Pennsylvania	yyau@ccon.upenn.edu

In Mauritania (via video conference)

Cherif Diallo	AFTH2	cdiallo@worldbank.org
Mohamed Ag Bendeck	World Bank	
Mounina Mint Abdellah	NUTRICOM	
Ba Oumar dit Foussala	NUTRICOM	
Sall Aliou Mamadou	NUTRICOM	
Sidi Ould Sidi Bouna	NUTRICOM	
Diabira Issaka	NUTRICOM	
Fatimetou mint Mohamed Abdallahi	NUTRICOM	
Niang Saidou	Ministry of Health	
Mohamed Lémine ould Moujtaba	UNFPA	
Thierno Coulibaly Ousmane	UNFPA	
Youssouf Koita	UNICEF	
Mohamed Lémine ould Mohamed	WHO	

wb64386

C:\Documents and Settings\wb64386\My Documents\Strengthening M&E - Nutrition.doc
October 28, 2003 12:06 PM



**Health, Nutrition
and Population**

World Bank
1818 H Street, NW.
Washington, DC 20433 U.S.A.



THE WORLD BANK