
Methods for

the Evaluation of

the Impact of Food

and Nutrition Programmes

Edited by David E. Sahn, Richard Lockwood,

and Nevin S. Scrimshaw



THE UNITED NATIONS UNIVERSITY

METHODS FOR THE EVALUATION OF THE IMPACT OF FOOD AND NUTRITION PROGRAMMES

**Report of a workshop on the Evaluation of Food and Nutrition Programmes,
sponsored by the United Nations University on behalf of the United Nations ACC
Sub-committee on Nutrition, held at the Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA, in September 1981.**

**Edited by David E. Sahn, Richard Lockwood,
and Nevin S. Scrimshaw**

THE UNITED NATIONS UNIVERSITY

© The United Nations University, 1984

The views expressed in this publication are those of the authors and do not necessarily reflect the views of the United Nations University.

The United Nations University
Toho Seimei Building, 2-15-1 Shibuya, Shibuya-ku, Tokyo 150, Japan
Tel.: (03) 499-2811 Telex: J25442 Cable: UNATUNIV TOKYO

Second printing, 1988
Printed in Singapore

WHTR-6/UNUP-473
ISBN 92-808-0473-1
United Nations Sales No. E.88.III.A.11
02500 P

The United Nations University is an organ of the United Nations established by the General Assembly in 1972 to be an international community of scholars engaged in research, advanced training, and the dissemination of knowledge related to the pressing global problems of human survival, development, and welfare. Its activities focus mainly on peace and conflict resolution, development in a changing world, and science and technology in relation to human welfare. The University operates through a worldwide network of research and post-graduate training centres, with its planning and co-ordinating headquarters in Tokyo, Japan.

Other UNU Titles of Interest

Effective Communications for Nutrition in Primary Health Care

Edited by Jean E. Andersen and Aree Valyasevi

Concentrating on the experiences of developing countries, this volume advocates the effective communication of nutrition and health information as a key component of primary health care. The book provides a framework for evaluating nutrition communications needs, methodological guidelines, and suggestions for programme evaluation.

WHTR-11/UNUP-634 ISBN 92-808-0634-3
220 pages, 16.4 x 23.9 cm, paper-bound, US\$20

Food Composition Data: A User's Perspective

Edited by William M. Rand, Carol T. Windham, Bonita W. Wyse, and Vernon R. Young

Knowledge of what is in the foods that are eaten around the world is of critical importance and at the same time extremely inadequate. Data on food composition underpin research and policy in a number of important areas of public health, dietetics, nutrition, and epidemiology; they are critical for key decisions of bilateral and international assistance agencies and play a major role in all phases of the food production and manufacturing industry, both locally and on a global scale. These needs stand in stark contrast to the availability and adequacy of food composition data. In this volume prominent workers in the field present their views and experiences concerning the importance of food composition data and its current problems and what must be done to improve the situation. It provides an essential introduction and survey of the field for anyone interested in or expecting to be involved with the gathering, compilation, or use of food composition data. It will also be a useful reference for university courses on food and nutrition.

WHTR-10/UNUP-633 ISBN 92-808-0633-5
240 pages, 16.4 x 23.9 cm, paper-bound, US\$20

Research Methods in Nutritional Anthropology

Edited by Gretel Pelto

A comprehensive manual of anthropological methodologies applicable to field studies in nutrition, this volume describes strategies of field research in nutritional anthropology, determinants and cultural components of food intake, methods for collecting and analysing data on energy expenditures, and statistical methods for nutritional anthropology.

WHTR-9/UNUP-632 ISBN 92-808-0632-7
In press, 16.4 x 23.9 cm, paper-bound, US\$20

PREFACE

The United Nations University agreed in early 1981 to take the lead, on behalf of the United Nations agencies represented on the Sub-committee on Nutrition of the UN Administrative Committee on Co-ordination, in convening a workshop to address the problem of how to evaluate the impact of supplementary feeding programmes targeted to vulnerable women, infants, and children. The original charge to those who prepared papers and participated in their review was to develop straightforward guidelines and procedures for evaluating programmes with nutritional goals. The objective of the workshop was to compile state-of-the-art knowledge, which could then be made available to planners and field workers involved in or contemplating the evaluation of the impact of food and nutrition programmes.

A number of conceptual issues provide a framework for the papers in this publication. In the first place, they are primarily focused on the issue of determining the impact of an intervention. We recognize that, more often than not, evaluations are process-oriented, examining service delivery, project plans, and so forth; and it is not our intention to suggest that process evaluation can be ignored in the search for programme impact. However, there is a need now to design methods and measures to determine whether, and to what extent, programmes have an impact.

A second issue is the challenge of how to do evaluations in the absence of adequate baseline data, or where there are financial constraints that militate against a rigorous evaluation effort. Outside evaluators are often called on to perform their task under less than optimal conditions, confronted by severe constraints on data availability and financing, and especially by a lack of baseline data. It is our strong belief that all projects should have evaluation procedures integrated into the planning and implementation cycle, but the reality is that such a concept will take time to gain acceptance. The workshop participants and authors of this book agreed that, in the interim, something has to be done to evaluate the impact of large-scale projects already under way.

A third important issue is the recognition that a food and nutrition programme can have a variety of beneficial effects on the programme participants that go beyond traditional health impacts or that are not measurable using indicators and techniques usually employed in the field. Therefore, specific attention is given to the numerous social, economic, and political consequences that may accompany a feeding programme (see chapters 10, 11, and 13 respectively).

Furthermore, this book differs from previous reviews of evaluation methodologies by proposing and examining new methodologies that have not generally been part of nutrition and health surveys or evaluations. These include the use of such indicators as immunological competence (chapter 7), work performance (chapter 8), and cognitive performance and behavioural assessment (chapter 9). In addition, up-to-date information is provided on the use of the array of anthropometric and laboratory methodologies (chapters 3 and 4) as well as on the role of clinical data, vital statistics, and dietary methods (chapters 5 and 6).

As a corollary to this perspective of considering the range of possible measures of outcomes, it is the aim of this book not only to examine impact on targeted individuals actually enrolled in a programme but to broaden the purview to include the variety of benefits that may accrue to other household members. Such an approach is suggested because it is possible that examining impact on children in terms of traditional parameters may yield little information, but very tangible negative or positive effects on other targeted individuals or family members may be found if innovative indicators that go beyond traditional limited nutritional status measurements are employed in the field.

A further underlying premise is that before any well-designed evaluation effort can be implemented, it is necessary to perform preliminary reconnaissance in order to gain detailed knowledge of the political life-styles, cultural features, and household patterns as well as the demographic, geographic, and logistic characteristics of the population the programme is designed to affect. Similarly, other systematic steps, such as field trials and assessment of training, equipment, and administrative and personnel requirements, are also integral components of any well-performed evaluation. All these aspects of the evaluation protocol should be clearly and concisely delineated in the standard operating procedures, following the basic principles of programme evaluation and research to the extent feasible. These considerations, however – as well as such issues as the presentation, publication, and dissemination of results – are not addressed in detail here. The reader who is unacquainted with these fundamental aspects of the evaluation process should consult appropriate technical literature and seek expert advice.

The first two chapters do introduce the basic concepts for designing evaluations during programme implementation, and the stages of evaluation of ongoing programmes. Chapters 3-10 address the major themes of the book: what measurements and indicators should be employed to assess the impact of food and nutrition

programmes, and how and under what conditions should they be selected and used. These chapters cover the numerous measurements of nutritional and other possible impacts of supplementary feeding programmes. Chapter 12 considers issues related to data recording and processing. Chapter 13 provides a framework for assessing the impact and role of an intervention from a macro-policy perspective, as contrasted with the previous chapters that look at impact from the point of view of the individual or household. The last chapter details how and why to develop a built-in evaluation system ; it was included in response to the consistent urgings of the workshop participants that an internal evaluation system should be incorporated into all projects, regardless of whether or not complementary external evaluation procedures are undertaken.

CONTENTS

Preface	v
1. Basic Concepts for the Design of Evaluation during Programme Implementation J.-P. Habicht, John B. Mason, and H. Tabatabai	1
2. Stages in the Evaluation of Ongoing Programmes John B. Mason and J.-P. Habicht	26
3. Measuring the Impact of Nutrition Interventions on Physical Growth Reynaldo Martorell	46
4. Measuring Impact Using Laboratory Methodologies Barbara A. Underwood and Abraham Stekel	65
5. Measuring Impact Using Clinical, Morbidity, and Mortality Data Lincoln Chen and Rachel Feilden	94
6. Measuring Impact by Assessing Dietary Intake and Food Consumption Miriam Chavez and Ruth Huenemann	127
7. Measuring Impact Using Immunologic Techniques Ranjit Chandra	144
8. Measuring Impact on Physical Activity and Physical Fitness John Durnin	153
9. Methods for the Behavioural Assessment of the Consequences of Malnutrition Ernesto Pollitt	179
10. Anthropological Methodologies for Assessing Household Organization and Structure Gretel Pelto and Pertti Pelto	204
11. Micro-economic Analysis in the Evaluation of Supplementary Feeding Programmes Beatrice Lorge Rogers	226
12. Data Recording and Processing Michael Guzman, Ricardo Sibrian, and Rafael Flores	233
13. Policy Evaluations David F. Pyle and Mitchel B. Wallerstein	252
14. Built-in Evaluation Systems for Supplementary Feeding Programmes – Why and How Roy I. Miller and David E. Sahn	265
Participants	289

1. BASIC CONCEPTS FOR THE DESIGN OF EVALUATION DURING PROGRAMME IMPLEMENTATION

J.-P. Habicht, John B. Mason, and H. Tabatabai

Introduction

Evaluators often face restrictive conditions; for example, they may be called on to evaluate the effects of a programme some years after the programme has begun. Often there are negligible baseline data and no control groups. Much of this and the next chapter are therefore written from the perspective of the unfortunate reality of having to evaluate an ongoing programme in the midst of implementation. It is hoped, however, that evaluations may be better planned so as to relax many of the constraints often encountered.

Present practice is to collect data on programme participants, and possibly non-participants, then to use statistical manipulations to investigate associations between programme delivery and outcome variables. These methods tend to be expensive and may be difficult to apply in developing countries as routine procedures; moreover, consideration of the questions that should be addressed before applying such methods reveals that they often turn out to be unnecessary. We, therefore, propose procedures that can be widely applicable to a range of programmes in operation.

When a nutrition intervention programme is proposed, the following sequence of questions (adapted from ref. 1.) need to be addressed to establish whether the intervention can in principle affect the performance, health, and survival of individuals—that is, whether the intervention is in fact justified.

1. Is a deficit of food or specific nutrients causing disease, decreased performance, or untimely death in individuals?
2. How detrimental is this deficit to individual performance, health, or survival? In other words, what is the dose-response relationship?
3. Is it possible to decrease or eliminate the deficit (or its effects) in individuals?
4. How prevalent is the deficit (and its effects) in the population? Is the problem increasing or decreasing?
5. What proportion of ill health, decreased performance, and untimely death in a population may be ascribed to this deficit now and in the future?

6. Is it possible to decrease or eliminate the deficit (or its effects) in the population?
7. What are the expected benefits, costs, and side-effects of the proposed intervention on the deficit (or its effects) in a large population, given the results of intervention trial studies in small populations? How long would the intervention need to run?
8. What are the actual benefits, costs, and side-effects of the intervention undertaken in the large population on the deficit (or its effects)? Is the actual benefit worth the actual cost and the actual side-effects?

Before expanding further on proposed procedures for the evaluation of ongoing programmes, it is worthwhile to introduce some fundamental terminology in this context. Specifically, the literature differentiates “(basic) research,” which is done to ascertain basic scientific facts independently of their applications to programmes, from “evaluative research,” which is done to assign a probability statement of causality to the relationships of an intervention in a community context and the observed nutritional and health impacts, in order to determine the viability and replicability of a given programme design and plan of operation, and from “operational programme evaluation,” which ascertains whether an ongoing programme is attaining its objectives (see; e.g.; 2; 3). Operational programme evaluation is the subject of this and the next chapter.

The evaluations of ongoing programmes considered here do not attempt to answer questions concerning the scientific basis for interventions (e.g. questions 1 to 4 above). These are best tackled through experimental research. However, unless these have been addressed in designing the intervention, the evaluation may be pointless. Similarly, the logic for performing evaluative research to determine whether a pilot intervention or demonstration project can successfully apply scientific principles in a field setting must be borne in mind before a large-scale project is undertaken and an operational programme considered (e.g., questions 5 to 8 are examples of areas of evaluative research).

In practice there is no clear distinction between research and evaluation in the methods used. In fact, many of the researcher’s tools can be used to answer questions relevant to evaluation. There is, however, a major difference in the kind of questions being addressed, and in the appropriate combination of methods to address each question. It is a grievous mistake to call research “more scientific” than evaluation. Good science is fitting the appropriate methods to seek an answer; in this sense, a good evaluation is as scientific as good research is.

Unfortunately, most works on operational programme evaluation in nutrition have been misdirected in emphasizing the researcher’s concern to substantiate the probability of causality while neglecting the other important questions that managers, administrators and funders ask. This misdirected research methodology often neglects the fact that operational programme evaluation can be divided into (a) “summative

evaluation," which examines the outcome of a programme, and (b) "formative evaluation," which monitors procedures and activities so as to improve the programme design and the delivery of services. We refer to the former as outcome or impact evaluation and to the latter as process evaluation. The document by Sahn and Pestronk(3) provides a careful review of the theoretical literature in health and human services about programme evaluation, and presents abstracts of many evaluations of nutrition intervention field programmes—clearly differentiating process from impact evaluation and evaluative research from operational programme evaluation. Other evaluative research studies of nutritional programmes are reviewed by Habicht and Butz (4).

This chapter discusses a number of basic concepts needed for designing evaluations appropriate to the decisions to be taken based on their findings. These concepts are intended to provide the basis for the practical stages in evaluation given in chapter 2.

Costs of Evaluation

Useful evaluations entail activities that apparently are not necessary to manage the logistics of a programme, and hence incur additional costs. Riecken has said, "... my experience with evaluation is that there are few bargains, and usually you get no more than you pay for," and "When an evaluation is cheap and quick, it is often also not very good" (5, p. 365). Without entering into this debate, we can nevertheless learn something from the actual levels of expenditure on evaluation. On the one hand, the World Bank review of monitoring and evaluation of projects in East Africa in 1979 for example, gives a figure of US \$12.8 million for monitoring and evaluation of 28 projects averaging US \$460,000 each, or somewhere around 0.5 to 5 per cent of total project costs (6). The US Government estimates that 1 per cent should be used to evaluate programs in health. On the other hand, Kielmann et al. consider that "... it should be neither uncommon nor unreasonable to budget 20 per cent to 40 per cent of total project costs for analysis of project-generated data and project evaluation" (7, p. 33). Evidently, the scale of expenditure of the project itself has some influence on these calculations. Similarly, the purpose of the evaluation of the project is important—while relatively high expenditures may be essential for pilot or experimental projects, it seems unlikely that more than perhaps 5 per cent of project costs would be made available for routine evaluations of large-scale service projects. Such expenditures would often be sufficient to allow useful evaluation, if it is recognized that causality may not be established.

Ultimately, determining the percentage of programme expenditures spent on evaluation is a subjective judgement, as it depends upon which costs are assigned to the delivery of programme services and which to the distinct effort labelled evaluation. In fact, the best evaluation may appear to cost nothing, as it would be an integral component of programme design and implementation.

Purposes of Evaluation

The question "How do we tell if a programme has an effect?" is incomplete without knowing why one needs to know. Common reasons are :

- to decide whether to continue the existing programme or not,
- to redesign the programme if necessary, or
- to decide whether to do similar programmes elsewhere.

Those involved in the programme often have different expectations about the purpose and results of an evaluation. It is important that the decisions which will be made using evaluation findings be clearly understood and agreed on. The evaluators must then tailor not only the design of the evaluation to the purpose of the evaluation but also the presentation of the results to that purpose. For instance, results presented as if the purpose were to decide on the continuation or termination of a programme are inappropriate if the purpose of the evaluation is to improve the programme. Evaluation cannot be seen in isolation from who asks the question. It is not so much that the principles and practice of evaluation of ongoing programmes are unsatisfactory but that the whole decision-making process in nutrition and food aid programmes needs improvement.

The purposes and issues addressed in evaluation depend on who is asking the questions. The following sequence of basic issues to be addressed are of particular interest to different audiences :

- Is the intervention performing as expected? (Programme managers, administrators, and funders)
- Is the intervention worth continuing? (Administrators and funders)
- Should it be extended? (Administrators and funders)
- Is it causally linked to improved nutrition? (Researchers, scientists, and others concerned with basic mechanisms of cause and effect)

This sequence begins with considering whether the programme is performing adequately, and can progress to seeking to ascribe causality between the intervention and the outcome. The sequence approximates the changing concerns of project management and administrators and the researchers concern with causality; however, causality, if it can be shown, is also important to all aspects of management, programme design, and policy. However, causality is difficult and expensive to establish, and the more that certainty on causality can be dispensed with, the easier the evaluation becomes. Project management can often, in fact, get by with the knowledge that the beneficiaries are improving, even if they cannot be sure this is due to the programme.

Part of the information needed to address questions such as those given above can be obtained by evaluating project design and from process data. Moreover, these data can be used to screen out those projects that are unlikely to have any important effect on outcome, and thus are not worth evaluating further. This procedure is set out in

subsequent sections. Other decisions required in establishing purposes of an evaluation centre on the degree of certainty required in linking outcome to programme delivery, and these need to be explained in more detail at this stage.

Different purposes of evaluation demand varying degrees of plausibility or certainty for the conclusions reached from the evaluation. The purposes, in the order of increasing need for higher levels of certainty (elaborating from the sequence of questions just given) are :

1. improvement in programme management,
2. continuation of funding,
3. replication of the programme in similar conditions,
4. replication of the programme in dissimilar conditions,
5. finding basic research results about cause-effect relationships.

The methodological and data requirements of responding to the differing needs of these purposes for certainty and plausibility entail, in order of increasing expense and difficulty :

- a. collection of data on process and outcome for participants only (programme data),
- b. collection of data through *ad hoc* surveys,
- c. advanced statistical analysis,
- d. control group(s) of some kind,
- e. collection of before-after data,
- f. highly-standardized measurements,
- g. randomized intervention,
- k. double-blind research designs (blind intervention and blind assessment).

It would be useful to consider how these two lists could be matched. Each item in the first list is taken up individually. (This discussion is summarized in table 1.1.)

1. Evaluation for programme management seeks to determine whether programme services are being delivered as planned to the intended target groups and whether the (gross) outcome is acceptable. The objective of this "adequacy evaluation" is to reveal the possibilities for improvement in programme management. It does this by relying on programme data relating to process and participants (method a). It may, on occasion, require survey data as well (method b).
2. The decision whether to continue funding of a particular programme often requires adding advanced statistical analysis (method c) to the requirements for adequacy evaluation (method a) and, possibly, survey work (method b).
3. Replication of the programme in similar conditions usually requires data on some form of control groups (method d), and/or surveys (b), in addition to the requirements for purpose 2.
4. Replication in dissimilar conditions entails, at the minimum, both control groups (d) and survey data (b), as well as advanced analysis (c). Sometimes it may be preferable to use a quasi-experimental design (see section on Design below), e.g. to add before-after data (method e), to standardize measurements carefully (method f), and/or to use a randomized design (method g).

TABLE 1.1. Appropriate Data Collection and Analysis for Different Decisions

		Decisions				
Data and analysis needs		Increasing need for confidence				
		(1) Manage- ment	(2) Continuation of funding	(3) Replication in similar conditions	(4) Replication in different conditions	(5) Basic research on causality
Increasing difficulty ↓	(a) Process data and outcome for participants only	+	+	+	+	+
	(b) <i>Ad hoc</i> surveys	(+)	(+)	+	+	(+)
	(c) Advanced statistical analysis		+	+	+	+
	(d) Some kind of control group(s)			+	+	+
	(e) Before/after data				(+)	+
	(f) Highly standardized measurements				(+)	+
	(g) Randomized intervention				(+)	+
	(h) Double-blind trials					+?

(+) = occasionally.

5. Basic research involves most of these requirements, and sometimes it may even be possible to employ the ideal research design, the double-blind randomized trial (method h).

The confidence with which the conclusions in each of the above cases is reached can be considerably improved with strong theory relying on good scientific evidence from elsewhere.

Setting Programme Objectives as a Basis for Evaluation

Process evaluation demands formulating implementation and performance objectives against which the programme can be evaluated. For the manager's questions (e.g. is the programme performing as expected?) and usually for the funder-administrator's questions (e.g. is the programme worth continuing, extending, etc.?), the comparison is between the procedures and activities of the programme and some preset standards, generally set out in the programme work-plan. The first prerequisite is, therefore, that the essential activities be stated in objectively measurable units. This is possible even for such an amorphous exercise as curative primary health care (1). Actual performance relative to these standards is ascertained through process evaluation.

A requirement for outcome evaluation is to establish objectives prior to assessment. These must be explicitly formulated as an acceptable difference from a standard, or as

a minimum improvement from some baseline. These quantitative standards of achievement should correspond to the implicit objectives of the programme and should be understood and agreed on by those who must use the results of the evaluation. Experience shows that the exercise of making stated and implicit objectives more explicit will often reveal hidden objectives, some of which are even contradictory. This is why a consensus about the programme objectives is one of the necessary first steps to an evaluation.

Almost inevitably, programme objectives change as a programme evolves. However, changing definition of objectives during evaluation of single projects should be avoided because it is rare that the design of the evaluation can deal with new objectives. For example, a recent review of supplementary feeding programmes discussed whether the more important effects of these programmes were in terms of income distribution since the supposed objectives of improving child nutrition were seldom reached (8). However, no comparison was made with quite different programmes that might be more efficient in changing income distribution. While this may be a reasonable question in general, changing objectives for an individual programme requires more fundamental decisions.

Once the underlying outcome is identified conceptually, the next step is to identify the measurable variables related to the outcome of concern. The major portion of this book discusses that step relating desired outcome (e.g., improved nutrition) to a measured variable (e.g., anthropometry). Subsequent chapters develop the relationship between the conceptual outcome and the measurements more fully.

Finally, the statistical test used to judge the reality of a measured difference (either between treatment and control groups, or between treatment and a standard) results in a statement that most of the time (which is usually specified as 95 to 99 per cent of the time) such a measured difference will be found if the true difference is not smaller than some quantity. In designing an evaluation one must further state how one is willing to miss identifying a true difference of more than a specified magnitude. This statement refers to power analyses (see, for example, [9]). These steps of specifying procedural and impact objectives, translating those objectives into measurable variables, specifying the minimum or maximum acceptable difference in that variable, and doing the power analysis are prerequisites for *any* quantified evaluation.

The sad fact is that the research giving scientific justification to a programme is often so lacking that these steps are impossible. Experiments in the precise setting of the proposed programme may not always be needed (or possible). However, there needs to be a marshalling of the evidence from previous evaluations, experiments, and scientific knowledge, to serve as a basis for designing a relevant evaluation. Unfortunately, this is all too seldom done.

Investigating Causality

Investigating causality involves exploring whether there is a link between programme activities and outcome. The logical sequence of questions in evaluating whether a measured outcome is plausibly caused by programme activities may be summarized as follows:

- Is the outcome adequate? (Of concern to programme managers, administrators, and funders.)
- Is there a statistical association between intervention and outcome? (Of concern to administrators and funders and to researchers.)
- Is the outcome due to the intervention? (Of concern to administrators and funders and to researchers.)

If there is no statistical association or if the outcome is inadequate:

- Were the statistical methods used correct?
- Was the intervention relevant and adequate?
- Were the measurements of outcome valid and reliable?
- Were the recipients likely to benefit?
- Was the sample size adequate?
- Was there negative confounding?

If there is statistical association and the outcome is adequate:

- Is the association likely to be causal? Discard confounding. (See table 1.2 on internal validity.)
- What direction was the causality?
- What mechanisms linked the intervention to the outcome?
- Can the findings be extrapolated to the population as a whole (i.e. do they have external validity)?
- What was the cost of the effect and of the marginal effect?

The basic question is whether there is a statistical association between the putative cause (the intervention) and the outcome or effect (e.g. improving nutritional status). Seeking associations is useful only when one needs some level of certainty that the programme is causally related to the outcome. Showing an association requires comparison of measured outcome in at least two groups that receive different intensities of programme intervention. This may mean comparing two groups such as control and treatment, showing correlations between different levels of programme delivery and outcome, estimating regression coefficients between programme activities and outcome, statistically controlling for other influences, and so on. Controlling for influences on the outcome other than the programme which mimic a programme effect is called "controlling for confounding influences."

The questions to be asked about a programme and its evaluation are very different depending on the results of the statistical tests of association. Although these tests