REVISION 1

METHODS OF SELECTING, DESIGNING AND EVALUATING FOOD AID PROJECTS

by

J.B. MASON and J.G. HAAGA

Cornell Nutritional Surveillance Program
Ithaca, New York 14853

November 1983

Paper prepared under contract for World Food Programme

The CNSP is supported by Cooperative Agreement # AID DSAN CA-0240 between the Office of Nutrition, Bureau for Science & Technology, USAID, and the Division of Nutritional Science, New York State Colleges of Human Ecology & Agriculture and Life Sciences, Cornell University, Ithaca, New York 14853.

A report of research of the Cornell University Agricultural Experiment Station

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report is the result of a two-week consultancy, during which WFP project documents were reviewed and discussions were held with WFP staff. As a contribution to an overall study of evaluation of project food aid, we assess information needs for project selection, design, evaluation, and re-design, and propose ways to meet them. The changes are meant to be incremental, not radical, and feasible as goals for the next five or so years.

The life-cycle of WFP supported food aid projects is described and the points at which major decisons are made (II).* We concentrate on project selection and design, and interim evaluation and possible re-design of projects. The major conclusions are as follows.

1. At the project selection and design phase, there should be a greater element of choice (III). This requires a portfolio of possible projects, rather than just responses to isolated requests. In this way, WFP could concentrate its resources on projects most likely to benefit the malnourished.
2. The unique advantage of food aid is that it is of most value to the very poor, whether directly consumed by recipients or sold for additional cash (much of which, in the case of the poor, will be used for higher food consumption anyway). Accordingly, WFP, while cooperating with other donor agencies, need not attempt to duplicate their efforts by supporting the most "bankable" projects. Even more than other agencies, WFP should evaluate proposed or existing projects in terms of humanitarian (or immediate consumption) goals, rather than conventional criteria of economic return. WFP's special expertise should be how donated food is supposed to help achieve project goals (IIIB, IIIC). This has important implications for WFP's information needs. To capitalize on its advantage, WFP needs information about the proposed beneficiaries of projects - who they are, how the benefits are supposed to reach them, and how much effect the WFP contribution could have. This information is needed both for selection of projects and for guiding data collection for evaluation (IIID). It is not currently presented, as least in systematic form, in WFP project documents. We review WFP Plans of Operations and Interim Evaluations; at present they contain little information on who is targetted,

* Numbers in parentheses refer to report sections.

how much is spent per recipient, or precise goals and outcomes of programs (VIA). We offer examples of the types of information required, ways in which it could be organized, and how it can be collected during the course of a project (IIIE, IV, VIB).

3. The current structure of the WFP project cycle - in particular the emphasis on major interim evaluations - seems useful; it allows for more flexibility in project re-design than do the procedures of most agencies. To make the best use of the interim evaluations, though, more attention should be paid to building into the project from the start routine monitoring of deliveries, distribution, numbers and types of recipients, food prices, achievement of development goals (where applicable), cause of delays or modifictions, etc. Though implementing this suggestion could depend on many country- and project-specific factors, we propose some examples of types of information needed and ways to present it (IVA, IVB).

4. Often this information is already generated, but must be arranged to bring out important points about program coverage and the degree to which programs are targetted to the malnourished (VIB). We also propose that greater use be made of existing reporting systems in recipient countries, partly to expand the range of information available for WFP evaluators, but also to help strengthen administrative practices likely to be useful in other development projects. As an illustration, the potential uses of the nutritional surveillance system in Botswana is discussed (VII). The implications of our proposal for greater emphasis on continuous assessment, rather than the one-shot interim evaluation, are that some of the funds for the latter should be shifted to the former, and more resources overall should be devoted to monitoring and evaluation. This could be justified if the payoff, both to WFP and to recipient governments, were better project selection and more adaptive management.

5. Impact evaluations are needed, but properly done, require more resources for data collection and analysis than are typically available. We propose that WFP not attempt thorough-going impact evaluations for all or even most projects, since most decisions that need to be made during the course of a project can be supported with much less information than impact

evaluations demand. We discuss the theory and practice of impact evaluations, and propose that WFP generate, presumably with central funding, a series of case studies covering the major types of projects (vulnerable group feeding in clinics, school feeding, food-for-work, etc.) in selected countries at different levels of socio-economic development and administrative capacity (IVC, Annex IV).

6. WFP should have periodic "country reviews", preferably in cooperation with other donors. These, rather than specific project documents, would be the appropriate means for considering possible negative effects of food aid at the macro policy level. We suggest ways in which such country reviews could be carried out to contribute at several points to improved decision-making (V).

I. INTRODUCTION

The terms of reference for this report are given in Annex I. They require our recommendations on possible improvements in selecting, designing, and evaluating food aid projects. The work was undertaken in the more general context of a review of WFP methods of evaluation. We have therefore started by trying to define the problems in present procedures, as a prerequisite to suggesting solutions. Deficiencies in present procedures need to be spelled out (and we were encouraged to do this). We would stress that this is only in order to try to help produce more effective use of project food aid, and is not criticism for its own sake. A first draft of this paper was provided for discussion in August 1983, and the present draft has tried to take the resulting comments from WFP and from colleagues here in Cornell into account. The paper aims to complement other recent reports for WFP on project food aid, notably those by Hogan (ref. 1) and Katona-Apte & Maxwell (ref. 2). Points covered in those reports are not covered here in the same way, because there is agreement on most positions taken. This document moves ahead from these to suggest additional ways of assessing and evaluating projects. A particular focus is on the information needs. The scope of the paper is from project selection through to final evaluation. However, we intend to concentrate on certain aspects of these where innovations could be made, rather than to review the entire process comprehensively. The approach is incremental, not radical. Moreover, not much emphasis is put on where the procedures are trying finally to get to in, say, five or ten years' time - the principles are given in (ref. 1) and (ref. 2) and indeed in much other literature on project planning and evaluation.

Similarly, we have laid out concepts and procedures for monitoring and evaluation in several other documents (ref. 3-6) and for economy of effort, these will not be repeated in detail.

The objective is therefore to make suggestions that could be tested fairly soon, which would be logical first steps in modifying procedures for project selection, preparation, management, monitoring and evaluation. The intended audience for the paper is, first, WFP management - to get reactions on whether the suggestions are sensible, and possibly to modify or elaborate them based on

these reactions. Second, it is hoped that the suggestions can be tested by missions, field staff and government officials and others with whom WFP works.

A. Issues considered

Our interpretation, based on discussions in Rome and on documentation reviewed, is that WFP is concerned with the following issues. WFP feels that projects should be increasingly selected and planned to more cost effectively meet better specified objectives. The problem is seen as having several facets: the most suitable projects for food aid support are not always selected; once the type of project is selected, the design may be inadequate; and the design itself is not always implemented. Objectives should be better defined, and more closely related to feasible achievements given the available inputs and activities. Projects need better monitoring to ensure that plans are followed, or that the plans are changed in the light of experience and changing conditions. When implemented, projects do not always meet objectives, and indeed may have certain unintended negative effects. Throughout this, more and better information is needed to guide decisions.

B. Initial considerations

The different phases in the life of a project provide a natural basis for organizing our considerations. The planning and evaluation stages are referred to as: Selection - Preparation - Interim Evaluation - Re-design/Further Phases. The term "routine evaluation" is used refer to in-built procedures during project implementation, whose use is to provide information for management and re-design throughout the project. This life-span is illustrated (and expanded) in Figure 1.

The suggestions made in this paper refer to: selection and preparation; evaluation, re-design and management. These overlap. For example, many criteria for choosing **between** possible projects are similar to those for deciding the design of **one particular** project - since design involves choosing (at least implicitly) between alternative designs.

There are certain simplifying concepts applied throughout. First, the primary concern is with **food** and its effects. Second, conventional project and

evaluation criteria are appropriate, suitably modified, to food aid. Moreover, project objectives could usually be achieved using either food or money as the resource. (The major exception would be disaster relief). Third, some assumptions are reasonable, even if unproven in a particular project context: thus, we will assume that a positive direct effect on food consumption, and potentially on nutritional status, is usually achieved if enough of the right food reaches the right people at the right time. (However, this assumption should be checked with selected impact evaluations.) Fourth, producing information is only justified when specific decisions by identified people require it; and the information must be tailored to these needs. Fifth, many important potential negative effects (e.g. disincentive effects on food production) are better handled by a country review process than project-specifically. This is discussed further in Section V.

Two types of projects only are considered: public works (e.g. food-for-work) programmes and feeding (or nutrition - usually for vulnerable groups) programmes. Not all activities supported by food aid fall into these categories - incentives to attend clinics, or improve private land, for example. However, the principles involved are similar, and if useful suggestions can be made for these two project types, these suggestions can be modified to cover the other uses of project food aid. Emergency and relief projects are included insofar as they operate by public works or feeding programmes.

Two types of objectives are also considered: development objectives (for which economic indicators are used), and humanitarian objectives (for which nutrition is here used as the primary measure). It will be argued that food-supported public works projects should have both development and humanitarian objectives; feeding programmes should have primarily humanitarian objectives.

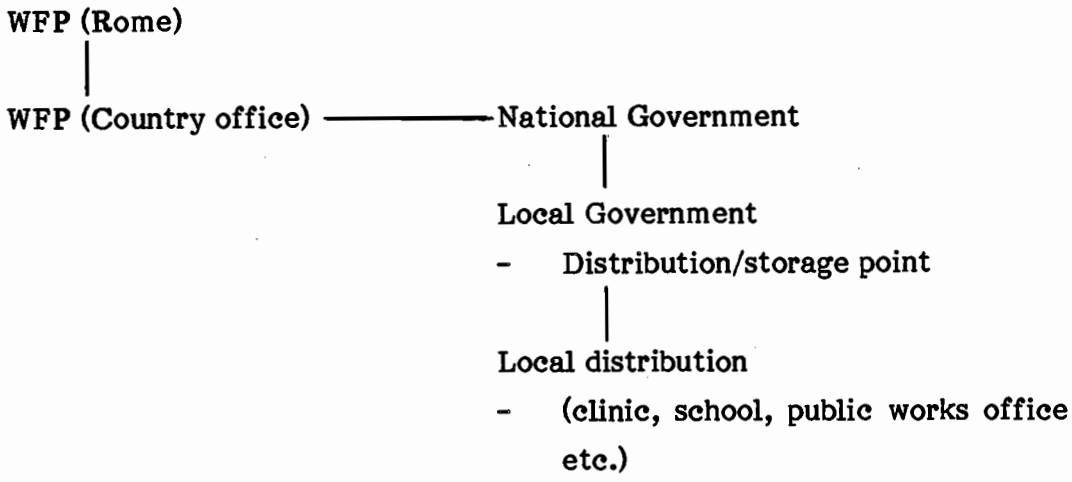
The possible ways in which the resources available - food or money - can be used to reach such objectives are illustrated in Figure 2. In principle, either food or money (if necessary, to buy food locally) could be used to reach WFP project objectives, both development and humanitarian. One option is to sell the food, and use the money obtained to fund projects - this is known as "programme food aid" (path B in Fig. 2). If food is provided as a grant, not

attached to specific projects, this is somewhat equivalent to general budgetary support from donors (path A in Fig. 2); the government then allocates funds freely according to its own priorities. An alternative possibility would be to allocate food for specific projects, but then convert the food to cash (by sale) to support that project (path D in Fig. 2). Finally, food can be used as the direct input into the project (e.g. as food-for-work) - (path E in Fig. 2). WFP is constrained largely to use food directly for projects (project food aid - path E), even though other resources (i.e. money) or other ways of using the food (i.e. programme food aid) might, other things being equal, be as good or better ways of reaching a particular WFP objective. The option of programme food aid is not considered - but this does not imply that changing the constraint on this might not in fact be the best approach to achieving certain objectives.

Finally, food is generally regarded as income-in-kind, except possibly for certain direct feeding programmes. However, the question of commodity choice is not addressed here. If the food aid substitutes for food normally bought, or is sold, the 'alpha factor' approach seems appropriate; in this approach it is suggested that the criterion for selecting donated commodities (e.g. DSM vs bulgur wheat) should be the optimum ratio of the local market value of the commodity to the cost of providing that commodity to the beneficiary. If food aid is consumed without substitution, cost per calorie, acceptability, etc, are the considerations. These ideas are well covered elsewhere (ref. 2).

II. POINTS AND PROCESSES OF DECISION-MAKING

The structure of decision-making and related information flow is taken as:



In words, decisions (illustrated in Table 1) on project selection and design are understood to involve requests or other information from local or national governments, leading to discussions with WFP country offices, hence to requests being forwarded to Rome for decision. The results of these decisions are a flow of food, to the national government and hence to distribution points. Finally, information on the arrival and distribution of the food should flow from the local level up through the system, to allow management decisions to maintain the system.

It needs to be stressed that project planning means making choices between possible projects. If there are no choices possible, no planning is possible. Hence improving the planning process may require expanding the possibilities for choice. Creating this space for choice, as a prerequisite to improving projects, may itself require some policy decisions. This policy is further discussed below. Our understanding of the way in which these decisions are made, and how they might be improved, is as follows.

(a) A request for food assistance arises in many different ways, often from discussions between WFP staff and central government officials. The

request is usually fairly imprecisely formulated, specifying the type of activity (e.g. public works in a certain area; support to MCH), with approximate numbers of beneficiaries. Usually, this is in the context of contribution to a larger-scale activity - a development programme with other inputs, or the running of regular government services (e.g. health or education). A preliminary decision is now made, between WFP - Rome and the WFP country office as to whether this is worth pursuing. If this decision is positive, a more detailed project proposal is prepared, in country by WFP and the government, probably with assistance from a mission sent by Rome. A second decision is then made, on whether the project is acceptable and should go ahead. This leads to a project summary being submitted to CFA, where it is almost invariably approved.

This process implies a choice between competing possible uses of the food, both between countries and between projects within countries. Only the latter choice is considered here. It seems that both the basis for this selection - how to rank the worthiness of projects -and the procedure for actually making the selection, could be improved.

Some possible improvements that might be considered are as follows:

- Discussing alternative uses of food aid at the earliest stage. This discussion could be based on a country review process where needs for, possible benefits of and problems with food aid are assessed.
- Participating with the government and other donors (who mostly support projects with money, not food) in selecting projects, for example as part of country programming exercises.
- Securing and analysing a minimum level of data on which to base this implicit decision on ranking and selection of projects.
- When a selection-in-principle is made, extending the analysis to optimise the project design in terms of feasible development and humanitarian objectives.

- Setting criteria for project preparation (including information requirements) and establishing the option of not deciding to support a project until such criteria are met.

Attention to these points early in project planning should lead to clarification of what decisions are currently feasible, and thus how much space there really is for planning. It seems that the decision not to support a request is rare. Although initially it seemed that the element of choice and (hence ability to plan) was very limited at present, our discussions did elicit areas of important potential choice which could be better identified and more effectively used, without in the first instance requiring far-reaching changes in policy. For example, (a) choices on project activities (e.g. a forestry project versus a rural development project) are sometimes made, and (b) selection of the geographical or administrative areas in which to place projects are more commonly made.

(b) As implementation of the project gets under way, WFP's primary responsibility is to get food into the country on time, and to support logistic costs. Some monitoring of food delivery to distribution points is carried out by reports from the government. Only very limited input is made by WFP into this process. On-going evaluation depends on whether the government has such a system, which is not usually the case. Management decisions at this stage by WFP are mainly concerned with delivery of food to the port-of-entry, and not with what happens thereafter.

An improved procedure would involve greater assistance by WFP to the government for producing and using information for routine evaluation, to assist programme management. This would be required especially in countries where administrative systems for project monitoring are poorly developed, that is, usually, the countries most in need of food aid. An example of a proposed system for Botswana, based on existing structures,

is given below in Section VII. Developing such systems will often require some investment of resources in producing the necessary capability for reporting, and analysing data, within the government. However, without this, little improvement can be expected.

(c) Towards the end of the first phase of the project, after two to three years usually, a request may be received for continuation or expansion of the project. Usually, this leads to sending a mission to carry out an interim evaluation. The data available at this stage vary. Logistic information and rate-of-disbursement of food are usually available. Less information is available on who benefits, and how far objectives (economic and nutritional, for example) have been reached. There is therefore still an inadequate informational basis for deciding on whether to extend, expand, re-design, or end the project.

An improvement on this process could be to capitalise on the common WFP procedure of preparation - first phase - interim evaluation -second phase etc, in order to build in flexible and adaptive project management. Essentially, this means that decisions will be made more frequently on project design and management. Mistakes can be corrected, and implementation move adaptively to more cost-effective ways of meeting objectives. But this will require more investment in information and management.

(d) Evaluations of impact - i.e. how far the project itself has caused objectives to be reached - are few. This is partly because after several years of implementation, especially without adequate data collection, it is extremely difficult to elucidate what effects the project may have had over time. Partly it is because the pressure of events prevents a sufficient priority being given to such evaluations. Their purpose may not be to improve the project itself, unless the project is being extended. But impact evaluations are essential for designing new projects - it is crucial to know what works, and what does not. In the absence of this

information, policies and project designs are made in a vacuum. We suggest that more attention be given to impact evaluations in selected cases, primarily to allow generalizations on effectiveness, and hence better decisions on future projects. Part of the information can come from futher analysis of data from on-going evaluations, provided that attention has been paid early on, to ensuring the necessary data are available for the relevant comparisons.

Since the benefit of an impact evaluation is primarily for future projects, the charge for the impact evaluation should not necessarily be made to the project itself. It would seem reasonable to find the costs for such evaluations from a number of different projects, or from central funds. Indeed, unless a policy decision is made on these lines, it seems unlikely that the required impact analyses will be done, simply because the cost will be too high to be found from any one project budget.

III. PROJECT SELECTION AND PREPARATION

A. Constraints

Decisions on project selection and preparation take place within certain critical constraint. These could possibly be overcome, which would do much to ease difficulties in project selection and preparation; but for the present we assume they will continue to apply. The constraints are as follows. First, there is pressure to commit food resources. This could mean that otherwise unpromising projects are accepted, and that time for adequate preparation is too limited. Second, human and financial resources for project preparation come from central funds, not from project funds. Unless central funds can keep pace with the resources to be committed, planning and evaluation efforts cannot easily be expanded to meet new project needs. In other words, the present staff available seems fully loaded with the existing number of projects to be dealt with. Increased food resources will only increase this load, unless they are linked with more funds to support planning and evaluation at least in the short term. Improvement in procedures is likely to require more staff time to establish per project, hence there is a very real constraint in terms of

funding to improving (or possibly even maintaining) the present procedures. Third, as mentioned earlier, WFP food cannot (usually) be used as programme food aid (i.e. converted to money and the funds thus generated used to finance projects).

In passing, the obvious step of concentrating resources - using more food on fewer projects - has presumably already been considered as a way to overcome the first two constraints. Given that present levels of food input do not seem to be having obvious impact, concentrating resources could make some sense. Since much of food aid is equivalent to income, there is no clear need to be bound by rations calculated on nutritional grounds.

Finally, a major influence on project selection is clearly government priorities, and this constraint is very frequently called upon to justify questioned project designs. However, the need to cooperate with government priorities does not in fact preclude more attention to dialogue and negotiation with governments, precisely as for other inputs by other agencies on this scale. Here too, a country review process, rather than negotiations over a specific project, may be the appropriate forum.

B. Concentration on food-specific aspects of project selection and preparation

WFP undertakes a wide range of project types (particularly regarding those with development objectives), thus an extensive variety of different expertise is needed for planning these projects. One option is for WFP staff to be expert in subjects ranging from, for example, forestry and rural development to maternal and child health. However, it seems more realistic for WFP to provide expertise specifically concerning food-related issues - the special considerations of using food to attain development and humanitarian objectives - rather than to aspire to cover itself the full range of necessary skills. Using food aid would not necessarily justify creating a generalized development agency, unless much greater resources could be devoted to project planning and management. At the moment, specialists are drawn from other agencies and institutions for project planning and evaluation. In fact, the institutional expertise is often completely relied on when food forms only one input to a

development project - if the project is considered sound by the investment agencies, the project's viability is not seriously questioned. This seems reasonable, and the discussion below is intended to imply **more** reliance on specialised institutions (e.g. in development and health) with WFP providing the food-related expertise. The suggestion is thus that WFP become more integrated with other agencies, for example, participating in their project preparation and selecting projects at the same time, indeed comparing project proposals with them.

C. Basis for selecting projects using food as the resource

Usually, money is an alternative resource to food for attaining WFP project objectives. For example, in public works projects, activities and development objectives are not peculiar to food-supported projects, only the use of food as the input is.

This comparison between food and money as the resource for supporting projects is crucial in selecting them - as will be discussed in more detail later. Although the value to be put on food inputs for comparative purposes is problematical, it seems logical that the decision to use food to support project activities in appropriate projects should be made only when money is not available to support these activities. The valuation of food determines to a large extent the expected cost-effectiveness (or expected cost-benefit) which could then be used for direct comparisons with money-supported projects. In theory, the cost of using food consists of the market value of the food itself, plus delivery costs; if food is not valued at the local market rate (being donated) this may make it appear a relatively cheap resource. In reality, there is some opportunity cost for the food (it could be used in other ways), but nonetheless paying wages with donated food may be cheaper from the government's standpoint than the equivalent in cash wages. The cost effectiveness depends also on a second set of considerations: it is widely thought that projects supported by food are less productive than those paying cash wages, partly because in these projects may be designed with less concern for productivity, and partly because food provides less motivation to be productive.

In any event, it is suggested that food aid projects should be selected by criteria comparable to those applied to other projects, and the suitability and availability of food or monetary resources for supporting them be made explicit. In principle there are the following combinations of possibilities:

Project Type	Suitability of.... Money	Food
A	Yes	No (e.g. buying equipment)
B	Yes	Yes, but may be less efficient than equivalent monetary input (e.g. public works)
C	Yes	Yes, may be equivalent to monetary input (e.g. school feeding)
D	Yes	Yes, has advantages over monetary input (e.g. where supply of food is inelastic)
E	No	Yes, (e.g. relief where no local food available).

The main questions arise concerning the examples of public works projects (type B) and direct feeding (type C). For public works projects money is an alternative to food, and if food is valued at the market rate is likely to be more effective. Direct feeding programmes (e.g. school feeding) are in fact open to similar questions: providing money to schools to buy food locally could be as effective in supporting the programme (depending again on the value put on the donated food) and have advantages over providing food directly, such as in stimulating local production. Finally, for some large-scale projects (type D) providing food directly may have advantages if rapidly increasing demand risks causing higher prices and localized food shortages.

The question of project selection arises particularly for types B and C, where food may be as good a resource as money, or at least would suffice. Most projects considered fall into these categories. Rationally, out of the range of possible projects to be supported, some ranking of priority is needed. It is suggested that possible projects be ranked based on (a) conventional economic criteria (e.g. internal rate of return) and (b) humanitarian criteria,

e.g. nutritional returns. The advantage of this is that the methodologies for assessing economic returns are familiar (ref. 7-8), and for nutritional returns becoming so (ref. 9). The major steps needed to achieve this are:

- to generate a selection of projects, or implicit designs for individual projects once selected;
- to get the information to provide some priority ranking.

The result, in principle, is that potential projects would have some level of expected cost-effectiveness assigned in advance, along the lines:

Project	Development Potential	Humanitarian Potential
1	High +	Low
2	High	Medium
3	High -	High
.	.	.
.	.	.
.	.	.
X	Medium	High
.	.	.
.	.	.
.	.	.
Y	Low	Medium

Those projects with high development potential requiring external resources can usually be picked up by other donors, using money. Unless food is a better resource (types D and E above) and even if it is as good a resource (type C above) this is reasonable. It is suggested that, to maximise cost-effectiveness, and use the unique resource of food, WFP should pick up those development projects where food is a good resource which do not (quite) qualify on economic grounds along for other donor funding, and which have a high potential for meeting humanitarian objectives. In practice this will mean selecting more

poverty oriented projects. Food aid will de facto be a resource for reaching the poorest. This conclusion is far-reaching: it requires acceptance that food does not compete with money for supporting projects, hence that the less "bankable" projects will be supported by food aid. The advantages are that food will be directed more effectively towards those most in need (but taking some account of development objectives). This policy is likely to have a greater effect on raising the food consumption of the poor, since the poor have a greater marginal propensity to consume food for any given increment in income.

For projects, such as vulnerable group and school feeding programmes, that have primarily or only humanitarian objectives, only the 'humanitarian potential' ranking is important.

These criteria apply for between project selection, and for selection of alternative designs within projects. Obviously, if there are no choices to be made, this or any other modification to selection procedures is irrelevant. However, often choices are implicit - one project is implicitly favoured over others, even if the others are not described. The main proposal here is that justification for a particular selection or design be made more explicit, based on some analysis. Some practical outcomes of these suggestions are now discussed.

(1) For projects with development objectives, attempts should be made to select projects from a range of possible projects put forward, whether these are intended for money-or food-support. This means participating in country programming exercises, country reviews, etc., alongside other donors. Food should be chosen to support certain projects that cannot qualify for financial support, yet have high potential for meeting humanitarian objectives. For projects with only humanitarian objectives, economic justification is clearly not relevant, but the trade-offs are still implicit. This and other aspects of project selection and design could be very usefully facilitated by a process of country reviews - where the major

opportunities for food investment, as well as the major concerns such as negative effects on food supply, can be reviewed on a non-project specific basis.

(2) Conventional methods for ex ante assessment of development potential, (i.e. economic return) and humanitarian potential (e.g. nutritional impact) be used (ref. 7-9). This means relying on those agencies and institutions able to make these assessments. In the early stages of project selection, these assessments will be qualitative and based on experience - but this is no different from any other project identification.

(3) As well as providing a justification for selecting a particular type of project, the project documentation should justify selecting the particular design or project (within type) over other possible uses of the food. This means that the project summary report - which may need to be modified in format - should include statements on:

- expected economic return of the project (which may be less than other investment projects supported by money - but may preferentially reach the poor)
- who benefits directly from the project, compared with other possible beneficiaries, relative to their need
- level of expenditure or food supply per beneficiary
- justification for using food rather than money to support the project activity.

(4) Project objectives should be specifically stated, in the project documentation, for both development and humanitarian objectives. This will include estimates on how far the poor (and/or malnourished) are included in the project. At the present state-of-the-art, such statements cannot be fully quantitative, with any confidence. Nonetheless, a move towards attempts to quantify expected objectives is needed, and the evaluation process discussed below should progressively enable these to become more credible.

(5) The analysis required for (3) and (4) above will be considerably more than is customary at present. However, if the cost-effectiveness is

to be improved, there is not alternative to devoting more resources to design, both pre-project as discussed here, and during project implementation, as discussed in the next section. The procedure of country reviews could be very useful, and give economies of scale (section V). By this means the potential for food aid in any one country could be analysed, and certain questions hitherto project-specific addressed in overall terms. Because of the objectives of food aid can be broad, it may be that these country reviews could form part of economic or sector surveys undertaken by other agencies (e.g. the World Bank). This would also deal more appropriately than on a project-specific basis with macro-level problems such as negative effects on food production, displacement of commercial imports etc.

(6) Methods for these analyses need to be developed, and supporting documentation produced. The resources required for application of these methods will vary depending on the data available, the complexity of the project, other analyses in hand (e.g. when food aid is part of a larger project) and so on. These studies should be done in-country, rather than by desk review. Perhaps some 2-4 person/months per project would not be unreasonable for such analyses for large-scale projects.

Given the existing schedules for preparation and submission of projects to CFA, devoting a few months to more careful project preparation would seem to be worthwhile.

D. Information Needs for Project Selection and Design

The relevant questions in selecting a project are:

(1) What are the proposed objectives, in economic and humanitarian terms? How do these compare with objectives of other possible projects?

(2) Is food a suitable resource for this project? Is it to be provided at a sufficient level that a significant effect could be expected?

(3) How far are unintended negative effects a risk (e.g. disincentive effects, disruption of services)?

It is suggested that these questions and their answers be explicitly covered in project documents, e.g. the 'project summary'.

The data required for assessing **economic** objectives are:

(a) cost of inputs (food and non-food)

(b) estimates of economic returns overall.

Estimating economic returns has been suggested using socially-weighted cost-benefit analysis (SCBA) (ref. 1-2). We suggest that the usual unweighted methods be used to calculate internal rate of return, or an equivalent indicator like payback period. We agree with the principle of weighting humanitarian objectives, but suggest that WFP use a cost/effect approach (i.e. effect measured by the outcome objectives without attempting to put these in monetary units) balancing economic and humanitarian objectives qualitatively rather than trying to combine these into a single number (as SCBA does). Combining the different outcomes into a single number has not proved useful in project decision making in most agencies, since in practice it has obscured the tradeoff between objectives.

The data required for assessing **humanitarian** objectives are:

(c) cost of inputs (food and non-food)

(d) relative need of proposed recipients (i.e. of target groups)

(e) estimates of direct consumption or income effects for recipients, for example relative to net food receipts in cash equivalents per head per time

(f) estimates of long-term consumption or income effects (e.g. from investment in public works), for example as income per head per time

In each case, the "cost of inputs" should include not just the direct budgeting cost to the government or WFP of resources used and staff time, but also any indirect costs attributable to the project.

The data for this assessment could be derived in various ways:

- from experience with other projects - e.g. from previous evaluations (b, e, f)

- from household budget, farm management etc. surveys, to estimate economic effects (depending on project type) (b,f)
- from socio-economic, nutritional, etc. data analysed to show relative need of different sections of the population (e.g. in country review exercise) (d).

Answers to the second question - is food a suitable resource? - derive again from experience with previous projects and from management and logistic considerations. However, it seems reasonable to propose that below a certain level of input into a multi-component project (e.g. less than 20% of wages) this could be ignored - that is, the input is likely to be too low to have much of an effect anyway.

Answers to the third question, on negative effects, such as on food production at national level, on production patterns at household level (e.g. dependency on food aid, hence disincentive effect on regular production, foor or otherwise or employment) or disruption of services etc, should, it is suggested, be generally handled at country level, again as part of a country review. This can then give guidelines to apply to individual projects.

Analogous information is required to design the project once its overall selection has been made. The question now becomes:

What are the outcome **objectives** (economic and humanitarian) of the proposed design, and how do these compare with other possible designs?

Beyond this, project implementation and evaluation requires further information, on:

Targetting: how are recipients defined, how many of them? How is the project oriented to the needy?

Level of delivery: what is the intended net receipt of benefits (e.g. the food ration) per capita?

(What proportion of income or wages does this represent? Is the value of the food (for consumption or even resale) high enough to make an impact on alleviating beneficiary poverty of malnutrition?

Quantified information on objectives, targetting, and level of delivery assist both planning and subsequent monitoring and evaluation. In fact, the information needs form a continuum from planning to evaluation.

Economic objectives can be expressed as internal rate of return, etc, and humanitarian objectives as immediate and long-term consumption, with relative need. An example is given in Table 2.

Targetting objectives may sometimes be definable quantitatively again in terms of relative need. The simple classification given below is appropriate.

		Needy?	
		Yes	No
To Receive	Yes	a	b
Food Aid?	No	c	d

"Needy" can be defined either as children malnourished (below some cutoff point of i.e. height and weight indicators) or as households below a locally relevant poverty line. This classification allows indicators of coverage - percentage of needy reached (a/a+c) and focussing - percentage of needy in target group (a/a+b) compared with the percentage in the overall population (a+c/a+b+c+d). Clearly, a poverty oriented project should have a higher percentage of needy in the target group than in the population over all. These data require a socio-economic classification of the population, which can use nutritional status as an indicator.

Level of delivery objectives may be equivalent to 'immediate net effects', in the table, defined as amount of food provided per head or per household, in cash equivalent terms.

The statement (e.g. in the project summary) should therefore give details of these four aspects (outcome objectives, targetting, targetting, level of delivery, goals/activities). Two hypothetical examples follow.

E. Examples of Project Statements

For a public works project, a statement along the following lines could be imagined. "The government development plan envisages extensive efforts to

develop the northeastern province, the poorest in the country. A series of projects for water control, feeder roads, and increased production of drought-resistant crops has been proposed, with a total investment requirement of x million dollars. Of this, y million dollars is available from government and donor resources. A and B districts have suffered from repeated drought, and the population of 10,000 households is among the poorest in the province, with malnutrition well above the average. The proposed project in districts A and B was selected for WFP assistance as having moderate economic potential, but reaching those most in need. The expected internal rate of return is 8.0%. (This compares with 15-25% for projects in higher potential areas.) The project will provide for water control and feeder roads that otherwise would not be built.

The project will directly benefit 1,200 households who are 30% of the lowest income group in northeastern province. 60% of the targetted recipients are in the lowest income group for the province (note that this may have to be defined in occupational or other operational terms) compared with 40% for the province as a whole. The project is thus strongly poverty-oriented. The level of delivery of food to participants is to be $x per head per year, representing 20% of average current income." (There should follow a description of project activities, number of rations to be handed out, out etc.)

A vulnerable group feeding programme might be described in the following terms: "30% of pre-school children are moderately or severely malnourished in Region A; in absolute terms this means about 30,000 children. It has been agreed to devote X% of available food resources to pre-school child feeding through selected MCH clinics. In the six provinces out of the 12 in the region where malnutrition is greater than average, MCH clinics will be supplied with food. In these six provinces, there is an average prevalence of 40% malnutrition, and 60% of the malnourished children live in these six provinces. All mothers attending clinics will be given take-home rations equivalent to half the child's requirements, and the ration will be doubled for mothers with malnourished children. The expected effect on outcome is not known precisely, but an estimated reduction of 20% of the prevalence of malnutrition during the

3 year life of the project (that is, from 40% prevalence down to 32%) would be regarded as adequate."

Besides summarizing the expected rates of economic return and expected impact on incomes and/or food consumption of the poor, these statements should also clearly describe the mechanisms through which the benefits are supposed to occur. This description is useful for several reasons: 1) It can expose points at which factors internal to the project (poor logistics, failure to recruit the right participants, etc.) jeopardize its chances for success. 2) It can show how and where the success of the project depends for its success on factors outside the control of project managers (prices paid for export crops, completion of rural electrification before village industries can get underway, etc.). 3) It can identify intermediate indicators that should be watched during interim evaluations to see if the project may be achieving the expected results.

One widely adopted approach to this sort of careful description of proposed projects is the "Logical Framework" of the U.S. Agency for International Development. (This is fully described in ref. 10.) We find this approach generally useful, but there is a danger in practice that in its highly stylized form it becomes just a routine form to be filled out, instead of an occasion for clear formulation of the project at an early stage. (Also, the distinction drawn in this framework between "Goals" and "Purposes" of a project does not seem very crucial).

An alternate method, one that could be modified for each particular project to avoid the "recipe" approach, would be the familiar box-and-arrow diagram, relating project inputs (personnel time, use of facilities, donated commodities) to expected outputs (increased employment, higher incomes, higher food consumption). Figure 3 shows an example is based on a dairy development scheme. Drawing in the intermediate steps is important. In one interim evaluation report of a dairy development scheme in Bolivia, the argument simply runs as follows: "Many people in Bolivia consume too little animal protein. Milk is a good source of animal protein. Therefore, using donated milk powder to help a new dairy plant get started is a good idea, and WFP should contribute." (This is a summary of Bolivia 2358). This skips too many steps, as Figure 3 illustrates. What conditions are necessary for the

project to benefit the people with inadequate diets? Either they will be able to buy more milk (i.e. milk prices come down or, if prices are fixed, rationing ends), or they will have higher incomes (from working on or in the factory, or delivering milk, or selling milk from their own herds to the factory, or working for others as herdsmen). How might the project fail to benefit people with inadequate diets? Possibilities are if the big new plant does not have lower costs and a lower-priced output of milk, or if it does and some of the poor are currently small herdsmen who will be undercut by it, or if unskilled labor is displaced by conversion from arable to pasture to supply the new milk market, or simply if the pathways by which benefits were supposed to flow don't materialize and the poor don't participate, either as consumers, workers, or herdsmen.

A full impact study might be impossible, but as Fig. 3 suggests, intermediate indicators can give project managers an idea of what is going on. What is happening to milk prices in the capital? Is milk more available in slum shops? What happened to those who used to sell milk? Who is being hired on the dairy farms? Are the plant's operating costs really lower, or are continuing subsidies necessary to keep prices down? Who gets the jobs at the plant? The interim evaluation could address questions like these only if the pathways of benefits were laid out clearly in the early project documents and if those charged with routine monitoring were asked to look for these indicators in addition to the data on food deliveries. Otherwise, all that could be given in the evaluation documents is the records of delivery of commodities to the plant and assertions that the local agricultural economy has not collapsed. Greater specificity early in the planning stages allows for more useful evaluation studies; less needs to be taken on faith. In the next section, we discuss further the steps for evaluation after the project has begun. A very useful early step for WFP/Rome, we feel, would be to generate a series of case studies of projects - how they were selected, designed, and what happened. These would be a start in guiding the adoption of improved procedures.

IV. EVALUATION FOR MANAGEMENT, PROJECT DESIGN & REDESIGN

The previous section discussed setting quantified objectives for selecting and designing projects. These objectives are the same as, and essential, for evaluation. Evaluation centers upon keeping track of how far these objectives are likely to be being reached.

Terminology in evaluation is confused. A brief review of how terms are used in this context is given in Annex II.

Here we refer to: 'routine evaluation', meaning on-going information derived primarily from project sources, for management purposes; 'interim evaluation', in the same way as currently used by WFP, for reviewing project performance after a few years of implementation to decide about expansion, redesign, continuation, or ending of the project; and 'impact evaluation' to elicit net effects due to the project. These three ideally lead into each other, and there seems little argument that the long-run aim should be to build on-going evaluation procedures into project management. The data needs in the continuum from planning to evaluation are shown in Table 3. The three aspects are considered individually below.

A. Routine evaluation

The main questions to be answered in the context of a food aid project are:

(1) Is enough food getting to the right people at the right time?

Subsidiary questions are:

(2) Are development goals being achieved (for appropriate projects)?

(3) Are outcome indicators (e.g. improvement in nutrition) adequate?

Clearly, if the project design is correct, a positive answer to the first question should mean that the next two questions are also positively answered. Thus, answers to the first question are always needed, preferably for the whole project area; whereas answers to (2) and (3) can if necessary be sought on a sample basis. The difficulty, again, is that insufficient documented experience has been brought together to know whether project designs are reasonable (Figure 3 is an attempt to restate such designs to highlight the areas of

uncertainty). Thus attention to (2) and (3) is still widely needed. Even question (1) is only worth answering if the management capability exist to correct problems so identified. The primary purposes of routine evaluation are to help maintain and improve project management; and to allow adaptive redesign of the project as circumstances change, or as assumptions in the project design are re-assessed.

Before considering how to get the necessary data, it is important to define closely the minimum data output required. Dummy tables are the best way. The indicators for: 'Is enough food reaching the right people at the right time?' are:

(a) food delivered per head; (b) coverage etc. (see below); and (c) deliveries and coverage over time, e.g. monthly. In tabular form, examples are given below.

(a)

Month 1	Area	No. of Recipients	MT Food	kg/hd/month
	(1)	(2)	(3)	(3/2)
	A			
	B			
	.			
	.			

or, better,

Month 1	Area	No, % of Recipients $< x$ kg/hd	No, % Recipients $\geq x$ kg/hd
	A		
	B		
	.		
	.		

(x = level of delivery objective)

(b) **Month 1** **Area A**

		Recipients? Yes	No
Targetted?	Yes	a	b
	No	c	d

$\frac{a}{a+b}$ gives delivery, etc.

which summarizes to:

Month 1	**Area**	**% Targetted who received food**
	A	(a/a + b)
	B	
	.	
	.	

(c) is then the aggregation of (a) and (b) over time:

AREA	**MONTH** 1	2	.	.
A	Cell entries = kg/beneficiary = % targetted for received food			
B				
.				
.				

Finally, the marginal totals of (a), (b) and (c) need to be calculated at different administrative levels (e.g. provincial, national).

Indicators for achievement of goals (Question 2) are already in use in WFP documents. To these could be added the intermediate indicators, such as discussed in the section IIIE, relevant for a particular project design.

For outcome indicators such as prevalence of malnutrition (Question 3) the analogous tables are simply:

	MONTH			
AREA	1	2	.	.
A				
B	Cell entries = Prevalence of malnutrition			
.				
.				

This is discussed further in references 3 to 6. For now, we will concentrate on data needs for Question 1.

How to get these data? In fact, much of the raw data often exists in administrative records, or if they do not exist, accounting methods can be adapted to record them. An initial decision has to be made on the level at which to report, usually between individual or, for example, village. This depends on the targetting method: if villages, but not individuals within the villages, are targetted, the village-level reporting is appropriate. If there is individual selection, e.g. by screening malnourished children or landless workers, then the individual level may be appropriate. Certainly in the health sector for feeding programmes, weight charts are now so common that tallying from these can often provide the required data.

The source data are thus quite straight-forward, at their simplest:
for village-level (e.g. by month): No. of recipients; No. in village; tons of food distributed; hence kg per head per time, and percentage of population covered.
for individual level: meets eligibility criteria? Received correct ration?

The requirements to set up such a system involve the usual steps of designing, or modifying, reporting, tallying and summary forms; field testing these; training for their use (usually this means training those responsible for distribution of the food); setting up and training for analysis at different levels in the administrative hierachy; and so on.

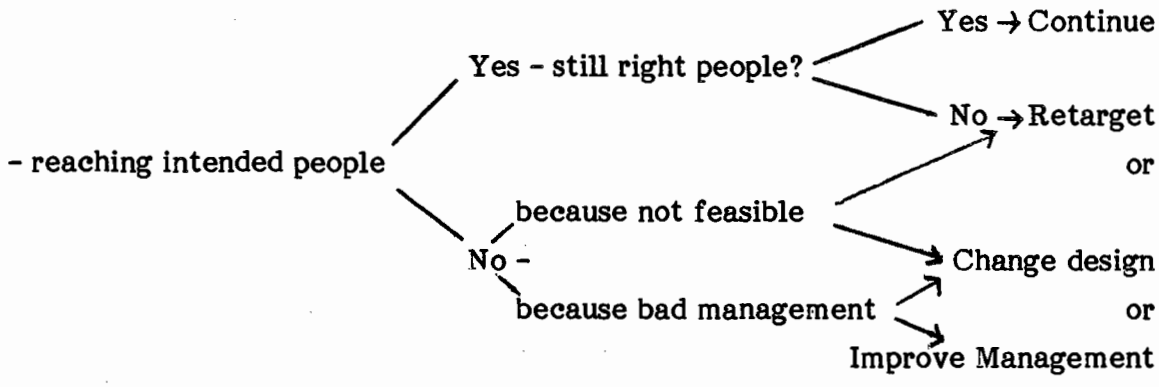
B. Interim Evaluation

The problem with interim evaluation arises primarily when there is no routine evaluation procedure to provide the necessary data. At present, this is the usual case. There are two alternative approaches, and the choice probably depends on the time and resources available for the interim evaluation. Redeployment of the resources currently used, i.e. for missions, might cover much of the necessary expense.

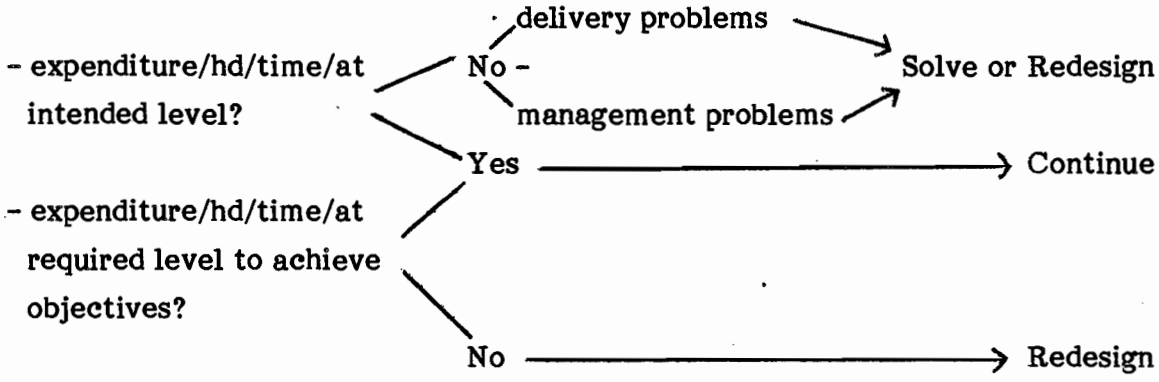
The first approach is to carry out the evaluation essentially as a cross-sectional exercise - as described in reference 4, a summary of which is attached as Annex III. This could involve a retrospective survey, and/or retrieving records and analysing these. The second is to use the resources available to begin routine evaluation, let this run for the minimum period of time necessary to get enough data to go on, and use this. At a guess, this might require a year overall - about six to nine months longer than a cross-sectional evaluation.

The use of the interim evaluation is to decide on continuation, expansion etc. of the project - equivalent to some degree of re-design: it thus overlaps in its requirements with the section on design, and in its indicators with both design and routine evaluation indicators. The decision process can be represented as follows:

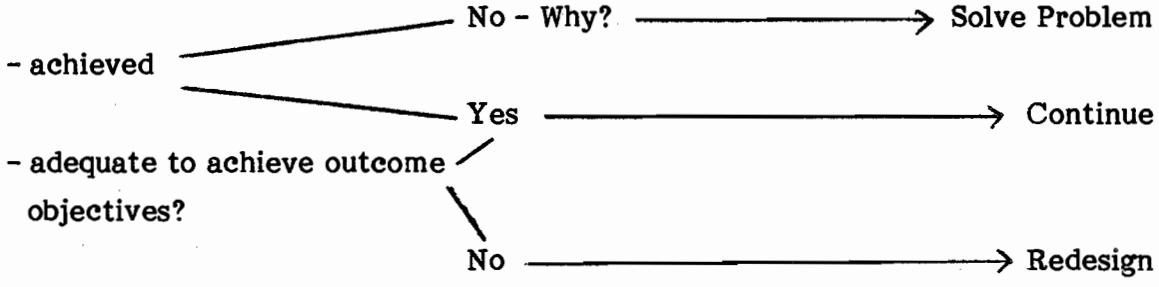
TARGETTING



DELIVERY



ACTIVITIES/GOALS



C. **Impact Evaluation**

The purpose of impact evaluation is to estimate the net effect of the food aid project on some outcome indicators related to the goals the project was intended to achieve. For vulnerable group feeding, the relevant indicators would measure the health and nutritional status of the targetted population. For food-for-work programs, the indicators might include both measures of the

humanitarian impact (e.g. contribution to incomes of those employed on the project) and of strictly economic impact (realized, as opposed to expected, rates of return in monetary terms)

The general method of impact evaluation is to try to answer the question "what would have happened in the population if the food aid project had not existed?" To answer this, the net effects of the project, those that can be causally attributed to it, must be sorted out from the gross effects, all the changes (in incomes, health, nutritional status, etc.) that are going on in the population independently of the project. Nearly always, the research design implies a contrast, comparing the project beneficiaries to a comparable population which did not receive the project, or to groups just entering the project, to national averages, or just to their own status before the project began. For example, the longitudinal information from a nutritional surveillance system such as that of Botswana described in section VII, could be used both for program area vs non-program area comparisons and before-after comparisons. Each of the possible designs has its drawbacks in terms of costs, feasibility, and the strength of inference that can be drawn from them. In ref. 5 (part of which is reprduced here as Annex IV) we discuss some of the practicable research designs and their advantages and limitations; other helpful sources are refs. 17-18. In this section, we will just add some considerations relevant to WFP planning.

Impact evaluations call for more resources than are typically available for WFP supported projects. The sort of activities outlined in the two preceeding sections (IV A and B) are quite enough to support most project-specific decisions. The information gathered to fill in the tables we have proposed is necessary, but not sufficient, to produce an impact evaluation, since evaluators would need at some stage to know what exactly happened in the project and who actually received the food in order to know where to look for an impact, and how much of an impact could reasonably be expected.

Selected impact evaluations should be supported by WFP, though. This will probably involve central funding, since the sort of research and information-gathering (especially on non-beneficiaries of the project) would greatly burden the reporting system for a single project without outside help.

Evaluation studies often suffer from misplaced perfectionism. No single study will be able to eliminate all possible alternate explanations for observed changes in incomes or nutritional status, leaving only the project activities as an explanation. But numerous studies in a range of settings, each with different strengths and weaknesses, can provide a good idea of the size of typical program impact, and the circumstances under which greater or lesser impacts could be expected. Beaton and Ghassemi have produced an extensive review of supplementary feeding programs (ref. 19) that could serve as a model for such reviews; a problem for that review was that the studies diverged greatly in purpose and method. If WFP were sponsoring its own series, they could focus more on measuring specifically the direct income and nutrition effects of projects.

The information about programme effectiveness is really useful for WFP decision-making only when combined with at least rough cost calculations. There is some level of resources that, if poured into an area in sufficient quantities for a long enough period, would eventually reach almost any project's objectives; the question is whether at realistic levels the project does enough good to outweigh whatever negative effects or costs are associated with it. For WFP-supported projects, the cost measures should include all resources contributed by WFP and the government, not just a value ascribed to the donated commodities. In such a cost-effectiveness exercise, it is more important to be comprehensive than to be exact.

For many WFP supported development projects, the WFP contribution is a fairly small portion of the overall total. If care is not taken to apportion to food aid only a proportional share of the overall project impact in a cost-effectiveness comparison, then the large multi-donor projects would artificially look better than smaller-scale development projects. This would seem a basic point, but several of the interim evaluations reviewed had long lists of achievements (thousands of acres of land reclaimed, miles of road rebuilt, etc.) that needed to be discounted by some factor to get an assessment of the (possible) effect of the WFP resources to make it comparable to reports on smaller, solely WFP projects.

Again, as we have stressed, WFP acceptance of a mission as a uniquely poverty-oriented donor agency would require a focus on impacts in terms of increased incomes and food consumption of the poor. We have urged that project design documents and interim evaluations have this focus; similarly these should be the major indicators for impact evaluation comparisons.

V. COUNTRY REVIEW PROCESS

Some of the information needs for project selection and design could be met most effectively not on an ad hoc, single-project basis but as part of a regular country review process. This would contribute both in selection and in evaluation. It would allow some consideration of alternate uses of project food aid, in the context of the country's overall development policy, helping to avoid the practice of responding to proposals for commodity distribution one-by-one.

Toward this end, WFP could participate with other donor agencies in regular donor group meetings at which portfolios of projects are considered (like the "Blue Book" for Indonesia) and funding responsibilities are allocated. The purpose is not to make WFP like the multi-purpose aid agencies and development banks, but to allow it to capitalize on its particular speciality in development projects - the use of commodities for payment in kind to benefit the poorest and most vulnerable groups. WFP, in this view, would not just be topping up funds for the economically viable projects that are attracting other donors and private banks anyway. Rather its representative would be looking at a whole list of proposals (laid out or rearranged to look like the list on p. 13), then deciding which projects could reach people most likely to increase their inadequate food consumption with the help of donated foodstuffs, or which projects would most likely use workers for whom donated food would be a significant part of their total wages. By both criteria the direct beneficiaries would be the poor, since it is they who spend the highest proportion of any increment to their incomes of food, and they for whom food (of the right kinds), is most nearly the same as cash. Some such projects will also attract other funders; some will not. In any case, WFP could be the poverty/employment oriented funding agency par excellence; such a use of project food aid would dovetail well with related uses for direct feeding and emergency relief. The

point is that identifying the projects for which food aid is logical and most likely to be appropriate will require advance work and consideration of a portfolio of project ideas; given limited WFP resources and the fact that other agencies are doing at least some of the groundwork already, perhaps WFP could usefully participate more actively in sectoral review and donor group meetings to achieve its goals.

A country review should improve WFP's ability to consider possible disincentive effects of food aid. At present, the interim evaluations contain a sort of check list of negative side effects, and the reports contain summary statements that each either does not exist or is negligible. This all-or-nothing approach diverts attention from the cost-benefit calculations that government and WFP decision makers ought to be making. As with any development programme, the question is not so much whether anyone conceivably loses out - someone nearly always does, if only indirectly - but whether the gains from the project sufficiently outweigh the losses, and who gains and who loses. The checklist approach does not add to the government/WFP ability to rank alternate uses of project food aid, and it probably doesn't satisfy persistent critics either.

The questions about disincentive and dependency effects are most logically considered as part of a country's overall food policy, not as elements of the design of a single project. The same considerations arise for both project and programme food aid, so they could usefully be treated together in a country review process. A logical way to do this might be analogous to the World Bank sector studies for a Country Economic Memorandum. WFP evaluation staff, working with the country representative and consulting outside experts as needed, could address for each of at least the major recipients of food aid the important questions about its positive and negative effects, direct and indirect. Where project food aid is an insignificant portion of the total food supply, there is no need to treat at great length its possible effects on local production. In many countries, such a review would reveal that price policies and foreign exchange rates are discouraging local food production. If, as international agencies are more frequently recommending, a country decides to raise food prices to encourage greater production, a good use of food aid might

well be to mitigate the harmful effects on the poorest consumers by targetted distribution or subsidizing the budgetary costs of a dual price policy for some basic good. The point is that the mere fact that local food production is not at a maximum is not by itself a sufficient argument against the use of food aid. Other, much more important policies are often responsible, and project food aid might even be a way to make change possible with minimal disruption for the poor. These possibilities are not currently discussed or explored in interim project evaluations, and we would argue that these evaluations are not really the appropriate place to do this sort of policy analysis. Regular country studies, either in cooperation with other institutions or solely as a WFP initiative, would be appropriate.

The other purpose of a country review process would be to provide the basis for selection and design of projects, by identifying needs and opportunities for project food aid; defining those groups in the population that most need to benefit from donated food, and giving an indication of the sorts of projects most likely to benefit these people. Important information will be to produce a socio-economic classification of the population in terms of nutrition. This classification can usually be assembled from existing data, and should essentially give priority rankings, for different groups in terms relevant to targetting. Usually this means identifying who is in need (e.g. malnourished) and how many by at least administrative area, and often by other factors such as accessability and employment status, in addition to the more common definitions by biological status (age, sex, pregnancy, etc). Such a classification can guide the design of a whole range of projects, and is much more efficiently produced on a national basis often at fairly long intervals, rather than project-by-project.

VI. EXAMPLES OF INFORMATION CURRENTLY AVAILABLE

A. In WFP plans of operations and interim evaluation reports

Four types of information have been recommended: on targetting and coverage; on level of delivery or expenditure per head; on activities and goals; and on outcome objectives. The availability of this information was reviewed from the documents available to us. These gave information from Plans of

Operations from 18 projects, and from Interim Evaluations on these plus a further 8 projects, totaling 26. The results are summarized in Table 4. Most of the information that is being both planned and actually collected and evaluated concerns the achievement of goals - for example number of rations distributed, or kilometers of roads built. The other three types of data are rarely planned for, and hence seldom collected. These findings are implied in the earlier discussion, and are presented here only to emphasize that a fairly important shift in the type of data collected, analyzed, and presented is called for in our recommendations. We suggest that emphasis should be given to the first two types of data, that is on targetting and coverage, and on expenditure per head. Outcome data is more difficult to collect, but where available (see example in section VII) should be used. The results however do emphasize that without this investment in additional data, it is difficult to see how the processes of evaluation and indeed eventually in planning, can be improved.

B. **In an evaluation study of a supplementary feeding program for children (India - 2206)**

The report of an evaluation sponsored by WFP on India - 2206 (ref. 11) was carefully reviewed. Before giving a summary of the results of this review, we should stress that the main recommendation is that in the future the required data outputs (again along the lines of the four types of data indicated above) be specified in advance. In fact in this case, much of the data required for this was in fact collected, although not presented. Simple calculation from the figures in the report alone can give much of the required information on targetting and coverage, and on expenditure per head. Too much emphasis on trying to assess impact was evident, and this merits some elaboration.

The evaluation was to be accomplished in a six-month period, using data collected only at that time, within a limited budget, covering much of the programme area. Usually, it will be impossible to gain information on impact in this way. Discussion with those involved indicates that there was pressure felt to give some answers on impact, hence for a design where "control" groups were selected, although the validity of the comparisons then made is not established in the report. A better procedure would be (a) to limit the

objectives of the evaluation to checking that "the right food was getting to the right people at the right time" i.e. process evaluation, (b) restricting any assessment of impact - if it is considered essential - to a small sub-sample, not usually therefore generalizable to the whole program, but possibly nonetheless informative, (c) stating clearly the limitations of any comparisons and hence conclusions on impact if they are to be made. However, this emphasis on outcome seems to have obscured some very valuable results, which if given more emphasis could have been calculated and found useful. Without any assessment of impact, certain important questions could have been answered from the report. Questions such as: What are the objectives of the program, and are they realistic? What is the coverage? Expenditure per head, cost per case of malnutrition prevented? Degree of targetting to the most needy? and so on can be answered by the data available, with relatively minor additional calculations and reformatting. Table 5, column C shows the calculated coverage by state of the child population; it ranges from below 1 to 5 percent of children. Evidently only a small proportion of the children at risk or overall were intended to be reached by the program. This has important implications for the programme objectives, and for the method of targetting the programme. If, as in the case of India, estimates are available of the prevalence of malnutrition in the population as a whole, the coverage figures can be transformed into estimates of the proportion of malnourished children covered. It appears that the prevalence of malnutrition both in the recipients of the program, and in the comparison groups chosen in the evaluation (ref. 11), is substantially less than in the population as a whole (according to the National Nutrition Monitoring Board - ref. 20). Specifically, in these seven states the prevalence of second and third degree malnutrition is estimated as around 22% in those receiving the program; 30% in the evaluation comparison group (ref. 20 table 19); and 31% in the child population at large in these states. Some of the difference between the evaluation control group and the NNMB survey sample could be explained by the fact that the former were measured in 1982 and the latter (the most recent figures available to us) are from 1979, but the improvement in nutritional status in 3 years was unlikely to account for all of the difference. Hence the coverage of malnourished by the program is of the

order of 2% overall. The prevalence of malnutrition in those receiving the programme (table 5, col. G) can be compared with an estimated prevalence for the population overall (table 5, col. H), conveniently expressed as a ratio (table 5, col. I): since this ratio is generally less than 1.0, it appears likely that the programme is somewhat targetted away from the malnourished. (It should be noted that such cross-sectional analyses cannot distinguish whether or not this is due to a positive effect of the programme - but nonetheless retargetting should be considered). Whether the figures in table 5 are exact is not the point, but that indicators such as these can readily be derived and do lead to important conclusions on the management of the program. Finally, some independent evidence that the matched control groups are comparable - such as literacy rates - is essential. Expenditure per head can also readily be calculated from the available figures. It is of the order of $20 per recipient per year, probably well within the range where an important effect on nutritional status could be expected - in fact it may be on the high side for this.

The results quoted above alone raise important issues for the program. First, there is evident scope for retargetting by area: both the "control" group and the population measured in an independent survey have a higher prevalence than the experimental group and hence there is potential benefit for moving the program towards those with a higher prevalence of malnutrition. Second, screening individual children on the basis of nutritional status into the program is worth considering: if only 2% of the malnourished are now being covered, some way of focussing resources is clearly essential, if the objective is to have an impact on the prevalence of malnutrition in the population as a whole. Third, the cost per head is high: $20 per head per year should have a very marked effect on nutrition (pilot programs have indicated a marked effect on nutritional status with an expenditure - although not primarily on food - of around $5 per head per year - see ref. 12). Use of some of the resources for other activities, or a lower expenditure per head giving a higher coverage integrated with other services, as well as better targetting, could well produce a larger effect.

As mentioned, these results were not presented as such in the WFP - commissioned report (ref. 11), although they could be readily derived from the

results that were presented. The main conclusion is not to criticize the valuable report which was put forward, but to emphasize that the results of interest to WFP should be specified by WFP guidelines in advance, before the evaluations are undertaken, in sufficient detail (possibly along the lines suggested in this report) to produce conclusions which are useful for program design and management.

VII. EXAMPLE OF A MONITORING AND EVALUATION SYSTEM FOR FOOD AID: DESCRIPTION OF SYSTEM AND ITS POSSIBLE DEVELOPMENT IN BOTSWANA

This section discusses changes in the WFP monitoring and evaluation system, along the lines of the foregoing discussion, as they might be implemented in a country that has been a major recipient of project food aid. Botswana would be a good test case for the development of new methods for monitoring and evaluation for several reasons.

- Much of the system is in place anyway, and recently the necessary steps have been started to make the system even more effective. It can be anticipated that given the necessary support an excellent model of an evaluation system will be functioning and can serve as a basis for replication elsewhere.
- The levels of project food aid have been large enough so that both the beneficial effects on nutrition, if any, and the negative effects on domestic production and food policy and other hidden costs, if any, would be discernible in Botswana. Even before the current drought, in the mid-1970's, Botswana received more aid per capita from WFP than did any other country (ref. 13). Besides continuing projects (primarily WFP 324, a supplementary feeding program), Botswana has recently been a major recipient of WFP emergency funds - only the Gambia and Mauritania in Africa received more per capita in 1982.
- The government of Botswana has a lively interest in issues surrounding food aid and how it should fit into a national food policy. During the current drought and earlier ones, the government has chosen not to institute food-for-work projects, relying instead on

Labor-Based Relief Projects with cash wages, for fear of creating over-dependence on government deliveries of food. Though the over-riding concern this year is with efforts to mitigate the effects of the drought, there is also lively concern among policy makers about the long-term consequence of the various food aid measures adopted. This was shown in recent discussions with officials of several ministries (ref. 14, 15). The concern, like project food aid itself, will probably still be there after this drought, so there are good reasons to expect interest and cooperation from the government in any attempts that WFP might make for improving the flow of information to monitor and evaluate projects.

A. Description of Existing Information Systems and Possible Improvements

The information currently available is as follows. The Food Resources Department in the Ministry of Local Government and Lands (FRD) collects data on a monthly basis from each of seventeen sub-depots on beginning stocks, amounts of commodities shipped to the feeding points (clinics, schools, and health posts), and closing stocks. The depot managers who are doing the reporting are not themselves employees of the FRD; rather, they are seconded from the Supplies Department of the Ministry of Finance.

The depot managers also report to the FRD data they are supposed to collect from the feeding points on the number of beneficiaries in various categories (school children, pre-school-aged children, pregnant women, "drought destitutes") that have received nations for the month. It does not appear that this information is used for any purpose by the Government of Botswana other than to fulfill a contractual obligation to WFP by adding up the numbers every quarter and sending them to the WFP country representative on a prescribed reporting form. One official within FRD handles the whole procedure. For three years, one of the seventeen depots has not reported either set of figures, and no one has attempted to fill the gap. This suggests that the numbers are not meeting a need perceived by decision makers.

The statistics on deliveries from depots could be made much more useful for purposes of monitoring the delivery system if they were combined with

information on distribution from feeding points. For purposes of evaluation, one would also want information on the nutritional situation both in communities where food aid is being regularly delivered and in communities where food aid is intermittent or non-existent. In Botswana, there is the potential for such a useful combination of data reported from different sources, since there is a well established nutritional surveillance system operated by the Ministry of Health in clinics and health posts throughout the country.

Several ministries have recently been exploring the idea of collating and analyzing data pertaining to food and human nutrition from their separate reporting systems, among other reasons, to improve drought relief planning. This idea could usefully be applied to the monitoring and evaluation of food and projects, during and after the drought, since nutritional surveillance will continue.

The surveillance system is based on regular reporting from the clinics, health posts, and mobile health teams of the weights and ages of children attending the clinics. Changes in children's weights (compared to age-group standards) provide a fairly reliable and inexpensive indicator of trends in the nutritional status of the community, since small children are the most vulnerable to any deterioration in conditions of life. The reverse is also true, that one would expect to see any real improvements in the nutritional status of the community reflected in increasing weight-for-age scores of pre-school-aged children.

As originally devised, the surveillance system proved somewhat cumbersome. As a result of recent evaluations, the system will be streamlined and the burden of measurement and reporting eased. The clinics will begin sending to their regional health officials and to the Central Statistics Office forms on which children's weights and ages are recorded, along with simple tally marks in certain columns of the reporting sheet: one to indicate whether the child is classified as "at-risk", one to indicate whether the child had gained weight in the last month, etc. One of these columns could be used to indicate whether the child received supplementary feeding, since the clinics now serve as the distribution points. Alternately, the clinic staff could be asked to report on the forms whether food and had arrived in the community and had been

distributed as planned within the last month. This would provide a check on food aid receipts, to compare to the information from the depots on how much was sent out from warehouses. This could be used to monitor the delivery system past the first link in the chain, which is the only level of information currently available. Since the nutritional surveillance systems already provides information on the location and severity of malnutrition in the country, the combination of data would allow evaluators to tell whether food aid is reaching the people who need it, in time, and in amounts, to forestall prolonged deterioration in their nutritional status. This would be a considerable improvement over what is now available.

The use of such an information system for monitoring might be fairly simple. If depot managers report that commodities are leaving the warehouses, but some drought-affected areas supposed to be served from those warehouses report no deliveries or distribution in a given month, then the breakdown in the distribution system can be located and, one hopes, corrected. Besides providing some information below the level of the depot, the new system would provide an independent check on the depot managers' reports. Presumably the Food Resources Department would be responsible for collecting and analyzing the data. Interpretive reports - saying what types of problems were encountered with the food delivery system, and what could be done to correct them - could be produced eventually, rather than just onward transmission of undigested numbers to the WFP.

B. Uses of Information for Evaluation Studies

For purposes of evaluation, the data would presumably be analyzed by the Central Statistics Office (which produces the nutritional surveillance reports) or the FRD. The basic question whether food aid was reaching the right people could be answered with a table of the following form:

		Child's weight-for-age:	
		Below 80% of standard	Above 80% of standard
Child (or family) receiving food aid?	Yes	a	b
	No	c	d

(This corresponds to the generic table in Section IV.A)

The timeliness of deliveries could be evaluated by comparing reports from drought areas over several months to see if there are places where children's nutritional status is deteriorating for months while food aid deliveries are delayed or infrequent.

A thorough impact evaluation would not be possible with these data alone, but one could use them for an "adequacy" evaluation (see Annex II for definitions). Over time, one would expect to see improvement in nutritional status in areas where food aid projects were continuing, or at least not to see the seasonal deteriorations that otherwise would be expected. The magnitude of improvement could be estimated roughly by comparison with the pre-project situation in the area or with similar areas where projects either don't exist or are just beginning. Besides the continuous, routine system, special studies, approximating more closely the ideal impact evaluation, might be useful. These could use more careful measurements of nutritional status in project and comparison areas, perhaps by household surveys, to reach clinic non-attenders and to gather information an socio-economic variables related to nutritional status might be confounded with program effects. (See Annex IV for a discussion of methods).

What would be the incentive to the Ministry of Health and the Food Resources Department to cooperate in such a monitoring and evaluation scheme? If like the present system it served only an audit function for WFP,

the answer, of course, is none. But both agencies have shown great interest in policy regarding food aid. If the information from a monitoring and evaluation system could be used to modify the design of projects, improve implementation, and contribute to analyses in other ways, there might be a very real incentive to participate. The Ministry of Health, for example, is concerned about the opportunity costs of the use of its personnel and resources in food aid distribution, possibly to the detriment of regular maternal and child health programs. Any information that would help in better design of future relief efforts would be of benefit to the MOH. It might be shown that aid could be delivered more efficiently by targetting it to the most seriously affected communities, which would be possible with a regular monitoring and evaluation system.

This sort of selectivity would allow food aid to Botswana to be used where its benefits are most likely to outweigh its costs, including the potential negative side-effects. Without information on the adequacy of the effort and the need for it in different communities, feeding programs almost have to blanket the country in order to be sure that the needy are reached.

There are some concerns, both of policy makers in Botswana and of WFP, that would require more focussed study than the routine system alone would allow. Useful work could be done by anthropologists and other social scientists in studies of the effects of food aid on the micro-economy of different areas in Botswana, to estimate its possible effects on traditional methods of coping with drought, on migration patterns, or on attitudes to the health care system (ref. 15).

Another concern is with costs, both direct and indirect, associated with the use of different methods and sites for delivering food in feeding programs. Data from the monitoring and evaluation system would be necessary, even if not sufficient, for such studies, simply to establish what project activities actually existed in different places, for how long. This would enable the focussed studies to be more than collections of anecdotes, since systematic comparisons could be made.

The staff and skills for the routine evaluation system and for special evaluation studies are probably not available in the agencies of the Government of Botswana that would most benefit from them; WFP could contribute some of

its evaluation resources to provide training and technical cooperation. The payoff would be not only to the WFP, in terms of case studies of project food aid costs and effectiveness, but also to the government agencies, since the methods and even the reporting system could be useful in other types of project evaluation.

REFERENCES

1. Hogan, Edward B (1983) 'Scope of Work for a Study of Potential Improvements in the WFP Programme System' (consultancy report).

2. Katona-Apte, Judit and Maxwell, Simon (1983) WFP Project Design Review (Prepared for WFP Rome).

3. Mason, J.B.; Habicht, J-P; Tabatabai H. and Valverde, V. (1983/4) 'Nutritional Surveillance'. WHO (In Press) Chapter 5: Nutritional Surveillance for Programme Management and Evaluation.

4. Mason, J.B.; Habicht, J-P and Tabatabai H (1982) 'Principles for Evaluation of On-going Programs'.

5. Mason, J.B. and Haaga, J.G. (1983) 'Note on a Framework for Monitoring and Evaluation of UNICEF/WHO Nutrition Improvement Programs.'

6. Mason, J.B. (1983) 'Provisional Guidelines for Designing Evaluation for JNSP Country Programmes' (Draft).

7. Gittinger, J. Price (2nd edition 1982) 'Economic Analysis of Agricultural Projects.' The Economic Development Institute, International Bank for Reconstruction and Development (The Johns Hopkins University Press).

8. Stokey, Edith and Zeckhauser, Richard (1978) 'A Primer for Policy Analysis.' Chapter 9: Project Evaluation: Benefit-Cost Analysis. (W.W. Norton & Co., New York).

9.a. FAO, Rome (1982) 'Integrating Nutrition into Agricultural and Rural Development Projects'. Nutrition in Agriculture No: 1. A Manual. Nutrition Planning, Assessment and Evaluation Service. Food Policy and Nutrition Division.

b. Pinstrup-Andersen, Per (1981) 'Nutritional Consequences of Agricultural Projects: Conceptual Relationships and Assessment Approaches'. World Bank Staff Working Paper No: 456.

10. U.S. Agency for International Development (1976). 'Evaluation Handbook' (2nd ed.) Washington, D.C.

11. Swaminathan, M.C. et al. (1983) 'World Food Programme Assisted Special Nutrition Programme in India: an Evaluation Study.' National Institute of Nutrition, Indian Council of Medical Research, Hyderabad, India.

12. Gwatkin, Davidson R; Wilcox, Janet R. and Wray, Joe D. (1980) 'Can Health and Nutrition Interventions made a Difference?' Monograph No: 13. Overseas Development Council, Washington D.C. 20036.

13. Stevens, Christopher (1978) 'Food Aid and Nutrition: The Case of Botswana.' Food Policy, Vol. 3, No: 1 February 1978 pp. 18-28.

14. Haaga, J.G.; Quinn, V.; Mason J.B. and Williams, K. (1983) 'Recommendations for Development of Nutritional Surveillance in Botswana' Report on visit to Botswana, 5-15 July 1983. A publication of the Cornell Nutritional Surveillance Program, Division of Nutritional Sciences, Savage Hall, Cornell University, Ithaca, NY 14853.

15. Tabor, Steven (1983) 'Drought Relief and Information Management: Coping Intelligently with Disaster'. A publication of the Cornell Nutritional Surveillance Program, Division of Nutritional Sciences, Savage Hall, Cornell University, Ithaca, NY 14853.

16. WHO (1981) 'Health Programme Evaluation'. WHO Health for All Series No: 6.

17. Casley, D.J. and Lury, D.A. (1981) 'A Handbook on Monitoring and Evaluation of Agriculture and Rural Development Projects'. World Bank November 1981.

18. Freeman, H.E.; Robbi, P.H. and Wright, S.R. (1979) 'Evaluating Social Projects in Developing Countries' Development Centre Studies. Paris OECD.

19. Beaton, G.H. and Ghassemi, H (1982); 'Supplementary Feeding Programmes for Young Children in Developing Countries'. Am. Journal of Clinical Nutrition. 35: (4 supplements) pp. 863-916. April 1982.

20. National Nutrition Monitoring Bureau (1980) Report for the year 1979. National Institute of Nutrition Indian Council of Medical Research, Hyderbad, India.

TABLE 1

EXAMPLES OF DECISIONS FOR FOOD AID PROJECTS

(a) **Project Selection and Design**

- Accept/modify project proposal in principle

If accepted:

- Accept/modify proposed project design

Design requires decision on:

- outcomes objectives (eg. economic and nutritional)
- targetting (nos. beneficiaries; Who are they?)
- level of input (eg. food per head per time)
- types of activities and goals

(b) **Implementation and Management**

- Is overall delivery schedule maintained?
- Are targetting, level of input, activities/goals on track?

If yes, continue. If no, decide on corrective action

(c) **Interim Evaluation and Redesign**

- Has project been implemented as planned

(have targetting, level of input, activities/goals been achieved?)

- Were outcomes objectives adequately reached?

Hence decide on redesign

(d) **Impact Evaluation**

- Were there changes in outcome indicators (effects) due to project?
- What were effects/unit cost? Were these costs acceptable?

Hence decide future project designs

TABLE 2

EXAMPLE OF ESTIMATED OBJECTIVES FOR ALTERNATIVE PROJECT DESIGNS

INPUT $1,000,000

	DESIGN		
	A	B	C
1. Economic Returns (internal rate of return)	12%	8%	8%
2. Targetting			
No. people	10,000	20,000	10,000
Relative need	Moderate	High	High
3. Level of Delivery			
Immediate net effect	$100/hd/yr	$50/hd/yr	$100/hd/yr
% Estimated income	10%	10%	20%
4. Long-term net effect	$10/hd/yr	$5/hd/yr	$10/hd/yr
% income	1%	1%	2%

Data: Household economic survey for 1,3,4

Socio-economic classification for 2.

TABLE 3

Summary of data needs for design and evaluation

SUBJECT (participants only)	DESIGN	ROUTINE EVAL.	INTERIM EVAL.	IMPACT EVAL.	INDICATORS (examples)
1. OUTCOME					
Economic	X	-	X	X	Income/Expenditure
Nutritional	X	-	X	X	Nutritional status
2. TARGETTING	X	X	X	X	Focussing, coverage etc.
3. DELIVERY	X	X	X	X	$/hd
4. ACTIVITIES/GOALS	X	X	X	X	Process data
(Non participants)					
5. OUTCOME	-	-	-	X	As 1

TABLE 4

Information Contained in WFP Plans of Operations and Interim Evaluations

	Information							
	Targetting/Coverage		Expenditure		Goals		Outcome	
	Planned	Evaluated	Planned	Evaluated	Planned	Evaluated	Planned	Evaluated
Given	1	1	2	0	13	17	0	0
Partly Given	5	1	0	0	0	0	0	1
Not Given	12	24	16	26	5	9	18	25
TOTAL	18	26	18	26	18	26	18	26

TABLE 5

Selected results from an evaluation of a WFP supported supplementary feeding program in India (India 2206).

State	A No. Children Covered by Program (thousands)	B Total Population of Children 0-6 Years (thousands)	C Percent of Population Covered (A - B)	D Number of Program Children Malnourished* (thousands)	E Number of Non-Program Children Malnourished* (calculated from reference 20)	F Coverage-Percentage of Malnourished Children Covered by Program* D - (D + E)	G Focussing Percentage of Children in Program Who are Malnourished* (D ÷ A)	H Percentage of Children in population malnourished* (reference 20)	I Focussing Population Prevalence (G ÷ H)
KERALA	202	4,320	4.7	31	1,367	2.2	15.3	33.2	0.46
MADHYA PRADESH	123	8,860	1.4	43	4,998	0.9	35.0	57.2	0.61
MAHARASHTRA	122	10,660	1.1	38	5,301	0.7	31.1	50.3	0.62
ORISSA	225	4,470	5.0	53	2,063	2.5	23.5	48.6	0.48
RAJASTHAN	69	5,800	1.2	12	(N.A.)	—	17.4	—	—
UTTAR PRADESH	36	18,850	0.2	8	5,117	0.2	22.2	27.2	0.79
WEST BENGAL	322	9,260	3.5	130	3,504	3.6	40.4	39.2	1.03
TOTAL OR MEAN	1,427	62,220	1.8%	315	22,350	1.4	22.1%	30.8	0.72

Sources: Ref. 11, Tables 2, 19; Ref. 20, Table 26
*"Malnourished" defined as below 75% standard weight for age.

FIGURE 1

PROJECT LIFE SPAN

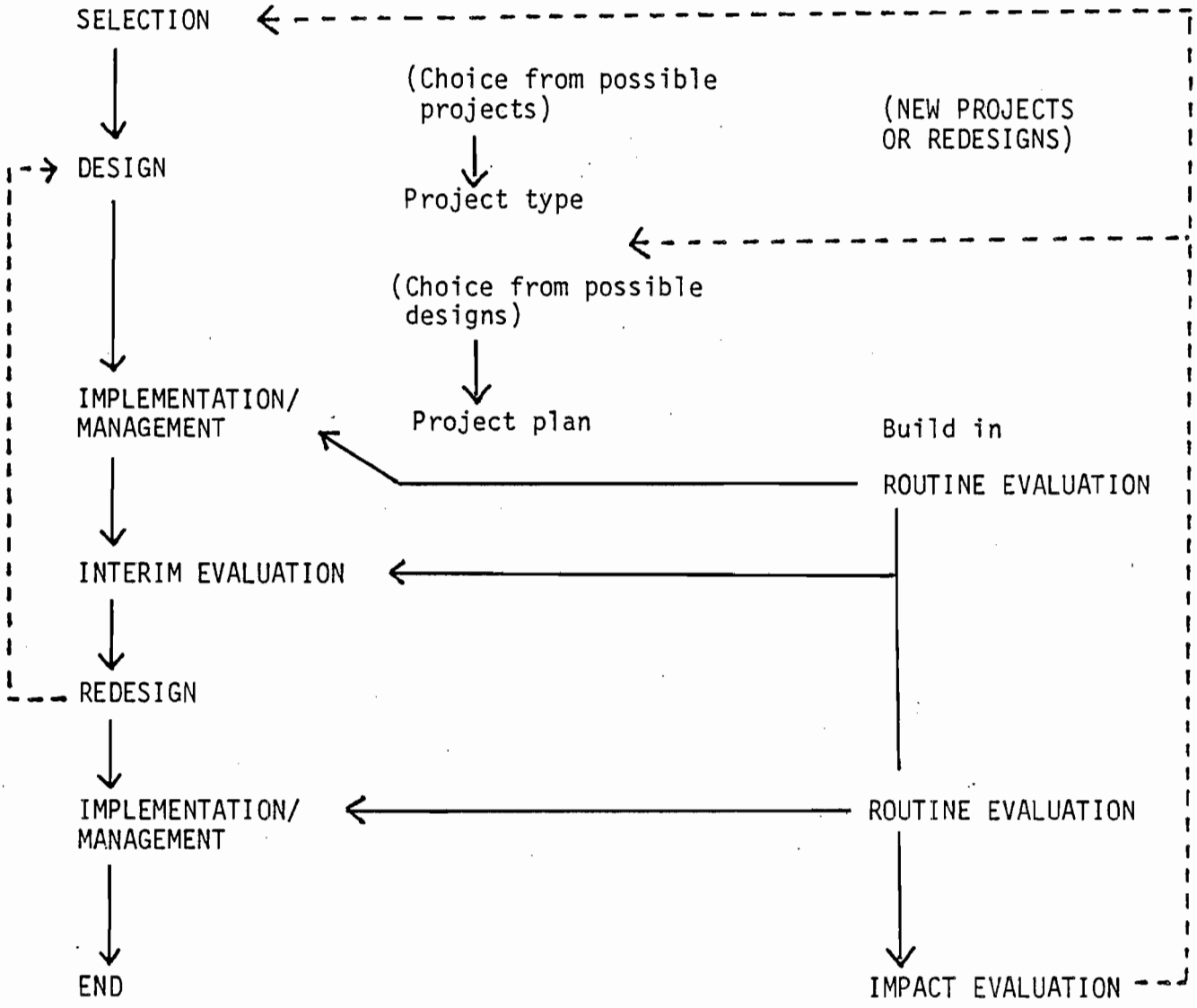
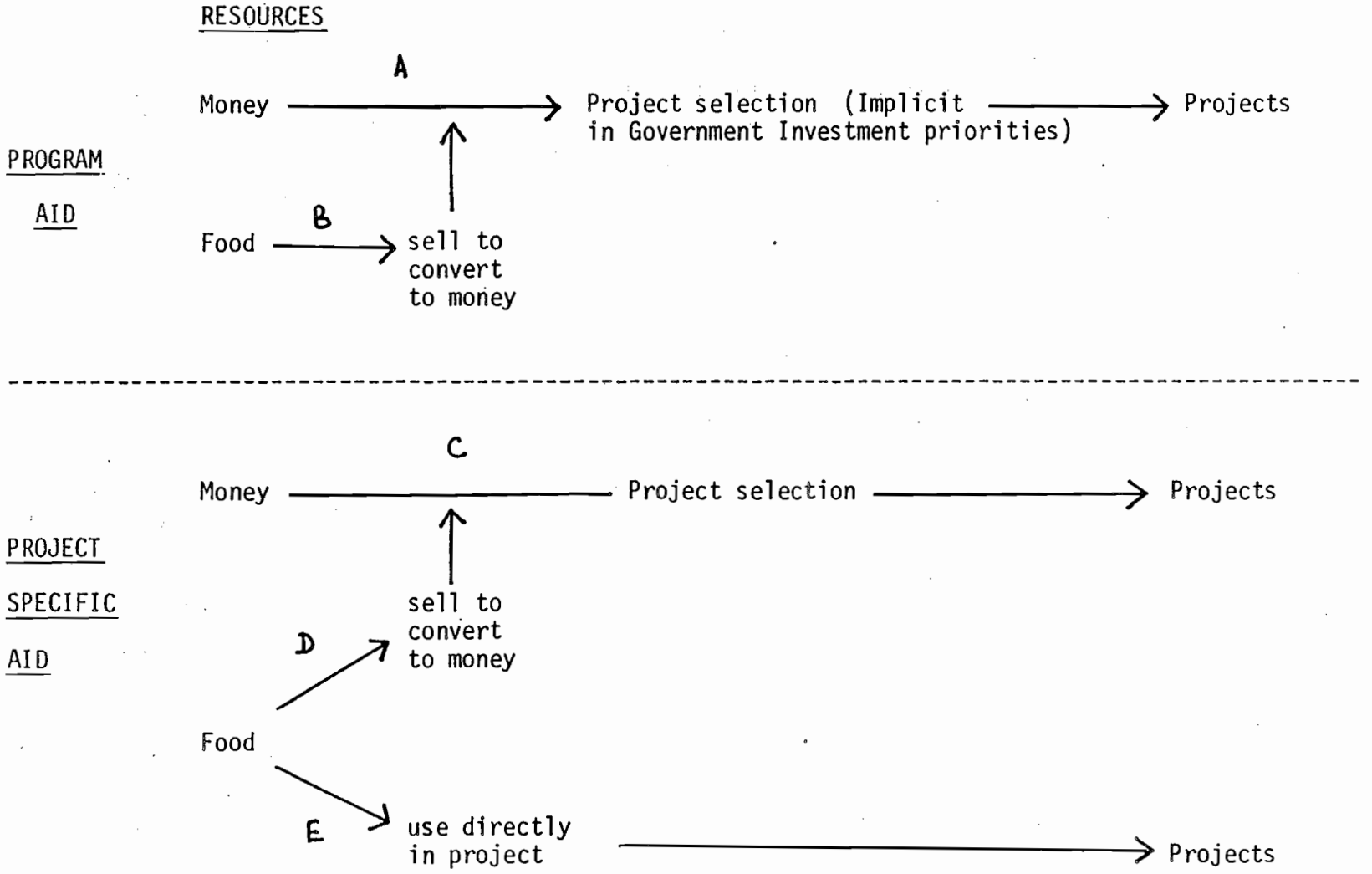


FIGURE 2

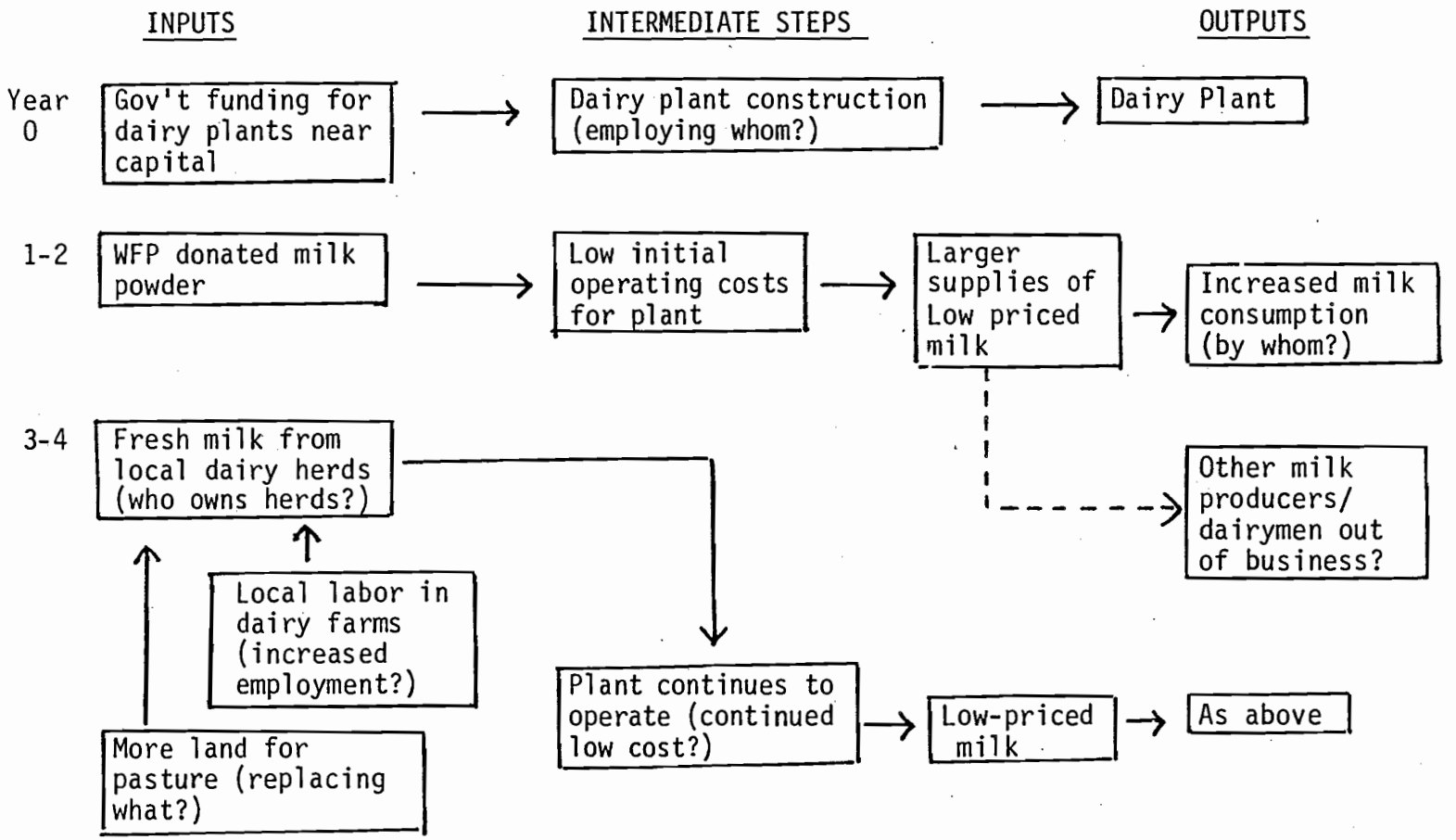
Uses of Resources to Reach Objectives



WFP Project Food Aid follows path E

FIGURE 3

BOX AND ARROW DIAGRAM OF A DAIRY DEVELOPMENT SCHEME



ANNEX I

TERMS OF REFERENCE

1. In the context of a more general review of WFP programming methodology and operating modes/procedures, which is currently going on or under preparation in a number of sectors (in-depth evaluation of selected projects, review of programming cycle, definition of reporting requirements for WIS-5, introduction of the logical framework methodology, holding or preparation of various seminars), the three consultants will **provide recommendations to improve**
   i) **definition of objectives and targets** of WFP-assisted projects, and in particular of the role and objectives of the WFP input to these
   ii) the analysis of the **process** (criteria for selection, food delivery and other support systems etc.) for achieving these objectives and targets in food aid under different conditions or of alternative approaches;
   iii) **the cost-effectiveness, relevance and appropriateness** of project food aid under different conditions or of alternative approaches;
   iv) essential indicators for project monitoring and evaluation;
2. In order to meet these obligations, the consultants will:
   i) examine the current WFP project design and documentation methodology;
   ii) analyse the documentation of approximately 20 projects (to be selected by WFP);
   iii) undertake consultations with WFP staff at Headquarters; and
   iv) prepare a final report for the delivery of the end-product as defined in para. 1 above, which may provide guidelines as to ways in which WFP might improve overall management procedures.

ANNEX II

(From Mason, J.B.: 1983 - Ref. 6)

EVALUATION TERMINOLOGY

A wealth of different terms is in use for planning and for evaluation, and it is impossible to choose one convention that will be familiar to everyone. Worse, there is such a confusion in the use of the different terms that a "constraint that is often encountered is a certain built-in resistance in principle to accepting evaluation and its results as a valid management tool" (ref. 16, para. 9).

The **"evaluation"** procedures discussed here are primarily (but not only) an integral part of programme menagement. Their purpose is the same as that for health programmes (ref. 16, para 6): "Evaluation is a systematic way of learning from experience and using the lessons learned to improve current activities and promote better planning... for future action". The relation to other terms in current use is geven briefly below.

Evaluation (here) = Evaluation (ref. 16) = Monitoring & Evaluation (ref. 17) = Operational Programme Evaluation (other literature).

We distinguish here between **"routine evaluation"**, aimed at ensuring satisfactory programme management, and **"impact evaluation"** aimed at assessing the net effect of a programme. Before explaining this, other concepts need to be clarified.

(a) There is a difference between checking on an activity ("monitoring" it). e.g. number of children fed; and examining its effect or outcome, e.g. nutritional status. The information relating to activities is known here as **"process"**. The information relating to effects or outcome is known as **"outcome"**. In other terminologies the equivalents are as follows:

Process = Review of Progress, Efficiency, Effectiveness (ref. 16, paras 38, 56-70). = Inputs and Outputs, or Monitoring (ref. 17).

Outcome = Output (ref. 16, para. 67) = Effect and Impact; Assessment of Results (ref. 17).

55

(b) There is another essential difference between examining changes in outcome (e.g. improvement in nutritional status) for programme participant, and in considering whether this is **due to the programme.** One differnece is obviously the changes that anyway would have occurred with or without the programme; one difficulty in assessing this may be due to the way in which programme participants are selected, and so on. The overall change, **not** allowing for changes that might anyway have occurred, is known as **"gross outcome".** If attempts are made to determine how much change is due to the programme, this is reffered to as **net outcome".** The expression net outcome is synonomous with **"impact".** The relationship may be illustrated as follows:

Net Outcome = Gross Outcome + Changes not due to Programme etc.
(or impact)

This concept is discussed further in (ref. 3 and ref. 18). It is suggested that a portion of the effort put into evaluation be devoted to impact evaluation, since this is the only way that some estimates of the actual effectiveness, and possibly cost/effectiveness, of the programme can be asessed.

"Routine evaluation" refers to using information on process and gross outcome to reach conclusions and decisions useful to programme management. Routine evaluation is similar in concept to a management information system. This procedure checks if programme implementation is adequate, by comparison with operational objectives, and if gross changes in (eg.) health and nutrition are adequate, in comparison with impact objectives. This is equivalent to "routine **adequacy** evaluation" in (ref. 5). **"Impact evaluation"** uses information on process and net outcome: it is equivalent to "assessment of impact" in (ref. 16, para. 71).

It is common to distinguish **"impact objectives"** (e.g. reduce infant mortality) and **"operational objectives"** (e.g. deliver X amount of food). These objectives relate directly to outcome indicators, to assess impact; and to process indicators, to assess progress towards meeting operational objectives. Quantification of these objectives in the planning stage gives the criteria against which progress can be assessed, by evaluation. This specification should be directly related, again initially in the planning stage, to the management or

policy decisions that are to be made if these objectives are or are not met (in management, usually the latter). This is illustrated in table A-1. The evaluation procedure is thus a way of making these objectives meaningful, and is yet another reason that planning a programme whould include planning the in-built evaluation. The level of detail reached in planning the programme, whether prior to implementation or, often better, as the programme is implemented, determines the level of detail feasible in designing the evaluation. Ideally, these should all be part of the same process.

Impact objectives need to be distinguished into "net outcome objectives" and "gross outcome objectives". Operationally, the outcome objectives needed for routine evaluation (e.g. reduce pre-school malnutrition by so many cases per 100 per year) are "gross outcome"; but the impact objectives are "net outcome". It may be decided in the programme planning that gross and net outcome objectives should be set as the same reduction in malnutrition, but the distinction is crucial for subsequent evaluation. This point becomes clearer when translating impact objectives into outcome indicators as shown in table A-1.

TABLE A-1

RELATIONSHIPS BETWEEN OBJECTIVES AND EVALUATION INDICATORS

Operational Objectives	Process Indicators to Assess Progress to Objective
(a) Deliver X amount food/month	(a) Delivery of food
(b) Provide supplementary food all families in area	(b) Number of household/ villages receiving food in Area B.

Impact Objectives	Outcome Indicators to Assess Progress
(a) Improve child growth so that proportion of children less than 80% wt/age is reduced by 3 cases per 100 per year.	(a) Prevalence of children less than 80% wt/age
(b) Reduce infant mortality from 200 per 1000 live births to 150 per 1000 live births after five years.	(b) No. of infants dying in relation to number of births, per year.

ANNEX III

STEPS IN CROSS-SECTIONAL EVALUATION

(From Mason, J.B., Habicht, J-P and Tabatabai, H: 1982 - Ref. 4)

STAGE 1: PRELIMINARY TASKS

1. **Decide:**

   - Who is doing the evaluation?
     (e.g. outsiders, management, funders)

   - For whom?
     (management, administrators and funders, research bodies)

   - For what purpose?
     (see Table I.2) e.g.:

     - continuation or modification of delivery of the program?

     - replicate the program?

     - estimate the net effects of the program?

2. **Reach consensus on the objectives of the program.**

   - Make underlying objectives explicit and resolve contradictions.

3. **Scouting**

   - Get initial impression

   - Make explicit how the program is expected to achieve its effects

4. **Plan the evaluation**

STAGE 2: EVALUATING THE PLAN OF THE PROGRAM

1. **Examine overall objectives**
   - What are they?
   - Are objectives specified in terms of:
     - quantities of inputs?
     - target groups (numbers, characteristics)?
     - permissible deviations in targetting and delivery?
   - Are inputs compatible with expected outcomes (was there a sound basis for the program design)?
   - If above cannot be specified - stop

2. **Evaluate implementation objectives**
   - Budget, work-plan
   - Planned expenditure per head
   - Compare with expected effect per unit cost

3. **Evaluate targetting objectives**
   - Who, how many?
   - Calculate planned focusing and coverage (see Table II.8)

4. **Evaluate outcome objectives**
   - Apply to who (recipients, targetted, total population)?
   - Do objectives match inputs?
   - Is planned effect/unit cost reasonable?
   - Define adequate effect.

5. **If adequate effect not expected from evaluation of plan**
   - stop

STAGE 3: EVALUATING IMPLEMENTATION

1. **Does program reach intended target group?**

   Assess targetting as:

   - Numbers targetted who are recipients, hence calculate indicators of delivery and leakage (Table I.8)

   - If outcome data available, (i.e. number needy who are recipients), calculate indicators of actual focusing and coverage (Table I.8), compare with targetting objectives.

2. **Assess level of delivery as:**

   - Expenditure or other measure per caput of recipients

   - Compare with implementation objectives.

3. **Do deviations from objectives affect expected outcome** (for target group or population)?

   - If to point where adequate effect could not be reached - stop

4. **How should implementation be improved?**

   - Reassess targetting, delivery (e.g. expenditure per caput)

STAGE 4: EVALUATING GROSS OUTCOME

1. **Choose outcome indicators.** Consider:

   - Responsiveness (Table I.9) and suitability (Table I.10)
   - Feasibility and cost of collection

2. **Measure gross outcome.** If data only obtainable on program participants, try:

   - Time-in-program method
   - Rapid collection of time-series data
   - Correlational analysis

   If data obtainable on non-participants and confounding variables see Stage 5, Table III.5.

3. **Evaluate gross outcome**

   - Is it adequate? (do outcome indicators meet preset standards of adequacy? - see Stage 2 point 4) - if so, usually stop for routine evaluation.
   - Is there reason to believe that further data collection or analysis would increase the estimated effect (e.g. evidence for negative confounding): if not, stop. If so, and answer worth the cost, continue to assess net outcome (Stage 5)

STAGE 5: EVALUATING NET OUTCOME

A. **If data estimating gross outcome are available:**

- Use data on (i) varying program delivery
  (ii) possible confounding factors (e.g. socio-economic status)

  to:
  - cross-tabulate by these factors and compare groups
  - control statistically for confounding and correlate with program delivery using multiple regression

This may give more plausible inferences on association of program with outcome (i.e. estimate net outcome). If these data are **not** available, or more certainty is needed, and surveys are considered worthwhile, go to B. If not **stop.**

B. **If survey to be carried out:**

- Decide on design (see Table I.5)
- Examples are:
  - interrupted time- series
  - non-equivalent control groups
- Analyze by:
  - matching
  - cross tabulation (as A)
  - control statistically and correlate with program delivery analogously to A

This may give additional plausibility to inferences on association of program with outcome.

C. **Evaluate estimates of net outcome**

- Is net outcome adequate?
- Is effect per unit cost adequate?
- How should program be modified (e.g. targetting, delivery, design)

STAGE 6: DATA PRESENTATION AND DISTRIBUTION

To whom?

- Already prepared principals (manager or funder) should have been involved in draft of report

How?

- Often iterative presentations to improve usefulness of report by better addressing needs of recipient both as to what is addressed and how (requires well done Step 1 in temporal organization of evaluation)

Feedback to other than principals?

- Do it with principals; ethics of confidentiality and censorship.
- busy people need short summaries of decision options + results + recommendations; careful people need everything relevant —fill these two needs separately.

ANNEX IV

(From Mason, J.B. and Haaga, J.G.: 1983 - Ref. 5)

IMPACT EVALUATION

1. **Purpose** - to estimate net effect of program on outcome indicators. Presupposes, but goes beyond, Routine Evaluation, to investigate the links in a causal chain connecting program activities with changes in health and nutritional status of intended beneficiaries. That is, goes beyond asking "Did activities happen as planned?" and "Are outcome indicators satisfactory in beneficiary population?" Information on effectiveness is potentially useful for decisions about expansion or replication of a pilot program. If combined with cost information (cost-effectiveness analysis), it can be used to set priorities, and to support resource allocation decisions at fairly high levels: among components of PHC, for example, or between PHC and rest of health sector. The audience for impact evaluation overlaps but it is not exactly the same as for adequacy evaluation. (1 and 2) In the UNICEF/WHO case, assessment of impact of different elements of the overall programme is crucial.

2. **General Method** - Evaluations try to compare the situation with the program or program component to what the situation would have been in the same population without the program. The latter is impossible to measure directly. Various research designs can give better or worse approximations, leading to greater or less confidence in results. Evaluators can never prove X caused Y, but they can strengthen their inference about the program's effect by taking into account competing explanations for observed changes. In an experiment (strictly defined), subjects are assigned randomly to treatment or control groups. This is not usually feasible for field interventions (as opposed to clinical trials). Quasi-experimental designs are more likely to be used in evaluating the components of pilot PHC programs.

66

3. **Quasi-Experimental Research Designs**

Note: In the following abbreviated descriptions of research designs, 0 = measurement of outcome indicators, with or without other variables

X = program actually begins operation. For further discussion, see Mason, Habicht and Tabatabai (1982 - ref. 4), table 1.6 and pp. 71 to 73.

A. Comparison with baseline survey on the same population.

Treatment group: $0_1 \rightarrow X \rightarrow 0_2$

This allows virtually no inferences on impact. The major problem is that too many other things change, while the program is going on. It is difficult to ascribe causation. Also, this research design won't pick up program effect "masked" by other trends. The health status of the population may be the same or worse at point 0_2, despite a positive net impact of the program, if the program served to prevent worse deterioration. Evaluations must take into account migration, and selection biases which would make the "before" and "after" group not truly comparable. Seasonal variations or natural disease cycles are often mistaken for program effects. Any effort to assess net impact should attempt to go beyond this design.

Repeated measurements can be made before and/or after starting the program (trend analysis).measurements are made before and/or after starting the program. $0_1 \rightarrow 0_2 \rightarrow 0_3 \rightarrow X \rightarrow 0_4 \rightarrow 0_5 \rightarrow 0_6$

If indicators show a clear break in health or nutritional status trends after initiation of the program, the inference that the effect is due to the program is stronger than comparisons with the baseline alone would allow. One still can't uncover "masking" effect of adverse changes in non-program factors, and migration and selection biases may still be operating, though their detection may be facilitated by this kind of monitoring. (Ref. 4, p. 71)

B. Comparison with national or regional averages.

Treatment group: $X \rightarrow 0_2$

Rest of population: $0_1 \rightarrow 0_2$.

This design is frequently used, especially in informal evaluations. Before making an inference about program effects, though, one would need to know or to estimate how comparable the two groups are that are being compared. It is easy to mistake a program effect for a difference in health status caused by non-program factors.

C. Use of comparison group receiving no (or different levels of treatment).

Treatment group: $0_1 \longrightarrow X \longrightarrow 0_2$

Comparison group: $0_1 \longrightarrow 0_2$

This design is often administratively difficult, but it can lead to strong inferences, depending on how similar groups were to start with, and how well changes in other factors related to health/nutritional outcomes can be monitored. Even if the groups are not exactly the same in levels of income, landholding, access to clean water, other services, age distribution, etc., or if they are differentially affected by environmental changes not related to the program, these differences can be partially controlled for in the analysis, if they are measured.

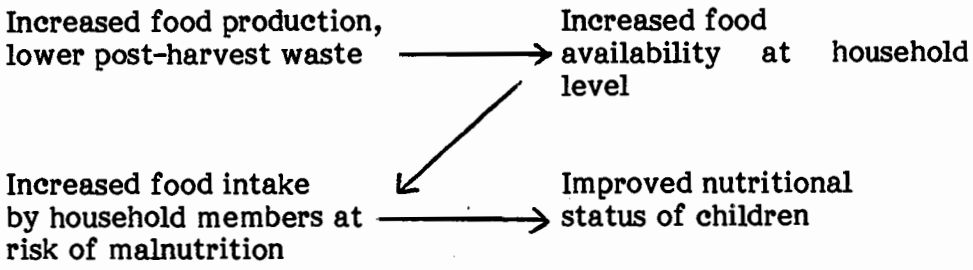
Comparison with incoming groups is a variant of C in which the comparison groups are new areas to which the program is being expanded. Most programs do not start everywhere at same time so baseline or first-contact data from the new areas can serve for comparisons with areas in which the program has been established longer. Again, measurement of other, non-program factors affecting health outcomes will strengthen the comparison.

Alternatively, in many situations program delivery itself may vary naturally. If this can be measured (e.g. number of patients seen, supplies delivered) the association between delivery and outcome can be investigated. Moreover, in programs not start gradually, the first measurements (outcome and delivery) taken with the program start may be able to substitute for the baseline survey.

4. **Example**

Evaluation of Impact of Household Food Production and Storage Program on Nutritional Status

A. Hypothesized causal links:



B. Research design: Compare villages where the activity is taking place to incoming villages in succeeding years.

The unit of analysis will be the village, on the assumption that this particular program component is targetted to entire villages rather than households.

Timing of Measurements:

Group	No. of villages	Year 1	2	3	4	5
A	30	A1	A2	A3	A4	A5
B	30		B2	B3	B4	B5
C	30			C3	C4	C5
D	30				D4	D5
E	30					E5

(Underlined entries indicate baseline surveys)

The effect of one year of program activities can be estimated by comparing:

A2 vs A1 **also**	A2 vs B2
B3 vs B2	B3 vs C3
C4 vs C3	C4 vs D4
D5 vs D4	D5 vs E5

Such a design allows larger sample sizes than before - after comparisons alone. It also makes it possible to control for extraneous events like an unusually poor harvest in one year, an epidemic, etc. If no baseline surveys are available, one can still estimate effects by comparing B3 vs A3, C4 vs B4, D5 vs C5, etc. Also one can look for longer-term impacts: A4 vs C4, A5 vs C5 and

D5, etc. The variation across 150 villages in the intensity and duration of program activities allows estimation of a "dose-response" curve: Do greater inputs always lead to improved outcomes? Under what circumstances? By how much? Up to what point of diminishing returns?

Potential confounding (competing explanations for differences observed in simple comparisons).

1. Accessibility - If the 30 villages in group 1 got services first because they were the most accessible (and accessibility is related to other factors that affect nutritional status- income from crops, availability of medical care), the single comparisons will over-state the program impact, if any.
2. Program or recipient selection bias - If the 30 villages in group 1 are somehow selected as those where service most needed, or if the most needy villages in each group are more likely to participate, then simple comparisons will under-state the true program impact. However, more commonly self-selection in the direction of over-estimating impact, because the self-selected participants are those most likely to improve.
3. More generally, non-comparability (in level and sources of income, age distribution and pre-program health status of population, etc.) of villages in the different groups will bias the estimation of program impact. This bias could be in either direction, exaggerating or masking the net impact of the program.
4. Differences in participation, and levels of program activities among and within groups of villages complicate evaluation studies. A very common problem is simply specifying what actually happened in the "treatment" villages, that is, what exactly is being evaluated. Especially when management decisions are decentralized, the evaluations need to be careful to distinguish between "X was tried but didn't have an effect" and "X was not in fact tried".

Simple between group comparisons are greatly strengthened if confounding factors can be accounted for (partially) in cross-tabulations or multivariate analysis.

In the case of a program which is reaching primarily the worst-off villages or households, then controlling for non-program variables may reverse the algebraic sign of a simple 2-way comparison of program and non-program groups. This applies with particular force when screening for selection into a program is used. The program participants should appear worse-off on simple comparison with non-participants. As another example: Suppose villages after two years in household food production and storage program still have a higher overall prevalence of underweight children than a comparison group of villages just entering the program.

Simple Comparison of Outcome Indicators:
% children 80% weight-for-age

Group 1 (2 years on program)	29% (30 villages, 1300 children)
Group 3 (just entering)	27% (30 villages, 1300 children)

But the villages in Group 1 have less access to adequate, clean water (and this is a presumed greater incidence of diarrheal diseases).

Comparison of Outcome Indicators Controlling for Water Supply:
% children 80% weight-for-age

	Water from standpipes (clean)	**Water from canals (dirty)**	TOTAL
Group 1	20% (n = 130)	30% (n = 1170)	27% (n = 1300)
Group 3	25% (n = 1040)	35% (n = 260)	26% (n = 1300)

The inference that the program brought about a 5% difference between Group 1 and Group 3 would be strengthened if the intermediate step (incidence of diarrheal disease) could be compared; and especially if similar results appear when other factors that affect attained weight-for-age are controlled.

Regression toward the mean complicates analysis if program participants selected by low values on a screening test. Statistical controls can only partly take the place of basic similarity between the treatment and comparison groups.

In the case of a program that doesn't reach the worst-off (because the truly destitute are unable to participate, or because it is administratively difficult to reach the destitute until after the program is well under way), controlling for non-program factors affecting growth similarly strengthens any inference about program effect (positive or negative) or lack of it. This time, the difference in outcome indicator between groups is less pronounced as more confounding factors are controlled, but the inference about the program's effect is strengthened as it "survives" multiple controls.

Simple Comparison of Outcome Indicators:
percent of children 80% weight-for-age

Group 1	22% (n = 1300)
Group 2	33% (n = 1300)

Comparison of Outcome Indicators Controlling for Water Supply
% children 80% WA

	Clean water	Dirty water	TOTAL
Group 1	20% (n = 1000)	30% (n = 300)	22% (n - 1300)
Group 3:	25% (n = 300)	35% (n = 1000)	33% (n - 1300)

These are two extreme cases: one where the children in the program villages appear to be worse off than the children in the villages without the program, but the net impact of the program is in fact positive; and the other where the confounding tends to exaggerate, not to mask, the program's impact. In practice, the situation is not always so stark, since some confounding factors may offset others to some degree, and since variation of the confounding factors between groups may not be so great as in these hypothetical examples.

(That is, the villages in the comparison group often happen to be broadly similar to those in which the intervention was first tried, particularly if there was some conscious effort to make them so.) In our examples, the program had a positive net impact, but the same types of confounding could operate either to mask or to exaggerate a net negative impact or to create an illusion of an impact where in fact there was none. Analyses are futher complicated if there are threshold effects of program components or interactions between components, (e.g. if measles immunization only has a strong effect if coverage is more than 80% over a 2-year cycle of transmission, or if the food production/storage program only has effect when combined with land redistribution or nutrition education.)

The variables usually needed include:

1. Factors to be controlled in analyses (e.g. other health services, water supply, household incomes, etc.)
2. Measures of program inputs (what services actually delivered? when?)
3. Outcome indicators (dependent variables in analyses).

Where possible, especially in small-scale studies of new interventions, various intermediate indicators related to steps in the causal chain connecting inputs to outputs will also be gathered. In our example above, intermediate indicators might include household food supplies or even dietary surveys.

5. Problems with impact evaluation

A. Expense - Impact evaluation cannot be instituted on routine basis, like the first two kinds of evaluation. The data and analytic requirements exceed those of adequacy evaluation. Accordingly once it has been established that an intervention does cause improvement in certain populations under certain conditions, administrators then monitor program effectiveness through types 1 and 2 evaluation until conditions change in such a way that the earlier conclusion may no longer hold. The oft-posed question which type of evaluation is more useful is beside the point - it all depends on the stage of the program, the degree of uncertainty about causation, and the decisions to be made. In deciding the level of resources to be devoted to evaluation as a whole and among the different types of adequacy and impact evaluation, program planners need to perform an informal cost-effectiveness analysis of evaluation research itself; some study designs may simply cost more than the expected value of the

information to be derived. (The more usual case, however, is that too few resources are devoted to evaluation, analysis, and reporting.)

B. Uniformity of "treatment" - As discussed above, this is a problem especially when management is flexible, or when the intervention consists of a complex bundle of services. Impact evaluation may be feasible only for discrete components of the overall package.

C. The choice of outcome indicators requires specification of the goals of the program and the purpose of the evaluation. Adequacy evaluations are often mistaken for impact evaluations, then the "outcome" indicators used are not actually measures of health status, but are measures of service delivery. The measures used as outcome indicators need to be sensitive enough in the relevant range to detect the impacts expected of a program in the sample size available. Mortality rates, especially cause-or-age-specific mortality rates, often cannot be used because of random fluctuations in small samples. Morbidity rates often suffer from "reporting bias" - the number of cases of infectious diseases reported in clincs or even household surveys may increase, despite decrease in true incidence, if program causes more cases to be detected and diagnosed. Growth rates are usually a good proxy for health/nutritional status of children, but may not be the outcome of interest for some PHC components.

D. Migration into or out of program areas. If those people adversely affected, or unaffected, by the program emigrate, the outcome indicators in the remaining population may not show the total net impact of the program. For example, some types of agricultural extension program may decrease employment opportunities for casual workers or seasonal migrants, or take over commons formerly used by nomads, etc. If the landless, migrants, or nomads are not around to be measured, evaluation studies could miss this effect.

6. **Conclusions** - Impact evaluation happens informally anyway, as part of health policy "folklore". (e.g. "X never works and costs too much" or "Y is an essential part of any PHC program") But planners need more specific guidance from past experiences: Under what circumstances might X work? or How much Y, and for whom? Impact evaluations, with their design and assumptions spelled out, could make experience of the UNICEF/WHO programs interpretable (and cumulative), - In the long-run one of the most important outcomes would

be to finally give solidly based guidance on what works, under what circumstances, and how much does it cost.