PNAAQ 006
ISN: 34622
7004 1146
62

A METHODOLOGY FOR DIAGNOSING CAUSES OF MALNUTRITION AT THE COMMUNITY LEVEL - THE CALI APPROACH

James C. Eckroad

January 1981

Office of Nutrition
Agency for International Development
Washington, D.C. 20523

Community Systems Foundation
1130 Hill Street
Ann Arbor, Michigan 48104

FORWARD

During the decade of the 1970s, many countries undertook nutrition planning activities with the objective of integrating nutrition concerns into national and sector-level policies and programs. These experiences led to a growing recognition of the need for micro-level analysis of nutrition problems and their causes and the desirability of regional and community involvement in planning activities. In response to this insight, the U.S. Agency for International Development undertook a program to stimulate the development of diagnostic and analytic tools for community-level nutrition planning.

As part of this program, A.I.D. provided funding for a project which would develop and test a methodology for diagnosing the causes of malnutrition at the community level and prescribe interventions that could be best implemented within the community environment.

A contract was given to the Fundacion Para la Educacion Superior (FES), a Colombian, non-profit organization which has worked cooperatively from time to time with the Community Systems Foundation of Ann Arbor, Michigan. Under this contract a University of Valle research team was assembled and this team and Community Systems Foundation staff together carried out the task of developing and field testing a methodology. The work took place in Cali, Colombia and the surrounding Cauca valley from 1974 to 1978.

In mid-1978 an evaluation team reviewed the methodology and field experiment and concluded that an approximate methodology had been developed and that it was ready for application under real life conditions. That methodology has been incorporated into this manual or field guide which will be of interest and use for those interested in the further application and testing this or any other community-level diagnosis methodology.

This document should not be viewed as a model or recommended approach but rather as a description and analysis of the approach used in Cali. It may be experimented with and adapted to different conditions and hopefully it will be modified and improved with use. It is anticipated that the field guide will also serve as a useful training tool.

The Office of Nutrition would welcome reports of experiences with its application.

Martin J. Forman
Director
Office of Nutrition
U.S. Agency For International
Development

Table of Contents

		Page
Table of Contents		i
List of Figures and Tables		ii
Preface		iii
INTRODUCTION		1
Chapter I.	THE DIAGNOSTIC METHODOLOGY	4
II.	COMMUNITY DEFINITION AND SELECTION	9
III.	SAMPLING PROCEDURE	12
IV.	SURVEY DESIGN	17
V.	DATA MANAGEMENT	33
VI.	ANALYSIS	39
Appendix A.	THEORY	46
B.	NUTRIENT GAP - NUTRIENT REQUIREMENTS	51
C.	SAMPLING PROCEDURES	56
D.	FAMILY COMPOSITION SURVEY	60
E.	ANTHROPOMETRIC SURVEY	64
F.	24-HOUR RECALL CONSUMPTION SURVEY	70
G.	FAMILY MARKET SURVEY	82
H.	HEALTH STATUS SURVEY	89
I.	ENVIRONMENTAL SANITATION SURVEY	94
J.	INDIVIDUAL CONSUMPTION SURVEY	99
K.	AGRICULTURAL PRODUCTION SURVEY	104
L.	COMMUNITY MARKET SURVEY	108
M.	ERROR ANALYSIS	112

List of Figures and Tables

		Page
Figure 1	- NUTRIENT FLOW MODEL	5
2	- BIOLOGIC UTILIZATION SUBSTYSTEM	6
3	- SAMPLE MAP SKETCH OF COMMUNITY	14
4	- SEQUENTIAL LOGIC BRANCHING MECHANISM	19
D-1	- FAMILY COMPOSITION SURVEY FORM	63
E-1, E-4	- MEASURING AND WEIGHING INSTRUMENTS	68
E-5	- ANTHROPOMETRIC SURVEY FORM	69
F-1	- 24-HOUR RECALL CONSUMPTION SURVEY FORM	81
G-1	- FAMILY MARKET SURVEY FORM	88
H-1	- HEALTH STATUS SURVEY FORM	93
I-1	- ENVIRONMENTAL SANITATION SURVEY FORM	98
J-1	- INDIVIDUAL CONSUMPTION SURVEY FORM	103
K-1	- AGRICULTURAL PRODUCTION SURVEY FORM	107
L-1	- COMMUNITY MARKET SURVEY FORM	111
Table B-1	- RECOMMENDED INTAKE OF NUTRIENTS	55
C-1	- TABLE OF RANDOM NUMBERS	59
F-1	- LIST OF FOODS	77-79
F-2	- LIST OF PREPARATIONS	80

Preface

The methodology presented in this manual evolved from the work and ideas of many individuals over a period of many years. While it is impossible to credit all of those who have participated, it would be unjust not to briefly recount the history of the work, mentioning some of the principal characters. There were actually three separate phases: (1) the initial period, during which most of the general ideas were formulated that led to an Office of Nutrition, United States Agency for International Development financed project with the Fundacion para la Educacion Superior (FES) and Community Systems Foundation (CSF); (2) the period during the project when the actual diagnostic procedures were developed; and (3) a final phase, which began during the project and extended beyond it, during which many of the ideas have been extended to other countries and have been modified for different applications.

The initial period corresponds to activities which were carried out in the Department of Systems Engineering at the Universidad del Valle in Cali, Colombia. Problem-oriented teaching methods were used by having groups of students work on solving real problems in the University Hospital. This eventually led to the identification of malnutrition as one of the most serious health problems in Colombia. The students then began to apply their "systems approach" to better understanding of the causes of this problem. This evolution is important because it resulted in a much broader definition of the problem than might have been expected from the traditional food and nutrition fields.

During this time the Universidad del Valle was heavily involved (as it still is) in very innovative work in the area of primary health care; there were also several international agencies in Cali working on related problems. Thus, there are quite a few persons (students, faculty and others), from a variety of backgrounds, who helped in the initial formulation of the concepts. The result was a proposal to AID, which Professor Wilson presented through FES and CSF, to develop a diagnostic methodology based on the "nutrient flow model."

AID decided to fund the project through the Office of Nutrition as "basic research" which may have applicability worldwide. During this

four-year project (1972-1976) the general concepts were converted into a practical field methodology. Because of the importance of this second phase, an attempt is made to name the principal persons who participated as project staff and contributed to the formulation of this methodology: Lusi Fajardo and Dean Wilson (co-directors), Oscar Bolaños, Alberto Pradilla, James Eckroad, Ramiro Muñoz, Fany Victoria, Bety de Ramirez, Gloria Quintero, Giovanni Acciarri, Arturo Erazo, Carlos Sistiva, Magda de Hernandez, Maria Luisa de Franco, Susana Villa, Aida Tello, Olga Estrada, Martha Cifuentes and Frankelina Viveros.

Also during the project there were many visitors from several countries in Latin America and Africa as well as from the United States. These visitors often led to activities in other countries which are attempting to apply parts of the methodology. Activities in the Dominican Republic, Chile, El Salvador, Honduras and Zaire have had some direct relationship to the project staff, and have helped to enrich the methodology.

In addition to the innumerable persons who contributed their ideas in the development of the methodology, the following persons have worked together to describe it in this manual in a way that will capture the environment in which it evolved: Richard Paullin, Dean Wilson, Luis Fajardo, Donald Schön and Giovanni Acciarri.

James C. Eckroad
Ann Arbor, Michigan
December 1980

INTRODUCTION

The nutrition planner must often deal with difficult intervention choices complicated by economic constraints and political pressures. What should he do? Should his resources be used for projects in sanitation and health, or should he select projects which are designed to increase nutrient consumption (e.g., food production, reduction of food losses, education on nutritious foods, etc.)? How should the planner allocate scarce human and financial resources among projects? What portion of his resources should be assigned to which program areas in order to maximize impact?

Seldom, if ever, does the planner enjoy the advantage of basing decisions on complete data; rather, he must usually rely on only partial information about the nutritional status of the population, and about the feasibility and comparative efficacy of various interventions. Compounding his dilemma is the seemingly endless array of diverse interventions from which to choose. These would include projects in such areas as: agricultural production, food supplementation, food processing and preservation, land reform, food stamps and/or rationing, nutrition education, potable water, sanitation, diarrhea control and even immunization. Moreover, in selecting projects or an intervention strategy, the planner is confronted with (often) considerable pressures from other institutions and interest groups to select a specific strategy.

This manual has been prepared to assist nutrition planners in selecting appropriate interventions for implementation at the national, regional or local/community level. Based on the experiences of the project team in Cali, Colombia, the manual presents a detailed description of a diagnostic methodology which has proven useful and practicable for planners to: (a) assess expeditiously and economically the nutritional status of a population; (b) determine the principle cause of malnutrition; (c) identify and compare various intervention strategies for that population; (d) select from among the possible interventions a program for improving nutritional status; and (e) specify in quantifiable amounts the probable benefit of the intervention program.

In addition, since the selection of a program -- using this diagnostic methodology -- is based on the analysis of certian explicit variables, these same variables can also be utilized as basic indicators for the

monitoring and evaluation of the program. Thus, the methodology has dual applications, in: (1) program selection, and (2) program monitoring and evaluation.

Because of the complexity of the nutrition problem and because the critical factors and relationships can be expected to change from one environment to the next, we do not expect this methodology to be applied uniformly in all situations. What works for the Cauca Valley in Colombia is not likely to be appropriate for the West African Sahel. Nevertheless, we believe that the basic diagnostic methodology is universally applicable; only the appropriate procedures will require modification to the specific environment.

Application of this methodology requires flexibility and creativity on the part of the planner. Therefore, it is important to understand the basic concepts incorporated in its formulation. While these concepts are discussed in more detail in Appendix A, we have summarized them here:

- The "scientific method" underlies the methodology. Quite simply, this is a procedure for learning more about phenomena which are only partially understood. It begins with what is known, postulates what is thought to be the case, and then submits explicit postulations to quantitative experimentation and verification.
- The "systems view" provides a framework within which the scientific method can be applied to specific instances. The "systems view" recognizes that the nutrition problem is the manifestation of a complex social system, and calls for the explicit postulation of those components which are relevant to the problem. Equally important, the "systems view" calls for quantification of the relationships existing between and among the various components of the nutrition system.
- The "inductive approach" to the planning process is the third fundamental concept incorporated in the methodology. This approach holds that the diagnosis of nutrition problems in different communities can lead to national level nutrition strategies which are more flexible in testing potential solutions and more responsive to local conditions.

It will be useful for the reader to keep these concepts in mind in reading this manual. The methodology presented here cannot be called a proven planning tool, but our experience in Colombia indicated that it has the potential for rationalizing the decision-making process for nutrition planners. However, that experience also makes it clear that much more experience is needed in applying the methodology in order to obtain reliable results. If the basic concepts are understood then the necessary experimentation with this methodology is much more likely to be fruitful.

The first chapter is a general discription of the diagnostic methodology, including a discussion of the "nutrient flow model" -- the basic framework for the methodology. Subsequent chapters provide general descriptions of each component of the diagnostic procedure. More detailed descriptions, based on examples from the Colombian experiences, are provided in the Appendices. These detailed descriptions are provided for the purpose of demonstration and are not to be considered as being universally applicable.

I. THE DIAGNOSTIC METHODOLOGY

One of the first steps in the development of the diagnostic methodology was to define and describe the nutrition problem as a system in order to facilitate the process of learning about and quantifying the complex relationships between and among the various components. This led to the elaboration of the "nutrient flow model."

This model is a description of the nutrition system that charts the "flow" of nutrients; it begins with agricultural production and culminates with the final biological assimilation of nutrients in the human body. The end result of this flow of nutrients is the nutritional status of the community. In its simplest form (see Figure 1, p.5) the nutrient flow model is merely a schematic depiction of the main components of the nutrition system: production, transportation, importation, exportation, storage, marketing, intrafamily distribution, consumption, biological utilization and nutritional status.

Using the basic schematic of nutrient flow, it is possible to model in much more detail any of the subsystems consisting of one component or a set of two or more components. For example, much work has gone into developing agricultural models. Given certain inputs such as climate, soil type, level of labor, level of technology, etc., it is possible to predict (with a reasonable level of confidence) the output in terms of quantity produced for any particular crop. Another example is the synergistic relationship between infectious disease and nutritional status. This relationship can be described in a more detailed submodel (see Figure 2) of the consumption, assimilation and nutritional status components, incorporating additional variables and the various types of infectious disease.

This process of elaborating details can be continued until an appropriate level of disaggregation is reached for a particular situation. The significance of the nutrient flow model in this iterative process cannot be over-emphasized: it provides the vital continuity and perspective for succesive attempts to arrive at the level of understanding needed to identify and design specific solutions to the nutrition problem.

In some respects the nutrient flow model is a misnomer since, in its current form, it does not contain the degree of rigor and detail expected

Figure 1

NUTRIENT FLOW MODEL

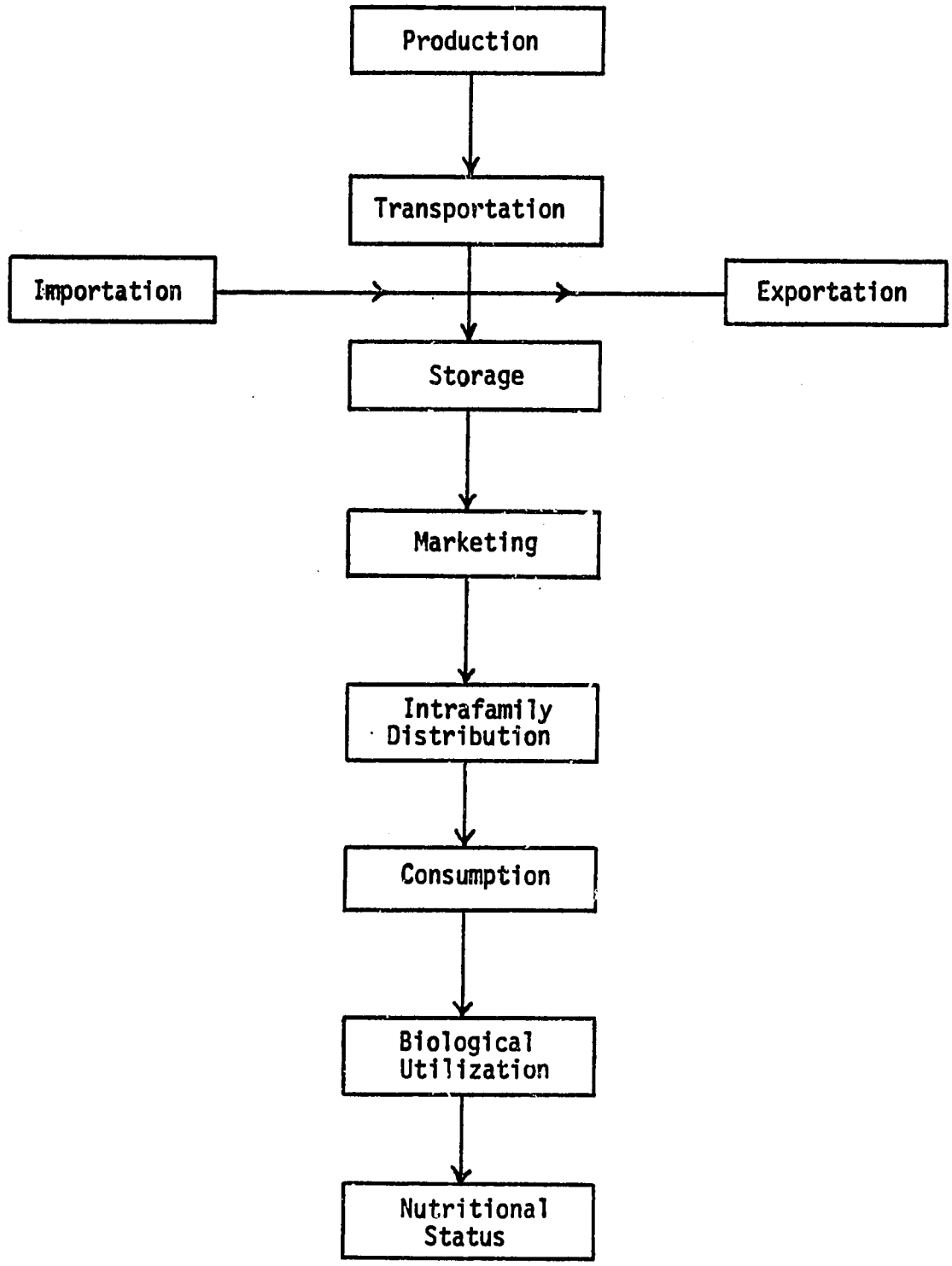
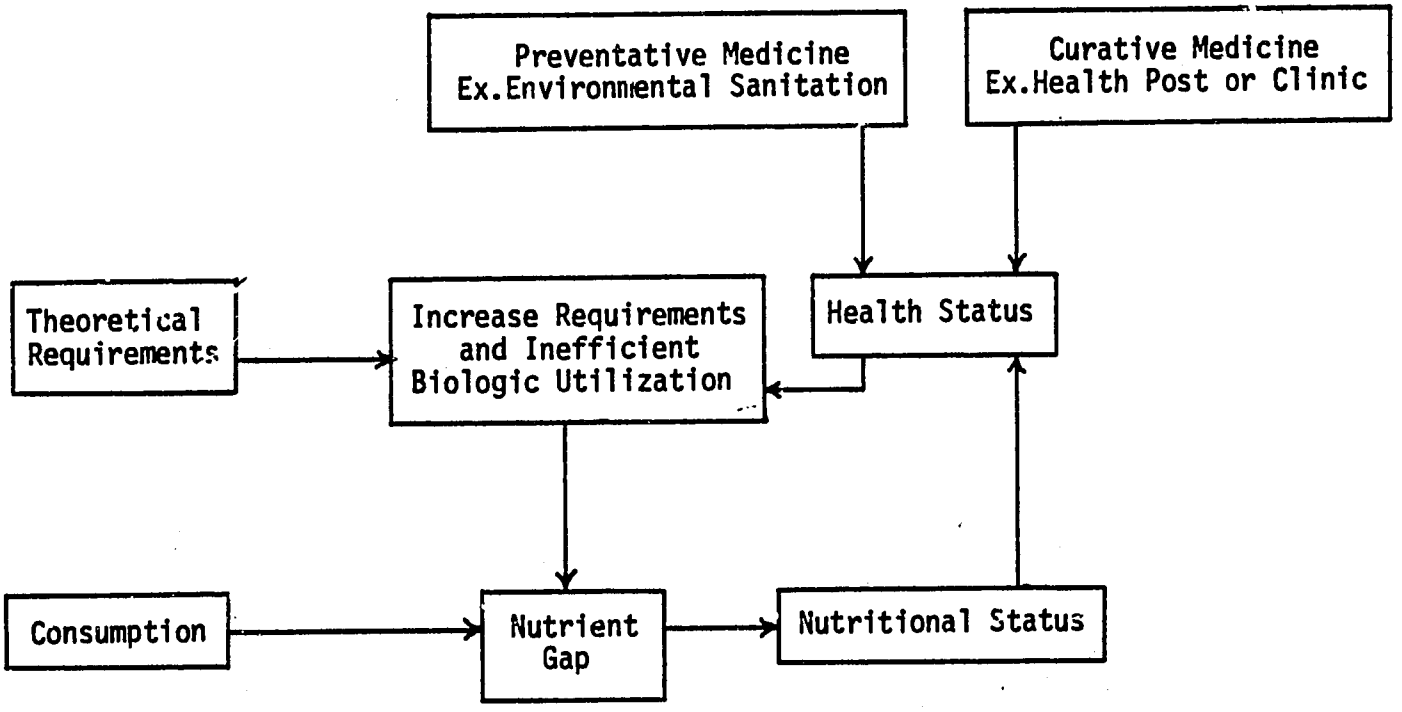


Figure 2

BIOLOGIC UTILIZATION SUBSYSTEM



of a "model." Nor would it be possible to construct a truly rigorous model of the nutrition system which would be applicable beyond the specific situation for which it was designed. For example, a model which is applicable to a remote rural community living primarily by subsistence agricultural would emphasize (and therefore include more detail of) the production and storage components; on the other hand, a model of a community comprised of a low income urban population would need to emphasize the marketing component together with a microeconomic model.

Thus, the nutrient flow "model" should be thought of as a general framework which helps to identify the salient variables that need to be included in a more rigorous model applicable to a particular environment. This framework serves two functions: first, as a tool to help design the more specific theories and models which will apply to individual situations; and second, to serve as a means to aggregate what is learned in specific cases at the local level into a more general body of knowledge from which decisions may be formulated at the regional or national level. That is, the nutrient flow model provides the structure for the inductive planning process -- it helps to organize the knowledge gained from application of the scientific method, and thereby provides for a dynamic approach to nutrition planning.

Once the nutrition system has been defined -- the variables and their interrelationships have been explicitly stated in a model or theory -- the scientific method calls for experimentation to determine if the relationships are stated correctly. Testing the theory requires that:

(1) indicators be selected to represent the variables;

(2) practical measurement techniques be adopted to provide the necessary quantification of the indicators; and

(3) an analytical procedure be developed to convert the measurement data into reliable conclusions that will serve to reinforce or modify the model.

In effect, this three-step process constitutes a diagnostic methodology for predicting the nutritional impact which might be expected from different interventions.

The first step in developing the diagnostic methodology for the nutrient flow model is to identify measurable indicators at various points throughout the nutrition system. This is done to quantify the amount of a particular nutrient or nutrients available at a given point (the actual points selected will depend on which variables are most relevant for a particular situation). A common denominator must be developed to facilitate the evaluation and comparative analysis of the quantity of nutrients available at the different points: this common denominator has been termed the "nutrient gap."

The nutrient gap is the difference between the required amount of a particular nutrient and the amount which is available. A similar concept is that of "adequacy," defined as the amount available divided by the amount required. These concepts can be applied usefully at the individual, family and community levels. This provides the flexibility to quantify the community nutrient gap at a point such as production; the family gaps at the point of food acquisition; and the individual gaps at the point of consumption. In this way it is possible to compare the nutrient gap at various points throughout the nutrient flow to establish the amount of deficit (or surplus) at each point as well as the change in nutrient deficit/surplus from one point to the next. (The concepts of nutrient requirement, nutrient gap and adequacy are described in more detail in Appendix B.)

The primary objective of the diagnostic methodology described in this manual is to be able to identify where, along the flow of nutrients, the problems (i.e., nutrient losses) are most severe. This, in turn, suggests identifying interventions, or areas of program activity, that might yield the greatest impact in terms of improved nutritional status. To be successful with this methodology, however, requires that the objectives be specified and purpose(s) stated explicitly and clearly <u>before</u> the diagnosis is undertaken. When the objectives for using the methodology remain unclear, the probability increases that the information gathered will either be misused or not used at all.

II. COMMUNITY DEFINITION & SELECTION

Definition

This diagnostic methodology, and its use as a tool for inductive planning, is based on the premise that the causes of malnutrition may vary from one "community" to the next (where initially "community" is taken to mean any grouping of inhabitants from a few families to an entire region or country). If this premise is valid then it follows that the size of the community should not be so large that it incorporates groups of families in significantly different circumstances with respect to the variables included in the model. Thus, major considerations in defining the community (or group to be studied) are: (1) homogeneity; (2) functional operation; (3) size; and (4) practicality for field work.

The concept of homogeneity is crucial in deciding the level of aggregation which is desirable for separating a larger population into groups of "communities." For example, if urban populations are to be included then care should be taken not to define a whole city as a "community." Rather, the city should be divided into sections (such as neighborhoods or areas within the city) comprised of families in approximately similar conditions with respect to housing, sanitation, and social and economic status. Likewise, the separation of a larger region into rural communities should tend to group families with similarities in environmental conditions as well as in employment and income patterns.

In addition to delineating homogeneous groups, definition of communities should try to insure that the family groupings do function as a community. This implies relative physical proximity as well as shared social services and activities. That is, families within a defined community would tend to use the same marketing structure, health services, schools, etc. Careful consideration should also be given to the size of the community -- it should not be so large that it is unmanageable for sampling and surveying; and to the practicality of conducting field work -- usually geophysical or political boundaries will provide readily identifiable limits.

Regardless of which specific characteristics are selected as relevant to the definition of community, these limits must be predetermined and

clearly defined so that there is no ambiguity over which families do/do not belong to the sample.

Selection

Once the definition of community has been established, there are several important considerations in the selection of communities for surveying under the diagnostic methodology. Since information gathered at the community-level is the basis for this inductive approach to planning, it follows that the communities should be representative of the different types of situations for which conclusions will be drawn. This requires that some sort of theory be developed about how different populations within the country (if the planning is at the national level) can be expected to exhibit different kinds of nutritional problems.

A classification should be devised to easily separate the population into groups. This classification system can be based on a variety of factors such as urbanism, geophysical characteristics (e.g., plains versus mountains), principal economic activity, and principal racial or cultural mix of the population. The primary considerations in developing the classification system are:

(1) that it be simple and easily applied using existing information; and
(2) that it be based on premises of how the nutrition system might be expected to vary among groups.

As diagnoses are performed for communities representing different groups, the classification system can be modified to the extent that the original premises are not confirmed.

The number of communities selected for the diagnostic procedure will depend on the needs of the inductive planning procedure and on the classification system. Several communities should be selected from each of the classification groups so that it can be determined whether or not a particular classification does in fact demonstrate similarities in the variables being measured.

In addition to the inductive planning requirement that the communities selected be representative of a larger population, there are other important

considerations. For example, if the nutrition planning activities are directed primarily at the most disadvantaged, then a stratification may be useful which helps select communities exhibiting extreme nutritional problems. Selection of "worst case" communities may also be helpful in that it will be easier, in the beginning, to quantify relationships which are very strong or predominant in a large portion of the community.

When the Colombian Department of National Planning used the methodology it first selected several departments (states) where, based on political objectives, it wanted to initiate its nutrition planning activities. It then selected at random forty municipalities which would be representative "communities."

Additional factors that can influence the selection process may not always be dictated by the objectives of nutrition planning. For example, some communities may be so remote that the cost of including them cannot be justified unless they represent a significant portion of the total population. Likewise, special interests and political considerations may influence the selection process. When factors extraneous to the planning needs enter into -- and substantially affect -- the selection process, it is important that they be explicitly identified in order to estimate their effect on the information gathered and on the conclusions drawn.

III. SAMPLING PROCEDURE

After a community has been selected, the next step is to develop a sampling procedure for gathering information representative of the community. First, the basic sampling unit must be defined. The diagnostic methodology depends on the family as the primary unit of observation and information collection, although we recognize that there may be cultural circumstances elsewhere that make it necessary to define a different sampling unit.

As in the definition of community, the description of the basic sampling unit -- the family -- must be clear and concise. It is necessary to specify exactly who/what constitutes a family. While "All persons living under the same roof or in the same living unit" may seem to be a useful guideline, the problem with this definition is that two different families may share the same residence. The definition found to be most useful in Colombia was to include "All persons who eat from the same pot." This proves to be a particularly functional definition since the allocation of resources for food and decisions about consumption usually apply to that group as a unit. Using this definition, the sample may initially be selected at the level of identifiable residences; when a situation is encountered in a residence where two or more groups are eating separately (i.e., from different pots), they are treated as separate families.

The planner may also find it helpful to define the sampling unit as only those families who have one or more preschoolers (children from zero to six years of age). This is suggested because preschoolers have high growth rates and thus it is easier to measure nutritional status in this group. Since nutritional status is the dependent variable, the end point in the nutrient flow model, the sampling unit should logically consist of families who can provide such data.

Having established the definition of the sampling unit, the actual sampling procedure must be designed. In some cases where the community is relatively small, measurements can be made on the entire population. If the community is large and complex, it may be necessary to employ more sophisticated sampling techniques (such as stratification or cluster sampling). However, in most situations the preferred method will be that of simple

random sampling. The following discussion considers this simpler technique.

To select a sample, all of the units (families) within the population (community) must be located and identified. If the community is defined by political boundaries, census information may provide a list of all the households with addresses or locations. Other agencies, such as the one responsible for malaria control, may also have local area maps that locate all of the houses.

In communities where such information is not available, it will be necessary to generate the data by riding or walking through the area and sketching the streets, roads and paths along with any important land marks to help orient the field personnel. Next, each of the residences is located on this map. This map does not need to be exactly to scale nor highly detailed (see Figure 3), but all of the houses must be included, and it must be possible to distinguish among the houses and locate them in the field.

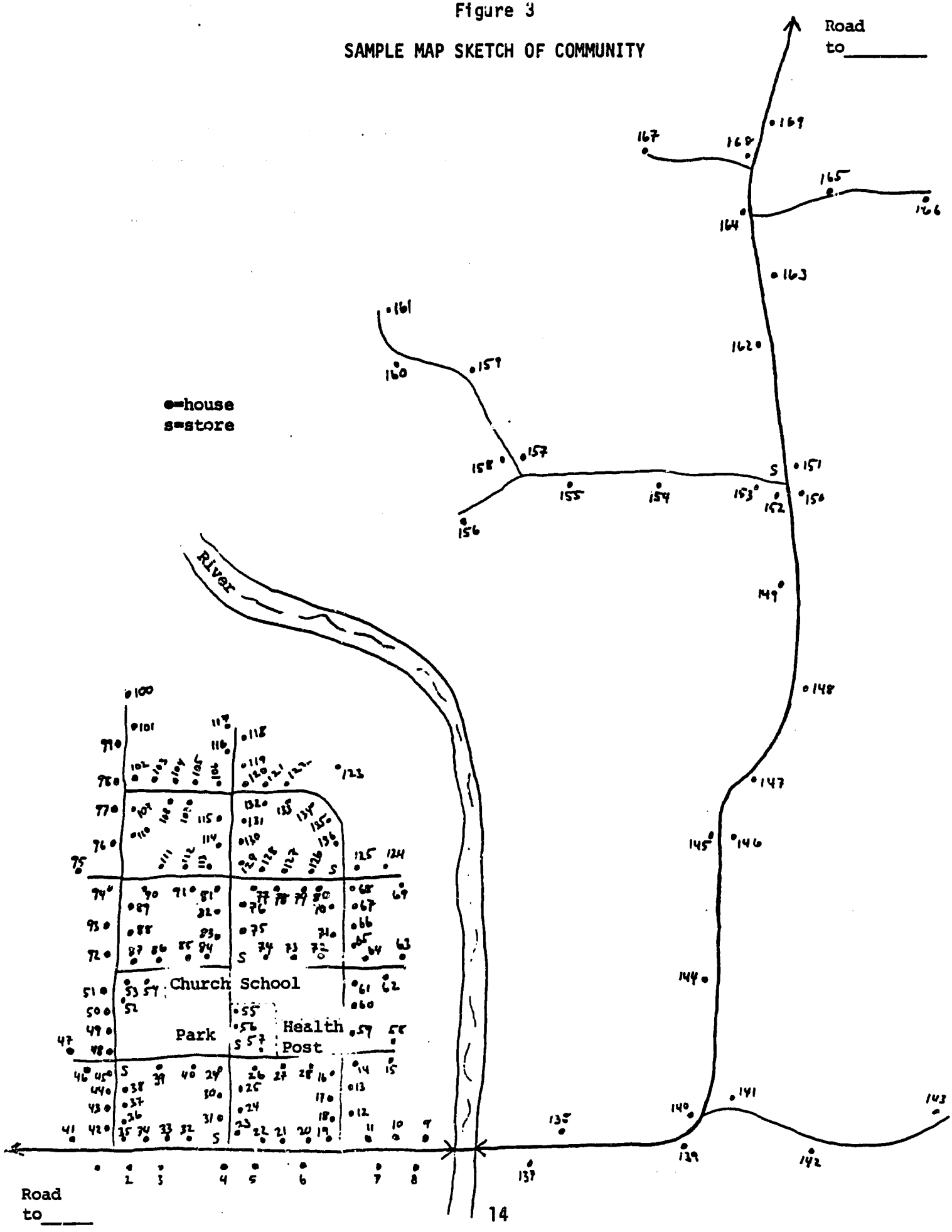
Once a complete list of all the residences or families has been compiled, the residences should be numbered consecutively, beginning with '1' and ending with the total number in the community. The number of families needed for the sample can then be easily selected from a list of random numbers. (A table of random numbers, with instructions on its use, is included in Appendix C; an alternative method, quasi-random sampling, is also described in Appendix C.)

Before the sample can be selected, a decision must be made as to the size of sample. There are two aspects to this decision. First, what sample size is needed to obtain the desired accuracy of conclusions? Second, how many families must be selected in order to complete the sample size?

There are several reasons why the final sample will be smaller than the number actually selected. As mentioned earlier, it is convenient to include in the sample only those families with preschool children; therefore, only about half of the families selected may actually be included in the sample. In practice, it will be necessary to select and make initial contact with a much larger number of families, eliminating from the sample those who do not have any preschool children. Secondly, some families may

Figure 3

SAMPLE MAP SKETCH OF COMMUNITY



not wish to participate. Since most of the measurement methodologies depend heavily on the cooperation of the families, the families must consent to participate. Some families will reject efforts to collect information, but extreme care must be taken to keep the number of rejections to a minimum. A significant level of rejections within the community could become a serious source of bias in the sample, since the characteristics of families who reject the surveys may be different from the characteristics of families who agree to be surveyed. Entrance into the community and the way in which the community views the activity are critical factors in the success of the sampling procedure.

The appropriate size of the final sample depends on the level of accuracy required for the analysis. There are statistical methods (see Appendix C) for arriving at the optimal size, but they are complex and may not be practical in many situations. If this is the case then the natural tendency is to make the sample "sufficiently large" to ensure a "reasonable level" of accuracy. While this may be a more practical strategy, it contains the risk of selecting a sample size that is much larger than necessary, thus resulting in a more costly survey and analysis.

To temper this tendency toward unnecessarily large samples, the planner may wish to consider another strategy, called "sequential sampling." In its extreme case, sequential sampling begins by selecting one family at random. This family is surveyed and conclusions are drawn with respect to the nutritional problems of the community. Although these conclusions would be tentative and highly uncertain, they would serve as a basis from which to predict the results of surveying a second family selected at random. If the predictions about the second family are correct, confidence in the conclusions can be increased; to the extent that the predictions are inaccurate, attempts should be made to modify the conclusions. This process of surveying additional families selected at random can be continued until it is possible to predict the results with the desired level of confidence.

In its extreme case sequential sampling is not practical because it would be too costly to have to continually return to survey one more family. While this may be impractical from an efficiency standpoint, its consideration re-emphasizes the importance of the concepts set forth in the

discussion of the scientific method (Appendix A). The use of sequential sampling in a particular community will reinforece the "learning" aspects of the diagnostic methodology. Although sampling, in groups of ten to twenty families at a time, may be feasible. If this is the case, then sequential sampling will help ensure that the total sample size is only as large as it needs to be, and it will also help to close the gap between data collection and data analysis.

There is no single definitive guideline to establish sample size. The proper sample size is a function of the characteristics of the community and the specific objectives of the diagnosis being undertaken. Nevertheless, for the work in Colombia it was found that a sample of from 100 to 200 families per community generally provided sufficient data to perform the desired analysis.

IV. SURVEY DESIGN

The most critical step in the diagnostic procedure is the design of the survey instrument -- a compendium of all the indicator measurements which will become part of the diagnosis. If the survey is designed correctly, the later stages of data management and analysis will be much easier to accomplish. The basic principle is that no data should be collected without a clear idea of its purpose and function; therefore, the survey design and the design of the analysis must be undertaken simultaneously. While the primary objective is to identify which points in the nutrient flow system are critical contributing factors to malnourishment in a particular community, the specific objectives of the diagnosis must be explicitly stated at the outset in order to limit data collection and analysis to only what is needed to respond to these objectives.

Sequential Logic

To facilitate simultaneous consideration of the design of the survey and analysis, it is helpful to discuss the concept of "sequential logic." This concept implies that data collection and analysis will be carried out on a step-by-step basis, first by selecting one factor or variable and then selecting and measuring the corresponding indicator; based on the analysis of this information, the next variable can be selected.

This iterative procedure continues, with each decision for additional data being based on the analysis of previously collected information, until the objectives of the diagnosis are satisfied. Subsequent variables are selected using this same sequential strategy. That is, existing information is analyzed and a working hypothesis is formulated; the working hypothesis serves as the basis for identifying which piece of information is needed next and for predicting the results of the information.

The nutrient flow model is the focal point for applying the sequential logic to the survey design. The first step is to modify the nutrient flow model so that it more closely represents the specific community under consideration. This modification represents the initial hypothesis describing that community's nutrition system; it consists of eliminating those components which are not applicable and, perhaps, expanding those components which are of greater importance.

Once the modified nutrient flow model has been elaborated, the procedure of sequential survey design begins with the selection of a starting point. The extent of malnutrition in the community is the first variable. If there is not a serious nutrition problem then it may not be necessary to proceed further. If there is a serious nutrition problem, a decision is needed as to which point in the nutrient flow should be measured next.

An established search routine is useful to provide an orderly and efficient method of selecting successive variables for measurement. The branching mechanism presented in Figure 4 is an example of the kind of sequential logic which might be used in the design of the survey instrument and the analysis. Rather than continuing step-by-step through each component of the nutrient flow model, a more efficient routine is to "cut" the system in the middle and then proceed forward or backward based on the results of the measurement at that point. An ideal point for cutting the system is at family consumption (or food availability). If the family gap (nutrients required minus nutrients available) is relatively large then there is insufficient food available and further inquiries should be made in the direction of food production and marketing. If the family gap is small, individual consumption and assimilation should be considered.

In most situations it will be too costly and time-consuming to apply the sequential logic to the actual execution of the survey. The value of sequential logic is in formulating the design of the survey instrument so that it is well-integrated with the kind of analysis to be performed.

The design of the survey instrument will not always include all of the variables deemed useful. In certain situations some variables are very difficult to measure. One example of this difficulty, the importation and exportation of food to and from the community, occurs when many families acquire food from markets outside the community and many farmers sell their produce individually to buyers outside of the community. This type of problem can be addressed indirectly by measuring variables on either side of the variable in question. In the community of Villa Rica, Colombia, a large quantity of soy was produced but not consumed, and a large amount of rice was consumed but not produced. Although it was not possible to directly measure either exportation or importation, it was possible to deduce

SEQUENTIAL LOGIC BRANCHING MECHANISM

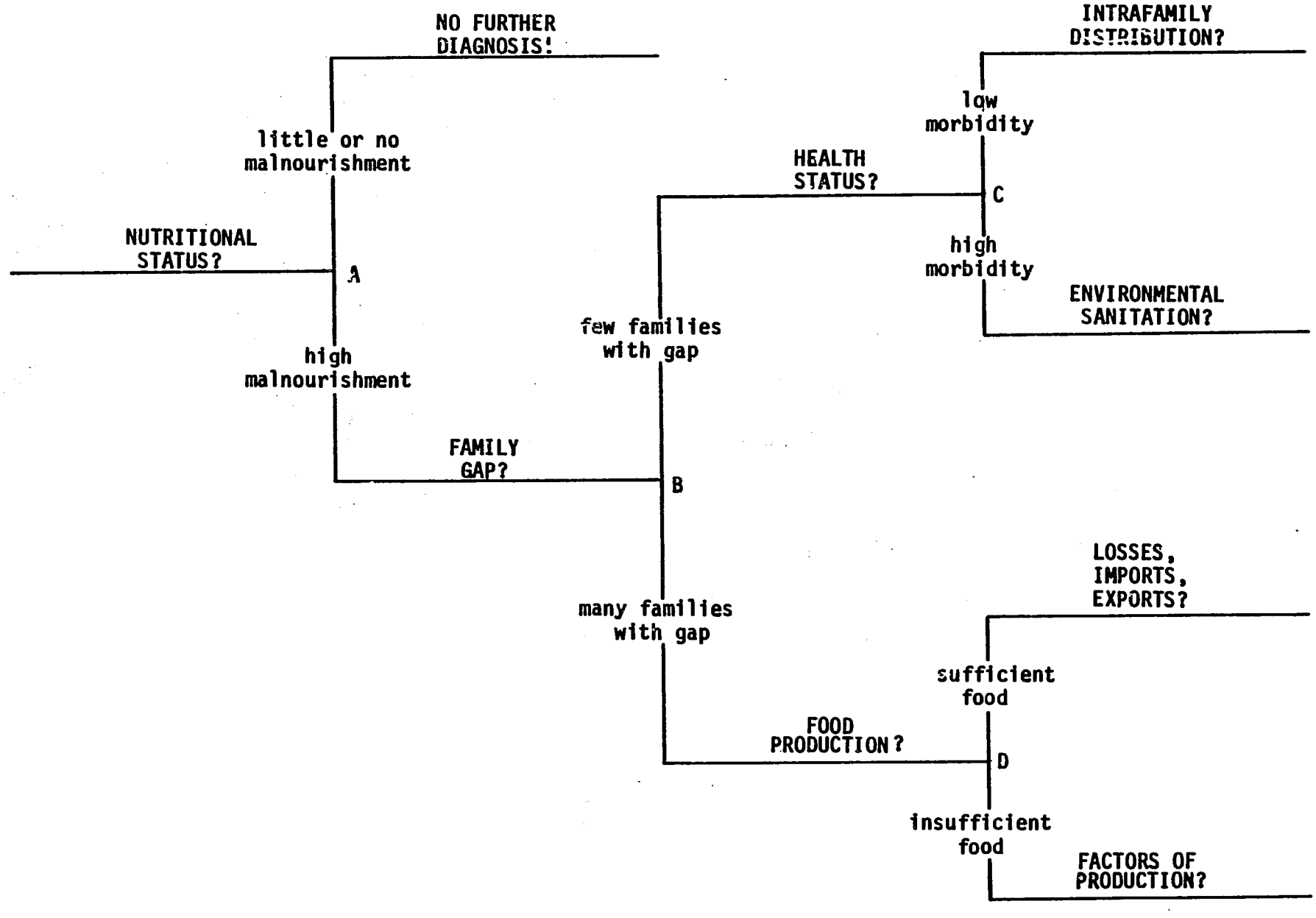


Figure 4

how much soy was exported (equal to the amount produced) and how much rice was imported (equal to the amount consumed).

An important consideration, then, is not only which variables are needed, but also which variables have indicators that can be measured within practical limits. The following is a discussion of the various components that comprise the nutrient flow system, and of the measurable indicators corresponding to them.

Family Composition

The first step in carrying out the surveys reflecting the different components of the nutrition system is to gather certain basic information on the composition of the families comprising the sample population of the community. Each member of the family is listed with information on their age, sex, relationship in the family and physiological status (in the case of pregnant or lactating women). It is then determined whether or not the family contains any preschool children; if it does not, the Family Composition Survey is completed and no further surveys are scheduled.

The Family Composition Survey (see Appendix D) is extremely important since it will be the first contact the survey team has with the individual families. While it is very important to have advance discussions with and acceptance by the community leaders, it is even more important that each family receives an adequate explanation of the diagnostic activities and their purpose in order to mitigate the inevitable spread of rumors.

Rumors can seriously affect the results of the diagnosis in two ways: (1) they can greatly increase the number of families who refuse to participate in the surveys, and (2) they can subtly bias responses to the various survey questions. Every effort must be made to avoid this situation; as such, the initial contact with the families is crucial. Once the explanations are complete, cooperation is requested and the additional surveys or data gathering activities can be scheduled.

Nutritional Status

This is the most important variable since it represents the output or endpoint of the nutrient flow system. As part of the diagnostic procedure, it is necessary to quantify this variable in order to determine the

magnitude of the nutrition problem. Quantification is also important in later stages of planning to know if an intervention produced the predicted change.

The most reliable and practical indicator of a community's nutritional status is the anthropometric parameters of its preschool children. The three basic parameters are height, weight and age; arm circumference may be useful where simplicity is required, but this is not usually as accurate. Height, weight and age can be combined to provide three different types of classifications:

(1) comparison of a child's height to the standard height for the child's age provides what is called "stunting" or "chronic malnutrition" classificaiton;

(2) comparison of a child's weight to the standard weight for the child's height provides what is called a "wasting" or "acute malnutrition" classification; and,

(3) comparision of a child's weight to the standard weight for the child's age provides what is called a "combined" (i.e., wasted and stunted; cronic and acute) or "global malnutrition" classification.

In general, the weight-for-height classification will provide the most useful indicator of nutritional status because it is a more precise indicator of the present condition and thus will be more sensitive to any change in nutritional status. Also, in some cultures it may not be possible to accurately determine the birthdate (or age to the nearest month) of children. Although the weight-for-height classification is the preferred indicator, it is also useful to collect data on age -- where possible -- because this may add an important dimension to the analysis of the nutrition problem. It is important to note that some experts feel that for very young children (e.g., up to one year old) the weight-for-age classification is the most accurate.

The anthropometric data for a population are evaluated by comparison with various national and international standards: We recommend the U.S. National Center for Health Statistics (NCHS) reference standards which are used by WHO and AID. There is considerable debate, however, over which national or international standard to apply in a given situation.

We have observed that the differences among the various standards are usually significant only when attempting to classify populations into such categories as "well-nourished vs. malnourished,: or "first degree" - "second degree" - "third degree" malnourished. These terms are merely labels representing arbitrary cut-points; the planner will have to establish his own cut-points which will result in classifying as malnourished that portion of the population with which the planner can realistically work. That is, if there is widespread malnutrition then a stringent cut-point, which classifies 80 percent of the population as malnourished, may not be useful if the programs cannot reach even half of the population. A less stringent cut-point which classifies as malnourished the 50 percent with a more serious problem, would be more operational.

Whatever cut-points are established, the planner may find it useful to express the classifications not as general labels but, rather, as percentages of the mean reference standard, or as percentiles or standard deviations (z-scores) with respect to the reference population. This will result in the following kinds of statements:

- a specific child weighs 72% of the mean reference weight for his height;
- a specific child is in the 10th percentile for his weight at a given age according to the reference population;
- the average weight of 1 year old children in a community is only 85% of the mean reference standard;
- 30% of the population is in the lower 10th percentile for the reference population.

The anthropometric classification of preschoolers is a relatively straighforward procedure (a detailed description of the procedure used in the Colombia project is presented in Appendix E). The important thing to remember is that the classification of preschoolers is being used as an "indicator" of the nutritional status of the community. The assumption is that the nutritional status of the children accurately reflects the status of the entire community. Although this assumption is relatively sound and widely accepted, conditions may exist in certain communities whereby a special nutritional deficiency in older children and adults will not be

reflected in the nutritional status of the preschoolers. If such a condition is expected, it may be necessary to check the nutritional status of these groups in a subsample of the population.

Family Food Availability/Consumption

Technically there is a difference between the food available to a family (i.e., that food in their possession) and the food they actually consume; this difference can be attributed to wastage and losses in preparation. In practice, however, such losses (particularly for poorer families) are minimal. Thus, food availability and consumption represent two slightly different approaches to obtaining the same information: the amount of food which a family has available for consumption. This component is important because, traditionally, nutritional problems have been associated primarily with food shortages and insufficient intake. Measurement of the nutrient flow system at this point will establish the extent to which this assumption is valid for a particular community.

Consumption measurements are generally difficult to perform and yield results of questionable accuracy. Techniques involving direct observation are slow (perhaps only one family per day per investigator) and usually represent a considerable inconvenience to the families involved. On the other hand, secondary techniques, such as questionnaire surveys, are relatively inaccurate. Nevertheless, the importance of the consumption variable requires that an effort be made to obtain this information. A possible solution to this dilemma is to gather the information by questionnaire and then to verify it by direct observation of a small subsample of families.

The most common method for measuring consumption is the 24-hour Recall Consumption Survey. This survey consists of asking the mother (or other responsible family member) to recall all of the items of food consumed by the family during the previous 24-hour period. This is usually done by first determining the dishes (or items) which were consumed and then determining the quantities of the various ingredients that went into the preparation of each dish. For a roll or piece of bread sold in uniform sizes, quantification is fairly simple; prepared dishes, however, often consist of many ingredients, making precise measurement difficult. For example, in many cultures the mainstay consists of a soup, stew or gruel containing whatever was available at the time.

Imprecise measures, coupled with inaccurate responses due to poor memory, make it difficult to quantify precisely the nutrients consumed by the family. However, to the extent that many of the errors are random (non-systematic), their impact may be minimal when the results are considered for the community as a whole. The 24-hour recall survey may be most useful to determine the quality of the diet and the way the different food items are being combined and prepared. This helps to establish the quality of protein in the diet and provides important information on eating habits.

It is important to note that the 24-hour recall survey as is used here refers only to the consumption of the entire family and does not distinguish individual consumption among family members. Although the 24-hour recall method has been used to measure individual consumption, we believe this application exceeds the limits of its accuracy and practicality. It must be recognized, however, that young children (0-2 years old) often will not participate in the regular family diet; they are either breastfeeding or eating special weaning diets. The 24 hour recall survey should be adapted to include the separate recording of consumption practices and dietary intake for this age group.

The 24-hour recall methodology for measuring family consumption was used as part of the diagnostic procedure in Colombia; a detailed description of its use is presented in Appendix F.

Family food availability is essentially a measure of the food acquisitions of the family over a specified period of time. The actual indicator and form of measurement may vary considerably depending on the characteristics of the community. If the families of the community are well-integrated into a market economy and purchase most of their food, the quantification of food purchases and market habits is appropriate. In this case, a market survey can be designed in which the family member who does the marketing is asked to recall the amounts of food purchased in the preceding period. The definition of the period should be consistent throughout the survey and should be of sufficient length to include the complete buying cycle of all the families. In the low income families that are likely to be the object of the diagnosis, the frequency of purchases is high and the quantities are low. Also, buying cycles in many cultures tend to be based on specific market days which occur once or twice a week. A period of one week would

probably include most of the average purchases for most families; however, a period of two weeks may be desirable to include more occasional purchases.

A number of related variables may also be included in the Family Market Survey; for example: information on family income, although often difficult to collect, has a very obvious relationship to food purchases. The portion of total income that purchases food is often a valuable indicator of the impact of a family's economic status on its nutritional well-being.

Other information relevant to the market survey concerns buying patterns. This includes the frequency of purchase of different items (Is rice purchased daily in small quantities at a higher price?), and the form of payment (Is food purchased on credit at a higher price?). A detailed description of the Family Market Survey developed for the Colombia project is presented in Appendix G.

If the families are not well-integrated into the market economy but, rather, consume a significant amount of the food they produce, then it will be necessary to develop a different or modified methodology. This type of situation is exemplified by the Mapuche Indians in southern Chile and marginal populations in northern El Salvador. The Mapuches produce and store wheat which accounts for a large part of their diet throughout the year; the same is true in the marginal areas of El Salvador where corn and beans are produced and stored. In both cases, the produce is stored by each family, so there are several ways the investigator can determine the availability of these staple food items. One method is to measure the stock at two different points in time (e.g., two weeks or a month apart). The reduction in stock then is equivalent to what was consumed during that period. Another method is to measure the stock at one point in time, then divide quantity by the estimated number of days remaining until the stock can be replenished with the next harvest. Care must be exercised to take into account non-human consumption (i.e., animal feed) and other losses which may occur in storage.

Nutrient Assimilation

If family food availability as well as dietary intake of the younger children appear adequate, and still there is empirical evidence of serious malnutrition, then the diagnosis should explore the efficiency with which nutrients are absorbed and utilized. The actual nutrient requirements of

an individual (healthy or not) may differ from the theoretical requirements (derived through a statistical treatment of the requirements of "healthy" individuals similar in age, sex and physical activity). This difference is further magnified by the effects of infectious diseases and other traumas that increase an individual's actual nutrient requirements.

It is widely believed that infectious disease is one of the principal determinants of malnutrition; since no practical method exists for measuring actual requirements in the field, it is necessary to identify another variable that will indirectly reflect this situation. The health status of the individual may be an effective alternative variable, but to perform complete health assessments of sample populations is time consuming and expensive. In addition, to determine the effect on nutrient requirements of a wide variety of health problems could require complicated calculations; for example: Is five days of common cold equivalent (in terms of increased nutrient requirements) to a case of the measles?

An alternative to the complete health assessment is to select one or more indicators of overall health status that also represent conditions likely to have a significant impact on nutrient assimilation. One such variable is diarrhea, which can be a symptom of a variety of different illnesses. For the purposes of the diagnosis, diarrhea is a good indicator of health status because it is one of the most common health problems in lesser developed countries, and it has a clear impact on nutrient requirements via two different channels: First, the presence of diarrhea usually represents the presence of infection, and when infection is present the body requires additional nutrients to combat the infection and replace cells which are destroyed; secondly, the existence of diarrhea usually indicates that the problem is located in the digestive tract -- nutrients which are consumed will not be efficiently assimilated; some portion of the nutrients will pass through the body without being assimilated.

There are some disadvantages in using diarrhea as an indicator for this component of the diagnosis, primarily because diarrhea is very difficult to measure quantitatively. Nor is it easy to arrive at a precise definition of diarrhea that can be agreed upon by the experts and, at the same time, be functional in the field. Likewise, it is difficult to grade the severity of diarrhea in a way that quantifies its impact on nutrient

requirements. Finally, seasonal variations in diarrheal rates may have to be considered, as well as the presence of other non-diarrhea illnesses which also increase requirements (and complicate the interpretation of data).

In spite of these problems, diarrhea remains the main variable for this component of the diagnosis. The information is gathered with a questionnaire asking the respondent to recall the episodes of diarrhea over a previous period of time. Since memory is unpredictable, the period should not exceed two weeks (one week is more realistic.) For the same reasons that the nutritional status indicator is much more responsive with children, it may be desirable to collect diarrhea information for children only; this is done by questioning the person responsible for caring for the child.

If diarrhea was present during the previous period, additional information should be gathered to determine: (1) the duration of the episode (number of days with diarrhea during the period); (2) the frequency (average number of depositions per day); and (3) whether blood or mucus was present. This information can then be used to construct a scale of severity.

In addition to diarrhea, it will be useful to try to incorporate other simple health variables, such as the presence of fever or colds (symptoms of upper respiratory infection), in the Health Status Survey. A detailed description of the procedure used in Colombia for gathering this data is presented in Appendix H.

Variables that reflect environmental conditions may also help to explain the existence of health problems; therefore, it may be desirable to gather data on a few simple indicators of environmental sanitation. The availability and quality of water, and the method for disposal of human wastes are usually the most helpful indicators; a discussion of these indicators is included in Appendix I -- the Environmental Sanitation Survey.

Individual Consumption

Probing this component of the nutrient flow system is done to determine if the individual's consumption is significantly different than the family's consumption. Although a particular family may be consuming nutrients equal to its requirements, some family members may be consuming more than their requirements and others consuming less. In fact, this possibility is the basis of many nutrition interventions (particularly food

supplementation programs) that target specific family members (e.g., pre-school children, pregnant and lactating women, school children, etc.).

It will not be worthwhile to include an Individual Consumption Survey unless information already exists indicating that intrafamily distribution is an important factor, or unless existing or projected programs emphasize the targeting of specific age groups or family members. This is because individual consumption is even more difficult and awkward to measure than family consumption since it requires a higher degree of discrimination among individual consumption patterns. The compromise mentioned earlier of measuring family consumption together with weaning practices and dietary intake of younger children should be considered before the decision is made to measure the individual consumption of all family members.

As mentioned during the discussion of family consumption, the 24-hour recall method can be used to measure individual consumption but its ability to differentiate consumption levels among family members is limited by its lack of accuracy. The alternative is to measure individual consumption by direct observation. This requires that the observer be present when meals are served. The observer actually weighs each portion (or measures volume) and accounts for the ingredients of each meal item. For multi-ingredient dishes such as stews, it may be necessary to assume that the proportion of ingredients is the same in each portion, but this should be attempted only if there is a strong indication that intrafamily distribution is a cause of malnourishment.

This method represents a tremendous imposition on the families, who often will resent having a stranger come into their home and "play" with their food before they eat it. If this survey is used, care must be taken to schedule the visit in advance and to explain the procedures. . If individual consumption data are needed it may be sufficient to perform this survey on a small subsample; since this information can only be collected at meal times, surveying the entire sample would take much longer to complete. A more complete description of this methodology is presented in Appendix F.

Food Production

The community's food supply must be considered when family-level food availability is insufficient. Food produced and consumed locally is more likely to benefit the poorer families since it will pass through few (if

any) intermediaries, and hence be relatively inexpensive. The importance of the local food supply to nutritional status will be even more pronounced if a significant portion of the community is directly involved in the production process and has some control over the factors of production -- particularly the land.

In structuring the Agricultural Production Survey it is useful to consider this basic relationship: production of a particular crop is equal to the land devoted to that crop multiplied by the productivity of that crop under the specific conditions of soil, climate, technology, etc. (Production = Land area x Productivity). This approach helps to identify several methods for determining production and verifying the data.

Production can be determined directly by observation and measurement at the time of harvest, or indirectly by asking the farmer; it can also be determined by multiplying land area by productivity. Land area, in turn, can be established by asking the farmer or by measuring the plot. This direct measure can employ relatively precise surveying techniques or simply involve a compass and a long tape measure. Productivity can be established by asking a local expert (such as an agriculture extensionist); it can also be determined by measuring sample harvest results.

The actual techniques employed in the Agricultural Production Survey will vary according to the local crops and production patterns. Corn, beans, rice or wheat, which are cropped singly and harvested all at once, represent a fairly straightforward measurement problem that can be solved using any combination of the techniques mentioned.

A different approach to measurement will have to be adopted for communities that practice mixed-cropping of perennials and harvest continually. For example, a traditional production pattern in Colombia is to plant coffee, cacao and plantains on the same plot, with fruit trees around the edges or near the house. The coffee is harvested over a relatively short period (several weeks), the cacao over a longer period, and the plantains and fruit are harvested throughout the year. The coffee and cacao are used primarily as cash crops while the plantains and fruit are used primarily for home consumption. In cases such as these where production patterns are complex, it will be necessary to depend on the production estimates of the farmer or agricultural extensionist, and perhaps verify these estimations by directly measuring a small subsample.

The precise production survey methodology will also be influenced by the importance of the particular crop under consideration. If the crop represents a major portion of the nutrient content of the diet (or has the potential to be) then it will be desirable to select relatively accurate techniques. Appendix K provides a detailed description of some of the techniques and of the survey used in Colombia.

Transportation, Storage, Importation and Exportation

Transportation and storage will often be functions beyond the control of the community; they may represent sources of losses for the regional or national food system, but only manifest these losses locally in the form of higher costs. This will not always be the case as, for example, in communities in El Salvador and the Mapuches in Chile where on-site storage is of great importance. Likewise, there are isolated communities where exports of produce and imports of supplies represent direct losses of time, resources and (in some cases) nutrients. Nevertheless, it is usually sufficient to describe these two components qualitatively during the first cycle of the diagnosis; later, if more information is required, techniques can be developed for quantifying these processes.

The combination of community importation and exportation represents a net gain or net loss of nutrients. If it is determined that these two components appear to be critical to nutritional status in the community, it will be necessary to design methods which will provide precise data. However, it is often possible to obtain a good approximation of these variables by deduction from other surveys. For example, the Family Market Survey records the source of different foods, and the Agricultural Production Survey records the destination of the products. This information, together with qualitative descriptions of the export and import mechanisms, may be sufficient during the first cycle of the diagnosis.

Income and Marketing

When the community is not primarily involved in agricultural production and most of the families are wage earners, the importance of food entry and distribution increases. The Family Market Survey can provide some basic information on families' incomes and purchasing patterns, as well as on the sources of food purchases. However, if food distribution is critical then more detailed information will be required.

To collect detailed information on the marketing system, first identify the different food establishments and classify them into groups. A community will usually have two types of stores: (1) larger ones that bring the bulk of the food into the community and wholesale it to smaller stores, in addition to some direct sales to the public; and (2) smaller stores that buy from the larger ones and sell to their own clientele, who are determined by a variety of factors including friendship or familial ties, location, price, availability of credit, etc. In addition to the regular stores there may be a market where once or twice a week there are a large number of sellers, or there may be other, non-location-specific sellers who vend certain specialty items such as eggs of milk.

Once this qualitative description of the marketing system is completed it is possible to design a survey to help quantify the importance of each type of commercial operation. Such a survey should determine, for each type of store, the kinds and volumes of foods (particularly the ones of significant nutritional importance to the community) as well as the frequencies and sources of supply for the different items sold. It will also be helpful to know the profit margins on the different kinds of commercial transactions, although this will be a very sensitive subject. Appendix L provides a detailed description of the Community Market Survey used in Colombia.

A Family Income and Expenditure Survey was not developed in Colombia because the diagnosis did not indicate that this significantly affected the community's nutritional status. However, the design of such an instrument is fairly straightforward once it is clear what kind of information is needed. On the one hand, it is necessary to determine all of the sources and amounts of income each family received during the preceding period (perhaps a month); on the other, information is needed on how much of that income was expended on different categories of purchases. Some categories such as food expenditures will need more detail than others. It is important to recognize that in many communities and for many families income will vary considerably from one period to another; therefore, it may be necessary to survey the community several times.

A carefully designed Family Income and Expenditure Survey will produce specific information that will contribute to an improved understanding of

the distribution component of the nutrient flow system. First, it will provide data on the amount of income that is available within the coummunity and how it is distributed among the various families; also, it will provide valuable information on how the income is budgeted among competing needs. By comparing the spending patterns of families at different income levels, it may be possible to better understand how economic decisions are made at the margin. That is, how would an additional income increment be distributed among competing needs? This information is useful in predicting the nutritional impact of certian economic-related interventions.

V. DATA MANAGEMENT

In the previous chapter, emphasis was given to the importance of simultaneously designing the survey and the analysis; it is equally important that at the same time, the plan for processing the data also be worked out in detail. This approach will reduce the long delays between data collection and data analysis. Data management includes a variety of steps, from designing survey forms (Chapter IV) that conform with the order and logic used to analyze the data; to the actual computer programming for data cleaning as well as the data processing which will make up the analysis.

We assume in this chapter that the planner will need access to some sort of computer to facilitate the data processing and analysis. Although data manipulation and analysis may, under some circumstances, be done without a computer, the computer performs two basic functions with great speed and accuracy: (1) sorts and merges (groups) data according to the values of certain variables, and (2) executes complex mathematical calculations. If the quantity of data is modest then manual sorting is possible; likewise, calculations can be performed manually or with a pocket calculator (particularly one of the more advanced models with statistical functions). These manual techniques may have particular relevance when the diagnosis is being done using the concepts of sequential sampling and/or sequential surveying and analysis.

Manual data processing, however, is extremely slow (even for modest amounts of data) and prone to a large amount of "human error." In most situations, therefore, computers are essential. The kind of computer and the arrangements for its use will depend on the particular situation. One possibility is to gain access to part-time use of one of the large computers which (1) are very fast, (2) have capacities for efficiently processing large amounts of data, and (3) have available a wide range of "pre-programmed packages" which can reduce significantly the amount of programming that would normally be necessary. Such a computer may exist within the planning department or some other government agency such as the department of census; access may also be available through a university or through the local offices of one of the large computer companies.

While there are many advantages in increased efficiency and sophistication with a large computer, recent advances in computer technology have

greatly increased the attractiveness of smaller computers. For example, it is now possible to select from micro- and mini-computers whose prices range from less than $500 to around $100,000. Given such a wide range of prices and capabilities, one of the smaller computers may well meet the minimal requirements for processing the data and, at the same time, be within the budget constraints of the project. Moreover, the "full-time" use of a smaller computer offers significant advantages over partial access to a large one:

- a large computer will have competing users and time constraints;
- there will be pressure to use the large computer efficiently (i.e., minimally) and not "play" with it in experimenting with new hypotheses or testing different explanations of a particular phenomenon;
- large computer facilities are usually physically separated from the project team's office(s), further complicating logistical considerations.

These factors can be eliminated by obtaining a small, "in-house" computer. If the diagnostic procedure is truly to be a learning process, then the diagnostic team, as a whole, should be integrated into the analysis process. To accomplish this it is necessary to reduce the "mystery" of the computer. "On-Line" data processing at the console of a mini- or microcomputer can enhance team interest in, and understanding of, the analysis in ways not possible with the "batch" processing common with the large computers. Although the analysis is planned in advance, it is also useful to encourage spontaneous querying of the system for answers to questions which arise in the normal course of analysis. This spontaneous involvement will only occur if the computer is readily accessible.

We recommend that whenever possible a small computer should be purchased for the diagnostic analysis, although even the larger mini-computers will not have the core memory capacity for simultaneous analysis of large amounts of data such as might be required for analyzing several different variables from several different communities. Also, some of the more sophisticated statistical methods may not be possible on the smaller computers, whereas they are pre-programmed in the larger ones. If a complex, multi-community diagnosis is anticipated then these considerations should be taken

into account; in this situation, the ideal solution will be to obtain a micro- or mini-computer with the ability to interface with a large computer: the smaller computer can be used for data entry and cleaning, as well as simpler analyses, while the more sophisticated analyses can be performed on the large one.

Once the computer arrangements have been made, the next step is planning how the data will be entered into the computer. The data must be in "machine-readable" form for the computer to be able to read or interpret them. Most commonly, data are stored in machine-readable form by key punching onto 80-column computer cards; or data are stored magnetically on tapes or discs (for larger computers), or on casettes or diskettes (for micro- and mini-computers). The selection of the form for data depends on the input devices available for the computer being used. With the large computers it is generally more efficient to first keypunch the data on cards and later "read" the data onto a magnetic form. With micro- and mini-computers it is more common to enter the data directly at the console where they are stored on a magnetic device.

When data are stored they are organized into "files." The master file or primary file contains all of the information as it was collected in the field. Later, secondary files are created which contain information in a more useful form. The data for the secondary files are constructed by performing basic operations on the primary data. For example, a child's weight is not very useful because it is difficult to interpret, but if it is compared to a standard for the child's height or age and reported as a percent of normal then it is easier to interpret and analyze/compare with other data.

The design of the primary file should correspond closely to the design of the survey(s). The process of putting data into machine-readable form should be as straightforward as possible. Whenever data are transferred manually from one form to another there is the chance for human error. The most direct method of transferring data from surveys to machine readable primary files is by using "precoded" survey forms. These forms have sequentially numbered boxes (up to 80 if the data are to be keypunched) where the data are recorded directly in the field. Each box may contain only one letter; the sequence in which the data are entered in the boxes is

the precise sequence which will make up the logical record.

When precoded survey forms are used the data can be keypunched or entered at the console directly from the survey forms. Some surveys may be too complex for precoded forms; in these cases, the data are collected in the field on regular survey forms, then transferred onto standard coding forms at a later time. From the coding forms the data go directly into machine-readable form. If the same person who collects the data in the field also does the coding in the office, errors in interpretation of the data forms are reduced. Example of precoded survey forms are found in Figures D-1, E-6, H-1, I-1 and K-1, while surveys requiring separate coding are described in Figures F-1, G-1 and J-1.

Once the data are stored in machine-readable form on the primary file, the next step is to "clean" the data. This consists of running logical checks on the data to determine if each of the values falls within certain limits of what is possible. For example, it is not possible for a six month old child to weigh 60 kilograms. If this value has been entered, the original survey forms must be reviewed; a common error is a misplaced decimal point. If a child's recorded weight is less than 55% or greater than 140% of normal weight, then it is probably in error. Consistency checks should also be made between variables; for example, a family without running water in the house is unlikely to have a flush toilet. While consistency checks such as these will not catch all of the errors, they will help to eliminate large ones which would tend to have more of an effect on the analysis.

It is not possible to give specific instructions for designing the consistency checks for cleaning data. Each variable must be considered and a decision made as to whether there are some values that would be impossible or <u>very</u> unlikely for that variable. When a particular value is outside the acceptable range, an attempt should be made to determine the "correct" value by looking for transcription errors from the survey form. If the problem is not found in the survey form then the variable must be remeasured in the field (if practical) or the value must be eliminated altogether. It is important, therefore, for the limits to be set at a level that will assure that suspect values are in fact wrong and do not represent unusual -- but actual -- situations. Eliminating values that accurately represent extreme situations seriously alters the analysis.

When the data in the primary file have been cleaned, work can begin on generating the secondary files. The first step is to enter into the computer (i.e., create files in machine-readable form) the additional information needed to construct the secondary variables. This information includes: (1) anthropometric standards for classifying the nutritional status of individual children; (2) nutrient requirements of individuals based on sex, age and physical activity; and (3) the nutrient content of the various foods which will be encountered in the different surveys.

The actual design of the secondary files (the organization of the logical records) is more important than the design of the primary file. The design objective for the primary file is to be able to transfer data in the most direct and simple way from the survey forms to the file.

However, the secondary files will be the ones used extensively in the analysis process; as such, the configuration of the records will have a great affect on the flexibility and efficiency with which the analysis is executed. The two main determinants for the secondary file design are: (1) the kind of analysis to be done (how many and which variables are to be analyzed together); and (2) the capacity of the computer (how much data can be stored at one time in the "core memory").

Core memory size is, in many ways, the key to the size and number of variables which the computer can work with at one time. If the core memory is relatively small (for example, less than 64,000 bytes), it will be necessary to move data back and forth between the core memory and the external storage devices (magnetic discs or tapes). For much of the analysis it will be desirable to know which variable values correspond to which individuals and/or families. If the variables are entered individually, additional space is needed to include the identification code; however, if a "package" of variables is entered as a logical record then the identification code is only needed for each record. Therefore, when there is insufficient core memory space, it is convenient to have files with record configurations containing all of the variables applicable in a particular analytic procedure. While this may result in inefficient external storage by requiring that certain variables be repeated in several different record configurations, and will lead to slower analysis, it will also provide a

less expensive way to gain flexibility from a small computer.

Within a file (primary or secondary) data are packed in a basic unit of data storage called the logical record. Each logical record must include a code or set of numbers which tells the computer: (1) the format in which the information is stored, (2) the community and family to which the information pertains, and (3) which variables are included in the record.

It will be noted that, for each of the surveys described in the appendices, the coding instructions call for the same kind of information in the first thirteen columns. The first column contains the survey code that, based on predetermined programming instructions, tells the computer which variables are found in what order and according to what format. Columns two through five identify the sector code and columns six through eight identify the family code. These two numbers indentify the family and distinguish it from all other families. If a particular family is surveyed more than once (as might be the case with the 24-Hour Recall Consumption Survey or the Individual Consumption Survey) then the survey date, columns nine through eleven, will serve to separate the different rounds of the survey. In the event that the data are stored on 80-column key punch cards then it may be necessary to divide the logical record among more than one card; columns twelve and thirteen will insure that the cards are in the correct sequence.

VI. ANALYSIS

The Sequential Logic

The analysis component of the diagnostic methodology consists of asking the "right questions" in the context of a sequential logic. To better understand this process, it is helpful to use the Sequential Logic Branching Mechanism (Figure 4, page 19) for selecting the questions; responses to these questions can be drawn from the various surveys described in the appendices.

The first question (point A in the Branching Mechanism) is: "Is there a significant malnutrition problem in the community?" This can be answered by examining the data from the Anthropometric Survey (Appendix E). Using the weight-for-height indicator and a "two standard deviations" cut-point, then the more specific question becomes: "What percentage of the children in the sample have weights which are two standard deviations or more below the normal weight for their heights?" The answer to this question must then be compared to the results of other populations (other nations, other regions within the country or other communities within the region) to establish the relative priority of this community. If the priority is low (e.g., only 5% of the children are classified "malnourished") then no intervention may be necessary for that community; however, the planner may wish to periodically repeat the Anthropometric Survey as a means of surveillance.

On the other hand, if the priority is high (e.g., 42% of the children are classified "malnourished") then further analysis will be necessary. At this point, the planner may also want to examine more carefully other aspects of the nutritional status component such as: (1)"Do other malnourishment classifications of the community (e.g., height-for-age and weight-for-age) correlate with and confirm the results of the weight-for-height classification?"; or (2) "Are there unusual distributions of malnourishment by age and/or by sex?"

The next point (B) in the branching mechanism suggests the following question: "Are there a significant number of families in the sample who have a deficit of nutrients?" This question is first answered by calculating the energy (kilocalorie) gap for each family using the results of the Family Market Survey (Appendix G). Consider two hypothetical responses:

(1) Only 15% of the families are consuming less than 90% of their requirements; or (2) over half (52%) of the families are consuming less than 90% of their requirements.

Before drawing conclusions from these results it is best to confirm them with other indicators. For example: Do the same families that have an energy gap also have a protein gap? Is the magnitude of the protein gap similar to that of the energy gap? The response for these questions will indicate whether the analysis can be continued based on the energy gap or whether protein consumption is the critical variable.

It will also be useful to calculate the energy and protein gaps using data from the 24-Hour Recall Consumption Survey (Appendix F). Do these correlate well with the gaps calculated from the Family Market Survey? If not, then a decision must be made on which indicator to base the continued analysis. One method is to average the results from the two surveys; another is to determine which survey yields more accurate information.

Assuming that the results of the energy gap for the Family Market Survey are valid, the the two hypothetical responses indicate which step should be taken next. If only 15% of the families have a deficit of kilocalories greater than 10%, it is unlikely that food availability is a major problem and the analysis can procede to point C in the branching mechanism.

However, if 52% of the families have deficits, two additional questions should be asked before moving to point D in the branching mechanism. First, is there a correlation between energy gap and nutritional status? That is, are the families with malnourished children the same families that have the energy gap, and do the worst cases of malnourishment belong to those families with the largest energy gap? If the correlation is weak then both points C and D should be considered; if the correlation is strong then a second question is asked before moving to point D: Is the average energy gap for the community as a whole also large? The response to this question will determine whether the deficits of half the families could be made up (in part) by surpluses of the other half of the families. If this is the case then the data from the Family Market Survey should be examined in more detail. Is there a difference between the two groups (families with deficits versus families with surpluses) in terms of resources (income) available or in terms of buying habits?

At point C we ask the question: "Are there health problems in the community which contribute to the malnourishment?" The first response comes from calculating the average prevalence of diarrhea from the data on the Health Status Survey (Appendix H). This will usually indicate whether diarrhea appears to be a problem. If average prevalence is high then morbidity should be considered to determine whether a relatively small number of children suffer from diarrhea every day, or if episodes are spread out over many of the children. It is also important to determine whether there is a correlation between: (1) those children with diarrhea and (2) those who are malnourished but whose families do not have a deficit.

If average prevalence of diarrhea is high and morbidity appears to be related to malnourishment, then the analysis should ask whether there are conditions indicated in the Environmental Sanitation Survey (Appendix I) which could explain the high prevalence. For example: Is there a correlation between diarrhea morbidity and distance to the water source? If diarrhea does not explain the malnourishment, consideration should be given to other health indicators such as the presence of colds or fever.

When the community's health status does not adequately explain the malnourishment attributed to this point, consideration must be given to the possibility that the food distribution among family members is not equal. The analysis must then consider the question: "Are there malnourished children whose families have sufficient food but who are not receiving their fair share?" This question can be answered by analyzing the data from the Individual Consumption Survey (Appendix J). If the "index of relative adequacy" for the preschool age members of the family is significantly less than one (e.g., 0.8), then this could explain part of the nutrition problem.

Point D in the branching mechanism should be considered when the number of families with an energy gap is relatively large. The question here is: "Does the agricultural sector of the community produce enough food to meet the community's kilocalorie and (protein) requirements?" This is answered by calculating the energy gap using the data from the Agricultural Production Survey (Appendix K). If only 60% of the community's energy requirements are being produced locally then the factors of production (resources available) should be examined carefully to determine the feasibility of increasing production.

If, on the other hand, the community is producing a sufficient amount of kilocalories (and protein), it should be determined whether this remains available to the community as food. One way to test this is to determine if the food items which are being produced are the same as those which contribute most significantly to meeting the requirements according to the Family Market Survey and the 24-Hour Recall Consumption Survey. For example, the community of Villa Rica in Colombia produces several times its requirements in both kilocalories and protein but this is primarily attributable to the production of sugar cane and soy beans. Both of these are considered cash crops and are exported from the community in their entirety.

Another possibility which should be examined is that the same kinds of food items are produced as are consumed, but they are produced by a relatively few members of the community and there is not sufficient economic demand to prevent the food from being exported. This would suggest interventions such as a food subsidy (e.g., food stamps) or income supplementation.

In addition to an unfavorable balance between imports and exports, other community losses should also be considered. There may be excessive post harvest storage losses or there may be inefficiencies due to large amounts of grains being allocated to livestock production.

Interpretation of Results

The above discussion attempts to demonstrate how, using a sequential logic, appropriate questions can be selected and, using the information from the surveys, responses formulated to indicate where malnourishment can be attributed in the nutrient flow system. It is important, however, to point out that this discussion may be misleading in that it may give the impression that the responses are usually simple to formulate and interpret. This is not likely to be the case when the analysis is done with real data.

There are several reasons why analysis of data collected from communities under field conditions will not necessarily lead to identifiable causes. One is that no community is likely to have only one cause of malnutrition. Most communities have a combination of causal factors; it is common to find food deficits and high morbidity. This requires that several iterations of the analysis be made as the magnitudes and probabilities of each response are compared and weighed.

Another reason is that, until experience is gained in the use of the methodology, interpretation of the values of the different indicators is difficult. For example: What is the significance of 40% of the families having an energy deficit greater than 10%, and how does that compare (in terms of nutritional impact) to an average prevalence of diarrhea of 30%? In the beginning intuition will have to determine the answers to these kinds of questions, but as the learning process progresses the answers will be based on what was found in other communities or in the same community at an earlier time.

Other reasons complicating the analysis are related to what might be termed "errors" in the measurement procedures and data processing. The term "error" as it is used here refers to all cases where an "observed value" is different than an "actual value." Errors are not limited to careless mistakes by project staff; rather, to a large extent they are the result of difficulties inherent in trying to quantify variables and relationships in the area of social science. Because the problem of errors is so complex and at the same time so important, it is discussed in detail in Appendix M. In general, however, all errors will either be biased (systematic) or random. Random errors will tend to dilute or hide relationships; biased errors will tend to produce false relationships which may lead to contradictions of logic.

Program Design

Each point in the nutrient flow model and each question formulated in the analysis suggest various kinds of interventions and programs. By defining and quantifying the nutrition system in terms of nutrient gaps, the diagnostic methodology provides performance specifications for selecting and designing interventions to address a community's malnutrition problem(s).

The concept of performance specification can best be understood by considering specific examples. The response to the first question (point A: "Is there a significant malnutrition problem in the community?) provides a performance specification at a very general level: for example, a program is needed to reduce the number of malnourished children from 42% to 5%. Point B in the branching mechanism is more specific: For example, a program is needed to reduce the number of families with a 10% or greater energy deficit

from 52% of the population to 15%.

As the analysis progresses the performance specifications become more specific: (1) a program is needed to increase production of kilocalories by 35% or increase production of potatoes by eight metric tons per year; (2) a policy is needed to reduce the community's export of specific food items by 20%; (3) a program is needed to reduce the average prevalence of diarrhea from 40% to 5%; or (4) a program is needed to provide safe drinking water within 10 meters of the house for the 80% of the families who do not already have this service.

Performance specifications such as noted above provide the planner with criteria for selecting and designing specific programs. Clearly there are other criteria which may act as constraints on the specification. Economic considerations may impose constraints that a particular program cannot cost more than a certain amount. There may be political constraints -- a program with potential for meeting a performance specification may not be politically feasible to undertake for a particular government. For example, a program of land reform may be a way to reduce exports by shifting the crop mix from cash crops to food crops, but land reform may not be an acceptable option. Also, there may be stability considerations; for example, it may be possible to reduce the family gaps by large food supplementation programs, but over the long run these would create undesirable dependencies. Likewise, there are time constraints whereby a program (such as land reform) would reduce the family gaps but the effects would not be realized for five years; this may suggest complimentary strategies such as a food supplementation program to cover the period in which the land reform is being implemented.

Once the planner has organized the performance specifications from the diagnostic methodology together with the other criteria, he/she is ready for final project selection and design. It is at this point that the expert is called. The engineer is told what is needed in the way of environmental sanitation; the agronomist receives instruction on the increases in agricultural production that are needed; the public health doctor is advised of the need for reductions in morbidity; the economist is asked to devise policies for income transfers or food subsidies. By providing these professionals with very explicit criteria for developing programs, the planner is able to convert the results of the diagnosis into interventions aimed at

specific identifiable aspects of the community's nutrition problem. The expert is expected to present specific program plans together with their estimated costs. These costs, when considered with the performance expected of the programs, will provide a cost/effectiveness measure for comparing and selecting among projects. In the implementation phase the performance specifications which were used in the design of the programs, together with the estimated costs of the programs, will become the criteria by which the programs can be evaluated.

The analysis is first performed on individual communities. However, as performance specifications and program designs are developed for each community, certain similarities will begin to emerge for groups of communities. These similarities are fed back into the context of the inductive planning approach and conclusions are generalized to formulate a nutrition strategy for larger populations (e.g., regions within a country, or an entire country).

THEORY

The Scientific Method

The "scientific method" underlies the methodology. Quite simply, this is a procedure for learning more about phenomena which are only partially understood. It begins with what is known, postulates what is thought to be the case, and then submits explicit postulations to quantitative experimentation and verification.

Nutritional problems have existed since the beginning of time, but only recently has there been widespread interest in attacking "the nutrition problem" in a concerted, multisectoral manner. Many of the activities undertaken in the name of "nutrition planning" are little more than the programming of interventions thought to affect the nutritional status of the population. Planning for nutritional improvement is a complicated process, inextricably meshed with the immense complexity of development itself. Little is known about either; both confront planners with the unenviable task of selecting interventions or policies on the basis of incomplete information.

Much of what is known about nutrition planning is of a qualitative nature and therefore insufficient to assist in the selection of "good" projects in an environment of scarce resources. For example, it is generally accepted that a water supply project will have a positive impact on nutritional status, and the same is true of a supplemental feeding program; but which will have a greater impact in a particular community is unknown.

An approach to nutrition planning that provides for a rapid response to a servere problem is essential: the planner and the population for which he is responsible cannot await the outcome of long studies. At the same time, however, this planning approach must somehow contribute to a better understanding of the problem as the response activities are being conducted. Such an approach must view nutrition planning as a learning process. Moreover, in order for this learning process to be effective the planner must be able to say that he is learning: it is very difficult to learn if one is required to pretend to be certain. The first step is to admit that the solutions are not already known.

In many different situations the best, time-tested, learning tool has been the "scientific method." Essentially, the scientific method is a process; the process begins with the formulation of a theory which is based on current knowledge. Next, observations are made of certain variables identified in the theory and an action is designed which the theory predicts will cause changes in the variables. Finally, the action is carried out and new observations are made to determine if the predicted change(s) in the variables actually occured. If the predicted changes occur, then the theory is reinforced; if they did not, then the theory is modified or reformulated.

The scientific method does not produce instant knowledge; rather, it provides a mechanism for developing an ever more sophisticated and precise understanding of the problem by systematically utilizing and improving on existing knowledge. The diagnostic methodology described in this manual has its origins in the scientific method, and that method makes up one of the principal characteristics of the methodology. The main purpose of this manual is to provide a decision-making tool for nutrition planners, to serve as a guide for selecting actions and as a framework for learning as the activities are undertaken. Hopefully, this will satisfy both the need for immediate action and the need for more complete knowledge.

The Systems View

The "systems view" provides a framework within which the scientific method can be applied to specific instances. The "systems view" recognizes that the nutrition problem is the manifestation of a complex social system, and calls for the explicit postulation of those components which are relevant to the problem. Equally important, the "systems view" calls for quantification of the relationships existing between and among the various components of the nutrition system.

One of the principal considerations in nutrition planning is that the nutrition problem is multifaceted. It is what has become known as an "inter-disciplinary" or "multisectoral" problem, thus suggesting that its solution requires a broadly-based approach incorporating many disciplines. But in practice this multisectoral approach is not well understood.

Some believe that the multisectoral approach means attacking the problem simultaneously in each of the relevant sectors; others argue that the attack must also be coordinated and/or integrated. Yet few have successfully applied these approaches in practice, for solutions to the nutrition problem are not to be found as much within the various sectors as at the interfaces between the sectors. For example, a policy to increase the production of food can have little impact if there are not complementary policies in the areas of transportation, storage and distribution. Likewise, it does little good to produce more food if those in the greatest need are not able to afford to purchase it.

Another impediment to realizing an effective multisectoral approach is the way that professionals tend to define problems. The doctor has his training in medicine and most of his experience is in that field. The agriculturalist's training is focused on agricultural production and much of his experience is limited to that area. The same is true of the educator, the engineer, the nutritionist and the economist. Understandably, it is difficult for these professionals to perceive the nutritional problem in a context which extends beyond their own experiences and interests; each will tend to define the primary emphasis within his own profession or sector.

For multisectoral planning to be effective requires a framework which forces recognition of the interconnections among the various factors that affect or contribute to the nutrition problem. This has led us to the "systems view": the systems view recognizes that the nutrition problem, like other problems, can be viewed as a system of factors or variables which are interrelated in specific and quantifiable ways. Using this approach requires that all relevant variables are specified in advance, and also that the relationships among the variables are quantified in some consistent manner. In application, this systems view of the nutrition problem allows the planner to test for the causes of malnutrition within the system (i.e., within a specific community), and to compare -- in quantifiable terms -- various intervention strategies for maximum impact.

The Inductive Approach

The "inductive approach" to the planning process holds that the diagnosis of nutrition problems in different

communities can lead to national level nutrition strategies which are more flexible in testing potential solutions and more responsive to local conditions.

Most government planning activities occur at the centralized level; policies are formulated from offices in the capital city for implementation in regional or local settings. This type of planning can be characterized as the "deductive approach," since the policies are often based on a few premises or on a generalized description of the nature of the problem from which conclusions are drawn with respect to specific situations.

Although this deductive approach is most common in planning and is quite useful in many circumstances, several important disadvantages exist. Often the policies in such a center-periphery system must be simple and relatively uniform because of administrative limits on the variation and complexity of policies for implementation in the peripheral localities. A more important disadvantage is that the deductive approach limits the amount and variety of experimentation possible. Since policies are based on generalized premises, only those limited premises can be tested. Because interventions are formulated from a generalized theory of the problem, the deductive approach does not allow for testing the possibility of different causal factors being predominant in different localities. The nature of the nutrition problem requires a planning approach that allows decision makers the opportunity to try many different activities in a variety of different communities to maximize learning possibilities.

The converse of the deductive approach to planning is the "inductive approach," which begins by looking at the localities to draw conclusions with respect to the general situation. This approach promotes experimentation with different kinds of interventions in different localities. The general theory and overall policy are induced from actual situations in specific localities. This inductive approach to planning provides more flexibility for learning about the complex problem of malnutrition.

This manual applies the inductive approach to nutrition planning, with the community as the basic unit of analysis. The purpose of this approach is to diagnose the nutritional status of a particular community.

There are several reasons why the community is appropriate for carrying out inductive planning. Policies are often implemented at the community-level, especially when they call for such traditional interventions as: health delivery systems (health posts), environmental sanitation (water supply and waste disposal), food supplementation (school feeding, MCH and food-for-work programs), and non-formal education (health and agricultural extension services). Even those programs that are characteristically national in scope (such as import-export and price support policies) can be made more effective when the dynamics of their effects at the community-level are better understood.

Another reason for concentrating on diagnosis at the community-level is the need for defining a "workable" group. Although it may, in some cases, be useful to make the diagnosis at the family level, experience has shown that the measurement instruments involved do not have the necessary accuracy to disaggregate conclusions to the individual families. On the other hand, undertaking a diagnosis at the state or regional level defeats the purpose of the inductive approach to nutrition planning (i.e., to promote experimentation with different kinds of interventions in different localities), because it introduces substantial variation.

Our basic assumption is that the causes of malnourishment can be expected to vary from one population group to the next. As such, it follows that a homogeneous population group is the ideal unit for study and experimentation.

NUTRIENT GAP - NUTRIENT REQUIREMENTS

The concept of nutrient gap (or adequacy) is important throughout the nutrient flow system. By calculating the nutrient gap at the various points in the system, it is possible to make comparative evaluations of the nutritional situation at each point. Essentially it is a way of comparing the amount of food available to the amount of food required at any particular component of the system.

The general idea of the comparison of requirements to availability can be defined in slightly different ways to convey different information; in other words, there are different kinds of nutrient gaps. For example, total requirements less total availability of a particular nutrient is called the "absolute gap." Dividing the absolute gap by the total requirements yields the "relative gap." If a particular individual's daily energy requirements are 3000 kcal but his actual consumption is 2300 kcal, then his absolute gap and relative gap are calculated as follows:

- absolute gap = requirements - availability = 3000 - 2300 = 700 kcal
- relative gap = $\frac{\text{requirements - availability}}{\text{requirements}} = \frac{3000-2300}{3000} = .233$ or 23.3%

Thus, it is possible to calculate the nutrient gap at the community, family or individual level. For those components pertaining to the community (such as production, import/export or marketing), the gap would be reported at the community level; however, by dividing the absolute gap by the number of familes or by the number of individuals it is possible to report this information at the family or individual level. This is helpful in determining the average shortfall of nutrients per family or per individual. Family food availability or consumption data would first be calculated as family gaps, but this information could also be disaggregated as individuals gaps.

The fact that there are two types of gaps (absolute and relative) and three different levels (community, family and individual) provide a wide range of possibilities for reporting this information. This allows the flexibility to consider the nutrient gap in the form most useful for a particular analysis. For example, if community food availability (production + imports - exports - losses) is determined for k calories, then useful

forms for reporting this would be the relative gap for the community (e.g., a 25% shortfall) and the average per capita absolute gap (e.g., each person, on the average, needs an additional 500 kcals). It is also useful to compare the average per capita absolute gap at the point of production and at family food availability.

As stated earlier, the concept of nutrient gap is essentially the comparison between nutrient availability and nutrient requirements at different points in the nutrition system. The nutrient availability is determined differently for each of the survey components (production, marketing, family consumption, etc.) which are described in more detail in other Appendices. However, the concept of nutrient requirements is complex and needs further explication.

Because it is not feasible to measure requirements directly in a field setting, it has been necessary to estimate requirements based on the statistical treatment of data taken from groups of individuals under controlled or laboratory conditions. These estimates result in average values for different sex and age groups, and different environmental and working situations. The values thus derived are said to be "theoretical" requirements in that they represent requirements of normally healthy persons. Sick persons will have considerably greater requirements since illness often affects the efficiency of metabolic processes and additional nutrients are needed to combat the illness and replace damaged tissues.

The accompanying table (B-1) on recommended nutrient intakes provides values for estimating the theoretical requirements of most nutrients. While information is included for a variety of different nutrients, experience demonstrates that energy (kcal) and proteins are the most important variables because deficits for these two nutrients usually accompany any deficits for other nutrients, and because energy-protein deficits are generally the most damanging. The gap analyses are generally carried out only for energy and proteins; other nutrients can be checked by examining the quality of the diet of a subsample of families. As long as families are deficient in kilocalories and proteins it is not appropriate to divert attention to deficiencies of other nutrients unless there exist obvious clinical signs that such deficiences are a problem in the community.

Estimating energy and protein requirements is a complex topic with many subtle issues which cannot be adequately discussed here. Nevertheless, there are some important considerations which should be kept in mind as the values in Table B-1 are applied to specific situations. For example, energy requirements are highly dependent on the level of activity of the individual. The energy requirements of a moderately active adult male is 3000 kilo calories, but it could range from 2700 kilocalories for slightly active to 4000 kilocalories for the exceptionally active. Likewise, an adult female's requirements could range from 2000 kcal for light activities to 3000 kcal for exceptional activity. It may be necessary to adjust the values in the table to conform with the activity levels of a particular community, keeping in mind that a low activity level may in fact be the result of deficient energy intakes.

In addition, there are two points to be considered when applying the recommendations for intake of protein. First, the values given in the table are based on "high quality" egg or milk protein. The quality of proteins actually consumed will depend on the typical diet for the specific region or country. The quality of diets in "rich countries" is generally equivalent to about 80% of egg or milk protein, while the quality of diets in "poor countries" may be equivalent to about 70% of egg or milk protein. There may be some communities where most of the protein comes from grains, in which case the quality may be as low as 60%. This can be taken into account by dividing the tabled protein requirements by the appropriate factor (normally .6 to .7 for most communities in developing countries).

The second point concerning the tabled protein requirements is that they are based on what the Joint FAO/WHO Expert Committee has defined as "safe levels of intake." The safe level of intake is designed to assure that the recommended requirements will encompass nearly all individuals (defined as two standard deviations beyond the norm which includes 97% of all individuals). Even after sex, age (size) and activity are taken into account, the average protein requirements will underestimate the needs of some individuals due to natural variation. The coefficient of variation has been established to be approximately 15%. In order to cover two standard deviations beyond the norm, the protein requirements listed in the table are actually 30% greater than the average requirements. While this "adjustment" may make sense when recommending protein consumption levels for individuals,

if the increased values are used for calculating the requirements of large groups (such as entire communities) then these requirements will tend to be 30% higher than may be necessary.

Once the decision is made as to what requirements are to be used for each nutrient to be analyzed (normally kilocalories and protein) then a computer data file is created with the information for each of the age/sex/physiological (pregnant or lactating) groupings. This data can be called upon to calculate the requirements of each individual for inclusion in the individual's secondary data file. The data can also be aggregated to determine the requirements of specific families or the community.

RECOMMENDED INTAKE OF NUTRIENTS

Age	Body weight	Energy (1)		Protein (1,2)	Vitamin A (3,4)	Vitamin D (5,6)	Thiamine (3)	Riboflavine (3)	Niacin (3)	Folic acid (5)	Vitamin B_{12} (5)	Ascorbic acid (5)	Calcium (7)	Iron (5,8)
	kilograms	*kilocalories*	*megajoules*	*grams*	*micrograms*	*micrograms*	*milligrams*	*milligrams*	*milligrams*	*micrograms*	*micrograms*	*milligrams*	*grams*	*milligrams*
Children														
< 1.....	7.3	820	3.4	14	300	10.0	0.3	0.5	5.4	60	0.3	20	0.5-0.6	5-10
1-3.....	13.4	1360	5.7	16	250	10.0	0.5	0.8	9.0	100	0.9	20	0.4-0.5	5-10
4-6.....	20.2	1830	7.6	20	300	10.0	0.7	1.1	12.1	100	1.5	20	0.4-0.5	5-10
7-9.....	28.1	2190	9.2	25	400	2.5	0.9	1.3	14.5	100	1.5	20	0.4-0.5	5-10
Male adolescents														
10-12.....	36.9	2600	10.9	30	575	2.5	1.0	1.6	17.2	100	2.0	20	0.6-0.7	5-10
13-15.....	51.3	2900	12.1	37	725	2.5	1.2	1.7	19.1	200	2.0	30	0.6-0.7	9-18
16-19.....	62.9	3070	12.8	38	750	2.5	1.2	1.8	20.3	200	2.0	30	0.5-0.6	5-9
Female adolescents														
10-12.....	38.0	2350	9.8	29	575	2.5	0.9	1.4	15.5	100	2.0	20	0.6-0.7	5-10
13-15.....	49.9	2490	10.4	31	725	2.5	1.0	1.5	16.4	200	2.0	30	0.6-0.7	12-24
16-19.....	54.4	2310	9.7	30	750	2.5	0.9	1.4	15.2	200	2.0	30	0.5-0.6	14-28
Adult man														
(moderately active).....	65.0	3000	12.6	37	750	2.5	1.2	1.8	19.8	200	2.0	30	0.4-0.5	5-9
Adult woman														
(moderately active).....	55.0	2200	9.2	29	750	2.5	0.9	1.3	14.5	200	2.0	30	0.4-0.5	14-28
Pregnancy														
(later Half)...		+350	+1.5	38	750	10.0	+0.1	+0.2	+2.3	400	3.0	30	1.0-1.2	(9)
Lactation														
(first 6 months months).....		+550	+2.3	46	1200	10.0	+0.2	+0.4	+3.7	300	2.5	30	1.0-1.2	(9)

(1) Energy and Protein Requirements. Report of a Joint FAO/WHO Expert Group, FAO, Rome, 1972. - (2) As egg or milk protein. - (3) Requirements of Vitamin A, Thiamine, Riboflavin and Niacin. Report of a Joint FAO/WHO Expert Group, FAO, Rome, 1965. - (4) As retinol. - (5) Requirements of Ascorbic Acid, Vitamin D, Vitamin B_{12}, Folate and Iron. Report of a Joint FAO/WHO Expert Group, FAO, Rome, 1970. - (6) As cholecalciferol. - (7) Calcium Requirements. Report of a FAO/WHO Expert Group, FAO, Rome, 1961. - (8) On each line the lower value applies when over 25 percent of calories in the diet come from animal foods, and the higher value when animal foods represent less than 10 percent of calories. - (9) For women whose iron intake throughout life has been at the level recommended in this table, the daily intake or iron during pregnancy and lactation should be the same as that recommended for nonpregnant, nonlactating women of childbearing age. For women whose iron status is not satisfactory at the beginning of pregnancy, the requirement is increased, and in the extreme situation of women with no iron stores, the requirement can probably not be met without supplementation.

SOURCE: Passmore, R. & others. Handbook on Human Nutritional Requirements. Rome, Italy: Food and Agriculture Organization of the United Nations, 1974. (FAO Nutritional Studies #28/WHO Monograph Series No. 61). P. 68-69.

Table B-1

SAMPLING PROCEDURES

Random Numbers

The Table of Random Numbers (Table C-1) is quite simple to use. At random, select some number in the matrix; then read the numbers from that point proceeding row-by-row or column-by-column. If the number of sampling units in the population is between 10 and 99 then the random numbers are read off two digits at a time; if the population is between 100 and 999 then the random numbers are taken three digits at a time.

To better understand this procedure, take the example of a community with a total of 83 houses and assume that a random sample of 10 houses is required. For this example, begin at the intersect of Column 9 and Row 13 (i.e., 6), and procede along the row by reading the digits two at a time; the random numbers would be: 67 (6-from Column 9, Row 13; 7-from Column 10, Row 13), 36, 11, 25, 07, 48, 55, 45, 91 (9-from Column 25, Row 13; 1-from Column 1, Row 14), 88, 44, 69. Note that twelve numbers are needed altogether; 91 and 88 cannot be used because there are only 83 houses in the sample.

Quasi-random Sampling

A sampling method which is often easier to apply in the field is known as quasi-random or systematic sampling. Essentially this consists of systematically proceeding through the community selecting every k^{th} house, where k is the sampling fraction equal to the number of houses in the sample divided by the number of houses in the community. This method could obviate the need to map and number the houses before taking the sample. As such, it becomes possible to enter the community and begin the survey process at the same time that the sample is being taken.

For this method to be practical, prior knowledge is needed of the approximate size of the community as well as the size of the sample which is needed. It is also helpful if all of the houses are located on streets, roads or on an identifiable network of paths in order to faciliate systematically working through the community without missing houses. If, for example, a community is thought to have about 400 families and a sample of 80 is needed, then every fifth house would be selected. Starting at one end

of the community, and arbitarily deciding to begin with the third house, a systematic procedure is established selecting the third, eighth, thirteenth, etc. houses.

This systematic sampling will tend to distribute the sample more evenly throughout the community than would normally occur in a purely random sample, thus assuring that all areas in the community are "represented." However, care must be taken when assuming equivalence to a simple random sample for the purpose of calculating significance of results.

Sample Size

It is difficult to establish an optimal size for the final sample because of the many factors which enter into the decision. Clearly the level of precision needed in the conclusions is a major factor.

From the statistical equation for determining sample size, $n = S(y)^2/S(\bar{y})^2$, we know that the sample size, n, is proportional to the square of the standard deviation, S(y), of the variable y for the population. Likewise, the sample size is inversely proportional to the square of the standard deviation of the mean $S(\bar{y})$, for the sample. The standard deviation of the mean -- also called the standard error of the mean -- is a measure of the probable precision or accuracy of an estimation of the value for a given sample size. The standard deviation S(y) is a characteristic of the population with respect to the variable y. Normally it is not known and must be estimated. A "first estimate" of the standard deviation of a particular variable in a population can be made based on previous knowledge about the variable and the population. This can later be verified by applying a "best estimate" formula to the data collected in the survey. Another possibility for estimating the standard deviation of the population is by carrying out a pre-test or pilot survey.

While it is helpful to keep in mind the above discussion, its usefulness in estimating the desirable sample size may be limited under actual field conditions. First, there may be too little experience or knowledge upon which to base the first estimate. A more important constraint is that the methodology calls for the quantification of not just one, but a number of variables. It may not be practical to estimate the standard deviation for each variable -- with the required precision needed for each -- in

order to select the sample size that would satisfy the constraints of each variable.

Table C-1

TABLE OF RANDOM NUMBERS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	5	8	3	8	4	1	7	2	1	8	9	6	0	2	4	8	1	6	7	8	9	3	6	5	9
2	8	1	8	5	9	3	6	8	7	0	8	6	0	8	7	1	7	4	9	1	0	4	0	5	8
3	2	3	5	6	7	9	8	3	6	8	9	6	8	3	7	5	2	1	6	6	9	4	9	2	7
4	7	5	7	7	2	7	5	2	9	4	7	4	5	1	2	1	3	3	4	6	0	5	6	5	6
5	1	5	2	4	6	1	3	0	4	0	1	6	5	7	6	9	6	9	4	7	3	7	7	7	6
6	7	0	4	1	6	4	0	6	5	5	6	9	1	3	8	3	9	3	3	4	4	7	7	3	3
7	1	0	8	2	1	6	5	0	2	1	9	0	6	3	7	2	9	1	9	1	3	5	2	4	0
8	7	2	6	4	7	4	1	0	8	1	0	2	7	9	6	4	9	1	4	7	3	9	2	4	7
9	6	9	5	9	7	2	0	9	5	5	0	2	1	5	6	7	5	2	9	5	5	9	9	6	7
10	3	3	9	7	2	9	5	7	2	4	3	0	4	0	2	0	1	6	6	3	6	3	2	5	3
11	1	7	2	5	9	8	1	3	0	9	8	0	5	5	2	9	3	6	3	9	8	8	4	3	2
12	1	1	6	7	7	6	5	5	8	8	6	6	3	3	8	6	1	7	0	2	6	2	7	6	2
13	0	5	1	9	4	3	0	2	6	7	3	6	1	1	2	5	0	7	4	8	5	5	4	5	9
14	1	8	8	4	4	6	9	4	2	7	8	7	0	1	3	9	0	5	6	6	4	9	3	7	3
15	4	1	9	8	3	8	0	9	5	1	9	8	7	0	8	0	2	6	9	6	6	6	0	6	9
16	9	3	6	5	9	6	6	9	8	8	3	8	9	5	0	2	4	5	7	3	5	6	6	5	7
17	7	0	7	5	9	8	9	2	8	1	0	9	7	2	1	7	2	6	5	7	1	1	9	8	3
18	9	7	6	9	0	2	2	8	2	4	4	7	7	1	9	5	1	0	8	6	5	3	6	4	3
19	0	0	2	9	6	1	0	1	4	3	1	2	2	3	8	4	9	6	5	5	9	5	8	0	6
20	3	3	8	8	1	5	5	0	7	4	0	3	5	8	2	6	7	6	1	1	3	6	6	3	5
21	3	4	8	3	2	1	6	0	4	7	7	9	8	2	0	6	3	1	0	8	3	0	8	0	6
22	5	8	5	6	7	7	4	1	4	3	1	0	5	8	4	8	4	2	9	3	8	8	6	3	8
23	7	0	8	2	3	2	4	4	9	9	7	9	2	6	1	8	9	1	8	7	2	2	3	4	8
24	7	4	9	9	5	2	9	4	2	9	6	5	5	3	7	1	0	5	2	0	6	3	7	1	9
25	8	8	5	9	7	2	5	0	0	9	8	2	4	0	0	3	7	1	7	9	7	3	4	8	6
26	6	1	7	6	3	3	0	6	2	7	2	9	1	6	7	6	4	4	5	5	7	5	2	2	3
27	4	5	8	1	2	6	8	2	7	2	3	8	2	7	6	2	7	8	6	6	4	3	1	0	1
28	3	0	7	1	0	1	1	8	5	0	0	5	7	8	0	9	5	3	3	1	7	3	1	2	0
29	5	0	1	7	9	2	5	4	5	7	5	5	5	4	4	0	5	1	0	3	1	0	8	9	7
30	0	6	4	0	1	2	7	7	8	0	5	0	5	4	7	6	4	9	9	3	8	5	0	7	0
31	3	6	5	2	5	8	6	1	1	2	1	6	5	3	7	4	3	5	6	5	3	0	8	4	2
32	1	5	1	5	6	8	9	3	7	8	8	2	6	0	6	2	7	2	8	0	5	5	3	2	4
33	1	7	8	3	6	3	5	0	9	2	7	1	2	6	7	1	4	8	9	5	9	9	4	3	6
34	0	5	9	8	7	9	0	4	7	5	2	6	9	6	6	4	9	5	3	1	9	2	8	2	1
35	8	9	4	0	0	5	8	5	0	1	5	2	8	9	8	9	1	8	4	8	6	0	1	9	4

FAMILY COMPOSITION SURVEY

Discussion

The Family Composition Survey is the first survey conducted and is very important since the information on the composition of the family becomes the basis for the other surveys. It is generally carried out at the time of initial contact with the family when the purpose of the surveys is explained and participation by the family is requested. For this reason it is important that the person responsible for this activity have a complete understanding of the entire project and an excellent facility for communicating with community members. If the family composition contains preschool-age children then the other surveys are scheduled during this first visit; if there are no preschool-age children then the Family Composition Survey is completed but no additional information is gathered.

This survey consists of listing all of the members of the family -- their names along with information on age, sex and physiological condition or family relationship. A two-digit individual code number is assigned to each family member; this same two-digit number is used throughout the other surveys to identify the individual.

The Family Composition Survey also assigns to each family a family code and a sector code. For the purpose of this discussion the family code will consist of a three-digit number and the sectoor code a four-digit number, although this can be adjusted to fit the specific situation. The sector code serves to identify the population from which the family has been selected. If there is more than one community in the diagnosis then the first one or two digits are dedicated to a code identifying the community; the remaining two or three digits can be used to identify the area or section within the community. The family code is a number unique to that particular family within the community and area identified by the sector code. The individual code, family code and sector code combine to form a unique nine-digit number which distinguishes an individual from any other individual in the diagnosis population.

A Family Composition Survey Form (Figure D-1) is included here along with instructions for completing it. It is a pre-coded form designed so that the data can be entered directly on console or onto computer punch cards.

Field Instructions

These are instructions for completing the Family Composition Survey Form (Figure D-1):

Card Type I

- Column 1 — contains the survey code, a number unique to the Family Composition Survey which distinguishes it from the other types of surveys.

- Columns 2-5 — contain the sector code to identify the population; if there is more than one community, the first one or two columns are dedicated to a community code; the remaining two or three columns can be used for a code identifying the area or section within the community.

- Columns 6-8 — contain the family's code -- a number unique to that particular family within the sample; together with the sector code it constitutes a unique identification number which distinguishes that family from all other families in all other communities. This allows data from various communities to be merged for one analysis.

- Columns 9-11 — contain the date the survey was taken; the first two columns contain a number from 01 to 12 indicating the month, and the last column contains the last digit of the year (0 if the year is 1980).

- Columns 12-13 — contain the card sequence number which gives the card order for the particular survey for that particular family; the Type I card will always have 01 in these columns.

- Columns 14-60 — contain the family's address or other descriptive phrase to identify the family's residence; it can be entered in alpha-numeric mode.

Card Type II

- Column 1 — contains the survey code, a number unique to the Family Composition Survey which distinguishes it from the other types of surveys.

- Columns 2-5 — contain the sector code to identify the population; if there is more than one community, then the first one or two columns are dedicated to a community code; the remaining two or three columns can be used for a code identifying the area or section within the community.

- Columns 6-8 — contain the family's code -- a number unique to that

particular family within the sample; together with the sector code it constitutes a unique identification number which distinguishes that family from all other families in all other communities. This allows data from various communities to be merged for one analysis.

- Columns 9-11 contain the date the survey was taken; the first two columns contain a number from 01 to 12 indicating the month, and the last column contains the last digit of the year (0 if the year is 1980).

- Columns 12-13 contain the card sequence number which gives the card order for that particular survey for that particular family; the first Type II card will be 02, the second will be 03, etc.

- Columns 14-15 contain the individual's code--a number unique to that particular individual which distinguishes him/her from other members of the family. There is one Type II card for each member of the family; the head of the household is usually listed first.

- Column 16 contains the category code which provides the following information: 1 = preschool child (0-72 months old); 2 = school-age child (6-14 years old); 3 = mother; 4 = father; 5 = pregnant woman; 6 = lactating woman; 7 = other adult woman; 8 = other adult man.

- Column 17 contains a code indicating the sex of the individual: male = 1; female = 2.

- Columns 18-19 contain the individual's age. If the individual is a preschooler the age is recorded in months (0-72); otherwise the age is recorded in years.

- Columns 20-60 contain the name of the individual recorded in alphanumeric mode.

Figure D-1

FAMILY COMPOSITION SURVEY FORM

1	Surv. No.
2	Sector Code
3	
4	
5	
6	Family Code
7	
8	
9	Month
10	
11	Year
12	Card Sequence
13	
14	Indiv. Code
15	
16	Category
17	Sex
18	Age
19	
20	Name
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

1	Surv. No.
2	Sector Code
3	
4	
5	
6	Family Code
7	
8	
9	Month
10	
11	Year
12	Card Sequence
13	
14	Address
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	

Sex: 1 = Male
2 = Female

Category: 1 = Preschool; 2 = School; 3 = Mother; 4 = Father;
5 = Pregnant Woman; 6 = Lactating Woman; 7 = Adult Woman; 8 = Adult Man

ANTHROPOMETRIC SURVEY

Discussion

The measurements involved in the Anthropometric Survey consist of three basic variables (weight, height and age) which are measured for preschool-age children (0-72 months of age). This survey can be implemented using relatively unskilled personnel, but it is very important that they be conscientious and that they take their task seriously. Generally, two-person teams are sufficient for collecting the data, although three persons may be more convenient in terms of organizing and transporting the equipment from house to house.

The equipment required is relatively simple but may present problems if it must be transported over rough terrain. For children weighing up to about 12 kilograms (approximately 2 years old), a fairly sensitive scale is needed. The ideal instrument is a standard baby scale (see Figure E-1), but is is a rather delicate instrument; if the babies are weighed at a central location (such as a health post) then it should be satisfactory. More commonly, however, the situation will require that the children be weighed in their homes -- often dispersed and hard to reach. In this case, a market-type hanging scale can be used in conjunction with a sling made from cloth such as a diaper (Figure E-2).

For children over two years old, a hanging type scale is needed which has a range of at least 26 kilograms. These older children can be supported from the scale by fabricating a sitting sling out of a pair of men's pants. For these heavier children it may be necessary to hang the scale from a sturdy support such as a tree limb.

Height is also measured differently for younger and older children. Children who are not able to stand erect on their own (generally less than 2 years old) should be measured in the prone position (i.e., flat on their backs). For the younger (shorter) children it is very important that their bodies be absolutely straight and that the measurement be taken from head to feet perpendicular to the body; a height meter (Figure E-3) should be used. This instrument can be fabricated from wood; it consists of a board with one fixed perpendicular end and another perpendicular end which slides back and forth. A Steel tape is attached to the board so that a pointer

on the sliding end indicates the height. One person steadies the child with its head flush against the fixed end while the other person holds the knees flat against the board, slides the end up against the bottoms of the feet, and reads the height from the steel tape. The major drawback to this method is that the height meter is heavy and awkward to carry if it is necessary to walk (or ride horses) over long distances in rough terrain.

For older children, height can be measured with a flexible plactic "seamstress" tape by having the child stand erect against a wall (Figure E-4). One person holds the end of the tape to the floor and steadies the child's feet and knees against the wall. The other person stretches the tape up the wall above the child's head and reads the height by placing a book or clip board level across the top of the child's head. If there is a flat perpendicular wooden surface such as a door jam, then the tape can be tacked to it in order to facilitate this process

Care must be taken to remove all of the clothing when the children are weighed. If this is forbidden then only a light pair of shorts should be permitted. The height should be measured when the child is barefoot.

The determination of the child's age will depend on what method will yield the most accurate and consistent results in a given cultural setting. It may be necessary to estimate the age based on number of years, seasons, months or days and then convert the total to the age in months. If at all possible, however, the child's birth date should be established and the computer should be left to calculate the age in months.

Since this survey involves directly recording measured values it is convenient to use precoded survey forms. The sample survey form included here (Figure E-6) is based on a maximum record length of 80 columns, on the assumption that the data will be entered on computer punch cards. This creates the potential need for more than one record (card) to include all of the data.

Two types of cards are used. The Type I card identifies the family and records the number of preschool children. This tells the computer how many Type II cards to look for -- one Type II card for up to four children, two Type II cards for five to eight children, etc. If the data were to be entered directly in magnetic storage form it would be possible to design records to be as long as necessary.

Field Instructions

These are instructions for completing the Anthropometric Survey Form (Figure E-6):

- Column 1 — contains a code unique to the Anthropometric Survey which distinguishes it from the other types of surveys.

- Columns 2-5 — contain the sector code to identify the population; if there is more than one community, then the first one or two columns are dedicated to a community code; the remaining two or three columns can be used for a code identifying the area or section within the community.

- Columns 6-8 — contain the family's code -- a number unique to that particular family within the sample; together with the sector code it constitutes a unique identification number which distinguishes that family from all other families in all other communities. This allows data from various communities to be merged for one analysis.

- Columns 9-11 — contain the date the survey was taken; the first two columns contain a number from 01 to 12 indicating the month, and the last column contains the last digit of the year (0 if the year is 1980).

- Columns 12-13 — contain the card sequence number which gives the card order for the particular survey for that particular family; the first card will have 01 in these columns; if there is a second card it will have 02 in these columns, etc.

- Columns 14-15 — contain the number of preschool (0-6 years old) children in the family; this number should also correspond to the number of children for whom anthropometric data are collected. These columns may be left blank after the first card.

- Columns 16-17 — contain the individual child's code which identifies him/her within the family; this number must be the same code the child was given in the Family Composition Survey.

- Column 18 — identifies the child's sex: male = 1; female = 2.

- Columns 19-20 — contain the child's age in months; if the child is a year and a half old than 18 is entered.

- Columns 21-23 — contain the child's birth date. The month is entered in the first two columns and the last digit of the year is entered in the last column (e.g., September 1976 = 096).

- Columns 24-27 — contain the child's weight in decagrams or hundredths of a kilogram (e.g., 6.5 kilograms = 0650).

- Columns 28-30 contain the child's height in centimeters (e.g., 1.5 meters = 150).

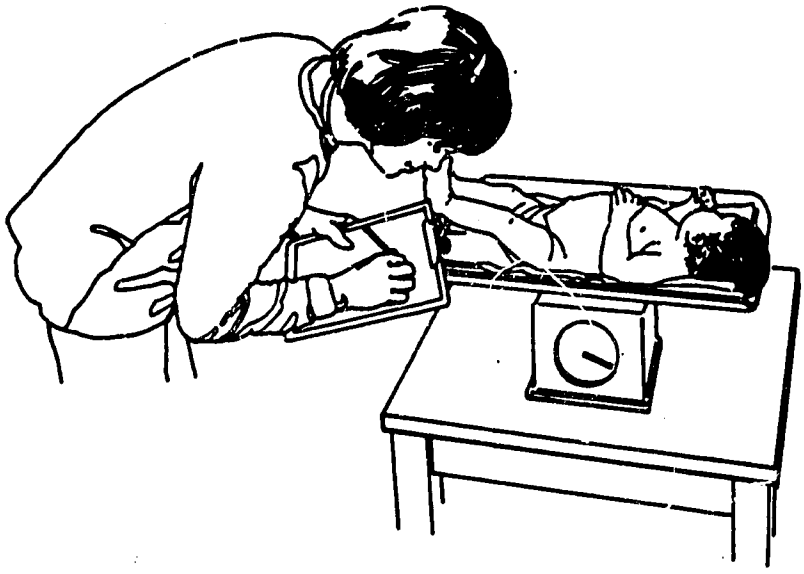
- Columns 31-45 contain data on the second child (if there are two or more preschoolers) in the same order and format as contained in columns 16-30 for the first child.

- Columns 46-60 contain data on the third child (if applicable) in the same order and format as columns 16-30.

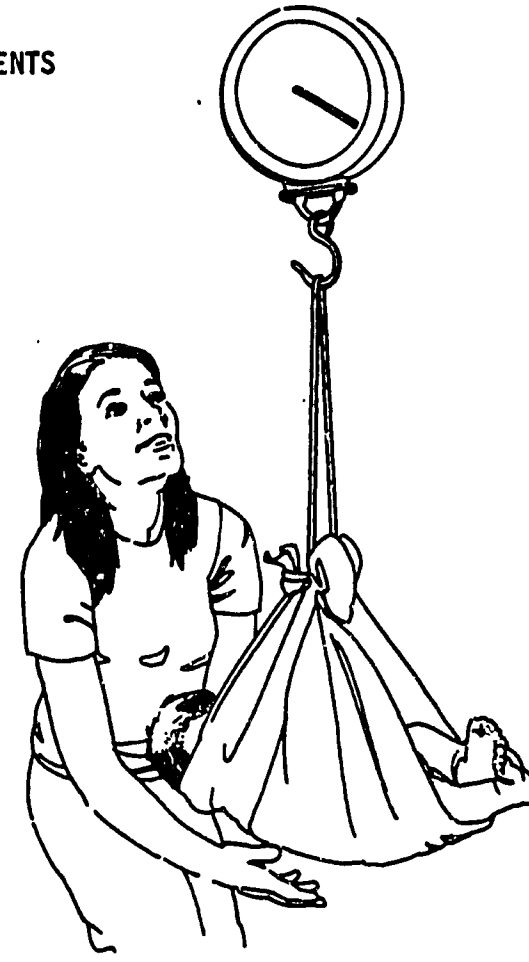
- Columns 61-75 contain data on the fourth child (if applicable) in the same order and format as columns 16-30.

Note: if there are more than four preschool children in the family then a second card (Card Sequence = 02) is filled out using the same format as the first card; if there are more than eight children in the family then a third card (Card Sequence = 03) is needed, etc.

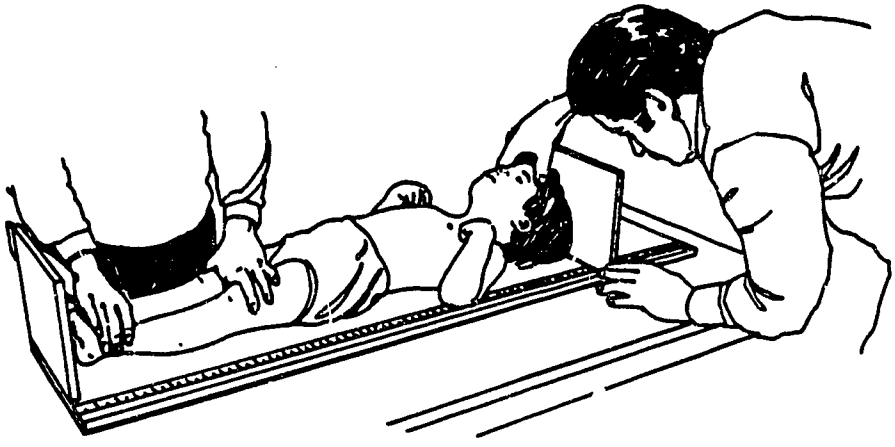
MEASURING AND WEIGHING INSTRUMENTS



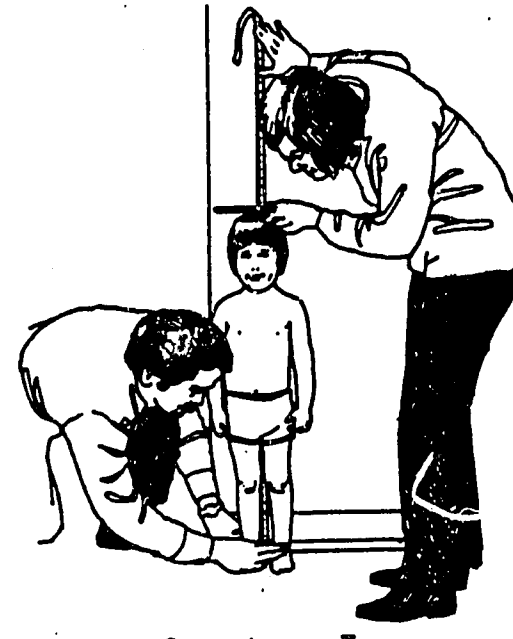
Standard Baby Scale
Figure E-1



Hanging Scale
Figure E-2



Height Meter
Figure E-3



Seamstress Tape
Figure E-4

Figure E-5

ANTHROPOMETRIC SURVEY FORM

			1			1	Survey Number
			2			2	Sector Code
			3			3	
			4			4	
			5			5	
			6			6	Family's Code
			7			7	
			8			8	
			9			9	Survey Date: Month
			10			10	
			11			11	Survey Date: Year
			12			12	Card Sequence
			13			13	
			14			14	Number of Children
			15			15	
	46		16		46	16	Individual's Code
	47		17		47	17	
	48		18		48	18	Sex
	49		19		49	19	Age in Months
	50		20		50	20	
	51		21		51	21	Birth Date: Month
	52		22		52	22	
	53		23		53	23	Birth Date: Year
	54		24		54	24	Weight in Decagrams
	55		25		55	25	
	56		26		57	26	
	57		27		57	27	
	58		28		58	28	Height in Centimeters
	59		29		59	29	
	60		30		60	30	
	61		31		61	31	Individual's Code
	62		32		62	32	
	63		33		63	33	Sex
	64		34		64	34	Age in Months
	65		35		65	35	
	66		36		66	36	Birth Date: Month
	67		37		67	37	
	68		38		68	38	Birth Date: Year
	69		39		69	39	Weight in Decagrams
	70		40		70	40	
	71		41		71	41	
	72		42		72	42	
	73		43		73	43	Height in Centimeters
	74		44		74	44	
	75		45		75	45	

24-HOUR RECALL CONSUMPTION SURVEY

Discussion

The 24-Hour Recall Consumption Survey is one of the most complex surveys in the diagnostic procedure. It requires that information be gathered, prior to the survey, on the different kinds of food consumed in the community as well as traditional preparations the families use to combine the different food ingredients. Generally this survey should only be undertaken by an experienced nutritionist who is trained in the techniques of consumption surveys and is familiar with the food habits of the region. It may be helpful to carry a small scale and graduated container in order to gain a more precise estimation of the quantity of food consumed.

The first step in designing the recall survey is to compose a list of all the food items found in the region. Each food item is assigned a unique two-digit code (assuming less than 100 items) to identify it in the computer analysis. Additional information for each food item must also be stored in the computer in order to perform certain basic calculations such as determining the nutrient content of the meals. Also, to facilitate computer programming and data manipulation it is necessary to provide for consistency in the units of measure.

The List of Foods included here (Table F-1) is made up of the food items used in the Colombia project. Column 1 contains the code and Column 2 contains the name of the food item. Column 3 contains the standardized unit of measure for reporting each food item. These units of measure correspond to what is most likely to be convenient for recording quantities of each of the foods consumed. For example, whole milk is a liquid measure and it is easiest to estimate individual servings in terms of cubic centimeters. On the other hand, meats and grains are usually measured and sold by the "pound"* and it is easiest to report these servings accordingly. Some items such as eggs, vegetables and fruit are most conveniently reported by the unit (e.g., 1 egg, 3 carrots, 1 banana). There also may be some items which are measured primarily by their value such as when bread is baked in different shapes and sizes but has a fairly consistent price-to-quantity ratio. In these cases the unit of measure is the value of the item, such as 50 cents worth of bread.

---

*In Colombia the term pound is used extensively to indicate a half of a kilogram, equal to 500 grams.

Column 4 contains a conversion factor for each food item. These are the constants which are needed to convert the different measurements units into grams. In some cases establishing this factor is fairly straightforward as in the case where the measurement unit is a weight unit (e.g., in the case of the 500 gram pound, the conversion factor is 500). When the measurement unit is a volume such as a cubic centimeter or a tablespoon it is necessary to determine the weight in grams of the particular measure for that particular food item (e.g., 1 cc of whole milk weighs 1g; 1 tablespoon of powdered milk weighs 5 g). When the food item is measured by the unit (such as one egg or two tomatoes) it is necessary to experimentally determine how much, on the average, a unit of each weighs. There may be some food items -- such as bread -- which are not normally measured by weight or volume and are therefore measured by value. For these items it is necessary to determine the number of grams in the amount of the food item which can be purchased in the community for a unit of money.

Columns 5 and 6 contain the percent of protein (grams of protein per gram of food) and the number of kilocalories per gram of the food item. These are the factors which, when multiplied by the quantity (in grams) of food consumed, will yield the amount of protein and kilocalories, respectively, consumed in that particular food item. These values (along with similar values for other nutrients such as vitamins and minerals) are determined by complex laboratory procedures but they can be obtained by consulting food value tables published for different regions of the world by the Food and Agriculture Organization (FAO) of the United Nations. Food value tables are also available for some individual countries; for example, tables established by the Colombian Department of Family Welfare were used for the work which was done in Cali.

Columns 7 and 8 contain additional factors needed for some of the food items. Column 7 contains the fraction of the food, as normally measured, which is edible. For example, if chicken is measured in pounds, it probably includes the bones which must be discounted before the proteins and calories are calculated. Also, since the values for proteins and calories for most foods are based on the weight after cooking, a factor is needed to convert the recorded (uncooked) weight to the weight that the food will have after it is cooked. Some items such as grains and beans will increase their

weight considerably when cooked while other items such as meats and some vegetables will lose weight. This change in weight in cooking is a result of a change in moisture content. The values that should be used for the "edible fraction" and the "percent change in cooking" can be found in the food value tables as part of the assumptions associated with the nutrient content of the foods.

The information in Columns 1, 4, 5, 6, 7 and 8 are filed in the computer in matrix form and are later called upon when quantities of specific foods are entered from the results of the 24-hour recall survey. For example, if the family ate a 2.5 pound chicken then the code (05) and the quantity in pounds (2.5) would be entered in the computer. The quantity in pounds is multiplied by the appropriate conversion factor (500) to yield the quantity in grams (1250). This is then multiplied by the edible fraction (0.60) and adjusted by the appropriate percent change in cooking (-0.39 to yield a net weight of 457.5 grams. The net weight is then multiplied by the percent of proteins (20.2) and the kilocalories per gram (1.78) to yield 92.4 grams of protein and 814 kilocalories in the chicken consumed by the family. Similar calculations are made for the other food items consumed by the family; the proteins and kilocalories are totaled to determine the nutrient intake for the 24-hour recall period.

Once the list of food items is complete, a second list is compiled of the different food recipes prepared by the families in the region. Table F-2 is an example of the preparations which are common in Colombia. The purpose of this list is to experimentally determine -- based on the results of the 24-hour recall survey -- how the various food items are combined and which items are consumed together. This information on eating habits is useful to determine the overall quality of the diet and to identify ways in which the eating habits could be improved (e.g., by substituting or adding more nutritious ingredient(s) to certain recipes).

The actual survey form (Figure F-1) requires considerable skill to solicit complete and unbiased responses. This form is not precoded since it is easier to concentrate on recording the responses as they are given and assigning the appropriate codes at a later time. Thus, there are two sets of instructions included with the sample survey form; one set explains how to complete the survey form in the field and the other set explains how to

transfer the data to standard 80-column coding sheets.

Since one 24-hour period may not provide an accurate picture of the average consumption of a family, it may be necessary to perform the survey several times in order to more closely approximate the normal situation. The survey date will separate the different rounds and distinguish whether it is the first, second or third time that the family has been surveyed. Also, since there are likely to be more than two meals for a particular survey it will be necessary to use more than one sheet. If two sheets are needed the first should be labeled "1 of 2" and the second "2 of 2."

The 24-hour recall survey should be pretested before being executed as part of the diagnostic methodology. This will help familiarize the surveyor with the form and the types of responses, and will help to determine if the list of foods and the list of preparations are reasonably complete. Even with the pretest, however, it may be necessary to add foods and preparations to the lists (as well as other modifications) after the survey is in progress.

Field Instructions

These are instructions for completing the 24-Hour Recall Consumption Survey Form (Figure F-1):

Sector Code - This is the same four-digit sector code assigned to the family on the Family Composition Survey; it identifies the community and area within the community.

Family Code - This is the same code assigned to the family on the Family Composition Survey; it is the unique number which identifies the family.

Date - The month and year of the survey is recorded; the first two boxes contain the month (e.g., February = 02) and the third box contains the last digit of the year (e.g., 1980 = 0).

Type of Meal - A code is entered in this box to represent the meal for which the survey is recorded. The codes are: Breakfast = 1; Lunch = 2; Dinner = 3; Between meal snacks = 4.

Family Non-Participants - This refers to family members who did not participate in the consumption of the meal. The age (years) and sex (male = 1; female = 2) of each non-participating family member is recorded in one of the sets of three boxes (e.g., if a nine year old boy did not eat that particular meal, then 091 would be entered). This information will make it possible to subtract the nutrient requirements of those

non-participants from the total requirements of the family in order to calculate the actual nutrient gap for that particular meal.

Non-Family Participants - This refers to persons who participate in the meal but who are not members of the family and hence do not normally eat with the family. The age and sex of each non-family participant is recorded in one of the sets of three boxes. This information will make it possible to add the nutrient requirements of those non-family participants to the total requirements of the family in order to calculate the actual nutrient gap for that particular meal.

Preparation/Food/Quantity - Under Preparation is listed each of the food preparations or recipes. It is repeated for each basic food item which it contains. Opposite the preparation name is recorded the basic food item and its quantity. A preparation of lentil soup would be recorded as follows:

lentil soup	lentils	200 grams
lentil soup	tomatoes	3 units
lentil soup	onions	2 units
lentil soup	salt	1 tablespoon

If a basic food item is eaten separately (e.g., bread) then it is listed in both the preparation column and the food column, and the amount (e.g., 5 units) is listed in the quantity column:

bread	bread	5 units

Note: in some cases a single preparation may be served for more than one meal. For example, the lentil soup may be prepared in the morning and served at lunch and at dinner. In this case the preparation should be listed under both meals but for the second meal the quantities should be recorded as zero. This allows for a description of the eating patterns but does not result in double accounting of the nutrients consumed in the 24-hour period.

Coding Instructions

These are instructions for transferring the data from the 24-Hour Recall Consumption Survey Form (Figure F-1) to standard 80-column coding sheets:

Card Type I

- Column 1 contains the survey code, a number unique to the Family Consumption Survey which distinguishes it from the other types of surveys.
- Columns 2-5 contain the sector code, transferred directly from the survey form.
- Columns 6-8 contain the family code, transferred directly from the survey form.

- Columns 9-11 contain the date the survey was taken, transferred directly from the survey form.

- Columns 12-13 contain the card sequence number which gives the card order for the particular survey for that particular family; the Type I card will always have a 01 in these columns.

- Column 14 contains the number of meals the family consumed during the previous day and for which data exists; this information is used to signal the computer how many sets of Type II and Type III cards to look for.

Card Type II

- Columns 1-11 contain the same information recorded in Columns 1-11 on the Type I card.

- Columns 12-13 contain the card sequence number which gives the card order for the particular survey for that particular family; the first Type II card will be 02 but the second Type II card (for the second meal) will depend on how many Type III cards there are corresponding to the first meal.

- Column 14 contains the meal code indicating which meal data is being recorded, based on the following codes: 1 = breakfast; 2 = lunch; 3 = dinner; 4 = between meal snack. This number is transferred directly from the survey form.

- Columns 15-16 contain the number of food items which made up that particular meal; this information is used to signal the computer how many Type III cards to look for (one if there are less than nine items; two if there are more than eight but less than seventeen, etc.) before looking for the Type II card corresponding to the next meal.

- Column 17 contains the number of family non-participants for that particular meal; this information is used to signal the computer how many three-digit age/sex data sets to look for before it begins looking for non-participants.

- Column 18 contains the number of non-family participants for that particular meal; this information is used to signal the computer how many three-digit age/sex data sets to look for following the family non-participant data sets.

- Columns 19-80 contain the age and sex information for the family non-participants and the non-family participants listed consecutively; if there are two family non-participants, their age/sex data is recorded in Columns 19-21 and 22-24; if there is also one non-family participant, that data is recorded in columns 25-27 and the remaining columns are blank.

Card Type III

- Columns 1-11 contain the same information recorded in Columns 1-11 on the Type I and II cards.

- Columns 12-13 contain the card sequence number which is the number immediately following the card sequence number for the Type II card corresponding to the same survey, family, round and meal; if it is the first meal then the card sequence would be 03.

- Columns 14-15 contain the preparation code for the first food item listed for that particular meal; if the item is not an ingredient in a larger preparation (i.e., consumed separately) then 00 is entered.

- Columns 16-17 contain the food code number for the first food item listed for that particular meal.

- Columns 18-21 contain the quantity consumed of the first food item listed; the quantity must be recorded in the units specified for that particular food item and must be right-hand justified within the four columns.

- Columns 22-29 contain the preparation code, food code and quantity for the second food item recorded in the same format as was the first food item recorded in Columns 14-21

- Columns 30-37 contain the preparation code, food code and quantity for the third through eighth food items, recorded in the same format as Columns 14-21.

Note: If there are more than eight food items in that particular meal then an additional eight can be recorded in a similar fashion on a second Type III card. This second Type III card would have a card sequence number of 04 if it were the first meal in the first round. In this case the card sequence would be 05 for the Type II card for the second meal.

Table F-1

LIST OF FOODS

1 Code	2 Food Name	3 Unit of Measure*	4 Conversion Factor	5 Percent Protein	6 Kilocalories per Gram	7 Edible Fraction	8 % Change in Cooking
1	Whole Milk	mililiters (c.c.)	1.00	3.4	0.60	1.00	0.00
2	Powdered Milk	tablespoon	5.00	25.2	4.07	1.00	0.00
3	Cheese	pound	500.00	25.0	3.87	1.00	0.00
4	Meat	pound	500.00	18.7	2.32	1.00	-39.00
5	Chicken	pound	500.00	20.2	1.78	0.60	-39.00
6	Fish	pound	500.00	17.9	1.01	0.50	-38.00
7	Organs (e.g. liver)	pound	500.00	15.2	1.43	1.00	-38.00
8	Eggs	unit (egg)	45.00	12.8	1.63	0.90	0.00
9	Green Peas	pound	500.00	8.2	1.16	0.40	101.00
10	Dry Peas	pound	500.00	23.9	3.08	1.00	101.00
11	Beans	pound	500.00	20.4	3.02	1.00	101.00
12	Lentils	pound	500.00	23.5	3.15	1.00	172.00
13	Soybeans	pound	500.00	34.00	3.66	1.00	101.00
14	Garbanzo Beans	pound	500.00	19.3	3.39	1.00	101.00
15	Protein Mix	tablespoon	15.00	23.2	3.40	1.00	0.00
16	Duryea	tablespoon	17.00	28.0	3.45	1.00	0.00
17	Carrots	unit	90.00	0.7	0.36	0.85	-24.00
18	Melon	unit	2000.00	0.9	0.39	0.65	-22.73
19	Spinach	unit of $1	232.50	3.5	0.27	0.60	0.00
20	Tomato	unit	240.00	0.9	0.17	0.80	0.00
21	Cabbage	unit	800.00	2.2	0.24	0.85	0.00
22	Lettuce	unit (head)	250.00	1.1	0.13	0.55	0.00
23	Lima Bean	pound	500.00	2.1	0.29	0.90	0.00
24	Sugar Beat	unit	170.00	1.4	0.42	0.80	0.00

*Note: In Colombia the term pound is used extensively to indicate a half of a kilogram, equal to 500 grams.

Table F-1 (continued)

LIST OF FOODS

1 Code	2 Food Name	3 Unit of Measure	4 Conversion Factor	5 Percent Protein	6 Kilocalories per Gram	7 Edible Fraction	8 % Change in Cooking
25	Ullucos (fruit)	pound	500.00	1.5	0.58	0.95	0.00
26	Guayaba (fruit)	unit	40.00	0.9	0.36	0.75	0.00
27	Mango (fruit)	unit	140.00	0.5	0.58	0.60	0.00
28	Oranges	unit	220.00	0.7	0.35	0.70	0.00
29	Bananas	unit	140.00	1.2	0.84	1.00	0.00
30	Rice	pound	500.00	7.8	3.59	1.00	154.00
31	Oats	pound	500.00	11.0	3.48	1.00	0.00
32	Barley	pound	500.00	10.2	3.11	1.00	0.00
33	Corn	pound	500.00	8.7	3.55	1.00	161.00
34	Corn on Cob	pound	120.00	4.7	1.36	0.60	0.00
35	Wheat Flour	pound	500.00	14.5	3.43	1.00	0.00
36	Corn Flour	pound	500.00	9.1	3.22	1.00	0.00
37	Corn Flour Biscuit	unit	90.00	3.4	1.46	1.00	161.00
38	Cassava Bread	unit	50.00	10.4	2.88	1.00	0.00
39	Bread	unit of $1	50.00	9.0	3.37	1.00	0.00
40	Fortified Noodles	pound	500.00	10.7	3.50	1.00	300.00
41	Corn Starch	tablespoon	12.00	0.6	3.47	1.00	0.00
42	Arracacha	unit	100.00	1.0	0.94	0.85	-13.00
43	Potato	unit	120.00	1.9	0.91	1.00	-8.30
44	Cassava	unit of $1	235.00	0.8	1.46	0.80	-8.30
45	Plantains	unit	200.00	1.2	1.42	0.65	-32.00
46	Refined Sugar	pound	500.00	0.0	3.84	1.00	0.00
47	Unrefined Sugar	unit	583.00	0.5	3.12	1.00	0.00
48	Chocolate	Unit (cube)	8.00	3.8	4.41	1.00	0.00

Table F-1 (continued)

LIST OF FOODS

1 Code	2 Food Name	3 Unit of Measure	4 Conversion Factor	5 Percent Protein	6 Kilocalories per Gram	7 Edible Fraction	8 % Change in Cooking
49	Vegetable Oil	tablespoon	10.00	0.0	8.84	1.00	0.00
50	Lard	tablespoon	15.00	0.0	8.92	1.00	0.00
51	Instant Coffee	tablespoon	10.00	0.3	0.05	1.00	0.00
52	Ground Coffee	tablespoon	10.00	1.0	0.30	1.00	0.00
53	Soft Drinks	bottle	270.00	0.0	0.44	1.00	0.00
54	Onion	unit	120.00	1.2	0.26	0.45	0.00
55	Salt	tablespoon	50.00	0.0	0.0	1.00	0.00
56	Bone	pound	500.00	0.0	0.0	0.00	0.00
57	Citron	unit	900.00	1.5	0.38	0.00	0.00
58	Tortillas	unit of $1	190.00	2.8	1.22	1.00	0.00
59	Infant Formula	tablespoon	15.00	2.0	3.09	1.00	0.00
60	Breast Milk	feedings	120.00	1.1	0.71	1.00	0.00
61	Bullion Cubes	cubes	15.00	2.8	0.18	1.00	0.00
62	Berries	unit of $1	100.00	0.6	0.23	0.90	0.00
63	Pineapple	unit	1400.00	0.4	0.51	0.55	0.00
64	Avocado	unit	360.00	1.6	1.27	0.60	0.00
65	Apples	unit	120.00	1.1	0.46	0.30	0.00
66	Mellons	unit	2300.00	0.4	0.12	0.40	0.00
67	Starch	tablespoon	12.00	0.2	3.56	1.00	0.00
68	Lulos (fruit)	unit	70.00	0.6	0.23	0.60	0.00
69	Tree Tomato	unit	70.00	1.4	0.30	0.60	0.00
70	Cucumber	unit	150.00	0.5	0.09	0.75	0.00

Table F-2

LIST OF PREPARATIONS

Code	Preparation Name	Code	Preparation Name
00	Other preparation	27	Noodle soup
01	Boiled meat	28	Spaghetti or noodles
02	Fried meat	29	Lentils
03	Broiled meat	30	Lentil soup
04	Meat stew	31	Fried cassava
05	Mazamorra (Morocho)	32	Boiled cassava
06	Corn pie	33	Cassava soup
07	Arepa	34	Chicken stew
08	Corn soup	35	Cuchuco
09	Steamed rice	36	Sugar water
10	Rice soup	37	Chocolate
11	Rice drink	38	Oat drink
12	Boiled potato	39	Fried fish
13	Fried potato	40	Fish soup
14	Potato soup	41	Boiled fish
15	Vegetable stew	42	Ullucos
16	Baked plantain	43	Tortilla soup
17	Fried plantain	44	Pea dish
18	Plantain drink	45	Cob corn dish
19	Bean dish	46	Coffee (milk sugar)
20	Benor soup	47	Barley dish
21	Fried egg	48	Mixed soup
22	Boiled egg	49	Barley Soup
23	Scrambled eggs	50	Locro
24	Wheat flour pie	51	Wheat soup
25	Wheat flour soup	52	Barley drink
26	Wheat drink		

24-HOUR RECALL CONSUMPTION SURVEY FORM

Page ___ of ___

Sector Code ☐☐☐☐ Family Code ☐☐☐ Date ☐☐☐

Type of Meal ☐	Age	Sex				Type of Meal ☐	Age	Sex			
Family Non-Participants						Family Non-Participants					
Non-Family Participants						Non-Family Participants					

Preparation	Food	Quantity	Preparation	Food	Quantity

Figure F-1

FAMILY MARKET SURVEY

Discussion

Information is collected in the Family Market Survey for two main objectives: (1) to determine the amount of nutrients which were available to a family for consumption over a specified period, and (2) to obtain additional information on the conditions under which the food was acquired.

In many cultures there exist specific patterns of cyclic food purchases, each food having its own cycle. This survey is intended to discover the length of the cycle for each food and to measure the amount of food purchased (or made available) in that cycle. In poorer families there is usually very little variety in food acquisitions and marketing becomes a very automatic process; thus it should be fairly easy for the person who does the buying to recall the amounts of each item acquired during the previous 7-day period. In many situations the Family Market Survey can be expected to provide more accurate estimates of family level nutrient consumption than will the 24-Hour Recall Consumption Survey.

Like the 24-hour recall survey, the Family Market Survey requires a certain amount of advance preparation to organize a list of foods with codes and units of measure together with local market locations and buying patterns. The same List of Foods that was prepared for the 24-hour recall survey (Appendix F, Table F-1) can also be used for this survey since the food items should be the same; the only difference for the market survey are some of the units of measure because for some items the most convenient unit of measure for preparing food will not be the same as the most convenient unit for purchasing the food. For example, a tablespoon is a convenient unit for measuring powdered milk in preparing it for consumption, but it would not be purchased by the tablespoon. It is common that some foods come prepacked in different size containers. When this is the case (i.e., when the unit of measure at purchase is different from that given in Column 3 of Table F-1) the size (in grams) of the package must be noted on the survey form. The other information in the 24-hour recall list of foods (food codes, protein and kilocalorie content, edible fraction and percent change in cooking) can also be used in the Family Market Survey without modification.

In addition to the basic information on the various food items, it will

also be necessary to assign codes for: (1) reflecting the frequency of the purchase; (2) the source or location where the purchase was made; and (3) the preference or reason for buying at that location. There will likely be situations where some items are purchased almost on a daily basis but not every day; in these cases, it will be necessary to note on the survey form the number of days during the previous 7-day period.

Before codes can be assigned to the purchase locations, the types of stores and their locations must be classified. One useful distinction is whether the markets lie outside the community being diagnosed (e.g., in a nearby city or other regional center of commerce); another distinction may be between the local traditional market and the local stores. It may also be useful to classify the local stores into two or more groups such as small retail and larger retail/wholesale (see discussion on page 30; also Appendix L). In addition, provision must be made for including food which is not purchased such as that produced by the family or received as donation.

The survey should also seek to determine why the family patronizes a specific location as its principal market source. This will provide insight into why some families purchase food under less favorable economic conditions which could help to explain problems of interfamily distribution of nutrients. For example, some families may purchase most of their food at a small, more expensive store because they are able to obtain credit or because of greater distances to less expensive sources. This information will be useful later if attempts are made to change purchasing patterns or if food subsidies are to be implemented.

An attempt is also made in this market survey to determine the resources (in the form of income) available for acquiring food. It will then be possible to calculate the portion of family income allocated to food -- an important indicator of the family's flexibility to improve its nutrient intake by adjusting its buying habits for non-food items. Moreover, this information will allow for testing the hypothesis that the families with the lowest incomes are the ones with the highest nutrient deficit. Note, however, that in many cultures income is a sensitive (private) issue and care must be taken in designing the question. This particular survey distinguishes between two types of income: (1) that which family members earn on a regular basis, and (2) that which the family earns on an occasional or seasonal

basis.

The trained nutritionist who carries out the 24-Hour Recall Consumption Survey should also do the Family Market Survey at the same time. Because this survey uses may different codes, the survey form (Figure G-1) is not precoded. Separate instructions are included for the field survey process and for transferring the data to standard, 80-column coding sheets.

Field Instructions

These are instructions for completing the Family Market Survey Form (Figure G-1):

Sector Code - This is the same four-digit sector code assigned to the family on the Family Composition Survey; it identifies the community and area within the community.

Family Code - This is the same three-digit code assigned to the family on the Family Composition Survey; it is the unique number which identifies the family.

Date - The month and year of the survey are recorded in these three boxes; the first two boxes contain the month (e.g., February = 02) and the third box contains the last digit of the year (e.g., 1980 = 0).

Market - These items correspond to the primary marketing activity of the family. Normally a family will regularly shop once or twice a week (perhaps on "market days") and then buy other items more or less frequently. In the space provided, the frequency, source and preference (reason for using that source) are recorded. Also recorded is the individual code (from the Family Composition Survey) of the person in the family who does the shopping.

Income - First, the amount and frequency (e.g., $30.00/week) of regular income is recorded for the family head, the spouse and other members of the family (or other regular sources). If the income is a daily rate, then the number of days per week must be noted. Next, the amount and frequency (e.g., $225, three times a year) is recorded for farm income and for other occasional or seasonal income.

Food Item - In this column is recorded the name of each food item acquired during the recall period (previous seven or fourteen days). Each item is only listed once although it may have been acquired a number of times during the period.

Source - This column identifies the location where that particular food item was purchased. If it was purchased several times at different places during the period then the "usual" or "normal" source should be recorded.

Frequency - In this column is recorded the frequency with which the food item was purchased during the period.

Days - If the item was purchased almost daily then the number of days in the week (e.g., 5 days out of 7) should be noted in this column. If the frequency is two or less times per week then this column can be left blank.

Quantity - This column contains the average amount that was purchased each time during the period. If rice were purchased biweekly, once 2 kilograms and once 3 kilograms, then "2.5 kilograms" would be entered. Note that both the quantity and the unit of measure must be recorded.

Unit/package - If, in the previous column, the measure is a package (e.g., 1 bag of refined sugar) where the weight is not obvious then the weight (in grams) should be recorded in this column.

Cost - This column contains the unit cost of the food item where the unit is the same as the unit of measure used in the "Quantity" column.

Coding Instructions

These are instructions for transferring the data from the Family Market Survey Form (Figure G-1) to standard 80-column coding sheets:

Card Type I

- Column 1 contains the survey code, a number unique to the Family Market Survey which distinguishes it from the other surveys.

- Columns 2-5 contain the sector code, transferred directly from the survey form.

- Columns 6-8 contain the family code, transferred directly from the survey form.

- Columns 9-11 contain the date the survey was taken, transferred directly from the survey form.

- Columns 12-13 contain the card sequence number giving the card order for the particular survey for that particular family; the Type I card will always have 01 in these columns.

- Column 14 contains the code for the frequency of the family's "regular" market activity as recorded on the upper part of the survey form; example frequency codes are: 1 = daily; 2 = semi-weekly; 3 = weekly; 4 = biweekly; and 5 = monthly.

- Column 15 contains the code for the source or location of the family's "regular" market activity; example source codes are: 1 = outside commercial center; 2 = local traditional market; 3 = local small store; 4 = local large store; 5 = home produced; and 6 = donated.

- Column 16 contains the code indicating preference for the "regular" market source; example preference codes are: 1 = credit available; 2 = less expensive; 3 = close to home (i.e., convenient); 4 = better quality; 5 = better selection; and 6 = other reasons.

- Column 17 contains the individual code for the person who performs the "regular" market shopping; the code is recorded in the upper part of the survey form and corresponds to the person's individual code in the Family Composition Survey.

- Columns 18-20 contain the weekly income (no decimals) of the family head; this information is found on the upper part of the survey form but may have to be converted to a weekly figure (e.g., if the survey figure is a daily rate, it is multiplied by the number of working days in the week; if it is a monthly rate it is divided by four).

- Columns 21-23 contain the weekly income of the spouse, as transferred or calculated from the survey form.

- Columns 24-26 contain space for recording the weekly income of an additional family member.

- Columns 27-29 contain space for recording the weekly income from yet another family member or other income source.

- Columns 30-31 contain the number of different food items required during the period and listed on the survey form; this information is used to signal the computer how many Type II cards to look for.

Card Type II

- Columns 1-11 contain the same information recorded in Columns 1-11 on the Type I card.

- Columns 12-13 contain the card sequence number giving the card order for the particular survey for that particular family; the first Type II card will have 02 in these columns; if a second Type II card is needed it will be 03, etc.

- Columns 14-15 contain the code for the first food item listed on the survey form; the code number for that food item is found in Appendix F, Table F-1.

- Column 16 contains the code for the source of the first food item as recorded in the first row, second column of the survey form; example source codes are the same as those listed for Column 15 of the Type I card.

- Column 17 contains the code for frequency with which the first food

item was acquired. The information is found in the first row, third column of the survey form; the sample frequency codes are the same as those listed for Column 14 of the Type I card.

- Column 18 contains the number of days in the week if the previous frequency was weekly (i.e., Column 17 = 3); otherwise this column can be left blank. The information is transferred from the first row, fourth column of the survey form.

- Columns 19-22 contain the quantity of the first food item; the information is found in the first row, fifth column of the survey form.

- Columns 23-26 are left blank if the quantity in the previous columns are in the same units of measurement as found in Column 3 of Table F-1 for the particular food item; if this is not the case, then a conversion factor must be entered for converting to grams the quantity in the previous columns. This information, if needed, is found in the first row, sixth column of the survey form.

- Columns 27-29 contain the unit cost of the first food item; this information is found in the first row, seventh column of the survey form.

- Columns 30-45 contain the codes for the food item, source, frequency, # days, quantity, grams/package and cost for the second food item (second row of the survey form); this data is recorded in the same format as the first food item in Columns 14-29.

- Columns 46-61 contain the codes for the food item, source, frequency, # days, quantity, grams/package and cost for the third food item (third row of the survey form); this data is recorded in the same format as the first food item in Columns 14-29.

- Columns 62-77 contain the codes for the food item, source, frequency, # days, quantity, grams/package and cost for the fourth food item (fourth row of the survey form); this data is recorded in the same format as the first food item in Columns 14-29.

Note: The same information for the fifth through eighth food items is recorded in the same manner on a second Type II card. This is continued until all the food items have been coded.

Figure G-1

FAMILY MARKET SURVEY FORM

Sector Code [ ][ ][ ][ ] Family Code [ ][ ][ ] Date [ ][ ][ ]

Market:

frequency _____ source _____ preference _____ person _____

Income:

family head _____ spouse _____ other _____ other _____

farm income _____ other occasional income _____

Food Item	Source	Frequency	#Days	Quantity	Grams/package	Cost

HEALTH STATUS SURVEY

Discussion

This survey consists of a questionnaire designed to quantify a few health indicator variables. The survey is conducted by a person who should be literate, conscientious, and have the appropriate personality and attitude for encouraging open and honest answers from the respondents. No special equipment is required.

Data are gathered only for preschool age (less than six years old) members of the family. This is done because usually younger children are most susceptible to illness and because the symptoms of health problems are more pronounced in young children, thus making the measurement task easier. Furthermore, this is the same age group for which the malnourishment indicator (Anthropometric Survey) is measured; therefore, it is possible to correlate the data for two of the most important variables: health status and malnourishment.

Information on health status should be obtained from the mother or other adult family member who is responsible for the care of the preschool child(ren). It is recommended that the questionnaire encompass only the previous 7-day period. That is, the mother is asked to recall the various aspects of the child's health during the preceding seven days. Definition of a longer recall period will decrease the accuracy of the measure because of a tendency not to remember specific occurrences when they are fairly common to begin with.

Diarrhea is considered to be the "best" health indicator from the standpoint of measuring the impact of health problems on the nutritional status of a community. Most of the information gathered in this survey relates to this variable. Presence of diarrhea is defined to be more than three liquid or watery stools per day. Care should be taken in explaining to the mother this definition of diarrhea in an attempt to ensure data consistency from one family to the next.

For each preschool child, information is obtained on the number of days (if any) that she/he had diarrhea. If there was diarrhea, the average number of stools per day is noted along with the presence of any blood or mucus in the stools. Also noted are the number of days (if any) in which the

child had: (1) spells of vomiting, (2) fever, or (3) cold symptoms such as coughing and congested sinus.

The survey is designed to provide data on:

(1) the morbidity of diarrhea: the ratio of sick-to-well children;

(2) the average prevalence of diarrhea during the 7-day period: the number od diarrhea-days divided by the total number of children-days; and

(3) the severity of each episode of diarrhea.

The concepts of morbidity and prevalence, as used here, can best be understood by considering an example where: five children are surveyed and one is found to have diarrhea six out of the seven days; one is found to have diarrhea four out of seven days; and the other three children had no diarrhea during the 7-day period. In this case:

- <u>morbidity</u> is 2 ÷ 5 = 40%; and
- average <u>prevalence</u> is (6+4) ÷ (5 x 7) = 10 ÷ 35 = 29%

Normally, of course, the morbidity and average prevalence would be calculated for a much larger group of children such as the sample for the entire community.

Severity of a particular episode is determined by the frequency of stools per day (e.g., more than eight stools per day) and/or the presence of blood and/or mucus.

The information which is gathered in the Health Status Survey can be recorded directly on a precoded questionnaire form. The survey form included here (Figure H-1) is based on maximum record length of 80 columns in case the data are to be entered on computer punch cards; this necessitates more than one card in the event that there are more than eight preschool children in the family. If the data are to be entered directly onto a magnetic form then this limitation does not exist and the ninth, tenth, etc. children can be entered without repeating the basic data in Columns 1-15.

Field Instructions

These are instructions for completing the Health Status Survey Form (Figure H-1):

- Column 1 contains the survey code, a number unique to the Health Status Survey which distinguishes it from the other types of surveys.

- Columns 2-5 contain the sector code to identify the population; if there is more than one community, then the first one or two columns are dedicated to a code which identifies the community and the remaining two or three columns can be used for a code identifying the area or section within the community.

- Columns 6-8 contain the family's code -- a number unique to that particular family within the sample; together with the sector code it constitutes a unique identification number which distinguishes that family from all other families in all other communities. This allows data from various communities to be merged for one analysis.

- Columns 9-11 contain the date the survey was taken; the first two columns contain a number from 01 to 12 indicating the month, and the last column contains the last digit of the year (0 if the year is 1980).

- Columns 12-13 contain the card sequence number which gives the card order for the particular survey for that particular family; the first card will have 01 in these columns; if there is a second card it will have 02 in these columns, etc.

- Columns 14-15 contain the number of preschool (0-6 years old) children in the family; this number should also correspond to the number of children for whom health status is collected. These columns may be left blank after the first card.

- Columns 16-17 contain the individual child's code to identify him/her within the family; this number must be the same code that the child was given in the Family Composition Survey.

- Column 18 contains the number of days (if any) during the preceding seven days in which that child suffered from diarrhea.

- Column 19 contains the average number of stools per day during those days that the child had diarrhea.

- Column 20 records whether or not there was blood and/or mucus present in the diarrhea; neither blood nor mucus = 0; blood only = 1; mucus only = 2; and blood and mucus = 3.

- Column 21 contains the number of days (if any) during the preceding seven days in which that child had bouts of vomiting.

- Column 22 contains the number of days (if any) during the preceding seven days in which that child had a fever.

- Column 23 contains the number of days (if any) during the preceding seven days in which that child suffered from cold symptoms such as congestion and coughing.

- Columns 24-31 contain data on the second child (if there are two or more preschoolers) in the same order and format as contained in columns 16-23 for the first child.

- Columns 32-39 contain data on the third child (if applicable) in the same order and format as contained in columns 16-23 for the first child.

- Columns 40-47 contain data on the fourth child (if applicable).

- Columns 48-55 contain data on the fifth child (if applicable).

- Columns 56-63 contain data on the sixth child (if applicable).

- Columns 64-71 contain data on the seventh child (if applicable).

- Columns 72-79 contain data on the eighth child (if applicable).

Note: If there are more than eight preschool children in the family then a second card (Card Sequence = 02) is completed using the same format as the first card.

Figure H-1

HEALTH STATUS SURVEY FORM

			1			1	Survey Number
			2			2	Sector Code
			3			3	
			4			4	
			5			5	
			6			6	Family's Code
			7			7	
			8			8	
			9			9	Month / Survey Date
			10			10	
			11			11	Year
			12			12	Card Sequence
			13			13	
			14			14	Number of Children
			15			15	
	48		16		48	16	Individual's Code
	49		17		49	17	
	50		18		50	18	Days of Diarrhea
	51		19		51	19	Stools per Day
	52		20		52	20	Blood/Mucus
	53		21		53	21	Days with Vomiting
	54		22		54	22	Days with Fever
	55		23		55	23	Days with Cold Symptoms
	56		24		56	24	Individual's Code
	57		25		57	25	
	58		26		58	26	Days of Diarrhea
	59		27		59	27	Stools per Day
	60		28		60	28	Blood/Mucus
	61		29		61	29	Days with Vomiting
	62		30		62	30	Days with Fever
	63		31		63	31	Days with Cold Symptoms
	64		32		64	32	Individual's Code
	65		33		65	33	
	66		34		66	34	Days of Diarrhea
	67		35		67	35	Stools per Day
	68		36		68	36	Blood/Mucus
	69		37		69	37	Days with Vomiting
	70		38		70	38	Days with Fever
	71		39		71	39	Days with Cold Symptoms
	72		40		72	40	Individual's Code
	73		41		73	41	
	74		42		74	42	Days of Diarrhea
	75		43		75	43	Stools per Day
	76		44		76	44	Blood/Mucus
	77		45		77	45	Days with Vomiting
	78		46		78	46	Days with Fever
	79		47		79	47	Days with Cold Symptoms

ENVIRONMENTAL SANITATION SURVEY

Discussion

It is widely acknowledged that a strong relationship exists between the transmission of infectious disease and a community's environment. However, the specific variables involved and the dynamics of the relationship are not well understood. Nevertheless, there are several variables included in this survey which are thought to represent the more important components in the relationship; these variables are: (1) water quantity, (2) water quality and (3) human waste disposal.

Often it is not practical to directly measure water quantity and quality; surrogate indicators must be used for the survey. Theory states that the quantity of water is important because of its ability to clean the environment, thus diluting the concentration of pathogens by increasing washing of: hands and body, food, cooking and eating utensils, floors, etc. Since in most situations it will not be possible to measure the quantity of water being used, the convenience of the water source serves as a surrogate indicator of the dilution function.

Factors which determine the convenience of the water supply are associated primarily with location in relation to point of use. Also relevent is whether or not the water is dispensed under pressure at a valve of faucet since this requires less time and effort than drawing water with a hand pump. The most convenient situation is one in which the water is piped under pressure into the house with a faucet in the kitchen and perhaps elsewhere, such as in a bathroom. The next most convenient situation is one in which there is a pressurized faucet or a well in the patio or yard close to the house. If neither of these situations exist, convenience will then be a function of the distance to the water source as measured by the time needed to walk to it.

The quality of the water source is of primary importance with respect to the water which is directly ingested, although it may also have lesser importance in terms of residual contamination from washing hands or food. In some situations it may be feasible to take samples of water and have them analyzed in a laboratory or tested in the field using a portable system (e.g., a Millipore which measures coloform). Even if one of these

methods is possible, costs may prohibit performing the tests on sample from each of the families.

It is possible, however, to obtain an approximation of the quality simply by noting the source of the water supply and the method of transmitting the water to the point of use. For example, if the water is pumped from a deep well and transmitted to the house in a closed-pipe system, then it is probable that the water is not contaminated. On the other hand, if water taken from a river is transported to the house and dispensed in open containers, it is likely to be highly contaminated unless it is boiled or treated before ingestion. While this empirical approach will not result in a cardinal scale of water quality values, it will be possible to associate each family with a group on an ordinal scale from one to ten, where the group with a value of one would have a very risky water supply and those with a value of ten would have a very secure water supply. The basic information with respect to the source of water supply is recorded directly on the survey form and the classification scheme is applied later as part of the data processing.

The disposal of human wastes can be classified into one of three basic categories: (1) sewer connection; (2) latrine; or (3) open-air. Included in the first category would be those families who are connected to a common or public disposal system, or who have private septic tank systems. The principal characteristics of this first category are that it includes the use of a "water seal" toilet, and that the human wastes and waste water are transported in closed pipes away from the living unit and disposed of underground or away from a population concentration. If families are to be placed in this category it should be ascertained that the system is functioning properly according to sound sanitary engineering principles.

The actual effect on the transmission of disease for the latrine and open-air categories is also relative to the specific situation. For example, if the latrine is not kept clean with the pit covered, then much of its advantage is lost, especially if it is located close to where food is prepared. Likewise, the usually negative effects of waste disposal in the open air (fields, etc.) are mitigated considerably if population density is very low and the houses are quite dispersed.

For the purpose of this survey no further information is gathered beyond distinguishing among the three categories of human waste disposal. However, if it is determined that infectious disease is a major component of a particular community's nutrition problem, then it will be necessary to design and carry out a more detailed investigation of the environmental factors which affect the transmission of disease. Such a detailed investigation should also gather data on the kinds and frequencies of social contacts among the families of a community (e.g., at schools, restaurants, etc.), since this is an important mechanism in transmitting infection within a community.

The Environmental Sanitation Survey can be recorded directly on a precoded survey form such as the sample form included here (Figure I-1).

Field Instructions

These are instructions for completing the Environmental Sanitation Survey Form (Figure I-1):

- Column 1 — contains a code unique to the Environmental Sanitation Survey which distinguishes it from other surveys.
- Columns 2-5 — contain the sector code to identify the population; if there is more than one community, then the first one or two columns are dedicated to a code which identifies the community; the remaining two or three columns can be used for a code identifying the area or section within the community.
- Columns 6-8 — contain the family's code -- a number unique to that particular family within the sample; together with the sector code it represents a unique identification number which distinguishes that family from all other families in all other communities. This allows data from various communities to be merged for one analysis.
- Columns 9-11 — contain the date the survey was taken; the first two columns contain a number from 01 to 12 indicating the month and the last column contains the last digit of the year (0 if the year is 1980).
- Column 12 — indicates the number of spigots or faucets (if any) inside the house or kitchen area.
- Column 13 — indicates the number of spigots or faucets (if any) in the yard or patio area.

- Column 14 indicates the number of showers (if any) which the family has.

- Columns 15-16 indicate the distance to the water supply measured in minutes needed to bring water to the house; if the source of water is in the house or yard then 00 is placed in these columns.

- Column 17 indicates the source of water supply for general use such as washing clothes, dishes, floors, etc.; the codes to be used for different kinds of sources are listed at the bottom of the survey form (Figure I-1).

- Column 18 indicates the source of the water supply used for direct consumption such as drinking and food preparation; the same codes at the bottom of the form are used to identify the source.

- Column 19 indicates if the consumption water supply receives some form of treatment; the codes for the different kinds of water treatment are also listed at the bottom of the survey form.

- Column 20 indicates the kind of system used for the disposal of human wastes; the codes for the different kinds of sewage disposal are also listed at the bottom of the survey form.

- Columns 21-23 contain a "crowding" indicator by noting the approximate distance in meters to the nearest neighboring house; if it is more than a kilometer then 999 is entered in the columns.

- Columns 24-28 contain space for entering results of water quality tests (if applicable); the units of entry will depend on the specific kind of test.

Figure I-1

ENVIRONMENTAL SANITATION SURVEY FORM

	1	Survey Number	
	2	Sector Code	
	3		
	4		
	5		
	6	Family's Code	
	7		
	8		
	9	Month	Survey Date
	10		
	11	Year	
	12	Number Spigots in House	
	13	Number Spigots in Yard	
	14	Number of Showers	
	15	Minutes to Water Supply	
	16		
	17	Water Supply - General	
	18	Water Supply - Consumption	
	19	Treatment of Water	
	20	Sewage Disposal	
	21	Distance to Closest Neighbor	
	22		
	23		
	24	Water Quality	
	25		
	26		
	27		
	28		

Water Supply Codes:

1 = Piped, pressurized supply, private
2 = Piped, pressurized supply, public
3 = Closed well, private
4 = Closed well, public
5 = Open well, private
6 = Open well, public
7 = Stored rainwater
8 = River, surface water

Sewage Disposal Codes:

1 = Sewer or septic tank
2 = Latrine
3 = Open air

Water Treatment Codes:

1 = Boiled water
2 = Filtered water
3 = Chemical treatment

INDIVIDUAL CONSUMPTION SURVEY

Discussion

The Individual Consumption Survey is used to evaluate the adequacy of the distribution of nutrients among the individual members of a family. An index of relative adequacy is calculated for each person according to the following formla: I = (Ci/Ri)/(Cf/R/f), where I= the index; Ci= nutrients consumed by the individual; Ri= the nutrients required by the individual; CF= the nutrients consumed by the entire family; and Rf = the nutrients required by the family.

If the index is greater than unity then that individual is receiving relatively more food than the rest of the family; if the index is less than unity, he/she is receiving relatively less. Although this is calculated for the individual, a similar index can be calculated for age/sex groups within the community by averaging the indexes for all individuals in each group.

This particular survey is the most difficult in terms of the inconvenience it imposes on the family because it involves actually weighing the food after the portions have been served for the individual family members. Care must be taken to make this procedure as simple and innocuous as possible. A scale is needed which will weigh quantities up to a kilogram, and which has the ability to adjust for the tare of the container (plate, bowl or cup) holding the food. It is also useful to have a supply of light, disposable (paper or plastic) containers. The portions can be served and weighed in these containers, then transferred to the individuals' plates.

When the dish being served contains several different food components (as would be the case with a stew) there are two ways to procede. The more accurate method is to separate and weigh each ingredient for each of the individual portions. This is time consuming and may not be acceptable to the family, in which case the separate weighing of each ingredient can be performed on one serving and the assumption (not necessarily valid) can be made that the rest of the servings contain the same amount of each ingredient in proportion to its total weight.

The survey form consists of a two-way matrix with the food items listed down one axis and the individual family members listed across the other axis.

There is space for twenty-four food items and ten individuals; if this is not sufficient then a second sheet is used. The individuals are identified by their respective codes taken from the Family Compostion Survey; the food items are the same as those listed in Table F-1, Appendix F. The food codes, percent protein and kilocalories per gram are taken from Table F-1 for use in the computer calculations; the other information in the table is not needed since the food is weighed directly in grams and is ready to eat at the time of weighing. The survey form is not precoded; it is filled out in the field and the data are transferred to coding forms back at the office.

Normally a family is only surveyed for one meal during a particular day; hopefully, this will be sufficient and other families can be surveyed for the other meals. Bear in mind that daily intake for one family will not be calculated; the calculation will be performed for a group of families--some with breakfast data, some with lunch data and some with dinner data. If circumstances necessitate data from all meals for each family then the date of survey will serve to separate the data from one meal to the next.

Field Instructions

These are instructions for completing the Individual Consumption Survey Form (Figure J-1):

Sector Code - This is the same four-digit sector code assigned to the family on the Family Composition Survey; it identifies the community and area within the community.

Family Code - This is the same code assigned to the family on the Family Compostion Survey; it is the unique number which identifies the family.

Date - The month and year of the survey is recorded; the first two boxes contain the month (e.g., February = 02) and the third box contains the last digit of the year (e.g., 1980 = 0).

Meal Code - A code is recorded in this box which indicates the type of meal for which the information is being collected: 1 = breakfast; 2 = lunch; and 3 = dinner.

Individual's Code - In each two-box set is recorded the individual code (from the Family Composition Survey) for each of the family members who participated in the meal.

Food Item - All of the food items which make up the meal are listed in this column; the food items should be the same as those in Table F-1.

Food Weight - Across from each food item and under each individual code is recorded the weight in grams of that food which was served to that

individual. One digit is placed in each box; if the portion weighs 150 grams then 0-1-5-0 is entered.

Coding Instructions

These are instructions for transferring the data from the Individual Consumption Survey Form (Figure J-1) to standard 80-column coding sheets:

Card Type I

- Column 1 — contains the survey code, a number unique to the Individual Consumption Survey which distinguishes it from the other types of surveys.
- Columns 2-5 — contain the sector code, transferred directly from the survey form.
- Columns 6-8 — contain the family code, transferred directly from the survey form.
- Columns 9-11 — contain the date the survey was taken, transferred directly from the survey form.
- Columns 12-13 — contain the card sequence number which gives the card order for the particular survey for that particular family; the Type I card will always have 01 in these columns.
- Column 14 — contains the meal code indicating the type of meal which was surveyed; the information is transferred directly from the survey form.
- Columns 15-16 — contain the total number of food items which were served and are listed on the survey form; this tells the computer how many Type II cards to look for.
- Columns 17-80 — contain the individual codes (in ascending order) of each family member participating in the meal; e.g., if the family has five members (according to the Family Composition Survey and individual number four was not present at the meal, then columns 17-24 would contain 01, 02, 03, 05, and the remaining columns would be blank. This tells the computer how many four-digit food weights to look for on each Type II card.

Card Type II

- Columns 1-11 — contain the same information recorded in columns 1-11 on the Type I card.
- Columns 12-13 — contain the card sequence number; the first Type II card will be 02, the second 03, etc.

- Columns 14-15 contain the code (from Table F-1) for the food items whose weights are recorded on that particular card; there is a separate Type II card for each food item which was served in the meal and is listed on the survey form.

- Columns 16-19 contain the weight in grams of that food item which was served to the first individual (the individual whose code appears in columns 17-18 on the Type I card).

- Columns 20-23 contain the weight in grams of that food item which was served to the second individual (the individual whose code appears in columns 19-20 on the Type I card).

Note: Food weights of the third, fourth, fifth, etc. individual are recorded in columns 24-27, 28-31, 32-35, etc., respectively, for each individual participating in the meal. On the next Type II card the weights of the next food item are recorded for each individual in the same format as the first Type II card.

INDIVIDUAL CONSUMPTION SURVEY FORM

Sector Code ☐☐☐☐ Family Code ☐☐☐ Date ☐☐☐ Meal ☐

Food Item	Individual Code/Food Weight in Grams

Figure J-1

AGRICULTURAL PRODUCTION SURVEY

Discussion

The objective of this survey is to determine (1) the amount of land which is available for agriculture, (2) the kinds of crops which are being produced, and (3) the amounts which are being produced and the proportion of that which is sold. The survey is a recall-type questionnaire where the person responsible for the agriculture activities provides the information. The data on production and land extension can be considered a first estimate and, if more precise data is needed, then it can be verified using other techniques mentioned in Chapter IV (page 29).

The survey is carried out on the same sample as the other surveys but in fact need only be completed for those families which own or control land within the community which is large enough for some agricultural activity. That is, the survey need not be completed for families who just have a house and a yard. If this does not yield a large enough sample, it may be necessary to draw an additional or complementary sample which better represents the agriculture sector.

The preliminary work which is needed before carrying out the survey consists of defining codes for the different types of crops and the different locations where the crops are sold. The list of crops will include some of the same items which appear on the List of Foods (Table F-1, Appendix F); it will not include the processed foods. Additional items must be added to include non-food crops or crops which become food only after processing. Nutrient content must be established for each item in order to calculate the gaps for the production component. The categories for place of sale should distinguish whether the produce is sold within the community or outside. It may be of interest to include other information such as whether or not the buyer receives the produce as payment on a loan since this would help define the relationships between buyer and seller.

The survey form which is included here is precoded; it is filled out directly in the field.

Field Instructions

These are instructions for completing the Agricultural Production Survey Form (Figure K-1):

Card Type I

- Column 1 — contains the survey code, a number unique to the Agricultural Production Survey which distinguishes it from the other types of surveys.

- Columns 2-5 — contain the sector code to identify the population; if there is more than one community, the first one or two columns are dedicated to a community code; the remaining two or three columns can be used for a code identifying the area or section within the community.

- Columns 6-8 — contain the family's code -- a number unique to that particular family within the sample; together with the sector code it constitutes a unique identification number which distinguishes that family from all other families in all other communities. This allows data from various communities to be merged for one analysis.

- Columns 9-11 — contain the date the survey was taken; the first two columns contain a number from 01 to 12 indicating the month, and the last column contains the last digit of the year (0 if the year is 1980).

- Columns 12-13 — contain the card sequence number which gives the card order for the particular survey for that particular family; the Type I card will always have 01 in these columns.

- Column 14 — contains the number of plots which the family has and for which data will be collected; a plot is defined to be an area of land with a single crop or set of mixed crops.

- Columns 15-18 — contain the total size of land area recorded in tenths of a hectare.

- Columns 19-21 — contain the amount of land (also recorded in tenths of a hectare) which is used as yard or area around the house.

- Columns 22-24 — contain the amount of fallow land (recorded in tenths of a hectare); this is the land which, during the period of measurement; was idle or not in production.

Card Type II

- Columns 1-11 — contain the same information, recorded in the same format, as in columns 1-11 on the Type I card.

- Columns 12-13 — contain the card sequence number; the first Type II card (data on the first plot) is 02.

- Column 14 — contains the plot code -- a unique number which distinguishes that particular plot from other plots of the family; the first plot is 1 etc.

- Columns 15-17 contain the size of that plot recorded in tenths of a hectare.

- Column 18 contains the number of different crops produced on that plot (1 if a single crop, 2 or more if mixed crops.) This tells the computer how many crops to look for.

- Columns 19-20 contain the crop code to identify which crop was grown on the plot; the code is taken from the List of Crops which was discussed above.

- Columns 21-24 contain the amount produced during the year measured in units which are common for the particular crop.

- Columns 25-28 contain a conversion factor to change whatever local units were used into kilograms; kilograms will be the standard unit of measure for calculations on the production data.

- Columns 29-31 contain the percent of that crop which was sold and not retained for use by the family.

- Column 32 contains a code which indicates the kind of place where the crop was sold (e.g., 1 = local, no credit repayment; 2 = local, for credit repayment; 3 = outside community, no credit repayment; 4 = outside community, for credit repayment).

- Columns 33-46 contain the data for the second crop (if applicable) produced on that plot; this is the same information as columns 19-32 contain for the first crop, recorded in the same format.

- Columns 47-60 contain the data for the third crop (if applicable) recorded in the same format as the first crop.

- Columns 61-67 contain the data for the fourth crop (if applicable) recorded in the same format.

Note: If the family has a second plot under production then a second Type II card (card sequence 03) is completed with the same kind of information in the same format as the first Type II card for the first plot; if the family has a third plot in production then a third Type II card is completed, etc.

Figure K-1

AGRICULTURAL PRODUCTION SURVEY FORM

Card Type II

		1				1	Survey Code	
		2				2	Sector Code	
		3				3		
		4				4		
		5				5		
		6				6	Family Code	
		7				7		
		8				8		
		9				9	Month	Date
		10				10		
		11				11	Year	
		12				12	Card Sequence	
		13				13		
		14				14	Plot Code	
		15				15	Hectares ÷ 10	
		16				16		
		17				17		
		18				18	No. of Crops	
47		19		47		19	Crop Code	
48		20		48		20		
49		21		49		21	Units Produced	
50		22		50		22		
51		23		51		23		
52		24		52		24		
53		25		53		25	Kilograms/Unit	
54		26		54		26		
55		27		55		27		
56		28		56		28		
57		29		57		29	Percent Sold	
58		30		58		30		
59		31		59		31		
60		32		60		32	Place Sold	
61		33		61		33	Crop Code	
62		34		62		34		
63		35		63		35	Units Produced	
64		36		64		36		
65		37		65		37		
66		38		66		38		
67		39		67		39	Kilograms/Unit	
68		40		68		40		
69		41		69		41		
70		42		70		42		
71		43		71		43	Percent Sold	
72		44		72		44		
73		45		73		45		
74		46		74		46	Place Sold	

Card Type I

	1	Survey Code	
	2	Sector Code	
	3		
	4		
	5		
	6	Family Code	
	7		
	8		
	9	Month	Date
	10		
	11	Year	
	12	Card Sequence	
	13		
	14	No. of Plots	
	15	Total Land Hectares ÷ 10	
	16		
	17		
	18		
	19	Yard Hectares ÷ 10	
	20		
	21		
	22	Fallow Land Hectares ÷ 10	
	23		
	24		

COMMUNITY MARKET SURVEY

Discussion

The purpose of this survey is to identify and quantify the flow of important food items through the community's market system. It attempts to determine the portion of business which corresponds to the various stores and what the main characteristics are for carrying out their food related commercial activities.

The first step in undertaking the Community Market Survey is to locate all of the stores or persons involved in selling food items within the community. These are then divided into groups based on a functional classification (such as was discussed in the Income and Marketing Section of Chapter IV-page 30). For the purpose of this survey we have defined three types of stores or enterprises: (1) large wholesale/retail stores which import most of the food in the community and then sell it to other stores as well as to the general public; (2) small retail stores which buy from the larger stores and sell to the general public; and (3) enterprises (either in a specific location or mobile) which sell only a few specialty items such as milk, eggs, honey, etc.

Once all of the stores have been located and placed in one of the classifications then a decision is made whether to survey all of the stores or only a sample of them. If there are a large number of the small retail stores it is probably sufficient to only survey a sample of them while surveying all of the stores in the other two classifications.

The survey is a recall-type where the proprietor is asked to report which food items he sells in the store. The food items and their respective codes can be transferred from Table F-1, Appendix F. It may not be necessary to survey all of the items in the store which are on the Food List (Table F-1); rather, a sub-list can be defined containing only those food items which are nutritionally significant in the community's diet as indicated in the 24-Hour Recall Consumption Survey (Appendix F) or Family Market Survey (Appendix G). The proprietor is asked to recall the average quantity, frequency, source and price of each item based on the most recent purchases (e.g., during the preceding month).

Since the survey is precoded it will be necessary to determine in advance the categories and codes for the frequency and source responses. The frequencies of purchases are similar to those in the Family Market Survey but also provide for longer range frequencies of up to six months. The source categories should indicate whether the food was acquired within the community's boundaries or outside, and whether it was purchased from a wholesaler or directly from a producer. The actual codes are included in the instructions.

Field Instructions

The following are instructions for completing the Community Market Survey Form (Figure L-1):

- Column 1 — contains the survey code, a number unique to the Community Market Survey which distinguishes it from the other types of surveys.

- Columns 2-5 — contain the same sector code used in the family surveys; if there is more than one community, the first one or two columns are dedicated to a community code; the remaining two or three columns can be used for a code identifying the area or section within the community.

- Column 6 — contains a code to identify the type of store: large wholesale/retail store = 1; small retail store = 2; and specialty store or enterprise = 3.

- Columns 7-8 — contain the store code, a number unique to that particular store which distinguishes it from other stores of the same type within the community. The store type, together with the store code, make up the store identification number which is analogous to the family identification number.

- Columns 9-11 — contain the date the survey was taken; the first two columns contain a number from 01 to 12 indicating the month, and the last column contains the last digit of the year (0 if the year is 1980).

- Columns 12-13 — contain the card sequence number. The first card is 01; if there are more than four food items a second card, 02, will be needed; if more than eight items, a third card, 03, etc.

- Columns 14-15 — contain the total number of food items reported on the survey form; this tells the computer how many to look for. After the first card these columns can be left blank.

- Columns 16-17 contain the code (from Table F-1) of the first food item.

- Column 18 contains the average frequency with which the store purchases the first food item: 1 = daily; 2 = semi-weekly; 3 = weekly; 4 = biweekly; 5 = monthly; 6 = every two months; 7 = every three months; 8 = every four months; and 9 = every six months.

- Column 19 contains the usual source where the store purchases the first food item: 1 = wholesaler within the community; 2 = producer within community; 3 = wholesaler outside community; and 4 = producer outside community.

- Columns 20-22 contain the average quantity (number of units) which the store purchases per time period indicated in Column 18.

- Columns 23-26 contain the weight in kilograms of the unit which was used in Columns 20-22.

- Columns 27-30 contain the average price for the unit used in Columns 20-22.

- Columns 31-45 contain the code for the second food item together with information on its frequency, source, quantity, weight and price; this data is recorded in the same format as Columns 16-30.

- Columns 46-60 contain the code for the third food item together with information on its frequency, source, quantity, weight and price; this data is recorded in the same format as Columns 16-30.

- Columns 61-75 contain the code for the fourth food item together with information on its frequency, source, quantity, weight and price; this data is recorded in the same format as Columns 16-30.

Note: The fifth through eighth food items are recorded on the second card in the same format as the first through fourth food items are recorded on the first card; the ninth through twelfth food items are recorded on the third card, etc.

Figure L-1

COMMUNITY MARKET SURVEY FORM

	1	Survey Number
	2	Sector Code
	3	
	4	
	5	
	6	Store Type
	7	Store Code
	8	
	9	Month / Date
	10	
	11	Year / Date
	12	Card Sequence
	13	
	14	Number of Items
	15	
46	16	Food Code
47	17	
48	18	Frequency
49	19	Source
50	20	Quantity
51	21	
52	22	
53	23	Kilograms/Unit
54	24	
55	25	
56	26	
57	27	Price/Unit
58	28	
59	29	
60	30	
61	31	Food Code
62	32	
63	33	Frequency
64	34	Source
65	35	Quantity
66	36	
67	37	
68	38	Kilograms/Unit
69	39	
70	40	
71	41	
72	42	Price/Unit
73	43	
74	44	
75	45	

ERROR ANALYSIS

The following discussion provides more detail on the potential effects of error resulting from the methodology. The broadest definition of error is used in this discussion: it includes anything which produces a value for a variable which is different from the actual value. This can include a wide range of communication problems; for example, problems of interpretation or logic within the measurement procedure, and a variety of different transcription problems.

Consideration of a specific component of the diagnostic methodology will help to understand the range of possibilities where the measured value can vary from the actual value of the variable in question. The concept of family-level nutrient gap (requirements less consumption; see discussion, Appendix B) will be considered since it is central to the diagnostic procedure: it indicates, for example, whether the nutrition problem is primarily attributable to food availability issues, such as production and among-family distribution, or whether the problem is more likely to be related to problems of health status resulting in increased requirements.

The first step in determining the gap is to establish the theoretical nutrient requirements of the family under the assumption of good health. No practical method exists for measuring the actual nutrient requirements of individuals in the field, so it is necessary to deduce a value for the requirements based on a statistical treatment of the problem. This is done by taking the values from the FAO/WHO recommendations for specific classifications of age, sex and physical activity (Table B-1, Appendix B).

Since the actual value of requirements varies among individuals (even within a specific classification), the use of an average value will result in overestimating or underestimating the actual value for many individuals. For a particular individual this difference could be quite significant, but since it is a random error the difference diminishes when individuals are considered as a group. Nevertheless, it is quite possible that for a given family group of 5 or 6 individuals, the requirement values deduced from Table B-1 will be significantly different than the actual requirements needed to nourish that family under conditions of good health.

Once the potential error in the requirements has been reviewed, the next

step is to consider the consumption aspect of the nutrient gap variable. It is in the measurement of the family-level consumption that there exists a wide range of possible ways that the measured value can differ from the actual value. This can be demonstrated by considering the 24-hour recall survey as a measure of family-level consumption. The following are examples of the ways that the measured value can differ from the actual value:

- The person responding to the survey deliberately falsifies responses due to some preconceived perception of the type of response that "should" be given.
- The person surveyed responds incorrectly because he/she does not know or remember the correct value.
- The person conducting the survey incorrectly interprets the response, perhaps due to problems with measurement units (e.g., a "cup" may not be 8 ounces).
- Whenever the response is manually transferred from one form to another (e.g., from survey form to coding form) there is the potential for some transcription error.
- Even when the response is made, received and recorded correctly, there is still the possibility that the value obtained is different than the actual value of the consumption variable; this is true to the extent that the family's consumption during the 24-hour period surveyed is not the same as their average or normal consumption.

Similar types of errors can enter into each of the surveys. The potential for error may be unavoidable; it is not the result of carelessness on the part of the investigator but, rather, it is inherent in the nature of measuring socially and culturally determined phenomena. While it may not be possible to avoid much of the error, it is very important to acknowledge its existence. This acknowledgement should be explicit -- perhaps in the form of an error model to identify the types of potential error in the different components and to predict the likely impact these errors will have on the conclusions drawn from the diagnosis.

The first step in an error analysis is to predict the various places

where error might enter into each of the diagnostic components. Next, means must be established for detecting the existence of errors at the various points. It is possible to speak of two different methods for error detection: (1) evaluation based on an extrinsic standard, and (2) evaluation which is intrinsic to the measurement methodology.

Error detection based on an extrinsic standard is possible when certain results contradict any "logical" explanation. This type of error detection is similar to the process of "cleaning" data which was described briefly in Chapter V on Data Management (page 36).

An example of extrinsic error detection occurs when a frequency graph of children's age demonstrates inordinate "peaks" at six months, twelve months, twenty-four months, thirty-six months, etc. The existence of a disproportionately high number of children with "even year" or "half year" ages is usually an indication that there is a significant amount of "rounding-off" occurring in the procedure for determining children's ages.

Another example occurs when the analysis can demonstrate no relationship (or worse, an inverse relationship) between consumption and nutritional well-being. That is, those families who are not consuming their full requirements have the same (or better) nutritional status as those families who are fulfilling their nutrient requirements. There may be some logical explanation for this phenomenon, but since the relationship between food consumption and nutritional status is such an important aspect of the nutrition theory, lack of any direct relationship may well indicate a high level of error in the consumption data and/or in the requirements data.

While extrinsic error detection is an "after-the-fact" procedure which is closely associated with the analysis process, intrinsic error detection requires that "checks" be conducted with (or as part of the measurement procedure. Generally, this type of error detection involves testing the accuracy of the standard measurement procedure by repeating the measurement several times or by applying an additional, more precise measurement technique. This is done on a small subsample for the explicit purpose of evaluating the standard measurement procedure. For instance, if the 24-Hour Recall Consumption Survey is repeated five times for the same family, the results would provide an estimation of the amount of variation that can be expected

from that indicator. The variation may represent a composite of several different kinds of error; for example, it would provide insight into how much a family's consumption fluctuates from one day to another.

Repetitive measures, however, will not uncover systematic (or biased) errors such as when persons consistently overstate their consumption due to embarrassment over their impoverished situation, or understate their consumption due to anticipation of receiving some assistance. In order to test for systematic errors, it is necessary to employ a more precise or direct measurement technique on the subsample. For example, it may be useful to undertake a detailed, direct-observation study to record the eating habits of a few families over a period of a week or more. Results of such a study, when compared to the 24-hour recall survey for the same families, could provide a better understanding of the full range of potential errors inherent in the 24-hour recall methodology.

These intrinsic methods of error detection can be applied "after the fact," but they can also be incorporated as a part of the regular measurement methodology. A common procedure for detecting transcription errors in keypunching data onto computer cards from coding sheets is to have two persons independently enter the data; the machine will reject the cards which do not match. This verification process is considered a regular part of the keypunch procedure.

Once the presence of error has been detected there are two possible responses: (1) correction and (2) interpretation. There will be some situations where it is possible to correct some of the errors. However, the opportunity for this type of response will be limited since much of the error detection procedure is applied only to a relatively small subsample of the data, as is the case when a more precise measurement of family consumption is made to check the results of the 24-hour recall survey. Another limitation to error correction is apparant in the example of inordinate age peaks at half years and even years. In this case there is little doubt of the existence of a significant number of erroneous values, but it is not possible to identify exactly which values were wrong and what the correct values should be.

A more significant type of correction may result when the detection of

errors leads to a change in the measurement procedures aimed at reducing or eliminating the errors. Continuing with the example of the erroneous ages, a change in the way that the question is asked may eliminate the problem; or it may be possible to obtain confirmation of dates through documents such as birth certificates.

While it will be possible to gradually improve the measurement procedures (thus improving their accuracy), the diagnostic methodology described in this manual -- by its very nature -- will be subject to a certain level of imprecision in its data. As such, the most important aspect of an error analysis is the interpretation of the effect which the errors may have on the conclusions. In some situations the effect may be negligible, while in others the conclusions could differ in direction and in strength.

The existence of errors will normally be defined in terms of a "range" of possible values. That is, for a particular value, the exact amount of error is not known but it is suspected of lying within certain limits. These limits are analogous to the "confidence limits" in statistics but the determination of the limits is not necessarily done statistically. The determination of the probable range of errors will depend primarily on the way in which their existence has been detected.

Once the limits of possible error have been established, a "sensitivity analysis" can be performed to determine what effect different values of a particular variable will have on the outcome of the analysis. A sensitivity analysis is the determination of how sensitive the decision is to change in a particular variable. If the extreme values (assumption of maximum error) are used in the diagnostic analysis and yet the final conclusions remain unchanged, then it can be said that the analysis is "insensitive" to the possible error in the particular variable. Stated differently, the sensitivity analysis determines whether or not more precision is needed for a particular variable.

An example using the calculation of the nutrient gap will help to clarify this concept. Suppose that the average per capita requirements in a community are calculated (based on population structure) to be 2400 calories; the average percapita consumption is measured and found to be 2000 calories. The conclusion is that there is a significant deficit in the consumption of

calories in the community. If it is estimated that the calculation of requirements is subject to a 2% error and the measurement of consumption is subject to a 5% error, then the possible range for requirements is 2352-2448; for consumption it is 1900-2100. Even if the requirements were overestimated (actually 2100), the conclusion remains unchanged: there is a deficit in the consumption of calories. On the other hand, if it is estimated that the consumption measurements may have as much as a 20% error, then the actual consumption could be as high as 2400 calories. This means that the community may be consuming sufficient calories and that the existence of a deficit is questionable.

As with most other topics in this manual, the actual kind and amount of error analysis will depend on the specific circumstances under which the methodology is being used. If only one community is being diagnosed and the results are reasonable and conform approximately with the *a priori* impression of the community's nutrition problem, then it is probably not necessary to expend much effort on analyzing the errors. However, if the methodology is being used extensively, or if the initial results imply a drastic change in the original theory or impressions of the community, then it is worthwhile to examine closely for errors affecting the diagnosis.