lib

DRAFT
DO NOT QUOTE

Filename: GUIDE4.DOC

Revised: January 21, 1990

GUIDELINES FOR THE USE OF STRUCTURED OBSERVATIONS IN HEALTH BEHAVIOR INTERVENTION STUDIES

This publication will be revised by IRC during the [illegible]

Prepared by

Margaret E. Bentley, Ph.D.
Rebecca Y. Stallings, M.H.S.
Joel Gittelsohn, Ph.D.

for the

World Health Organization

CDD Programme

The Johns Hopkins University
Department of International Health
School of Hygiene and Public Health
615 North Wolfe Street
Baltimore, Maryland 21205

January 1990

Table of Contents

	Page
INTRODUCTION	1
What are Structured Observations?	1
Historical Development of Structured Observation Methods	1
Purpose and Use of the Guidelines	2
Types of Health Behavior Studies	3
Risk Factor Studies	3
Health Behavior Intervention Studies	5
Structure of the Guidelines	6
PART I. TYPES AND ADVANTAGES OF STRUCTURED OBSERVATIONS	7
Types of Structured Observations	7
Continuous monitoring observations	7
Spot check observations	8
Environmental/individual ratings-checklists	9
Advantages of Structured Observations	11
Why is Survey Data Often Inaccurate or Misleading?	12
PART II. RESEARCH DESIGN ISSUES	14
The Need for an Interdisciplinary Team	14
Preliminary Qualitative Research	15
"Marker" Behaviors: When Direct Observations are Inappropriate	17
Sampling Issues	18
Location or Person-Based Observations	21
Frequency and Duration of Observations	21
Reactivity	24
Time of Day	25
Seasonality	26
Defining and Operationalizing Variables	26
Recording Multiple Behaviors	27
Training of Data Collectors	28

Page

PART III. SPECIFIC STEPS FOR THE USE OF STRUCTURED OBSERVATIONS IN INTERVENTION STUDIES 29

Step 1. Decide if Structured Observations are Necessary and Can Be Done Considering Project Resources 29
Step 2. Generate a List of Potential "Key" Behaviors through Literature Review and Interviews with Knowledgeable Informants 31
Step 3. Conduct Qualitative Research to Refine Hypotheses and Identify Behaviors to be Observed in a Structured Format 33
Step 4. Identify Actors, Locations, Times and Events Associated with Specific Behaviors to be Observed 37
Step 5. Estimate Reactivity and Variability of Key Behaviors 40
Step 6. Choose Type of Structured Observation to Conduct 44
Step 7. Design Instruments and Data Sheets 47
Step 8. Determine Data Collection Schedule 53
Step 9. Train Data Collectors and Pretest Instruments 55
Step 10. Conduct Reliability Tests to Reduce Inter- and Intra-Observer Variability 57
Step 11. Implement Data Collection and Data Management 59
Step 12. Clean Data Sets 61
Step 13. Operationalize New Variables 63
Step 14. Conduct Data Analysis 72

Tables 75

Figures 76

APPENDIX I. SAMPLING STRATEGIES 88

APPENDIX II. SAMPLE SIZE CALCULATION 90

APPENDIX III. DIETARY MANAGEMENT OF DIARRHEA ETHNOGRAPHIC GUIDE 92

APPENDIX IV. A SAMPLE "SCRIPT" 98

APPENDIX V. CALCULATION OF VARIABILITY AND RELIABILITY OF KEY BEHAVIORS 101

APPENDIX VI. GENERAL SUGGESTIONS REGARDING THE APPROPRIATE CONSTRUCTION OF DATA COLLECTION FORMS 104

Page

APPENDIX VII. GENERAL SUGGESTIONS REGARDING THE APPROPRIATE ENTRY OF DATA 105

APPENDIX VIII. RELIABILITY ANALYSIS OF SCALES 106

APPENDIX IX. DATA ANALYSIS 107

References Cited 113

INTRODUCTION

This document provides guidelines for the use of structured observations in household and community health behavior studies. The audience for these guidelines is an interdisciplinary team that wishes to consider structured observation techniques as part of a health behavior study, especially if the study is in support of a proposed or existing health intervention project.

What are Structured Observations?

Structured observations are a quantified record of a behavior or behaviors collected by a trained observer, through use of a precoded or partly coded data collection instrument. A simple example is an observation of a hygiene behavior, such as hand washing. Trained data collectors might observe and code several possible features of the behavior, such as whose hands are being washed, where, with what, and for how long. Variables and codes are carefully defined to minimize observer bias or error, and the data are amenable to statistical analysis.

Historical Development of Structured Observation Method

The historical use of structured observations derive from clinical and developmental psychology (Whiting and Whiting 1975; Rogoff 1978), from studies of animal behavior (ethology) (Altmann 1973), and from time allocation studies conducted by anthropologists (Rappoport 1967; Lee 1969; Johnson 1975) and economists, of which several reviews have been completed (Minge-Klevana 1978; Gross 1984; Muller and Caro 1985; Messer, 1989).

Types of Health Behavior Studies

There are two types of health behavior studies in which the use of structured observations may be useful. These include

1. risk factor studies, which seek to identify behavioral transmission routes of a particular infectious disease, such as diarrheal disease (Briscoe 1984; Stanton and Clemens 1986; Clemens and Stanton 1988).

2. intervention-related health behavior studies, which act to support intervention projects with the primary goal of disseminating messages to change health behavior and to achieve positive changes in morbidity and/or mortality rates (Khan 1982; Black et al 1981; Torun 1985; Stanton and Clemens 1987, a,b). These studies may have as their purpose either assisting project design, implementation, or evaluation (which may use structured observations to assess the success of an intervention, through a significant change in observed behavior (Stanton and Clemens 1987,b)).

An example of how structured observations have been used for these different types of studies is discussed below.

Risk Factor Studies

A risk factor study of health behavior attempts to show a relationship (either causal or non-causal) between a particular behavior or set of behaviors and the risk of disease. For example, there is now good evidence, obtained through several risk factor studies, of the association between the quantity of water available and used within a household and reduced diarrheal incidence (Esrey et al. 1985).

achieve behavioral change, in a randomized trial Stanton and Clemens (1987,b) measured both change in behaviors and diarrheal morbidity among intervention and non-intervention communities. Structured observations, using the same instrument developed for the risk factor study, were made of the three behaviors that were related to the promotional messages. One 3-5 hour visit was made for each family, and comparisons of hand washing, fecal deposit, and fecal disposal behaviors were made between intervention and non-intervention communities. The evaluation showed that significantly more mothers washed their hands in the intervention villages, but there were no differences in where a child defecated or in garbage/feces disposal. Diarrheal incidence rates were 26% lower in the intervention communities for the six month period of the intervention (22% for the year after follow-up).

Structure of the Guidelines

These guidelines are divided into four parts. Part I describes the different types of structured observations and considers some of the advantages of structured observations compared to other methods of data collection, particularly surveys. Part II focuses on general research design issues that should be considered when using structured observations, including cultural sensitivity, the need for preliminary qualitative research, subject reactivity and an interdisciplinary team approach. Part III is a step-by-step guide for formulating and implementing structured observations for health behavior studies.

Health Behavior Intervention Studies

A health behavior intervention study uses research as a tool to assist in the design, implementation (operations research, monitoring) and/or evaluation of a health project.

For the purposes of project design, health behavior studies are concerned with identifying modifiable behaviors that are significantly associated with disease rates. An example is the Stanton and Clemens study which had the ultimate objective of developing culture-specific messages for promotion through a community intervention. Once they identified the behaviors (through the use of structured observations) that were found to be highly associated with incidence of diarrheal morbidity, they designed messages to promote behavior change (1987,b). The messages were directly related to the three behavioral risk factors that were identified (fecal deposit, fecal disposal, and hand washing).

A second use of health intervention research is for project implementation. Such research may have two purposes: 1) to solve specific program-related problems as they arise (ie. operations research); and 2) to provide ongoing monitoring of project activities. An example of the former might be an infant feeding study already underway that promotes a local weaning food, but that shows poor usage rates. A piece of quick "trouble-shooting" research could be developed to ascertain the reasons for the low rates and lead to program improvements.

Finally, health intervention research may be used for project evaluation. To assess whether the hygiene intervention messages did equipment in place of a trained observer. The use of videotape to record actions and behaviors is a common technique in psychology. Following the use of film by anthropologists Mead and Bateson (Bernard 1988:278), videotape has been been used for observation of non-verbal behavior (Dehavenon 1978). Videotape has the advantage of recording actual actions and events, and observer effects may be minimized. It would be important to determine beforehand that videotaping was both culturally acceptable and did not cause subjects to alter their behavior. As well, a system for operationalizing variables and transforming the video images into quantified observations must be developed.

Spot check observations

A spot check observation is a particular type of structured observation, whereby the data collector records an activity or appearance of an individual or thing at the first moment of observation (Mulder and Caro 1985). For example, for time allocation studies, spot observations often note the exact activity of all individuals within a household at a given (usually random) time.

Spot observations are, by definition, a rapid assessment tool, and for health behavior studies the information is usually recorded, on a precoded data collection sheet, immediately upon arrival of the data collector to a household. This may have the advantage of making an assessment that is "naturalistic" and less biased than continuous monitoring observations, which are more invasive due to the possible interruption of natural events by the presence of the data collector in

PART I TYPES AND ADVANTAGES OF STRUCTURED OBSERVATIONS

Types of Structured Observations

There are two main types of structured observations: continuous monitoring and spot check. Both methods can be enhanced by the addition of an environmental/individual ratings checklist.

Continuous monitoring observations

Continuous monitoring structured observations involve the presence of a trained observer at a household (or other designated locale) for an extended period of time. Activities and behaviors of interest are recorded in a structured format (either with notes or through use of a pre-coded data sheet) along with associated temporal and spatial characteristics.

Continuous monitoring observations of health events may focus on defecation disposal patterns, hand washing behavior, weaning food preparation behavior, infant feeding and so on. Within each category of behavior, a more detailed description may be desired. For example, to explore weaning food preparation behaviors, it may be insufficient to record merely that the food was prepared. An investigator may wish to know how the weaning foods are prepared and whether weaning foods are prepared fresh each time or not. If they are not prepared fresh each time and reheating of foods occurs, the investigator may wish to examine how and for how long weaning foods are stored. The observer may spend many hours in the household obtaining this information.

One form of continuous monitoring involves the use of video environmental/individual ratings checklist. The ratings checklist requires the observer to make a judgement on some aspect of environmental or actor condition, and record his judgement as part of the structured observation.

For example, during a structured observation, ratings of a household hygiene environment might focus on the presence or absence of visible dirt or other types of contamination in different areas of the house and on different household members (see Figure 1). Another example relates to child caretaking behavior, where an observation might focus on whether a small child's face, hands, or clothes are "clean" or "dirty."

The use of environmental/individual ratings checklists does not require a data collector to wait for an event to occur before information is recorded, as an observer assessment is made immediately in many cases. Thus, this method is particularly suited as an addition to spot check observations. They can be "one-time" only ratings, which might result in an ordinal hygiene score for the household or individual, or they may be repeat ratings of the same environment or actor across time to deal with issues of variability or seasonality. In the case of the former, a spot observation checklist of a large number of variables might be filled in at the same time socioeconomic and demographic information is collected. In the case of the latter, repeat spot observations might occur at the beginning, middle, and end of a study, or could be done once during each season for the duration of the study.

the home. However, it is not possible to directly observe behaviors of limited frequency and duration (such as hand washing) through use of spot observations, since it is unlikely that an individual will be washing his hands at the precise moment a spot observation is scheduled. Spot observations are most useful for documenting overall patterns of activities for one or more individuals. However, in combination with a technique described below, they can be used for making a quick, objective assessment of the environment or of a particular person.

In some instances, it may be useful to do timed spot observations (e.g. every 10, 15 or 30 minutes) during a longer observation period (time-sampled spot check observations). Decisions about the number and frequency of spot observations should be made based upon research objectives, data analysis capacity, the variability of the phenomenon in question, and how strongly it relates to the outcome measure. An important trade-off to recognize is that a one-time observation will be much easier to handle in the statistical analysis, but may not capture the variability of the phenomenon. A dirty kitchen on Monday morning may be a clean kitchen on Wednesday afternoon.

Environmental/individual ratings checklists

Both continuous monitoring and spot observations involve recording individuals and their activities as they occur, without value judgements. The entry: "woman washes her hands" is a pure observation, while the entry: "woman washes her dirty hands" requires a judgement by the observer on the cleanliness of the woman's hands. Both methods of structured observation can be strengthened through the addition of an studies lies in the depth and breadth of this methodology. Not only can the researcher observe the frequency and duration of key behaviors and events, but he or she can observe associated activities and behaviors. Thus, structured observations may help us to understand why a mother does not wash her hands before preparing food on one occasion and why she does on another.

Why is Survey Data Often Inaccurate or Misleading?

Reported behaviors have often been found to provide inaccurate data, since people often do not do what they say they do, or what they think they should do (Stone and Campbell, 1984; Stanton and Clemens 1987c; Huffman et al. 1988; Hornick 1989). There are several reasons why this is so. First, every culture has "rules" for behavior, called "cultural norms," which influence how people dress, how and what people eat, and how they should behave. Cultural norms are learned and are passed down from generation to generation. People's adherence to cultural norms, however, will usually vary, depending on many individual and social factors. Whether people adhere to cultural norms or not, they usually know the "rules." Therefore, if a North Indian mother is asked whether she massages her baby with oil every day (as "good" Indian mothers do) she may report that she does, when in fact she does so only occasionally. This may be because she is often too busy or because she does not always have the oil, or for some other reason. The important point here is that she knows what she should do and what she may even wish to do, but other factors influence what she actually does. Still, because of strong cultural beliefs and traditions that dictate

Advantages of Structured Observations

There are many advantages to using structured observations for intervention-related research. First and foremost, structured observations of human behavior provide information on what people actually do, rather than on what they say they do (or did). For example, if we are interested in knowing whether a mother washes her hands before handling food for the family, there are several ways in which this question might be investigated:

a) The mother could be asked, through use of a structured (quantitative) or unstructured (qualitative) interview, whether she washes her hands before food preparation. Data obtained this way provides information on reported behavior.

b) The mother could be observed, through continuous monitoring in the home, for handwashing before preparing food. Data obtained this way provide information on actual behavior.

c) The mother's hand washing behavior could be inferred through use of a spot observation, combined with a ratings checklist. The ratings will provide a "proxy" for the behavior of interest. For example, if a mother's hands are observed five times and they are clean each time, it is not unreasonable to infer that this mother does regularly wash her hands. It is not possible to really know how she cleans them or whether she does so before meal preparation, however, unless the actual event has been observed, or unless reliable information can be obtained through interview. Thus, these data provide information on inferred behavior.

A second advantage of structured observations for intervention

Nepal evaluated the accuracy of a KAP (Knowledge, Attitudes and Practices) survey of rural family planning and contraceptive practices by conducting in-depth ethnographic research in the same households. They found that many questions were culturally reinterpreted. For instance, the KAP survey found that 37% of the respondents had never heard of abortion. Stone and Campbell found through their own in-depth interviews that these individuals had heard of abortion, but had reinterpreted the question on "heard of abortion" as a question on knowledge of technique or as knowledge of who had had an abortion.

PART II RESEARCH DESIGN ISSUES

This part of the guidelines introduces and briefly discusses issues relating to the use and design of structured observations for intervention-related research.

The Need for an Interdisciplinary Team

An interdisciplinary team approach for conducting health behavior studies or interventions is essential. Although the mix of team members may vary depending on the specific research question or on local resources, a typical team composition would include an epidemiologist, a social scientist, and if an intervention is planned, a health educator or communications specialist.

The team should work together from the beginning, including the development of hypotheses and the research plan, the design and "appropriate" behavior, she may feel compelled to report the idealized behavior.

Suppose it is important for an investigator to know with some degree of certainty whether a mother massages her baby with oil during a 24-hour period. Based upon a lack of knowledge about the variability of this behavior, but aware of the strong cultural belief that baby massage is an important care taker responsibility in this culture (Reissland and Burghart 1987), the investigator is unsure whether just asking the mother if she massaged her baby yesterday will provide true information. In this case, the best option may be to have a data collector stay with the mother for a reasonable period of time to see if the event occurs or not.

Second, it is not uncommon for people to report to a data collector (often a complete stranger, who may be more educated or urban than the respondent) what he or she thinks the data collector wants to hear. For example, imagine that a community intervention has intensively communicated messages for people to brush their teeth twice daily. After six months of message dissemination, an evaluation to assess behavioral change queries respondents as to whether or not they brushed their teeth that morning. In such an evaluation of reported behavior, it is likely that some respondents will report what they think the data collector wants to hear — that they did brush their teeth — when actually they did not.

Other studies of the efficacy of survey research for measuring human behavior have identified problems with the interpretation of questions by the respondents. A study by Stone and Campbell (1984) in free, and so on. In addition, it is necessary to have some estimate of the variability of the behaviors. If, for example, it is found through ethnographic interviews and observation that a particular food storage behavior shows little variability across households (even if it has the potential of being an important transmitter of pathogens), then it would be foolish to choose this behavior for measurement through structured observations, since it would not emerge as a risk factor in the data analysis (On the other hand, it would still be important to measure this behavior in some other way in order to describe the study population). Likewise, if the data revealed that weaning foods in the study area were rarely saved and re-fed to infants, then the hypothesis would need to be reformulated.

A number of social science methods can be used to conduct the qualitative research. For example, several applied health projects have used ethnographic techniques of unstructured interviews with a wide variety of respondents and a few key informants. Ethnographic techniques also rely on some amount of observation within the natural setting, to complement the interview data. Focused ethnographic research, which should be conducted through the use of a detailed instructional guide, can usually be completed in one to two months, providing rich detail of a setting and a problem (Brown and Bentley 1988; Bentley et al 1988; Scrimshaw and Hurtado 1986). General references for conducting ethnographic research are also available (Bernard 1988; Kirk and Miller, 1987; Spradley 1979; Werner and Schoepfle 1986; Morgan 1988).

Another technique, which has recently been adapted by practitioners pretesting of instruments, the review of secondary and primary data, implementation, and through data analysis. A compartmentalized approach, whereby each team member does his or her "own thing," is not true interdisciplinary research.

Preliminary Qualitative Research

It is the premise of this document that structured observations of health behavior, regardless of the study design, should never be done without preliminary qualitative research of the key study questions. Indeed, if "good science" is the goal, preliminary qualitative research is required to refine hypotheses, finalize the protocol, and design the instruments.

For example, returning to the weaning food storage study, what kind of information might be required? A principal investigator would probably have at least some idea of how food is prepared, served, and stored within households in the setting where our study will take place. However, unless the behaviors of interest are understood in detail it is not possible to formulate a good working hypothesis or to design data sheets for the use of structured observations.

For this example, the kind of qualitative information that may be required includes the frequency and times of day when food is prepared for an infant, who the usual child caretaker/feeder is, where food is prepared, the types and recipes of weaning foods prepared, the technology that is used, the types of utensils used for food preparation and within which food is stored, whether it is common to leave food uncovered or not, and if so, where food is stored, whether animals run by each investigator.

When a behavior itself cannot be observed, for sensitivity or for other reasons, often another "marker" behavior can be noted. For example, in Nepal, adults pour a small pot of water to take with them when they go to defecate. This behavior could be used as a "marker" for adult defecation. Careful attention should be paid to identifying culturally appropriate behaviors for direct observation and those which must be observed indirectly through the use of "marker" behaviors.

Sampling Issues

Two main sampling issues need to be resolved by investigators interested in conducting structured observation research: selection of an appropriate sampling method and determining a suitable level of selection.

The selection of sampling methods depends primarily on the purpose of the structured observations, specifically, whether these observations are intended to be used as explanatory or predictive variables, or as the outcome of interest. Several common sampling strategies are discussed in Appendix I, including simple random sampling, stratified random sampling, and cluster sampling.

The second key sampling issue is how large a geographical area to cover for the research. Sampling schemes operate on many levels, depending on the requirements of the study or project. The levels that may occur in health intervention projects are four: country, city/village, household/site within city/village, and individual. Since doing accurate structured observations requires an in-depth cultural of "social marketing," uses focus group interviews as a key research tool (Manoff 1985; Griffiths et al 1988; Morgan 1988). Focus group interviews have the advantage of generating a large amount of information in a relatively short time, and the group dynamic often provides information that may not come out through person-to-person, in-depth interviews. Focus group interviews, however, require skilled facilitators, and the interpretation of the information collected may be difficult.

In summary, there is a growing literature on the use of qualitative research methods for public health purposes, and these should be consulted before attempting to design the preliminary research. As stated previously, it is recommended that this work be implemented with the full collaboration of a social scientist, preferably someone who has done previous applied health research, and is familiar with both qualitative and quantitative research methods.

"Marker" Behaviors: When Direct Observations are Inappropriate

There are some behaviors or events that are "off-limits" for the use of structured observations. For example, although most investigators of family planning would like to know the frequency and effective use of condoms, this is not an event that can or should be observed. Likewise, although it is possible to observe children defecating within a household or compound, there are few cultures where it would be appropriate to actively observe adults defecating. Sensitivity around the use of structured observations within a specific culture or for specific behaviors is an issue that should be addressed

Examples 3 & 4 can also be classified as cluster sampling problems because the unit being randomized (i.e. the village) is comprised of multiple observable units (e.g. the household).

III. Choosing multiple sites within cities/villages (may or may not be chosen randomly)

Example 1. - Project is directed toward all households with children under 3 years of age. A census should be conducted and all such households should be invited to participate in the project. This is a comprehensive, non-random sample, although there may be a bias if households can refuse to participate.

Example 2. - Intervention is to be implemeted by volunteers, each of whom is responsible for teaching n participants. This is a non-random sample where serious bias may be introduced because of the self-selection of the volunteers and their selection of participants.

Example 3. - Resources have limited the investigator to selecting n households for intensive in-home study. A census should be conducted and n households should be chosen from among all of those eligible and willing to participate. This is simple random sampling. Again, there may be a bias if participation is voluntary.

IV. Choosing a target person(s) within the site (may or may not be chosen randomly)

Example 1. - Sites are defined as compounds, which are randomly selected. Families live jointly in compounds, and several families have children under 3 years of age. All children under 3 are included in the sample. This is a comprehensive, non-random sample and fits the definition of a cluster sample.

Example 2. - Limited resources require that the research team choose only one child per compound for intensive observation. All children in the compound under 3 should be listed and one target child should be randomly chosen from the list. This is simple random sampling.

When a sampling method and level of selection have been identified, knowledge of the study area, and as most countries show great internal cultural differentiation, the level of selection for structured observations is likely to be at the city/village level or lower. Some examples follow:

I. Choosing multiple countries (usually <u>not</u> chosen randomly for practical or political reasons)

II. Choosing multiple cities or villages (may or may not be chosen randomly)

Example 1. - A national government has designated a target area; all villages in the target area are to be included in the project. This is a comprehensive, non-random sample.

Example 2. - A national government has designated a target area with participation conditional on agreement of the village elders. This is also a comprehensive, non-random sample, although there is a potential here for a lack of representativeness of the villages (bias) with respect to the target area.

In the above two examples, there is no need to discuss sampling issues, because the sample is predetermined.

Example 3. - Project design needs n_1 villages for "exposed" and n_2 villages for "non-exposed" groups (e.g. villages with and without latrines). A list of possible villages to choose from will be devised, designating each as "exposed" or "non-exposed". N_1 villages will be randomly selected from the "exposed" and n_2 villages will be randomly selected from the "non-exposed". This is an example of stratified random sampling.

Example 4. - Project design needs n_1 villages for "intervention" and n_2 villages for "non-intervention" or "control". A list of possible villages to choose from will be devised. N_1 villages will be randomly chosen for "intervention" and n_2 villages will be randomly chosen for "control" status. This is a case of simple random sampling with random assignment.

"leftover" weaning foods are stored or not, for how long, and in what condition, before they are fed to the child in a subsequent feeding. This is a rather complicated research question to investigate, and it is probably not amenable to spot observation techniques, for the reasons cited above. To investigate this issue, information on reported behavior could be collected from the child caretaker, through an unstructured or structured interview. If structured observations are desired, however, it will be necessary for a data collector to sit in the household for a period of time to observe several events. A possible protocol for this research might instruct a data collector to arrive early in the morning for the child's first feed and to remain until the second feed is completed, recording a number of behaviors of interest (such as meal preparation, feeding, storage, utensil cleaning, etc.). Depending on the locale, this might require 5-6 hours for each observation period/household. A longer period of observation time might be required if the investigators were interested in storage duration and patterns of food reuse.

Previous hygiene studies that have used continuous monitoring observations often limit the number of observation periods (or visits) to one or two[1]. For example, in the Stanton and Clemens study (1987b), one three-hour observation was done per household. An obvious reason why the number of observation periods in most studies is small is because of the labor-intensive nature of continuous monitoring observations. Another reason, is because data analysis is less complex

---

[1] For a discussion of why this would be a suboptimal policy, see the discussion of "Reactivity".

calculations can be performed to determine the appropriate sample size. Sample size calculations are described in Appendix II.

Location or person-based observations

A key research design decision is whether to focus the structured observations on a person or persons, or on a specific location. In the case of the former, the data collector would "follow" the targeted person, perhaps moving from location to location. For example, the investigation of child caretaking would focus on the child, regardless of who was doing the caretaking. In this case, the child would be followed, and observations would be person-based. On the other hand, if the focus is on water storage, observations might be focused on the water jug, to establish whether it is properly covered or not. In this case, several individuals might move into the location and become part of an observation, but structured observations would be location-based.

Frequency and duration of observations

An important set of decisions has to do with when and how many times an environment, person, or event is observed, and in the case of continuous monitoring observations, the duration of each observation period. These decisions will depend upon a number of factors, many of which can be assessed through preliminary qualitative research or secondary data sources.

For example, imagine a risk factor study that is focused on the relationship between weaning food storage behavior and diarrheal morbidity of weanlings. The key feature of interest is whether

Returning once again to the weaning food storage study, if uncovered food in a household is observed during one observation period, can it be inferred that food in this household is usually not covered before it is re-served? Should the event be observed two or three separate times? What are the cost implications for multiple observation periods? What are the implications for research validity if observations are not done during multiple periods?

There are no easy answers to these questions. Much will depend on the specific research issue, on the sociocultural context, on the duration of the study, on whether it crosses seasons, and on other factors. One way to address this issue is to make a preliminary rapid assessment of the variability of the key behaviors, either through survey or ethnographic interview and observational techniques. These data can be used to help determine the number of observations required. An ethnographic study can also help the investigator assess the problem of subject 'reactivity' when performing key behaviors.

Reactivity

Reactivity refers to the "observer effect" during structured observations. It is quite possible that subjects being observed will 'react' to the presence of an observer and alter their behaviors. For instance, when Gittelsohn (1989) observed meals in Nepal, he found that young women, shy in his presence, would make efforts to avoid being seen eating by sitting in an obscure or darkened corner of the room. In subsequent observations in the same household, this avoidance behavior ceased in most households.

when there is no need to sum observations across different visits. However, a major risk is to sacrifice reliability and validity for logistical ease or cost, essentially ignoring the intra-individual or household variability of many events.

Not all structured observation studies have skimped on time periods. Bentley et al.'s (1989) study of infant feeding practices during diarrheal episodes conducted continuous structured observations during three consecutive 12-hour days of an illness episode rather than on only one day, since it was felt that child feeding behavior during illness would be highly variable. As the purpose of the same study was to compare feeding behavior during illness with convalescence and health, additional continuous observations were made on two days each of convalescence and health. The same study included spot-check and ratings list methods as well to consider other behavioral issues. As preliminary investigation indicated that women clean up after meals in a consistent fashion, it was decided that one need only observe this process once per household, marking certain designated behaviors on a ratings checklist, in order to examine "cleaning up after meals" as a potential risk factor. During the continuous monitoring observation of the target child, as the investigators were interested in mother's time allocation (specifically the relative amount of time spent in childcare activities) a short checklist of possible activities was marked every 30 minutes throughout the day (i.e. time-sampled observations). The time-sampled checklists were repeated for each of the 3 health conditions.

Unfortunately, there is no rule-of-thumb to judge how often a particular event should be observed to reflect natural variability.

should be conducted all day long? Clearly this is an impossible task and is not necessary. Key behaviors of interest may occur during limited and regular periods during the day (i.e. meals). A period of initial ethnographic research with trial observations should suffice to identify **when** a key behavior is most likely to occur. On the other hand, if the researcher is attempting to ascertain the proportion that a particular behavior or set of behaviors constitutes of an entire day of activity, observation periods would have to be expanded.

Seasonality

Just as morbidity rates are affected by seasonality, so too are many behaviors. One of the ways in which maternal health behavior may be affected is through changing time allocation patterns and activities across seasons. This is particularly true in areas where women have major roles in agriculture. During planting, harvesting, and agricultural crop processing periods, maternal food preparation, and domestic and child caretaking activities will probably be different or compromised. At the very least, it is important to know how seasonality may affect the research question. If a large effect is expected, a choice can be made to limit data collection among all households to only one season, so that seasonality does not confound the results. If the research protocol spans more than one season, care must be taken to control for seasonal variability.

Defining and Operationalizing Variables

It is essential that variables and codes be defined and

Reactivity to structured observations can be reduced in two ways. First, prior to the study, the presence of observers in the study community for a significant period of time (ie. at least a few weeks) appears to reduce people's unease. Second, reactivity appears to be reduced by repeated observations, as the observed grow accustomed to the presence of the observer (Mulder and Caro 1985; Bernard 1988:271). Through pretesting and the initial ethnographic period, it may be possible to identify a point in time where reactivity decreases significantly (e.g. a "reactivity threshold"). Figure 2 shows a hypothetical graph of change in subject reactivity with repeated observations. Note that reactivity to the presence of the observer never fully disappears. Clearly, this reactivity threshold will be one factor in determining the optimum number of structured observations per household. It may be possible, through extensive ethnographic research, to identify types of behaviors that are associated with higher and lower subject reactivity. A reactive behavior would be expected to either increase or decrease continuously with time, while a non-reactive behavior (the key behavior) will fluctuate according to natural variability but should not show steady decline or incline.

Time of Day

Choosing the appropriate time of day during which to conduct a structured observation is an important issue. Obviously, people are not equally active over a 24-hour period, and of course the nature of the activities that they are involved in will change throughout the day (Martin and Bateson 1986). Does this imply that structured observations simultaneous activities, which subsequently increases the complexity of the analysis phase.

Training of Data Collectors

Adequate time for training of data collectors and the development and pretesting of instruments should be allowed. For most projects, this process should take at least one month, depending on the number and type of structured observations and the number of data collectors to be trained.

The training period should allow an orientation to the overall objectives of the research, the technical aspects of the project, and detailed instructions on the use of structured observations. The specific research hypotheses, however, should not be shared with the data collector. For example, in the weaning food storage research described above, data collectors should not be informed about the hypothesis regarding the relationship of weaning food hygiene and diarrheal morbidity. Such knowledge has the potential of introducing bias to their observations.

operationalized to eliminate problems of inter-observer variability. Fisher et al. (1983) state,

> Operational definitions serve two very essential purposes: (1) they establish the rules and procedures the research investigator will use to measure variables; and (2) they provide unambiguous meaning to terms that otherwise can be interpreted in different ways.

In other words, the goal should be to define variables and codes so that, if a pair of data collectors observe the same event, they will "see" and code the event in exactly the same way. This task is easier for some variables compared to others. Observing whether a hand washing event took place, and whether water or water and soap were used, is fairly straightforward. It is more difficult, however, to define whether an environment or child is "clean" or "dirty," as might be required for an individual ratings checklist. For a data collector to know which code to choose, he or she must be provided with a precise definition of each possible choice.

Recording Multiple Behaviors

Individuals are capable of doing several things at once, and when several of these simultaneous behaviors are significant to the research, investigators must set up a clear system for recording them. Bernard (1988) recommends recording all possible behaviors observed in the order of their primacy, according to the observer's best judgement at the time of observation. For instance, if a woman was observed caring for her children and preparing the evening meal, the observer may judge her primary activity to be food preparation, and secondarily child care. Alternatively, some data collection forms allow for the recording of collection required for intervention design and evaluation may need to be of broader scope than that required for implementation research, where the research may be focused on answering one or more specific operational questions.

1b. **Determine if structured observations will fulfill those needs.** Does the project require knowledge of **actual** human behavior for design, implementation, and/or evaluation purposes? Or will reported behavior suffice? Structured observations of events for health studies are complicated to conduct and analyze and are often expensive, when compared to survey or ethnographic data. The choice of technique should depend upon the requirements of the study and upon the financial and research capacity to implement a particular methodology.

1c. **Decide if project resources can support structured observation research.** Can the logistic requirements of structured observations be met by project resources, including cost, time, personnel, and community acceptance? The relatively high cost of structured observations has already been mentioned. Structured observations, especially continuous monitoring, are a very time and personnel intensive methodology. Clearly, an interviewer could inquire as to the hand-washing behavior of mothers, a process which would take only a few minutes. Continuous monitoring or even spot checks (with associated ratings checklists) would take considerably longer. At this point, it is probably not possible to estimate

PART III SPECIFIC STEPS FOR THE USE OF STRUCTURED OBSERVATIONS IN INTERVENTION STUDIES

This part of the guidelines provides step-by-step instructions on the process of developing, pretesting, implementing and analyzing structured observations within the context of a health intervention study. Each step is further broken down into one to six key goals. Examples are drawn from hypothetical and real research projects that have investigated health behaviors. It should be emphasized that these "steps" describe one possible scheme for incorporating structured observations into intervention projects; variations on this design are possible. As well, the application of these guidelines for non-intervention study research is possible.

Step 1. Decide if Structured Observations are Necessary and Can Be Done Considering Project Resources

KEY GOALS:

a. **Determine research needs of intervention project**
b. **Determine if structured observations will fulfill those needs**
c. **Decide if project resources can support structured observation research**

1a. **Determine research needs of intervention project.** It is assumed that the user of these guidelines is engaged in an intervention project and would like to conduct research in order to assist in: intervention design, intervention implementation, and/or intervention evaluation. It is likely that the types of data category, specific behaviors can be listed.

There is a large hygiene literature that can be drawn upon to assist in making the list (see Feachem 1984; Esrey et al 1985; Esrey and Feachem 1989). The international hygiene literature, however, may not address the regional or local setting, and there may be a gap of information. Epidemiological data on diarrheal incidence and prevalence rates should, of course, be reviewed.

In addition to the hygiene literature, it is equally important to research the social science or ethnographic literature relevant to the setting. The social scientist on the team should be able to locate pertinent reports and documents. Embedded within ethnographic reports may be useful pieces of information that detail local descriptions of hygiene, food preparation, or child care behavior.

It is also recommended to arrange meetings with a wide variety of people who are knowledgeable about the selected population and who may be able to provide information about local behavioral patterns of interest. These may include sociologists, anthropologists, village health workers, PHC clinic doctors or nurses, or individuals involved with community interventions through NGOs and PVOs. These meetings should be unstructured and free flowing, almost a "debriefing" exercise of expert "key informants." It should not be surprising, however, if different individuals provide contradictory information; an urban doctor posted to a rural clinic for only one year will no doubt have a different story to tell than that of the local, village health exact costs.

Step 2. **Generate List of Potential "Key" Behaviors through Literature Review and Interviews with Knowledgeable Informants**

KEY GOALS:

a. **Review the literature and "brainstorm"**
b. **Develop preliminary list of hypothesized behaviors**

2a. **Review the literature and "brainstorm."** Even the brightest, most sure-handed and quick of eye observer cannot record all events that are likely to occur during an observation. Moreover, even if it were possible to record everything that occured during an observation, it would be impossible to enter and analyze the massive quantity of data that would be produced. Whatever the purpose of the research (design, implementation, or evaluation), a subsample of key behaviors must be identified as appropriate to record during structured observations.

A preliminary step is to generate a list of "candidate" behaviors. This exercise should be done by the entire team of investigators. The objective is to "brainstorm" about possible behaviors associated with the outcomes in question (eg. diarrheal mobidity and mortality, weight-for-age, etc.). At this point, there should be no concern about the length of the list. For instance, if the study is attempting to investigate behavioral factors linked to the transmission of diarrhea, one way to begin is to list categories of factors, such as "food handling," "personal hygiene," "weaning utensils" or "infant feeding." Within each and key informants to locate for interview, the recording and summation of data, and so on. An example of an ethnographic guide that was developed for the Dietary Management of Diarrhea Project (Brown and Bentley 1988) is shown in Appendix III.

A decision should be made about whether to conduct focus group interviews. As stated above, focus group interviews have the advantage of generating a substantial amount of information about reported behavior in a short period, but require skilled individuals to conduct them and to interpret the data.

For observational data collected during the preliminary ethnographic period, an effective strategy is for a data collector to remain in a household or compound for several hours, taking notes about what he or she sees in the form of a "script." Obviously, the "scripting" of each and every behavior, event, or actor may not be feasible or necessary, but detailed descriptions of relevant events and their sequence can provide a rich data source. An example of a "script" of a mealtime observation is given in Appendix IV.

Although the data collected from this preliminary research is qualitative, it does provide an opportunity to establish rough estimates of the variability of certain behaviors of interest. For example, if an investigator is interested in knowing the frequency and duration of breastfeeding, this information could be collected and rough percentages calculated. Although not precise, these estimates will help in the refinement of hypotheses and the choice of behaviors for structured observations.

worker.

2b. **Develop preliminary list of hypothesized behaviors.** After these tasks have been completed, the team should finalize their preliminary list of hypothesized behaviors, perhaps giving "weight" to behaviors that appear to be particularly important (for example, behaviors that were frequently cited in the literature or by key informants, or behaviors that intuitively suggest a major transmission route for disease). The list will form the basis of the qualitative data collection.

Step 3. Conduct Qualitative Research to Refine Hypotheses and Identify Behaviors to be Observed in a Structured Format

KEY GOALS:

a. **Develop ethnographic field guide**
b. **Through qualitative research, refine list of behaviors to be observed in a structured format**
c. **Determine behavioral markers, if needed**
d. **Determine reactivity threshold of observed behaviors**
e. **Ascertain ethnic, economic, religious heterogenicity of study population**
f. **Hypothesis generation: analysis and interpretation of qualitative and secondary data**

3a. **Develop ethnographic field guide.** Based upon the list of hypothesized behaviors, organized by categories, a guide for conducting unstructured interviews and observations regarding relevant behaviors should be developed. The guide instructs the field workers, who ideally should be trained ethnographers and/or focus group facilitators, on the types of questions that should be investigated and observed, the types of households, respondents, has been created, it will be necessary to ascertain if all may be directly observed. As stated earlier, it may be culturally inappropriate or insensitive to observe many behaviors. During the initial ethnographic phase, an attempt should be made to identify appropriate "marker" behaviors. What are the characteristics of a good "marker" behavior? It should be both specific and sensitive to the actual behavior. In other words, the "marker" behavior should consistently occur when the actual behavior occurs and it should not occur when the actual behavior does not.

3d. **Ascertain ethnic, economic, religious heterogenicity of study population.** The main part of the qualitative data should not be collected within the exact same location or among households where the structured observations will be done, since interviews about health behavior might influence what people subsequently do (Chisholm 1985). The area(s) selected, however, should be similar culturally, environmentally, or socioeconomically. An effort should be made to ascertain the ethnic, religious, environmental, and economic heterogenicity of the two study areas for this purpose. Additionally, this information will serve an important function when deciding on the type of structured observations to conduct. Much of this information may be obtainable from secondary data sources (i.e. censuses).

3e. **Hypothesis generation: analysis and interpretation of qualitative and secondary data.** At this point, a review of all available data

The amount of time allocated to the qualitative data collection will depend on the complexity of the research project, the number and variability of sites, the number of data collectors, and available resources.

The ethnographic data may be entered and organized for analysis on a microcomputer. A number of software packages are available for the entry, organization and analysis of textual information, including SCIMATE, ANTHROPAC, ZyINDEX, and NOTEBOOK (Pfaffenberger 1988). A report of the qualitative data should be prepared by the lead ethnographer or social scientist for review by the entire team.

3b. **Through qualitative research, refine list of behaviors to be observed in a structured format.** The preliminary list of behaviors that has been developed will structure the content of the preliminary qualitative data collection task. The collection of primary, qualitative data, through use of ethnographic and focus group techniques, seeks to provide detailed information that will help the team to identify new behaviors and refine existing behaviors to be observed in a structured manner. An example of the successive phases in the identification and refinement of a list of key observable behaviors is presented in Table 1, which describes the behaviors that were identified by Gittelsohn (1989) for his study of intrahousehold food distribution in Nepal.

3c. **Determine behavioral markers, if needed.** Once a list of behaviors ethnographic data to identify and refine key behaviors for the structured observations. This step uses the same data to describe the context in which these behaviors occur. As an example, consider an interesting hypothesis investigated by Stanton and Clemens in Bangladesh (1986) regarding urban mothers' use of a sari and the possibility that it might be a vector of diarrheal disease transmission.

Through fieldwork implemented for a sanitation project, the investigators noted that "in Bangladesh women used the sari for purposes other than to clothe the body and we speculated that these uses might be health hazards for their children" (1986:485). Thus, a hypothesis was developed and a study undertaken to focus more intensely on mothers' use of a sari, and to correlate specific behaviors with diarrheal incidence rates of their under 3 year old children.

4a. **Identify actors responsible for key behaviors.** An actor in structured observations refers simply to the individual carrying out the behavior of concern. Just as it is important to reduce the number of key behaviors to observe to a manageable level, so it is important to focus in on those actors responsible for the behavior. A decision must be made about who to observe: in the sari example, it will be a mother of an infant, or other female caregivers.

4b. **Identify locations where key behaviors occur.** Prior to commencing structured observations, it is important to determine where to observe. This decision is important, because it will affect the physical placement of the observer. Behaviors which are actor-should be conducted. The objective of the review of primary (ethnographic) and secondary (archival) data is to identify specific behaviors for structured observation and use them to refine research hypotheses. A well-formulated testable set of research hypotheses is critical for intervention planning.

Every investigator begins a research project with assumptions about "what is going on" within the population to be studied. Indeed, it is these "hunches" that are the beginning of good science, and which contribute to the formulation of original hypotheses. The information collected in Steps 2 and 3, however, will allow a sharpening or correction of research hypotheses.

For example, suppose that an investigator hypothesizes that a mothers' hands are an important transmitter of diarrheal pathogens. Without detailed information about local maternal domestic work patterns and hygiene behavior, however, the investigator cannot hypothesize about specific behavioral patterns that might contribute to pathogen transmission through dirty hands. All pertinent information should be summarized into concise reports for review by the team.

Step 4. Identify Actors, Locations, Times and Events Associated with Specific Behaviors to be Observed

KEY GOALS:

a. Identify actors responsible for key behaviors
b. Identify locations where key behaviors occur
c. Identify times that key behaviors occur
d. Identify events (groups of time-associated behaviors) associated with key behaviors

The previous step was concerned mainly with utilizing the has been "missed"; rather, it has not been performed.

The seasonality issue must also be addressed when making decision about when and how often to observe a household. As mentioned earlier, the ethnographic data should provide information on seasonal variability of activities and behavioral patterns of interest. Depending on the specific research protocol, hypotheses, and project resources, it may be necessary to observe each household during one or more seasons, and within each season, more than one time.

4d. **Identify events associated with key behaviors.** Structured observation events refer to clusters of key behaviors associated with common actors, locations, and time. A meal is an event, and so is the period of time preceding the meal when food preparation occurs. It may not be possible to identify focal events for structured observations. In the sari example, use of the sari for cleaning purposes is likely to occur all day long.

Step 5. Estimate Reactivity and Variability of Key Behaviors

KEY GOALS:

- a. **Select test sites.**
- b. **Plot variation in key behaviors.**
- c. **Compare plots of the same behavior across units.**
- d. **Compare conclusions for the different behaviors with other team members.**

The extent of reactivity and the variability of the behaviors of interest will determine the number of repeated time units (e.g. days or partial days) needed per observation unit (e.g. person or household).

linked may require the observer to follow the actor, while location-linked behaviors may necessitate the observer remaining in one or two key spots for the duration of the observation. In the dirty <u>sari</u> example, most observations would be conducted within the house or compound, but would involve following the mother as she goes about her daily tasks.

4c. **Identify times that key behaviors occur.** The duration of the observation period is a critical issue, which should be based upon knowledge about the occurrence of the events of interest. For example, if the use of the <u>sari</u> as a "towel," to clean dirty hands, dishes, children, or to blow her nose, are the behaviors to be observed, it is necessary to have some idea of when and how often mothers use their <u>saris</u> in this fashion. If the preliminary data suggest this behavior occurs infrequently, then the event may even be missed during a 3-5 hour observation period, and a longer observation period should be planned. The timing of key behaviors should also be assessed during the period of ethnographic data collection. Key behaviors may only occur during certain times of day, therefore structured observations should focus on those time periods. Therefore, if the focus is on <u>sari</u> use <u>before</u> food preparation, then an observation period can be structured around a known meal time. In this case, a data collector could visit the home one hour preceding either the morning or evening meal, and stay until the food is served. If <u>sari</u>-use events do not occur before food preparation within a household, this is not because it by sequence of observation should be made, marking values of the behavior on the vertical axis and sequence # on the horizontal axis (see Figure 3.). Basically, there are two main strategies for quantifying behaviors. First, the investigator may plan to count the number of times per day that a behavior was observed (e.g. washing hands) or sum the number of minutes per day that a behavior was observed (e.g. time spent preparing food). Either of these would yield a continuous measurement. Second, she may be treating each observation as a separate piece of information. For example, each time the child defecates, she records the mother's response, using one or more behavioral items. The specific behavioral item would be coded either dichotomously or multichotomously.

If there are several observations per day, separate the days from each other with extra space on the horizontal axis. Examine each plot, and ask the following questions:

a) Do the responses change with time in a relatively uniform manner? If so, there may be a problem with <u>reactivity</u> to the observer. Discuss this issue with the data collectors and other members of the study team. Are there ways to reduce the possible reactivity? For example, could the investigator spend more time at the site explaining her intentions beforehand? It may be necessary to return to each site and collect additional observations (say, 5 more), then repeat this exercise. Hopefully, a "threshold" point at which reactivity becomes negligible would be found (for further discussion, see prior discussion of 'Reactivity'. The data collection plan would need to include a sufficient number of observations per unit past this threshold point.

b) Do the responses appear to be relatively consistent over time? If so, there may <u>not</u> be a problem with reactivity and the behavior may have little within-unit variability, thus enabling the observer to collect only one or two observations per unit. Refer to Appendix V for calculation methods.

c) Do the responses fluctuate significantly over time? If so, the

In the case of spot checks, the number of observations needed per day also needs to be determined. How can these variables be estimated?

It is essential to conduct preliminary observations in sites similar to those being included in the sample, using the latest draft of the data collection form. During the "scripting" portion of the initial ethnography, the degree of subject reactivity to the presence of the observer can be ascertained. Effort should be made to determine at what point during the observation(s), the subjects stop overtly reacting to the presence of the observer. Identifying the "reactivity threshold" may require repeated visits to the same household. It is also important to determine how the subjects react. Specifically, do they appear to alter key behaviors as well as other behaviors. For instance, direct observation of breastfeeding practices of mothers may elicit some sort of modesty response (eg. turning away from the observer), but may not significantly affect the key behavior (ie. the act of breastfeeding).

5a. Select test sites. A resonable number of sites should be chosen (perhaps 10) and repeated visits made to each site (perhaps 5). An observer should not go to the same site on the same day of the week and if he is doing partial day observations, he should not go to the same site at the same time of day. (Obviously, if the behavior of interest occurs only on certain days of the week or at certain times of the day, the latter caution would not apply).

5b. Plot variation in key behaviors. A series of simple plots for each individual observation unit (site or person) of each key behavior

Step 6. Choose Type of Structured Observation to Conduct

KEY GOALS:

a. Use flow chart to select type of structured observation

6a. Use flow chart to select type of structured observation. The next step is to make a decision about the type(s) of structured observation(s) that will best "capture" the behaviors and events of interest. Figure 4 presents a flow chart to assist in choosing one type of structured observation over the other (continuous monitoring or spot checks). Six yes-no questions must be answered. A positive response to any of the questions results in 1-2 points being added in favor of spot check observations, while a negative response results in 1-2 points added in favor of continuous monitoring.

The first question, "Do the key behavior(s) account for more than one hour of the actor's time per day (based on initial ethnographic assessment)?" is weighted more heavily than the other five. In many cases, this will be the decisive variable, as behaviors that occur infrequently are unlikely to be picked up via spot checks.

The second question, "Is relative amount of time spent on an activity required (versus the actual amount of time)?" relates to the intended use of the dataset. Intervention projects are often concerned with the added burden their promoted behavior or activity might place on the time constraints of adult women. Spot checks are not ideal for calculating the exact amount of time spent on an activity or behavior (although time can be estimated, see investigator may need to evaluate the extent of this <u>within-unit variability</u>. If the investigator does not have at least 10 observations of the behavior per unit, she may need to collect additional observations. Otherwise, proceed as described in Appendix V.

5c. **Compare plots of the same behavior across units.** Attention should be paid as to using the appropriate unit. (i.e. sites or individuals). Do <u>most</u> of the plots have the same pattern (a,b,c)? If not, are there reasons why particular units should have been different? Discuss these plots with the actual observers and reread any unstructured notes that were included with the structured observations for clues. Try to classify the behavior into one of the three patterns. If this seems unreasonable, it may be necessary to collect data on additional units and/or times per unit and repeat these steps. The investigator should also consider the possibility that this behavior needs to be recorded differently, i.e., by revising the data collection form.

5d. **Compare conclusions for the different behaviors with other team members.** Can an overall consensus for the data collection tool regarding how frequently observations should be made at each site be reached? If no consensus is clear, reconsider the behaviors explored so far. Would revisions of the data collection form, such as revising the way in which a particular behavior is recorded, possibly alleviate some of the problems?

determined a sample size, using methods described in the "Sampling Issues" section or is working in a small community and so knows an approximate sample size. The figure of 200-400 households is somewhat arbitrary. Continuous monitoring, which often requires several hours (or more) of observation per visit to each household and multiple visits, is time, money and personnel consuming. Even two hundred households is an extremely large number of households to be continuously monitored, several times each household. Spot checks, on the other hand, are more useful for gathering data on larger samples.

The fifth question, **"Is the study population very heterogeneous (ethnic, economic, religious)?"** is closely related to the fourth question, and should have been determined during the qualitative data collection phase. A study population that is homogeneous requires a smaller sample to achieve representativeness then one that is heterogeneous. Spot checks, because they allow a smaller time input per sample household, permit coverage of a wider range of populations.

The sixth question, **"Are the number of key behaviors to be observed less than 15?"** is concerned with the degree of detail required for the structured observations. The cut-off point of 15 key behaviors is also arbitrary. Structured observations that require observers to identify large numbers of distinct behaviors cannot be done with spot checks, particularly if many or all of the behaviors are similar. Using Table 1 as an illustration, had Gittelsohn required a level of detail depicted in stage one, spot

Data Analysis section). Rather, they give a sense of the relative amount of time spent on various activities.

The third question, **"Is determining other behaviors associated with key behaviors unimportant?"** refers to the context in which the key behaviors occur. If the researchers are interested only in the counting and recording of key behaviors, the what and the how many, then other behaviors associated with the key behaviors are unimportant. If the researcher is interested in knowing why a key behavior occurred, then it is often important to know what happened before and after the key behavior. Spot checks, even in combination with ratings checklists, are unlikely to catch the whys of human behavior.

In the sari example, the best strategy would be to do continuous monitoring focused on when the mother used the sari for any of the specific behaviors (e.g. wiping nose, cleaning child's anus, wiping hands, etc.) within the observation period. In this example, the activities that occur immediately after the event will be of special interest. For example, suppose that, within a particular household, a mother cleans up after her child defecates, rinses her hands in water and then dries them on her sari (which may or may not be dirty). After a short period, she prepares her child's weaning food. The sequence of events that occurs is important information that should be recorded, either through a time notation of the events or a notation of the sequence.

The fourth question, **"Is the study sample greater than 200-400 households?"** assumes the researcher has either already checks (if focused on meal times) might have been sensitive enough to pick up those simple behaviors. However, the level of attention required to discern stage three behaviors require the presence of an observer on an on-going basis (ie. continuous monitoring).

Step 7. Design Instruments and Data Sheets

KEY GOALS:

a. **Operationalize and define key variables**
b. **Design continuous monitoring/spot check instruments**
c. **Develop ratings checklist and observation summary**

7a. **Operationalize and define key variables.** What does it mean to give an operational definition to a variable or code, and why is it a necessary step? When a set of key behaviors has been established, it is necessary to clearly operationalize under what circumstances the behavior will be considered present or absent, and the different forms of the behavior that will be recognized. For instance, investigations of hygiene behavior may be interested in observing hand-washing among members of a group of households. It would be critical to clearly define what constitutes handwashing: a mere rinsing with water, actual rubbing of the hands together while rinsing, the use of soap? In this case, it may be necessary to operationalize several different types of handwashing.

Still other aspects of the handwashing event might be important to record: Did the person wash his own hands or were they washed by someone else (in the case of a mother washing a small child's hands, this might be useful as an indicator of quality of child care)? Where did the hand washing event take place? What water was used for washing hands (ie. drinking cup, water container, etc.)? What was done with the waste water (fed to animals, left on ground)?

7b. **Design continuous monitoring/spot check instruments.** Once key behaviors have been operationalized into measurable (ie. observable) variables, design of the structured observation instruments may begin. General suggestions regarding the proper construction of data collection forms can be found in Appendix VI. There are two main types of instruments for coding structured observations: 1) precoded data sheets and 2) uncoded, but structured data sheets. Both methods make use of a codebook that lists, defines, and operationalizes all variables and codes for the data sheet.

An example of a **precoded data sheet** for structured observations is presented in Figure 5. Key behaviors are printed on the data collection sheet, so that indicating a behavior has occurred involves little more than checking a box. Precoded data sheets are a simple and rapid method of data collection that are well-suited to spot check structured observations. As the number of different behaviors to be recorded must fit on one or two sheets of paper, this method is especially appropriate when a short, well-defined list of key behaviors has been developed. However, the use of precoded data sheets may lead observers to categorize ambiguous behaviors into inappropriate categories, even if blank spaces are left on the data sheet for "other" behaviors.

Figure 6 presents an example of an uncoded, but structured data sheet. Blank spaces are left to be filled in by the observer with the appropriate behavioral codes as they occur. The observer will need to refer frequently (especially in early phases of the data collection) to a codebook. Uncoded data sheets are suited for types of structured observations of complex events where numerous behaviors occur and in a rapid sequence. Great complexity and range of behaviors can be best explored through continuous monitoring structured observations. Uncoded data collection instruments are also recommended when information is required on why a certain behavior has occurred, as the additional behaviors that must be included will be too numerous for a precoded data sheet.

Data should be coded at two levels: the observational episode (the visit) and the key behavior. At the level of the observational episode, items that may require coding are the date and time of the beginning of the observation period, the identification of the observer, the identification of the household and actor(s) observed, and the time of each observation or the sequence of observations. Minimally, at the level of the individual behavior, both types of structured observation require the following data: time the behavior is observed, actors involved, and the behavior observed. Additionally, other kinds of data can be recorded at the behavioral level if needed. For instance, a study of infant feeding behavior might record the type and quantity of food given to a child during a food serving.

7c. **Develop ratings checklist and observation summary.** When the type of structured observation instrument has been selected, a decision should be made about the inclusion of a supplementary ratings checklist. Recall that this is an instrument complementary to that created for the previous goal. The main structured observation instrument involves a relatively objective recording of specific observed behaviors. On the other hand, the supplementary ratings checklist requires more subjective value judgements and general description by the observer. There are two potential components of this instrument: 1) a ratings checklist and 2) an observation summary, and/or expanded qualitative notes.

A **ratings checklist** involves an observer making a qualitative judgement (ie. ranking, rating) based on a set of predetermined criteria for a group of clearly operationalized variables. For example, in the hand-washing example, if hand-washing takes place in an unobservable locale or was relatively infrequent, it might be necessary to develop an assessment of the cleanliness of an individual's hands. In this case, it is important to specify exactly how the measurement of cleanliness/dirtiness will be made, and to define precisely what the variable means. The variable must be defined in terms that are "observable by the senses" (Fisher et al 1983), and be coded in such a way that a data collector can discriminate among all choices.

A decision must be made whether to code the variable "hand cleanliness/dirtiness" as a dichotomous outcome, such as 1 = clean and 2 = dirty, or to develop a scale of cleanliness, such as 1 = most clean, 2 = relatively clean, and 3 = dirty. Each code must be defined in such a way that a data collector would be able to choose, with certainty, the code that conforms to what is observed. A possible operational definition for the dichotomous outcome of a rating of hand cleanliness is shown below:

Are the subject's hands "clean" or "dirty"?

1 = clean	There are no visible stains or dirt on the hands, no dried food or other organic material
2 = dirty	Hands are visibly dirty with soil, stains, or dried food or other organic material

The definitions of each code are clear, and the definitions are mutually exclusive, e.g. the two categories do not overlap. It should be possible for a data collector to choose a code based upon these definitions.

Developing a scaled definition for this rating would be more difficult than the dichotomous code, since it might be difficult to achieve agreement among data collectors. In general, dichotomous outcomes are easier to define and operationalize, and are more reliable for achieving inter-observer reliability (Martin and Bateson 1986).

It is advisable at the end of an extended structured observation (i.e. a continuous monitoring) to have the observer respond to a series of open-ended questions about the observation. This observation summary with expanded qualitative notes can serve many purposes. It can permit the observer to judge his own performance and effect on the subject's behavior. For instance, in Gittelsohn's (1989) study of mealtime behavior in rural Nepal, he had the observers rate their effect on mealtime behaviors (after the meal) along a scale of 0 (no effect) to 5 (great effect) and asked them to describe the reasons for their ratings.

The data sheet for structured observations may also include some structured questions that are to be asked after the observation has been completed. Sometimes it is important to have specific pieces of information that will help evaluate the observation, or that require an informant's reply. For example, when observing dietary intakes, dieticians often inquire, at the end of the day, whether or not the family was celebrating a religious or special day that might influence their daily dietary choice. Additionally, depending on the objectives of the study, at the end of the observation it may be possible for the observer to ask the subject to clarify and explain unfamiliar behaviors, including whether he or she changed his or her behavior because of the presence of the data collector.

Pre-coding of a data sheet does not preclude the opportunity to collect qualitative data. In fact, qualitative notes should be taken if it helps to explain events that have been observed. For example, in the handwashing example, a note might expand upon the appearance of dirty hands: "the mother's hands were covered with what appeared to be dried animal dung." Although this information would not be used for the quantitative analysis, it provides the investigators with contextual and interpretive data.

Step 8. Determine data collection schedule

KEY GOALS:

a. **Determine the number of days of data collection available**
b. **Determine the mean number of observational episodes per day that can be conducted by one observer**
c. **Determine the number of observers required to conduct structured observations**

The key goals in this step represent a series of logistic decisions about how the collection of structured observation data should be conducted. Rough estimates for most of the parameters can be calculated from the formula depicted in Figure 7. Note that the equation can be manipulated algebraically to permit estimates of several different logistic parameters, when the values of the other factors are known or can be estimated. This equation should not be utilized to estimate sample size or the number of observational episodes required per household (see section on "Sampling Issues").

8a. **Determine the number of days of data collection available.** The number of days of data collection available for the structured observation can be estimated using the formula in Figure 7. This parameter will depend largely on program resources and timing. Ideally, behaviors should be examined many times over a long period of time to account for daily, weekly, and seasonal fluctuations and to derive more reliable estimates of frequencies of key behaviors. However, research that is meant to assist in program implementation is likely to be needed quickly.

8b. **Determine the mean number of observational episodes per day that can be conducted by one observer.** The equation in Figure 7 can be manipulated to show the minimum number of observations per day that must be completed to meet sampling conditions, with a given number of observers and a given number of days of data collection available. However, the actual number of observational episodes an observer can conduct in one day should be estimated during the initial ethnography (Step 3) and refined during the pretest of the data collection instruments (Step 8). A general basis for estimating this parameter lies in the type of structured observation to be conducted. As continuous monitoring observations usually require a minimum of several hours per observational episode, it is unlikely that more than 2-3 such observations could be conducted in a single day. Spot checks, on the other hand, have much shorter time requirements, ranging from a few minutes per observational episode to perhaps 15 minutes if a lengthy ratings checklist has been devised. Therefore, 10-20 spot checks are possible in one day. Obviously these estimates will vary depending on the distance between households, means of transportation available, weather conditions and so on.

8c. **Determine the number of observers required to conduct structured observations.** Using the formula in Figure 7 and given estimates of the other parameters, an estimate of the number of observers required for the structured observations can be calculated. A rough estimate can be based additionally on the type of structured observation to be employed. Generally, fewer (2-5), more highly trained observers should be used for continuous monitoring observations, especially when uncoded, but structured instruments are utilized. On the other hand, more (4-10), less highly trained observers can be used for spot check observations, especially when precoded instruments are utilized. Continuous monitoring is a highly intensive type of data collection that requires a lot of personal judgement and initiative from observers. No matter how well key behaviors have been operationalized and coded, new situations will come up during data collection. Observers doing continuous monitoring observations must be able to decide when to write down new, uncoded behaviors or variants of coded behaviors. Thus, there is a bit of "art" to continuous monitoring observations.

Step 9. Train Data Collectors and Pretest Instruments

KEY GOALS:

a. **Involve observers in key behavior formulation and instrument design steps**
b. **Develop field manual and codebooks**
c. **Include observers in pretesting of instruments**

9a. **Involve observers in key behavior formulation and instrument design steps.** Training data collectors (observers) to do structured observations and recording of data can begin with the task of defining variables. Indeed, the data collectors can help to define the mutually exclusive codes for each variable. For example, returning to the handwashing example, one strategy would be to take a group of two or three data collectors to several households to conduct spot observations. Upon arrival to a household, they have been given instructions to write down whether they believe the mother's hands are clean or dirty, and why. Following the household visits, a debriefing session can be done to compare observations and notes. This exercise should provide information on the most feasible operational definitions, and on whether problems may lie ahead for inter-observer variability of the spot observations.

9b. **Develop field manual and codebooks.** Once the data sheets have been developed in a preliminary form, a training period should precede the pretesting of the instruments. A manual should be prepared that outlines the entire protocol, including the timing and frequency of observation periods, rules for how the data collector should present herself and interact with household members, and details for filling in each data sheet, with operational definitions for each variable provided. During the training period, the supervisor of the data collectors should go over each data sheet in detail, explaining how variables and codes are operationally defined, and procedures for recording the data. Data collectors should be encouraged to ask questions during the training period. It is important, however, not to share the hypotheses with the data collectors, and no information about expected or "hoped for" observations or outcomes should be revealed. In other words, the data collectors should be prepared to be as objective as possible before the data collection. Likewise, the data collectors should be instructed on what they may and may not say to individuals who will be observed. Role playing is another observer training technique that could be employed.

The amount of time required for training will depend upon the complexity of the research project, the number of data sheets, variables, and structured observations to be done, and the number and experience of the data collectors. A minimum would be one week of training. An investment in time for training, however, will pay off by producing reliable data for quantitative analysis.

9c. **Include observers in pretesting of instruments.** Once the training exercises have been completed, the data sheets can be pretested in an adjacent community to the project site, or among households within the same community where data will not be collected. Pretesting of the instruments should provide information on any problems with the data sheets in general or specific variables or codes. It is during this period that inter-observer tests for reliability can be implemented.

Step 10. Conduct Reliability Tests to Reduce Inter- and Intra-Observer Variability

KEY GOALS:
- a. Test for inter-observer reliability
- b. Test for intra-observer reliability

10a. **Test for inter-observer reliability.** Structured observation data must be reliable <u>between</u> observers, so that independent observers code behaviors in the same way. A test of the variability of agreement between data collectors should be done prior to formal data collection. The best method for this test is to pair data collectors for observation of the same observational episode, and then to compare codes (Whiting and Whiting, in Mulder and Caro 1985). Each data collector should be paired with each of the other data collectors (presuming there are no more than 4-5 data collectors in total). The paired data collectors should be blinded to each other while recording the events. An alternate method is to prepare a training videotape of a series of behaviors and have the observers record their observations separately. Until there is 80-90% agreement among all data collectors, for all of the variables, the codebooks, data sheets, and training of observers are not final, and data collection should not begin. During this exercise, it will be possible to note which variables or codes cause the most confusion or disagreement, signaling the need to revise operational definitions. It will also be possible to note whether a particular data collector is having difficulty, signaling a need for more intensive training.

10b. **Test for intra-observer reliability.** A second reliability test is of intra-observer consistency over time (Anastasi, in Mulder and Caro 1985). It is possible that an observer will alter his pattern of coding over time. It is recommended that a consistency check be done at regular intervals, with increased surveillance during the beginning of data collection. Consistency checks can assess whether the observer is including many new codes or omitting old codes. Significant variation in ether direction is cause for concern.

Step 11. Implement Data Collection and Data Management

KEY GOALS:

a. **Conduct short pilot study and determine strategies to reduce reactivity**
b. **Review data sheets and store properly**
c. **Initiate data entry early**

11a. **Conduct short pilot study and determine strategies to reduce reactivity.** When the data sheets are finalized and reliability tests have been completed, data collection can be implemented. It is recommended that a short pilot study be done on non-sample households or in a contiguous community. A pilot study will identify unforeseen problems with the protocol or with the data sheets, and revisions can be made before data collection on the sample begins.

If it has not been established in the pretesting period, this phase can be used to establish a reactivity threshold for the key behaviors in question. Data collectors should be asked whether they perceive possible changes in behavior because of their presence, and what strategies should be utilized to minimize subject reactivity. However, it is likely that reactivity will only be reduced by time, and that the first observations in a household will be highly reactive. If a multitude of spot check observations are being conducted, the first few observations per household could be eliminated from the sample (up to the reactivity threshold). If continuous monitoring observations are being conducted, one possible way to reduce reactivity is to conduct one or two short (1/2 - 1 hour) "test observations" in each household.

11b. **Review data sheets and store properly.** During the first weeks of data collection, data sheets should be reviewed nightly by the field supervisor, and consistency checks done on a weekly basis. In addition, data collectors should be queried about any difficulties they may be experiencing in making observations or recording data. Data sheets should be stored in an area safe from rain, insects and other forms of damage, and should be entered directly onto the computer or copied as numeric codes on columned coding paper if direct data entry is not possible. This last method is flawed in that the copied data are subject to transcription errors.

11c. **Initiate data entry early.** If the data are recorded on pre-coded sheets, data entry onto a computer can and should begin as soon as possible. Numeric data should be double-entered onto a computer and the two datasets compared for inconsistencies. A more detailed discussion of alternate data entry methods is presented in Appendix VII. Preliminary data analysis can identify variables with a low frequency or a narrow distribution, and a decision can be made whether or not to continue observing and recording these events. For example, if mothers are never observed to cover the jug in which water is stored, then jug covering, a likely behavioral risk factor for diarrheal disease transmission, may not need to be observed further. This information about the variability of specific behaviors, however, should have been discovered during the qualitative research, avoiding the need to "throw away" useless data.

Step 12. Clean Data Set (s)

KEY GOALS:

a. **Conduct range checks.**
b. **Conduct consistency checks.**

Once the data has been entered and the investigator is satisfied with the degree of accuracy, the data should be read using a statistical software package. There are two basic steps in data cleaning: range checks and consistency checks.

12a. **Conduct range checks.** Obtain simple frequency distributions for every item on the data file. The researcher should be able to define a set of possible correct values (or codes) for each item. For example, if an item was supposed to be coded "1=none 2=some 3=all 8=not applicable 9=unknown", the researcher should circle any occurrences of values other than 1,2,3,8, or 9 on the printout. Sometimes with continuous measurements, it may be difficult to define what values are "out-of-range". In such instances, it is best to use a reasonable guess and to flag values outside the lower and upper boundaries. Once the questionable values have been marked, the researcher needs to identify the records containing those values. Almost any statistical software package will allow this type of investigation.

Once the "bad" records have been identified, you can "pull" the original paper records. If the questionable code is simply a data entry error, just write down the correct value. Otherwise, see if you can tell from the "context" (i.e., related items), what the correct code would be. In the case of suspicious "out-of-range" values, you may get a sense of the believability of that value from the "context". If the investigator decides on a new code, write it down. A last resort is to recode all suspicious items as "unknown".

The researcher is now ready to make changes to the data file. It is a good idea to return to the data entry software to make these corrections rather than to make the changes through the statistical software.

12b. **Conduct consistency checks.** In certain situations, there will be items in the data collection form that can be cross-checked against each other. For example, there may have been recorded an answer to "Where is the mother?". If the mother was "away" and could not be observed, the answer to each of three items related to the way in which the mother responded to the child when he defecated should be coded "not applicable". Any other code would be inconsistent. Go through the form carefully to look for items that can be cross-checked in this manner and write out the correct relationships that are possible between the codes of the items. A program can then be written using the statistical software package that will identify any records with inconsistently coded items. Pull the original paper records and evaluate each occurrence. If it is not clear where the error is, both items involved may need to be recoded as "unknown". Return to the data entry software to make any new corrections.

Step 13. Operationalize New Variables

KEY GOALS:

a. Determine frequencies of key behaviors.
b. Determine amount of time spent on key behaviors.
c. Create behavioral scales or scores.

Structured observations gather data in the form of numerous behavioral units, differentiated by time, space, specific action involved and the actor performing the behavior. At the level of the individual behavioral unit, very little in the way of analysis can be performed, other than a listing of the different types of behaviors. Thus, prior to data analysis, it is important to create new composite variables based on the combination of discrete behavioral units. Note that in this discussion, the unit of analysis is the individual, although it could be a household, clinic, etc..

13a. Determine frequencies of key behaviors. An initial simple type of variable to create is based on the counting of key behavioral units identified in Step 3b. If the data are recorded in a continuous observation, it is assumed that each person is observed for approximately the same amount of time. If a few people are observed for a different length of time compared to the larger sample, those records should be deleted from the analysis. If a large number are so affected, it is be necessary to make an adjustment to reflect the length of time. In this case, rates could be used instead of simple counts (e.g. # times behavior X was observed / 12 hours for person i).

When time-sampled spot observations are used, it is important to note that the observation is of a period of behaviors rather than a behavioral event. Hence, depending on the frequency of the spot checks and the duration of the state, the same event could reoccur (both begin and end) between spot checks, resulting in an undercount. This potential bias is likely to be constant across people, however, so that the analysis can proceed with comparisons between people.

Furthermore, it is assumed that each person is "spotted" the same number of times. If this is not true, proportions should be used instead of simple counts (e.g. # times behavior X was observed / n_i where n is the # of spot checks for person i).

In general, counts of behaviors form the basis for the estimation of time allocation of household members to different tasks, e.g. what proportion of a woman's time is spent on domestic tasks, what proportion on agriculture, and so on. However, counts can have more specific foci as well. For instance, an infant feeding project concerned with the feeding of different types of foods to infants could conduct four six-hour observations on 30 households of children 6-12 months. They may end up with a table of the following counts of key behaviors:

Type of Food:	Count:	Percent:
Breastmilk	266	42%
Other Milks	175	27%
Paps, Porridges	142	22%
Family Diet	56	9%
	639	100%

From these counts it appears that, on the average, breastmilk is more frequently given to infants in that age range than other foods. Other milks and paps, porridges are given with similar frequency and foods from the regular family diet are infrequently given.

13b. **Determine amount of time spent on key behaviors.** For continuous observations, it is a simple matter to calculate time spent in a particular activity if the starting and stopping times are recorded each time the activity is observed. Such recordings allow the investigator to calculate duration for each observation of the activity, which can then be summed together over the observation period to yield total time spent in activity X. The mean duration of activity X may also be computed by dividing the total time spent by the number of times the activity is observed. In both cases, it is assumed that observation periods are of approximately equal length for all people. If only a few people are observed for a much shorter or much longer period of time, those records should be deleted from the analysis. Otherwise, a proportion should be substituted for simple time spent (e.g. total minutes in activity X

/ total minutes observed) and the mean duration should be analyzed separately for subsets of people with similar observation periods (e.g. if people were observed 5-16 hours, they could be divided into subsets of people with 5-8, 9-12, and 13-16 hours of observation).

When using time-sampled spot checks, the assumption is usually made that the proportion of checks during which activity X is observed directly estimates the proportion of time spent in the activity. For example, is spot checks are done every 30 minutes over a 10-hour period (hence, 20 spot checks) and the mother is "spotted" caring for her child 8 times, the proportion of time spent caring for the child is estimated as 8/20, or 40%. it must be recognized that, using time-sampled spot checks, activities of short duration will tend to be undersampled and activities of long duration will tend to be oversampled. Therefore, estimates of the proportion of time spent on activity X may be biased, but the bias will usually be constant across people, allowing comparisons between people.

Clearly the duration of a behavior is as important a characteristic as how often it occurs. Using the infant feeding example described above:

Type of Food:	Mean # Minutes Per Feeding:	Mean # Minutes Per Day:
Breastmilk	8.6	57
Other Milks	2.3	12
Paps, Porridges	5.6	34
Family Diet	6.1	7

In this example it appears that the contribution of paps and porridges to the infant's diet may be larger than originally perceived based upon key behavior counts. Of course, it may also mean that it takes longer to feed paps and porridges rather than other foods.

13c. **Create behavioral scales or scores.** At times, information may be collected on a set of related behaviors which the investigator intends to combine into a scale that is a proxy for a characteristic that cannot be measured directly. For example, Bentley et al. (1989) collected information on specific observable behaviors of young children and their mothers during feeding episodes throughout a 12-hour observation day. The 4 child items and 5 mother items were combined into scales measuring "child's acceptance of food" and "mother's encouragement to eat", respectively. Following are brief descriptions of some methods for combining related items.

Simple additive scales can be created by combining frequencies of two or more categories of related behaviors. Combining related measures eliminates redundancy and reduces the number of variables used in the final analysis. In addition, combining several unreliable measures can sometimes yield a single, reliable measure (Kraemer, 1979). Picking the behaviors to be included in an additive scale is often done intuitively, or on the basis of other knowledge (Martin and Bateson 1986).

These scales are "unweighted," i.e., each item is considered to be equally important. The scale should always be evaluated using a reliability analysis (see Appendix VIII). Such methods allow you to discern "weak" items which can dilute the strength of your scale. The investigator may find, however, that no combination of items available will produce a reliable scale. If so, it may be due to the fact that the items are related to each other in an ordered way (see Guttman scaling) or may need to be weighted (see Factor Analysis). The former was the case for Bentley's feeding behavior data, where reliability as measured by a KR-20 statistic was never sufficiently large, but a very strong scale was produced under the assumption of ordered relationships among items, tested with the Guttman procedure.

Guttman scaling are used in the case of a set of related dichotomous items, when the investigator is able to hypothesize an ordered relationship of the items to each other. If so, take advantage of this special "pattern" to increase the strength of the scale. One method of doing this was developed by Guttman, whose method orders "both items and subjects with respect to some underlying cumulative distribution" (McIver & Carmines, 1981). In a perfect Guttman scale, an individual's response to each scale item can be predicted by the total scale score. To illustrate, a perfect Guttman scale with 4 items would contain the following possible patterns and scores:

	Item				
Subject	1	2	3	4	Score
A	1	1	1	1	4
B	1	1	1	0	3
C	1	1	0	0	2
D	1	0	0	0	1
E	0	0	0	0	0

Hence, in this perfect Guttman scale, a score of 3 can only be obtained when items 1, 2, and 3 are all positive, and so forth. Since it is unlikely that such "perfect" relationships will be encountered, techniques have been developed to evaluate the extent of "imperfection" in the scale, allowing a judgement of its usefulness. Refer to the McIver & Carmines publication for an in-depth presentation. [Note that these procedures are available with some statistical software, including mainframe SAS).

Factor analysis is the final type of scaling that will be discussed. A factor can be regarded as a unifying construct, not directly observable, which can be derived from the measurement of a set of directly observable related items, Factor analysis is "a broad category of approaches to conceptualizing groupings (or clusterings) of variables and an even broader collection of mathematical procedures for determining which variables belong to which groups" (Nunnally, 1978). Nunnally further distinguishes between component factors, which can be dirrectly derived from the data set, and common factors, which can only be estimated from the data set. In general, component factors are more commonly discussed and the following comments refer to that classification.

Factor analysis may be useful when there is a relatively large number of related items, and either: 1) the investigator has hypothesized a scheme for categorizing these items into a few "constructs" or "factors" and wish to test hypotheses; or 2) the investigator wishes to mathematically derive a few "factors" from the available items which can be subsequently evaluated for their intuitiveness. The former situation is considered "confirmatory" analysis, while the latter may be called "exploratory" analysis.

In either case, the basis for the mathematical techniques is a set of standardized items (mean=0 and variance=1) available for each person (or household, village, etc.) from which linear combinations can be determined. Hence, there is an implicit assumption of "normality" of the items, i.e., the items should all be measured at the interval level and be more-or-less normally distributed. In practice, non-normal interval and ordinal items are often included. There are a variety of methods that can be employed to determine the linear combinations, or factors, which differ in terms of their weighting schemes. The ultimate goal, in any case, is to reduce the items to a manageable number of factors for which each person can then be assigned a score.

Usually, a score derived by one of the above approaches (or others) is assigned to each observation. For example, in Bentley's study, the child was observed for particular behaviors at each feeding episode. Using Guttman scaling procedures, a score representing "acceptance of food" was assigned for each feeding episode. If the preferred unit of analysis is the day or, perhaps, a set of roughly adjacent days, it would be easiest to obtain one "score" per person per day (or set of days).

This can be accomplished in a situation in which the number of observations per person per day (or set of days) is roughly equal. Thus, if each child was observed about 6 feedings per day for 3 adjacent days of diarrhea, we could use an "average" of his 18 feeding scores to represent his usual food acceptance level during diarrhea. This average could be a simple mean, median, or mode, depending on the distribution of the scores assigned at the observation level.

Unfortunately, a situation may be encountered in which the number of observations per person varies considerably. In Bentley's study, the number of feeding episodes per child per day ranged from 1 to 11. This marked variability was in large part explained by the age range of 4-36 months (i.e. younger children had fewer non-breastmilk feedings), as well as the fact that some children were offered many small meals, while others were offered a few large meals. The decision was made to analyze this data set at the meal level, rather than the day (or set of days) level, and to statistically adjust for the varying number of observations per child in the final analysis step.

Factor analysis is more likely to be applied to a set of day-long or period-long related behaviors. For example, counts of specific maternal behaviors over a 12-hour day could be used as the items for a factor analysis performed to delineate a few meaningful constructs, each of which will yield a daily score.

Step 14. Conduct Data Analyses

KEY GOALS:

a. Determine if key behaviors and created variables are to be used as dependent or independent variables or both.
b. Do descriptive analyses.
c. Do bivariate analyses.
d. Do multivariate analyses.

Methods of data analysis to select depend primarily on the purpose of the structured observations, specifically, whether these observations are intended to be used as explanatory or predictive variables, or as the outcome of interest. The type of analysis used will also vary according to the number of variables being examined and the level of measurement of the variables (i.e. nominal, ordinal, interval, ratio).

In risk factor or project design studies, there is an "outcome" of interest, e.g. diarrhea incidence, to which the investigator would like to relate specific behaviors. In this example, the analysis plan is largely determined by focusing on the best way to obtain and examine diarrhea incidence, so that the structured observations become a secondary issue.

When structured observations are used for project implementation or evaluation, these observations may become the "outcome" of interest. For example, suppose an investigator determined that mothers' hand washing methods were associated with diarrhea incidence in their infants. A health education intervention could focus on teaching mothers how and when to "properly" wash their hands, followed by an evaluation study using structured observations of mothers in households with and without the intervention. In this case, the primary focus is on the best way to obtain and examine hand washing events.

Note that this step describes data analysis issues in general terms. Specifics and actual methods of calculation can be found in Appendix IX of this document. The key goals in this step are ordered on the basis of increasing methodological complexity. This is the one step in the guidelines where it is not necessary to complete all of the key goals.

14a. Determine if key behaviors and created variables are to be used as dependent or independent variables or both.

What is the outcome of interest? If it is a key behavior (or a new variable created from a set of key behaviors), make a list of all of the variables which you would like to examine in relation to the outcome. If the key behavior is not the outcome, list the outcome(s) and the key behavior.

14b. Do descriptive analyses.

Examine the frequency distribution of each variable on your list, including the outcome. Determine the level of classification of each variable:

Nominal - dichotomous or multichotomous?

Ordinal - Are there relatively few categories (e.g. 5 or less) or many (may be able to treat as interval)?

Interval/Ratio - What does the distribution look like? Is there evidence of digit preference or more than one modal value? Is there an especially long tail (on either the left or right side)?

14c. Do bivariate analyses.

Are there repeated measurements on the same observation unit (i.e. household or person) over time? If yes, are you trying to make statements regarding changes over time between or among groups or simply within the entire study sample? See Appendix IX sections for "Repeated Measures - single Group" or "Repeated Measures - 2 or more groups". If there are no repetitions, see Appendix IX section for "Non-repeated measures".

14d. Do multiple varible analyses.

Are you interested in examining the simultaneous effects of a set of independent predictors or explanatory variables on the outcome? Refer to Appendix IX section for "Multiple variable models".

Table 1: Refinement of Food-Serving Behaviors in Rural Nepal

STAGE I: (Based On Literature)	STAGE II: (Based on Key Informant Interviewing)	STAGE III: (Final, Based on Analysis of "Scripts"):
Serves Food	Serves Food	Serves Food, without Asking
		Server Asks, then Serves Food
		Server Asks, but Recipient Refuses Food
		Server Shares Own Food
	Forces Person to Eat Food	Forces Consumer to Eat Food (already served)
		Forces Consumer to Eat Food (new)
Serves Self	Serves Self	Serves Self (from family pot)
		Serves Self (from other's plate)
Asks for Food	Asks for Food, and Receives Food	Asks for Food, and is Served Food (from family pot)
		Asks for Food, and is Served Food (from other's plate)
	Asks for Food, and is Refused	Asks for Food and is Refused (discriminatory)
		Asks for Food and is Refused (non-discriminatory)
Breastfeeds infant	Breastfeeds infant or child	Breastfeeds infant or child

Figure 1: Example of a Ratings Checklist for Household Sanitation

Format: The observer approaches the household and immediately records ratings on household sanitation

GENERAL CLEANLINESS RATINGS:

Definitions:

1 = Clean, no visible stains or dirt on object/area observed, no dried food or other organic material

2 = Moderate, visible stains or dirt on object/area observed, possibly some dried food or other organic material

3 = Dirty, a lot of stains or dirt on object/area observed, presence of dried food or other organic material

Instructions: Observer circles appropriate observation code

CLEANLINESS RATING:			OBJECT/AREA OBSERVED
Food Preparer/Child Caretaker:			
Clean	Moderate	Dirty	
1	2	3	Hands
1	2	3	Clothing
Target Child:			
Clean	Moderate	Dirty	
1	2	3	Hands
1	2	3	Clothing
1	2	3	Rest of Body (incl. face)

Inside House:

Clean	Moderate	Dirty	
1	2	3	Cooking Area Floor
1	2	3	Cooking Area Countertops/Tabletops
1	2	3	Eating Area Floor
1	2	3	Eating Area Tabletops
1	2	3	Food Storage Containers--Outside
1	2	3	Food Storage Containers--Inside
1	2	3	Water Containers--Outside
1	2	3	Water Containers--Inside
1	2	3	Cooking Pots and Utensils
1	2	3	Serving Dishes and Eating Utensils

Outside House:

Clean	Moderate	Dirty	
1	2	3	Adult Defecation Area
1	2	3	Courtyard
1	2	3	Perimeter of House Excluding Courtyard
1	2	3	Animal Shelters/Pens
1	2	3	Perimeter of Animal Shelters/Pens

Communal Areas (may be observed seperately from household observations):

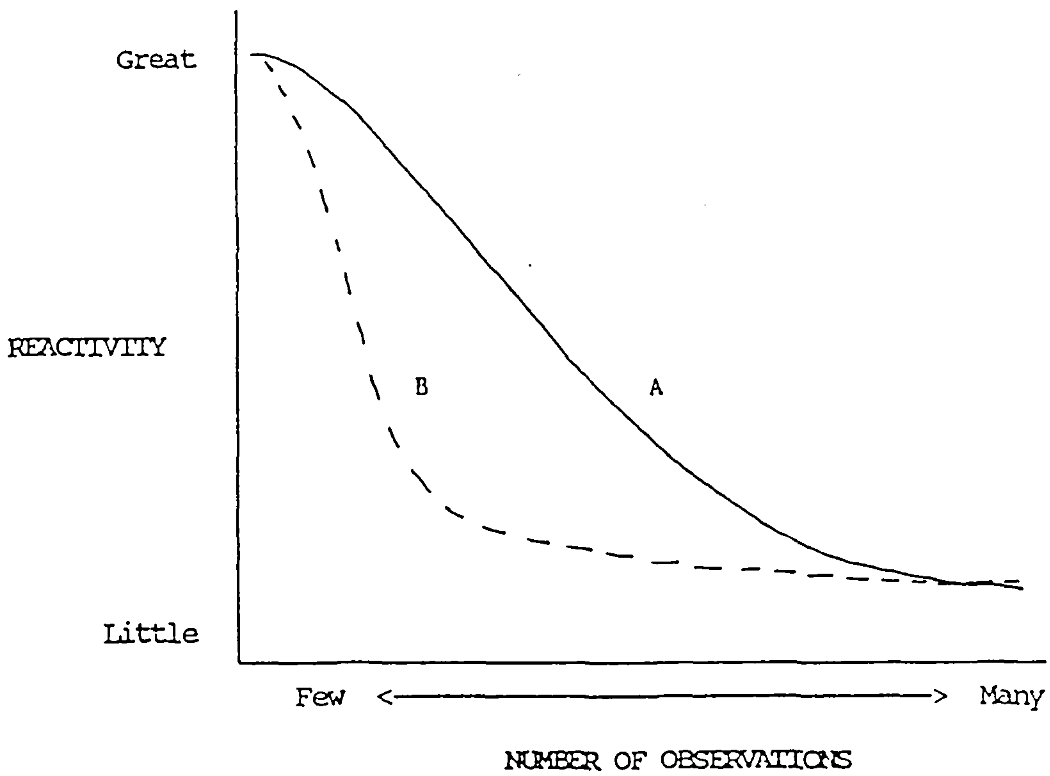
Clean	Moderate	Dirty	
1	2	3	Adult Defecation Area
1	2	3	Place Where Water is Collected
1	2	3	Place Where Laundry is Done
1	2	3	School Play Areas

SPECIFIC CLEANLINESS RATINGS:

Instructions: Observer answers a series of yes-no questions about sanitary conditions inside and outside the household, responses are recorded by checking the appropriate box.

YES	NO	
_____	_____	Are the water containers covered?
_____	_____	Are there human feces on the interior of the house?
_____	_____	Are there human feces within 20 feet of the exterior of the house?
_____	_____	Is there soap visible in the house?
_____	_____	Is uncovered cooked food plainly visible in the house?
_____	_____	Are flies or other insects plainly visible in the cooking/eating areas?

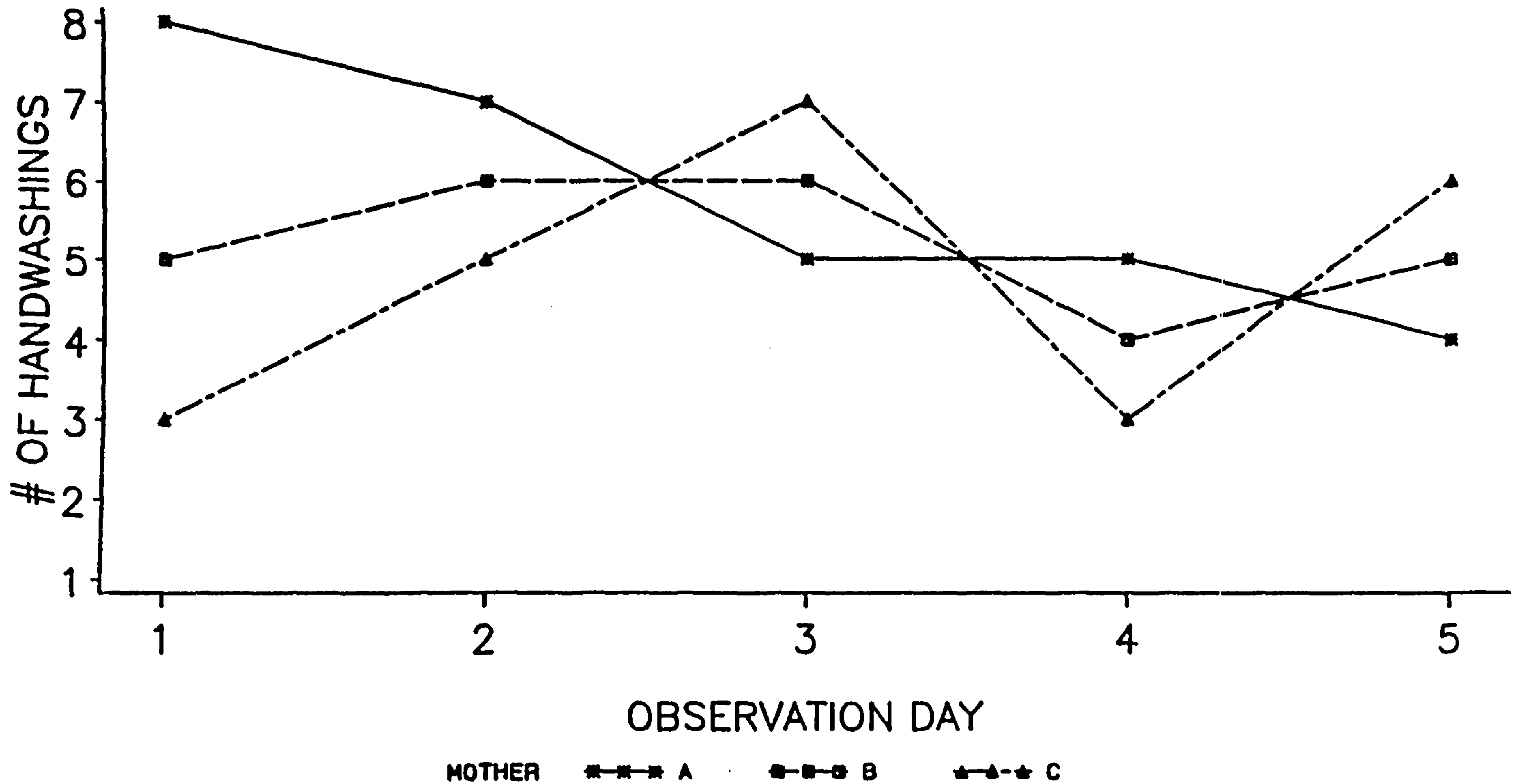
Figure 2: Hypothesized Effect of Repeated Observations on Subject Reactivity



A. Gradual decrease in reactivity with number of observations

B. Rapid decrease in reactivity with number of observations (observeable reactivity threshold)

FIGURE 3.
EXAMPLE: EXAMINING A PLOT FOR REACTIVITY
COUNT OF MOTHER'S HANDWASHINGS BY OBSERVATION DAY



PATTERNS: MOTHER A IS REACTIVE; MOTHER B IS CONSISTENT;
MOTHER C IS FLUCTUATING

Figure 4: Flow Chart to Assist in Selecting Type of Structured Observation

Do the key behavior(s) account for more than one hour of each actor's time per day (based on initial ethnographic assessment)? ---No---> Add 2 points to Continuous Monitoring

Yes

Add 2 points to Spot Check

Is relative amount of time spent on an activity required (versus actual amount of time)? ---No---> Add 1 point to Continuous Monitoring

Yes

Add 1 point to Spot Check

Is determining behaviors associated with key behaviors unimportant? ---No---> Add 1 point to Continuous Monitoring

Yes

Add 1 point to Spot Check

Is the study sample greater than 200-400 households? ---No---> Add 1 point to Continuous Monitoring

Yes

Add 1 point to Spot Check

Is the study population very heterogeneous (ethnic, economic, religious)? ---No---> Add 1 point to Continuous Monitoring

Yes

Add 1 point to Spot Check

Are the number of key behaviors to observe less than 15? ---No---> Add 1 point to Continuous Monitoring

Yes

Add 1 point to Spot Check

Total Points for Spot Check & Select Structured Observation Method:	0	1	2	3	4	5	6	7
	V	V	V	V	V	V	V	V
	CM	CM	CM	CM/SC	CM/SC	SC	SC	SC

CM-continuous monitoring

SC-spot check

CM/SC-either or both

Figure 5: Example of a Precoded Data Sheet Used for Spot Check Observations

FORM 13: TIME ALLOCATION SPOT CHECKS PAGE: ______

1. Village/Ward: ______________
2. Neighborhood: ______________
3. Household ID No.: ______________
4. Household Name: ______________
5. Observer: ______________

DATE A: ______ TIME A: ______

DATE B: ______ TIME B: ______

DATE C: ______ TIME C: ______

DATE D: ______ TIME D: ______

Observation No.: __ __ __ __

ACTIVITY	HOUSEHOLD MEMBER ID: ID No., Name, Rel to MHH											
	A	B	C	D	A	B	C	D	A	B	C	D
Observer/Informant												
Cooking Food												
Serving Food												
Eating Food												
Breastfeeding												
Collecting Fodder												
Herding Animals												
Ploughing												
Harvesting												
Construction Work												
In School												
Washing Dishes												
Washing Clothes												
Getting Water												
Watching Children												

Please Turn Over --->

FORM 13 CONTINUED

ACTIVITY	HOUSEHOLD MEMBER ID: ID No., Name, Rel to MHH											
	A	B	C	D	A	B	C	D	A	B	C	D
Collecting Firewood												
Bathing												
Sleeping												
Conversing												
Not Known												
Gone to:												
Gone to:												

Instructions:

Date and time of observation are selected randomly. This form permits the collection of four seperate observations on three members of a single household. Households larger than three members will require additional forms. Household composition data (including ID number, name, and relationship to male head of household (MHH)) are derived from an earlier household survey.

Upon arrival at the household for the first observation on a form, record the date and time in the lines corresponding to DATE A and TIME B. Determine the activity of each household member. If you are able to determine the activity of the household member through direct observation, write your initials in the first row of boxes (Observer/Informant) under the appropriate column (A, B, C or D depending on observation number). If you are unable to directly observe the activity of the household member, ask another present adult household member where the individual is and what s/he is doing. Record the ID number of the individual who gave you the information in the appropriate Observer/Informant row-column. Now mark the box of the observed or reported activity in the appropriate row-column for the individual. If the individual is performing an activity which cannot be categorized under one of the headings given, write the name of the activity in one of the blank boxes provided.

Figure 6: Example of an Uncoded, but Structured Data Sheet Used for Continuous Monitoring Observations

FORM 15: Direct Observation of Nepali Meals DATE: ______ PAGE: ___

1. Village/Ward: ______________ 4. Observer: ________________

2. Household ID No.: __________ 5. Observation No.: __________

3. MHH Name: __________________ 6. Meal Observed: _____________

OBS	TIME	ACTOR ID	RECIP ID	ACTIVITY CODE	FOOD	CONDITION	QUANTITY /UTENSIL

INSTRUCTIONS FOR FORM 15:

1. Arrive at sample household.

2. Ascertain if majority of household members are present and the meal uneaten. If either of these conditions are not met, move on to another sample household.

3. Record the initial activity of each household member.

4. Record the type and quantity of already cooked foods.

5. If the morning meal is being observed, or the evening meal without an associated 24-hour recall, perform a short-term recall on each household member (i.e. foods consumed in the past three hours).

6. Record all activities of interest, especially food-related.

   As each KEY BEHAVIOR occurs, record the TIME it occurs, the identity of the individual performing the behavior (ACTOR ID), and the person who is the subject of the activity (RECIP ID). The key behavior itself should be recorded under the ACTIVITY CODE column with codes drawn from the codebook. If foods are involved, the type of FOOD is recorded as well as it's CONDITION (burnt, fresh, raw, etc.) and the QUANTITY served.

7. Record conversations of interest, especially if food-related.

8. Leave only after meal is finished and clean-up has begun.

Figure 7: Formula for Calculating Data Collection Parameters

$$\text{A. Number of Observers Required} = \frac{(\text{B. No. households[2] in sample}) \times (\text{C. No. observational episodes required per household})}{(\text{D. No. days of data collection available}) \times (\text{E. Number of observational episodes per day by one observer})}$$

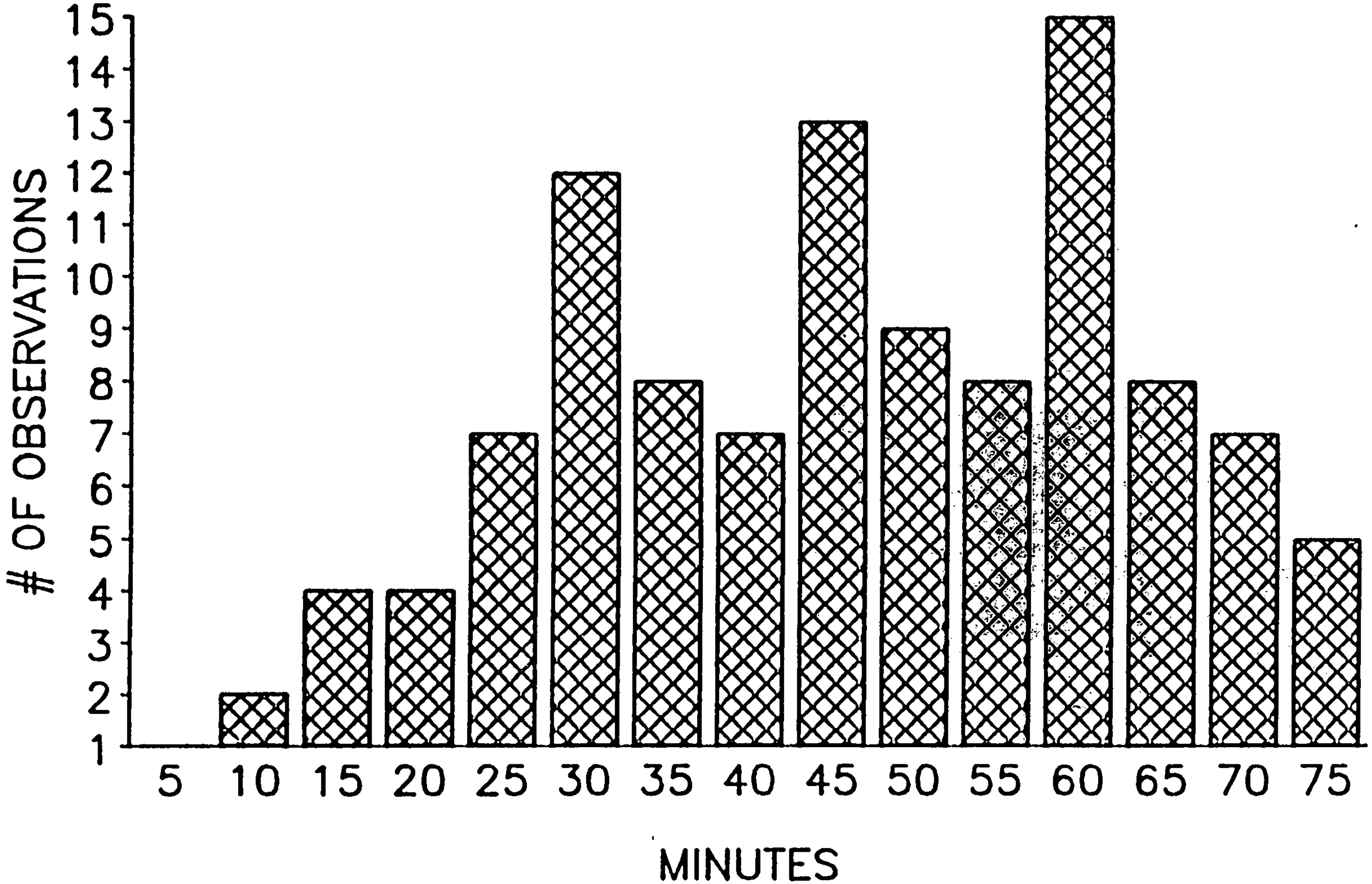
$$\text{D. Number of days of data collection required} = \frac{B \times C}{E \times A}$$

2 The location of the observational episode is described here as the household. Clearly, many other places would be appropriate locales for conducting structured observations (eg. hospitals, schools, etc.).

FIGURE 8.

EXAMPLE: EXAMINING A FREQUENCY DISTRIBUTION FOR IRREGULARITIES

MINUTES SPENT PREPARING MID–DAY MEAL



REMARKS: DIGIT PREFERENCE SEEN AT 30, 45, AND 60 MINUTES

APPENDIX I: SAMPLING METHODS

Simple Random Sampling

"In its simplest form, random sampling means that every member of the population has an equal chance of appearing in the sample, independently of the other members that happen to fall in the sample." (Snedecor and Cochran 1967). The underlying assumption is that the sample selected at random will be "representative" of the population from which it was chosen.

Sampling may be done with or without replacement. Sampling without replacement means that, each time a unit is chosen to be in the sample, that same unit is removed from the remaining pool of candidates. This is the usual procedure in research studies. The sample is generally drawn using a table of random digits or, when available, through a computerized selection algorithm. Formulae and examples for calculating sample size for structured observation studies have been provided in Appendix II.

Stratified Random Sampling

In stratified random sampling, "every member of the population has a <u>known</u> probability of coming into the sample, but these probabilities may not be equal or they may depend, in a known way, on the other members that are in the sample" (Snedecor & Cochran, 1967). Basically, the population of interest is divided into strata based on one or more criteria. For example, the villages in a rural area may be divided on the basis of access to piped water, and further divided within the villages on the basis of access to a latrine. Within each of these four strata, the investigator would randomly choose n_h households for his sample. The number of units per stratum (n_h). may be the same in all strata (known as <u>proportional allocation</u>) or may reflect both the standard deviation of the units within the stratum and the cost per unit of sampling within the stratum (known as <u>optimum allocation</u>).

There are several advantages to stratified sampling. First, if the population is heterogeneous, precision is gained in estimating the mean of the outcome variable by dividing the population into more homogeneous strata. Second, the size of the sample drawn from any stratum can be controlled, focussing more attention on groups of special interest which might not otherwise be well represented in a simple random sample. Third, it may be necessary at times to use different techniques for composing lists of units in different settings, making it awkward to combine the lists for sampling puposes. Stratified in this way, units at random can be drawn from within each setting.[3]

Cluster Sampling

In cluster sampling, the unit being sampled is a cluster of elements to be observed. For example, a unit may be a village comprised of many households, where the actual observations will take place at the household

---

[3]For more detail and formulae for sample size requirements, refer to Snedecor and Cochran, 1967, or another comprehensive statistics text.

level. Clusters generally consist of elements in close physical proximity. From a list of all potential clusters, a sample of clusters is drawn, usually by either simple random sampling or stratified random sampling. All eligible elements (e.g. households) within the cluster unit (e.g. village) are included in the sample.

When cluster sampling is used, the probability of an element of a cluster being selected is equal to the probability of that cluster being selected. The advantages of cluster sampling are: 1) there is no need to make a listing of all elements (e.g. households) prior to sampling, only a listing of clusters and their approximate sizes; 2) since elements are usually in close physical proximity within the cluster, it is usually more efficient and less costly to carry out the field work.

The disadvantages are: 1) a cluster sample will, in general, yield less precise estimates than a simple random sample of the same size; the variance increases with the average cluster size; 2) the analysis of cluster data is more complex, if done correctly.

More explicitly, it is assumed that the behaviors to be observed are more similar for elements within a cluster (e.g. households) than for elements from different clusters (e.g. villages). Technically, the correct level of data analysis under cluster sampling is the cluster rather than the element which is actually observed. If the analysis is done at the level of the element, as is usually the case, the power to detect an effect is reduced.

The extent to which behaviors within clusters are similar relative to behaviors between clusters largely determines both the number of clusters and the number of elements per cluster needed to provide an adequate sample size to test the hypotheses of interest (e.g. "exposed" vs "non-exposed" or "intervention" vs. "control" differences in behaviors). Overall, the best "rule of thumb" is to keep the number of clusters large and the average cluster size small[4].

---

4 Refer to Donner et al. (1981) for further discussion and formulae for sample size requirements when randomizing by cluster and comparing either means or proportions.

APPENDIX II: SAMPLE SIZE CALCULATION

GENERAL SAMPLE SIZE FORMULAS FOR SIMPLE RANDOM SAMPLES:

A. Sample size to estimate a population mean.

You would like to estimate the population mean, μ, within a certain number of units, L, of the scale on which it is measured, with a certain probability, Z_α. Then the size of the sample needed is given by:

$$n = Z_\alpha^2 \sigma^2 / L^2$$

where σ^2 is the population standard deviation. (The latter must be guessed to the best of your ability, usually based on the literature or pilot studies).

Example.

You are interested in estimating the mean number of times per day that mothers wash their hands with 95% certainty. A pilot study estimated the standard deviation at 2.0. You wish to estimate the number of times to within +/- 1 hand washing. Substituting these values,

$$n = (1.96)^2 * (2.0)^2 / (1)^2$$

yielding n=15.37, or n=16.

B. Sample size to estimate a population proportion.

You would like to estimate the population proportion, p, whose standard deviation is $\surd$ (pq/n), with a probability of Z_α. The allowable error, L, is $Z_\alpha * \surd (pq/n)$, so that:

$$n = Z_\alpha^2 \, p \, q / L^2$$

Example.

You are interested in estimating the proportion of mothers who wash their hands "correctly", using your own definition, with 95% certainty. A pilot study estimated p to be 0.3, and you wish to estimate the proportion to within +/- 0.05. Substituting, you obtain:

$$n = (1.96)^2 * (0.3) * (0.7) / (0.05)^2$$

yielding n=322.69, or n=323.

[NOTE: This is a large sample approximation. Please refer to statistics texts for other formulae if the n you compute is more than 10% of the population of interest].

C. Sample size to compare means of two independent samples.

[NOTE: It is assumed that the reader if familiar with hypothesis tests, and understands the concepts of α and β, as used below. If not, please refer to any basic statistics text].

You are interested in testing the hypothesis that the difference in means between two groups is at least δ units. You wish to test this hypothesis with a power of $1 - \beta$ and significance level α. Then

$$n=(Z_\alpha + Z_\beta)^2 * 2\sigma^2 / \delta^2$$

Again, you must guess a value of σ^2. Note that Z_α is the normal deviate corresponding to either a one- or two-tailed probability α, whereas Z_β corresponds to a two-tailed probability β.

Example.

You wish to test the hypothesis that the number of daily hand-washings by the mother is significantly different by +/- 2 hand-washings in households in which there is a "low" incidence of diarrhea compared to those in which the incidence is "high". A pilot study estimated σ to be 2.0, and you have chosen to do a two-tailed test with α=.05 and $1 -\beta$=.90. Substituting, you obtain for each group:

$$n=(1.96 + 1.282)^2 * 2 * (2.0)^2 / (2)^2$$

yielding n=21.02 or n=22 per group.

D. Sample size to compare proportions from two independent samples.

You are interested in testing the hypothesis that two proportions, p_1 and p_2, are significantly different, at a signficance level of α and power $1 - \beta$. You will need to guess values of p_1 and p_2. The sample size for each group is computed as:

$$n=(Z_\alpha + Z_\beta)^2 * (p_1 q_1 + p_2 q_2) / (p_2 - p_1)^2$$

Example.

You wish to test the hypothesis that the proportion of mothers who wash their hands "correctly" differs significantly between households with "low" incidence of diarrhea compared to those with "high" incidence, using a two-tailed test with α=.05 and power, $1-\beta$=.80. Pilot data estimated the overall p=0.3. You guess that p_1=0.4 for mothers in low incidence households and p2=0.2 for mothers in high incidence households. Substituting these values,

$$n=(1.96 + 0.842)^2 * [(0.4)(0.6) + (0.2)(0.8)] / (0.2 - 0.4)^2$$

yielding n=78.51 or n=79 per group.

[NOTE: This is a large sample approximation. Please refer to statistics texts for additional formulae].

APPENDIX III

DIETARY MANAGEMENT OF DIARRHEA ETHNOGRAPHIC GUIDE

Brown KH and Bentley ME. (1988) "Improved Nutritional Therapy of Diarrheal Diseases: A Guide for Program Planners and Decision Makers. PRITECH, 1988.

APPENDIX II ETHNOGRAPHIC FIELD GUIDE COLLECTION OF INFORMATION ON FEEDING PRACTICES DURING DIARRHEA

INFORMATION TO BE COLLECTED

This guide covers the breadth of information that should result from interviews with key informants. Interviews with others, which must necessarily be of shorter duration and of a non-repeated nature, should not go into great depth on general issues. The key issues of diarrhea management should be the emphasis for non-key informant interviews.

The guide should be read through carefully several times before conducting the first interview and may be helpful for the first several interviews. The goal should be to "know" the material well, so that interviews will be complete in content, but spontaneous in form. The sequence that is presented here need not be strictly followed. However, all the key elements should be covered in each topical area.

EXPLANATION OF STUDY

Introduce yourself. Explain your program association and that you are interested in learning about child feeding, health, and illness. Stress that you are not an expert in any of these areas, but that your work is to understand what people do.

GENERAL ILLNESS TAXONOMIES

Begin by asking, "Can you give me the names of all the kinds of illness people have around here?" *First, simply acquire a list of all the names. Try* prompting the respondent by suggesting seasonal illnesses: "What kinds of illness occur here in winter?" After compiling the list of illnesses, go back to the beginning of the list and ask, "What is the symptom of this illness?" Probe for detail here. For example, the respondent may give the name *gripe*, with the following associated symptoms: fever, sore throat, and runny nose. Ask, "If there is only fever and sore throat, but not a runny nose, is it still called *gripe*, or would that illness have another name?" For this primarily taxonomic exercise, there is no need to ask about causes or treatments.

CHILD ILLNESS

Refer to the list of general illnesses. Ask, "Which of these illnesses do children experience?" Make a new list of child illnesses. With this list, repeat the exercise used for general illnesses. Find out what the symptoms for each illness are. Once this has been accomplished, ask whether any of these illnesses are serious, and under what circumstances. For example, if the respondent answers, "Measles is very serious," you should ask, "What can happen to a child with measles? Are there any circumstances when measles is not so serious?"

For some of the childhood illnesses, ask about causes and treatments. Don't do this for *all* illnesses, as it would require too much time. The objective is to be able to make some distinctions between diarrhea and other illnesses. For example, ask, "When the child has *gripe*, what do you do? If he or she doesn't get better, then what do you do?"

CHILD DIARRHEA — GENERAL

If diarrhea has been mentioned as a childhood illness, return to it by saying, "You mentioned that diarrhea is a childhood illness, what is diarrhea? How do you know the child is having diarrhea?" How diarrhea is *perceived* by the respondent is one of the key questions. Ask if there are any special names that are given for diarrhea. Are there different "types?" What are the names and physical characteristics (descriptions) of each type? What is the cause of each type? What is the best treatment for each type?

Ask if the respondent considers diarrhea to be a serious thing? If not, probe for why not. If yes, find out the reason. Ask if any one type of diarrhea is more serious than another. Determine why. Ask what can happen to a child with diarrhea. Here probe for cognitive beliefs about a developmental sequence for a diarrhea episode. Get good descriptions of the symptoms and physical characteristics of each stage. Probe for distinctions between the different types of diarrhea and possible alternative outcomes.

Refer to the causes that have been given for each diarrhea type. Ask if the respondent thinks diarrhea can be prevented. If yes, how? If no, why not? Refer to the treatments that have been mentioned for each type. Probe for more detail on treatment. Find out when during an episode a particular treatment should be done, and why. Do people use both home and health/medical treatments at the same time? What kind of advice do their doctors and health practitioners give them about what to do during diarrhea? (Ascertain what kinds of health care is available, what people prefer, etc.)

CHILD FEEDING — NORMAL, DURING DIARRHEA, AND AFTER DIARRHEA

Determine what people normally feed their infants and children. Ask about breastfeeding and weaning practices: When is the best time to wean, and why? What are the best foods to offer, and why? When should other foods (list them) be offered, and why? When should breastfeeding be stopped completely, and why?

Ask about how food availability and food production influences what a child eats, and determine the amounts available. Find out if there are any times of the year when there is not enough food. If so, do people have enough money to buy from the market? Do they ever experience hunger? Are there times of the year when mothers think their children don't get enough to eat?

For some interviews with mothers who have small children, get a 24-hour recall of all foods that were consumed. Begin by asking, "Yesterday when the baby or child woke up, what was the first thing that he or she ate?" Find out what the mother gave the child throughout the day until the child went to bed. Don't worry about recording *amounts* of foods, just find out which types of foods the child consumed. Be sure to get these lists for a wide range of ages of children.

Ask about feeding during diarrhea. Start with a general question: "When a child has diarrhea, should he or she be fed differently?" Let the respondent answer spontaneously, then probe for details about changes in the amounts and kinds of foods that should be given during diarrhea. Find out the reasons for these changes. If the respondent believes that some foods are useful or harmful during diarrhea, get a list of these along with the reasons.

After discussing feeding during diarrhea, go back to the matrix of the different "types" of diarrhea. For each type, ask whether there should be changes in the amounts or kinds of foods that should be given. Find out when changes in feeding during a diarrhea episode should occur: At the onset? After a few days? Probe, but don't lead. For those foods that have been listed as useful, get precise recipes (if such preparation is required). Ask about child feeding right after the diarrhea has stopped. Are there any changes in the amounts or kinds of foods given? Does a child appear more or less hungry?

THE LAST DIARRHEA EPISODE

Ask about the last diarrhea episode that occurred in the household. Find out the age and sex of the child. Ask the respondent to describe everything that happened, beginning with the physical description and symptoms (e.g. number, color of stools, vomiting, etc.). What do they think caused the diarrhea? Ask about treatments that were given. When during the episode, and why, was each treatment done? Who and how was the decision made for each treatment? Were they happy with the outcome of the treatment/s?

Ask about feeding during diarrhea. Find out amounts and kinds of foods given. Was more or less food given? Did the child seem more or less hungry? When during the episode, and why, did changes in feeding occur? What were the reasons for these changes? Were the changes perceived as beneficial?

WOMEN'S WORK/CHILD CARE

Find out what kind of work women do during the day, and how this changes throughout the year. Begin by talking about the current season. Ask what women do when they first wake up: Then what? Then what? Next, ask how this changes during the next season. Probe to see if women perceive conflicts between their non-domestic work and their domestic work— including child care. If women do work substantially outside of the home, who takes care of their children?

FOOD PRODUCTION/AVAILABILITY/ CONSUMPTION

Ask what foods are currently being grown. Find out who does the agricultural work. What happens to the food when it is harvested? Is the food processed in the home? By whom? Is the food grown for local/household consumption, or for the market? What foods are grown during different seasons? Within a household, do all family members eat these foods — how is the food that is grown distributed? Which foods that the family consumes must be bought in the market place? Which foods are obtained not from food production or the market, but from alternative sources (e.g. food aid, exchange/barter). Make lists. In the homes of your key informants, do a food inventory (list all the foods, and amounts, that are currently in stock).

COMMENTARY ON INTERVIEW GUIDE

In the everyday activities and experiences of people, childhood diarrhea — its perceived causes and consequences, and what to do about it — is but one small footnote. Understanding how people "get their groceries," or provide resources for their families, is of key importance. For women who have multiple work roles but finite time, a description of their usual activities (and how these vary seasonally) is a necessary prerequisite to understanding how they manage an illness like childhood diarrhea. Therefore, one important goal of the preliminary ethnography is to "paint a picture" of women's work and how this changes throughout the day and year.

Although the focus of a DMD project is on feeding during and after diarrhea, these behaviors cannot be separated from the larger cultural context of childhood diarrhea. Similarly, it is necessary to understand how diarrhea as a child illness is perceived in relationship to other child illnesses, such as measles, chickenpox, upper respiratory infections, etc. It is possible that some respondents will perceive that diarrhea episodes follow a somewhat predictable sequence of developmental stages, and that specific symptoms will trigger a change in behavior — an action. For example, an increase in the number of stools may result in the mother (or other caretaker) *making a change in feeding patterns, or the child* may be taken to the doctor. Similarly, the combination of fever and vomiting may precipitate specific actions. In the interviews, first probe to see if diarrhea can be described developmentally, and if so, link these descriptions or "stages" to beliefs about what the appropriate response (or action) should be.

Probing for cultural definitions and subcategorizations of diarrhea, beliefs about cause and treatment for each "type" of diarrhea, and differences in feeding for each diarrhea "type" is of key importance. The matrix shown below structures the key information. In writing the final report, it should be possible to fill in this matrix from information gathered during the ethnography. It is assumed that the ethnography will be done in multiple sites, where ecological and/or cultural differences exist. The matrix, therefore, will no doubt be different for each site.

The names for the different "types" of diarrhea may or may not reflect its characteristics. Probe in detail for the descriptions. There may be more than one characteristic for each diarrhea type. The perceived cause for each diarrhea type should be described in detail. Often, the perceived cause may lead to clues about intervention barriers. For example,

MATRIX OF SUBCATEGORIES OF DIARRHEA

	Name of Diarrhea	Characteristics (Description)	Cause	Preferred Treatment	Feeding Practices
1					
2					
3					
4					
5					

In some cultures, teething diarrhea (which is perceived to be caused by teething and related to a developmental stage in a child's life) is considered a "rite of passage," and to some mothers may not be seen as requiring or responding to an intervention. It may not be seen as serious, but rather a positive benchmark of growth. This is the kind of description that should be brought out in the interviews. It is not enough to know that one type of diarrhea is called "teething diarrhea" and that it is caused by teething. The context around each finding must be determined.

Beliefs about what are thought to be the best or most appropriate treatments should be described for each type of diarrhea. Often what people think are the best things to do are not what they actually do. One way of finding out what people do is to ask about the last diarrhea episode. Find out about actions that were taken then. Treatments should include both home treatments (e.g. herbal) or treatments outside of the home (e.g. medical doctor, village health guide, chemist, exorcist, etc.). It is expected that for different "types" of diarrhea, there will be different treatments.

The information about feeding during diarrhea is key to the project. A significant portion of the interview should address the feeding issue. In order to contrast differences in feeding patterns during normal or healthy times compared to when the child is having diarrhea, normal infant and child feeding patterns must be understood: When should a child be weaned? What kinds of solid foods should be given first? etc. In reference to feeding during diarrhea, questions should be asked about amounts and kinds of foods given. Foods that are considered useful and harmful during diarrhea should be listed, and the reasons for these beliefs. It may be that the different types of diarrhea have different lists of foods, or that the amount of food that should be given will vary by type. Determining this will require extensive questioning. For those foods that are considered "useful" or that are often given during diarrhea, *obtain a detailed recipe for its preparation.*

Feeding after diarrhea, during what is called the convalescent stage, is important. Find out if mothers feed differently during this time, compared to during the episode or when the child is healthy. Amounts and kinds of foods should be listed, along with reasons for these beliefs and practices.

LOGISTICAL ISSUES

It is assumed that the selection of sites for the preliminary ethnography will be made in the field by the project team. The sites will be chosen based upon ecological, agricultural, cultural/linguistic, population demography, etc. criteria. Given time and logistical constraints, a small number of sites will be chosen for the preliminary ethnography. Fieldwork should be carried out for about two weeks in each site.

The entry process for each site must be done carefully and sensitively. The first people to approach are the village leaders. A letter of authorization from the Ministry of Health should be shown, and the project explained. It is essential to gain the goodwill of the important influencers. Without it, there is little chance of conducting free-flowing interviews over an extended period.

If it is possible to live directly in the village or site where the preliminary ethnography is to be done, this is optimal. However, this may not be practical or acceptable to the villagers, and discretion is required in making a decision about where to base the ethnography.

Within each site, some fairly systematic criteria for choosing respondents should be established. Decisions about who to interview should be based upon one important principle: *capture the variation.* For example, suppose the ethnography will take place in a village of one thousand population, with approximately 200 households. On the first day of ethnography in this village, some quick demographic surveillance should be done to "map" or stratify the village into important divisions. For example, is there a clear socioeconomic stratification? One way of measuring this is to map the physical "types" of houses: Which are made with more expensive materials? Which have tile floors as opposed to mud floors? Are the different types clustered in one area of the village? Rapid surveillance and quick sketches will provide a structure for selection of respondents. For the example provided above, it is important to choose respondents from all the levels of strata.

Within each strata (and socioeconomic status is often the most important as it covaries with a number of other important variables) identify key informants. Key informants are respondents who may be particularly knowledgeable about the issues. Lengthy, repeated interviews should be done with key informants. In a DMD project, key informants could include a midwife or village health guide or mothers who have several young children. Again, choose enough key informants to capture the variation. Although knowledgeable, a village health guide may know too much about the "scientific" way to manage diarrhea, and may be out of touch or judgmental about what other people do. The only way to find key informants is to talk to many different kinds of people, and return for second interviews with key informant candidates. For each of the two sites, there should be about five key informants.

Along with key informant interviews, which are characterized by their more intensive, longitudinal nature, one-time only interviews should be done with a wide variety of people. Don't limit the choice to mothers of young children only, although this is certainly the target group (the focus should be on mothers with children three years and under). Don't forget grandmothers and mothers-in-law, who are important influencers, especially if they live in an extended family where young children are being raised. If it is possible to talk to fathers, carry out some interviews with them. It is not unlikely that they are important decision-makers in the household, and it is important to know what they believe. In many settings, fathers are the household members most likely to make a decision to take a child to the doctor, and often they, not the mother, take the child to the practitioner, clinic, or hospital.

In many cultural settings, group discussions (focus group interviews) provide valuable information. Often, focus group interviews may allow a more free-flowing and open discussion, and information may come out which would not in a person-to-person depth interview. However, it is possible that the composition of the group will actually inhibit spontaneity. An example of this could be when a mother-in-law and daughter-in-law are both in the same focus group, and the daughter-in-law is overshadowed or inhibited by her mother-in-law. Make your own evaluations on whether group interviews are useful. In many settings, it may be impossible to conduct interviews that are not in some sense focus group, as people will wander in and sit down to talk. When this happens, find out who is there: Mothers of young children? Do they work outside the home? Write this information in your notes.

When conducting interviews, it is important to cover the same topics for each interview. Notes should be taken during each interview. In the late afternoon and evening, read through the notes and add details. Remember, this is the only data and it must be result in *usable* and *reliable* information. It is very likely that project investigators will want copies of the notes. They should be legible and understandable. Make sure to allot enough time to transform field notes for this purpose.

ANALYSIS OF DATA AND REPORT WRITING

Use the format of the guide as an outline for your data analysis and write-up. Given time constraints, don't worry about elegant style, but focus on pulling out the key pieces of information.

Where there are differences of opinion between respondents, point these out. And list all of the different responses, giving weight to those that are more commonly mentioned. Beware of giving "normative" descriptions and don't make statements like, "mothers believe this or that...." There are many different kinds of mothers.

One of the most important tasks is to look for the variation. Be careful not to overemphasize interesting or "exotic" results at the expense of less interesting but more relevant data. For example, do not give undue attention to mothers who take their children to exorcists. In fact, the percent of mothers who do this may be very low, but because it makes for interesting discussion and reading, there may be a temptation to dwell upon such a finding in the report.

Summarize the data in tables, and provide frequency distributions of pertinent information. Disaggregate the tables by site or urban/rural categories. Tables help in the organization of notes and are useful to the interdisciplinary team members.

APPENDIX IV: A SAMPLE "SCRIPT"

A Kaami (Low-Caste) Meal in Rural Nepal

On July 8, 1987 an evening meal (**ratiko kharcha**) was observed in the household of Nokhi Ram Kaami. The household consists of nine individuals: Nokhi Ram, 43, the male head of household; Dhami, 41, his wife; Chandra Ram, 23, his elder son; Kul Bahadur, 22, his second son; Bodhi Kumari, 13, his elder daughter; Indra Kumari, 8, another daughter; Rurn Kali, 22, his daughter-in-law (Chandra Ram's wife); Khuma Kumari, 4, his granddaughter (Chandra Ram's daughter); and Suk Lal, 2, his grandson (Chandra Ram's son). All household members, except Chandra Ram, were present for the observed meal. The meal was observed by Joel Gittelsohn.

I arrive at 6:35 p.m.. Nokhi Ram is out at a construction project, doing wage labor to build a new primary school. His wife, Dhami, is away herding cows. Chandra Ram is almost a day's walk away in Tharmare, doing some plastering for wages. Kul Bahadur is resting inside, having recently returned from his studies in Khalanga. Bodhi K. is chopping wood for cooking the evening meal. Indra K. is outside playing. Rurn Kali, the daughter-in-law, is winnowing rice, while Suk Lal lies on some old clothes sleeping next to her. Rurn Kali's daughter, Khuma K., is sitting next to her mother. It is raining lightly outside.

The house is relatively poor, with only one story, a few small windows, and a thatched roof. It is located in Gairagaun, near the commercial center of the **panchayat**. There is only one cooking area, with a slightly raised platform off to the side. All household members eat on this platform, except Rurn Kali, her son Suk Lal, and Bodhi Kumari, who eats on the floor near the **chulo** (a low stove made of mud and stones).

At 6:41, Bodhi K. returns with kindling and began preparing potato **taarkhari** (a vegetable stew). Dhami returns at 6:48 with some greens and some bananas she had purchased from a nearby house. Bodhi K. says, "give the greens to me Mom, I will cook them."

At 6:52, Dhami serves herself a wheat **roti** (a kind of unleavened flat bread) (medium-sized) and a half-teaspoon of salt-chili mixture (a common kind of **achar** (sauce condiment) used by poor households for seasoning). Bodhi K. serves herself a banana. Dhami gives one sixth of her **roti** to her granddaughter Khuma K.. She then serves Suk Lal a small **roti**. Bodhi K. begins chopping the spinach greens (7:00). At 7:09, Dhami serves Khuma K. a banana. Kul B. serves himself a banana, as does Bodhi K.. Dhami then begins preparing the rice. Her daughter-in-law brings in some firewood. Bodhi K. goes out to herd the goats (7:16). Dhami serves Indra K. a banana, then splits a second banana between herself and Rurn Kali.

At 7:22, Rurn Kali begins washing dishes. Khuma K., still hungry for bananas, begins hunting for more around the kitchen. Dhami tells her, "there's none left," and she begins to cry. Dhami speaks with me about some of the difficulties of living in Nepal, and asks if I would give her grandson an injection. Very upset, Khuma K. hits her grandmother for not giving her a banana. She begs for a banana. Dhami hits Khuma on the head, causing her to cry loudly.

At 7:35, Rurn Kali returns and begins to stir the rice. Indra K. asks for, and is given water. Rurn begins breastfeeding Suk Lal. Rurn Kali serves 1/6th cup of rice water (**maardh**) to Khuma, and the same amount to Indra. Khuma refuses to eat for a bit; she is still holding out for a banana, but finally drinks the rice water. Dhami cuddles her granddaughter. Dhami hits Khuma lightly on the head as punishment for playing with the dishes with her feet (7:47). Kul pours himself some water and goes outside to bathe his hands and feet before the evening meal.

Rurn serves Suk Lal rice (1 1/2 cups) and potato **taarkhari** (1 tbl.) in a bowl, then begins to feed the child by hand. Dhami sweeps the eating area (8:01). The family had been waiting for the male head of house to return from his work, but it is late and he has not returned, so the younger members of the family begin the evening meal. Rurn serves rice and **taarkhari** onto several plates at once. Rurn serves Khuma rice (2 1/2 c.) and **taarkhari** (2/3 c.) (8:04). She serves Kul Bahadur rice (4 1/2 c.) and **taarkhari** (1 1/2 c.) (8:05). Rurn Kali serves Indra rice (2 c.) and **taarkhari** (2/3 c.). She serves Suk Lal more rice (1/2 c.). At 8:06, she serves Bodhi rice (3 c.), **taarkhari** (1/2 c.), and a chili. At 8:08, she serves Khuma and Kul Bahadur each one chili. Rurn then serves Kul some more **taarkhari** (1/3 c.). Seeing this, Indra asks for more **taarkhari** and is served (1/4 c.).

All served individuals began eating quietly. As in most poorer Dadagaon homes, the meal is eaten with all household members sitting directly on the ground. Foods are eaten with the right hand, as the left is used for toilet functions and is considered polluted. Rurn, as the food server, should wait until all other household members are finished before she serves herself. Dhami is waiting for her husband to return home. At 8:12, Rurn asked Kul Bahadur if he would like more rice, but he refuses. Rurn serves **taarkhari** juice to Suk Lal (2 tbl.) and Bodhi (1/4 c.). At 8:13, Nokhi Ram returns home.

8:15. Kul Bahadur asks Bodhi for some water and is served. Kul has finished his meal, with no leftover food, and goes outside to wash his hands and mouth. Bodhi has also finished, but leaves 2 cups of rice ana a chili remaining on her plate. Nokhi Ram goes outside to wash prior to his meal. Indra K. finishes her meal, leaving 1 cup of rice on her plate. Rurn combines Indra and Bodhi's leftover food onto one plate. The small children eat out of individual plates, while adults had plates for rice and bowls for the **taarkhari**.

At 8:18, Rurn serves Nokhi Ram rice (6 c.), **taarkhari** (2 c.), and two chilies. At 8:20, Khuma finishes her meal, with 1 1/4c rice and 1/3 c. **taarkhari** remaining. Rurn puts all the **jutho** (polluted by hand-mouth contact) food near the fire to keep it warm. At 8:22, Nokhi Ram finishes his meal and goes outside to wash. He has left 1 c. of rice and 2/3 c. of **taarkhari** remaining. Dhami serves herself her husband's **jutho** and that of Indra and Bodhi Kumari. Rurn serves Dhami 1/2 c. of **taarkhari** juice.

At 8:25, Rurn serves herself Khuma's unfinished food, eating off of Khuma's original plate. Additionally she serves herself the burned rice at the bottom of the pot (3 c.) and **taarkhari** (2/3 c.). Nokhi Ram teases Khuma by saying I will cut her ear if she is not good. Suk Lal has finished his meal, with 2/3 c. rice and 1/5 c. **taarkhari** leftover. Rurn serves herself this leftover food. Rurn occasionally feeds little bits of food to Suk Lal. Rurn serves her mother-in-law **taarkhari** juice (1/6 c.) and then herself the same (1/5 c.). Rurn serves herself water (1/2 c.). At 8:33, Rurn has finished, with no leftover food, and washes her hand in her plate. Dhami finishes shortly afterward, leaving 1 cup of rice unfinished, and also washes her hand in her plate. Rurn cleans Nokhi Ram's place up with her hand. Dhami pours out seed millet and soybeans to be sown when the rains improve.

APPENDIX V: CALCULATION OF VARIABILITY AND RELIABILITY OF KEY BEHAVIORS

The behavior in question may be recorded in one of several different formats: dichotomous (e.g. mother assisted child as he ate - yes/no), multichotomous (e.g. extent to which mother assisted child as he ate - none/some of the time/most of the time/all of the time), or continuous (e.g. number of minutes mother spent assisting the child as he ate). The format dictates the procedure.

DICHOTOMOUS

This is the simplest situation. For each unit (e.g. person), count the number of times the behavior was observed (n) and the number of times it was coded "positive" (f). Using a table for the "95% confidence interval for the binomial distribution" (as can be found in a standard statistics text), find the entry corresponding to your values of f and n. The values shown are the lower and upper bounds of the observed fraction, f/n. If your bounds include 0.5 (or 50 %), you can conclude that the fraction of "yes" responses could easily have been observed by chance alone. If you suspect some reactivity and you have a sufficiently large number of observations on the same unit, you can divide the responses by time and see if the estimated fraction changes over time. For example, if you recorded 30 feeding episodes for a given child, divide the episodes into thirds timewise. Calculate f/n for each third and look up their confidence intervals. If they overlap, you can conclude that the estimated fractions do not change significantly over time (i.e. no evidence of reactivity).

MULTICHOTOMOUS

A multichotomous format may be either ordinal or non-ordinal. If the format is ordinal, the easiest solution is to dichotomize the responses and proceed as described above with a dichotomous format. For example, if you observed the extent to which the mother assisted the child while eating, you could combine "none" with "some of the time" and "most of the time" with "all of the time", yielding a dichotomy.

If you are dealing with a non-ordinal format, it is more difficult to evaluate the variability. For example, if the behavior of interest is child's activity, you may have a variety of possible codes to choose from which hold no intrinsic order of importance with regard to your study question (e.g. sleeping, eating, running, crying, playing, etc.). The simplest approach would again be to dichotomize the responses, repeating the procedure for each of several key codes of interest (e.g. "sleeping vs. not sleeping", "crying vs. not crying", etc.). Then proceed as with a dichotomous format.

CONTINUOUS

We will assume that the behavior of interest is derived from an approximately "normal" continuous distribution. (Consult any introductory statistics book if you do not understand this concept).

If you are attempting to evaluate reactivity, one simple approach would be as follows:

a) Divide the observations into equal segments of time (e.g. halves, thirds, or fourths). Calculate the mean value of the behavior (X) for each person (i) for each time segment (t), i.e., X_{it}.

b) Use standard formulae to compute a paired t-test across persons for each successive pair of time segments (i.e. time 1 vs. time 2; time 2 vs. time 3, etc.). If there are no statistically significant differences between paired times and the means are not steadily increasing or decreasing over time, it is unlikely that reactivity is present.

In a situation in which the same unit is observed multiple times, we can follow a random effects model for the analysis of variance:

$$X_{ij}=\mu+A_i+\epsilon_{it} ,$$

where A_i is a random variable assumed to be approximately normally distributed with mean 0 and standard deviation σ_A and the ϵ_{it} are assumed to be approximately normally distributed with mean 0 and standard deviation σ. The A_i and ϵ_{it} are further assumed to be independent.

Thus, the value of the behavior, X, for unit i at time t is equal to a mean value of X (μ) for all of the units plus a deviation from the population mean for each unit (A_i) plus a deviation from that unit value each time the unit is measured (ϵ_{it}), including simple measurement error.

Suppose that each mother was observed the same number of times, n, feeding her child and that the number of minutes spent assisting the child was recorded each time. The mean square between mothers (S_M^2) computed from all observations for all mothers is an unbiased estimate of the two components of variance,

$$S_M^2 \approx \sigma^2 + n\sigma_A^2 ,$$

and the mean square within mothers estimates σ^2. Through algebra, you can compute σ_A^2.

The variance of the sample mean (X..) is

$$V(X..)=\sigma_A^2/a + \sigma^2/an .$$

Thus, it is possible to evaluate the benefit of increasing the number of mothers (a) vs. increasing the number of observations per mother (n). In a field study, it will generally be less expensive and/or more efficient to minimize the number of mothers rather than the number of observations per mother. If this is not the case, then we simply want to find integer values of a and n which satisfy the above equation when we have set the variance of the sample mean (i.e. the standard error) equal to a predetermined value. The right-hand side of the equation should not exceed this value.

However, if we do have an impression of the approximate cost of observing different mothers (c_1) and observing the same mother repeatedly (c_2), we can incorporate this information into our decision process. Since

$$cost=c_1*a + c_2*a*n ,$$

we can combine the two equations, yielding

$$VC=(\sigma_A^2/a + \sigma^2/an) \ (c_1*a + c_2*a*n) .$$

If we minimize this equation, the minima occurs when

$$n = \sqrt{(c_1 * \sigma^2 / c_2 * \sigma_A^2)} ,$$

which represents the best number of observations per mother. The best value for "a", the number of mothers, can be found by solving either the cost equation or the variance equation for "a", making use of this new value of "n".

APPENDIX VI: GENERAL SUGGESTIONS REGARDING THE APPROPRIATE CONSTRUCTION OF DATA COLLECTION FORMS

Data collection forms should be designed to minimize the decision-making of the field worker and to ease the task of data entry.

1) Allowances should be made for additional codes to be added as data collection proceeds. The best way to do this is to include room next to each item for the field worker to record information which he cannot comfortably classify into any of the codes provided. The supervisor can keep tract of these "exceptions" over time and decide if they appear with a frequency worthy of creating a new code, or if they can be combined with other codes, etc..

2) Whenever possible, allow for the recording of simultaneous codes. For example, in spot checks of mother's activity, it is best to design your form such that every activity in which the mother is engaged at the moment of the spot check can be checked off. Otherwise, you force the field worker to make a subjective decision about the relative importance or intensity of an activity. This could result in biased data and will result in a loss of information on activities.

3) Whenever possible, be consistent in designating codes for items. For example, if many items are "yes/no", make all such items codable as "0=no, 1=yes". If there are situations in which a response can be "not applicable" or "refused" or "unknown", establish a code for each of these which will work with most items. By being consistent, you will help field workers make fewer recording errors and also make data cleaning go more smoothly. It is especially important not to leave blanks for items where the answer is "unknown". Data entry and data analysis software vary in how they handle blank fields, and you may end up with a situation in which your blanks have been redefined and possibly combined with another legitimate code.

4) If your data collection form is precoded (as is advisable whenever possible to reduce transcription errors), lay out each page such that the flow of item responses is simple to follow for data entry purposes, e.g., all keystrokes can be made by following the boxes down the lefthand margin, or by reading each row of underscores across the page.

5) If you are recording the same information repeatedly on the same unit, as with repeated spot checks, each "check" should be entered as a separate record in the data entry process. Each record would thus need to repeat identifying information and to contain other variables such as date and time. Note that if the data recorded at each "check" is not substantial, you may be able to collect the information for many "checks" on the same form, but the data entry should still be done treating each "check" as a separate record.

APPENDIX VII: GENERAL SUGGESTIONS REGARDING THE APPROPRIATE ENTRY OF DATA

The optimal way to perform data entry is to have each data form entered twice, using two different people. Some data entry software packages contain built-in "alarms" which beep or freeze the screen when a different value is entered for a record the second time around. Most software is not that sophisticated, however, and the next best thing is to have two people enter the data on different storage devices (e.g. diskettes), and then have a programmer write a program to compare the two data files and to print a list of any records with different keystrokes. Even this may be beyond your project resources.

Another, but time consuming, process would involve having a person review each record, either on the screen or on a printout, against the original form, making note of errors to be corrected. This is usually considered to be an unacceptable use of manpower, so the next best choice is to pick a sample of the records to validate in this manner. The percent chosen depends on your salary budget; people often review 5-10 % of the records, unless the study is extremely large. If many errors are found (say, more than 5%), you should seriously consider reviewing all records.

Keep track of the errors that you find because you may detect some pattern. For example, certain items may have a high frequency of entry errors, or records filled out by a particular field worker may be problematic (poor penmanship, etc.), or records entered by a particular data entry person may have many more errors than those entered by other people. If you do your comparisons periodically throughout the data collection period, this information could allow you to make changes to personnel or to your forms so as to reduce the error rate for the remainder of the data entry.

APPENDIX VIII: RELIABILITY ANALYSIS OF SCALES

A scale should be both valid and reliable. Validity means that the scale measures what it is intended to measure. It may be possible to test this assumption after you have created your scale if you have available to you another measure that you believe should correspond strongly to the characteristic that you are trying to approximate. For example, Bentley used calories consumed at the meal to validate her "child's acceptance of food" scale.

We assume that there is an hypothetical universe of items that measure the characteristic of interest. The items we are using in our scale are thus a sample from this universe. A scale is reliable (or stable, repeatable) if there is a strong relationship between the score assigned to an individual based on it and scores that would have been assigned based on scales comprised of different equally-sized samples of items drawn from that same universe.

Cronbach's Alpha (α) is a frequently used reliability coefficient based on the "internal consistency" of the scale (either the average correlation or the average covariance of the scale items). The items are assumed to be positively correlated, thus yielding an α with range 0 to 1 (a negative α would thus be a warning that this assumption has not been met). It is important to remember that the α depends both on the strength of the item correlation and on the number of items in the scale. Hence, a large number of items with moderate correlation could yield a sizeable α, or a small number of items with high correlation could yield a moderate α.

The level of measurement normally assumed for the items in the scale is interval, although it is often used with ordinal items. In addition, there is a reliability coefficient based on the Kuder-Richardson 20 formula which is analagous to Cronbach's α. The KR-20 is used when all items are dichotomous. Hence, if your ordinal items have very few possible values or most observations in your data occur at only a few values, you should dichotomize the items and use the KR-20 instead of the Cronbach's α.

Reliability procedures (refer to Nunnally, 1978) basically allow you to determine if certain items weaken your scale and thus should be eliminated and if the resulting scale is a "sufficiently" reliable measurement of the characteristic of interest. [Note that these procedures are available in some software packages, including the SPSS PC + Advanced Statistics Module].

APPENDIX IX: DATA ANALYSIS

This section reviews some of the basic analytic methods that can be utilized to look at structured observation data. The choice of statistical method depends on the study design and the distributional characteristics of the outcome variable and the explanatory/predictor variable. (Note that structured observations can be used for either role). Following are brief descriptions of some of the available methods.

In general, the correct way in which to approach the analysis of any particular structured observation item depends on: the presence of repeated observations of this item on the same observed unit (e.g. person or household); whether or not this item is an outcome or an explanatory/predictor variable; and its level of measurement[5]. These three things must be considered simultaneously.

Repeated measurements refers to a situation in which the investigator wishes to make comparisons of the same individual over time or in different circumstances. Thus, a household could be observed before and after a handwashing education intervention to see if the intervention was effective. Non-repeated measurements involve the use of a single measurement, although that variable may be a summary of many observations, as described in Part III., Step 13c.. Thus, the investigator may wish to relate a daily count of handwashings or a daily average "handwashing effectiveness score" to the incidence of diarrhea over a 3-month period.

---

[5] In this document, level of measurement refers to two broad categories: continuous and discrete distributions.

Within the continuous category are interval and ratio variables. Interval implies that distances between classes are defined by fixed and equal units, but with no inherently determined zero point. Therefore, one can examine differences but not proportionate magnitudes. Ratio variables, however, do have inherent zero points.

Within the discrete category are nominal and ordinal variables. Nominal implies that the value of a class is merely a label, and that no assumptions are made regarding order or distances between classes. Ordinal implies that all classes can be ranked, but no assumptions can be made regarding distances between classes.

Dichotomous variables (those with only two classes) may be either nominal (e.g. male or female) or ordinal (e.g. dirty or clean; absent or present).

In practice, many researchers treat many ordinal variables as interval variables in their analyses. This is usually done only when the ordinal variable in question contains more than a few classes (say, > 5) and when the analysis is exploratory, rather than confirmatory, in nature.

In the following pages, procedures described for continuous variables may therefore, at times, be cautiously applied to ordinal variables. Ordinal variables can always be correctly treated with the discrete variable procedures, especially those which specifically incorporate properties of order (as will be noted).

Non-Repeated Measures

1. Continuous Outcome vs. Continuous Predictor

    If both the outcome and the predictor/explanatory variable are continuous (ratio, interval, and possible ordinal), a Pearson's or Spearman's correlation coefficient can be used to estimate their strength of association. Pearson's coefficient assumes that the two variables are approximately "normally" distributed[6], whereas the Spearman's coefficient allows for more skewness in the distributions.

2. Continuous Outcome vs. Discrete Predictor

    If the outcome variable is continuous, ratio or interval level and the predictor/explanatory variable is discrete and multichotomous (>2 classes), one-way analysis of variance can be used to test the null hypothesis that the mean values of the outcome for each level are approximately equal. In addition, the investigator can apply a multiple comparison procedure to further examine the interrelationships of the various class means (i.e., which class means differ significantly from one another). There is a wide variety of such procedures, some common ones being Bonferroni's t-test, Duncan's multiple range test, Scheffe's multiple-comparison test, and Fisher's least-significant-difference test.

    When the outcome variable is continuous, ratio or interval level and the predictor/explanatory variable is discrete and dichotomous (only 2 classes), the one-way ANOVA reduces to the familiar Student's t-test.

    In the case of an appropriate ordinal level outcome, it may be preferable to avoid the parametric assumptions underlying the use of ANOVA by using a non-parametric method. The Wilcoxon rank sum test (or the Mann-Whitney U) involves ranking observations from 2 classes and comparing the sums of the ranks from each class (adjusted for class size). An extension to >2 classes is provided by the Kruskal-Wallis test. In both cases, the tests are used to detect shifts in location. (For more information on non-parametric procedures, one good reference is Hollander and Wolfe, 1973).

3. Discrete Outcome vs. Continuous Predictor

    If the outcome variable is discrete (either dichotomous or multichotomous), it is necessary to force the continuous predictor/explanatory variable into discrete categories as well, then proceed with discrete vs. discrete analyses (see below). The decision for dividing the continuous variable into classes may be based on several approaches:

    a) Precedent - In the literature, there may be commonly used cutpoints for a particular variable. For example, birthweight (grams) is often divided into groups such as <=1500, 1501-2500, 2501-5000, >=5001.

---

[6] It may be necessary to "transform" the distributions first to make them more approximately normal (please refer to Velleman and Hoaglin (1981) for details on transforming data). A distribution that is highly irregular, such as one in which there is obvious digit preference or more than one modal value (see Figure 8.) cannot be transformed. Instead, the distribution should be partitioned into reasonable segments and treated as an ordered categorical variable.

response variables across time". (Their methods can be implemented indirectly through the Statistical Analysis System (SAS) software).

If the measurement is an ordered categorical response (e.g. mild, moderate, severe), Agresti (1989) reviews three models which "describe simultaneously the dependence of marginal response distributions on values of explanatory variables and on the occasion of response." (Again, SAS can be used to conduct these analyses).

Repeated Measures; Two or More Groups

The following comments apply to a situation in which a measurement is repeated and the investigator wishes to compare groups of individuals across repetitions.

1. Continuous

   Again, for a ratio or interval level measurement with an approximately normal joint distribution, MANOVA is the most correct approach.

   If the repetitions are based on time and are not comparable across individuals (including situations in which the number of repetitions per individual varies), recent stand-alone Fortran software has been developed at Harvard in conjunction with work done by Laird and Ware (1982) on the use of restricted maximum likelihood procedures which aide in handling these issues.

2. Discrete

   In the same reference for Landis et al. noted above, a second methodology involving the "fitting of variational models to summary functions of the correlated marginal distributions across time using a weighted least squares algorithm" is presented for comparison of groups.

Refer also to the Agresti paper if there is an ordered categorical response.

Multiple Variable Regression Models

When analyses of bivariate relationships are complete, it is often desirable to combine a set of seemingly important predictor or explanatory variables into a single model. The model is chosen based on the form of the outcome variable. Some suggestions follow (there are many others):

Linear - for an approximately normal ratio or interval (and possibly ordinal) level outcome. The Y variable may first need to be "transformed", as mentioned previously.

Logistic - for a discrete dichotomous outcome.

[Linear and logistic regression procedures are available in a large number of statistical software packages].

Multinomial logit - for a discrete non-ordered outcome (>2 classes). [Available in mainframe SAS. Other loglinear procedures are available in various microcomputer packages].

Cumulative logit - for a discrete ordered outcome (>2 classes).

[Available in mainframe SAS. The procedure can also be done using a standard logistic regression program; see Agresti].

The procedures mentioned previously for "Repeated Measures; Two or More Groups" also can incorporate multiple variables. These procedures are, in general, designed for a small number of repetitions.

In the case of a linear or binary outcome repeated numerous times, there is another technique which can be used. It is necessary to treat the repetitions as separate (i.e. independent) observations and then modify the parameter estimates to reflect the actual lack of independence among observations from the same individual (i.e. the repetitions).

For example, suppose the investigator observed mothers washing their hands on 4 occasions before, during, and after an educational intervention. Each time, the mother was rated on handwashing technique as unacceptable or acceptable. A logistic regression model could be employed in which all handwashing scores were regressed on a set of 3 dummy indicator variables defining the occasions (4 minus one occasion used as the reference category). The investigator would obtain estimates of each dummy indicator plus its standard error, ignoring the fact that each mother appears in 4 observations.

Recent statistical techniques have been developed to make corrections to these estimates by imposing one of several potential "correlation structures" on the data set, and using this theoretical structure to "improve" the estimates. Zeger and Liang (1986) have worked out procedures based on generalized estimating equations for a variety of mean-variance relationships (Gaussian, Poisson, binary, and Gamma). [Macros compatible with either the mainframe or microcomputer version of the SAS statistical package and a function called through the "S" microcomputer statistical package which implement their technique are available through the Johns Hopkins Dept. of Biostatistics].

b) Irregularities in the frequency distribution obtained from the sample - The investigator may note "blips" or "peaks" in the distribution which may be explained (e.g. digit preference) or unexplained (see Figure 8). The investigator may choose to center her groupings at these points.

c) Percentiles - If there is no precedent and there are no sizeable irregularities, it is always acceptable to divide the distribution into percentiles. It is common to use halves (50%) or quarters (25%), or to combine the middle quarters (top 25%, middle 50%, bottom 25%) to yield 3 classes.

4. Discrete Outcome vs Discrete Predictor

If the outcome is discrete, the investigator can examine these relationships using simple crosstabulations and appropriate tests of association. There are a wide variety of tests available, depending primarily on the number of rows and columns and whether or not the row or column variable is ordered. Some suggestions follow (there are many others):

a) Any crosstabulation - Pearson's chi-square statistic (with a continuity correction if the investigator has a 2 x 2 table)

b) 2 rows x 2 columns - Fisher's exact test; odds ratio (with case-control design); relative risk (with cohort design)

c) >2 ordered rows (or columns) x 2 columns (or rows) - In addition to Pearson's chi-square, the investigator may wish to try a test for a linear trend in proportions (see Snedecor and Cochran, 1967).

d) >2 ordered rows x >2 ordered columns - Gamma; Kendall's tau-b; Stuart's tau-c

Two Repeated Measures; Single Group

The following applies to a situation in which the investigator wishes to compare measurements of an individual at two timepoints (e.g. before and after an intervention) or under two conditions (e.g. during illness and health). Thus, both measurements will have the same scale. In addition, for these examples the investigator is not interested in examining differences over time between individuals (i.e. there is a single group).

1. Continuous

When the measurements are continuous, ratio or interval level, the paired t-test can be used. Refer to the caution stated above in "continuous outcome vs continuous predictor".

If the measurements are repetitions of an appropriate ordinal level variable, non-parametric tests for paired data can be used. One such test is the Wilcoxon signed rank test, wherein the absolute values of the differences between 2 repeated observations are ranked, their signs (positive or negative) are restored, and the counts of signs are compared for their deviation from equality.

2. Discrete (dichotomous)

Simple crosstabulations with McNemar's chi-square for paired tests of proportions can be used for a 2 x 2 situation.

3. Discrete (multichotomous)

Dixon and Massey (1967) describe a test for symmetry in a paired crosstabulation of order k. Fleiss presents a kappa statistic for quantifying concordance which is corrected for chance agreement (Fleiss 1981).

Three or More Repeated Measures; Single Group

1. Continuous

If the measurements are continuous, ratio or interval level, their joint distribution is approximately normal, and the repetitions took place under similar conditions for each individual, the data meet the conditions for multivariate analysis of variance (MANOVA). "Similar conditions" implies that, for instance, child activity was measured at approximately ages 3, 6, 9, and 12 months for each child or that child feeding behavior was measured during illness, convalescence, and health using rigid criteria to define these 3 states. In MANOVA, contrasts between different repetitions or combinations of repetitions are statistically evaluated. For example, does child activity from 3 to 6 months of age differ from that at 9 to 12 months? Does feeding behavior differ between diarrhea and convalescence or between diarrhea and health?

Although MANOVA is a relatively complex statistical method, software is available for microcomputers (including SPSS PC+ and BMDP). The reader is strongly advised to refer to a textbook which addresses this topic in detail. (Suggested references include: Hand and Taylor, 1987; Chapman and Hall, 1975; and Winer, 1971).

Another approach is to use a random effects, 2-way ANOVA model, including terms for individual, repetition, and individual by repetition. Both the individual and individual by repetition terms would be considered "random" effects, while repetition would be considered "fixed". Although 2-way ANOVA is commonly found in microcomputer statistical software, the researcher will usually need to hand calculate the F statistic from the error estimates provided by the software because the standard F is based on a conventional "fixed" model (see Snedecor and Cochran, 1967, or another standard statistics text).

For repetitions of an appropriate ordinal level measurement, the non-parametric Wilcoxon signed rank test has been adapted to data with >2 repetitions (see Hollander & Wolfe 1973). Multiple comparisons can be made among the repetitions.

2. Discrete (dichotomous)

Cochran's Q test may be used for >2 dichotomous repeated measurements. [Note: McNemar's chi-square for a pair of repeated measurements is, in fact, a special case of the Q statistic].

3. Discrete (multichotomous)

Unfortunately, there are no straightforward procedures for dealing with >2 repetitions of a categorical measurement.

Landis et al. (1988) present a "generalized randomization model approach for tests of interchangeability of the distributions of the

References Cited

Agresti, A. 1989
A survey of models for repeated ordered categorical response data. Statistics in Medicine 8:1209-1224.

Altmann, J. 1973
Observational study of behavior: Sampling methods. Behavior 44:227-265.

Bentley, M., et al. 1988
Rapid ethnographic assessment: Applications in a diarrhea management program. Social Science in Medicine 107-116.

Bentley, M., et al. 1989
Maternal feeding behavior and child acceptance of food during diarrhea, convalescence, and health in the central Sierra of Peru. Manuscript submitted for publication.

Bernard, H.R. 1988
Research Methods in Cultural Anthropology. Sage Publications.

Bock, D. 1975
Multivariate Statistical Methods in Behavioral Research. McGraw-Hill Book Company.

Black, R.E., et al. 1981
Handwashing to prevent diarrhea in day-care centers. American Journal of Epidemiology 113:445-451.

Briscoe, J. 1984
Intervention studies and the definition of dominant transmission routes. American Journal of Epidemiology 120:449-55.

Brown, K. and Bentley, M. 1988
Improved Nutritional Therapy of Diarrhea: A Guide for Program Planners and Decision Makers Involved in CDD Programs. Technologies for Primary Health Care (PRITECH) Project, Management Sciences for Health, Arlington, Virginia, USA.

Chisholm, J.S. 1985
Commentary in Mulder and Caro (listed).

Dehavenon, A. 1978
Superordinate Behavior in Urban Homes: A Video Analysis of Request-Compliance and Food Control Behavior in Two Black and Two White Families Living in New York City. PhD dissertation, Columbia University.

Dixon, W. and Massey, F. 1969
Introduction to Statistical Analysis (3rd edition). McGraw-Hill Book Company.

Dumont, J-P. 1985
Commentary in Mulder and Caro (listed).

Esrey, S., et al. 1985
Interventions for the control of diarrhoeal diseases among young children: Improving water supplies and excreta disposal facilities. Bulletin of WHO 63(4):757-772.

Esrey, S. 1987
The Effects of Improved Water Supplies and Sanitation on Child Growth and Diarrheal Rates in Lesotho. PhD dissertation, Cornell University.

Esrey, S. and Feachem, R. 1989
Interventions for the Control of Diarrhoeal Diseases Among Young Children: Promotion of Food Hygiene. WHO/CDD/889.30.

Feachem, R. 1984
Interventions for the control of diarrhoeal diseases among young children: promotion of personal and domestic hygiene. Bulletin of WHO 62(3):467-476.

Fisher, A., et al. 1983
Handbook for Family Planning Operations Research Design. The Population Council, New York.

Fleiss, J. 1981
Statistical Methods for Rates and Proportions (2nd edition). John Wiley & Sons.

Gittelsohn, J. 1989
Intrahousehold Food Distribution in Rural Nepal. PhD Dissertation, University of Connecticut (Storrs).

Griffiths, M., et al. 1988
Improving Young Child Feeding During Diarrhea. A Guide for Investigators and Program Managers. The Weaning Project, Manoff International, Inc., Washington, D.C..

Gross, D. 1984
Time allocation: A tool for the study of cultural behavior. Annual Review of Anthropology, 13:519-558.

Hand, D. and Taylor, C. 1987
Multivariate Analysis of Variance and Repeated Measures. Chapman & Hall.

Hollander, M. and Wolfe, D. 1973
Nonparametric Statistical Methods. John Wiley & Sons.

Huffman, S., et al. 1988
Applied research for a dietary management of diarrhea program in Peru: comparison of anthropological and survey methods. Manuscript submitted for publication.

Johnson, A. 1975
Time allocation in a Machiguenga village. Ethnology 14:301-310.

Khan, M.U. 1982
Interruption of shigellosis by hand washing. Transactions of the Royal Society of Tropical Medicine and Hygiene, 76:164-168.

Kirk, J. and Miller, M. 1986
Reliability and Validity in Qualitative Research. Qualitative Research Methods, series 1, Sage Publications.

Laird, N. and Ware, J. 1982
Random effects models for longitudinal data. Biometrics 38:963-974.

Landis, J.R. et al. 1988
Some general methods for the analysis of categorical data in longitudinal studies. Statistics in Medicine 7:109-137.

Lee, R.B. 1969
Kung Bushman subsistence: An input-output analysis. In Vayda, A., editor, Environment and Cultural Behavior. Natural History Press.

Manoff, R. 1985
Social Marketing: New Imperative for Public Health. Praeger Publishers.

Martin, P. and Bateson, P. 1986
Measuring Behavior: An Introductory Guide. Cambridge University Press.

McIver, J. and Carmines, E. 1981
Unidimensional Scaling. Quantitative Applications in the Social Sciences, series 07-024. Sage Publications.

Messer, E.
The relevance of time allocation analysis for nutritional anthropology. Unpublished manuscript.

Morgan, D. 1988
Focus Groups as Qualitative Research. Qualitative Research Methods, series 16, Sage Publications.

Mulder, M.B. and Caro, T.M. 1985
The use of quantitative observational techniques in anthropology. Current Anthropology 26(3):323-335.

Nunnally, J. 1978.
Psychometric Theory (2nd edition). McGraw-Hill Book Company.

Pelto, P. and Pelto, G. 1979
Anthropological Research: The Structure of Inquiry. Cambridge University Press.

Pfaffenberger, B. 1988
Microcomputer Applications in Qualitative Research. Qualitative Research Methods, series 14, Sage Publications.

Radin, P.
In Bernard (listed), p. 23.

Rappoport, R. 1967
Pigs for the Ancestors. Yale University Press.

Reissland, N. and Burghart, R. 1987
The role of massage in South Asia: Child health and development. Social Science and Medicine 25(3):231-239.

Rogoff, B. 1978
Spot observation: An introduction and examination. Quarterly Newsletter of the Institute for Comparative Human Development, Rockefeller University 2(2C):21-26.

Scrimshaw, S. and Hurtado, E. 1987
Rapid Assessment Procedures for Nutritional and Primary Health Care. Anthropological Approaches to Improving Programme Effectiveness. The United Nations University, Tokyo.

Snedecor, G. and Cochran, W. 1967
Statistical Methods (6th edition). Iowa State University Press.

Stanton, B. and Clemens, J. 1986
Soiled saris: A vector of disease transmission? Transactions of the Royal Society of Tropical Medicine and Hygiene 80:485-488.

Stanton, B. and Clemens, J. 1987a
An educational intervention for altering water-sanitation behaviors to reduce childhood diarrhea in urban Bangladesh I. Application of the case control method for development of an intervention. American Journal of Epidemiology
125(2):284-291.

Stanton, B. and Clemens, J. 1987b
An educational intervention for altering water-sanitation behaviors to reduce childhood diarrhea in urban Bangladesh II. A randomized trial to assess the impact of the intervention on hygienic behaviors and rates of diarrhea. American Journal of Epidemiology 125(2):292-300.

Stanton, B. and Clemens, J. 1987c
Twenty-four-hour recall, knowledge-attitude-practice questionnaires, and direct observations of sanitary practices: a comparative study. Bulletin of WHO 63(2):217-222.

Torun, B. 1983
Environmental and educational interventions against diarrhea in Guatemala. In Chen, L.C. and Scrimshaw, N.S., editors, Diarrhea and Malnutrition: Interactions, Mechanisms, and Interventions. Plenum Press.

Velleman P. and Hoaglin D. 1981
The Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press.

Werner, O. and Schoepfle, G.M. 1987
Systematic Fieldwork. Two volumes. Sage Publications.

Winer, B. 1971
Statistical Principles in Experimental Design. McGraw-Hill Book Company.

Whiting. J.M., *et al*. 1976
Field Guide for a Study of Socialization. John Wiley & Sons.

Zeger, S. and Liang, K. 1986
The analysis of discrete and continuous longitudinal data. *Biometrics* 42:121-130.