A METHODOLOGY TO REVIEW PUBLIC HEALTH INTERVENTIONS: RESULTS FROM NUTRITION SUPPLEMENTATION AND WATER AND SANITATION PROJECTS

by Steven A. Esrey, Jean-Pierre Habicht and William P. Butz

1. Measurement of Health and Nutrition Effects of Large-Scale Nutrition Intervention Projects
   by Jean-Pierre Habicht and William P. Butz

2. The Impact of Improved Water Supplies and Excreta Disposal Facilities on Diarrheal Morbidity, Growth and Mortality Among Children
   by Steven A. Esrey and Jean-Pierre Habicht

Cornell International Nutrition
Monograph Series
Number 15 (1985)

Division of Nutritional Sciences
N.Y.S. College of Human Ecology
N.Y.S. College of Agriculture and Life Sciences
Statutory Colleges of the State University
at Cornell University
Ithaca, NY 14853

Cornell International Nutrition Monograph Series

1 (1974) Priorities in Dealing with Nutrition Problems in Indonesia, by Soekirman

2 (1975) The Promotion of Bottle Feeding by Multinational Corporations: How Advertising and the Health Professions Have Contributed, by Ted Greiner

3 (1975) Using the Method of Paulo Freire in Nutrition Education: An Experimental Plan for Community Action in Northeast Brazil, by Therese Drummond

4 (1977) Regulation and Education: Strategies for Solving the Bottle Feeding Problem, by Ted Greiner

5 (1977) Nutrition Planning and Policy for African Countries, Summary Report of a Seminar Held in Nairobi, Kenya, June 2-19, 1976, Edited by Michael C. Latham and Sidney B. Westley

6 (1979) The Economic Value of Breastfeeding (with results from research conducted in Ghana and the Ivory Coast), by Ted Greiner, Stina Almroth and Michael C. Latham

7 (1980) Nutritional Problems and Policy in Tanzania, by Olyvia Mgaza

8 (1981) Nutrition in Policy Planning for the Rural Sector, by Patricia M. O'Brien Place

9 (1981) The Impact of Socioeconomic Development and Ecological Change on Health and Nutrition in Latin America, by Giorgio R. Solimano and Georganne Chapin

10 (1982) The Decline of the Breast: An Examination of Its Imapct on Fertility and Health, and Its Relation to Socioeconomic Status, Edited by Michael C. Latham (Three Papers: I. The Relationship of Breastfeeding to Human Fertility, by Michael C. Latham; II. Appropriate Strategies to Improve Infant and Young Child Feeding, by a Cornell Working Group, III. Infant Feeding Options for Bangkok Professional Women, by Penny Van Esterik)

11 (1982) The Decline in Breastfeeding: An Analysis of the Role of the Nestle Corporation from Two Perspectives, by Carolyn E. Campbell

12 (1983) A Comparison of Growth Standards: Similarities Between NCHS, Harvard, Denver and Privileged African Children and Differences with Kenyan Rural Children, by Lani S. Stephenson, Michael C. Latham and Ad Jansen

13 (1984) The Energy Trap: Work, Nutrition and Child Malnutrition in Northern Nigeria, by Richard Longhurst

14 (1985) Determinants of Young Child Feeding and Their Implications for Nutritional Surveillance: I. Child Feeding Practices: Knowledge, Research Needs and Policy Implications, by R. Martorell et al.; II. Intra-Family Food Distribution: Its Relevance for Maternal and Child Nutrition, by P. Van Esterik

Copies of the Monographs may be obtained from: Dr. Michael C. Latham, Division of Nutritional Sciences, Savage Hall, Cornell University, Ithaca, New York 14853, U.S.A.

Price: U.S.$3.00 including postage surface mail. Make check payable to Cornell University.

Cornell University
Program in International Nutrition

A METHODOLOGY TO REVIEW PUBLIC HEALTH INTERVENTIONS: RESULTS FROM NUTRITION SUPPLEMENTATION AND WATER AND SANITATION PROJECTS

by Steven A. Esrey, Jean-Pierre Habicht and William P. Butz

1. Measurement of Health and Nutrition Effects of Large-Scale Nutrition Intervention Projects

   by Jean-Pierre Habicht and William P. Butz

2. The Impact of Improved Water Supplies and Excreta Disposal Facilities on Diarrheal Morbidity, Growth and Mortality Among Children

   by Steven A. Esrey and Jean-Pierre Habicht

Cornell International Nutrition
Monograph Series
Number 15 (1985)

Dr. Michael C. Latham, Editor
Division of Nutritional Sciences
N.Y.S. College of Human Ecology
N.Y.S. College of Agriculture and Life Sciences
Statutory Colleges of the State University
at Cornell University
Ithaca, NY 14853

Program in International Nutrition

Copies may be obtained from:

Dr. Michael C. Latham
Division of Nutritional Sciences
Savage Hall, Cornell University
Ithaca, New York 14853, U.S.A.

Price: U.S. $3.00 including postage surface mail

We regret to inform our overseas buyers that only U.S. checks or currency are acceptable. We are not able to accept foreign checks except for amounts over $10.00.

PREFACE AND SUMMARY

Two papers appear together in this monograph because they evaluate benefits to child health following large-scale public health interventions. They provide an algorithm to examine studies beyond the published substantive findings. Both papers develop a similar conceptual model to deal with confounding issues related to experimental design, measurement of variables, and statistical analysis. One paper deals with nutritional factors while the other focuses on issues related to disease. The first paper by Habicht and Butz discusses the literature on nutrition interventions on birthweight, morbidity, growth and mortality among children. The second paper by Esrey and Habicht reviews the literature on water supply and sanitation facilities and their impact on morbidity, growth, and mortality among children.

The paper by Habicht and Butz reviewed the criteria necessary to judge the adequacy of nutrition interventions, in contrast to a later review by Beaton and Ghassemi (Supplementary feeding programs for young children in developing countries, AJCN 1982; 35:864-916). Habicht and Butz concluded, contrary to Beaton and Ghassemi, that properly conducted evaluations can show benefits following nutrition supplementation programs. In addition, the choice of indicators for evaluation needs to be carefully chosen to prevent evaluations from being falsely negative. Of the five health indicators reviewed, only height and weight could show health impacts. Furthermore, control of non-nutritional influences through intervention, design, and statistical procedures is indispensable to avoid false negative evaluations of large-scale nutrition interventions. This review originally appeared in the book Evaluating the Impact of Nutrition and Health Programs, edited by Robert E. Klein, et al., 1979. It is reprinted here with the permission of the publisher, Plenum Press.

The second paper in the monograph addresses the continued controversy over whether health benefits occur following water and sanitation interventions. The review comes at the mid-point of the International Drinking Water Supply and Sanitation Decade (1981-1990). Esrey and Habicht used a similar algorithm as developed in the first paper to judge the impact of water and sanitation interventions. They conclude that the better studies consistently show health benefits, whereas studies reporting negative findings were all flawed in one or more major ways. Further, the health impact due to sanitation was larger than the impact due to water supplies. Water quality was of marginal importance in improving child health. Targeting of areas prior to intervention and adherence to specific epidemiologic principles during evaluation need to be followed to enable researchers to identify health impacts.

Steven A. Esrey
Jean-Pierre Habicht
William P. Butz

TABLE OF CONTENTS

MEASUREMENT OF HEALTH AND NUTRITION EFFECTS OF LARGE-SCALE NUTRITION INTERVENTION PROJECTS

by

Jean-Pierre Habicht and William P. Butz

THE IMPACT OF IMPROVED WATER SUPPLIES AND EXCRETA DISPOSAL FACILITIES ON DIARRHEAL MORBIDITY, GROWTH AND MORTALITY AMONG CHILDREN

by

Steven A. Esrey and Jean-Pierre Habicht

LIST OF TABLES

MEASUREMENT OF HEALTH AND NUTRITION EFFECTS OF LARGE-SCALE NUTRITION INTERVENTION PROJECTS

by

Jean-Pierre Habicht and William P. Butz

THE IMPACT OF IMPROVED WATER SUPPLIES AND EXCRETA DISPOSAL FACILITIES ON DIARRHEAL MORBIDITY, GROWTH AND MORTALITY AMONG CHILDREN

by

Steven A. Esrey and Jean-Pierre Habicht

Page

LIST OF FIGURES

MEASUREMENT OF HEALTH AND NUTRITION EFFECTS OF LARGE-SCALE NUTRITION INTERVENTION PROJECTS

by

Jean-Pierre Habicht and William P. Butz

THE IMPACT OF IMPROVED WATER SUPPLIES AND EXCRETA DISPOSAL FACILITIES ON DIARRHEAL MORBIDITY, GROWTH AND MORTALITY AMONG CHILDREN

by

Steven A. Esrey and Jean-Pierre Habicht

MEASUREMENT OF HEALTH AND NUTRITION EFFECTS OF LARGE-SCALE NUTRITION INTERVENTION PROJECTS[1]

Jean-Pierre Habicht[2] and William P. Butz[3]

---

[1] Reprinted from "Evaluating the Impact of Nutrition and Health Programs", edited by Robert E. Klein, et al., 1979, with the permission of the publishers, Plenum Publishing Corporation. Some minor changes have been incorporated into the present publication. The authors express sincere appreciation to Marilyn Ward for preparation of the manuscript in its present form.

[2] Address: National Center for Health Statistics, Hyattsville, MD; Current address: Division of Nutritional Sciences, Cornell University, Ithaca, NY.

[3] Address: Rand Corporation, Santa Monica, CA; Current address: U.S. Bureau of the Census, Department of Commerce, Washington, D.C.

INTRODUCTION

Increasingly the nutritional impact of complex intervention packages is being sought. Typically these complex interventions include intentional changes in the economic, social and political environment as well as in medical and nutritional factors. We are concerned in this paper with the critical issues of the choice of indicators necessary to evaluate these complex interventions with multiple treatments.

We first will argue that large scale, complex interventions require knowledge about indicator variables that is likely to emerge only from careful evaluations of simple interventions. We then discuss a set of optimal properties for field indicators of nutrition and health. These properties have implications for the design and evaluation of interventions. In light of these implications we will review the few evaluations of nutrition interventions reported in the literature in order to assess the knowledge now available for structuring complex, large scale interventions. Finally, we will explore problems of design and measurement peculiar to interventions with multiple treatment.

OPERATIONAL DEFINITIONS OF HEALTH AND NUTRITION

The objectives of the program to be evaluated must be clearly defined and the evaluation must make the appropriate comparisons given these objectives. In the context of this volume, which focuses on very poor populations, good health may be defined as the absence of symptomatic illness, the absence of life-shortening processes and the absence of pathological constraints on performance. Although this definition is broader than that of the clinician, it is narrower than many definitions of health which cannot be translated operationally. More importantly, this definition concentrates mainly on the soundness of the body, little on the soundness of the mind, and not at all on the soundness of the soul. When conditions which threaten the integrity of the body have been obviated thanks to social and economic development, the fostering of the non-physiological components of health can receive a higher priority.

Also within the context of poor populations, good nutrition may be defined as food intake which is adequate enough so that neither health, performance, nor survival are impaired for lack or excess of food or of its components. Again, this definition is broader than that of classical nutrition, but is narrower than that of many nutrition demagogues. It may be considered a transient definition because other socially desirable components may come to be viewed as necessary as a population's physiological needs are met.

In the context of these operational definitions one should be able to infer from an improvement in indicators of health and nutrition that there is an associated improvement in performance, a decrease in overt illness, and/or longer survival. This requires previous demonstrations that these benefits are directly tied to improvements in the chosen indicators of health and nutrition. For instance, a reduction in infant mortality appears clearly to be a health benefit, while an increase in nutrient ingestion may or may not be. So much has been said about the interrelationships between nutrition and health that the interrelationships are presumed understood, at least at most practical levels. Unfortunately, this is only true in severe malnutrition and very poor health. Among those with moderate degrees of ill-health and malnutrition, characteristics of the vast majority of the poor, too few competent field studies have been done to establish the implications for performance, health and survival of changes in indicators of health and nutrition.

LEVELS OF EVALUATION OF FIELD STUDIES AND PUBLIC HEALTH PROGRAMS

There is a logical sequence of evaluation studies depending upon previous scientific and administrative knowledge (1). In general, these include field intervention studies, public health pilot projects, evaluations in conjunction with implementations of an intervention, and ultimately continuous monitoring of program impact. Each of these is discussed more fully in the paragraphs that follow. The knowledge resulting from each is not only important in deciding whether a given type of intervention is likely to be useful or not, but it also will indicate which experimental design to use and which variables to measure in an evaluation of new or on-going nutrition and health interventions.

A. Field Intervention Studies in Individuals

Some field intervention research is directed to identifying physiological and behavioral responses of individuals to public health intervention. It is not concerned with macrochanges at the level of the community. One might think that such individual responses could be done in a laboratory setting where conditions can be controlled. However, clinical and laboratory studies cannot substitute for a field study because the natural ecology of infections and nutritional stresses cannot be duplicated in metabolic wards.

An example of the value of such a field study is presented by Yarbrough and Habicht (2). Their population consisted of older preschool children whose dietary staple was maize, in whom they quantified the relative contribution to growth resulting from different doses of calories as contrasted to similar doses of calories combined with protein. Two unexpected results developed. The first was that a small increment in calories without added protein was as effective in improving growth as was a similar increment in calories with protein. The second was that even large improvements in protein-calorie nutrition could not improve growth rates to the levels seen in developed

countries, in all probability because of recurrent and frequent diarrheal disease. These findings contradicted the inferences drawn from research in laboratory and clinical settings which could not take the village ecology into account (3).

This kind of prospective experimental epidemiological field study requires a large array of measures and a rigorous experimental design to assure the comparability of data from those individuals receiving the different kinds of interventions. Only by such careful control can the inferences be sufficiently strong and generalizable to individuals in other populations to be of use to scientists and clinicians in that they are assured of the effectiveness of the intervention for individuals.

B. Community Field Intervention Studies

Field studies which show a benefit in individuals are not necessarily designed to show an effect at the community level. Yarbrough and Habicht, for instance, reported a clear benefit to some individuals without any measurable impact in the entire population. Another type of field research, which must be based on knowledge, such as that acquired under the type of study discussed in (A) above, is directed to evaluating community or a population response to an intervention. Are there enough individuals who respond sufficiently to the program so that one can identify a response from measurements aggregated at the community level? This research is also expensive, above all because it requires replication at the community level and must have as rigorous an experimental design as in (A). A good example is the group of fluoridation evaluation studies done by the U.S. Public Health Service which examined many indices of fluoride nutrition and many kinds of outcomes, including possible adverse side effects (4). Collectively, these clearly demonstrated the effectiveness of water fluoridation to prevent caries in communities.

C. Public Health Pilot Studies

Only when the results of field intervention studies have demonstrated effectiveness and thus promise a likely benefit from public health programs, is it worth investigating the feasibility and the cost-efficiency of large scale public health interventions. We call these public health pilot studies. On the basis of the knowledge gained from studies such as those described in A or B above, these pilot studies can utilize the best and most sensitive measures of mediating variables and of outcomes. Similarly, they only need to measure the few confounding variables which have been shown in study A or B to be important. However, they must also measure cost in such a way that the three components of the intervention can be examined separately to permit cost-efficiency analyses. No such analyses can be done without an experimental design which is at least as rigorous as in study (A) for cost-efficiency per individual benefited, or as rigorous as in study (B) for cost-efficiency of community impacts.

D. Evaluation During Implementation of Large-Scale Interventions

Once a pilot study has shown an intervention to be feasible, effective, and efficient, it may be extended more widely. A well-planned study needs to be conducted concurrently with this extension. It must be intensive and rapid, measuring those variables which are likely to change rapidly as the intervention is introduced. This implementation evaluation compares baseline data with that obtained later, or it compares geographic areas where the program has begun with those where it is about to begin.

The kinds of measurements are less numerous and more selective than those needed in the pilot study described in (C). The analysis of data should proceed quickly for each new geographic area benefiting from the introduction of the intervention to assure that the results correspond to those found in (C). If they do not correspond in spite of corrective action, either the chosen introduction is not feasible as a large scale public health activity, or the conditions which permitted the program to have an impact in study (C) do not hold at this large scale level.

E. Evaluation Through Monitoring

Once a program is implemented on a large scale, the only concern is that it results in an improvement compared to past health and nutritional status and that these results persist. This requires the establishment of a monitoring system which must be functioning adequately before the initiation of the public health intervention program (5). It requires no rigorous control group and a minimum of crucial measurements.

CHOICE OF EVALUATION INDICATORS

The choice of an indicator depends upon whether the evaluation is directed towards change in individuals or in communities.

Indicator Sensitivity in Individuals

An indicator of nutritional or health status of individuals must be responsive to the improved health or nutrition for which the proposed intervention is designed. In other words, the indicator must be responsive over the range of improvement expected.

This requires that there is an abnormal value for the indicator in individuals before intervention. Furthermore, the abnormality must be due to that element of nutrition or health which is to be improved. Thus, for example, if the intervention only improves protein quality of the diet in a population which is stunted because of inadequate caloric intake, the intervention will not improve growth (6). Historically, the health and nutritional factors responsible for abnormal

levels of indicators have often been incorrectly identified on the basis of descriptive studies which were not buttressed by intervention studies of the type described previously.

Even when an indicator's abnormal value is related to, or due to, the factor which the intervention is designed to improve, the indicator often may not be sensitive to improvement. This occurs because many indicators of health and nutritional status have been derived from comparisons between healthy, well-nourished individuals and clinically ill or malnourished patients. However, the majority of individuals in the usual target populations for large-scale interventions are not suffering extreme malnutrition or ill health.

The consequences of moderate malnutrition can often not be predicted from severe malnutrition. For example, the severe protein deficiency syndrome of kwashiorkor is accompanied by a deterioration of the body's defense mechanisms against infection and by impaired intestinal function, both of which result in diarrhea. However, protein deficiency sufficient to stunt growth does not result in increased diarrhea (7). Therefore, one may not presume that a strong effect on performance, health and survival, during severe malnutrition will necessarily lead to proportionately reduced indicator values under less severe malnutrition. In fact, trying to demonstrate the effectiveness of an intervention by using indicators demonstrated effective only under extreme conditions will usually fail.

Observations such as the above suggest that in many situations the dose-response curve may not be linear. Indeed, in those rare studies where one has looked for a dose-response on performance, health and survival, through improved nutrition in man, one finds a significantly lessened benefit as one improves nutritional state even at levels of nutrition universally accepted as inadequate (2,8). This means that for many indicators of performance, health and survival, one may not expect much improvement after intervention, unless the levels of the indicators in the malnourished population are quite different from normal levels in well-nourished regions (c.f. Figure 1).

Measured dose-response depends not only on the physiological response to dose consumed, but also, and sometimes importantly in field conditions, on the vagaries of measuring the intervention indicators and the response or outcome indicators. The larger and more frequent the errors of measurement of intervention and of outcome, the less sensitive will be the measured dose-response. For certain types of intervention (e.g., Vitamin A fortification of a food eaten occasionally by everybody) and for certain outcomes (e.g., an increase in fat folds) these vagaries in measurement can conceal any significant association between intervention and outcome. Where such errors are considerable, certain measurement strategies and statistical manipulations can help (9). It is much better to assure that the intervention and outcome variables chosen have little intrinsic variability and that the measurements are done carefully (10).

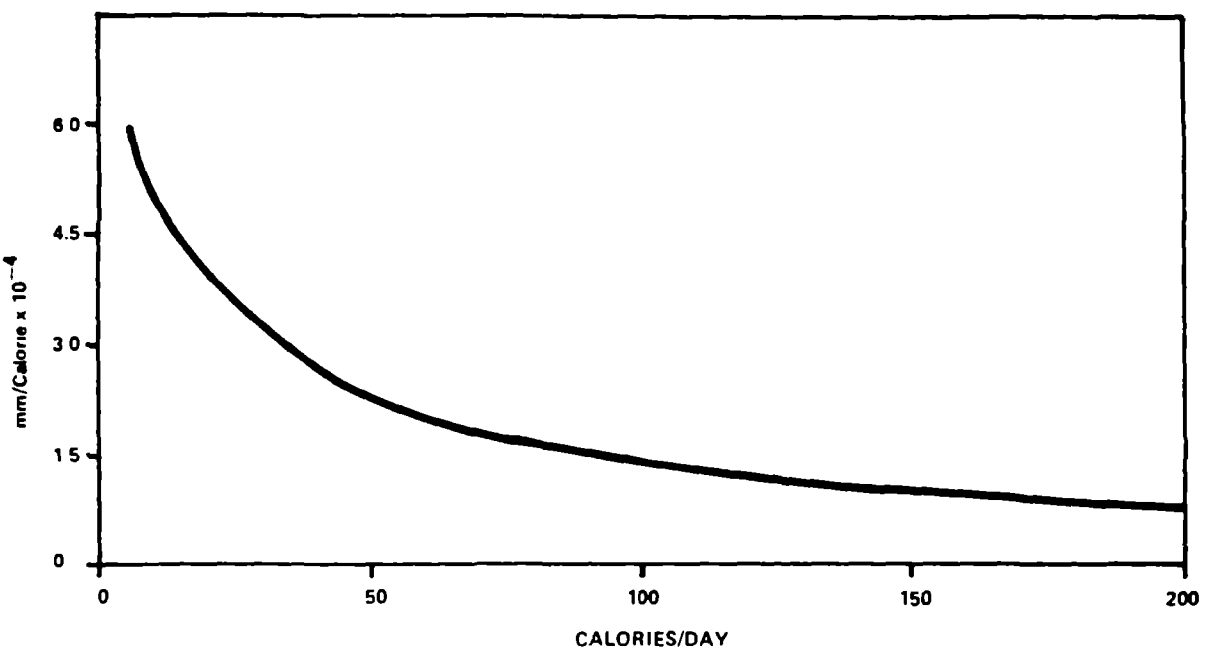

FIG. 1. Efficiency of calorie conversion to growth at different levels of calorie intake in young children. Based on two year increments in growth at different levels of supplementation in one year old children (2).

Indicator Sensitivity in Communities

So far we have discussed the effects of a health and nutrition intervention on indicators of performance, health and survival in individuals. Now we turn to problems which arise when one wishes to evaluate the results of an intervention on a population, rather than on individuals.

Evaluation of interventions in individuals usually depends upon sequential (longitudinal) measurements in the same individuals. The comparison is between improvement in those individuals who receive the intervention with the lack of improvement in those who do not. In contrast, the evaluation of interventions in populations often depends on sampling individuals at different points in time. Where the variability between individuals in some indicator is large compared to the expected response of that indicator to intervention, measuring different individuals each time instead of the same individuals longitudinally, will result in a marked decrease in the sensitivity of the indicator, similar to that which we noted will occur if the measurements are done poorly in individuals. The decision as to whether the improved sensitivity of evaluation acquired by longitudinally measuring the same individuals in populations is or is not worth the added cost and difficulty, as compared to sampling different individuals each time, can and should be calculated before intervention is started.

Usually, knowing the dose-response curve in individuals does not permit predictions about the effectiveness of an intervention in the community. One reason for this discrepancy can be found when a certain critical reduction in disease or disease-causing agents results in eradication of the disease from the community because a cause-effect chain is interrupted, as in malaria prevention programs. In such a case, a greater response is obtained than would be anticipated when looking at individuals alone. A similar situation can be postulated for the effect of natality of reducing infant and childhood mortality. If there is a sudden marked decrease in child mortality, it may be that the birth rate will decrease more rapidly than with an equal but more gradual fall in child mortality. A sudden increase in the number of infants and toddlers in the family may be more evident to the parents than would be a slow increase in the proportion of children who survive.

A more general reason why individual response rates to a given intervention do not predict population response rates to that same intervention, lies in the fact that the population response depends upon the characteristics of beneficiaries of the intervention compared to the rest of the population. Thus, one can expect different dose-responses in similar populations depending on the way the intervention is distributed. For instance, nutrition supplementation appears to be consumed in some nutrition programs inversely to the individual's needs (11). The impact of such a supplement will be negligible compared to a program with identical coverage which also assures that maximum supplementation is ingested where it is most needed.

For the above reason, many programs direct their interventions to those most likely to benefit. Other programs may cover the whole population but select for evaluation those who will most benefit. This selection is done on the basis of indicators of probable benefit. For this purpose, one must choose an appropriate "cut-off point" on that indicator which will permit the best selection (12). We call this characteristic the "selectivity of the indicator's 'cut-off point'". 1/

No variable is perfectly "selective". A child who is small for his age may be genetically stunted or he may be stunted for nutritional or health reasons. In the individual case one can ascribe a probability to the genetic and non-genetic possibilities if one knows two of the following three distributions: the distribution of growth of all children in the population of which the child is representative; the genetic distribution of sizes; or the distribution of stunted children. Figure 2 shows the probability of environmentally stunted growth at different weights among five-year-olds in a mixed population, half environmentally deprived (14) and half well-nourished (15). The smaller the child in this population, the greater the probability that the child's growth was stunted for non-genetic reasons.

Selectivity of an indicator's "cut-off point" does not refer, however, to the individual's probability of being malnourished or ill, but refers to the number of individuals who fall below that "cut-off point" because they suffer from malnutrition or ill health rather than because of genetic factors. Figure 2 also shows how this selectivity changes with different "cut-off" points in this same population of five-year-olds. Selectivity depends upon three characteristics of a dichotomous diagnostic variable at a specified "cut-off point": the measured or true prevalence of the disease; the proportion of all correctly diagnosed as ill for this disease (sensitivity of diagnosis); and the proportion of non-ill persons correctly diagnosed as not ill with this disease (specificity of diagnosis). Only one of these characteristics, the sensitivity of diagnosis, can be expected to remain constant under standardized conditions across different populations. It is obvious that the prevalence of disease can change. The specificity of diagnosis will also change with the changing prevalence of factors other than the disease or nutritional cause against which the intervention is addressed. Therefore, selectivity has to be estimated for each population.

Such estimates of selectivity will often show that an intervention can only have a modest impact even when it improves markedly the condition for all those who can benefit from it, because only a few of

---

1/ Clinical pathologists with a concern for prognosis have described the identical characteristic and called it "predictability" (13). We have tried to use the term "predictability" in the context of public health but found it so confusing that we have regretfully retained "selectivity" for this presentation.

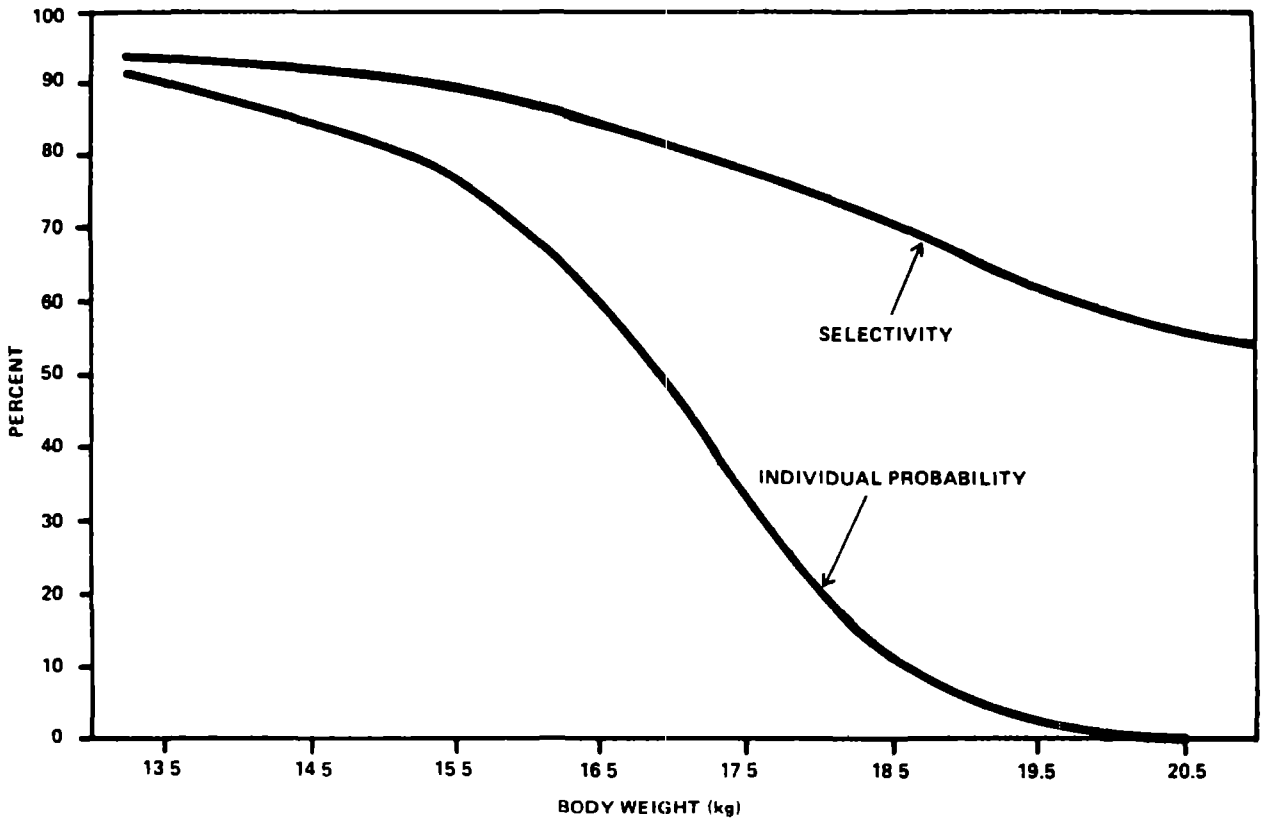

FIG. 2. Individual probability as contrasted to selectivity of environment stunting. Based on data for five-year-old boys of a mixed population with the same proportion of environmentally stunted boys (14) and well-nourished boys (15).

those selected can benefit. For instance, in the U.S.A., the official hemoglobin "cut-off point" for anemia is 12g. in Black women. This "cut-off" delivers a prevalence of 20% anemics among Black women, all presumed to be iron deficient based on the literature. In fact, probably less than 10% of those classified as anemic would benefit from iron therapy (16)--the selectivity of this hemoglobin "cut-off point" is, therefore, only about 10%. If each of those Black women who could benefit from iron therapy responded to an iron fortification program by raising their hemoglobin 2g., this increase in hemoglobin would be diluted to a mean 0.2g. increase among those classified as anemic. Such results would not indicate iron therapy to be an effective intervention if the selectivity were not known beforehand.

Thus, whether or not an indicator is sensitive at the level of the population, depends not only on its sensitivity at the level of the individual, but equally important, upon the selectivity of the indicator.

It is clear from this discussion that, to ensure that evaluation does not result in spurious negative results, no large-scale intervention program should be evaluated unless certain facts about the intervention, the measurements, and the population's probable response to the program are known beforehand. The easiest and safest way to elucidate these facts is by conducting carefully designed and implemented intervention studies in similar populations beforehand. Tables I.a-e present the data reported in nine such intervention programs. The specifics of these tables are discussed in the next section.

These Tables present the results of nutritional interventions in populations of preschool children with malnutrition and they reveal that growth in height is the most sensitive indicator, increments in weight are less sensitive, and improvements in health and post-infant survival are so insensitive that they cannot be used as indicators of nutritional status or to measure the effect of nutritional interventions in populations.

We do not review here the results of nutrition intervention studies in pregnant women because we reviewed this literature previously (25) and concluded that birthweight and duration of pregnancy probably were not related to nutrition of the mother except in sever maternal deprivation. We have since persuaded ourselves otherwise, at least as far as birthweight is concerned (26,27). We will try to justify our conversion at the end of the next section. We hope soon for a similar justification in the literature for thinking that infant mortality is sensitive to maternal and infant nutrition--but that is not yet available.

This evidence about the sensitivity of indicators can only come about from careful, well-designed intervention studies such as those described previously.

Intervention, Outcome and Intermediary Indicators

Later we will review the evidence for the sensitivity of outcome (impact) indicators which reflect cellular responses to improved

Table Ia. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON TOTAL DIETARY INTAKE

Reference: Years Publication	Supplement	Supervised/ Measured Ingestion	Replacement Estimated	Age in Months	Intake Before Intervention or Compared to Control	Intake after Intervention	Change in Intake
(17)-1963	391 Kcal.* 13.4 g[2]	No	Yes	48-96	From previous study 1580Kcal 37 g.	Not reported. Claim no substitution effect	?
(18)-1965	101-284 Kcal 9.8-10.1 g.	No	No**	6-12	Not measured	Not measured	?
(19)-1967-9	Not specified	No	Yes	0-59	678 Kcal. 20 g.	1040 Kcal. 30 g.	+362 Kcal. +10 g.
(20)-1970	250 Kcal. 12.5 g.	No	Yes*	36-96	Not reported	Not reported	?
(11)-1970	300 Kcal. 10 g	?	Yes	12-60	Not reported	Not reported. Claim no substitution effect	+300 Kcal.? +10 g.?
(21)-1970	?	Yes	Yes	35-59	486 Kcal. 13.7 g.	1181 Kcal. 35.3g.	+695 Kcal. +21.6 g.
(22)-1973	310 Kcal. 3 g.	Yes	Yes	12-60	700 Kcal. 18 g.	1010 Kcal. 21 g.	+319 Kcal. + 3 g.
(23)-1973	800-1300kJ. 6.7-11.8 g.	?	Yes	24-72	3700 kJ.	Not reported. Claim no substitution effect	+800-1300kJ.? +6.7-11.8g.?
(2)-1977	200 Kcal 14 g.	Yes	Yes	12-36	78% of recommended energy intake	Claim 10% substitution effect	+180 Kcal. +13.5 g.

Legend: * = Energy intake/day and Protein intake/day
** = Cause for doubting author's inferences
? = Uncertain

Table Ib. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON HEIGHT

Reference: Year Publication	Type of Height	Age in Months	Best-Worst = R **	Before Intervention or Compared to Control		After Intervention		Change in		Statist. Signif. ($p<.05$)
				Level	% of R	Level	% of R	Level	% of R	
(17)-1963	1 year increment in cm/year	48-96								
		Boys	6.0-5.6=0.4	4.3*	0%	4.4*	0%	0.1	-	No
		Girls	6.0-5.3=0.7	3.7*	0%	4.1*	0%	0.4	-	No
(18)-1965	6 month increment in cm/year	6-12	14-6.0 =8 0	7.4-7.6	18-20%	6.9-7.9	11-24%	-0.1	-1%	No
(19)-1967-9	3 year increment in cm/3 year	0-11	41-20=21.0	17.6*	< 0%	16.7*	< 0%	-0.9	-	No
		12-48	30-17.5=12.5	19.9	19%	22.0	36%	2.1	17%	Yes
(20)-1970	Attained developmental age	36-96	1.0-0 5=0.5	0.7	40%	0.7	40%	0.0	0%	No
(11)-1970	Attained height cm after two years supplementation	24-35	90-75= 15 0	77.1	14%	78.3	22%	1.2	8%	No
		36-47	99-81= 18.0	82.0	6%	84.6	20%	2.6	14%	No
		48-59	106-88=18.0	91.0	17%	90.9	16%	-.1	-1%	No
		60-71	113-95=18.0	97.4	14%	98.7	21%	1.3	7%	No
(21)-1970	6 month increment in cm/6 month	35-59	3.5-2.9= 0.6	1.94*	<0%	2.7	33%	0.8	133%	Yes
(22)-1973	14 month increment cm/14 month	12-23	12.5-6.8=5.7	6.5*	<0%	9.3	44%	2.8	49%	Yes
		24-35	9.5-6.8= 2.7	7.8	37%	9.5	100%	1.7	63%	Yes
		36-47	8.5-6.8= 1.7	7.4	•35%	9.1	135%	2.0	118%	Yes
		48-60	7.5-6.8= 0.7	7.3	71%	8.4	229%	1.1	157%	Yes
(23)-1973	6 month increment in cm/6 month	24-72	3.5-2.9= 0.6	2.0*	<0%	3.2	50%	1.2	200%	Yes
(2)-1977	2 year increment in cm/2 year	12	20-12= 8.0	15.7	46%	18.3	79%	2.6	33%	Yes

Legend: * This growth rate is lower than the lowest extrapolated from the most stunted population reported in the literature (24).

** R = Physiological Range (see text, page 155).

Table Ic. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON WEIGHT

Reference: Year Publication	Type of Weight	Age in Months	Best-Worst = R *	Before Intervention or Compared to Control		After Intervention		Change in		Statist. Signif. ($p < .05$)
				Level	% of R	Level	% of R	Level	% of R	
(17)-1963	1 yr. increment as kg/year	48-96								
		Boys	2.0<1.6=0.4	1.6	0%	2.1	125%	0.5	125%	No
		Girls	1.4=0.6	1.3	<0%	2.6	200%	1.3	217%	Yes
(18)-1965	6 month increment presented as kg/yr	6-12	4.0-1.5=2.5	2.2-2.3	29-32%	2.2-2.5	26-35%	.04	1%	No
(19)-1967-9	Regression Coefficient (kg/year)	0-11	6.0-3.5=2.5	3.6	4%	3.9	16%	0.3	12%	No
	Regression Coefficient (kg/3 yrs)	12-48	6.0-4.5=1.5	5.2	47%	5.7	80%	0.5	33%	Yes
(20)-1970	Attained Developmental Age	36-96	1.0-0.5=0.5	0.68	36%	0.68	36%	0.0	0%	No
(11)-1970	Attained weight after two yrs. supplementation (kg)	24-35	14.5-8.0 =6.5	8.8	12%	9.2	18%	0.4	6%	No
		36-47	15.5-9.0 =6.5	10.1	17%	10.9	29%	0.8	12%	Yes
		48-59	17.5-11.0=6.5	12.0	15%	12.0	15%	0.0	0%	No
		60-71	19.5-13.0=6.5	13.3	5%	14.0	15%	0.7	11%	Yes
(21)-1970	6 months increment (kg/1/2 yr)	35-59	1.0-0.7=0.3	1.23		2.28		1.0	350%	Yes
(22)-1973	14 month increment (kg/14 months)	12-23	2.6-1.7=0.9	1.74	4%	2.35	72%	0.61	68%	Yes
		24-35	2.3-1.7=0.6	1.71	1%	2.34	71%	0.63	70%	Yes
		36-47	2.3-1.7=0.6	1.58	<0%	2.04	38%	0.46	51%	Yes
		48-71	2.3-1.7=0.6	1.38	<0%	1.86	18%	0.48	53%	Yes
(23)-1973	6 month increment (Kg/1/2 year)	24-71	1.0-0.7=0.3	0.2	<0%	1.5		1.3	433%	Yes
(2)-1977	2 yr increment (kg/2 yr)	12	4.0-3.0=1.0	3.67	67%	4.50	150%	0.83	83%	Yes

Legend: * R = Physiological Range (see text, page 155)

Table Id. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON MORBIDITY

Reference Year Publication	Type of Morbidity	Age in Months	Before Intervention or Compared to Control (Level)	After Intervention (Level)	Change in Level	Statist Signif. ($p < .05$)
(18) 1965	Illness score based on effect of illness on growth rate	6-12	0.5 - 0.6	0.6 - 1.2	-0.3 Better	No
(19) 1967-9	Average days ill per year	0-59				
		No Intervention	13	22	+9 Worse	*
		Medical Interv.	71	48	-25 Better	-
		Nutrition "	10	46	+36 Worse	-
(11) 1970	% children with symptoms of protein-calorie malnutrition	12-60	23.0	11.3	-11.7 Better	Yes
			17.0	5.6	-11.4 Better	Yes

Legend: * No statistical significance testing done

Table Ie. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON MORTALITY

Reference Year Publication	Type of Mortality	Age in Months	Best-Worst R = *	Before Intervention or Compared to Control		After Intervention		Change in		Statist. Signif. ($p < 0.5$)
				Level	% of R	Level	% of R	Level	% of R	
(19) 1967-9	Infant (deaths/ yr/1000 births)	0-11	16-200=184							
		No Intervention		186	8%	191	4%	-5	Worse	No
		Medical Intervention		136	35%	88	60%	46	25%	No
		Nutrition Intervention		182	10%	146	29%	36	20%	No
	Preschool (deaths/ 1000 children)	12-48	0.3-90=89.7							
		No Intervention		81-9	10%	50-40	45%	31	35%	Yes
		Medical Intervention		50-40	45%	35-55	61%	15	17%	No
		Nutrition Intervention		56-34	38%	24-66	74%	32	36%	Yes
(20) 1970	Infant (deaths/ yr/1000 births)	0-11	16-200=184	135	34%	48	83%	87	47%	Yes
	Preschool (deaths/ 1000 children)	12-48	0.3-90=89.7	40	56%	22	76%	18	76%	No

Legend: * R = Physiological Range (see text, page 155)

nutrition because these indicators alone provide evidence of physiological benefit from a public health intervention. No evaluation study can, however, rely on such outcome indicators alone. These outcome indicators must be complemented by indicators which measure the intervention itself and its intermediary results. In field intervention studies the intermediary results are crucial for substantiating that the intervention caused the outcome. For this purpose the intermediary variables chosen will be those which biology indicates should change together, and they will be analyzed for such congruity of response. This analysis is imperative to substantiate causality between an intervention and a coincidental outcome.

The evaluation of all intervention studies is greatly facilitated if one knows how much intervention various members of the population receives relative to their needs. To achieve this, the intervention indicators should be as unambiguously tied to the intervention as possible. For instance, including in food supplements a tracer that can be measured in the urine permits one to ascertain who is consuming the food supplements. This and similar strategies for evaluating interventions are particularly essential when an intervention does not succeed in improving health, performance or survival. In such circumstances, one must differentiate between the question, "Was the failure because the intervention failed to reach those who needed it?" as contrasted to the question "Was the intervention itself inappropriately chosen?".

EVIDENCE THAT CONVENTIONAL HEALTH AND NUTRITION INDICATORS ARE SENSITIVE TO NUTRITION INTERVENTION

Establishing the Specificity of Indicator Response

To document the sensitivity of an indicator of nutritional status requires nutrition intervention studies, which demonstrate that the indicator responds to improved nutrition. Demonstration of such a response includes exclusion of the probability that the response was caused by non-nutritional factors. This exclusion, which assures the specificity of response to the nutrition intervention, can only be achieved by carefully designed and implemented intervention studies.

This section presents the criteria necessary to judge whether a response in an indicator was likely to be due to nutrition, in which case the indicator is sensitive to changes in nutrition, or whether the change could have been due above all to non-nutritional influences. In the context of testing the sensitivity of an indicator, these non-nutritional influences are "confounding" factors in statistical parlance.

The need to control for confounding factors is of course as important in evaluating the success of an intervention as in identifying sensitive indicators. Therefore the considerations reviewed in this section are important for designing all evaluations. This is especially true where the biological response to an intervention is under investigation as in Section (A) and (B) cited earlier, but

controlling for confounding becomes less important since evaluation is less concerned with proving intervention effectiveness and is more concerned with monitoring as one proceeds through the evaluations described in Sections (D) and (E). The reason for describing the control of confounding factors in this section, however, is not to prescribe experimental designs for intervention evaluation, but is rather to aid in judging whether a putative indicator of nutritional status has been demonstrated to be sensitive to changes in nutrition in individuals and populations where only a small minority suffer the florid clinical forms of kwashiorkor or marasmus.

We have discussed how the sensitivity of response is diminished by random errors of measurement and random variations in the indicators. These errors of measurement and other variations in the indicators are all due to factors other than those to which the intervention is addressed. They are, in that context, variations that are not specific to the purposes of the intervention. So long as these non-specific variations are random and their effects add up to zero, they only decrease the sensitivity of response. When, however, a non-specific influence changes the indicator among many individuals in the same direction, there is a danger that the resulting shift in the mean will be incorrectly attributed to the intervention.

There are basically three strategies to control for confounding factors: Controlled experimental designs; use of complementary indicators; and statistical analyses. The classical procedure is by experimental design where one compares the group benefiting from the intervention with a group similar in all relevant characteristics but which does not benefit from intervention (28). For instance, volunteering to participate in an intervention immediately introduces a bias if this group is to be compared to a control group which did not choose to take advantage of the intervention, because the factors which promote cooperation with the intervention program may also effect changes in the outcome indicators.

Good experimental design is the single most important factor necessary for successful evaluation. This depends upon careful formulation of the questions which the evaluation is supposed to answer. Defining the appropriate questions is facilitated if the practical consequences of alternative answers are specified. For instance, the question, "What are the correlates and consequences of participation by potential beneficiaries of a program?", is much less useful than asking, "Who needs the program? If these needy participate, is their performance, health or survival improved?" If not, "why not? What proportion of the needy participate? Why not?" Alternative answers to each of the questions in the latter series has immediate implications for program implementation. Experimental design is always slighted in compendia, such as this volume, because adequate treatment of the issue cannot proceed without addressing specific substantive questions. Generalizations on this issue have not been useful because apparently minor constraints on the use of "classical" experimental designs vitiate their usefulness and such constraints are the rule in field evaluations.

One particular constraint which results in falsely optimistic evaluations about a program's effectiveness, is the use of the same indicator to select those who are in need of the intervention, and to judge the response of the intervention in those selected. The use of such an indicator must correct for the indicator's inevitable regression towards the mean between the time of selection and the time of evaluation (29).

Evaluations which sample different individuals in a population instead of following individuals longitudinally must be particularly careful to ascertain whether population movements in and out of the intervention areas are not due to the immigrant's desire to cooperate with the intervention compared with the emigrants' indifference. In such a case, the immigrants may immigrate into the intervention area with better indicators of performance, health and survival than those of the emigrants, because those better indicators are associated with factors which promote cooperation with the intervention program, but are not due to the intervention program per se.

One of the great disappointments in evaluating intervention programs has been the discovery that comparisons between villages or regions often result in spurious differences due to non-specific influences which affect whole villages and regions. Often these effects cannot be explained, much less prevented (6). In this context, Gordon et al. stated that in the nutrition intervention studies they reported it was impossible to determine how much of the difference in effects observed between villages was due to the different interventions, to general secular trends which were different between the three villages, to sudden unexpected occurrences such as epidemics which infested villages differently, and to other unknown factors which might have affected the villages differently (19:VIII).

For instance, it is usual practice to ascertain through baseline surveys the comparability of villages with respect to the evaluation indicators. Figure 3, shows actual data on infant mortality rates for two villages chosen to be comparable in 1968 for a nutrition intervention that began in 1969. Comparing the rates in 1968 suggests that the two villages were quite similar with respect to this indicator. However, looking at the trends between 1960 and 1968 in these villages, sorely tries one's confidence in the comparability of future infant mortality data across the villages. Such confidence is, of course, a prerequisite for believing that the reduced mortality after the intervention seen in village A relative to village B is due to an intervention applied to village A and not to village B.

Therefore, any experimental design which does not randomly distribute the intervention and its control within a village or a region must have sufficient villages or regions covered by each treatment (replicates) so that one can estimate the probable contribution of non-specific influences at the village or regional level. Adjacent villages and regions must have different treatments, and the villages and regions should be so stratified that any other random non-specific influences are controlled for. Designs which show differences between regions or villages but do not have these required replicates must remain suspect.

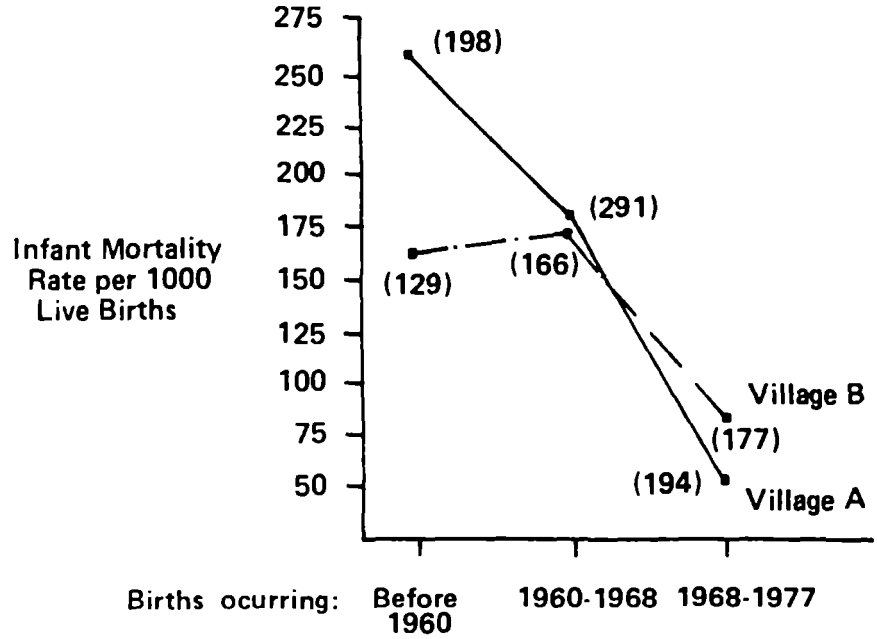

FIG. 3. Infant mortality rate trends in two villages chosen for a nutrition intervention. Source: Female Retrospective Life History Questionnaire from *INCAP-RAND Survey in Guatemala.*

Notes: Numbers of live births in each period are in parentheses. All rates are calculated from retrospective data for comparability. We can identify no reporting bias that would have differentially affected the two villages; nevertheless, this possibility exists.

The second and complementary strategy to avoid misinterpreting a change in levels of an indicator consequent to intervention is to measure various complementary indicators (30). Each indicator should measure some different step between the intervention and its outcome. For instance, if increased caloric intake of the pregnant mother is supposed to be responsible for a subsequent improved survival of the infant, then one should find that improved caloric intake leads to greater maternal skinfolds, to a greater maternal weight gain during pregnancy, to a greater birthweight of the infant and to a greater infant skinfold, as well as to improved infant survival. Furthermore, all these variables should show a statistical association with each other. The choice of these complementary indicators and their expected statistical associations requires a clear conceptualization based on previous demonstration of the expected effect of the intervention on performance, health and survival. If influences not related to the intervention's objective affect one of the indicators, it is unlikely that they will affect the whole chain of indicators. Thus, for instance, if improved caloric intake by pregnant mothers was accompanied coincidentally by improved medical care, and it was the medical care which improved infant survival, one would not find the statistical links between the intermediary variables linking improved maternal nutrition to improved infant survival. If the whole chain of indicators are congruously affected, one can assert that the nutrition or health has been improved by the intervention or physiologically similar influences. A decision as to whether or not it was due to the intervention itself depends on adequate statistical design.

The third strategy to control for confounding factors is to measure these and take them into account when analyzing the data (c.f. Habicht et al., 31). This requires the identification of the variables possibly confounding in the context of the intervention, and the indicators of outcome which will be evaluated. The first consideration must depend upon knowledge of the population's psychological, social, economic and environmental circumstances as they relate to the intervention and to this population's participation with the intervention. Estimates must then be made of the expected effects of these behavioral and environmental biases on the indicators. The appropriate measures of the confounding factors or their proxies can then be chosen.

It is not possible to provide a list of confounding variables that are relevant in every health or nutrition intervention. At the bottom of Table II, we list the possible confounding factors which, in our opinion, could have been measured and evaluated in the studies reviewed there. Butz and Habicht (32) give a more complete list and discuss methodological considerations that arise in evaluating their effects.

It is important to emphasize that none of these strategies: correct experimental design, measurement and analysis for congruence, and measurement and analysis for confounding factors can substitute for each other. They have been presented in their order of importance. There can be no possible useful conclusion without adequate experimental design. Interpretation of positive effects in an outcome

Table II. SUMMARY OF DESIGN AND INDICATOR CHARACTERISTICS FROM INTERVENTION PROGRAMS IN TABLE I.

References. Year of Publication	(17) (1963)	(18) (1965)	(19) (1967)	(20) (1970)	(11) (1970)	(21) (1970)	(22) (1973)	(23) (1973)	(2) (1977)
I. Choice of Intervention									
a. Unit of Intervention	Village	Child	Village	Village	Village	Child	?	Child	Child
b. Was choice of type and quality of intervention based on more evidence than dietary survey information?	No	No	Yes	Yes	?	?	Yes	Yes	Yes
II. Control of Intervention									
a. Did intervention reach central distribution center in adequate quantity?(weighed)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
quality?(by assay)	Yes	?	Yes	?	?	?	Yes	?	Yes
b. Did intervention reach home by documented record of distribution?	-	Yes	Yes	Yes	-	-	-	Yes	-
c. Did intervention reach target persons in adequate quantity by documented quantitative measurement by occasional dietary survey?	-	?	Yes	Yes	-	Yes	-	Yes	Yes
frequent record of ingestion of intervention?	Yes	?	No	No	No	Yes	Yes	No	Yes
d. Was replacement effect sought measured adequately and taken into account?	No	No*	Yes	No*	No	Yes	Yes	No	Yes
III Choice and Measurement of Indicators									
a. Was initial level low for Diet? Anthropometry? Morbidity? Mortality? Other?		See Table I							
b Was analysis made of variability due to measurement and short-term intrinsic variability?									
Diet?	No	No	No	No	No	No	No	No	Yes
Anthropometry?	No	No	No	No	No	No	No	No	Yes
Morbidity?	-	No	No	-	-	-	-	-	-
Mortality?	-	-	No	No	-	-	-	-	-
Other?	No	-	-	-	-	-	-	No	-
IV. Control of Confounding Factors									
a Adequate controls	Yes	Yes	Yes	Yes	No*	No*	?	Yes	No
b Replication	Some	Yes	No*	Some	No	-	?	Yes	Yes
c Analysis for Congruity									
Dietary Ingestion?	Yes	No	No	No	No	Yes	No	No	No
Anthropometry?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Morbidity?	-	No	Yes	-	-	-	-	-	-
Mortality?	-	-	No	No	-	-	-	-	-
Other?	Yes	-	-	-	-	-	-	No	-
d Analysis Stratified by									
Age	Yes	Yes	Yes	No*	No	No	Yes	No	Yes
Sex	Yes	No	Yes	Yes	No	No	Yes	No	Yes
Dose of Intervention	-	No	No	No	No	No	No	No	Yes
Indicators of Self-selection	No	No	No	No	No*	No*	No	-	Yes
Lactation	-	No	No	No	No	No	No	No	-
Food Supply or Home Diet	No	No	No	No	No	No	No	No	No
Income or Health	No	No	No	No	No	No	No	No	No
Education of mother or other family member	No	No	No	No	No	No	No	No	No
Medical care, disease risk or disease experience	Yes	No	No*	No*	No	No	No	No	No
Secular trend	No	No	Yes*	Yes	No	-	No	-	-

Legend - = not relevant, ? = not clear in report, * = cause for doubting authors' inferences Where * is associated with "No", it indicates that we consider this a serious defect which calls into question some or all of the authors' inferences. Where * is associated with "Yes", it indicates that the authors tried to take this into account but we judge either that their methodology was inadequate or that we come to a different inference from the results than they did.

indicator must be reinforced by analysis of other indicators for congruity to be persuasive. Analysis of possible confounding factors reinforces the assertion that a change in an indicator of nutritional status was, indeed, due to the intervention and not to some confounding factors (see Table II).

Presentation of Specific Intervention Studies which Tested the Sensitivity of Indicators of Nutritional Status

Tables I. a-e present results from nutrition intervention studies at the individual or population levels directed against protein-calorie malnutrition in preschool children who live at home. These tables present those variables which have been studied often enough to be tabulated. This is a rather small subset of the many variables proposed for evaluation studies (33-35). Other variables have not been reported frequently enough in intervention literature to be able to ascertain whether they will or will not respond to nutrition intervention programs directed against protein-calorie malnutrition under field conditions.

Where possible, we have tried in Table I to set limits or maximal ranges within which the indicator is expected to vary. The level seen or recommended in developed countries we call the "best" level. The "worst" level is that which we believe is compatible with survival of the population in its present numbers. 2/

We convert the levels of the indicator before and after intervention to a percent of the physiological range, R, by subtracting the value of the indicator from the "worst" level and dividing this difference by the range R to deliver a percentage. Thus, in Table I.c the first study measured one year increments in weight (17).

In developed countries this one year increment is 2 kg over the age period under consideration. Under the "worst" conditions in the literature the one year increment in boys is 1.6 kg. the range, R, is 0.4 kg. The control group of boys who received no supplement gained 1.6 kg per year, no better than the "worst" expected growth. Their percent of R was therefore (1.6-1.6)/0.4 = 0%. The boys who received

---

2/ In Table I, the Worst and the Best situations for height, weight and mortality were estimated in the following fashion:

Weight and height: attained and increment: The "best" levels were data taken from a well fed population (15). The "worst" are taken from a compilation (36). In preschool children, measuring the increment from these data produces the same result as measuring the mean increment in children followed longitudinally [see comparative data in Hansman (37)], which is not the case in adolescence.

Mortality: The best levels were those reported to the United Nations Organization (38). The worst levels were taken from our guess based on our experience in underdeveloped countries, where deaths are underreported.

the supplement gained 2.1 kg per year. Their percentage of R was therefore (2.1-1.6)/04 = 125%. Thus if the indicator is already at levels similar to those of developed countries, the percentage of range will approach 100%. If the indicator reflects conditions similar to those worst conditions reported in the literature the percent of the range will approach 0%. A comparison between the percentage of the adequacy before the intervention and that after intervention gives an idea of change during the intervention.

The last column of Table I indicates whether the authors reported a statistically significant change in the indicator subsequent to intervention as compared to a control group. The next step is to decide whether a negative finding in this column reflects insensitivity of the indicator or ineffectiveness of the nutritional intervention.

Evidence that the Nutrition Intervention Actually Improved the Diet

Unless a dietary intervention of adequate quality can be shown to have reached target individuals in sufficient quantity, the failure to find an effect on outcome indicators could be due to an inadequate intervention. Demonstration of adequate quantity of intervention requires that one show that individuals consume an adequate amount of supplement (Table II. Section II.c), and that this increase was not lost through a comparable reduction in the usual home diet consumption. The latter is referred to as dietary substitution (Table II. Section II.d). Only careful, well-designed surveys entailing large sample sizes of about 600 person-measurements in each comparison group can discard the possibility of physiologically important substitution of 5% or more of home diet. These calculations are based on the fact that the day to day variability in the intake of individuals in malnourished populations has a coefficient of variation of about 35% for protein and calories. Therefore, where the claim for no substitution effect was made on the basis of small dietary surveys this is considered an inadequate response to the question, "Was replacement effect sought, measured adequately, and taken into account?" In such cases the estimated "change in intake" is followed by a question mark in Table I.a. This table presents the dietary intervention data on the nine studies reviewed. Because none of the studies presented age distribution data, we could not estimate minimum protein-energy requirements for survival ("worst" case) or for maximum growth ("best" case) so as to judge how much the intervention would be expected to improve the diet. However, one study reported such a high protein-energy intake before intervention that one wonders whether one could expect any improvement from the intervention (17). For five of the studies the data presented does not permit an estimate of improved dietary intake (see last column). The four studies which adequately measured substitution effects represent interventions which should have resulted in some beneficial outcomes for the participating preschool children inasmuch as the baseline protein and especially the energy intakes were low in all, and the improvement of the protein-energy intake was substantial: from 45% to over 140% of the energy contained in the initial diet. The latter increase seems hardly

believable (21). Either the initial diet was not compatible with life or the increase was not compatible with physiological ingestive capacity.

Sensitivity of Height and Weight to Improved Nutrition

Tables I.b and I.c report the outcome indicators of growth in height and weight. Of the five studies in which one could not judge the quantity of the intervention, only one showed any consistent improvement in growth (23). Two others showed inconsistent results. On the other hand all of the four studies which presented adequate evidence of an appropriate nutrition intervention also reported statistically significant increments in growth for preschool children after infancy (2,19,21,22). Of these studies, only one measured infants and this study revealed no improved growth during infancy due to supplementation of the infants or of their lactating mothers. Thus, one may conclude that the outcome indicators of height and weight are sensitive in preschool children to factors associated with the intervention. Martorell, et al., reached the same conclusion in their review of post-infancy growth and protein-calorie supplementation (39).

Whether or not the effect on growth was due solely or even principally to the nutritional component of the intervention must be addressed by examining the adequacy of the control groups, the outcome of analyses for congruity, and inspection of concurrent changes in confounding variables (Table II. Sections IV. a-d). In this context one of the five positive studies selected the intervention group from volunteers and compared that group to non-volunteers (21). The results of the statistical analyses cannot, therefore, be ascribed to the nutrition intervention. Another of these studies reporting positive results is not specific enough concerning its comparison groups for one to be sure they were appropriate, although careful reading of the report inclines one to believe they were (22). A third study has inappropriate comparison groups but presents convincing data to show that the improvement in growth is not due to factors affecting consumption of the supplement (2).

Appropriate comparisons imply adequate estimates of random variability to decide if a difference is significant. One study (19) did not have enough replicate groups to do so; another (22) may have had such replicates, but did not present the necessary analyses for the reader to judge.

Congruity analysis of the five positive studies separately (Table II) results in discarding one (21) as presenting very unlikely results, and the discarding of the infant data in another (19) because the values of growth in height and weight are incompatible with infant physiology. The preschool data in the third (23) presents non-intervention data for weight increments which are so low as to appear incompatible with the survival of a stable population. The response to intervention overshoots normal growth enormously during a period of six months. This data is not, however, so incongruous with present knowledge that we reject it.

From a practical point of view, we conclude that at least certain studies demonstrate that height and weight are sensitive to nutrition intervention programs, especially if measured longitudinally in the same children. In such cases, height is considerably more sensitive to intervention in the sense of producing greater statistical significance than is weight--a fact which we have commented upon elsewhere (40).

Sensitivity of Morbidity and Mortality to Improved Nutrition

Table I.d presents the reported effects on morbidity of nutrition intervention in these studies. The only study (11) which presents evidence of any positive effect of nutrition on illness had comparison groups which were self-selected and could be expected to show this pattern without nutrition intervention. The study most cited (19) as showing a beneficial nutritional effect on mortality does not show it at all in the data as presented in Table I.d, and the argument in the report based upon a change in secular trends is not convincing, especially in view of the lack of replicability in the experimental design. This lack of a beneficial effect on the morbidity of free-living malnourished but ambulatory children corresponds to our findings of a marked effect of disease on child growth but no effect of growth stunting on the incidence of disease (7).

Table I.e presents the reported effect on mortality of nutrition interventions in two studies (19,20). The first of these is the study most cited as showing a beneficial effect on mortality from improved nutrition, but in fact it shows no greater improvement due to nutrition than is evident in one of the comparison groups. Again, the argument made in the report that the nutrition-supplemented group was in some way better is based on secular trends which are not replicated in the experimental design, a defect recognized and mentioned by the authors of these reports and brought to the attention of the reader in one (19:VIII).

Gordon et al., pointed out in reporting on his field study that in developed countries only about 30% of all infant deaths occur after the first months of life (postneonatal), whereas about 60% of all infant deaths are postneonatal in developing countries. However, this ratio of postneonatal to infant mortality is influenced by so many factors that it is probably not useful for evaluation of this type of intervention. For instance, in the village receiving medical care, the ratio worsened from 54% to 69% of postneonatal deaths over all infant deaths at the same time as the total infant death fell from 136 to 88 per 1000 live births.

In the other study (20), data are presented suggesting that infant mortality is improved by nutrition intervention, but the authors could show no effect whatsoever on growth. This lack of congruity makes us believe that the beneficial effect reported was due to other causes, such as differing medical care. In practical terms we must conclude that morbidity and mortality are not sensitive enough indicators to serve as evaluation indicators for the effect of nutrition intervention in preschool children.

Sensitivity of Birthweight and Infant Mortality to Improved Nutrition of the Mother

We have reviewed elsewhere our reasons for believing that birthweight can be a useful indicator to evaluate nutritional intervention in malnourished mothers (26,27). Although many reports in the literature would contradict this belief, we found that these reports either did not substantiate improved dietary intakes or that the studies were done in well-nourished populations (25). Our conversion to thinking that birthweight could reflect individual and aggregate maternal malnutrition in malnourished populations was based on a study which controlled for many confounding factors by experimental design. The distribution curve of birthweight was more affected by maternal supplementation among the lower than among the higher birthweights, which is congruent with a physiological effect associated with supplementation rather than with biases in measurement (8). This study also examined the relative additive and synergistic effects of important influences, including nutrition, on birthweight (41), and explored which confounding factors might explain the association between intervention and outcome (41). This study found congruent dose-responses of the home diet and of the nutritional intervention (27). The dose-response was independent of when and of how long the supplement was consumed during pregnancy (42). This is congruent with efficient maternal storage of energy during pregnancy for use when the energy requirements of the fetus are greatest towards the end of the third trimester. The birthweight was more significantly affected by maternal nutrition than was the length of the newborn, in contradiction to the pattern after birth, but congruent with a transfer of more energy from the better nourished mother to her child toward the end of pregnancy. There is to date only this one study which shows such clear-cut results that birthweight will increase if nutrition is improved in malnourished pregnant women. Belief based on the results of one study border on faith, and corroborative evidence is sorely needed from an independent research team.

This evidence for an effect of maternal malnutrition on infant mortality was vigorously and eruditely denied until the nutrition community came to believe that maternal nutrition affected birthweight on the evidence of the single study reported above. Early data from that study gave some basis for hope that the clear association between low birthweight and infant mortality was mediated by malnutrition, rather than by gestational prematurity, intrauterine infection or other similar non-nutritional courses (41). Until more definitive data is available, projected estimates of infant lives to be saved by national nutrition programs are premature. In particular it is probable that nutrition intervention which is not coupled with primary medical care (24) will not be cost-effective, nor perhaps even beneficial in terms of improved infant health and well-being (8).

In conclusion to this section, we were amazed at how little competent work has been done in the evaluation of nutrition interventions directed towards remedying protein-energy malnutrition in populations. Before much more can be said about choosing sensitive indicators for such evaluations, more candidate indicators must be tested in careful

field intervention studies as described earlier in (A) and (B). This research cannot be accomplished by evaluations of pilot or large-scale interventions because such interventions cannot assure the specificity of response necessary to identify sensitive indicators of nutritional status.

CAN INTEGRATED INTERVENTION PROGRAMS BE EVALUATED?

Our discussion so far has treated issues of measurement and evaluation of simple interventions: those in which the treatment consists of only one or a few changes introduced as part of a single intervention. As Table II indicates, successful application of optimal design and evaluation principles has been rare in areas of nutrition and health, even when interventions were of this simple type. These studies could have taken advantage of experimental design techniques for which the standard principles and techniques were developed.

More often than not, nutrition and health interventions are not of this simple type. Therefore, we turn now to discuss the special measurement and evaluation complications that arise in complex integrated interventions in which intentional changes are introduced in medical, nutritional, social, political and economic factors. Evaluation is complicated immensely in these situations. The greater the number of factors purposely changed, the more difficult it will be to estimate the separate effect of changes in each factor on chosen indicator variables. More importantly, replication of the intervention also will be more complicated.

Setting for Large Scale Integrated Interventions

Before outlining the nature of these measurements and evaluation difficulties, it is useful to ask why integrated interventions are becoming popular. Frustration of researchers as well as policy-makers with their meager results from simpler specific interventions is probably an important factor. Their frustration is partly due to several of the problems discussed above leading to failure of many simple interventions to produce measurable and important changes in indicators of health and nutrition in individuals and more frequent failure to produce significant results at the population level.

Health and nutrition professionals are also increasingly concerned that very specific interventions may be ineffective within the same institutional, technological and socio-economic environment that originally led to poor nutrition and health. The concern is well founded in light of the rich interactions that characterize biological, social, economic and agricultural systems in poor populations. It is clear that nutrition and health outcomes emerge from these interactions. However, a change introduced as an intervention may not affect these outcomes if other factors in the environment are in fact limiting improvements in health and nutritional status. In addition,

individuals can frequently take advantage of these interactions to turn the effect of an outside intervention away from that intended to a direction they prefer.

Let us illustrate the biological, socio-economic and agricultural interactions that commonly exist in economically poor environments. Consider that the nutrition and health of a family's members (and, therefore, of a population) in a poor area are influenced by the amounts and types of food produced or transported into the region, the types and distribution of food storage facilities, the prices of nutritious and non-nutritious foods, the prices and availability of medical care, and people's knowledge and beliefs concerning food and medical care. The family's income and wealth also directly influence the family members' consumption of food and their use of traditional and/or modern medical care. In addition, the complex of economic and other factors that influence how people spend their time can have significant indirect effects on nutrition and health. As an example, in communities where women have incentives to work away from home, breast-feeding is less common and those women with the higher work incentives tend to lactate the shortest period. These women may also spend less time in food preparation and home health care. As another example, changes in the amount of time children spend working and in school may affect both their own nutritional requirements and the effectiveness of their mothers in meeting these requirements. In poor populations family members' nutritional status also depends on patterns of food distribution within the family and on determinants of these patterns. We discuss elsewhere the role of many of these factors in influencing women's breastfeeding behavior (32).

Is is thus clear that nutrition and health are among the many outcomes of the biological, social, economic and agricultural systems that interact within families and communities. If these systems are equilibrating systems--that is, if the pattern of observed outcomes reflects an optimal allocation of the family's or community's resources given the biological, technological and economic conditions that exist--then particular changes in the environment may, indeed, be ineffective. For example, a particular intervention may only relax a non-effective constraint. Increasing food grain production will not increase rural people's grain consumption nor nutritional status if facilities for storing grain from plentiful to sparse seasons are already inadequate or if the high cost of transporting the grain to other communities prevents the farmers from selling their higher production for income. Reducing the price of existing means of storage and transport or introducing new technologies in these areas would enable the community to benefit from increased food production.

Similarly families may find it in their interest to reduce their own health- or nutrition-producing activities in response to an intervention that independently contributes to health or nutrition. A school supplementation program cannot be expected to increase children's food ingestion by the amount of supplementation. Poor parents respond to the school feeding program as though it were a decrease in the effective price of food and therefore encourage their children to consume somewhat more food at school. They then divert resources from children's food to food for other family members and to expenditures

on non-food commodities such as shelter or clothing. As long as these substitution possibilities exist, persons can be expected to make use of them in order to increase their perceived benefit from an intervention. The result is a smaller change in the indicator variables than might be expected. The less the people in a community value better health and nutritional status relative to other things, the more they will rearrange the allocation of their resources to transform a nutrition and health intervention into benefits that they value more highly.

For both these reasons there is considerable appeal to shifting from simple interventions of the classic experimental design toward integrated interventions that change a number of conditions thought to be limiting to better nutrition and health. Furthermore, an integrated intervention study may well be the most cost-effective way to elucidate critical facts necessary for effective public health policy for certain important questions.

Approach to Evaluating Integrated Interventions

When intentional changes are introduced in medical, nutritional, social, political and economic factors, evaluation is complicated immensely in these complex interventions. The greater the number of factors purposefully changed, the more difficult it is, in general, to estimate the separate effect of changes in each factor on chosen indicator variables. More importantly, replication of the intervention is also more complex. Finally, if the intervention is unsuccessful, it is more difficult to find out why--which factors were responsible.

Unfortunately, inadequate formulation, operation, measurement or evaluation results in programs of doubtful benefit and even more doubtful replicability. It is here that a clear perception is necessary as to the nature of the program. Is it an intervention study designed to prove biological relationships between an intervention and outcome in a free-living population whose characteristics are well defined? Alternately, is the program a pilot study which is based upon the proven results of a field intervention study and which tries to accomplish an outcome which is known to be sensitive to the intervention under the expected constraints of a large-scale public health program? Or perhaps it may be the large-scale public health program itself.

Formulating a comprehensive intervention requires considerably more knowledge about the structure of the complex system from which people's nutritional status and health emerge than does formulating a simple experimental design. In addition to following the important considerations discussed in previous sections, designers of integrated interventions must also use procedures that maximize the probability that the set of chosen interventions has a significant effect on the indicators while simultaneously minimizing the losses from undesirable side effects--and do this all on a budget.

The first goal is not so difficult. The list of things causally associated with good health and nutrition is long, and we know how to change many of these things, from food production and distribution to water, sewage, and hygiene. By intervening in enough ways, a significant result is nearly guaranteed. The evidence is all around, however, that significant undesirable side effects are nearly impossible to avoid in large interventions, and often very difficult to measure. Our limited scientific understanding of the linkages within and among biological, behavioral and agricultural systems does not facilitate identification of the many outcomes that may be affected by a single change in the environment, much less by multiple changes. These issues as well as the identification of promising integrated interventions, can only be determined by rigorous intervention studies. Only these will reveal integrated interventions that are likely to succeed within the resources available to public health programs. Intervention studies and pilot programs will reveal the initial outcome variables and side effects which must be measured in the evaluation of large-scale integrated public health programs.

After an integrated intervention has been formulated and tested in field studies, additional problems will arise during its operation as a pilot study. Chief among these is the temptation, sometimes explicitly encouraged, to alter the set of interventions in mid-course as experience accumulates. The decision is admittedly a difficult one. If it is clear that an additional change should have been added to the intervention set based on the emerging data concerning limiting conditions or the efficacy of the existing intervention, one would like to make the indicated change for the duration of the intervention. Similarly, the initial intervention set may have been well formulated, but conditions have changed, due either to natural responses to the intervention or to independent changes in the environment. Making the indicated change in mid-course increases the expected change in the indicator variables. The result of such an operational change is to hamper the measuring of the experimental treatment and of evaluating the intervention. The cost of the pilot study which revealed these deficiencies will, however, have been small in comparison to making these mistakes in the course of a large-scale public health program. And another pilot public health program can substantiate that the new integrated intervention is replicable and cost-effective.

Another pitfall in the operation of integrated interventions arises when part of the intervention consists of political, social or educational activities intended to organize people or redirect their attention toward goals of the intervention. The exact nature of such interventions invariably shifts and adjusts as the intervention proceeds, making it very difficult to measure and keep track of just what the intervention was at particular times. Even if accurate records are kept, evaluation of these parts of the intervention and possibly of the entire effort is in jeopardy because the form of the intervention has become endogenous and dependent on the population's responses. How then can one examine these or related responses to evaluate the effectiveness of the intervention? In addition, political and social interventions are very difficult to replicate since, among other reasons, their outcomes generally depend on personalities

and activities of the persons who are intervening. Hence, replicating the treatment across villages in the experimental design, as recommended above, is difficult. Replicating it later in other sites is even more difficult. However, adequate knowledge about the crucial linkages acquired through intervention studies and substantiation of the replication in pilot studies will increase the likelihood that the large-scale intervention will be successful and be adequately evaluated and monitored for maximum cost-effectiveness.

Proper experimental design is critically important in the case of complex integrated interventions. For simple experiments there is the possibility of making inferences about cause and effect through multivariate statistical analyses if the design is flawed, as long as most conditions in the population have stayed static. When many factors are intentionally changed, on the other hand, inadequate replication and controls leave one unable to untangle the mass of changes and make statements about nutrition and health effects due to the intervention. This problem is exacerbated if the effects of the components of the intervention are not additive. Indeed, one generally acts as if they are not, since one generally tries to combine complementary interventions in the hopes that their result will be greater than the simple addition of their effects (synergistic action).

In our opinion, the evidence that many kinds of behavioral and biological processes are interrelated does not imply the conclusion that interventions must be broad and complex to produce lasting changes in health and nutrition indicators. The implication instead, is that the specific interventions chosen must be those that change the conditions--institutional, economic, biological--that are limiting in a particular setting. By changing specific conditions in different experimental settings, by formulating the experiments wisely, and by measuring the treatments, major confounding factors, and indicator variables intelligently, we can accumulate understanding of the relevant mechanisms and the particular interventions that are effective in specific settings.

One should design the integrated intervention studies so that this synergistic effect is sufficiently understood so that it can be applied cost-effectively. For instance, perhaps one wishes to know which combination of medical care, environmental sanitation, nutrition education, and food subsidies will result in cost-effective results in health. Single-purpose intervention studies such as those reviewed earlier, will have delivered the best indicators for improvements in health and nutrition. By using these indicators to evaluate different combinations of interventions one can evaluate new additions to previous combinations, beginning with the least expensive and most feasible first, and progressing to more expensive additions later, until the benefits no longer justify further increments in cost. This procedure will not permit a teasing out of the synergistic from the additive effects of the intervention, but will reveal a good candidate for pilot testing and ultimately for large-scale intervention. The alternative is complex large-scale interventions that are costly to operate, costlier to replicate on a national scale, and from which little can be learned if the intervention fails and nothing can be safely changed in the future if it succeeds.

CONCLUSION

We conclude that an indicator must be sensitive to the intervention and be specific for that intervention. Knowledge about sensitivity must come from single purpose intervention studies. In these studies specificity of the indicator's responses to nutrition is assured by controlling non-nutritional influences through intervention, experimental design and statistical analyses. The procedures, therefore, are clear for amassing the knowledge necessary to implement and evaluate the nutritional and health impact of large-scale integrated public health programs. It is distressing that we have followed this procedure so little that most of the crucial indicators necessary for evaluation of interventions have not even been properly tested. In particular, only height and weight have been reliably shown to be sensitive to improved protein-calorie nutrition in preschool children, and height is more sensitive than weight. Birthweight as an indicator of maternal nutrition has so far only been shown to be sensitive to improved nutrition in one study--this finding must be replicated before it can be generalized. Morbidity and mortality appear to be poor and insensitive indicators of nutritional status in the preschool years, although infant mortality may ultimately prove to be a useful indicator where medical care is otherwise adequate.

The little experience to date in testing these very few of the many indicators suggested for evaluation should be a warning that pronouncements, even by renowned authorities, cannot replace validation by single purpose intervention studies. Relaxing "scientific" constraints to recommend the continued use of "commonly used" even unvalidated indicators will only result in falsely negative evaluations of valuable and useful programs. It would be better to recognize our limited knowledge about useful indicators and to use them gingerly when we must do so now. We should proceed as quickly as possible to focus single purpose intervention studies to validate "commonly used" indicators and to develop other better indicators.

ACKNOWLEDGEMENT

This is a report of research of the Cornell University Agricultural Experimental Station, Division of Nutritional Sciences.

REFERENCES

1. Habicht, J-P., and M.C. Latham. Ethics of population research in population. (Submitted for publication).

2. Yarbrough, C., J-P. Habicht, R.E. Klein, et al. Response of indicators of nutritional status to nutritional interventions in populations and individuals. In: S.J. Bosch and J. Arias (eds.). Evaluation of Child Health Services: The Interface Between Research and Medical Practice. DHEW Publication No. (NIH) 78-1066. Washington, D.C.: U.S. Government Printing Office, 1977. pp. 195-206.

3. Scrimshaw, N.S., and M. Behar. Malnutrition in underdeveloped countries. New Engl. J. Med., Part I. 272: 137-144, 1965; Part II, ibid. 272: 193-198, 1965.

4. McClure, F.J. Fluoride in Drinking Water. U.S. Public Health Service Publication No. 825. Washington, D.C.: U.S. Government Printing Office, 1962.

5. Habicht, J-P., J.M. Lane, and A.J. McDowell. National nutrition surveillance. Fed. Proc., (in press).

6. El Lozy, M., and G.R. Kerr. Results of lysine fortification of wheat products in southern Tunisia. pp. 113-133. In: H.L. Wilcke, (ed.). Improving the Nutrient Quality of Cereals. Washington, D.C.: U.S. Agency for International Development, 1977.

7. Martorell, R. Illness and Incremental Growth in Young Guatemalan Children. Ph.D. dissertation. Seattle: University of Washington, 1973.

8. Habicht, J-P., A. Lechtig, C. Yarbrough, and R.E. Klein. Maternal nutrition, birth weight and infant mortality. pp. 353-378. In: K. Elliott, and J. Knight, (eds.). Size at Birth. CIBA Foundation Symposium 27, London: Elsevier, 1974.

9. Habicht, J-P., and C. Yarbrough. Growth monitoring and nutritional assessment: anthropometric field methods; criteria for selection. In: R. Alfin-Slater, and D. Kritchvsky, (eds.). Nutrition: A Comprehensive Treatise. New York: Plenum Press, (in press).

10. Habicht, J-P. Estandarizacion de metodos epidemiologicos cuantitativos sobre el terreno. Bol. Of. San. Pan., 76: 375-384, 1974.

11. Swaminathan, M.C., D.H. Rao, R.V. Rao, et al. An evaluation of the supplementary feeding programmes for preschool children in the rural areas around Hyderabad City. Indian J. Nutr. Dietet., 7: 342-350, 1970.

12. World Health Organization. Methodology of Nutrition Surveillance. Report of a Joint FAO/UNICEF/WHO Expert Committee. WHO Technical Report Series No. 593. Geneva: World Health Organization, 1976.

13. Galen, R.S, and S.R. Gambino. Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses. New York: Wiley & Sons, 1975.

14. Yarbrough, C., J-P. Habicht, R. Malina, et al. Length and weight in rural Guatemalan ladino children: birth to seven years of age. Am. J. Phys. Anthropol., 42: 439-447, 1975.

15. National Center for Health Statistics. 1976 NCHS Growth Charts. Monthly Vital Statistics Report, 25(3) Supplement, 1977.

16. Meyers, L., J-P. Habicht, and C. Johnson. Components of the difference between black and white women in the USA of hemoglobin concentration in blood. Am. J. Epidemiol., 1979.

17. King, K.W., W.H. Sebrell Jr., E.L. Severinghous, et al. Lysine fortification for wheat bread fed to Haitian school children. Am. J. Clin. Nutr., 12: 36-48, 1963.

18. Bancroft, T., and K.V. Bailey. Supplementary feeding trial in New Guinea highland infants. J. Trop. Pediat., 11: 28-34, 1965.

19. Series of articles written with various coauthors by Gordon, J.E., N.S. Scrimshaw, M.A. Guzman, M. Behar, and W. Ascoli with alternating first authorship: Nutrition and infection field study in Guatemalan villages. Guatemala: Institute of Nutrition of Central America and Panama, (INCAP), 1959-1964.
I. Study plan and experimental design. Arch. Environ. Health, 14: 657-662, 1967; II. Field reconnaissance, administrative and technical; study area; population characteristics; and organization for field activities. Arch. Environ. Health., 14: 787-801, 1967; III. Field procedure, collection of data and methods of measurement. Arch. Environ. Health., 15: 6-15, 1967; IV. Deaths of infants and preschool children. Arch. Environ. Health, 15: 439-449, 1967; V. Disease incidence in preschool children under natural village conditions, with improved diet, and with medical and public health services. Arch. Environ. Health, 16: 223-234, 1968; VI. Acute diarrheal disease and nutritional disorders in general disease incidence. Arch. Environ. Health, 16: 424-437, 1968; VII. Physical growth and development of preschool children. Arch. Environ. Health, 17: 107-118, 1968; VIII. An epidemiological appraisal of its wisdom and its errors. Arch. Environ. Health, 17: 814-827, 1968; IX. An evaluation of medical, social, and public health benefits, with suggestions for future field study. Arch. Environ. Health, 18: 51-52, 1969.

20. Baertl, J.M., E. Morales, G. Verastegui, and G.G. Graham. Diet supplementation for entire communities. Growth and mortality of infants and children. Am. J. Clin. Nutr., 23: 707-715, 1970.

21. Kamalanathan, G., G. Nalinakshi, and R.P. Devadas. Effect of a blend of protein foods on the nutritional status of preschool children in a rural balwadi. Indian J. Nutr. Dietet., 7: 288-292, 1970.

22. Gopalan, C., M.C. Swaminathan, V.K. Srishna Kumari, et al. Effect of calorie supplementation on growth of undernourished children. Am. J. Clin. Nutr., 26: 563-566, 1973.

23. Rajalakshmi, R., S.S. Sail, D.G. Shar, and S.K. Ambody. The effects of supplements varying in carotene and calcium content on the physical, biochemical and skeletal status of preschool children. Brit. J. Nutr., 30: 77-86, 1973.

24. Habicht, J-P., and Working Group on Rural Medical Care. Delivery or primary care by medical auxiliaries: Techniques of use and analysis of benefits achieved in some rural villages in Guatemala. pp. 24-37. In: Medical Auxiliaries. PAHO Scientific Publication No. 278. Washington, D.C.: Pan American Health Organization, 1973.

25. Lechtig, A., G. Arroyave, J-P. Habicht, and M. Behar. Nutricion materna y crecimiento fetal. Arch. Latinoamer. Nutr., 21: 505-530, 1971.

26. Lechtig, A., J-P. Habicht, C. Yarbrough, et al. Influence of food supplementation during pregnancy on birth weight in rural populations of Guatemala. pp. 44-52. In: A Chavez, H. Bourges, and S. Basta, (eds.). Nutrition. Vol. 2. Basel: S. Karger, 1975.

27. Lechtig, A., J-P. Habicht, H. Delgado, et al. Effect of food supplementation during pregnancy on birthweight. Pediatrics, 56: 508-520, 1975.

28. Federer, W.T. Experimental Design. New York: MacMillan, 1963.

29. Davis, C.E. The effect of regression to the mean in epidemiologic and clinical studies. Am. J. Epidemiol., 104: 493-498, 1976.

30. Habicht, J-P., C. Yarbrough, and R.E. Klein. Assessing nutritional status in a field study of malnutrition and mental development: specificity, sensitivity, and congruity of indices of nutritional status. pp. 35-42. In: J. Cravioto, L. Hambraeus, and B. Vahlqvist, (eds.). Early Malnutrition and Mental Development. Symposia of the Swedish Foundation, Vol. XII. Uppsala: Almqvist and Wiksells, 1974.

31. Habicht, J-P., C. Yarbrough, A. Lechtig, and R.E. Klein. Relation of maternal supplementary feeding during pregnancy to birth weight and other sociobiological factors. pp. 127-145. In: M. Winick, (ed.). Nutrition and Fetal Development. New York: Wiley Interscience, 1974.

32. Butz, W.P., and J-P. Habicht. The effects of nutrition and health on fertility: hypotheses, evidence and interventions. pp. 210-238. In: R.G. Ridker, (ed.). Population and Development. Baltimore: Johns Hopkins University Press, 1976.

33. Jelliffe, D.B. The Assessment of the Nutritional Status of the Community (With Special Reference to Field Surveys in Developing Regions of the World). pp. 3-27. WHO Monograph Series No. 53. Geneva: World Health Organization, 1966.

34. Gordon, J.E., and N.S. Scrimshaw. Field evaluation of nutrition intervention programs. World Rev. Nutr. Diet., 17: 1-38, 1973.

35. Beaton, G.H., and J.M. Bengoa. Nutrition in Preventive Medicine. WHO Monograph Series No. 62. Geneva: World Health Organization, 1976.

36. Habicht, J-P., R. Martorell, C. Yarbrough, et al. Height and weight standards for preschool children. Lancet, 1: 611-615, 1974.

37. Hansman, C. Anthropometry and related data. pp. 101-154. In: R.W. McCammon, and V.E. Beales, (eds.). Human Growth and Development. Springfield: Charles C. Thomas, 1970.

38. United Nations. Demographic Yearbook 1974. New York: United Nations, 1975.

39. Martorell, R., A. Lechtig, C. Yarbrough, et al. Protein-calorie supplementation and postnatal physical growth: a review of findings from developing countries. Arch. Latinoamer. Nutr., 26: 115-128, 1976.

40. Yarbrough, C., J-P. Habicht, R. Martorell, et al. Anthropometry as an index of nutritional status. Adv. Exp. Med. Biol., 49: 15-26, 1974.

41. Habicht, J-P., C. Yarbrough, A. Lechtig, and R.E. Klein. Relationships of birthweight, maternal nutrition and infant mortality. Nutr. Rep. Int., 7: 533-546, 1973.

42. Habicht, J-P., A. Lechtig, C. Yarbrough, and R.E. Klein. The effect on birth weight of timing of supplementation during pregnancy. p. 71. In: P. Arroyo et al., (eds.). Sumaria - IX International Congress of Nutrition. Mexico, 1972.

THE IMPACT OF IMPROVED WATER SUPPLIES AND EXCRETA DISPOSAL FACILITIES ON DIARRHEAL MORBIDITY, GROWTH AND MORTALITY AMONG CHILDREN[1]

Steven A. Esrey[2] and Jean-Pierre Habicht[2]

---

[1]Supported in part by USAID grant number OTR-2199 and the Cornell Nutritional Surveillance Program Cooperative Agreement No. AID DSAN CA-0240 between USAID and Cornell University.

[2]Division of Nutritional Sciences, Savage Hall, Cornell University, Ithaca, NY 14853.

The authors express appreciation to Maureen Lewis (USAID) for useful suggestions, Penny Spingarn for editorial assistance, and Marilyn Ward and Elaina Jeddry for manuscript preparation.

ABSTRACT

Many studies have tried to associate water supply and sanitation facilities with health status. The published results are confusing and contradictory. This review examined the nature of the relationship between poor water and sanitation and diarrheal morbidity, growth faltering and mortality in children by systematically assessing the available reports according to their research design and health outcome. The evidence shows that improving excreta-disposal facilities has been more effective than improving water supplies for lowering diarrheal morbidity and mortality rates and improving children's growth. Water supply improvements by themselves appear to have limited effectiveness in improving health. A conceptual framework has been developed to reduce methodological problems when evaluating existing and future water or sanitation programs.

BIOLOGICAL LINK BETWEEN WATER, EXCRETA DISPOSAL, AND CHILD HEALTH

Diseases related to poor excreta-disposal facilities and inadequate or contaminated water, especially diarrheal diseases, are still a major health problem in the less-developed countries. Diarrheal diseases impose a heavy toll in terms of the number of children affected and the severity of the consequences. From 500 million to 1 billion episodes of diarrhea occur annually among African, Asian, and Latin American children (1). Diarrhea leading to dehydration, electrolyte imbalance, and sometimes fever is the major cause of death in young children: estimates range from 3.9 to 18 million deaths per year (1, 2). Diarrhea may also debilitate children slowly through anorexia, improper feeding, decreased absorption of nutrients, and increased catabolism. Growth faltering or, ultimately, death may occur if the effects of diarrhea are severe and protracted.

Most of the pathogenic agents causing diarrhea in infants and children are transmitted via the fecal-oral route. That is, fecal pollution of the environment leads to the oral ingestion of pathogenic agents. The fecal-oral route is also the pathway of some non-diarrheal disease agents, such as poliomyelitis and Ascaris lumbricoides. Inadequate waste disposal is associated with all of these diseases (figure 1).

Poor water quality and availability are also related to the persistence of diarrheal diseases. Contaminated water is thought to be mainly responsible for the spread of cholera, typhoid, and giardiasis. These diseases are called water-borne diseases (3, 4), because water is the vehicle for transmission for the disease agents, although they can also be transmitted by contaminated food and hands. Pathogens leading to diarrhea but related more to water availability are Shigella, Entamoeba histolytica, enterovirus, and rotavirus. These diseases are called water-washed diseases (3, 4) because it is likely that contamination of food, utensils and hands by these disease agents would be prevented if there were sufficient water, even of poor quality, for washing.

Improvements in water and sanitation are thought to affect health, mainly by reducing the ingestion of fecal-oral pathogens (figure 1). However, this is not always the case. For instance, hookworm infection, which may affect child growth, will not be affected by improvements in water alone, but may be reduced by sanitation improvements. On the other hand, water supply improvements may lead to improved health through home gardening, improved child care as a result of time savings, or other income generating activities. However, figure 1 does conform to the fact that reducing diarrhea is probably the major mechanism whereby children's health can be improved. Therefore, this paper examines this mechanism in detail.

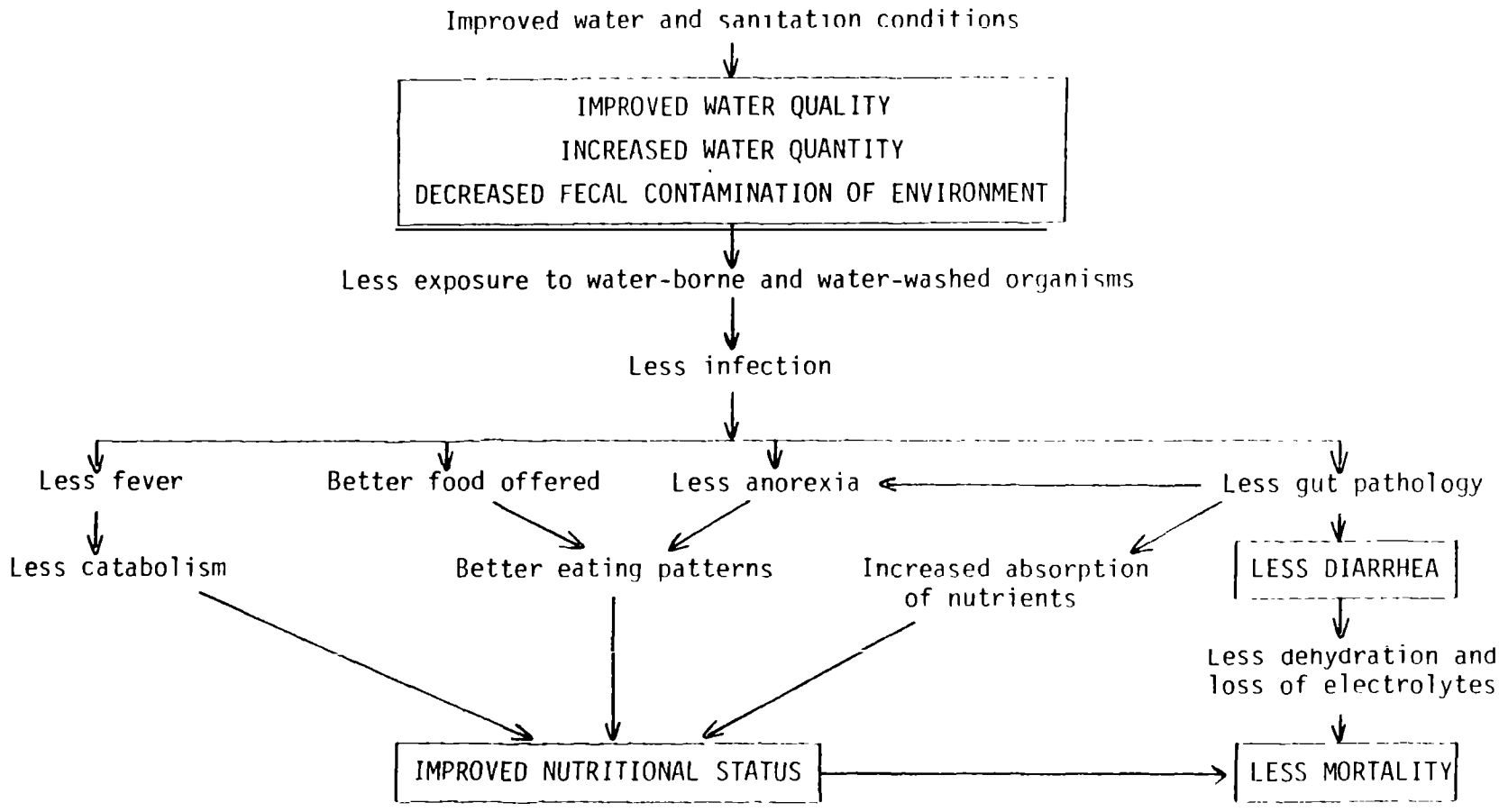

FIGURE 1. Schematic presentation of the sequence of expected benefits from water and sanitation improvements. (The most easily measured outcomes are boxed.)

The United Nations has launched a worldwide effort to break the fecal-oral transmission of diarrheal diseases by declaring the 1980s to be the International Drinking Water Supply and Sanitation Decade (5). The intention is to provide everyone with access to safe drinking water and waste-disposal facilities by 1990. This task is formidable, because of the great number of people who need this service, and because of the great cost. In 1975 only an estimated 22 percent of the rural population in less-developed countries, excluding China, had access to safe water, and only 15 percent had access to adequate waste disposal (6). Urban areas were somewhat better off: 77 percent of the people had access to safe water and 75 percent had access to adequate waste disposal. Thus, improved water and sanitation facilities need to be extended to 500,000 people a day during the 1980s in order to achieve the U.N. goal. The cost of accomplishing this within 10 years has been estimated at $60 billion for adequate worldwide water supplies and $300-600 billion for sanitation (7). The per-capita cost is $200-400, a cost that exceeds the yearly per-capita income of most of those in need.

Investments of this magnitude could be justified by showing that such improvements in water and sanitation have been cost-effective. These calculations are difficult to do with any precision because estimates of effectiveness are elusive. Despite the fact that the steps needed to determine cost-effectiveness appear to be straightforward, the true benefits of water and sanitation improvements are not known. This review evaluates the published studies that have looked at improvements in water and sanitation in terms of health benefits--specifically, lower morbidity and mortality, and better growth in children.

FOCUS OF REVIEW

The review evaluates published studies examining water and/or sanitation conditions that quantified differences in diarrheal morbidity or specific pathogens, nutritional status, or mortality. Improved sanitation measures include the provision of flush toilets and pit latrines. Improved water supplies include taps and handpumps. All of these improvements can be installed at the community or household level. The review attempts to be exhaustive of all studies since 1950; however, in a few cases, the data were too difficult to obtain or were too meager. Some of the studies included did not specifically analyze water or sanitation conditions but provided information on these conditions and health.

For review, all of the studies considered are categorized according to health outcome: diarrheal or pathogen morbidity (table 1), nutritional status (table 2), or mortality (table 3). The tables summarize the location, type of comparison, ages of children studied, and the results reported. The review section highlights only selected studies, to exemplify major methodological and design issues, that could explain the results--whether positive, negative, or inconclusive. These issues form the basis in the discussion section for

TABLE 1. The effect of water supply and excreta disposal conditions on intestinal infections, infestations and diarrhea

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1982 BGD (8)	cholera - percent positive	W - contaminated vs. uncontaminated	All ages					
		drinking		48.9	22 5	26.4	-54%	p < .05
		cooking		51.8	13.9	37.9	-73%	p < 001
		bathing		50.9	14 3	36.6	-72%	p < .001
		washing		50.9	14 3	36 6	-72%	p < .001
1980 BGD (9)	cholera - percent positive	W - non-tubewell vs tubewell users	All ages	11	12	-1 0	+9%	(NS)
		- contaminated vs. uncontaminated		11.3	1 2	10.1	-89%	p < 001
1977 BGD (10)	diarrhea/ dysentery - percent positive	W - ground vs tubewell	0-12	4.3	3.4	0 9	-21%	-
			12-120	4.7	5 8	-1.1	+23%	-
		- surface vs. tubewell	0-12	9.1	3.4	5.7	-63%	-
			12-120	11.0	5 8	5 2	-47%	-

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1977 BGD (11)	cholera - % pos.	W - other vs. tubewell	All ages	1 0	4 3	-3.3	+330%	p < 001
	Shigella - % pos.			1 5	0.9	0 6	-40%	NS
	diarrhea - per 1,000			117 7	133 5	-15.8	+13%	-
1977 BGD (12)	cholera - attack rate per thousand	W+E - non-tubewell and surface latrine vs tubewell and sanitary latrine	All ages	4 1	1 6	2 5	-61%	p < 01
1976 BGD (13)	cholera - incidence rates	W - non-tubewell vs. tubewell	All ages	8 4	14.2	-5 8	+69%	p = 08
	diarrhea - per 1,000			3 2	7.5	-4.3	+134%	p = .07
1972 BGD (14)	cholera - percent positive	W - tubewell far vs. near	12-48	39 1	12.5	26 6	-68%	(NS)
			48-60	20.0	0.0	20 0	-100%	p < .001
	- classical		0-12	17.6	27.3	-9 7	+55%	(NS)
	- el tor		12-48	40.0	43 8	-3 8	+10%	(NS)
1978 COL (15)	diarrhea - prevalence	W - unpiped vs. piped	0-72	42 4	45.3	-2 9	+7%	NS

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1978 COL (16)	diarrhea - prevalence	W+E - hygiene scale	school children	-	-	-	-44%	p < 0001
1965 CR (17)	diarrhea - morbidity rate per 1,000	W - worst vs best district	0-48	1515	1450	65	-4%	-
		E - worst vs best district		905	1456	-551	+61%	-
1954 EGY (18)	E. histolytica Giardia lamblia - percent positive	W - control + E (wells and latrine)	12-48	16	16	0	0%	NS
		- treatment		32	51	-19	+59%	-
1977 ETH (19)	gastroenteritis - days sick per year	W - community vs. private	0-24					
		< 10 lcd		25	30	-5	+20%	-
		> 10 lcd		27	9	18	-67%	-
		E - public vs. private						
		< 10 lcd		28	31	-3	+11%	-
		10-20 lcd		37	13	24	-65%	-
		> 20 lcd		30	8	22	-73%	-

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1978 GUA (20)	diarrhea - prevalance per 1,000	W - unpiped vs. piped	All ages	39.7	39 5	0 2	-1%	NS
1963 GUA (21)	diarrhea - incidence per 100 persons per year	W - public vs private faucet	0-60	38.9	32.4	6 5	-17%	NS
		E - without vs. with privy	0-12	52 8	80 7	-27.9	+53%	NS
			All ages	22.4	16.8	5.6	-25%	$p < .01$
1957 GUA (22)	Shigella - percent positive	W - community vs. private	0-120	9 4	6 3	3.1	-33%	$p < .05$
		E - areas with <50% vs >50% of houses with privies and flush toilets		11.2	4 8	6 4	-57%	$p < 05$
1980 HAI (23)	diarrhea - percent positive	W - <1 vs. >1 can per person per day	0-60	28.7	25.5	3.2	-11%	NS

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1978 IND (24)	gastroenteritis - percent	W - area with 27% vs. 48% house connections	All ages	4	0	4	-100%	(NS)
1977 IND (25)	diarrhea - incidence per 100 per year	W - well vs. taps in homes	0-60	21.5	23.5	-2	+9%	NS
		- street taps vs taps in homes		36 8	23 5	13.3	-36%	p < 01
	Shigella - incidence per 100 per year	W - well vs. taps in homes		10.3	6.0	4.3	-42%	p < .001
		- street taps vs taps in homes		15 3	6.0	9 3	-61%	p < .001
1971 IND (26)	gastroenteritis	W - pre-post chlorination	All ages	11 0	1 1	9 9	-90%	p < 01
		- control		10.6	17 7	-7 1	+67%	NS
	dysentery	W - pre-post chlorination		7 3	1 0	6 3	-86%	p < .01
		- control		7 4	10 5	-3.1	+42%	NS

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1970 IND (27)	diarrhea - 3 month moving averages	E - none vs bore-hole latrine	Children	9 4	7 4	2 0	-21%	-
1966 IRN (28)	diarrhea - percent positive	W - unpiped vs piped	0-84	48 7	36	12.7	-26%	p .001
	Shigella - percent positive	- unpiped vs piped		7.5	4 5	3 0	-40%	NS
1972 KEN (29)	diarrhea - percent positive	W - surface vs. piped	All ages	19	3	16	-84%	p .1
mdated KEN (30)	diarrhea -	W - far + unpure vs near + pure	0-12	23 3	18 8	4 5	-19%	-
			12-36	16 6	6 3	10 3	-62%	-
			36-72	7 0	7 3	-0 3	+4%	-
1978 LES (31)	diarrhea - cases of diarrhea all non-water cases	W - poor vs good quality	0-48	.52	.48	.04	-8%	NS
1981 LIB (32)	E histolytica - percent positive	W - poor vs good quality	All ages	13	11	02	-15%	-
	G. lamblia - percent positive			.009	009	0	0%	-
	Bacillary dysentary - percent positive			84	58	.26	-31%	-

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1971 PAN (33)	Shigella - percent positive	W - stream vs. piped	0-120	1 5	4 3	-2.8	+187%	-
		stream vs well		1 5	2 0	-0 5	+33%	-
		E - pit privy vs flush		2 9	5 0	-2 1	+72%	-
		ground vs pit privy		1.0	2.9	-1 9	+190%	-
1974 PHI (34)	cholera - rates per 1,000	W - poor vs. improved	0-48	542.2	213 7	328.5	-61%	-
		E - poor vs improved		542 2	321 1	221.1	-41%	-
		W+E - poor vs improved		542 2	193.1	349.1	-64%	-
1981 STL (35)	diarrhea - percent positive	W - public vs. household	7	35 4	35.9(23 7)	-.5(11 7)	+1%(-3%)	-(-)
			10	30.9	33 4(17 7)	-2 5(13 2)	+8%(-43%)	-(-)
		(W+E - public vs household)	13	19 8	16.1(7.2)	3 7(12.6)	-19%(-64%)	-(-)
			16	22 1	13.0(11 1)	9 1(11 0)	-41%(-50%)	-(-)
			19	19 1	10.1(3.8)	9.0(15.3)	-47%(-80%)	-(-)
			22	22 9	8.8(3.3)	14.1(19.6)	-62%(-86%)	-(-)
			25	5.7	5 4(5.2)	.3(0.5)	-5%(-9%)	-(-)
1978 SUD (36)	diarrhea - avg. days ill/person/6 months	W - <26 lcd vs. >40 lcd	All ages	15.6	7.5	8.1	-52%	p < .001

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1966 SUD (28)	diarrhea - % pos.	W - unpiped vs piped	0-84	26.6	21 8	4 8	-18%	(NS)
	Shigella - % pos.	- unpiped vs piped		14	6 4	7 6	-54%	p < .01
1966 UAR (28)	diarrhea - % pos	W - unpiped vs. piped	0-72	38	10	28	-74%	p < .0001
	Shigella - % pos	- unpiped vs piped		9.4	3.4	6	-64%	p < 01
1978 UK (37)	diarrhea - % pos.	W - pre (supply cut 17 hr) vs. post (no restriction)	School children	5 6	3 3	2 3	-41%	p < 001
		- pre (supply cut 12 hr) vs. post (no restriction)		4.9	3.2	1 7	-35%	p < .001
		- pre (no restriction) vs. post (no restriction		3.9	3 5	0.4	-10%	p < 005
1969 USA (38)	diarrhea - avg. clinic visit per child	W - indoor plumbing	0-12	2.0	.85	1.15	-58%	p < .05
		- control		3.1	2 60	0.50	-16%	NS

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Difference	Percent Reduction	Statistical Significance Reported (Calculated)
1960 USA (39)	diarrhea - attack rate per 100	W - low vs high quality	0-24	21 6	9 4	12.2	-56%	-
1959 USA (40)	Shigella - % pos	W - off vs. on premises	0-60	6.0	5.8	0 2	-3%	(NS)
		- out vs in premises		5.9	2.4	3.5	-59%	(p < 05)
		E - out vs. in premises		2.4	1 1	1.3	-54%	(p < 05)
1957 USA (41)	diarrhea - incidence rate per 1,000	E - latrine	0-120	17.3	10.4	6 9	-40%	(NS)
		- control		20 3	20 5	-0 2	+1%	(NS)
	Shigella - percent positive	- latrine		6.0	3.9	2.1	-35%	(NS)
		- control		2.8	7.3	-4 5	+161%	(NS)
1955 USA (42)	Shigella - percent positive	W - far vs near well vs. city	0-120	5.8	4 1	1.7	-29%	p < 025
				5 2	4.8	0.4	-8%	NS

Year Country[a] (Ref.)	Type of Indicator	Type of Comparison[b]	Age in Months	Before Intervention or Control	After Intervention or Treatment	Differ-ence	Percent Reduction	Statistical Significance Reported (Calculated)
1955 USA (43)	Shigella -% individuals positive	W - out vs. in homes (unmatched)	0-120	5.8	3.0	2.8	-48%	($p < .05$)
		- out vs. in homes (matched)		5 9	1.2	4 7	-80%	($p < 05$)
	-% families positive	- out vs. in home (unmatched)		11.0	6.2	4.8	-44%	($p < .05$)
1953 USA (44)	Shigella - percent positive	W - >15 vs <15 people per faucet (out of home)	0-120	9 2	5.3	3.9	-42%	$p < .0001$
1966 VEN (28)	diarrhea - % pos.	W - unpiped vs piped	0-84	43	39	4	-9 %	NS
	Shigella - % pos	- unpiped vs. piped		7 6	3 0	4 6	-61%	$p < 001$
1976 ZAM (45)	diarrhea - % clinic cases/1,000	W - unpiped vs. piped	All ages	338	212	126	-37%	-

[a] Country code BGD = Bangladesh, BRA = Brazil, COL = Colombia, CR = Costa Rica, EGY = Egypt, ETH = Ethiopia, FIJ = Fiji, GUA = Guatemala, HAI = Haiti, IND = India, IRN = Iran, KEN = Kenya, LES = Lesotho, LIB = Libya; MAL = Malaysia, NIG = Nigeria, PAN = Panama, PHI = Philippines, SRL = Sri Lanka, STL = St. Lucia, SUD = Sudan, UAR = United Arab Republic, UK = United Kingdom, USA = United States, VEN = Venezuela, ZAM = Zambia.

[b] W = water supply, L - excreta disposal

TABLE 2 The effect of water supply and excreta disposal conditions on nutritional anthropometry

Country Year (Reference)	Type of Indicator[a]	Type of Comparison	Age in Months	Physiological Range[b] = R	Before Intervention or Control		After Intervention or Control		Change		Percent Change[c]	Statistics Reported (Calculated)
					Level	%R[b]	Level	%R[b]	Level	%R[b]		
Colombia 1975 (55)	W/A - % > 90% standard	Excreta - latrine vs sewer system	0-32		26		47		21		+81%	p < .01
		Water - no bath vs bath			22		51		29		+132%	p < .0005
	H/A - % > 95% standard	Excreta - latrine vs sewer system			26		48		22		+85%	p .01
		Water - no bath vs. bath			20		53		33		+165%	p < .0005
Fiji 1983 (56)	W/A - mean percentile	Excreta - non-flush vs flush	0-60									
	Urban				102		100		-2		-2%	NS[d]
	Rural				95		102		7		+7%	p < .0001
	H/A - mean percentile											
	Urban				100		100		0		0%	NS
	Rural				99		103		4		+4%	p < .005
Nigeria 1978 (57)	W/A - % > 75% standard	Water - far vs near	6-48		50		69		19		+38%	p .05
	H/A[4] - % > 90% standard				80		69		-11		-14%	NS
	W/H - % > 80% standard				63		90		27		+43%	p < .01

Country Year (Reference)	Type of Indicator[a]	Type of Comparison	Age in Months	Physiological Range[b] R	Before Intervention or Control Level	Before Intervention or Control %R[b]	After Intervention or Control Level	After Intervention or Control %R[b]	Change Level	Change %R[b]	Percent Change[c]	Statistics Reported (Calculated)
St Lucia	Weight-growth	Water - public	3-6	7	1 32	31%	1 59	70	27	39%	+20%	p 01
1981	increment	vs private	6-9	7	92	74%	99	81%	07	10%	+8%	NS
(35)	(kg)		9-12	5	67	54%	69	58%	02	4%	+3%	NS
			12-15	6	70	83%	71	85%	01	2%	+1%	NS
			15-18	5	67	84%	71	102%	09	18%	+15%	NS
			18-21	3	66	153%	58	127%	- 08	-27%	-12%	NS
			21-24	4	66	115%	58	95%	- 08	-20%	-12%	NS
		Water + excreta -	3-6	7	1 32	31%	1 43	47%	11	16%	+8%	NS
		public vs	6-9	7	92	74%	99	84%	07	10%	+8%	NS
		private	9-12	5	67	54%	70	60%	03	6%	+4%	NS
			12-15	6	70	83%	63	72%	- 07	-12%	-10%	NS
			15-18	5	62	84%	63	86	01	2%	+2%	NS
			18-21	3	66	153%	67	157%	01	3%	+2%	NS
			21-24	4	66	115%	62	105	- 04	-10%	-6%	NS
	Height	Water - public	6		66		66 5		5		1%	NS
	obtained	vs private	12		72 5		73 5		1 0		1%	p 001
	(cm)		18		78		78 5		5		1%	NS
			24		82 5		83 0		5		1%	NS

[a] W/A = weight/age, H/A = height/age, W/H = weight/height
[b] R = Best-worst values of the indicator reported in the literature (58)
%R = the higher the value, the better the conditions or change in conditions
[c] A positive number indicates the percent improvement in anthropometry (or child growth), and a negative number indicates the percent decrease in anthropometry (or child growth) due to the improved water or sanitation condition
[d] NS = not significant at the 5% level

TABLE 3 The effect of water supply and excreta disposal conditions on mortality

Country[a] Year (Reference)	Type of Indicator	Type of Comparison[b]	Age in Months	Best-Worst = R [(c)]	Before Intervention or Control		After Intervention or Treatment		Change		Statistical Significance Reported (Calculated)
					Level	%R	Level	%R[c]	Level	%R[d]	
Brazil	child mortality	W - unpiped vs piped	children	-	-		-		- 290		p .0001
1983	regression	- water x education									
(58)		interaction							102		p .05
Brazil	infant mortality	E	0-12	-	-		-		+ 52		p .009
1980	correlation with										
(59)	sewage disposal										
Costa Rica	child mortality	E - latrine vs	0-23	-	- 0821		- 2163		-		p < .05
1982	regression coeff	septic tank									
(60)	on prob dying										
Egypt	infant mortality	W - poor vs. wells	0-11	184	175	14%	206	0%	-31	+18%	NS
1952	per 1,000	W+E - none vs wells			175	14%	203	0%	-28	+16%	NS
(61)	neonatal mortality	& latrines	0-1	225	93	64%	94	64%	-1	+1%	NS
	per 1,000				93	64%	85	68%	8	-9%	NS
India	death rates - ?	W - pre-post	all ages								
1962	diarrhea	chlorination			1 03		59			-43%	-
(62)	dysentery				39		30			-23%	-
	cholera				27		.07			-74%	-
	typhoid				22		.08			-63%	-
Kenya	mortality -	W - well	0-23						- 0106		p .1
1980	regression coeff	piped							0065		p > .1
(63)	for survival	E - pit latrine							0153		p < .05
		flush							0228		p > .1

Country[a] Year (Reference)	Type of Indicator	Type of Comparison[b]	Age in Months	Best-Worst = R [c]	Before Intervention or Control		After Intervention or Treatment		Change		Statistical Significance Reported (Calculated)
					Level	%R	Level	%R[c]	Level	%R[d]	
Malaysia	mortality -	W - unpiped vs									
1982	reduction in	piped									
(65)	mortality per	full BF[e]	2-6						4 6		NS
	1,000 live	partial BF							-1 8		NS
	births	no BF							-7 6		NS
		full BF	7-12						+ 6		NS
		partial BF							-4 3		NS
		no BF							-14 8		p < 01
		E - nonflush vs flush									
		full BF	2-6						+4 8		NS
		partial BF							-20 3		p < 005
		no BF							-67 4		p 001
		full BF							-1 3		NS
		partial BF	7-12						-10 8		p 05
		no BF							-18 4		p 05
Sri Lanka	mortality -	W - tap	0-12	184	89	89	41	95	-2	+1%	-
1980	regional per	well			91	58	26	95	65	-35%	-
(66)	1,000 live births	river			91	58	41	86	50	-27%	-
		E - latrine			36	89%	91	59	-55	+30%	-

Country[a] Year (Reference)	Type of Indicator	Type of Comparison[b]	Age in Months	Best-Worst = R [c]	Before Intervention or Control		After Intervention or Treatment		Change		Statistical Significance Reported (Calculated)
					Level	%R	Level	%R[c]	Level	%R[d]	
Sri Lanka	mortality -	E - *nonflush* plus									
1980	per 1,000	no educ. vs									
(67)	live births	flush plus									
		some educ									
		intervention	0-1		52		24		28		(p < .001)
			1-12		24		16		8		(p < .001)
			13-60		42		25		17		(p < .001)
		no facility vs	1-12		31		20		11		(p .001)
		buck or	13-60		31		16		15		(p .001)
		cesspool									
							12		21		-
Sudan	mortality -	W - < 40 lcd vs	0-240		33						
1978	deaths per	> 40 lcd									
(36)	household										

[a] See footnote a, table 1
[b] See footnote b, table 1
[c] See footnote b, table 2
[d] A negative number indicates the percent *reduction* in mortality and a positive number indicates the percent increase in mortality due to the improved water or sanitation conditions
[e] BF = breastfeeding

developing an algorithm, related to the study's design and reported health associations, covering the criteria necessary for establishing a cause-effect relationship between water or sanitation and health. The studies reviewed are evaluated against these criteria to identify those that succeed in showing the most plausible relationships.

REVIEW OF PUBLISHED RESULTS

Association of water and sanitation to diarrheal morbidity

Most studies of improvements in water and sanitation have tried to show effects on diarrheal disease or infection rates (table 1; (8-45)). The results are by no means consistent: Some studies obtained strong positive results; others, the reverse of the expected benefits.

A typical study is one conducted in two Indian villages in Arizona (38), one village with indoor plumbing (treatment village) and one without (control). Hospital admissions of infants with diarrheal disease reportedly declined in the village where indoor plumbing was installed. There are a number of major confounding issues in this study, but one in particular should be noted. There was no random assignment to the treatment group. One village simply cooperated when the Public Health Service wanted to construct indoor plumbing facilities, while the control village resisted all attempts to improve their condition. Therefore, the reported differences cannot be ascribed to plumbing conditions alone. They may also be related to characteristics of self-selection by the participants, especially their interest in or attitudes toward health.

In fact, none of the studies that have evaluated the effect of interventions on diarrheal morbidity, were randomized (18, 20, 26, 27, 30, 31, 32, 34, 35, 38, 39, 41, 45). Yet random assignment is crucial, for it is never possible to identify all the possible differences between comparison groups which may influence the results. Randomly assigning interventions to villages can transform the uncertainty about these differences into a statistical probability statement.

Another design problem is illustrated in a study from Guatemala (21), which reported no differences in diarrhea attack rates among infants from families with and without privies. On the other hand, children 1-5 years of age from families with privies were reported to have lower attack rates than children from families without privies. But the results are inconclusive. A protracted epidemic occurred while the study was conducted, which the authors believe may have been measles. Since measles has a high attack among infants, a measles epidemic, unrelated to water or sanitation, may have masked results in the infant age group.

Of course, it is not always possible to anticipate nonspecific problems, such as a measles epidemic that may affect the results of a study; nor is it always possible to identify these problems if they occur. But replicating the number of villages in a study can diminish the effect of nonspecific influences. Both random assignment and replication, neither of which can be substituted for the other, are important to the validity of health-intervention studies.

Studies are often compromised because health outcomes are difficult to measure. A study in Haiti (23) examining diarrhea prevalence among children from homes using less than one can of water per person per day or more than one can of water per day reported that the better-off group had 11 percent less diarrhea, but the difference was not significant. A major problem with this study was the measurement of diarrheal disease, because the disease recall period was over 240 days long. Underreporting of diarrhea increases as the recall period lengthens (46), so in this study the differences between the two groups were probably underestimated. Bias in the measure of the outcome indicator, diarrhea, may have prevented the investigators from finding true differences.

Underreporting in field surveys is only one way that a diarrheal morbidity indicator can mask benefits from improved water or sanitation supplies. A study from Lesotho (31) illustrates two other ways. This study examined reductions in the ratio of clinic-reported diarrhea to all non-water-related diseases among all age groups and among children under four years of age after improvement in village water supplies. Although children from improved villages were reported to have less diarrhea, this difference was not statistically significant.

The investigators used this ratio to provide a measure of disease which was independent of differential village attendance at clinics and population size. If children from unimproved villages were farther from a clinic, they may have underreported diarrhea more than those from improved villages (47). However, if the non-water-related diseases were not differentially underreported, or to a lesser extent, the ratio used to compare diarrhea rates between the two villages would underestimate true differences.

Underreporting could have also occurred even if the ratio corrected for village attendance and population size. That is, all diseases, including diarrhea, could have been higher among children from unimproved villages. In this case, reductions in diarrhea could have occurred, but use of the ratio precluded finding any differences.

Negative results were obtained in a study from Bangladesh (13) that examined cholera and non-cholera diarrheal rates among families that used coliform-free public tubewells differently. The three tubewells provided water for 88 families averaging 6.7 persons per family and 3.3 below the age of 15, and families were defined as tubewell users if they used tubewell water five times more often than other water sources (canal or tank). Diarrheal incidence data obtained from a hospital were correlated with water use for 11 years.

Tubewell users were not reported to have lower cholera or non-cholera diarrheal rates; instead, they had higher rates than non-tubewell users. The type of water supply made no difference even when the educational level (families with or without high school graduates) was taken into account, but educational level made a difference regardless of the water source.

There are several possible explanations for the negative association between tubewells and health. Each well had two or three mechanical breakdowns per year, and service from the well was interrupted for about 3 weeks during each breakdown. Thus, each tubewell was inoperative 9 weeks/year. In addition, some of the families classified as tubewell users reported using surface water (1800 coliforms/100 ml) for drinking. It seems clear that tubewell users were still consistently exposed to polluted water for domestic hygiene and drinking. It was not that the tubewells failed to reduce diarrhea, but that people were still exposed to fecally-polluted water.

It is worth noting that better education was associated with reduced cholera and non-cholera diarrhea. This suggests that certain extraneous factors could produce positive associations; they could just as easily mask true differences when they exist. Few studies in table 1 controlled for known extraneous factors which could have produced the findings reported.

Another study in which the results are questionable because extraneous confounding factors were not taken into account, compared differences in Shigella prevalence rates among individuals and families in migrant-labor camps in Fresno County, California, which had various types of water and sanitation facilities (43). The camps contained scattered subcamps that were either equipped with inside water faucets or had no indoor plumbing. Shigella prevalence was lower among individuals and among families in the subcamps with indoor faucets. When the authors matched these two types of subcamps for a number of factors, such as people per household, faucet, shower, and toilet, total child population and percent of total camp with inside faucets, the difference in Shigella prevalence between the two was reported to be greater than in the unmatched analysis.

Yet several factors could have confounded these results, particularly family composition. Families were classified as positive if they had one infected child under 11 years of age. The subcamps with no indoor facilities had many more children under 11 years of age than did subcamps with indoor faucets, 61 percent vs. 7 percent respectively. If the children in the subcamp without plumbing were younger than the children in the subcamp with plumbing, they and their families would be more likely to be classified as positive for Shigella since children under five have higher positivity rates than children over five (22). Age could have confounded the results from the family-level analysis even if Shigella positivity rates were constant over the first 10 years of life. Families with many young children would be more likely to be classified as positive since only one positive child classified the entire family.

Positivity rates among families may also have been affected by migration. During the 6-month period of investigation, the number of migrant workers declined from 175,000 to 25,000. If younger, less well-settled parents with more children move more often than older, well-settled parents with fewer young children, this could affect positivity rates. The relationship between Shigella prevalence and the parents' length of settlement was not investigated. It is possible that migration or family age distribution rather than water and sanitation were responsible for the lower Shigella rates found in the subcamps with indoor plumbing.

The problem of children's age and family composition was common among the studies of morbidity. A number of studies failed to control for age at all (8-9, 11-13, 20, 24, 26, 29, 32, 36, 45). Two studies did not observe children under 5 years of age (16, 37). In one study it was unclear what age was observed (27), and eight studies combined data on children below and above 5 years of age (15, 22, 28, 33, 41-44).

Similar problems affected a study in India (27) that examined the effect of bore-hole latrines on diarrhea prevalence among children. It was reported that diarrhea was lowered by the introduction of the latrines, although no statistical tests were done. But seasonal changes in diarrhea rates could have produced this reduction with or without the latrine improvements. Preintervention diarrhea rates were determined only during a time of year when the incidence of diarrhea rises and reaches its peak (48). Postintervention observations included the time of year when diarrheal incidence was falling and at its lowest. Thus, the different seasons alone could have accounted for this reduction.

An example of how some factors can mask true differences is illustrated from a study in Panama (33). The authors examined bacteriological evidence of Shigella, Salmonella, enteropathogenic E. coli, and Edwardsiella tarda among children under 10 years of age and compared the pathogen prevalence rates with types of waste-disposal sites and sources of drinking water. Flush toilets had a higher association with infection than any other type of excreta disposal site (ground, pit privy, and river or sea). Piped water was reported to be worse than streams or wells, but better than rain water.

This unexpected finding was explained by the fact that most of the faucets and flush toilets were located in substandard dwellings or multifamily tenement slums in Panama City. It is likely that the facilities were shared and not adequately maintained. The authors also cited lack of knowledge about personal hygiene as masking the expected benefits of plumbing. Not only population density but also feeding patterns and medical care are likely to differ between rural and urban areas, and these factors should be measured and included in rural/urban analyses.

Conflicting results from a single set of data were obtained from a study conducted in two villages in rural lowland Guatemala (20). This study examined, among other outcomes, the impact of piped water on diarrheal morbidity. In one village, each household initially

received a piped water supply, with health education and latrines in subsequent years. The control village received no improvements. Comparison of diarrhea rates before and after the intervention within the treatment village for all ages combined showed no differences using simple t-statistics. No data were presented for the control village. A reanalysis of these data (49), separately by different age groups using a Chi-square technique on the episodes of diarrhea, found statistically different diarrhea rates for all age groups except infants and those over 45 years of age.

A panel was convened to resolve this discrepancy (50). The panel concluded that no valid relationship between water and diarrhea could be found in this study, and discredited the use of the Chi-square technique. Chi-square testing requires independent observations, but episodes of diarrhea were not independent since some individuals suffered repeated bouts of diarrhea. Thus, the Chi-square statistic was artificially inflated and the probability of finding a significant difference by chance was not 5 percent, as was expected, but was higher. This issue points to the need to use the correct test statistic and to correctly calculate the chance of finding significant differences when none exist.

The panel noted that the study also suffered from some of the methodological problems already discussed, including inadequate standardization of measurement and failure to randomize and replicate the villages studied to control for confounding factors.

These nine studies illustrate the most important design and methodological problems in the studies of diarrheal morbidity listed in table 4: lack of randomization, inadequate replication of the unit of intervention, errors in the measurement of diarrhea, inability to verify usage of improved services, failure to identify known extraneous factors, failure to include those factors in a multivariate analysis, and inappropriate statistical analysis. Failure to account for these problems could have accounted for the differences reported in some studies or masked true differences not found in other studies.

Association of water and sanitation to nutritional status

The use of nutritional status, as measured by anthropometry, has a biological and statistical basis for inclusion in studies evaluating different water and sanitation conditions. There is good evidence that diarrhea contributes to growth faltering (51-53) and that anthropometric indicators may respond as readily to improved water and sanitation as diarrheal indicators (54). Four studies (35, 55-57) examined the association between water and sanitation and nutritional anthropometry (table 2); one of these (35) also examined diarrheal morbidity. Many of the problems already discussed apply to these studies but will be addressed in the context of using nutritional anthropometry as the outcome measure.

A longitudinal study in St. Lucia (35), in which about 75 babies in each of three valleys were followed for up to two years, reported an association between water and sanitation conditions and mean growth increments for height and weight. The "control valley" had one

standpipe for every 350 people. The "water valley" had an individual water supply system installed for each family. The "water-and-excreta valley" had an individual water supply system and water-seal latrines installed for each family. In the water group babies 3-6 months of age put on more weight when compared to control babies. Thereafter, no differences in growth between comparison groups were seen through the first two years of life. In fact, the water group appeared to be better off than the water and excreta group, although not statistically. The control group was reported to be significantly shorter and lighter than the other two groups from 5-16 months of age.

Failure to replicate the valleys receiving each treatment affects this analysis. Since no estimate of the between-valley variance can be calculated when only one replicate per treatment group is measured, the analysis had to use within-child variance. And since the within-child variance is lower than the between-valley variance, the test statistic is inflated. Between-valley variance would have had two components: variance due to children within a valley and variance due to valleys within a treatment. In this study, the probability of finding a significant difference was increased when in fact one did not exist.

Other data in this study render the results inconclusive. If water or sanitation does affect growth by reducing diarrheal morbidity, then this study should have found differences in diarrhea either concurrent with or preceding differences in growth. In two valleys, the control and water groups, growth differences were detected only during 3-6 months of age while no differences in diarrhea prevalence were detected during the first year of life. During the second year of life, when diarrhea rates in the two valleys were different, no growth differences were found. In fact, the mean differences in height and weight became smaller; that is, catch-up growth seemed to occur in the control group which was suffering more diarrhea. The lack of congruity between the growth and diarrhea data tends to cast doubt on the inference that improved water and sanitation improved health.

A study in Colombia (55) also reported that better water and sanitation were related to better growth among children under 5 years of age. Families connected to sewage systems had taller and heavier children than did families with only latrines, and families with a shower or bath had taller and heavier children than did families without these facilities. A number of other factors were also associated with children's height and weight: dwelling size, persons per bed, education of parents, and general cleanliness of the dwelling.

When a multivariate regression analysis was performed that included these socioeconomic factors, the sanitary scale (which was undefined) was still reported to significantly affect children's height, but not weight. But the multivariate analysis did not separate the effect of a better water supply from better excreta-disposal facilities, and the numbers of smaller, lighter children associated with poor water or poor excreta disposal were very similar. Consequently, no statement can be made regarding which component, water or excreta disposal, was more important. Since water and sanitation interventions may have different design requirements, costs, and

effects on health, it is important to identify the component of an intervention that is responsible for the desired outcome. It is also important to distinguish effects due to water quality from those due to water quantity.

A recent analysis from Fiji (56) examined how differences in height and weight among preschool children were affected by excreta-disposal facilities. In rural areas the presence of a flush toilet in the home was reported to significantly improve both height and weight of children. No differences could be detected in urban areas where the children could be considered to come from a well-off group and were not malnourished. But in this case, the absence of a relationship between toilets and health in the urban sample does not mean that the addition of flush toilets would not improve growth of malnourished children. Growth parameters would not be expected to respond to any improvement in sanitation where children are already well-off. It is important to measure differences in populations that can respond to the intervention.

The last study of nutritional outcomes was conducted in 1978 in Nigeria (57) and examined weight-for-age (W/A), height-for-age (H/A), and weight-for-height (W/H) in two groups of preschool children who used water from protected or unprotected wells. Protected water was obtained from dug wells 20-40 feet deep with at least 18 inches of surrounding parapet. This water was reported to be clean, plentiful, and easily accessible. Unprotected water came from wells without parapets, usually shallow, or from ponds or pools dug in river beds, all located at least one-half mile from the households.

The authors reported differences in measures of wasting (W/H and W/A) between the two groups, but no differences in stunting (H/A)--although stunting was prevalent in both groups. The analysis was done using Harvard standards instead of more powerful multivariate techniques to adjust for age and the H/A figures were inaccurately reported.

Thus, the four studies examining the effect of water and sanitation on children's growth suffered from some of the same problems identified in the preceding section on diarrheal morbidity: lack of randomization, inadequate replication, failure to identify, measure and control for extraneous factors, inappropriate statistical analysis, and failure to verify usage of services. Additional issues revealed by these studies are the importance of supporting findings with collateral evidence, of measuring populations that can respond to the intervention, and of calculating and using the correct statistical error term in the analysis.

Association of water and sanitation to mortality

The ten studies that have examined the association of water and sanitation with infant and childhood mortality (58-67) are reported in table 3. Although mortality is a more distal response indicator than morbidity and anthropometry to water and sanitation conditions (figure 1), reductions in mortality should follow improvements in water and sanitation since diarrhea and malnutrition are leading causes of childhood death in the less-developed countries.

Many of the problems that affected the studies of diarrheal morbidity and nutritional anthropometry also apply to the mortality studies: lack of randomization, inadequate replication, measurement errors, inappropriate analytical methods, and classification problems. Selected mortality studies shed further light on these issues.

A study in Kenya (63) included water supply and sanitation as possible determinants of childhood mortality at the district- and household-levels. The district analysis used the 1969 population census data to predict the probability of a child's surviving to age two. The source of data for the household analysis was the 1974 ILO/University of Nairobi Household Survey. A benefit from sanitation was reported at the household level.

In the district-level analysis the percentage of adult literacy (PAL) was used as a proxy for type of drinking water, toilet facilities, personal hygiene, and nutrition. It was reported that PAL had a statistically positive effect on children's life expectancy--except when the number of hospital beds per 1,000 persons was included in the regression analysis, which suggests that medical care may be more important than water supplies or toilet facilities.

An ecological study of this nature uses a geographic unit (e.g., districts) to measure exposure (e.g., toilets) and disease (e.g., mortality). But such an ecological study contains no information on the exposure of all individuals or of individuals who died, and it is not possible to attribute geographic characteristics to individuals. In the context of this study no causal statement may be made about PAL and mortality. Furthermore, since risk factors tend to cluster into geographic units, there is an increased risk of confounding in the results. Another ecological mortality study (66) has this problem.

In the household-level analysis, data on the type of water supply and sanitation facilities were available. Only pit latrines were reported to be associated with reduced mortality: This was significant at $p < .1$ level for the rural sample, but at $p < .05$ level when the rural and urban samples were combined, suggesting that sanitation may be more important in urban areas. The presence of a toilet was reported to have a greater magnitude of benefit than a pit latrine did, but this was not significant at the 5 percent level. Piped or well water did not appear to reduce mortality rates.

In this analysis household characteristics were measured and applied only for 1974, while births could have occurred as early as 1960. A retrospective study of this nature is subject to misclassification bias according to the exposure, presence or absence of toilets. Because an erroneous temporal sequence between toilets and mortality events cannot be ruled out, a spurious association between toilets and mortality could have occurred. In other words, the association between sanitation and mortality may be merely coincidental. Three other studies suffered this problem as well (58, 66, 67).

An Egyptian study (61) examined neonatal and infant mortality rates in a control village that had no improvements of any kind and in four other areas that had one or more of four improvements: wells and

latrines, fly-control measures, preventive medical care, and installation of refuse disposal. The author reported that only fly control seemed to reduce mortality rates.

This study suffered from several of the flaws already discussed: failure to randomize, to replicate villages within a treatment, and to verify exclusive use of improved facilities. Another possible explanation for the negative findings is the overall poor condition of all villages in the study. That is, an objective sanitation score indicated that despite the improvements examined in the study, poor conditions continued after the interventions. Any expected reduction in mortality rates may have been offset by the deplorable conditions in general. This is tantamount to saying that no intervention occurred, since no improvement was able to reduce pathogen ingestion sufficiently to reduce mortality.

A descriptive study in Sri Lanka (66) reported the association between regional variations in water and sanitation conditions and infant mortality rates. The rates were similar in the region in which only 2 percent of the households had tap water and the region in which 30 percent of the households had tap water. The situation was even more pronounced for latrines: the region with the highest percentage of households possessing latrines (42 percent) had the highest infant mortality rate of all regions in Sri Lanka, while the region in which only 3 percent of households had latrines was 60 percent lower. Regions which relied heavily on well water had lower infant mortality rates than regions which did not rely on well water.

Despite a number of flaws in this report, one particular problem should be mentioned. The infant mortality rate had been declining in Sri Lanka for several years before data on water and sanitation were collected. It is likely that regional differences in mortality reflected differential improvements in health during the preceding years rather than water or sanitation conditions at a particular point in time.

A retrospective study from Sri Lanka (67), in which the estate sector was analyzed separately from the rest of the country, analyzed the statistical interaction between types of sanitation, mother's literacy, and father's education at different periods in the infant's life. The presence of toilets was reported to be significantly associated with lower mortality. When the presence of toilets was examined along with data on mother's literacy and husband's education, a synergistic interaction was found: that is, the reduction in infant mortality was even greater than would result from adding the effects of toilets and good education.

A problem in interpreting results of this study, if the impact is real, is the likelihood of mortality underreporting. The study pooled mortality rates from 1950-1975 but used toilet conditions and educational levels only from 1975. If mortality underreporting in the worse-off group was greater than the better-off group, this would suggest that the actual differences in mortality were even larger than

the reported differences. In this case the magnitude of improvement to be expected from the installation of toilets would be underestimated, and so would the cost-effectiveness relative to other interventions.

Another retrospective study, this one from Malaysia (64-65), analyzed the effect of toilets and piped water on infant mortality. The effect of breastfeeding was also considered. The presence of toilets was associated with lower infant mortality, particularly in the postneonatal period. Piped water had a beneficial effect for 2-6-month-olds but not for other infants.

When the statistical interaction of toilets and breastfeeding was analyzed, the presence of a toilet was reported to be even more beneficial for children who were not breastfed. It did not add any benefit for children who were breastfed. Similarly, piped water was associated with lower infant mortality among babies who were partially breastfed, although the effect was smaller than the one found for toilets. It should be noted that mothers who breastfed tended to have poor sanitation facilities, and those who did not breastfeed generally had good facilities.

This analysis controlled for a number of extraneous factors: income, birthweight, parity, mother's age, preceding inter-pregnancy interval, child's age, and ethnicity. However, the analysis was done at the child level, not the community level, and regional controls were not included in the analysis. Because this design did not take into account the large clusters from which the data were collected, the chance of finding a statistically significant difference when one did not exist was actually greater than 0.05 as reported. Because intracluster correlations were not taken into account, the variance used when testing differences was artificially lowered.

In conclusion, methodological and design problems could have produced spurious associations in the studies reporting on water, sanitation, and mortality, as well as in the other studies of morbidity and nutritional status. These issues will now be systematically structured to judge the plausibility of all studies in tables 1-3, according to their design and the statistical associations they report.

DISCUSSION OF CRITERIA TO EVALUATE STUDIES

Overall, the studies examining associations between improved water or sanitary facilities and diarrheal disease have reported conflicting and inconsistent results. In the studies using nutritional anthropometry as an indicator, water appeared to have a greater effect on children's height and weight than sanitation conditions. The studies on childhood mortality, on the other hand, suggest that waste-disposal facilities may be more important than water facilities. But regardless of the type of association reported--whether a positive one, negative one, or none at all--each study must satisfy certain criteria for internal validity before any implications can be drawn.

Systematic criteria applied to research findings expose biases when they occur and help distinguish between causal and indirect associations, thus permitting an assessment of the plausibility of the findings. A previous algorithm (68) has been extended and used to show whether or not the findings associated with different water and sanitation conditions satisfy the criteria for validity.

For this evaluation, the studies are grouped according to whether a statistical association was reported and what study design was used (table 4). Studies reporting positive associations between water or sanitation conditions and health (figure 2) included experimental trials that evaluated interventions and observational surveys that examined nonrandomized differences or preexisting differences in water or sanitation conditions.

Experimental studies

Experimental trials provide more plausible results than observational studies, because experimental trials make probability statements of causality of association. All studies which reported positive associations and evaluated interventions were considered experimental trials and are included in table 5 along with the criteria necessary for judging experimental trials. The most important criterion is that the intervention is assigned in a strictly random fashion. This decreases the possibility that associations are due to some confounding variable inherent in the treatment or control group when the intervention is applied. Furthermore, a probability statement can be assigned to the possibility that the association described is not due to confounding, even when some possible confounding factors are not measured or even recognized (see table 5, part B1a). A randomized trial can use cross-sectional data or include before-after measurements; in the latter case it is called a before-after randomized trial.

Although randomized trials can assign a probability to the likelihood that an intervention caused a result, randomization alone cannot identify which component of an intervention was responsible. In order to identify the responsible component additional techniques must be used when designing experimental trials. First, randomization should be done across enough independent units so that statistical probability tests can be applied. In trials involving water and sanitation measures, the unit of intervention is not the child even though the objective of the intervention is to improve children's health. The appropriate unit of intervention that must be sufficiently replicated is the village, since children within a village are subject to any influences which affect the entire village. For instance, if a measles epidemic affected an entire village, it would mask or exaggerate the true effect of any intervention if only that village was included in the study (table 5, part B1b).

Second, if statistical testing is done using the incorrect unit of replication, then probability statements will invariably be wrong. Child-level variances are smaller than village-level variances and if child-level variances are used in the analyses, then statistical probability statements of association due to chance alone will be

TABLE 4. A list of questions asked about study findings depending upon the type of study design and whether or not differences in outcomes were tested and found

WAS THERE A STATISTICAL ASSOCIATION REPORTED?

YES		NO
EXPERIMENTAL STUDIES	OBSERVATIONAL STUDIES	EXPERIMENTAL AND OBSERVATIONAL STUDIES
A. Control for confounding 1. Was randomization properly executed? 2. Was analysis done with: a. proper unit of observation? b. proper variance term? 3. Was study blinded by a. assignment? b. assessment? 4. Was randomization confirmed? B. Analysis for congruity 1. Was intervention confirmed? 2. Were other concomitant outcomes measured and analyzed?	A. Control for confoundin 1. Were major confounding variables measured? 2. Was there matching by design? 3. Did statistical analysis deal with confounding? 4. Were measurements blinded? B. Analysis for congruity 1. Were measures of different water/sanitation conditions confirmed? 2. Were other concomitant outcomes measured and analyzed?	A. Control for confounding 1. Was intervetion (experimental studies) necessary? 2. Did intervention (experimental studies) occur? 3. Were differences large enough to produce a change in health (observational studies)? 4. Was randomization (experimental studies) carried out? 5. Were measurement biases controlled? 6. Was sample size adequate? 7. Did statistical analysis deal with confounding?

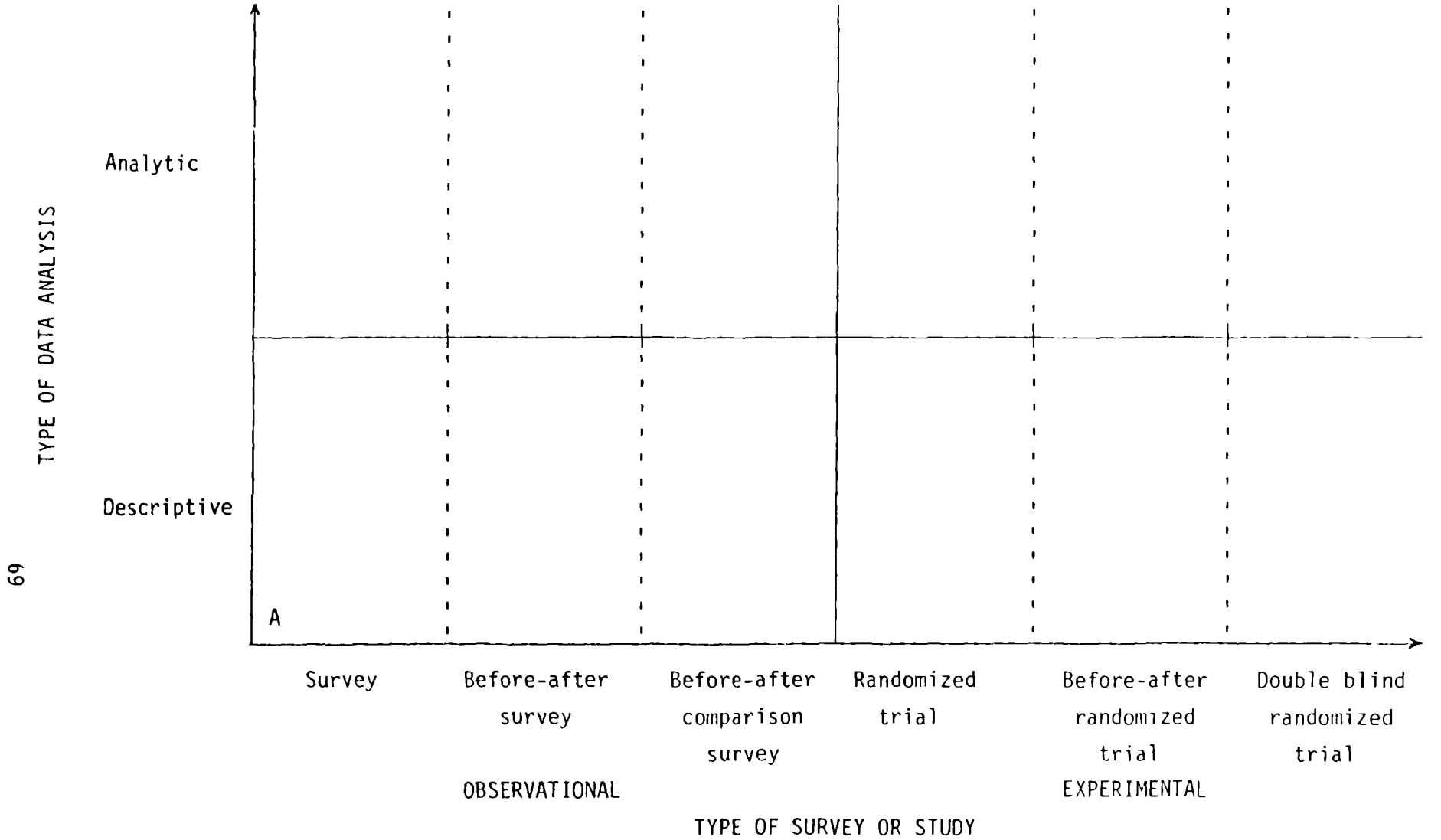

FIGURE 2. A schematic diagram of different surveys and trials. (Plausibility of associations reported increase the further a study is from point A.)

TABLE 5. Summary of ability to satisfy design characteristics for experimental trials showing benefits from water or sanitation improvements

Reference	38	41	35	26	34	32	30
Country[a]	USA	USA	STL	IND	PHI	LIB	KEN
Year of publication	1969	1957	1981	1971	1974	1981	unpub.
A. Description of study							
1. Type of intervention[b]	W	E	W	W	E, W W+E	W	W
2. Unit of randomization[c]	vi	vi	va	vi	vi	?	vi
B. Control of confounding							
1. Randomization							
a. Was intervention randomly assigned?	-*[d]	-*	-*	-*	-*	-*	-*
b. Were units of intervention sufficiently replicated?	-	-	-	-	-	?	-
2. Was analysis done with the variance of the unit of replication?	-	-	-	-	-	-	-
3. Was intervention assumed the intervention done?	-	-	+	+	+	+	+
4. Was study double blind?	-	-	-	-	-	-	-
a. If no, were measurement biases estimated or controlled?	-	+	-	-	+	-	-
b. If no, was participant self-selection after randomization controlled?	-	+	+	-	+	-	-

5. Was randomization confirmed through comparison of the distribution of confounding variables?

a. Exposure to infection

1) Education

2) Income

3) Rural/urban

4) Seasonality

5) Crowding

6) Secular trend

b. Susceptibility to infection

1) Age

2) Sex

3) Diet

4) Breastfeeding

c. Prevention of outcome

1) Medical care

C. Analysis for congruity

1. Measures of intervention confirmed

a. Water quality

b. Water quantity

c. Utilization of services

2. Measures for consistency

a. Morbidity

b. Nutritional status

c. Mortality

[a] See footnote a, table 1
[b] W = water supply, E = excreta disposal, W+E = water supply + excreta disposal.
[c] vi = village; va = valley
[d] - means no, + means yes, ? means unclear.
* Since random assignment did not occur, these studies will be considered as observation studies, table b.

higher than those reported. Since the statement regarding the probability of causality of association is high, the plausibility of the findings is diminished (see table 5, part B2).

A third important issue concerns the application of the intervention to the recipients. Water or sanitation interventions may be accompanied by additional inputs such as health education. In this case the assumed intervention (water or sanitation) was not the true intervention (water or sanitation and health education). Thus, the true benefit of water or sanitation cannot be estimated (see table 5, part B3).

Fourth, the study should be conducted in a manner that prevents systematic bias from occurring when measuring the outcome. For example, measurement of diarrhea might be biased if the enumerators know which villages received the intervention and which did not. This can be avoided by blinding the assessment of the outcome, in which the enumerators do not know which villages have received the intervention or ignore this knowledge (see table 5, part B4a).

A randomized trial should also control for behavioral differences associated with the intervention. If those villages receiving a water supply act differently because they know they received the intervention and are being observed, this could bias the study's results. An improvement in health might be caused by a change in behavior that would not be duplicated unless the intervention were applied in an experimental context. This can be controlled by blinding by assignment--that is, by applying a placebo intervention to the non-intervention groups, so that all groups think they are receiving an intervention but are blinded as to which one (see table 5, part B4a).

Even if an intervention is randomly assigned, there may be elements of self-selection in the results. If enumerators measure only participants who choose to be measured, they would obtain a biased estimate of the health outcome. This would diminish the plausibility of the findings (see table 5, part B4b).

Studies in which the unit of intervention is randomly assigned, sufficiently replicated, and blinded by assessment and by assignment are known as double-blind randomized trials. This is the only type of study that permits assigning a probability statement that the association found is a causal one.

Finally, experimental trials should measure known and potential confounding factors, to verify that randomization was successful in distributing confounding factors equally across comparison groups (see table 5, part B5). Experimental trials can increase the probability that an association is causal if confounding variables are identified and taken into account either in the design of the study (e.g., matching) or in the analysis of the results. Even if some potential confounding factors are missed, there is a greater possibility that associations found are not due to confounding.

Confounding factors that are likely to affect studies of water and sanitation conditions can be identified on the basis of the

biological pathways that lead from the intervention to the outcome (see figure 1). The most relevant confounding factors that might affect the outcomes can then be measured (figure 3). If all steps change concurrently as predicted by biological theory, the plausibility that the changes are due to the intervention is increased. Analytic studies that control confounding are more valuable than descriptive studies that only report associations.

These criteria can now be applied to the studies of water and sanitation interventions that reported a statistically significant association. Table 5 indicates that even though several studies evaluated interventions and were intended as experimental trials, none were conducted as randomized trials and, therefore, none can be considered truly experimental. Since these studies did not use true random assignment, they must be evaluated according to the criteria for observational surveys.

Observational surveys: Positive statistical association

Observational surveys of various types can only make probability statements of association (figure 2). Cross-sectional observational studies use survey data to examine the concurrent presence or absence of good water or sanitation facilities in relation to an outcome, such as diarrhea. Before-after surveys would look for a reduction in diarrhea following an intervention. Both types of surveys could find associations that are due to factors other than improved water and sanitation. For instance, a before-after survey could detect a decline in diarrheal rates due to a general regional improvement in the standard of living which would include better health and hygiene. In order to avoid this type of secular confounding, data from areas without the improvements can be included in before/after comparisons with data from the improved areas. This design, which is differentiated from the before-after survey by the presence of the comparison group, increases the plausibility of the findings. However, it is impossible to identify and measure every possible confounding factor that could affect the intervention and comparison groups, so there is always some doubt about the findings of observational studies. The factors most likely to confound results in observational studies of water or sanitation are listed in table 6, part B1.

Two techniques can be used to control for confounding factors. Comparison groups can be matched on certain identified factors, or these factors can be controlled by multivariate statistical techniques. Analytical studies, in which confounding is controlled by the above techniques, yield more plausible results than descriptive reports (see table 6, parts B2 and B3).

As in experimental studies, measurement errors must be minimized. Although blinding by assignment is not possible, blinding by assessment remains an issue and is the only way to ensure that potential biases are estimated and controlled. Measurements can be standardized by stratifying or randomizing the enumerators across comparison groups (see table 6, part B4).

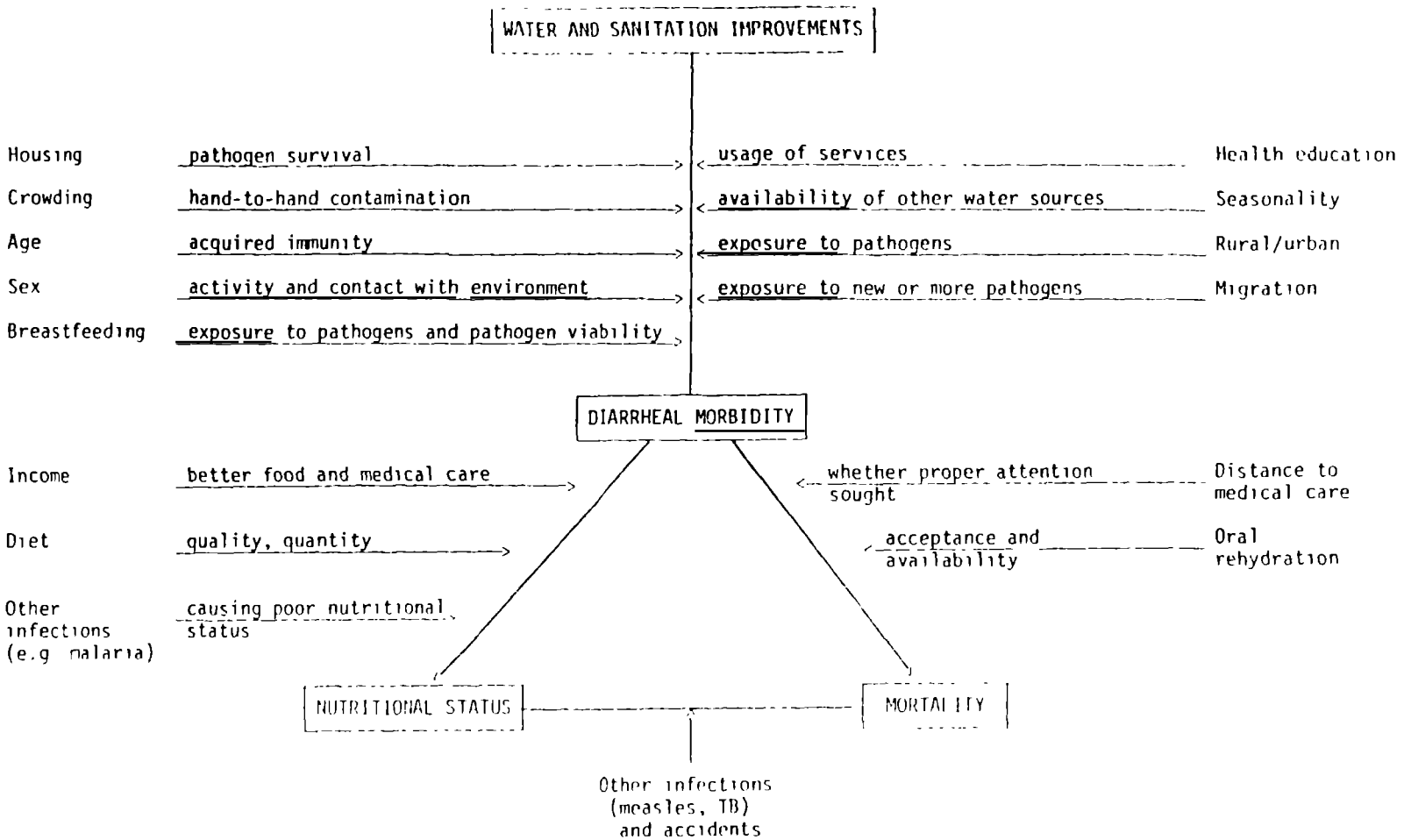

FIGURE 3 Potential confounding factors and their relationships to the health benefits expected from water and sanitation improvements

Observational studies of water and sanitation must be particularly aware that improved services may not always be used. There are many possible reasons, including breakdowns, lack of knowledge on how to use the services, easy access to other sources of water (e.g., streams or ponds), and sanitation problems (e.g., odors and darkness in improved pit latrines). If positive associations were reported but the usage of services was not measured, then the plausibility of the association is diminished (table 6, part C1). On the other hand, plausibility of an association is increased if several health outcomes are measured (see figure 1) and change concurrently as predicted by biological theory (see table 6, part C2).

Table 6 evaluates the observational studies that reported a statistical association between water or sanitation and health. In addition, a few studies reported a change in outcome but did not calculate statistical associations; if the change was greater than 50 percent, the study is included in table 6. Some studies reported both positive and negative findings, either for different indicators or for different ages. These studies are also included in table 6.

According to table 6, there are a few studies that do satisfy some of the criteria necessary to increase the plausibility of the associations they report. Most of these are mortality studies (60, 63, 64, 65, 67), and their results suggest that improving facilities for excreta disposal is more important than improving the water supply. This is in accordance with what is known about fecal-oral transmission of pathogens (figure 4). If excreta is effectively removed from the environment, fewer pathogens are available for ingestion. Once pathogens have entered the environment, they may take many pathways to ingestion and are thus more difficult to control.

The evidence in favor of excreta-disposal facilities is advanced when results from two studies examining nutritional status are considered (55, 56). In these studies children's body size was associated with excreta-disposal facilities even when other factors were controlled in the analyses. These two outcomes--anthropometry and mortality--offer collateral evidence that excreta-disposal facilities can affect health.

One study in diarrheal morbidity (16) reported that hygienic bathrooms were associated with less diarrhea among grade schoolers. In another morbidity study (42) diarrhea was reduced when water was more readily accessible, whether well or piped city water. A third study (19) also found sanitation to be a good predictor of diarrhea. In addition, the amount of water used per person was more important than the source of water used.

Observational surveys: Negative statistical association

The next question to be considered is whether the studies reporting a negative association or lack of association between sanitation and health counterbalance these inferences. Because no negative study was randomized, all are subject to the criteria for observational surveys. Some issues described above for positive studies are pertinent: documentation of usage of services, control of measurement

TABLE 6 Summary of ability to satisfy design characteristics for observational surveys showing benefits from water or sanitation improvements

Reference	42	16	55	56	60	63	64-5	67	39	41	40	43	19	25	26	62	57	37	35	8	9	22	12	29	34	58	44	28	38	78	30	36	10	21
Country[a]	USA	COL	COL	FIJ	CR	KEN	MAL	SRL	USA	USA	USA	USA	ETH	IND	IND	IND	NIG	UK	STL	BGD	BGD	GUA	BGD	KEN	PHI	BRA	USA	IRN	USA	VEN	KLN	SUD	BGD	GUY
Year of publication	1955	1978	1975	1983	1982	1980	1982	1980	1960	1957	1958	1955	1977	1977	1971	1962	1978	1978	1981	1982	1980	1967	1957	1972	1974	1983	1953	1966	1969	1966	unp	1978	1977	1983
A *Description of study*																																		
1 Type of comparison[b]	W/E	E	W+E	W+E	E	W+E	W+E	E	W	E	W+E	U	W+E	W	W	W	W	U	W+E	W	W	W	W	W	W+E	W	W	W	W	W	U	W	U	W+E
2 Unit of comparison[c]	b	s	n	v	f/k	h/d	h	f	p	t	v	p	v	v	v	v	h	s	v	n	n	v	p	v	v	c	p	v	v	v	d	f	v	v
B *Control of confounding*[d]																																		
1 Were major confounding variables measured?																																		
a *Exposure to infection*																																		
1) Education	-	+	+	+	+	+	+	+	?	-	+	-	+	?	-	-	-	+	+	-	+	-	-	+	-	+	-	?	+	?	+	+	-	+
2) Income	-	-	+	+	+	+	+	+	?	-	+	-	+	?	-	-	-	-	+	-	+	-	-	+	-	+	-	?	-	?	+	+	-	+
3) Rural/urban	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	?	+	?	+	+	+	+
4) Seasonality	?	-	+	-	+	-	-	-	+	+	+	+	+	+	-	-	+	+	+	+	?	+	+	+	+	-	+	?	+	?	-	-	+	+
5) Crowding	-	+	+	+	-	-	+	-	?	-	+	+	+	+	-	-	-	-	+	+	+	-	+	+	+	-	+	?	-	?	+	-	-	+
6) Secular trends	-	-	-	-	+	+	+	+	-	+	+	-	-	-	-	+	-	+	+	-	?	-	-	-	-	+	-	?	-	?	-	-	-	+
b *Susceptibility to infection*																																		
1) Age	-	-	+	-	+	+	+	+	+	-	+	+	+	+	-	-	+	+	+	-	-	+	-	-	+	-	+	+	+	+	+	-	+	-
2) Sex	-	-	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	+	-	-	-	-	?
3) Diet	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	+	-	-	-	-	?	-	?	+	-	-	-
4) Breastfeeding	-	NA	-		-		-	-	-	-	-	-	-	-	-	-	-	NA	+	-	-	-	-	-	-	-	-	?	-	?	-	-	-	?
c *Prevention of outcome*																																		
5) Medical care	-	-		-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	?	+	?	+	-	+	?
6) Other	-	-	+	+	+	+	+	+	?	?	+	+	+	-	-	-	-	-	+	+	+	+	-	+	+	-	+	?	+	?	+	-	+	+
2 Was there matching in design?	+	-	+	-			-	-	-		+	+		-	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3 Did statistical analysis deal with confounding?	-	-	+	+	+	+	+	+	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-		-	-	-	-
4 Were measurements blind?	?	-	+	+	-	-	-	-	-		-	+	-	-	-	-	-	-	-	-	+	+	-	-	+	-	-	?	-	?	-	-	-	-
If no, were potential biases estimated or controlled?	?	-	NA	NA	+	+	+	-	?	-	?	NA	+	-	-	-	?	?	?	-	NA	NA	-	-	NA	-	-	?	-	?	-	-	-	-

Reference	42	16	55	56	60	63	64-5	67	39	41	40	43	19	25	26	62	57	37	35	8	9	22	12	29	34	58	44	28	38	28	30	31	10	21
Country[a]	USA	COL	COL	FIJ	CR	KEN	MAL	SRL	USA	USA	USA	USA	ETH	IND	IND	IND	NIG	UK	STL	BGD	BGD	GUA	BGD	KEN	PHI	BRA	USA	IRN	USA	ZEM	KEN	SUD	BGD	GUA
Year of publication	1955	1978	1975	1983	1982	1980	1982	1980	1960	1957	1958	1955	1977	1977	1971	1962	1973	1978	1981	1982	1980	1967	1957	1972	1974	1983	1953	1966	1963	1966	unp	1978	1977	1980
C Analysis for congruity																																		
1 Measure of comparison confirmed																																		
a Water quality	-	-	-	-	NA	-	-	NA	+	NA	+	+	-	+	+	+	+	-	+	+	+	-	-	+	+	-	-	+	+	?	+	-	+	+
b Water quantity	-	-	-	-	NA	-	-	NA	-	NA	-	-	+	-	-	-	-	+	+	-	-	-	-	+	-	-	-	-	-	?	+	+	+	+
c Utilization of services	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	-	-	+	+	-	-	-	-	?	+	-	+	+
2 Measure for consistency																																		
a Morbidity	+	+	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
b Nutritional status	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
c Mortality	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	+	-	-

[a] See footnote a, table 1
[b] See footnote b, table 5
[c] v = village, t = town, f = family, d = district, z = zone, h = household, c = city, p = camp, n = neighborhood b = block, s = school, k = canton
[d] - means no, + means yes, ? mean unclear, NA means not applicable

biases, and control of identified confounding factors by either matching or statistical analysis. But studies that fail to show an association must also be evaluated in light of three additional criteria.

First, was the outcome likely to be changed by improvements in water and sanitation? That is, were conditions such that an outcome could respond to the intervention? For example, in an area where most children are breastfed, the addition of good water is not likely to reduce infant mortality rates, because breastfeeding reduces exposure to pathogens and confers immunity to ingested pathogens. This criterion is checked in table 7, part B1.

The second criterion deals with sample size. Were there enough replications of the unit of analysis so that the power of the test was sufficiently large to find differences in the groups compared? This criterion is checked in table 7, part C.

The third criterion is control for possible negative confounding. A negative confounder is a factor related to both the intervention and the outcome which would wipe out any true benefit from the intervention. For example, improved water supplies may draw many people to a particular area, creating a crowding problem. Overcrowding, which enhances the spread of pathogens directly from one person to another, will mask the true effects of improved water. Negative confounding may prevent an outcome from occurring, whereas positive confounding may allow an event to occur independently of the intervention. This criterion is checked in table 7, parts D and E.

Examination of the studies reporting negative associations or reporting a change in outcome of less than 50 percent (table 7) reveals that all suffer from one or more of these flaws. Only two studies attempted to control for confounding factors in the analysis (13, 14); both were in Bangladesh and examined the health impact of the tubewells. In these studies tubewell users used other polluted sources of water because of tubewell breakdown. The confirmed high exposure to pathogens from other sources would suggest that the actual level of tubewell use was insufficient to improve health.

In many of the other studies reporting negative associations the investigators failed to measure confounding, so it is difficult to know if negative confounding affected study results.

Only three studies included more than two villages in their comparison (11, 31, 33). In two of these (31, 33), the analysis was done at the individual level; both were confounded by other factors. One study (31) used an inappropriate measure of diarrhea and the other study (33) was confounded by urban/rural differences. The study which reported data at the village level failed to control for age differences.

In summary, most studies reporting negative associations are flawed by either not measuring age (11, 13, 24, 32, 45) or reporting on too large an age range (28, 33), or by an inability to specify which age was included in the analysis (27). The remaining studies

are flawed by failure of the intervention to occur (18, 61), regional differences in the comparison groups (17, 59, 66), or failure to control for identified confounding factors (15, 20, 23). As a result, there are no reported negative associations suggesting that water or sanitation will not improve health. These negative results have much less plausibility than reports of positive associations.

CONCLUSION

When all of the published studies of water or sanitation improvements are subjected to scrutiny, it becomes apparent that the true effect of such interventions remains difficult to estimate. There are a few studies (16, 19, 40, 42, 55, 56, 60, 63, 64, 65, 67) that: 1) support the hypothesis that excreta disposal is more important than water interventions, 2) support the hypothesized pathway between water and sanitation improvements and better health, and 3) do not have severe flaws (16, 42, 55, 56, 60, 63, 64, 65, 67).

Examination of these studies provides useful information for targeting areas (e.g. those with negative risk factors) most in need of these environmental interventions. The type of intervention (e.g. water quality, water availability and usage, or excreta disposal) which would result in the largest health impact of children can also be determined.

A model has been proposed which examines the relationship between the level of pathogen exposure and the percent reduction in diarrhea following an intervention (69). This model suggests that maximum health impacts can be achieved when exposure levels are highest. It did not address the issue of which type of intervention should be considered, particularly in the presence or absence of risk factors.

When the level of exposure to fecal contamination from the environment is high, the health impact due to sanitation would be expected to be greater than the impact due to water availability, which would in turn be greater than water quality (figure 4). On the other hand, when exposure is low this ranking may change.

All the major pathogens of diarrhea are shed via feces, therefore effective sanitation measures should remove these pathogens from the environment and result in large health impacts. Excreta disposal was a more important determinant of child health than were water related variables when the better studies were examined. This effect was seen for all types of outcomes, morbidity, child growth, and mortality. In most studies which reported an effect due to a water related variable, the effect due to sanitation was usually not controlled for in the analysis. When the effect of both variables were controlled for in the analysis, the effect of the water variable was greatly diminished and in most cases lost statistical significance (56, 63, 64).

The health impact due to the quality of drinking water may be larger when the level of fecal contamination from the environment is low compared to high levels of exposure. When exposure is high, water

TABLE 7. Summary of ability to satisfy design characteristics for studies reporting negative findings

Reference Country[a] Year of Publication	18 EGY 1954	61 EGY 1952	13 BGD 1976	33 PAN 1971	66 SRL 1980	24 IND 1978	45 ZAM 1976	31 LES 1978	20 GUA 1970	32 LIB 1981	17 CR 1965	23 HAI 1980	28 SUD 1966	15 COL 1978	27 IND 1970	59 BRA 1980	11 BGD 1977	14 BGD 1972	28 SRL 1966
A. Description of study																			
1 Type of Comparison[b]	W+E[c]	W+E	W	W+E	W+E	W	W	W	W	W	W+E	W	W	W	E	E	W	W	W
2 Unit of Comparison[c]	v	v	v	v	d	t	c	v	v	?	d	t	v	t	v	d	v	z	v
B Contribution of treatment or intervention																			
1. Was intervention necessary?	+[d]	+	NA	NA	NA	+	+	+	+	-	NA	NA	NA	NA	+	NA	+	NA	NA
2 Did intervention (treatment) occur by documentation?																			
a. Water quality	-	-	-	+	-	-	-	+	+	-	+	-	-	-	NA	NA	-	-	-
b. Water quantity	-	-	-	+	-	-	-	-	+	-	-	+	-	-	NA	NA	-	-	-
c Utilization of services	+	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
C Control of sample size																			
1 Was expected difference specified ?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2 Was variability within comparison groups described?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
3. Was sample size adequate?	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-
D. Control of negative confounding by design																			
1. Randomization assignment performed?	-	-	NA	NA	NA	-	-	-	-	-	NA	NA	NA	NA	-	NA	-	NA	NA
2 Were measurement biases estimated?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Reference	18	61	13	33	66	24	45	31	20	32	17	23	28	15	27	59	11	14	28
Country[a]	EGY	EGY	BGD	PAN	SRL	IND	ZAM	LES	GUA	LIB	CR	HAI	SUD	COL	IND	BRA	BGD	BGD	SRL
Year of Publication	1954	1952	1976	1971	1980	1978	1976	1978	1978	1981	1965	1980	1966	1978	1970	1980	1977	1972	1966
E. Control of negative confounding by statistical analysis																			
1 Were confounding variables measured?																			
a Education	-	-	+	-	-	-	-	-	+	-	+	+	?	+	-	+	-	-	?
b Income	-	-	+	-	-	-	-	-	+	-	+	+	?	+	-	+	-	-	?
c Rural/urban	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
d Seasonality	-	-	+	+	-	-	-	-	+	+	+	+	?	+	-	?	+	-	?
e Crowding	-	-	+	+	-	-	-	-	+	-	+	+	?	-	-	-	+	-	?
f. Secular trend	-	-	+	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
g. Age	+	+	-	+	+	-	-	+	+	-	+	+	+	+	-	+	-	+	+
h. Sex	+	+	+	+	-	-	-	-	-	-	+	+	-	+	-	?	-	+	-
i. Diet	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
j Breastfeeding	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
k. Medical care	-	-	+	-	-	-	-	-	-	-	-	-	?	+	-	-	+	-	?
l. Other	-	-	+	-	+	-	-	-	+	-	+	+	?	+	-	+	+	-	?
2 Matching for factors in E1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. Did statistical analysis deal with confounding factors other than age?	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-

[a] See footnote a, table 1.
[b] W = water supply, E = excreta disposal
[c] v = village, d = district, t = town, c = city, z = zones
[d] + means yes, - means no, ? means unclear

quality improvements may have little or no impact (figure 4). Since diarrhea is multifactorial in origin, drinking water constitutes only one source of pathogen transmission. Eliminating drinking water as a source of pathogen ingestion may, therefore, have no appreciable health impact. Since most of the studies examined improvements at relatively high levels of exposure, this partly explains the inability to find health improvements due to water quality in general. Of course, uncontaminated source water may become contaminated before ingestion thereby eliminating any possibility of an improvement (20, 25, 31).

On the other hand, when exposure is low, the quality of drinking water may play a larger role in transmitting diarrheal disease agents. There is indirect evidence to support this contention. The quality of drinking water in developed countries, where exposure is low, plays an important role in outbreaks of diarrheal disease. Many recent outbreaks of diarrhea have been traced to the quality of the drinking water.

In the case of water quantity improvements impacts would likely be larger when exposure is high. The use of water for domestic hygiene can help eliminate many sources of transmission, such as cooking utensils, hands, and objects. When exposure is low, the impact may depend more on the type of pathogens prevalent in the community (69).

Certain risk factors can influence the child's level of exposure and thereby determine the impact that an intervention can have on a child's health. For example, breastfeeding, literacy of mothers, and income interact with water and sanitation conditions. Poor, illiterate, non-breastfed populations can be assumed to be exposed to higher levels of contamination compared to their better-off counterparts.

In Malaysia (65) where infants were breastfed, the addition of a toilet or piped water supply did not significantly lower infant mortality rates. In areas where breastfeeding was not universal or of short duration, then the addition of water and sanitation facilities produced a large health impact. Breastfeeding can reduce exposure and confer immunity against infection.

Infant mortality rates are lower when mothers are literate. Therefore, if an improved water supply or sanitation facility were provided, one would expect the infant mortality rate to change more among the illiterate than among the literate. In a study from Sri Lanka (67), however, the reverse occurred. Infant mortality rates fell more when mothers were literate compared to the illiterate. This was particularly true among the neonatal population. Although this finding does not fit the exposure hypothesis, it does not nullify the hypothesis. If illiterate mothers do not use the new facilities or do not adequately maintain them, then the excreta disposal facilities will have a marginal impact at best. On the other hand, literate mothers may know how to use the new facilities and have a means of disposing of feces as well as refuse in general. This possibility needs to be researched in more detail.

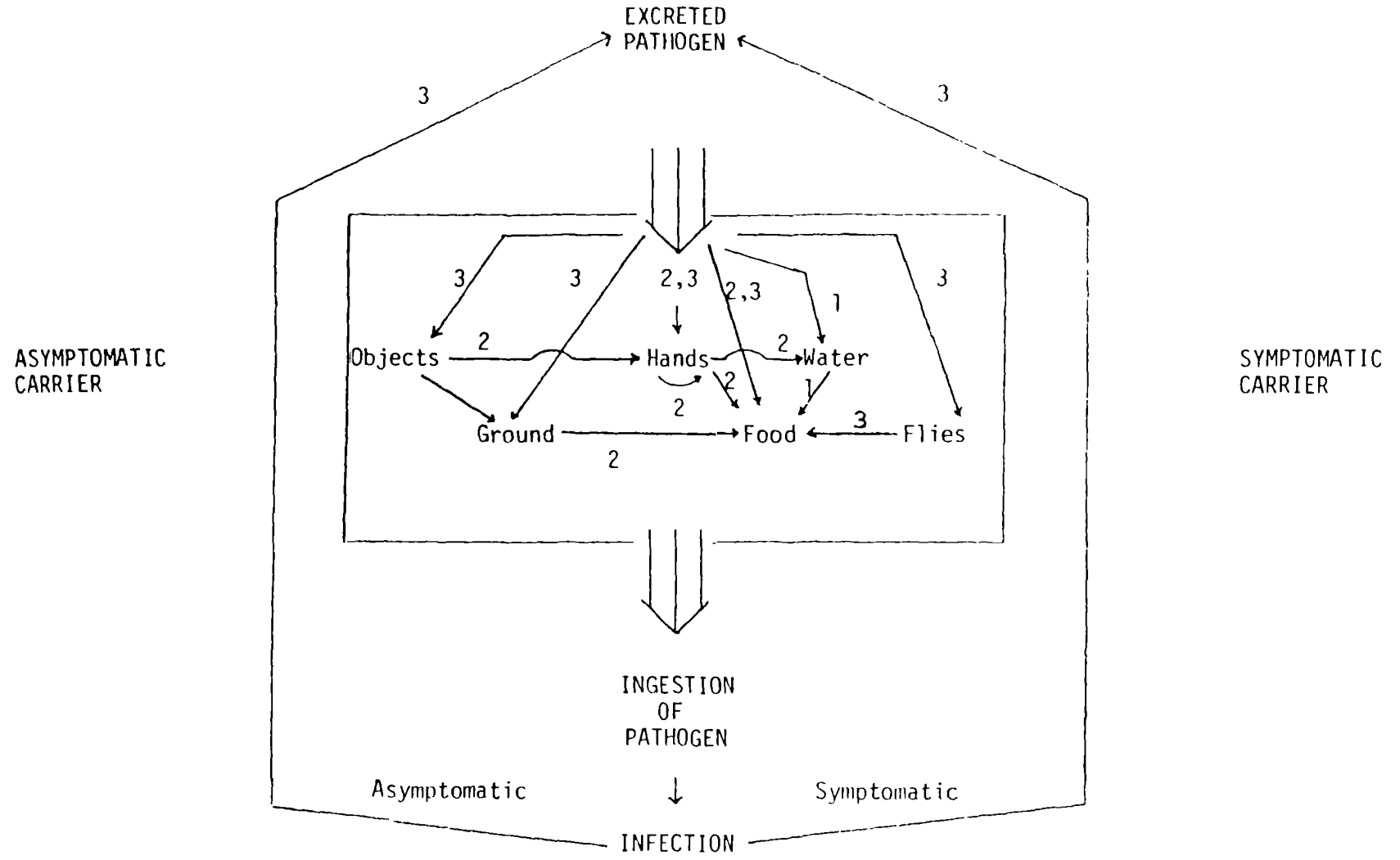

FIGURE 4. Schematic presentation of the fecal-oral transmission of pathogens in relation to water and sanitation interventions. (The points that water and sanitation interventions may break this transmission are numbered (1 = water quality; 2 = water availability; 3 = excreta disposal.)

In Fiji (56) toilet and income interacted among the Fijian population. Flush toilets were associated with improved child nutritional status among the low income group. This was also true among the high income group but to a much lesser extent. Thus, when exposure was high, toilets produced a measurable health impact.

In Ethiopia (19) sanitation was a more powerful predictor of morbidity than were cultural demographic characteristics of the household. During the first two years of life children had less diarrhea when they came from families with higher water usage per person compared to children from families with lower water usage per person. This effect was greatest when no latrine was present in the household. Thus when exposure is high increased use of water can reduce diarrhea rates among children. In this study the amount of water used per person was more important than the source of water used.

Despite the inferences drawn from these studies, it is certainly difficult to calculate accurate estimates of the magnitude of improved health, but this should be less difficult if the methodological issues described above are considered in the design and analysis of water and sanitation evaluations. Although it is unlikely that double-blind randomized trials can be designed for sanitation or water availability interventions because recipients and enumerators would know which villages received the intervention, double-blind randomized trials are possible for water quality interventions. For instance, improving piped water in communities selected at random among many communities with piped water of bad quality should be possible. Thus, recipients and enumerators could be blinded as to who receives the good or poor quality water. If the ideal study cannot be implemented, those parts which are possible should be implemented. For instance, randomization, albeit not double-blinded, could be a part of sanitation and water supply interventions. Most projects are implemented over a period of years. Therefore, randomization could be incorporated into the staging of the project.

The biological theory and epidemiological techniques are advanced enough to provide better estimates of the cost-effectiveness of water or sanitation programs. Failure to provide more accurate estimates could very likely lead to the reallocation of available funds away from water and sanitation projects and toward other projects. In short, many of the water and sanitation programs being planned and implemented for the International Drinking Water Supply and Sanitation Decade should be evaluated so that their cost-effectiveness relative to other interventions can be demonstrated.

REFERENCES

1. Rhode JE, Northrup RS. Taking science where the diarrhoea is. In: Elliott K, ed. Acute diarrhoea in childhood. Amsterdam: Elsevier/Excerpta Medica, 1976.

2. World Bank. Water Supply and Waste Disposal. Washington, D.C.: Transportation, Water, and Telecommunications Department, 1980.

3. White GF, Bradley DJ, White AU. Drawers of Water: Domestic Water Use in East Africa. Chicago: University of Chicago Press, 1972.

4. Saunders RJ, Warford JJ. Village Water Supply: Economics and Policy in the Developing World. Baltimore: Johns Hopkins University Press, 1976.

5. United Nations. Report of the United Nations Water Conference, Mar del Plata, 14-25 March. New York: (document E/CONF. 70/29), 1977.

6. World Health Organization. World Health Statistics Report: Water and Sanitation 1976; 29:544-603.

7. Kalbermatten JM, Julius DS, Gunnerson, CG. Appropriate Technology for Water Supply and Sanitation: Technical and Economic Options. Washington, D.C.: World Bank, 1980.

8. Hughes JM, Boyce JM, Levine RJ, et al. Epidemiology of eltor cholera in rural Bangladesh: Importance of surface water in transmission. Bull WHO 1982; 60:395-404.

9. Spira WM, Khan MU, Saeed YA, et al. Microbiological surveillance of intraneighborhood El Tor Cholera transmission in rural Bangladesh. Bull WHO 1980; 58:731-40.

10. Skoda JD, Mendis JB, Chia M. A survey in rural Bangladesh on diarrhoeal morbidity, water usage and related factors. Dacca: UNICEF (First Report), 1977.

11. Curlin GT, Azız KMA, Khan, MR. The influence of drinking tubewell water on diarrhea rates in Matlab Thana, Bangladesh. Dacca, Cholera Research Lab Working Paper #1, 1977.

12. Khan M, Curlin GT. Urban cholera study, 1974 and 1975. Dacca, International center for diarrhoeal disease research, Bangladesh Scientific Report #7, 1977.

13. Levine RJ, Khan MR, O'Sauza S, et al. Failure of sanitary wells to protect against cholera and other diarrhoea in Bangladesh. Lancet 1976; 10 July:86-9.

14. Sommer A, Woodward WE. The influence of protected water supplies on the spread of Classical/Inaba and El Tor/Ogawa cholera in rural East Bengal. Lancet 1972; 2:985-7.

15. Wray JD. Direct nutrition intervention and the control of diarrheal diseases in preschool children. AJCN 1978; 31:2073-82.

16. Koopman JS. Diarrhea and school toilet hygiene in Cali, Colombia. Amer J Epi 1978; 107:412-20.

17. Moore HA, de la Cruz L, Vargas-Mendez O. Diarrheal disease studies in Costa Rica: IV. The influence of sanitation upon the prevalence of intestinal infection and diarrheal disease. Amer J Epi 1965; 82:162-84.

18. Chandler AC. A comparison of helminthic and protozoan infection in two Egyptian villages two years after installation of sanitary improvements in one of them. Amer J Trop Med Hyg 1954; 3:59-73.

19. Freij L, Wall S. Exploring child health and its ecology. Acta Paediatrica Scandinavica 1977; Supplement 267:1-180.

20. Shiffman MA, Schneider R, Faigenblum JM, et al. Field studies on water sanitation and health education in relation to health status in Central America. Prog Wat Tech 1978; 11:143-50.

21. Bruch HA, Ascoli W, Scrimshaw NS, et al. Studies of diarrheal disease in Central America. V. Environmental fators in the origin and transmission of acute diarrheal disease in four Guatemalan villages. Amer J Trop Med Hyg 1963; 12:567-79.

22. Beck MD, Munoz JA, Scrimshaw NS. Studies on diarrhoeal diseases in Central America. I. Preliminary findings on cultural surveys of normal population groups in Guatemala. Amer J Trop Med Hyg 1957; 6:62-71.

23. Thacker JB, Music SI, Pollard RA, et al. Acute water shortage and health problems in Haiti. Lancet 1980; 1:471-3.

24. Raman V, Parhad NM, Deshpande AW, et al. Assessment and control of water quality in a town distribution system with reference to the incidence of gastrointestinal diseases. Prog Water Tech 1978; 11:65-71.

25. Rajaskeran P, Dutt PR, Pisharoti KA. Impact of water supply on the incidence of diarrhoea and Shigellosis among children in rural communities in Madurai. Ind J Med Res 1977; 66:189-99.

26. Trivedi BK, Gandhi HS, Shukla NK. Bacteriological water quality and incidence of water borne diseases in a rural population. Ind J Med Sci 1971; 25:795-801.

27. Kumar P, Sehgal BS, Singh R. Bore-hole disposal of excreta of children and diarrhoeal morbidity in a rural community. Env Hlth 1970; 12:155-9.

28. Van Zijl WJ. Studies on diarrhoeal diseases in seven countries by the WHO diarrhoeal disease advisory team. Bull WHO 1966; 35:249-61.

29. White GF, Bradley DJ, White AV. Drawers of Water. Chicago: Univ of Chicago Press, 1972.

30. Fenwick KWH. The short term effects of a pilot environmental health project in rural Africa: The Zaina scheme re-assessed after four years. Mimeographed undated.

31. Feachem R, Burns E, Cairncross S, et al. Water Health and Development. London: Tri-Med Books, 1978.

32. Ghannoum MA, Moore KE, Al-Dulaimi M, et al. The incidence of water-related diseases in the Brak area, Libya from 1977 to 1979, before and after the installation of water treatment plants. Zbl Bakt Hyg, I Abt Orig B 1981; 173:501-8.

33. Kourany M, Vasquez MA, Mata LJ. Prevalence of pathogenic enteric bacteria in children in 31 Panamanian communities. Amer J Trop Med Hgy 1971; 20:608-15.

34. Azurin JC, Alvero M. Field evaluation of environmental sanitation measures against cholera. Bull WHO 1974; 51:19-26.

35. Henry FJ. Environmental sanitation infection and nutritional status of infants in rural St. Lucia, West Indies. Trans Roy Soc Trop Med Hyg 1981; 75:507-13.

36. Bannaga S, Pickford J. Water-health relationships in the Sudan. Effluent and Water Treatment Journal 1978; Nov:560-9.

37. Burr ML. Diarrhoea and the drought. Public Hlth, London 1978; 92:86-7.

38. Rubenstein A, Boyle J, Odoroff Cl, et al. Effect of improved sanitary facilities on infant diarrhea in a Hopi village. Public Hlth Reports 1969; 84:1093-7.

39. Petersen NJ, Hines VD. The relation of summertime gastrointestinal illness to the sanitary quality of the water supplies in six Rocky Mountain communities. Amer J Hyg 1960; 71:314-20.

40. Schliessman DJ. Diarrhoeal disease and the environment. Bull WHO 1959; 21:381-6.

41. McCabe LJ, Haines TW. Diarrheal disease control by improved human excreta disposal. Public Hlth Reports 1957; 72:921-8.

42. Stewart WH, McCabe LJ, Hemphill EC, et al. The relationship of certain environmental factors to prevalence of Shigella infection IV. Diarrheal disease control studies. Amer J Trop Med Hyg 1955; 4:718-24.

43. Hollister AC, Beck MD, Gittelsohn RM, et al. Influence of water availability on Shigella prevalence in children of farm labor families. AJPH 1955; 45:354-62.

44. Watt J, Hollister AC, Beck MD, et al. Diarrhea diseases in Fresno county, California. AJPH 1953; 43:728-41.

45. Bahl MR. Impact of piped water supply on the incidence of typhoid fever and diarrhoeal diseases in Lusaka. Med J Zambia 1976; 10:98-9.

46. Martorell R, Habicht J-P, Yarbrough C, et al. Underreporting in fortnightly recall morbidity surveys. J Trop Ped 1976; 22:129-34.

47. Rahaman MM, Aziz KMS, Munshi MH, et al. A diarrhea clinic in rural Bangladesh: Influence of distance, age, and sex on attendance and diarrheal mortality. AJPH 1982; 72:129-34.

48. Kielman AA, Taylor CE, DeSweemer C, et al. The Narangwal experiment on interactions of nutrition and infection: II. Morbidity and mortality effects. Ind J Med Res 1978; 68 (Suppl.):21-41.

49. Dworkin D, Dworkin J. Water supply and diarrhea: Guatemala revisited. Washington, D.C.: AID Evaluation Special Study #2, 1980.

50. Cvjetanovic B, Chen L, Kronmal R, et al. Measuring and evaluating diarrhoea and malabsorption in association with a village water supply and sanitation. Washington, D.C.: Wash Tech Report #12, 1981.

51. Martorell R. Acute morbidity and physical growth in rural Guatemalan children. Amer J Dis Child 1975; 129:1296-301.

52. Mata LJ, Urrutia JJ, Albertazzi C, et al. Influence of recurrent infections on nutrition and growth of children in Guatemala. AJCN 1972; 25:1267-75.

53. Rowland MGM, Cole TJ, Whitehead RG. A quantitative study into the role of infection in determining nutritional status in Gambian village children. Brit J Nutr 1977; 37:441-50.

54. Esrey SA, Habicht J-P. Nutritional anthropometric indicators for evaluating water and sanitation projects. 1983; Paper presented at the International Workshop on Measuring the Health Impact of Water and Sanitation Programs. 21-25 Nov., Cox's Bazaar, Bangladesh.

55. Christiansen N, Mona JO, Herrera MG. Family social characteristics related to physical growth of young children. Brit J Prev Soc Med 1975; 29:121-30.

56. Yee V. Household level correlates of child nutritional status in Fiji. 1983; Master's Thesis, Div Nutr Sci, Cornell University.

57. Tomkins AM, Drasar BS, Bradley AK, et al. Water supply and nutritional status in rural Northern Nigeria. Trans Roy Soc Trop Med Hyg 1978; 72:239-43.

58. Merrick T. The effect of piped water on early childhood mortality in urban Brazil, 1970-1976. World Bank Working Paper #594, 1983.

59. Victora CS, Blank N. Epidemiology of infant mortality in Rio Grande de Sol, Brazil. J Trop Med Hyg 1980; 83:177-86.

60. Haines MR, Avery RC. Differential infant and child mortality in Costa Rica: 1968-1972. Pop Studies 1982; 36:31-43.

61. Weir JM. An evaluation of health and sanitation in Egyptian villages. J Egyptian Pub Hlth Assoc 1952; 27:55-114.

62. Zaheer M, Prasad BG, Govil KK, et al. A note on urban water supply in Uttar Pradesh. J Ind Med Assoc 1962; 38:177-82.

63. Anker R, Knowles JC. An empirical analysis of morbidity differentials in Kenya at the macro and micro levels. Econ Dev & Cultural Change 1980 29:165-85.

64. Butz WP, DaVanzo J, Habicht J-P, et al. Environmental factors in the relationship between breastfeeding and infant mortality: The role of sanitation and water in Malaysia. Amer J Epi, 1984; 119:516-25.

65. Butz WP, DaVanzo J, Habicht J-P, et al. Biological and behavioral influences on the mortality of Malaysian infants. Santa Monica: The Rand Corporation, 1982.

66. Patel M. Effects of the health services and environmental factors on infant mortality: The case of Sri Lanka. J Epi Comm Hlth 1980; 34:76-82.

67. Megeema SA. Socio-economic determinants of infant and childhood mortality in Sri Lanka: An analysis of post-war experience. London: World Fertility Survey Report #8, 1980.

68. Habicht J-P, Butz WP. Measurements of health and nutrition effects of large-scale nutrition intervention projects. In: Evaluating the impact of nutrition and health programs. Plenum Publishing Corporation, 1979.

69. Esrey SA, Feachem RG, Hughes J. Interventions for the control of diarrhoeal diseases among young children: Improving water supplies and excreta disposal facilities. Bull WHO 1985; 63:(4), in press.