A RAND NOTE

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

Jean-Pierre Habicht and William P. Butz

October 1980

N-1601-RF

Prepared For

The Rockefeller Foundation
This research was sponsored by the Rockefeller Foundation under Grant RF 75058.
A RAND NOTE

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

Jean-Pierre Habicht and William P. Butz

October 1980

N-1601-RF

Prepared For

The Rockefeller Foundation

Rand
SANTA MONICA, CA. 90406
PREFACE

This Rand Note reproduces the text of a paper published in Robert E. Klein et al. (eds.), Evaluating the Impact of Nutrition and Health Programs, Plenum Publishing Corporation, 1979. That collection is the proceedings volume of the Pan-American Health Organization's International Conference on the Assessment of the Impact of Nutrition and Related Health Programs, Panama, 1977. The research on which the Note is based was partially supported by Grant RF 75058 to The Rand Corporation from the Rockefeller Foundation.

Jean-Pierre Habicht is a consultant to The Rand Corporation; William P. Butz is a member of the Rand research staff.
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS OF LARGE-SCALE NUTRITION INTERVENTION PROJECTS

Jean-Pierre Habicht,
Cornell University
Ithaca, New York

and

William P. Butz
Rand Corporation
Santa Monica, California

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
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

upon previous scientific and administrative knowledge (1). In gen-
eral, these include field intervention studies, public health pilot
projects, evaluations in conjunction with implementations of an in-
tervention, and ultimately continuous monitoring of program impact.
Each of these is discussed more fully in the paragraphs that fol-
low. The knowledge resulting from each is not only important in
deciding whether a given type of intervention is likely to be use-
ful 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 re-
sponses 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 infec-
tions 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 dif-
ferent 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 pro-
tein-calorie nutrition could not improve growth rates to the levels
seen in developed countries, in all probability because of recur-
rent and frequent diarrheal disease. These findings contradicted
the inferences drawn from research in laboratory and clinical set-
tings 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 experimen-
tal design to assure the comparability of data from those individ-
uals receiving the different kinds of interventions. Only by such
careful control can the inferences be sufficiently strong and gen-
eralizable to individuals in other populations to be of use to
scientists and clinicians in that they are assured of the effec-
tiveness 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
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 benefitting 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 (19). 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).
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 sam-
pling 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 that would be anticipated when looking at individuals alone. A similar situation can be postulated for the effect on 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 not only 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/

1/ Clinical pathologists with a concern for prognosis have described the identical characteristic and called it "predictability". 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.
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 heights 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.

![Graph showing individual probability and selectivity](image)

**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).

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 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% anemias 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.
# MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

## Table 1a. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON TOTAL DIETARY INTAKE

<table>
<thead>
<tr>
<th>Reference: Years Publication</th>
<th>Supplement</th>
<th>Supervised/Measured Ingestion</th>
<th>Replacement Estimated</th>
<th>Age in Months</th>
<th>Intake Before Intervention or Compared to Control</th>
<th>Intake after Intervention</th>
<th>Change in Intake</th>
</tr>
</thead>
<tbody>
<tr>
<td>(17)-1963</td>
<td>391 Kcal.* 13.4 g.</td>
<td>No</td>
<td>Yes</td>
<td>48-96</td>
<td>From previous study 1300Kcal 37 g.</td>
<td>Not reported. Claim no substitution effect</td>
<td>?</td>
</tr>
<tr>
<td>(18)-1965</td>
<td>101-286 Kcal 9.5-10.1 g.</td>
<td>No</td>
<td>No**</td>
<td>6-12</td>
<td>Not measured</td>
<td>Not measured</td>
<td>?</td>
</tr>
<tr>
<td>(19)-1967-9</td>
<td>Not specified</td>
<td>No</td>
<td>Yes</td>
<td>0-59</td>
<td>678 Kcal. 20 g.</td>
<td>1040 Kcal. 30 g.</td>
<td>+362 Kcal. +10 g.</td>
</tr>
<tr>
<td>(20)-1970</td>
<td>250 Kcal. 12.5 g.</td>
<td>No</td>
<td>Yes*</td>
<td>36-96</td>
<td>Not reported</td>
<td>Not reported</td>
<td>?</td>
</tr>
<tr>
<td>(21)-1970</td>
<td>300 Kcal. 10 g.</td>
<td>?</td>
<td>Yes</td>
<td>12-60</td>
<td>Not reported</td>
<td>Not reported. Claim no substitution effect</td>
<td>+300 Kcal. +10 g.</td>
</tr>
<tr>
<td>(22)-1973</td>
<td>?</td>
<td>Yes</td>
<td>Yes</td>
<td>35-59</td>
<td>486 Kcal. 13.7 g.</td>
<td>1181 Kcal. 35.3g.</td>
<td>+695 Kcal. +21.6 g.</td>
</tr>
<tr>
<td>(23)-1973</td>
<td>310 Kcal. 3 g.</td>
<td>Yes</td>
<td>Yes</td>
<td>12-60</td>
<td>700 Kcal. 18 g.</td>
<td>1010 Kcal. 21 g.</td>
<td>+319 Kcal. +3 g.</td>
</tr>
<tr>
<td>(2)-1977</td>
<td>?</td>
<td>Yes</td>
<td>Yes</td>
<td>24-72</td>
<td>3700 KJ.</td>
<td>Not reported. Claim no substitution effect</td>
<td>+800-1300KJ. +6.2-11.8 g.</td>
</tr>
<tr>
<td>(2)-1977</td>
<td>200 Kcal 14 g.</td>
<td>Yes</td>
<td>Yes</td>
<td>12-36</td>
<td>78% of recommended energy intake</td>
<td>Claim 10% substitution effect</td>
<td>+180 Kcal. +13.5 g.</td>
</tr>
</tbody>
</table>

**Legend:**
- * = Energy intake/day and Protein intake/day
- ** = Cause for doubting author's inferences
- ? = Uncertain
Table 1b. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON HEIGHT

<table>
<thead>
<tr>
<th>Reference Year Publication</th>
<th>Type of Height</th>
<th>Age in Months</th>
<th>Best-Worst = R **</th>
<th>Before Intervention or Compared to Control</th>
<th>After Intervention</th>
<th>Change in Level % of R</th>
<th>Statist. Signif. (P&lt;.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(17)-1963</td>
<td>1 year increment in cm/year</td>
<td>Boy 6-12</td>
<td>6.0-5.6=0.4</td>
<td>4.3* 0% 4.4* 0% 0.1 - - No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Girl 6-12</td>
<td>6.0-5.3=0.7</td>
<td>3.7* 0% 4.1* 0% 0.4 - - No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(18)-1965</td>
<td>6 month increment in cm/year</td>
<td>6-12</td>
<td>16.0=8.0</td>
<td>7.4- 7.6 7.6 20% 7.9 24% -0.1 -1X No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(19)-1967-9</td>
<td>3 year increment in cm/3 year</td>
<td>0-11</td>
<td>41.0=21.0</td>
<td>17.6* 4% 16.7* 0% -0.9 - - No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-48</td>
<td>30-17.5=12.5</td>
<td>19.9 19% 22.0 36% 2.1 1X Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(20)-1970</td>
<td>Attained development age</td>
<td>36-96</td>
<td>1.0-0.5=0.5</td>
<td>0.7 40% 0.7 40% 0.0 0% 0% No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11)-1970</td>
<td>Attained height cm</td>
<td>24-35</td>
<td>90-75=15.0</td>
<td>77.1 14% 78.3 22% 1.2 8% No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>36-47</td>
<td>99-81=18.0</td>
<td>82.0 6% 84.6 20% 2.6 16% No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>48-59</td>
<td>106-88=18.0</td>
<td>91.0 17% 90.8 16% -.1 -1X No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>60-71</td>
<td>113-95=18.0</td>
<td>97.4 14% 98.7 21% 1.3 7% No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(21)-1970</td>
<td>6 month increment in cm/6 month</td>
<td>35-59</td>
<td>3.5-2.9=0.6</td>
<td>1.94* 4% 2.7 33% 0.8 13% Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(22)-1973</td>
<td>14 month increment in cm/14 month</td>
<td>12-23</td>
<td>12.0-8.0=5.7</td>
<td>6.3* &lt;.01 9.3 44% 2.8 40% Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24-35</td>
<td>25.0-21.0=4.0</td>
<td>7.8 37% 9.5 100% 1.7 63% Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>36-47</td>
<td>38.5-26.8=11.7</td>
<td>7.4 35% 9.1 135% 2.5 118% Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>48-60</td>
<td>7.5-6.8=0.7</td>
<td>7.3 71% 8.4 229% 1.1 157% Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(23)-1973</td>
<td>6 month increment in cm/6 month</td>
<td>24-72</td>
<td>3.5-2.9=0.6</td>
<td>2.6* &lt;.01 3.2 50% 1.2 200% Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)-1977</td>
<td>2 year increment in cm/2 year</td>
<td>12</td>
<td>20-12= 8.0</td>
<td>15.7 46% 18.3 79% 2.5 33% Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type of Weight</th>
<th>Age in Months</th>
<th>Best-Worst * R *</th>
<th>Before Intervention or Compared to Control</th>
<th>After Intervention</th>
<th>Change in</th>
<th>Statist. Signif.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>% R</td>
<td>Level</td>
<td>% R</td>
<td>Level</td>
<td>% R</td>
</tr>
<tr>
<td>(17)-1963</td>
<td>1 yr increment as kg/year</td>
<td>48-96</td>
<td>2.041.6±0.4</td>
<td>1.6  0.2</td>
<td>2.1   125%</td>
<td>0.5</td>
<td>125%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.4±0.6</td>
<td>1.3   40%</td>
<td>2.6   200%</td>
<td>1.3   217%</td>
</tr>
<tr>
<td>(18)-1965</td>
<td>6 month increment presented as kg/yr</td>
<td>0-11</td>
<td>6.0-3.5±2.5</td>
<td>3.6  4%</td>
<td>3.9   16%</td>
<td>0.3</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12-48</td>
<td>5.2   47%</td>
<td>5.7   80%</td>
<td>0.5   33%</td>
</tr>
<tr>
<td>(19)-1967</td>
<td>Regression Coefficient (kg/year)</td>
<td>36-96</td>
<td>1.0-0.5±0.5</td>
<td>0.68</td>
<td>0.68   36%</td>
<td>0.0</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(20)-1970</td>
<td>Attained Developmental Age</td>
<td>24-35</td>
<td>16.5-8.0±6.5</td>
<td>8.8  12%</td>
<td>9.2   18%</td>
<td>0.4</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>36-47</td>
<td>15.5-9.0±6.5</td>
<td>10.1</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48-59</td>
<td>17.5-11.0±6.5</td>
<td>12.0</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60-71</td>
<td>19.5-13.0±6.5</td>
<td>13.3</td>
<td>15%</td>
</tr>
<tr>
<td>(21)-1970</td>
<td>6 months increment (kg/1/2 yr)</td>
<td>35-59</td>
<td>1.0-0.7±0.3</td>
<td>1.23</td>
<td>2.28   10%</td>
<td>1.0</td>
<td>350%</td>
</tr>
<tr>
<td>(22)-1973</td>
<td>6 month increment (kg/14 months)</td>
<td>22-23</td>
<td>2.6-1.7±0.9</td>
<td>1.74</td>
<td>2.35   72%</td>
<td>0.61</td>
<td>68%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24-35</td>
<td>2.3-1.7±0.6</td>
<td>2.34</td>
<td>71%</td>
</tr>
<tr>
<td>(23)-1973</td>
<td>6 month increment (kg/1/2 year)</td>
<td>48-59</td>
<td>2.3-1.7±0.6</td>
<td>1.58</td>
<td>2.04   38%</td>
<td>0.46</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60-71</td>
<td>2.3-1.7±0.6</td>
<td>4.06</td>
<td>18%</td>
</tr>
<tr>
<td>(2)-1977</td>
<td>2 yr increment (kg/2 yr)</td>
<td>12</td>
<td>4.0-3.0±1.0</td>
<td>3.67</td>
<td>4.50   150%</td>
<td>0.83</td>
<td>83%</td>
</tr>
</tbody>
</table>

Legend: * R = Physiological Range (see text, page 155)
Table Ia. EFFECT OF PROTEIN-CALORIE INTERVENTION IN PRESCHOOL CHILDREN ON MORBIDITY

<table>
<thead>
<tr>
<th>Reference Year Publication</th>
<th>Type of Morbidity</th>
<th>Age in Months</th>
<th>Before Intervention or Compared to Control</th>
<th>After Intervention</th>
<th>Change in Significance (p&lt;0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(18) 1965</td>
<td>Illness score based on effect of illness on growth rate</td>
<td>6-12</td>
<td>0.5 - 0.6</td>
<td>0.6 - 1.2</td>
<td>-0.3 Better</td>
</tr>
<tr>
<td>(19) 1967-9</td>
<td>Average days ill per year</td>
<td>0-59</td>
<td>13</td>
<td>22</td>
<td>+9 Worse</td>
</tr>
<tr>
<td>(11) 1970</td>
<td>% children with symptoms of protein-calorie malnutrition</td>
<td>12-60</td>
<td>23.0</td>
<td>11.3</td>
<td>-11.7 Better</td>
</tr>
</tbody>
</table>

Legend: * No statistical significance testing done

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

<table>
<thead>
<tr>
<th>Reference Year Publication</th>
<th>Type of Mortality</th>
<th>Age in Months</th>
<th>Best-Worst R = **</th>
<th>Before Intervention or Compared to Control</th>
<th>After Intervention</th>
<th>Change in Significance (p&lt;0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(19) 1967-9</td>
<td>Infant (deaths/yr/1000 births)</td>
<td>0-11</td>
<td>16-200=184</td>
<td>186</td>
<td>191</td>
<td>4% -5 Worse</td>
</tr>
<tr>
<td></td>
<td>Medical Intervention</td>
<td>136</td>
<td>35%</td>
<td>88</td>
<td>60%</td>
<td>46 25% No</td>
</tr>
<tr>
<td></td>
<td>Nutrition Intervention</td>
<td>182</td>
<td>10%</td>
<td>146</td>
<td>29%</td>
<td>36 20% No</td>
</tr>
<tr>
<td>(20) 1970</td>
<td>Infant (deaths/yr/1000 births)</td>
<td>0-11</td>
<td>16-200=184</td>
<td>135</td>
<td>34%</td>
<td>48 83% 87 47% Yes</td>
</tr>
<tr>
<td>(19) 1967-9</td>
<td>Preschool (deaths/1000 children)</td>
<td>12-48</td>
<td>0.3-90=89.7</td>
<td>81-9</td>
<td>10%</td>
<td>50-40 45% 31 35% Yes</td>
</tr>
<tr>
<td></td>
<td>Medical Intervention</td>
<td>50-40</td>
<td>45%</td>
<td>35-55</td>
<td>61%</td>
<td>10 17% No</td>
</tr>
<tr>
<td></td>
<td>Nutrition Intervention</td>
<td>56-34</td>
<td>38%</td>
<td>24-66</td>
<td>74%</td>
<td>32 36% Yes</td>
</tr>
</tbody>
</table>

Legend: * R = Physiological Range (see text, page 155)
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 severe maternal deprivation. We have since persuaded ourselves otherwise, at least as far as birth weight 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 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.
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 affect 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 evaluation.

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
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 (37) give a more complete list and discuss methodological considerations that arise in evaluating their effects.
Table II. SUMMARY OF DESIGN AND INDICATOR CHARACTERISTICS FROM INTERVENTION PROGRAMS IN TABLE I.

<table>
<thead>
<tr>
<th>References:</th>
<th>(13) (18) (19) (20) (21) (22) (23) (24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Unit of Intervention</td>
<td>No No Yes Yes ? ? Yes Yes Yes</td>
</tr>
<tr>
<td>b. Was choice of type and quality of intervention based on more evidence than dietary survey information?</td>
<td>Yes ? Yes ? ? ? Yes Yes Yes</td>
</tr>
<tr>
<td>II. Control of Intervention</td>
<td>Village Village Child Village</td>
</tr>
<tr>
<td>a. Did intervention reach central distribution center in adequate quantity? (by assay)</td>
<td>Yes Yes Yes Yes Yes Yes Yes Yes Yes</td>
</tr>
<tr>
<td>b. Did intervention reach home by documented record of distribution?</td>
<td>Yes Yes Yes Yes Yes Yes Yes Yes Yes</td>
</tr>
<tr>
<td>c. Did intervention reach target persons in adequate quantity by documented qualitative measurement by:</td>
<td>Village Village Village Village Village</td>
</tr>
<tr>
<td>occasion survey recording of intervention?</td>
<td>Yes No No No Yes Yes Yes No Yes</td>
</tr>
<tr>
<td>d. Was replacement effect sought measured adequately and taken into account?</td>
<td>No No Yes Yes No No Yes No Yes</td>
</tr>
<tr>
<td>III. Choice and Measurement of Indicator</td>
<td>See Table I</td>
</tr>
<tr>
<td>b. Was analysis made of variability due to measurement and short-term intrinsic variability?</td>
<td>No No No No No No No Yes Yes</td>
</tr>
<tr>
<td>Diet? Anthropometry? Mortality? Mortality? Other?</td>
<td>No No No No No No No No No No No</td>
</tr>
<tr>
<td>IV. Control of Confounding Factors</td>
<td>Yes No No No No No No No No No No No No</td>
</tr>
<tr>
<td>a. Adequate controls</td>
<td>Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes</td>
</tr>
<tr>
<td>b. Replication</td>
<td>Yes No No No No No No No No No No No No</td>
</tr>
<tr>
<td>c. Analysis for Congruity</td>
<td>Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes</td>
</tr>
<tr>
<td>Dietary Insulation? Anthropometry? Mortality? Mortality? Other?</td>
<td>Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes</td>
</tr>
<tr>
<td>d. Analysis Stratified by:</td>
<td>Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes</td>
</tr>
<tr>
<td>Age Sex Dose of Intervention Indicators of Self-selection Lactation Food supply or home diet Income or health Education of mother or other family member Medical care, disease risk or disease experience Secular trend</td>
<td>Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes Yes</td>
</tr>
</tbody>
</table>

Legend: - = not relevant; ? = not clear in report; * = cause for doubting authors' inferences. Interven is associated with "No", it indicates that we consider this a serious defect which calls into question some or all of the authors' inferences. There * 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.
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 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/ 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 (35). 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.
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 the supplement gained 2.1 kg per year. Their percentage of R was therefore (2.1-1.6)/0.4 = 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 500 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, there-
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

fore, 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 (49).

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 1.6 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 (post-neonatal), whereas about 60% of all infant deaths are post-neonatal in developing countries. However, this ratio of post-neonatal 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 post-neonatal 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
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 (31). 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 towards 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 measurement and evaluation difficulties, it is useful to ask why integrated interventions are becoming popular. Frustration of researchers as well as policymakers with the 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
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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).

It 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 re-arrange 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
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS

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 confound-
ing factors, and indicator variables intelligently, we can accumu-
late 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 expen-
sive 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 repleci-
ate 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 inter-
vention and be specific for that intervention. Knowledge about
sensitivity must come from single purpose intervention studies.
In these studies specificity of the indicator's response to nutri-
tion 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 indi-
cators 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 general-
ized. 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 "sci-
entific" constraints to recommend the continued use of "commonly
used" but unvalidated indicators will only result in falsely nega-
tive evaluations of valuable and useful programs. It would be bet-
ter 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 Agricul-
tural 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 interven-
   tions in populations and individuals. In: S. J. Bosch
   and J. Arias (eds.). Evaluation of Child Health Servi-
   ces: The Interface Between Research and Medical Practice.
   DREW Publication No. (NIH) 78-1066. Washington, D. C.:
3. Scrimshaw, N. S., and M. Bēhar. Malnutrition in underdeveloped
   Health Service Publication No. 825. Washington, D. C.:
5. Habicht, J-P., J. M. Lane, and A. J. McDowell. National nutri-
   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 Internation-
MEASUREMENT OF HEALTH AND NUTRITION EFFECTS


   I. Study plan and experimental design. Arch. Environ. Health, 14: 657–662, 1967; II. Field reconnaissance,


MEASUREMENT OF HEALTH AND NUTRITION EFFECTS
