PN-AAS-283
ISN: 39058

CSF

EFFECT OF CLASSIFICATION ERRORS UPON EVALUATION OUTCOMES: NUTRITION PROGRAMS REVISITED

June 1984

USAID Contract Number DAN-1406-C-00-3045-00

by
William D. Drake
Roy I. Miller
Robert J. Timmons

Community Systems Foundation
1130 Hill St.
Ann Arbor, Michigan 48104
Phone: (313) 761-1357
Telex: 759414

EFFECT OF CLASSIFICATION ERRORS UPON EVALUATION OUTCOMES: NUTRITION PROGRAMS REVISITED

June 1984

USAID Contract Number DAN-1406-C-00-3045-00

by
William D. Drake
Roy I. Miller
Robert J. Timmons

Community Systems Foundation
1130 Hill St.
Ann Arbor, Michigan 48104
Phone: (313) 761-1357
Telex: 759414

ACKNOWLEDGMENTS

This study has received the support and assistance of many individuals and organizations. Three separate offices in USAID provided financial support: the Office of Nutrition, the Office of Food for Peace and the Office of Evaluation of the Asia Bureau. We were greatly helped by useful discussions with Martin J. Forman, Hope Sukin, Judy Gilmoré, Harold Rice and Maureen Norton, from these offices.

Additional substantive discussions were held with Pudjo Rahardjo, Director of Research, BKKBN, Indonesia, and Professor James Koopman of the School of Public Health at the University of Michigan.

EXECUTIVE SUMMARY

Over the last several years there has been increasing emphasis upon the evaluation of nutrition programs using impact indicators such as malnutrition prevalence. Many studies have been undertaken and much has been written about how to conduct such studies. However, there is one methodological issue which has not been dealt with adequately in the evaluation and policy literature but which is of central importance to nutrition evaluations--namely the effect of misclassifying the nutritional status of children due to imperfect measurement techniques. The importance of this misclassification arises from the effect it has upon reporting the changes that occur during the course of a nutrition program. Thus, it is warranted, and perhaps critical, to consider the consequences of misclassification and to derive procedures for adjusting for its presence.

Seven sources of error in classifying nutritional status have been identified: 1) imperfect correspondence between measure and malnutrition; 2) inappropriate standards and/or classification; 3) incorrect measurements due to faulty measurement devices, measurement variation between and within subjects and observers, and inaccurate estimates of age; 4) recording errors; 5) transcription errors; 6) coding or keying errors; 7) and errors introduced during analysis.

In Section II and III these sources of misclassification are structured mathematically in terms of sensitivity and specificity. Expressions showing the relationship between observed point prevalence or change in prevalence and the corresponding true prevalence are formulated under varying conditions.

Section VI provides whatever evidence exists at present regarding the range of error for each of these seven sources of misclassification. When sensitivity and specificity are

equal and constant over time, we underestimate the true change in prevalence; when sensitivity and specificity differ but remain constant over time, we also underestimate the true change in prevalence; and in the general case where sensitivity and specificity differ and vary over time, true prevalence can be under- or overestimated, or estimated in the opposite direction of the true change.

Without specific knowledge suggesting sensitivity and specificity change over time, it is perhaps most reasonable to assume they remain constant or do not vary significantly over time. If this is true, there will always be an understatement of the true changes in malnourishment during the course of a project. Until the necessary field work can be done, a reasonable estimate of sensitivity and specificity in typical nutrition programs is between .70 and .80. (These estimates are reasonable when clinical diagnosis is representative of truth. On the other hand, if an accurate measure of the field indicator, weight-for-age, were used to represent truth, these estimates of sensitivity and specificity may be construed as unusually low.) When sensitivity and specificity are .75 and constant over time, only half of the true change in prevalence is reported. It is, however, unlikely for specificity to be as low as .75 when the prevalence of malnutrition in the population is relatively low.

A reanalysis of the results from eight different nutrition evaluations was performed under differing assumptions of sensitivity and specificity. These findings bear out the impact misclassification has upon reported program impact and are shown as Tables VI through XIV.

The study concludes with guidelines for adjusting results obtained in an impact analysis. They are as follows: 1) select an approximation of values for sensitivity and specificity 2) determine if there is clear evidence pointing to a change in either sensitivity or

specificity over time; 3) using the best estimate for sensitivity and specificity over time, use the appropriate correction formula to compute the adjustments to prevalence rates; 4) reassess the resultant estimates of sensitivity and specificity; 5) and report both unadjusted changes in prevalence and corrected changes.

TABLE OF CONTENTS

ACKNOWLEDGMENTS i

LIST OF TABLES vi

LIST OF FIGURES viii

EXECUTIVE SUMMARY ii

I INTRODUCTION 1

II DEFINITIONS 3

III UNDERREPORTING CHANGE IN PREVALENCE 6

IV TESTING STATISTICALLY FOR CHANGE IN PREVALENCE 10

V A DISCUSSION OF SOME UNDERLYING ASSUMPTIONS . . 11

VI SOURCES OF ERROR LEADING TO MISCLASSIFICATION . 16

VII CHANGES IN SENSITIVITY AND SPECIFICITY OVER TIME 25

VIII IMPACT OF MISCLASSIFICATION ON EVALUATION OUTCOMES 30

IX GUIDELINES FOR CORRECTING EVALUATION RESULTS . 43

REFERENCES 46

APPENDIX A 51

APPENDIX B 57

LIST OF TABLES

Table 1. Portrayal of Misclassification 4

Table 2. Sources of Error in Classifying Nutritional Status 18

Table 3. Misclassification with Random Weighing Errors 22

Table 4. Field vs. Computer Determined Nutritional Grade 24

Table 5. Illustration of Consequences of Changing Sensitivity and Specificity 27

Table 6. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .90; Sensitivity after = .90; Specificity before = .90; Specificity after = .90) . . . 32

Table 7. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .85; Sensitivity after = .85; Specificity before = .85; Specificity after = .85) . . . 33

Table 8. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .80; Sensitivity after = .80; Specificity before = .80; Specificity after = .80) . . . 34

Table 9. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .75; Sensitivity after = .75; Specificity before = .75; Specificity after = .75) . . . 35

Table 10. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .80; Sensitivity after = .85; Specificity before = .80; Specificity after = .80) . . . 36

Table 11. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .80; Sensitivity after = .75; Specificity before = .80; Specificity after = .80) . . . 37

Table 12. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .85; Sensitivity after = .80; Specificity before = .85; Specificity after = .80) . . . 38

Table 13. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .80; Sensitivity after = .80; Specificity before = .80; Specificity after = .85) . . . 39

Table 14. Comparison Between Observed and Actual Malnourishment--Prevalence Rates Expressed in Percentages (Assuming Sensitivity before = .80; Sensitivity after = .85; Specificity before = .80; Specificity after = .85) . . . 40

LIST OF FIGURES

Figure 1. 13

Figure 2. 15

I INTRODUCTION

Over the last several years there has been increasing emphasis upon the evaluation of nutrition programs using empirical impact indicators such as malnutrition prevalence. Many investigations have been undertaken and much has been written about how to conduct such studies. However, there is one methodological issue which has not been dealt with adequately in the evaluation and policy literature but which is of central importance to nutrition evaluations--namely the effect of misclassifying the nutritional status of children due to less-than-perfect measurement techniques. In the ensuing investigation we are <u>not</u> concerned with biases introduced by improper sampling techniques or with consistent biases arising from improperly calibrated measurement instruments and/or biased personnel. (These kinds of biases are well understood and are treated extensively in the literature.) Rather we are concerned with the biases created <u>solely</u> by misclassification arising from unreliable and/or imperfectly valid measurement techniques.[1]

While this misclassification phenomenon generally has been avoided in the recent flood of evaluation and policy literature, some aspects of it have been discussed at some length and with a fair amount of confusion in the literature of statistics and statistical epidemiology. Our approach is (1) to focus upon clarifying some of the confusion in the literature and then (2) to develop the implications of misclassification on evaluation, especially as related to policy formulation, and (3) to provide for adjusting evaluation results to reflect more accurately the true

[1]The following citation discusses the implications of this phenomenon from an evaluation and policy perspective. William D. Drake, Roy I. Miller, and Margaret Humphrey, <u>Final Report: Analysis of Community-Level Nutrition Programs. Volume I.</u> (Ann Arbor: Community Systems Foundation, October 1980), pp. 89-94.

changes which occurred during the course of the intervention.

We will show that with reasonable assumptions concerning the rate of misclassification, the comparisons of prevalence rates in two populations (or in the same population at two points in time) most often <u>understates</u> the true difference in prevalence rates. From a policy perspective, it is important to note that under typical field conditions the amount of misclassification experienced leads to <u>substantial</u> understatement of results. In at least some instances this understatement has most certainly led policy makers to undervalue the true changes experienced during the course of the project and makes the correction for this phenomenon especially important at this time.

In the discourse that follows, we will consider a variety of assumptions with regard to the nature of misclassification in the context of nutrition. Specifically, we will be concerned with the errors in classification arising from the use of anthropometry in the measurement of nutritional status. In general, misclassification arises for two reasons: 1) the indicator or test used to detect some given condition does so imperfectly and 2) mistakes are made in the mechanical aspects of computation such as in measurement, arithmetic and recording. In statistical jargon, the correspondence between an indicator and an underlying phenomenon is called "validity" while the replicability of the procedures used in data generation, in the field, is called "reliability".

With regard to the classification of individuals as to their nutritional status using anthropometry, "validity" refers to the degree of correspondence between any given anthropometric score and the underlying phenomenon of malnutrition. "Reliability" refers to the ability while in the field to repeat the measurement process; if field personnel are unable to reproduce consistent results for a

single individual, it is more likely that any single observation is incorrect. In theory, improved field protocol could eliminate errors in measurement, computation and recording; that is, anthropometry could be perfectly reliable. In practice, however, this is highly improbable, if not impossible. On the other hand, due to the complex etiology of malnutrition and the numerous ways it manifests itself in individuals, no known classification, however costly, is completely valid. Thus it is important, perhaps critically so, to consider the natural consequences of misclassification and to derive procedures for adjusting for its presence.

II DEFINITIONS

To develop the mathematical equations necessary to estimate the magnitude of error due to misclassification under different assumptions, we need, first, to present some definitions. Central to our work is the concept of prevalence, for it is the change in prevalence which is so often understated in nutrition evaluation projects.

> *Prevalence*, denoted P in this report, is the proportion of cases having a disease or condition in a population at a given time. We often speak of a point prevalence rate: the ratio of the number of cases of a disease to the total number of people in the population.

The point prevalence of a disease is thus a proportion with a value between "0" and "1". Sometimes, it is expressed as a percentage, for example, in a statement that the rate of malnutrition is 82%. If all individuals in a population have equal probability of having the disease in question, the prevalence rate is also the probability that any single individual has the disease. (This is generally not the case.)

The prevalence of a disease in a population is generally determined through some diagnostic test

administered to a representative sample of the population. When a test indicates that an individual has the disease in question, its result is said to be "positive". For example, in the nutrition field, anthropometry is used quite frequently to estimate the prevalence of malnutrition, the test being the comparison of some anthropometric score for an individual to a predetermined "standard". If the score falls below some percentage of that "standard", called the cut-point, the individual is judged to be malnourished. We are most interested in two characteristics of a test.

> *Sensitivity*, denoted U in this report, is the chance, or probability, that a diagnostic test will be positive when applied to someone known to have the disease.
>
> *Specificity*, denoted V in this report, is the chance, or probability, that a diagnostic test will be negative when applied to a healthy person.

Often, these concepts are presented symbolically in the literature by portraying the classification possibilities in a 2 x 2 contingency table. This symbolic representation is reproduced in Table 1.

TABLE 1

PORTRAYAL OF MISCLASSIFICATION

	Truly Ill	Truly Healthy
Test Positive	a	c
Test Negative	b	d

Based on this table, we can offer the following definitions.

$$(1)\ \text{TRUE PREVALENCE} = P = \frac{a+b}{a+b+c+d}$$

$$(2)\ \text{OBSERVED PREVALENCE} = \hat{p} = \frac{a+c}{a+b+c+d}$$

$$(3)\ \text{SENSITIVITY} = U = \frac{a}{a+b}$$

$$(4)\ \text{SPECIFICITY} = V = \frac{d}{c+d}$$

Note, the probability of misclassifying a truly ill child is (1-U)=b/(a+b) and the probability of misclassifying a truly healthy child is (1-V)=c/(c+d). In much of the literature, especially that generated by non-epidemiologists, the mathematical presentations are made in terms of these probabilities of misclassification rather than the probabilities of correct classification as defined in the table above. This adds some confusion regarding the entire subject. One particularly interesting presentation of these concepts is made in terms of conditional probabilities and Bayes' theorem. In this work by Fleiss, however, specificity is defined as the probability of misclassifying a normal person while sensitivity is defined in the usual way.[2]

In the literature, repeated reference is made to the fact that both probabilities, sensitivity and specificity, are constant across different populations with different prevalences.[3] This is true if both the test and the

---

[2]Joseph L. Fleiss, Statistical Methods For Rates and Proportions (New York City: John Wiley & Sons, 1973), pp. 3-6.

[3]Walter J. Rogan and Beth Gladen, "Estimating Prevalence From the Result of a Screening Test," American Journal of Epidemiology, 107 (1978), p. 71. and Jean-Pierre

underlying condition are truly discrete. However, this is not the case in diagnosing malnutrition. Because of the fact that the classification is made from a score with an underlying continuous distribution--a distribution that should change in the presence of an intervention--this constancy of sensitivity and specificity in populations with different prevalence rates does not necessarily hold.[4] This case is discussed further in a later section.

III UNDERREPORTING CHANGE IN PREVALENCE

In most cases, the misclassification of individuals due to imperfect sensitivity and/or specificity of a test results in a bias when the observed prevalence is used as the estimate of the true prevalence. Using the notation introduced above, we can express the observed prevalence as a simple function of the true prevalence and the sensitivity and specificity of the test.

$$(5)\quad \hat{p} = PxU + (1-P)x(1-V)$$

In words, we observe a prevalence rate, for example, a malnutrition rate, made up of two components: the true rate multiplied by the proportion of those classified correctly (sensitivity) and the prevalence of normals (1-P) multiplied by the rate at which they are misclassified (1-specificity). Through algebraic manipulation of the equations used above for $\hat{p}$, P, U, and V, we could verify that equation (5) is, indeed, correct. Examination of equation (5) reveals that a

---

Habicht, Linda D. Meyers and Cavell Brownie, "Indicators for Identifying and Counting the Improperly Nourished," The American Journal of Clinical Nutrition, 35 (1982), p. 1242.

[4]C.M. Fletcher and F. D. Oldham, "Diagnosis in Group Research," in Medical Surveys and Clinical Trials: Some Methods and Applications of Group Research in Medicine, ed. L. J. Witts, C.B.E. (London: Oxford University Press, 1959), p. 35.

test of perfect specificity (V=1) will lead to an underestimate of prevalence by a factor equivalent to the sensitivity of the test. However, if prevalence (P) is less than .5 and the test is not "perfectly specific" (V<1), the multiplier "(1-P)", will be larger than the multiplier "P" and the effect of imperfect specificity will dominate the equation. (Note, in all that follows, that we assume our test is at least as good as a random assignment; that is, both sensitivity and specificity are greater than .5.) In most typical situations in the nutrition field, we overestimate prevalence as a result of imperfect sensitivity and specificity. For example, in a population with a prevalence rate of 30% (.3) and a test with specificity of .8 and sensitivity of .9, we would estimate prevalence as (.3x.9 + (1-.3)x(1-.8)) = .41 or 41%. (In this example specificity has been set lower than sensitivity since our concern is with screening for program entry. Had surveillance been our concern, an appropriate example would set specificity higher than sensitivity.) In the special case where (1-V)/(1-U) = P/(1-P), our estimate of prevalence would be correct (unbiased).

Using subscripts to denote the estimates of prevalence and sensitivity and specificity at particular points in time (for example, U_0 = sensitivity at time t_0), we can express the magnitude of error in estimating change in prevalence in a single equation.

$$(6)\quad \hat{p}_1-\hat{p}_0 = P_1 x U_1+(1-P_1)x(1-V_1) - (P_0 x U_0+(1-P_0)x(1-V_0))$$
$$= P_1 x(U_1+V_1-1) - P_0 x(U_0+V_0-1) - (V_1-V_0)$$

In this equation, we allow sensitivity and specificity to change at different points in time as well as to be different from each other. When so many elements of equation (6) are free to vary, situations exist where the

observed change, $\hat{p}_1-\hat{p}_0$, is greater than, equal to, and less than the true change, P_1-P_0. We will return to this general equation but, first, we can see that under some simplifying, but reasonable, assumptions underreporting rather than overreporting of change is most likely to occur.

Case 1: $U_1 = U_0 = V_1 = V_0$

Consider the simplest case of all, the case where sensitivity and specificity are equal to each other and remain constant over time. Using U to denote that single value, equation (6) reduces to:

$$(7)\quad \hat{p}_1 - \hat{p}_0 = P_1 x(2U-1) - P_0 x(2U-1) = (P_1 - P_0)x(2U-1)$$

If we classify individuals with a scheme somewhat better than a random assignment (the joint value of sensitivity and specificity is between .5 and 1), we observe a change in prevalence less than the true change by a factor of 2U-1. (If sensitivity and specificity are .8, we observe only (1.6 - 1) = .6, or 60%, of the true change in prevalence.

Case 2: $U_1 = U_0$ and $V_1 = V_0$.

A less restricted case is one where sensitivity and specificity differ from each other but remain constant over time. Once again, we can rewrite equation (6) in simplified form.

$$(8)\quad \hat{p}_1 - \hat{p}_0 = P_1 x(U+V-1) - P_0 x(U+V-1) = (P_1-P_0)x(U+V-1)$$

Again, if both sensitivity and specificity are between .5 and 1, we observe a change in prevalence less than the true

change. In this case, the correction factor is (U+V-1). A sensitivity of .9 and a specificity of .8 cause the observed change to be only .7 (70%) of the true change.

Case 3: $U_1 \neq U_0 \neq V_1 \neq V_0$.

This is the general case treated in equation (6). As noted, in this instance, we might not understate the change in prevalence--in fact, we might overstate it or even observe a change in the opposite direction from the actual change. Because the understatement of change results from the bias in the individual estimates of point prevalence, we can anticipate that specificity changes play the major role in determining the degree of misstatement in this case. This is best illustrated by examples.

Let $P_1 = .3$ and $P_0 = .4$.
Then $P_1 - P_0 = -.1$.
(Prevalence drops by 10 percentage points.)

A) Let $U_0 = .8$ and $V_0 = .85$
$U_1 = .9$ and $V_1 = .75$
Then, using equation (6):
$\hat{P}_1 - \hat{P}_0 = .3x(.9+.75-1) - .4x(.8+.85-1) - (.75-.85)$
$= .035$

B) Let $U_0 = .8$ and $V_0 = .75$
$U_1 = .9$ and $V_1 = .85$
Then, using equation (6):
$\hat{P}_1 - \hat{P}_0 = .3x(.9+.85-1) - .4x(.8+.75-1) - (.85-.75)$
$= -.095$

C) Let $U_0 = .9$ and $V_0 = .75$
$U_1 = .8$ and $V_1 = .85$
Then, using equation (6):
$\hat{P}_1 - \hat{P}_0 = .3x(.8+.85-1) - .4x(.9+.75-1) - (.85-.75)$
$= -.165$

When specificity declines, in "A", we observe an increase in the rate of malnutrition when, in truth, the rate declines. When specificity improves, in "B", we underestimate the change in the rate of malnutrition. Finally, in "C" we allow sensitivity to decline while specificity increases and we overestimate the change in

malnutrition rate.[5] These results depend on the magnitude of change in sensitivity and specificity, not only on the direction of change. (If we were to change U_1 to .85 in "B", then $\hat{p}_1-\hat{p}_0 = -1.025$.) Therefore, we could keep the direction of change in sensitivity and specificity the same and still alter our perception of prevalence change.

IV TESTING STATISTICALLY FOR CHANGE IN PREVALENCE

In most quantitative evaluations of nutrition programs, it is not sufficient to note the direction of a change in nutritional status and the magnitude of that change; it is equally important to demonstrate the statistical significance of that observed change. (Statistical significance is affirmed if the probability, or chance, of observing a particular outcome due to random factors is suitably small.) Misclassification might well change the results of hypothesis tests regarding changes in prevalence.

In the instance, case 2 above, the case where sensitivity and specificity are different from each other but remain constant over time or are the same in separate populations, the effects of misclassification on hypothesis testing have been well known for over thirty years. Irwin Bross writes:

> An especially important practical situation is the one where the same classification system is used in both samples. It will be shown that for

---

[5]The reader should be aware that these few examples may oversimplify the varied behavior of the extimate of the difference in prevalence as a result of changing sensitivity and specificity. In each of our examples, both sensitivity and specificity are allowed to change. When this is the case, estimation of the change in prevalence is not only conditional on the direction of change in sensitivity and specificity, but on the magnitude of their change as well. We can say, however, that when sensitivity is constant and specificity increases, the estimate of the change in prevalence will increase. And when specificity is constant and sensitivity increases, the estimate of the change in prevalence will decrease.

> this situation, the significance test remains valid even though the misclassification is ignored. The comfort that the research worker can draw from this result is somewhat mitigated by the fact that the significance test, though valid, becomes less powerful.[6]

The proofs of this contention are rather complicated and are beyond the scope of this paper. In effect, Bross' findings are consistent with intuition. In our Case 2, there is always underreporting of the change in prevalence; therefore, a larger sample is needed to avoid failure in identifying significant change in prevalence when it, in fact, is present. (In the jargon used by Bross, a larger sample is needed to make the test more powerful.) However, since the change is always in the same direction as the actual change, our conclusion regarding its direction will be the same.

Unfortunately, in the more general case, our Case 3, the effects of misclassification on hypothesis testing are as variable as the effects on the magnitude and direction of observed changes in prevalence. This more general case is discussed by Goldberg, who concludes that misclassification distorts the conclusions drawn from a test when the error rates ($1-U_i$ and $1-V_i$) are unknown but are known to be different in the two samples.[7]

V A DISCUSSION OF SOME UNDERLYING ASSUMPTIONS

Two tacit assumptions mark the mathematical presentation thus far:

---

[6] Irwin Bross, "Misclassification In 2 X 2 Tables," Biometrics, 10 (December, 1954), p. 483.

[7] Judith D. Goldberg, "The Effects of Misclassification on the Bias in the Difference Between Two Proportions and the Relative Odds in the Fourfold Table," Journal of the American Statistical Association, 70 (September, 1975), p. 562.

a) all healthy individuals (or diseased individuals) have an equal probability, called specificity (or sensitivity) of being classified correctly, and

b) disease is a discrete phenomenon in that an individual is either sick or healthy.

With regard to malnutrition and, in fact, most diseases, and other real-world phenomena, neither assumption holds.

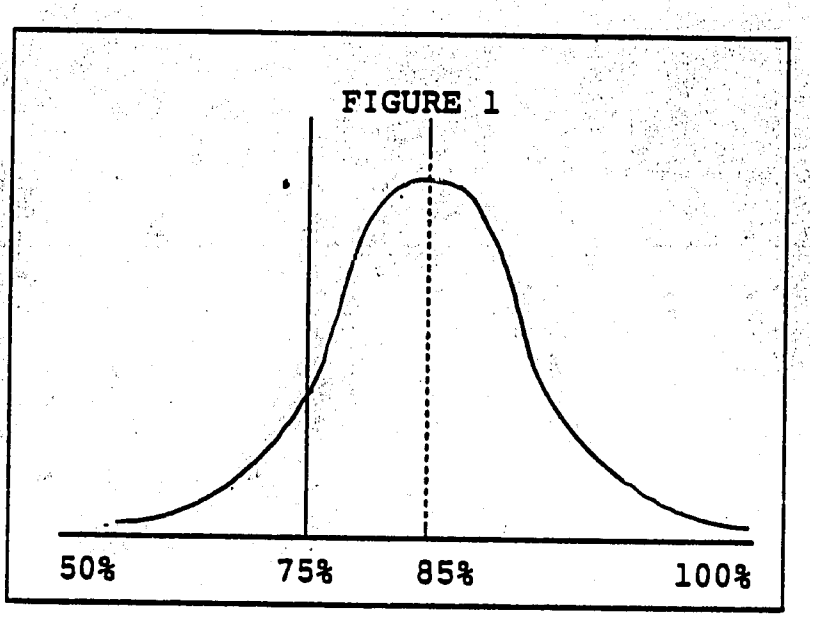
First, consider the implications of the first assumption regarding equal probability of correct classification. The most frequently used tests for malnutrition, tests based on anthropometric measurements, are derived from underlying continuous variables. (Continuous variables take on the full range of values on an interval scale.) The classification leading to a differentiation of nutritional status by category is predicated on the selection of ranges in those continuous variables corresponding to classes of health. Intuitively, it is more likely to misclassify individuals whose scores fall near the class boundaries than those whose scores fall clearly in the middle of the classes.

Consider the simplest case, a two-way classification of the nutritional status of preschool children into categories of "normal" and "malnourished" using the weight-for-age score. (The weight-for-age score is the ratio of the weight of a given child to a standard weight for children of the age of that given child. This ratio is often expressed as a percentage.) Traditionally, a cut-point of 75% of the standard is selected as the boundary differentiating between normal and malnourished individuals. (The 75% cut-point corresponds to the dividing line between Grade I and Grade II malnutrition in the near universally applied Gomez classification.[8]) If we assume that the weight-for-age

---

[8]Frederico Gomez, Rafael Ramos Galvan, Joaquin Cravioto and Silvestre Frenk, "Malnutrition in Infancy and Childhood, with Special Reference to Kwashiorkor," Advances in Pediatrics, 7 (1955).

scores in a given population are distributed normally, we can represent the classification process pictorially, as in Figure 1.



In this figure, all children whose scores fall below the cut-point of 75%, the solid vertical line, are classified as malnourished. Note, the prevalence of malnutrition is the area under the curve to the left of the cut-point divided by the total area under the curve. If the distribution is truly normal, with known mean and standard deviation, the prevalence is easily determined using tables of the standard normal distribution. If the mean and standard deviation of the curve in Figure 1 were 85% and 12% respectively, the prevalence would be approximately 20%.

Given the nature of errors leading to misclassification, discussed more fully in the next section, we would expect to make classification errors on those children whose scores fall near the cut-point. For example, using the NCHS-CDC standard,' the "normal" weight for an 18 month old boy is 11.5 kilograms. Seventy-five percent of

---

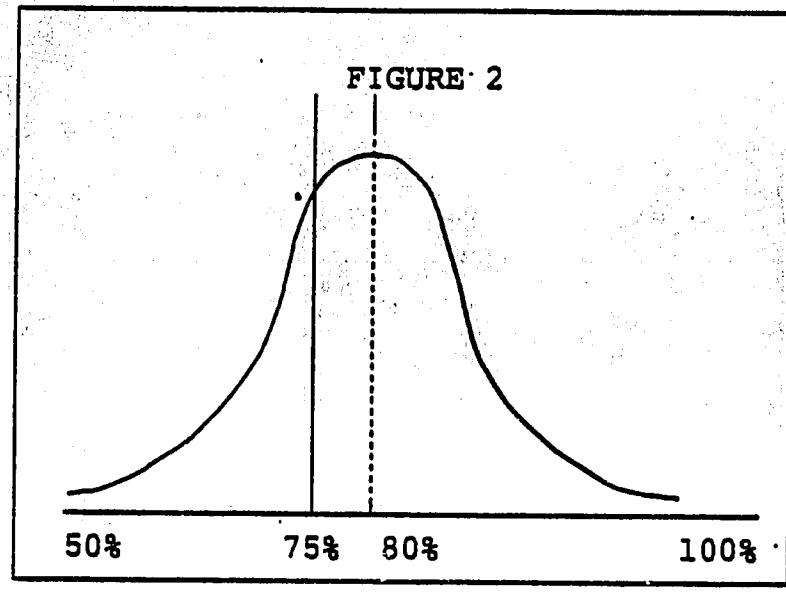
'Peter V. Hamill, Terence A. Drizd, Clifford L. Johnson, Robert B. Reed and Alex F. Roche, "NCHS Growth Charts, 1976," Monthly Vital Statistics Report, Vol. 25, No. 3 (1976).

that weight is 8.625 kilograms. One source of error frequently encountered in the field is the failure to "zero" the scale correctly. If such an error generated weights one-half kilogram less than actual weights, all 18 month old boys truly weighing between 8.625 and 9.125 would be misclassified. Boys with weights further from the cut-point weight, over 79.3% of standard, would still be classified correctly.

Clearly, the nearer one's score to the cut-point, the less error is needed to result in a misclassification, that is, the higher the probability of misclassifcation. One consequence of the higher probability to misclassify individuals whose scores fall near the cut-point is that the sensitivity and specificity of a test are related to the underlying distribution of the scores in <u>that given context</u>. If the scores cluster near the cut-point, more individuals will be misclassified and specificity and sensitivity will go down.

We noted earlier that many authors state that sensitivity and specificity are independent of the prevalence of the disease in the population under consideration. In the case under discussion, where a unique probability of misclassification can be assigned to each score, this no longer holds. For example, if the scores are normally distributed with known mean and standard deviation, the absolute number of cases falling near the cut-point varies with the mean of the distribution--as does the prevalence. Again, referring to the weight-for-age score, consider Figure 2. It is identical to Figure 1, except that the mean of the distribution is 80%. The prevalence of the disease is higher and the percentage of cases near, say within 3 percentage points of, the cut-point is also higher. (With mean 80% and standard deviation 12%, prevalence is 34%. More importantly, 18.1% of all scores fall within 3 percentage points of the cut-point while only 14.1% of all

scores fell within 3 percentage points of the cut-point in Figure 1, where the mean of the distribution was 85%.)



This suggests that we might define a Case 4 in which the probability of misclassifying any particular individual is a function of his/her score on the continuous variable underlying the classification scheme. A very general formulation of such a model is presented by Nissen-Meyer in which sensitivity and specificity are defined as the integrals over density functions describing the probability of a positive outcome in a single test for a given person under unique conditions. A distinctive characteristic of this model is that separate and distinct density functions apply to the "truly" healthy and the "truly" sick sub-populations.[10] As regards nutrition, the nature of those density functions is unknown and, more important, a procedure would have to be derived for estimating those functions in every possible setting for an evaluation or research study.

Consider now the implications of the second assumption of disease being a discrete phenomenon. Diseases,

---

[10]Sven Nissen-Meyer, "Evaluation of Screening Tests in Medical Diagnosis, Biometrics, (December, 1964), p. 733.

malnutrition included, are generally continuous phenomena. Individuals suffer from malnutrition to a varying degree; therefore, the classification into two catgeories is somewhat arbitrary and, perhaps, misleading. In an early work, J. Neyman suggests that more than two categories of disease be considered in partitioning populations. His work is presented in terms of three categories, the middle one being a moderately affected group. He points out:

> . . . we may expect . . . that the category of "moderately affected" splits itself into a continuous gradation of the intensity of illness, from very slight to very heavy, with the probabilities of the illness being detected increasing gradually from $1-\pi$ [π = the probability of correct diagnosis for entirely healthy individuals] to unity.[11]

A mathematical model that considers both the underlying continuity of test scores and the disease under scrutiny might look very similar to the general one presented by Nissen-Meyer. However, we would need a density function for the probability of positive outcome for each class specified or, if we treat the disease as truly continuous (rather than one defined by a multiple categorization), we would need a joint density function for an individual with disease of a given severity and a given score.

VI SOURCES OF ERROR LEADING TO MISCLASSIFICATION

We have shown that misclassification with respect to nutritional status can lead to bias in prevalence estimates which, in turn, generates errors in estimates of change in prevalence over time. The magnitude of the error in estimating change is dependent on the rate of misclassification. Misclassification arises from many sources. To gain an appreciation for the magnitude of

---

[11] J. Neyman, "Outline of Statistical Treatment of the Problem of Diagnosis," Public Health Reports, 62 (1947), p. 1450.

misclassification in nutrition evaluation studies, we need to consider these many sources in detail.

In our INTRODUCTION, we alluded to two general causes of misclassification: a lack of reliability of the measurement procedures and a lack of validity of the measures themselves. (In discussing error in evaluation studies, a third general cause of error is equally relevant: sampling error, which also creates bias. However, this error is not one of misclassification; therefore, we do not dwell on it here.) Table 2 lists seven specific sources of error. The first two refer primarily (though not entirely) to issues concerning validity; the remaining five refer primarily to issues concerning reliability.

A. Validity Issues

In general, nutritional status can be measured by a number of scores constructed from direct measurements of selected characteristics of individuals. These scores are often called indicators. Typically, in field situations, indicators are computed from anthropometric measurements. Each of these indicators captures a different dimension of nutritional deficiency. For example, "height-for-age" is regarded as an indicator of historical nutritional deprivation (chronic malnutrition or stunting) while "weight-for-height" is an indicator of current deficiencies (acute malnutrition or wasting). The most frequently used anthropometric indicator, "weight-for-age", is a composite of the two. Unfortunately, for persons concerned with the validity of a classification scheme for nutritional status, there is no single known indicator, no matter how difficult or expensive a procedure of measurement is invoked, which defines the rather abstract concept of nutritional status.

Thus, it can and has been argued that the selection of a particular indicator should be context specific--it should reflect the particular objectives and methodologies of the

TABLE 2

SOURCES OF ERROR IN CLASSIFYING NUTRITIONAL STATUS

A. Validity Related Errors
1. Imperfect correspondence between measures (indicators) and the underlying phenomenon of malnutrition.
2. Selection of inappropriate standards and/or cut-points in defining the classification.
B. Reliability Related Errors
1. Measurements (e.g. weights, heights, ages) are incorrect. a. The measurement device is biased. b. Variation exists between and within subjects. c. Variation exists between and within observers. d. Estimates (especially age) are inaccurate.
2. Recording of measurements is incorrect.
3. Transcription errors are made (e.g. weights recorded on growth charts are transcribed incorrectly to some data form used in analysis.)
4. Coding or keying errors or both are made (e.g. classifications based on raw measurements are done wrong, or, in computer analysis, errors are made in data entry.
5. Errors are introduced during analysis.

intervention under consideration. Also, the selection should be based on the use to be made of the indicators--diagnosis, prediction or evaluation.[12] Regarding misclassification and the consequences of that misclassification on comparative analyses, it is crucial

[12]Jean-Pierre Habicht, Charles Yarbrough, and Reynaldo Martorell, "Anthropometric Field Methods: Criteria for Selection", in Human Nutrition--A Comprehensive Treatise, Vol 2, eds. Derrick B. Jelliffe and E.F. Jelliffe (Plenum Publishing Corporation, 1979).

that the imperfect correspondence of each and every indicator to the underlying phenomenon of malnutrition be acknowledged and accepted. For our problem, accounting for such imperfect correspondence in quantitative analyses, the difficulty remains--there is no absolute measure of malnutrition to which anthropometric and/or other indicators can be compared. At this time we have no quantitative procedure to generate numerical estimates of the components of sensitivity and specificity reflecting lack of validity.

However, we can certainly argue that such components exist and can estimate likely bounds for them. For example, given the way in which classifications are generated by comparisons of observed ratios to standards, we know that even in the presumed "healthy" populations used to generate the standards, some individuals fall beyond the chosen critical score for defining malnutrition. Thus, at least some of the individuals classified as malnourished in a field setting can be presumed healthy. For example, seventy-five percent of the NCHS-CDC weight-for-age standard falls approximately 2.5 standard deviations from the median. Therefore, .625 percent of all children (6 per thousand) in a healthy environment will be malnourished according to the Gomez classification which defines second and third degree malnutrition as below 75% of the reference standard. In a study where 1000 children are weighed and 300 (30%) are said to be malnourished, the specificity of our test can be no better than 694/700 or 99%.

In our view, it would be extraordinary if any anthropometric test achieved such a high sensitivity--even if it were perfectly reliable. Issue 2 in our table, the standard and cut-point chosen to define the classification, continues to surface in the nutrition literature. In brief, some advocate the redefinition of both standards and cut-points in local contexts; others claim that the international standards can and should be applied to all.

The fact that such a debate persists is indicative of the imperfect validity of the use of anthropometric classification. Nonetheless, in field settings and in more carefully controlled evaluation studies, anthropometrics remain the most viable and, perhaps, most valid means of arriving at a classification.

B. Reliability Issues

1. Incorrect Measurements

Anyone who has observed the typical nutrition intervention in progress has first-hand knowledge of how difficult it is to take accurate measurements in the field. Scales break; scales are not zeroed out correctly; ages are unknown or deliberately misstated; children capable of standing slouch while being measured for height while pre-standers wriggle upon being stretched out. Moreover, measurements vary for individuals according to time of day or nearness to last meal. Finally, observers vary with regard to reading the scales, tightening an arm band, or stretching out a child for a height measurement.

Under relatively controlled circumstances, efforts have been made to estimate errors in the taking of anthropometric measurements. At INCAP, Martorell and his colleagues differentiated between field error and error in a more carefully controlled setting, an orphanage. They concluded that standard deviation of error for body weight, measured in kilograms in the field, is .29. However, even in the field portion of this study, outlying observations prompted remeasurement or elimination when they were "clearly greater than the maximum range of variability.[13] An

[13]Reynaldo Martorell, Jean-Pierre Habicht, Charles Yarbrough, Guillerma Guzman and Robert E. Klein, "The Identification and Evaluation of Measurement Variability in the Anthropometry of Preschool Children," American Journal of Physical Anthropology, 43 (1975), p. 348.

earlier study, by Malina, et al. showed far greater standard deviation of error in weight--1.17. Martorell speculates that this reflects greater between-observer variance (his own study had one observer) and that the interval between repeated measurements was longer.

The impact of a .5 or 1 kilogram deviation in weight readings on a classification will vary depending on the distribution of scores about the cut-points. In populations where weight-for-age scores cluster near the chosen cut-points, many children may be misclassified due to small errors in weight. However, in populations where scores are more uniformly distributed, errors in measuring weight will lead to fewer misclassifications.

To develop a sense of the magnitude of misclassification error due to weighing errors, a small simulation model was effectuated which generated normally distributed random errors in weights for a specified standard deviation of error. (A computer program listing of this model is shown as Appendix A.) The actual classification was then compared to the classification based on the simulated weights.[14] Three hundred and twenty-four children between the ages of 24 and 72 months of age were the subjects of the simulation. (These were all children in that age range weighed during the second semester of 1969 during an intervention in Candelaria, Colombia.) Nutritional grades were determined using the NCHS-CDC standard and a classification similar to Gomez's except for the lower bound for Normal (Grade III, less than 60%; Grade II, between 60% and 75%; Grade I, between 75% and 85%; Normal, over 85%). Table 2 contains the results for errors occurring with standard deviations of .5 and 1.0. (Slightly over 2/3 of all simulated erroneous weights fall within one standard deviation of the actual weight.)

---

[14] William D. Drake, Luis J. Fajardo, and Roy I. Miller, The Promotora Program in Candelaria: Volume IV. (Ann Arbor:

TABLE 3
MISCLASSIFICATION WITH RANDOM WEIGHING ERRORS

CASE A: STANDARD DEVIATION = .5 ACTUAL				
SIMULATED	NORMAL	GRADE I	GRADE II	GRADE III
NORMAL	173	16	0	0
GRADE I	18	69	4	0
GRADE II	0	7	35	1
GRADE III	0	0	0	1
CASE B: STANDARD DEVIATION = 1.0 ACTUAL				
SIMULATED	NORMAL	GRADE I	GRADE II	GRADE III
NORMAL	157	32	1	0
GRADE I	31	41	10	0
GRADE II	3	19	25	0
GRADE III	0	0	3	2

The entries in the tables correspond to the number of individuals classified according to the weight observed as indicated by the row but, who after adding random error were classified as indicated by the column. Thus, 19 Grade I children were classified as Grade II when error was randomly generated with a standard deviation of 1.0 kilogram.

We can compute estimates of sensitivity and specificity of erroneous readings relative to the actual readings by determining the probability of classifying normal and malnourished children correctly. In case A, sensitivity is 117/133 or .879 while specificity is 173/191 or .906. In case B, sensitivity is 100/133 or .75 while specificity is 157/191 or .821.

Community Systems Foundation, Project on Analysis of Community Level Nutrition Programs, November 1980).

2. Recording Is Incorrect

In field settings, record keeping is often haphazard or treated with disdain by field workers committed to achievement rather than reporting. Thus, errors are immortalized through incorrect recording of data. Quantification of the magnitude of the recording problem in a typical field setting is extremely difficult for, unless independent records are kept, there is no source of comparison for existing records.

There is some evidence that the accuracy of the data kept as part of ongoing programs is related to the type of use it receives. When data are used for local decision making, such as to determine treatment or coverage, accuracy is higher than when it is kept primarily for evaluation or program monitoring purposes. In a study conducted on the PRIMOPS Promotora program in Cali, Colombia[15] it was found that data used for local purposes improved with time while data not so used deteriorated in accuracy.

3. Transcription Errors

Records are often transcribed from one form to another. This is especially common when records for individuals are maintained in the field but summaries are generated for reporting or evaluation. Errors are often introduced during the transcription phase.

One common transcription problem arises when weight charts are maintained for individuals while summaries of community-level nutritional status are generated from the charts. A point is plotted on the chart and transcribed as a nutritional grade. In the data set from Candelaria, the nutritional grade of each child was determined in the field

---

[15]John D. Nystuen, The PRIMOPS Experience: Information Processing in the Design and Performance of a Health Care System: Volume V. (Ann Arbor: Community Systems Foundation, Project on Analysis of Community Level Nutrition Programs, November 1980).

by the Promotora using a weight chart. Though called the Candelaria standard, the standard used was equivalent to that used by Gomez in originating his classification. Using a computer, the nutritional grade of each child was recalculated. Table 4 compares the field work with the computer.

TABLE 4

FIELD VS. COMPUTER DETERMINED NUTRITIONAL GRADE COMPUTER					
FIELD	NORMAL	GRADE I	GRADE II	GRADE III	MISSING
NORMAL	7246	248	9	2	6
GRADE I	57	1548	113	0	5
GRADE II	4	16	411	6	0
GRADE III	1	0	4	49	0
MISSING	7	2	0	0	0

Relative to the computer, the test as applied by the field workers had a sensitivity of 2151/2408 or .894 and a specificity of 7246/7315 or .99. Of course, the errors may, in fact, be due to errors in keypunching or data processing and not be a reflection of transcription in the field.

4. Coding or Keying Errors

The use of computers in evaluation studies introduces a new source of error--error made while preparing data through coding for machine analysis and error in the data entry itself. In a study based on interviews combined with retrospective recording of data from weight charts, the data were keyed but not verified at the source.[16] (Verification is a process whereby one data entry operator reenters the work of another to "verify" that the first was correct).

[16] John O. Field, Roy I. Miller, and William D. Drake, Kottar: Malnutrition, Intervention and Development in a South Indian District. Volume III. (Ann Arbor: Community Systems Foundation, Project on Analysis of Community Level Nutrition Programs, January 1981).

Forty percent of the interviews had noticeable keypunch errors (fields out of line, etc.). These were correctable. No one knows how many other errors persisted despite the rigorous data cleaning procedures applied to the non-verified data.

5. Errors during Analysis

Finally, as analysts ourselves, we must acknowledge that our own procedures introduce errors. For example, with a computer, it is a relatively straightforward task to apply a classification to a set of "percentage of standard" scores. However, errors can still occur. Let us say our classification partitions the population at 60%, 75% and 85% of standard. A pertinent question is the assignment of those individuals whose scores are equal to those cut-points. Are children at precisely 85% of standard Grade I malnourished or normal? In the Candelaria data used to illustrate discrepancy between field computations and computer calculations, 29 children were classified normal with percent of standard scores exactly equal to 85%. Different computer operators, or even the same computer operator at different points in time, might conceivably assign those children whose scores fall precisely on the cut-point defining categories to either category. (We have made that mistake ourselves.)

VII CHANGES IN SENSITIVITY AND SPECIFICITY OVER TIME

To observe the nature of the fluctuation in reported results of intervention in cases where sensitivity and specificity vary over time, we can create a hypothetical example which incorporates all the possibilities. Suppose we observe an initial prevalence of 40.2 percent and a final prevalence of 35.5 percent; that is we observe a drop in prevalence of 4.7 percentage points. (The intervention appears to have had the desired effect.) If sensitivity (U) and specificity (V), both before and after the intervention,

were .8, we have a true drop in prevalence of 7.8 percentage points. That is, we underestimate the true change in prevalence. Line 1 of Table 5 summarizes this scenario.

In subsequent lines of the table, we allow first specificity, then sensitivity, and then both to deviate from the arbitrary constant value of .8. Lines 2,3 and 4 reflect improvements in specificity, sensitivity, and both, respectively. In lines 5 through 7, the same pattern is followed but the probabilities decline. Finally, in lines 8 and 9, they are manipulated in opposite directions.

Line 2 shows us that an improvement in specificity from .8 to .9 is sufficient to cause us to observe a drop in prevalence when, in truth, prevalence increased. However, when only sensitivity improves, as in line 3, we understate the true results by an even greater amount than in the case where both factors remain constant and equal. When both sensitivity and specificity improve, as in line 4, we observe a larger change in prevalence than the true change but at least in the same direction as that true change.

When specificity alone decreases, as in line 5, the observed drop in prevalence is far below the true decrease. But, when sensitivity alone decreases, as in line 6, we overstate the drop by a small amount. Line 7 demonstrates the large understatement of the decrease resulting when both sensitivity and specificity decline.

Line 8 illustrates the synergistic effect of a decline in sensitivity and a rise in specificity--an effect which leads to the reporting of change in the opposite direction from the true change. Line 9 shows the mutually supportive effect of a rise in sensitivity and a decline in specificity, the case marked by the largest underestimate of change in the entire table.

In considering this table, it is important to recall that the degree of under-, over- or misreporting of change

TABLE 5

ILLUSTRATION OF CONSEQUENCES OF CHANGING SENSITIVITY AND SPECIFICITY

	U_0	V_0	U_1	V_1	$\hat{P}_0$	$\hat{P}_1$	Observed Change in Prevalence	P_1	P_2	True Change	Result	Amount of Change
1	.8	.8	.8	.8	40.2	35.5	drop 4.7	33.7	25.8	drop 7.8	under	3.1
2	.8	.8	.8	.9	40.2	35.5	drop 4.7	33.7	36.4	rise 2.8	opposite	
3	.8	.8	.9	.8	40.2	35.5	drop 4.7	33.7	22.1	drop 11.5	under	6.8
4	.8	.8	.9	.9	40.2	35.5	drop 4.7	33.7	31.9	drop 1.8	over	2.9
5	.8	.8	.8	.7	40.3	35.5	down 4.7	33.7	11.0	drop 22.7	under	18.0
6	.8	.8	.7	.8	40.2	35.5	drop 4.7	33.7	31.0	drop 2.7	over	2.0
7	.8	.8	.7	.7	40.2	35.5	drop 4.7	33.7	13.8	drop 19.9	under	15.2
8	.8	.8	.7	.9	40.2	35.5	drop 4.7	33.7	42.5	rise 8.8	opposite	
9	.8	.8	.9	.7	40.2	35.5	drop 4.7	33.7	9.2	drop 24.5	under	19.8

Case 1: $U_1=U_0=V_1=V_0$ $\quad P_1-P_0 = \dfrac{\hat{P}_1-\hat{P}_0}{2U-1}$

Case 2: $U_1=U_0$ and $V_1=V_0$ $\quad P_1-P_0 = \dfrac{\hat{P}_1-\hat{P}_0}{U+V-1}$

Case 3: $U_1 \neq U_0 \neq V_1 \neq V_0$ $\quad P_1-P_0 = \dfrac{\hat{P}_1+V_1-1}{U_1+V_1-1} - \dfrac{\hat{P}_0+V_0-1}{U_0+V_0-1}$

in prevalence is a function not only of changes in sensitivity and specificity but also a function of the observed prevalence rates. Thus, the results in the table would differ if the magnitude of the observed prevalences were changed. The reader might find it instructive to recreate a similar table for another pair of observed prevalences.

Generalizations from Table 5 are difficult to make. It is readily apparent that an improvement in specificity leads to a situation where the observed drop in prevalence over time is greater than the actual drop. However, in most other cases, the observed drop in prevalence is less than the actual decrease--often by a staggering amount.

The question which must be confronted by the pragmatist is, "What are the typical trends for both sensitivity and specificity in field and/or research applications?" Unfortunately, we can offer conflicting scenarios; that is, there are logical reasons to believe that both of the key probabilities can move in either direction. We will present the logic of several scenarios here.

1) <u>Effective programs lead to improved sensitivities and specificities.</u> An effective program can be characterized as one in which the mean of the normal distribution of anthropometric scores for participants (for example, weight-for-age scores) increases. This moves many of the scores from below the cut-point used to define malnutrition to the other side of that point. If the distribution of scores is near normal, such a shift in the distribution of scores leaves fewer participants near the cut-point. Thus, errors resulting from mistakes in weighing or in erroneously computed ages will cause fewer scores to cross the cut-point. As a result, sensitivity and specificity should improve.

2) <u>Effective programs lead to changes in target population composition resulting in uncertain changes in sensitivity and specificity.</u> As field personnel perceive a decline in the number of seriously malnourished in a targeted population, they are likely to relax entry criteria (e.g., raise

the anthropometric cut-point or raise the age limitation), or expand the targeted population geographically.[17] Relaxing entry criteria by either example would probably lead to a decrease in sensitivity or specificity (other sources of error in field protocol held constant). On the other hand, geographic expansion of services would, in general, not produce predictable changes in sensitivity or specificity.

3) Any program with frequently scheduled activities devoted to taking measurements (for example, monthly weighings) may suffer a decline in sensitivity and specificity. The tedium of making most measurements under difficult field conditions is likely to lead to haste and carelessness in the making of those measurements. The effects of this tedium are compounded by the decline in the performance of the machinery used to make the measurements. Thus, data drawn from a series of frequent measurements are likely to deteriorate in quality with time. Sensitivity and specificity suffer accordingly.

4) A program which regularly uses measurements for local decision making may improve reliability with time. As exemplified by the PRIMOPS example mentioned on page 22 of this report, specificity and especially sensitivity may improve with time as the field staff find that the data are of use to them.

5) Skill levels of field workers increase over time resulting in an improvement in both sensitivity and specificity. Counter to the argument raised in 3, it is equally plausible that in selected circumstances, the capabilities of field workers in ongoing programs increase with experience. Thus, sensitivity and specificity should improve with time rather than deteriorate.

6) No noticeable change in sensitivity or specificity occurs over time. Little or no change in sensitivity and specificity can arise either because field and data-handling protocol has stabilized to a "steady state" condition or because different aspects described in the foregoing scenarios cancel each other out. Without solid evidence to the contrary perhaps the best assumption to make is this scenario.

---

[17]Although this is an example of selection bias which we said in the "Introduction" was of less interest to our study, we believe this particular example merits consideration at this time.

These scenarios illustrate a subset of the possible chains of events leading to changes in sensitivity and specificity over time. In any given analysis, be it for research or evaluation, it is essential that the analyst be aware of the conditions of his/her study that might alter the accuracy of classification. If nothing else, Table 5 illustrates how easily small changes in sensitivity and, especially, specificity can lead to the misstatement of the results of a study of a magnitude far greater than the expected result.

VIII IMPACT OF MISCLASSIFICATION ON EVALUATION OUTCOMES

We now turn to the question of what impact misclassification has upon the interpretation of nutrition program evaluations. In order to explore this question we retrieved the numerical results from eight different studies. A re-analysis of these results was undertaken for a variety of different sensitivity and specificity assumptions. The programs reanalyzed represent a broad cross-section of interventions in both Latin America and Asia. They include programs in Sri Lanka; Candelaria, Colombia; Tamil Nadu, India; Santarém, Brazil; Cali, Colombia; and Pesperi, Honduras. Because the number of different combinations of assumptions needing review is large, a computer program was written to assist in this re-analysis. (A listing of that computer program is shown as part 2 of Appendix B.) Tables 6 through 14 show a sample of the results of this re-analysis using what we believe to be very conservative (higher than anticipated) estimates for sensitivity and specificity. Tables 6 through 9 assume that both sensitivity and specificity remain constant over time and are equal to each other. (This is our case 1, page 7.) With a sensitivity (U) and specificity (V) equal to .9, only 80% of the true change in prevalence is reported for all the programs (Table 6). As U and V decrease, the percentage of true change which is reported diminishes so that by the time

U and V are each .75 only 50% of the actual results are shown by the analysis (Table 9).

In Section VI we attempt to summarize all the evidence which currently exists regarding the range over which sensitivity and specificity can vary. Because the basic field work has not yet been undertaken, these bounds are necessarily quite broad and the corresponding estimates imprecise. However, our judgement is that it is not at all unreasonable for U and V to be between .8 and .7 in real field situations. If we select the midpoint of that range, .75, then only 50% of the true results are reported under conditions of constant U and V (Table 9). Even a more conservative estimate of .8 yields only 60% of the true results (Table 8).

Tables 10 through 14 show the impact changes in sensitivity and specificity over time have on estimating changes in prevalence. The rationale for these changes is provided in the field level scenarios presented in section VII. Five percentage point variations were made around the more conservative estimate of .8 mentioned above. Table 10 holds specificity constant and allows sensitivity to increase from .8 to .85. The calculations indicate that greater underreporting of the true results occurred reported than if sensitivity had diminished over time, say from .8 to .75. In this case a higher percentage of the true results would be reported as shown in Table 11.

If both sensitivity and specificity were to diminish over time from say, .85 to .80, the percentage of the true results in our sample of interventions would diminish even more. This outcome, shown as Table 12 indicates that the average fraction of the true results drops to 49.9% and is comparable to a constant sensitivity and specificity of .75. Scenario three on page 29 describes one of the field conditions which could allow this to occur.

TABLE 6

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .90; Sensitivity after = .90; Specificity before = .90; Specificity after = .90)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	41.9	28.4	13.5 drop	2.7 under	80.0
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	32.0	23.4	8.6 drop	1.7 under	80.0
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	23.8	13.7	10.0 drop	2.0 under	80.0
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	26.1	23.9 drop	4.8 under	80.0
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	40.6	9.6 drop	1.9 under	80.0
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	47.5	28.8	18.7 drop	3.7 under	80.0
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	26.3	21.2	5.0 drop	1.0 under	80.0
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	23.6	16.4	7.2 drop	1.4 under	80.0
9	AVERAGE	39.5	29.9	9.6 drop	36.9	24.8	12.1 drop	2.5 under	79.3

TABLE 7

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .85; Sensitivity after = .85; Specificity before = .85; Specificity after = .85)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	40.7	25.3	15.4 drop	4.6 under	70.0
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	29.4	19.6	9.8 drop	3.0 under	70.0
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	20.0	8.6	11.4 drop	3.4 under	70.0
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	22.7	27.3 drop	8.2 under	70.0
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	39.3	11.0 drop	3.3 under	70.0
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	47.1	25.7	21.4 drop	6.4 under	70.0
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	22.9	17.1	5.7 drop	1.7 under	70.0
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	19.9	11.6	8.3 drop	2.5 under	70.0
9	AVERAGE	39.5	29.9	9.6 drop	35.0	21.2	13.8 drop	4.2 under	69.6

TABLE 8

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .80; Sensitivity after = .80; Specificity before = .80; Specificity after = .80)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	39.2	21.2	18.0 drop	7.2 under	60.0
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	26.0	14.5	11.5 drop	4.6 under	60.0
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	15.0	1.7	13.3 drop	5.3 under	60.0
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	18.2	31.8 drop	12.7 under	60.0
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	37.5	12.8 drop	5.1 under	60.0
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	46.7	21.7	25.0 drop	10.0 under	60.0
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	18.3	11.7	6.7 drop	2.7 under	60.0
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	14.8	5.2	9.7 drop	3.9 under	60.0
9	AVERAGE	39.5	29.9	9.6 drop	32.5	16.5	16.0 drop	6.4 under	60.0

TABLE 9

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .75; Sensitivity after = .75; Specificity before = .75; Specificity after = .75)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	37.0	15.4	21.6 drop	10.8 under	50.0
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	21.2	7.4	13.8 drop	6.9 under	50.0
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	unpredict-able	condition	-	-	50.0
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	11.8	38.2 drop	19.1 under	50.0
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.4	35.0	15.4 drop	7.7 under	50.0
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	46.0	16.0	30.0 drop	15.0 under	50.0
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	12.0	4.0	8.0 drop	4.0 under	50.0
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	unpredict-able	condition	-	-	50.0
9	AVERAGE	39.5	29.9	9.6 drop	36.1	14.9	21.2 drop	11.6 under	45.3

TABLE 10

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .80; Sensitivity after = .85; Specificity before = .80; Specificity after = .80)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	39.2	19.5	19.6 drop	8.8 under	55.1
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	26.0	13.4	12.6 drop	5.7 under	54.8
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	15.0	1.5	13.5 drop	5.5 under	59.3
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	16.8	33.2 drop	14.1 under	57.5
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	34.6	15.7 drop	8.0 under	49.0
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	46.7	20.0	26.7 drop	11.7 under	56.2
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	18.3	10.8	7.6 drop	3.6 under	52.6
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	14.8	4.8	10.1 drop	4.3 under	57.4
9	AVERAGE	39.5	29.9	9.6 drop	32.5	15.2	17.3 drop	7.7 under	55.5

TABLE 11

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .80; Sensitivity after = .75; Specificity before = .80; Specificity after = .80)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	39.2	23.1	16.1 drop	5.3 under	67.1
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	26.0	15.8	10.2 drop	3.3 under	67.6
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	15.0	1.8	13.2 drop	5.2 under	60.6
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	19.8	30.2 drop	11.1 under	63.2
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	40.9	9.2 drop	1.7 under	81.9
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	46.7	23.6	23.0 drop	8.0 under	65.2
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	18.3	12.7	5.6 drop	1.6 under	71.4
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	14.8	5.6	9.2 drop	3.4 under	63.0
9	AVERAGE	39.5	29.9	9.6 drop	32.5	17.9	14.6 drop	5.0 under	65.8

TABLE 12

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .85; Sensitivity after = .80; Specificity before = .85; Specificity after = .80)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	40.7	21.2	19.5 drop	8.7 under	55.4
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	29.4	14.5	14.9 drop	8.0 under	46.3
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	20.0	1.7	18.3 drop	10.3 under	43.7
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	18.2	31.8 drop	12.7 under	60.1
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	37.5	12.8 drop	5.1 under	60.2
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	47.1	21.7	25.5 drop	10.5 under	58.8
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	22.9	11.7	11.2 drop	7.2 under	35.7
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	19.9	5.2	14.7 drop	8.9 under	39.5
9	AVERAGE	39.5	29.9	9.6 drop	35.0	16.5	18.5 drop	8.9 under	51.9

TABLE 13

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .80; Sensitivity after = .80; Specificity before = .80; Specificity after = .85)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	39.2	27.2	11.9 drop	1.1 under	90.8
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	26.0	21.1	4.9 drop	2.0 over	140.8
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	15.0	9.2	5.8 drop	2.2 over	137.9
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	24.5	25.5 drop	6.4 under	74.9
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	42.3	8.0 drop	0.3 under	96.3
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	46.7	27.7	19.0 drop	4.0 under	78.9
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	18.3	18.5	0.1 rise	4.1 opposite	N/A
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	14.8	12.5	2.4 drop	3.4 over	241.7
9	AVERAGE	39.5	29.9	9.6 drop	32.5	22.9	9.6 drop	0.0	100.00

TABLE 14

COMPARISON BETWEEN OBSERVED AND ACTUAL MALNOURISHMENT--PREVALENCE RATES EXPRESSED IN PERCENTAGES
(Assuming Sensitivity before = .80; Sensitivity after = .85; Specificity before = .80; Specificity after = .85)

		a	b	c	d	e	f	g	h
	Name & Location of Program	Observed Prevalence Before Inter-vention	Observed Prevalence After Inter-vention	Observed Change in Prevalence	Actual Prevalence Before Inter-vention	Actual Prevalence After Inter-vention	Actual Change in Prevalence	Difference Between Observed and Actual Prevalence Change	Observed Change Expressed as Percentage of Actual Change
				(Col. b-a)			(Col. e-d)	(Col. f-c)	(Col.c/f)x100
1	Sri Lanka, 1982 MCH (children 1 to 5 years)	43.5	32.7	10.8 drop	39.2	25.3	13.9 drop	3.1 under	77.7
2	Sri Lanka, 1982 MCH (all children under 5)	35.6	28.7	6.9 drop	26.0	19.6	6.4 drop	0.5 over	170.8
3	Candelaria, Colombia (children under 6)	29.0	21.0	8.0 drop	15.0	8.6	6.4 drop	1.6 over	125.0
4	Candelaria, Colombia (Family planning-pregnancy rate)	50.0	30.9	19.1 drop	50.0	22.7	27.3 drop	8.2 under	70.0
5	Kottar Social Services Tamil Nadu, India (children under 5)	50.2	42.5	7.7 drop	50.3	39.3	11.0 drop	3.3 under	70.0
6	Esperança MCH Santarem, Brazil (children under 5)	48.0	33.0	15.0 drop	46.7	25.7	21.0 drop	6.0 under	71.4
7	PRIMOPS--Cali, Colombia (children under 6)	31.0	27.0	4.0 drop	18.3	17.1	1.2 drop	2.8 over	333.3
8	Community Development Pespire, Honduras (children under 5)	28.9	23.1	5.8 drop	14.8	11.6	3.3 drop	2.5 over	175.8
9	AVERAGE	39.5	29.9	9.6 drop	32.5	21.2	11.3 drop	1.7 under	85.0

In an earlier section we mentioned that change in specificity over time had an extremely volatile effect upon outcome. Tables 13 and 14 portray this phenomenon for our sample of nutrition programs. If sensitivity remains constant at .8 and specificity increases to .85, over-reporting occurs in three of our eight cases and in one instance the true direction is opposite from that which was reported (Table 13). If both sensitivity and specificity move from .8 to .85 over time, four of the eight cases experience overreporting, as shown in Table 14. Scenarios 1, 4, and 5, described on pages 28 and 29, provide field conditions under which this can occur.

In summary then, what can we say about the impact of misclassification due to imperfect sensitivity and specificity? First, without specific knowledge leading us to conclude there was a change, the most reasonable assumption is that sensitivity and specificity remain constant over time. (Equally plausible scenarios exist for no change, an increase or a decrease.) Given a constant sensitivity and specificity, there is always an under-statement of the true results.

Second, while the basic field work necessary to identify the magnitude of the values for both sensitivity and specificity remains to be done, a reasonable conservative estimate for the data gathered from typical field-level nutrition programs is between .70 and .80. (We have anecdotal evidence of much lower values.) Taking the midpoint of this range (.75) and assuming constancy over time, only half of the true change in prevalence is being reported by the typical program analysis. It is, however, unlikely for specificity to be as low as .75 when the prevalence of malnutrition in the population is relatively low.

Third, if there is evidence that changes in sensitivity and/or specificity do occur over time, it is extremely

important to attempt to estimate them--especially for specificity. While underreporting is most likely under conditions of change experienced in the field, it is possible to have overreporting.

Finally and perhaps most important, if comparisons are made among nutrition programs, using nutritional status as a comparative indicator, sensitivity and specificity estimates also must be considered. A program which has good data gathering and handling protocol will show more of the true change than another program which devotes less energy to this aspect of the program, simply because the sensitivity and specificity are higher. While it could be argued that good field protocol is one measure of a good nutrition program it is also possible that too heavy an emphasis was placed upon measurement in comparison to other program elements such as education, and preventative and curative health care. Research-oriented interventions may compare more favorably than they should with action-oriented programs for the same reason.

The logic outlined in the above paragraph is of course also applicable at a far more general level. Comparisons are made not only among nutrition programs but also between nutrition programs and interventions in other sectors such as agriculture, water, transportation, education and housing. If the program under comparison happens to have measures of effectiveness which are not subject to much misclassification, the program will "fare better" in any evaluation than one with difficult-to-measure indicators. We know that nutritional status is an extremely difficult phenomenon to measure and, for all the reasons given earlier in this paper, is subject to considerable misclassification. This fact may be an important reason why nutritional program evaluations worldwide have not yet provided much analytic evidence of nutrition improvement, especially when compared with other public-sector interventions.

IX GUIDELINES FOR CORRECTING EVALUATION RESULTS

One of the reasons that the implications of misclassification have not been widely recognized by both researchers and policy makers may be that the concepts, while simple on the surface, are subtle. They are also easily confused with sampling biases and other distortions arising from data gathering and analysis protocol. Furthermore, as demonstrated in the foregoing sections, the basic field-level knowledge about sensitivity and specificity values under varying field conditions does not yet exist. Nevertheless, since the effects of misclassification are so large, we will attempt to formulate a strategy and set of guidelines for interpreting results of nutritional studies, given our current limited state of knowledge.

The basic strategy proposed for adjusting these results is to attempt to estimate the magnitude of misclassification arising from the imperfect sensitivity and specificity found under field conditions and then to apply the appropriate correction formula. The appropriate correction formulas are presented in section III, pages 7 and 8 of this report for each of the basic conditions which can be experienced. Equation (6) is the general case in which sensitivity (U) and specificity (V) both vary from each other and over time. Equation (7) is the case under which U and V are assumed to be equal and constant with time, and equation (8) if for differing U and V which are constant with time.

Several different approaches were tried for presenting findings. Charts and graphs were constructed which we thought might help the researcher or policy analyst apply the formulas. However, it was concluded that the best and easiest way to proceed was simply to apply the formula to the particular case at hand. Unfortunately, if one wishes to see what effect variations in sensitivity and specificity have upon the outcome, the large number of computations

become somewhat laborious. It was therefore decided to write a computer program in a programming language available to virtually all computers (including personal computers) which would carry out these calculations and present results in a convenient tableau similar to that shown in Tables 6 through 14. Furthermore, in order to familiarize the user with the underlying concepts, a tutorial program was also written. A listing of these two computer programs is presented in Appendix B. They are written in Basic language for an Apple IIe computer and are capable of operation on virtually any other computer with some modification.

It was also felt that it might be helpful to suggest a set of specific guidelines or steps which could be followed in adjusting results obtained in an impact analysis. For the purposes of these guidelines it is presumed that there is a control population or "before intervention" population denoted by subscript 0 against which a comparison is being made. The guidelines follow.

1. Select an approximation of values for sensitivity and specificity. Table 2 on page 18 of this report presents the sources of error in classification of nutritional status. Assuming no bias in the measuring device, a reasonable selection of cut-points, no errors introduced during analysis, and field conditions found in typical operating nutrition programs, we roughly estimate sensitivity and specificity to take on values of between .7 and .8 or a median value of .75. If conditions are either better or worse than described above, this estimate could be adjusted accordingly.

2. Determine if there is clear evidence pointing to a change in either specificity or sensitivity between t_0 and t_1. The logic of the scenarios which could create changes in sensitivity and specificity is presented on pages 28 and 29. Since the logic for more than one scenario is often applicable to a single field condition we suggest that variations from the assumption of constancy be based upon fairly solid evidence to the contrary. This is especially important because (1) plausible scenarios are often in conflict with each other and (2) the adjustments

to results arising from this variation can be substantial.

3. Using the best estimate for sensitivity and specificity for both t_0 and t_1, use the appropriate correction formula to compute the adjustments to prevalence rates. These formulas are shown on pages 7 and 8 of this report. Since these are only rough approximations, it may be useful to vary estimates of U and V to see the magnitude of the impact upon results. If a computer is handy, especially an Apple IIe, it may be worthwhile using the programs listed in Appendix B to facilitate this sensitivity analysis.

4. Reassess your estimates of sensitivity and specificity. It may be that the estimates which were selected for U and V have such a large impact on outcome that the adjustments yield impossible results. This can mean that either the estimates of misclassification, (U-1) and (V-1), are too high or the point estimates of prevalence are biased and therefore incorrect for reasons unrelated to misclassification. In any event, it may be helpful to cycle back with analysis results in hand to check the assumptions and logic employed in the re-analysis.

5. Report both unadjusted changes in prevalence and corrected changes in prevalence. Because the values attributed to sensitivity and specificity are approximations based on good judgment, a researcher's presentation should value both unadjusted and corrected changes in prevalence. Confidence in corrected results can grow only from field studies of classification errors.

REFERENCES

Bross, Irwin. "Misclassification In 2 X 2 Tables." Biometrics, Vol. 10, December 1954, pp. 478-486. (Note by D. J. Newell, Biometrics, March 1963, pp. 187-188).

Buell, Philip and John E. Dunn, Jr. "The Dilution Effect of Misclassification." American Journal of Public Health, Vol 54, No. 4, 1964, pp. 598-602. 9 references. Appendix.

Cochrane, A. L. and W. W. Holland. "Validation of Screening Procedures." British Medical Bulletin, Vol. 27, No. 1, 1971, pp. 3-8. 37 references.

Copeland, Karen T., et al. "Bias Due to Misclassification in The Estimation of Relative Risk." American Journal of Epidemiology, Vol. 105, No.5, 1977, pp.488-495. 20 references.

Dawid, A P. and A. M. Skene. "Maximum Likelihood Estimation of Observer Error-rates Using the EM Algorithm." Applied Statistics, Vol. 28, No. 1, 1979, pp. 20-28. 7 references.

Diamond, Earl L. and Abraham M. Lilienfeld. "Misclassification Errors in 2 x 2 Tables with One Margin Fixed: Some Further Comments." American Journal of Public Health, Vol. 52, No. 12, 1962, pp. 2106-2110. 3 references.

Diamond, Earl L. and Abraham M. Lilienfeld. "Effects of Errors in Classification and Diagnosis in Various Types of Epidemiological Studies." American Journal of Public Health, Vol.52, No. 7, 1962, pp. 1137-1144. 8 references.

Drake, William D., Luis J. Fajardo, and Roy I. Miller. The Promotora Program in Candelaria: Volume IV. (Ann Arbor: Community Systems Foundation, Project on Analysis of Community Level Nutrition Programs, November 1980).

Drake, William D., Roy I. Miller, and Margaret Humphrey, Final Report: Analysis of Community-Level Nutrition Programs. Volume I. (Ann Arbor: Community Systems Foundation, October 1980), pp. 89-94.

Erdreich, Linda S. and Elisa T. Lee. "Use Of Relative Operating Characteristic Analysis In Epidemiology." American Journal of Epidemiology, Vol. 114, No. 5, 1981, pp.649-662. 24 references.

Field, John O., Roy I. Miller, and William D. Drake, Kottar: Malnutrition, Intervention and Development in a South Indian District. Volume III. (Ann Arbor: Community Systems Foundation, Project on Analysis of Community Level Nutrition Programs, January 1981).

Fleiss, Joseph L. Statistical Methods for Rates and Proportions. John Wiley and Sons. New York, 1973.

Fletcher, C. M. and P. D. Oldham. "Diagnosis in Group Research." Medical Surveys and Clinical Trials: Some Methods and Applications of Group Research in Medicine, L. J. Witts ed. London: Oxford University Press, 1959. pp. 23-38. 31 references.

Gart, John J. "Point and Interval Estimation of the Common Odds Ratio in the Combination of 2 x 2 Tables with Fixed Marginals." Biometrika, Vol. 57, No. 3, 1970, pp. 471-475. 18 references.

Gavan, James A. "The Consistency of Anthropometric Measurements." American Journal of Physical Anthropology, Vol. 8, 1950, pp. 417-426. 5 references.

Gladen, Beth and Walter J. Rogan. "Misclassification and the Design Of Environmental Studies." American Journal of Epidemiology, Vol. 109, No. 5, 1979, pp. 607-616. 18 references.

Goldberg, Judith D. "The Effects of Misclassification on the Bias in the Difference Between Two Proportions and the Relative Odds in the Fourfold Table." Journal of the American Statistical Association, Vol. 70, No. 351, September 1975, pp. 561-567. 33 references.

Gomez, Frederico, Rafael Ramos Galvan, Joaquin Cravioto, and Silvestre Frenck. "Malnutrition in Infancy and Childhood, with Special Reference to Kwashiorkor," Advances in Pediatrics, 7 (1955).

Green, Manfred S. "Use Of Predictive Value To Adjust Relative Risk Estimates Biased By Misclassification Of Outcome Status." American Journal of Epidemiology, Vol.117, No. 1, 1983, pp. 98-105. 10 references.

Greenhouse, Samuel W. and Nathan Mantel. "The Evaluation of Diagnostic Tests." Biometrics, Vol. 6. December 1950, pp. 399-412. 4 references.

Gullen, Warren H., Jacob E. Bearman and Eugene A. Johnson. "Effects of Misclassification in Epidemiologic Studies. Public Health Reports, Vol. 83, No.11, 1968, pp. 914-918. 1 reference.

Habicht, Jean-Pierre, Charles Yarbrough, and Reynaldo Martorell. Anthropometric Field Methods: Criteria for Selection. In Human Nutrition--A Comprehensive Treatise, Vol. 2, Nutrition and Growth, Roslyn Alfin-Slater and David Kritchevsky, eds. Plenum Publishing Corporation, 1979. 25 references.

Habicht, Jean-Pierre. "Some Characteristics of Indicators of Nutritional Status for Use in Screening and Surveillance." American Journal of Clinical Nutrition, Vol. 33, No. 3, March 1980, pp. 531-535. 12 references.

Habicht, Jean-Pierre. "Indicators for Identifying and Counting the Improperly Nourished." American Journal of Clinical Nutrition, Vol. 35, No. 5, May 1982, pp. 1241-1254, Supplement. 42 references.

Hamil, Peter V., Terence A. Drizd, Clifford L. Johnson, Robert B. Reed, and Alex F. Roche. "NCHS Growth Charts, 1976;" Monthly Vital Statistics Report, Vol. 25, No. 3 (1976).

Harkness, W. L. and L. Katz. "Comparison of the Power Functions for the Test of Independence in 2 x 2 Contingency Tables." Annals of Mathematical Statistics, Vol. 35, September 1964, pp. 1115-1127. 14 references.

James, I. R. "Estimation of the Mixing Proportion in a Mixture of Two Normal Distributions from Simple, Rapid Measurements." Biometrics, Vol. 34, June 1978, pp. 265-275. 15 references.

Jamison, Paul L. and Stephen L. Zegura. "A Univariate and Multivariate Examination of Measurement Error in Anthropometry." American Journal of Physical Anthropometry, Vol. 40, 1974, pp. 197-203. 14 references.

Kemper, H. C. G. and J. J. L. Pieters. "Comparative Study of Anthropometric Measurements of the Same Subjects in Two Different Institutes." American Journal of Physical Anthropology, Vol. 40, 1974, pp. 341-343. 8 references.

Keys, Ancel and Jaakko K. Kihlberg. "Effect of Misclassification on Estimated Relative Prevalence of a Characteristic: Part I. Two Populations Infallibly Distinguished. Part II. Errors in Two Variables." American Journal of Public Health, Vol. 53, No. 10, pp.1656-1660. 7 references.

Koch, Gary G. "The Effect of Non-Sampling Errors on Measures of Association in 2 x 2 Contingency Tables." American Statistical Association Journal, September 1969, pp. 852-863. 16 references.

Marshall, James R., et al. "On The Distortion Of Risk Estimates in Multiple Exposure Level Case-Control Studies." American Journal of Epidemiology, Vol.113, No.4, 1981, pp.464-473. 14 references.

Martorell, Reynaldo, Jean-Pierre Habicht, Charles Yarbrough, Guillerma Guzman and Robert E. Klein. "The Identification and Evaluation of Measurement Variability in the Anthropometry of Preschool Children," American Journal of Physical Anthropology, Vol. 43 (1975), p. 438.

Murphy, James R. "The Relationship of Relative Risk And Positive Predictive Value In 2 x 2 Tables." American Journal of Epidemiology, Vol. 117, No.1, 1983, pp. 86-89. 5 references.

Newell, David J. "Error in the Interpretation of Errors in Epidemiology." American Journal of Public Health, Vol. 52, No. 11, 1962, pp. 1925-1928. 6 references.

Neyman, J. "Outline of Statistical Treatment of the Problem of Diagnosis." Public Health Report, Vol. 62, October 3, 1947, pp. 1449-1456.

Nissen-Meyer, Sven. "Evaluation of Screening Tests in Medical Diagnosis." Biometrics, Vol. 20, December 1964, pp. 730-755. 7 references.

Nystuen, John D. The PRIMOPS Experience: Information Processing in the Design and Performance of a Health Care System: Volume V. (Ann Arbor: Community Systems Foundation, Project on Analysis of Community Level Nutrition Programs, November 1980).

Quade, Dana, et al. "Effects Of Misclassifications On Statistical Inferences in Epidemiology." American Journal of Epidemiology, Vol. III, No. 5, 1980, pp. 503-515. 9 references (SENIC Project, CDC, Atlanta, Ga. 30333).

Rogan, Walter J. and Beth Gladen. "Estimating Prevalence from The Results of a Screening Test." American Journal of Epidemiology, Vol. 107, 1978, pp. 71-76.

Trowbridge, Frederick L. "Clinical and Biochemical Characteristics Associated with Anthropometric Nutritional Categories." American Journal of Clinical Nutrition, Vol. 32, No.5, April 1979, pp. 758-766. 20 references.

Whaley, Fredrick S., et al. "Effects of Method Error on the Power of a Statistical Test." American Journal of

Epidemiology, Vol.111, No.5, 1980, pp. 534-542. 9 references.

Yerushalmy, Jacob. "Statistical Problems in Assessing Methods of Medical Diagnosis, with Special Reference to X-ray Techniques." Public Health Report, Vol. 62, October 3, 1947, pp1432-1449.

APPENDIX A

A program listing for a simulation model to test the effect of a random weight error on a sample of children.

```
C     ***   PROGRAM TO TEST EFFECT OF RANDOM WEIGHT ERROR
      REAL*8 AGE,HT,WT,HAC,WHC,WAC,HAZ,WHZ,WAZ,HAPM,WHPM,WAPM
      DIMENSION NGA(5),NGS(5),NGT(5,5)
      INR=0
      HT=0.0
      STD=.5
      DO 10 I=1,5
      DO 5 J=1,5
    5 NGT(I,J)=0
      NGA(I)=0
   10 NGS(I)=0
C          LOOP OVER FILE
      DO 500 J=1,324
      READ(1,1) ID,IORD,ISX,IAGE,KWT,CDCA
      XWT=KWT*.001
    1 FORMAT(I5,9X,I2,6X,I1,6X,I2,6X,I5,7X,F5.0)
      CALL NORML(XWT,STD,VAL,INR)
      WRITE(2,6) XWT,VAL
    6 FORMAT(1X,2F6.1)
C          CALL CDC STANDARD PROGRAM
      WT=VAL
      AGE=IAGE
      CALL PCTL9Z(AGE,ISX,HT,WT,HAC,WHC,WAC,HAZ,WHZ,WAZ,HAPM,WHPM,WAPM)
      IF (CDCA .LT. 30.) GO TO 100
      IF (CDCA .LT. 60.) GO TO 110
      IF (CDCA .LT. 75.) GO TO 120
      IF (CDCA .LT. 85.) GO TO 130
      IF (CDCA .LT. 200.) GO TO 140
  100 K=5
      GO TO 150
  110 K=4
      GO TO 150
  120 K=3
      GO TO 150
  130 K=2
      GO TO 150
  140 K=1
  150 CDCS=WAPM
      IF (CDCS .LT. 30.) GO TO 200
      IF (CDCS .LT. 60.) GO TO 210
      IF (CDCS .LT. 75) GO TO 220
      IF (CDCS .LT. 85.) GO TO 230
      IF (CDCS .LT. 200.) GO TO 240
  200 L=5
      GO TO 250
  210 L=4
      GO TO 250
  220 L=3
      GO TO 250
  230 L=2
      GO TO 250
  240 L=1
  250 NGA(K)=NGA(K)+1
      NGS(L)=NGS(L)+1
      NGT(K,L)=NGT(K,L)+1
```

25

```
0053        500 CONTINUE
0054            WRITE(6,21)
0055         21 FORMAT('1     SIMULATION OF ERROR IN TAKING WEIGHTS')
0056            WRITE(6,22)
0057         22 FORMAT('O     SIMULATED WEIGHT IS NORMALLY DISTRIBUTED')
0058            WRITE(6,23) STD
0059         23 FORMAT('      AROUND THE ACTUAL WEIGHT WITH STD. DEV. =',F5.2)
0060            WRITE(6,2)
0061          2 FORMAT('O          NORM  GR I GR II  GR 3   UNC')
0062            WRITE(6,3) NGA
0063          3 FORMAT(' ACTUAL  ',5I6)
0064            WRITE(6,4) NGS
0065          4 FORMAT(' SIMULA  ',5I6)
0066            WRITE(6,7)
0067          7 FORMAT('O     ACTUAL VS. SIMULATED')
0068            WRITE(6,8)
0069          8 FORMAT('         N    I   II  III  UNC')
0070            WRITE(6,11) (NGT(1,J),J=1,5)
0071         11 FORMAT(' N  ',5I5)
0072            WRITE(6,12) (NGT(2,J),J=1,5)
0073         12 FORMAT(' I  ',5I5)
0074            WRITE(6,13) (NGT(3,J),J=1,5)
0075         13 FORMAT(' II ',5I5)
0076            WRITE(6,14) (NGT(4,J),J=1,5)
0077         14 FORMAT(' III',5I5)
0078            WRITE(6,15) (NGT(5,J),J=1,5)
0079         15 FORMAT(' UNC',5I5)
0080            CALL EXIT
0081            END
```

65.000
66.000
67.000
68.000
69.000
70.000
71.000
72.000
73.000
74.000
75.000
76.000
77.000
78.000
79.000
80.000
81.000
82.000
83.000
84.000
85.000
86.000
87.000
88.000
89.000
90.000
91.000
92.000
93.000

SUBPROGRAMS CALLED

SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION
IBCOM#	138	NORML	13C	PCTL9Z	140	EXIT	144		

SCALAR MAP

SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION
HT	190	WT	198	AGE	1A0	HAC	1A8	WHC	1B0
WAC	1B8	HAZ	1C0	WHZ	1C8	WAZ	1D0	HAPM	1D8
WHPM	1E0	WAPM	1E8	INR	1F0	STD	1F4	I	1F8
J	1FC	ID	200	IORD	204	ISX	208	IAGE	20C
KWT	210	CDCA	214	XWT	218	VAL	21C	K	220
CDCS	224	L	228						

ARRAY MAP

SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION
NGA	22C	NGS	240	NGT	254				

FORMAT STATEMENT MAP

SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION
1	288	6	2D1	21	2DA	22	309	23	33B
2	371	3	39C	4	3AD	7	3BE	8	3DC
11	3FD	12	409	13	415	14	421	15	42D

```
*OPTIONS IN EFFECT*  ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,MAP
*OPTIONS IN EFFECT*  NAME = MAIN     , LINECNT =       57
*STATISTICS*    SOURCE STATEMENTS =       81,PROGRAM SIZE =     2548
*STATISTICS*  NO DIAGNOSTICS GENERATED

NO STATEMENTS FLAGGED IN THE ABOVE COMPILATIONS.
```

```
0001          SUBROUTINE NORML(EX,STD,VAL,INR)                                  1.000
        C          ROUTINE TO RETURN NUMBER FROM A NORMAL DISTRIBUTION          2.000
0002          SUM=0.0                                                           3.000
0003          DO 10 I=1,12                                                      4.000
0004          R=URAND(INR)                                                      5.000
0005       10 SUM=SUM+R                                                         6.000
0006          VAL=STD*(SUM-6.0)+EX                                              7.000
0007          RETURN                                                            8.000
0008          END                                                               9.000
```

SUBPROGRAMS CALLED

SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION
URAND	94								

SCALAR MAP

SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION	SYMBOL	LOCATION
SUM	9C	I	A0	R	A4	INR	A8	VAL	AC
STD	B0	EX	B4						

```
*OPTIONS IN EFFECT*  ID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,MAP
*OPTIONS IN EFFECT*  NAME = NORML    , LINECNT =       57
*STATISTICS*    SOURCE STATEMENTS =        8,PROGRAM SIZE =      460
*STATISTICS*  NO DIAGNOSTICS GENERATED
```

APPENDIX B

(1) A BASIC language listing of the tutorial program for explaining the effects of misclassification upon evaluation outcomes.

(2) A BASIC language listing of the program for performing the computations and printing the results of the application of the relevant correction formulas.

```
100  REM  *** UNDEREPORTING PROGRAM #1
105  DEF  FN TH(X) = ( INT ((X * 1000.) + .5)) / 1000.
110  HOME
120  PRINT "WE ARE ABOUT TO EXAMINE A"
130  PRINT "CURIOUS PHENOMENON ENCOUNTERED"
140  PRINT "IN EVALUATING NUTRITION INTERVENTIONS."
150  PRINT : PRINT "THE OBSERVED CHANGE IN PREVALENCE RATE"
160  PRINT "IS OFTEN INCORRECT DUE TO THE FACT"
170  PRINT "THAT SOME CHILDREN ARE MISCLASSIFIED AS"
180  PRINT "TO THEIR NUTRITIONAL STATUS."
190  GOSUB 2000
200  HOME
210  PRINT "TWO CONCEPTS, BORROWED FROM"
215  PRINT "EPIDEMIOLOGY, ARE USEFUL FOR"
220  PRINT "UNDERSTANDING THIS PHENOMENON."
230  PRINT : PRINT "THE PROBABILITY OF CORRECTLY"
240  PRINT "CLASSIFYING A MALNOURISHED"
245  PRINT "CHILD IS CALLED"
250  PRINT : PRINT "       *** SENSITIVITY ***"
260  PRINT : PRINT "PLEASE ESTIMATE THE SENSITIVITY"
270  PRINT "OF YOUR TEST--ENTER A NUMBER"
280  PRINT "BETWEEN .5 AND 1"
290  INPUT "? ";U
300  IF U < .5 THEN  GOTO 320
310  IF U < 1 THEN  GOTO 350
320  PRINT : PRINT "ENTRY NOT IN RANGE, TRY AGAIN."
330  GOTO 290
350  HOME
360  PRINT "SIMILARLY, THE PROBABILITY OF"
370  PRINT "CORRECTLY CLASSIFYING A TRULY"
380  PRINT "HEALTHY CHILD IS CALLED"
390  PRINT : PRINT "       *** SPECIFICITY ***"
400  PRINT : PRINT "PLEASE ESTIMATE THE SPECIFICITY"
410  PRINT "OF YOUR TEST--ENTER A NUMBER"
420  PRINT "BETWEEN .5 AND 1"
430  INPUT "? ";V
440  IF V < .5 THEN  GOTO 460
450  IF V < 1 THEN  GOTO 475
460  PRINT : PRINT "ERROR IN ENTRY, TRY AGAIN."
470  GOTO 430
475  HOME
500  PRINT "NOW, IN ORDER TO ESTIMATE THE"
510  PRINT "ERROR IN CHANGE IN PREVALENCE,"
520  PRINT "WE NEED TO KNOW THE OBSERVED RATE"
530  PRINT "BEFORE AND AFTER YOUR INTERVENTION."
540  PRINT : PRINT "PLEASE ENTER THE RATES IN"
550  PRINT "THE FORM OF PERCENTAGES, E.G. 40.2"
560  PRINT : INPUT "PREVALENCE BEFORE? ";R1
565 P1 = R1 * .01
570  INPUT "PREVALENCE AFTER? ";R2
575 P2 = R2 * .01
600  HOME
602  LET I$ = "INCREASE"
604  LET D$ = "DROP"
610  LET O = P2 - P1
620  IF O > 0 THEN  GOTO 660
630  LET P$ = D$
640  LET P =  FN TH(P1 - P2)
642  LET R = P * 100
```

```
645  PRINT "YOU HAVE OBSERVED A ";R;" (";R1;" - ";R2;")"
650  GOTO 685
660  LET P$ = I$
670  LET P =  FN TH(O)
675  LET R = P * 100.
680  PRINT "YOU HAVE OBSERVED A ";R;" (";R2;" - ";R1;")"
685  PRINT "PERCENTAGE POINT ";P$;" IN"
687  PRINT "THE PREVALENCE OF MALNUTRITION"
690  LET A = O / (U + V - 1)
692  LET A =  FN TH(A)
695  PRINT
700  PRINT "ASSUMING SENSITIVITY AND SPECIFICITY"
710  PRINT "REMAIN CONSTANT OVER TIME,"
720  PRINT "WE USE THE FORMULA": PRINT
730  PRINT "    A = O / (U + V - 1)"
735  PRINT
740  PRINT "WHERE U=SENSITIVITY, V=SPECIFICITY"
750  PRINT "       O=OBSERVED CHANGE, A=ACTUAL CHANGE"
760  PRINT : PRINT "IN THIS CASE, U =";U;" AND V =";V
770  PRINT "THEREFORE,"
775  IF A < 0 THEN A =  - A
777  LET AR = A * 100.
780  PRINT "THE ACTUAL ";P$;" IS ";AR;" NOT ";R
800  GOSUB 2000
810  HOME
820  PRINT "IN PRACTICE, SENSITIVITY AND"
830  PRINT "SPECIFICITY MAY NOT REMAIN CONSTANT"
840  PRINT "OVER TIME.": PRINT
850  PRINT "THE TRUE PREVALENCE AT A POINT"
860  PRINT "IN TIME CAN BE DETERMINED"
870  PRINT "FROM THE OBSERVED PREVALENCE"
880  PRINT "WITH THE FOLLOWING EQUATION:"
890  PRINT : PRINT "      TRUE = (OBS + V - 1) / (U + V - 1)"
900  PRINT : PRINT "FOR EXAMPLE,"
910  PRINT "WITH SENSITIVITY (U) = ";U
920  PRINT "AND SPECIFICITY (V)  = ";V
930  PRINT "AND OBSERVED PREVALENCE = ";R1
940 T1 = (P1 + V - 1) / (U + V - 1)
945  LET T1 =  FN TH(T1)
947  LET T1 = T1 * 100.
950  PRINT : PRINT "WE HAVE A TRUE PREVALENCE = ";T1
960  GOSUB 2000
1000  REM  FINAL SECTION
1010  HOME
1020  PRINT "WE CAN DETERMINE THE MAGNITUDE OF THE"
1030  PRINT "ERROR WE MAKE IN USING THE OBSERVED"
1040  PRINT "CHANGE IN PREVALENCE TO ESTIMATE THE"
1050  PRINT "TRUE CHANGE BY REPEATED APPLICATION"
1060  PRINT "OF THE FORMULA": PRINT
1070  PRINT "      TRUE = (OBS + V - 1) / (U + V - 1)"
1080  PRINT : PRINT "FOR EACH OF TWO TIME PERIODS."
1081  PRINT : PRINT "WE NEED ONLY CALCULATE THE TRUE"
1082  PRINT "PREVALENCES AT EACH POINT IN TIME"
1083  PRINT "AND SUBTRACT ONE FROM THE OTHER"
1084  PRINT "TO DETERMINE THE TRUE DIFFERENCE."
1086  GOSUB 2000
```

```
1087  HOME
1090  PRINT "SEVERAL FEATURES OF SUCH AN APPLICATION"
1100  PRINT "OF THIS EQUATION SHOULD BE NOTED:"
1110  PRINT : PRINT "  1) THE TRUE CHANGE IN PREVALENCE,"
1115  PRINT "      TRUE(T2) - TRUE(T1), IS A FUNCTION"
1120  PRINT "      OF BOTH OBSERVED PREVALENCE RATES"
1125  PRINT "      AS WELL AS THE SENSITIVITY AND"
1130  PRINT "      SPECIFICITY IN EACH TIME PERIOD."
1135  PRINT : PRINT "  2) WHEN SENSITIVITY OR SPECIFICITY"
1140  PRINT "      CHANGES OVER TIME, WE CAN UNDER"
1145  PRINT "      REPORT, OVER REPORT, OR IN THE"
1150  PRINT "      WORST CASE, REPORT CHANGE IN THE"
1155  PRINT "      OPPOSITE DIRECTION FROM THE TRUE"
1158  PRINT "      CHANGE."
1160  GOSUB 2000
1170  HOME
1175  PRINT : PRINT "  3) CHANGES IN SPECIFICITY TEND TO"
1180  PRINT "      PLAY THE MORE IMPORTANT ROLE IN"
1185  PRINT "      THE DETERMINATION OF BOTH THE"
1190  PRINT "      MAGNITUDE AND DIRECTION OF THE"
1195  PRINT "      ERROR MADE IN ESTIMATING CHANGE"
1197  PRINT "      IN PREVALENCE."
1216  GOSUB 2000
1220  HOME
1230  PRINT "TO EXPERIMENT WITH CHANGING"
1240  PRINT "SENSITIVITIES AND SPECIFICITIES, TYPE"
1250  PRINT : PRINT "RUN UND.REP"
1260  GOTO 3000
2000  PRINT : PRINT : PRINT
2010  PRINT "TO GO ON, PLEASE TYPE C"
2020  GET Z$
2030  IF Z$ <  > "C" THEN 2020
2040  PRINT
2050  RETURN
3000  END
```

```
100  REM  PROGRAM TO EXPLORE UNDERREPORTING
110  REM  VERSION 3
120  DEF  FN TH(X) = ( INT ((X * 1000.) + .5)) / 1000.
130  DIM A$(9)
140  LET M1$ = " RISE DROP "
150  LET M2$ = " UNDER OVER "
160  LET BL$ = "      "
170  LET C5 = 0
200  REM  INTRODUCTION
210  HOME
220  PRINT "WE WILL NOW ALLOW YOU TO SET"
230  PRINT "SENSITIVITY (INITIAL AND FINAL),"
240  PRINT "SPECIFICITY (INITIAL AND FINAL),"
250  PRINT "AND OBSERVED PREVALENCE (INITIAL AND"
260  PRINT "FINAL."
270  PRINT : PRINT "FOR EACH SET OF VALUES YOU ENTER,"
280  PRINT "THE COMPUTER WILL CALCULATE TRUE CHANGE"
290  PRINT "AND COMPARE IT WITH OBSERVED CHANGE."
300  PRINT : PRINT "WE WILL START WITH AN ARBITRARY"
310  PRINT "SET OF VALUES.  YOU MAY CHANGE ANY"
320  PRINT "OR ALL OF THEM AS YOU WISH."
330  PRINT : PRINT "AFTER, AT MOST, 9 CHANGES, THE"
340  PRINT "RESULTS WILL BE STORED IN A FILE, ON"
350  PRINT "DISK FOR LATER VIEWING."
360  GOSUB 4100
1000  REM  INPUT SEGMENT
1010  REM  SET INITIAL VALUES FOR ALL VARIABLES
1020  LET U1 = .8: LET U2 = .8
1030  LET V1 = .9: LET V2 = .9
1040  LET Q1 = 40.2: LET Q2 = 35.5
1050  HOME
1060  PRINT "WE HAVE:"
1070  GOSUB 4000
1080  PRINT : PRINT "TO MAKE CHANGES"
1090  PRINT "ENTER THE CODE AS INDICATED"
1100  PRINT "(1) INITIAL SENSITIVITY"
1110  PRINT "(2) INITIAL SPECIFICITY"
1120  PRINT "(3) FINAL SENSITIVITY"
1130  PRINT "(4) FINAL SPECIFICITY"
1140  PRINT "(5) INITIAL OBSERVED PREVALENCE"
1150  PRINT "(6) INITIAL FINAL PREVALENCE"
1160  PRINT "(7) TO QUIT"
1170  PRINT "(9) PROCEED TO CALCULATION"
1180  PRINT : INPUT "CODE? ";C1
1190  ON C1 GOTO 1200,1240,1280,1320,1360,1400,2970,1440,2000
1200  INPUT "NEW INIT. SENS. (.5 < U1 < 1)? ";U1
1210  IF U1 < .5 THEN  GOTO 1230
1220  IF U1 <  = 1. THEN  GOTO 1060
1230  PRINT "ILLEGAL VALUE": GOTO 1200
1240  INPUT "NEW INIT. SPEC. (.5 < V1 < 1)? ";V1
1250  IF V1 < .5 THEN  GOTO 1270
1260  IF V1 <  = 1. THEN  GOTO 1060
1270  PRINT "ILLEGAL VALUE": GOTO 1240
1280  INPUT "NEW FINAL SENS. (.5 < V1 < 1)? ";U2
1290  IF U2 < .5 THEN  GOTO 1310
1300  IF U2 <  = 1. THEN  GOTO 1060
```

```
1310  PRINT "ILLEGAL VALUE": GOTO 1280
1320  INPUT "NEW FINAL SPEC. (.5 < V2 < 1)? ";V2
1330  IF V2 < .5 THEN  GOTO 1350
1340  IF V2 <  = 1. THEN  GOTO 1060
1350  PRINT "ILLEGAL VALUE": GOTO 1320
1360  INPUT "NEW INIT. OBS PREV. (5 < P1 < 60)? ";Q1
1370  IF Q1 < 5. THEN  GOTO 1390
1380  IF Q1 <  = 60. THEN  GOTO 1060
1390  PRINT "ILLEGAL VALUE": GOTO 1360
1400  INPUT "NEW FINAL OBS. PREV. (5 < P2 < 50)? ";Q2
1410  IF Q2 < 5. THEN  GOTO 1430
1420  IF Q2 <  = 50. THEN  GOTO 1060
1430  PRINT "ILLEGAL VALUE": GOTO 1400
1440  PRINT "ILLEGAL CODE": GOTO 1180
2000  REM  WE HAVE ALL INPUTS
2010  REM  COMPUTE TRUE PREVALENCES
2020  REM  TIME 1
2030  LET O1 = Q1 * .01: REM  CNVRT TO DECIMAL
2040  LET T1 = (O1 + V1 - 1) / (U1 + V1 - 1)
2050  IF T1 > 0 THEN  GOTO 2100
2060  PRINT "TRUE PREVALENCE AT TIME 1 IS < 0"
2070  PRINT "SET NEW PARAMETERS"
2080  GOSUB 4100
2090  GOTO 1060
2100  REM  TIME 2
2110  LET O2 = Q2 * .01: REM  CNVRT TO DECIMAL
2120  LET T2 = (O2 + V2 - 1) / (U2 + V2 - 1)
2130  IF T2 > 0 THEN  GOTO 2160
2140  PRINT "TRUE PREVALENCE AT TIME 2 IS < 0"
2150  GOTO 2070
2160  REM  PRINT TABLEAU
2170  GOSUB 4000
2180  LET R1 =  FN TH(T1): LET R2 =  FN TH(T2)
2190  LET R1 = R1 * 100.: LET R2 = R2 * 100.
2200  PRINT "TRUE PREV.  =    ";R1,R2
2210  PRINT
2220  REM  ESTABLISH DIRECTION OF CHANGE
2230  REM  IN OBSERVED PREVALENCE
2240  LET OD = O2 - O1
2250  IF OD > 0 THEN P1$ =  LEFT$ (M1$,6): LET F1 = 1: GOTO 2270
2260 P1$ =  RIGHT$ (M1$,6): LET F1 = 2: LET OD =  - OD
2270  LET OD =  FN TH(OD): LET QD = OD * 100.
2280  PRINT "OBSERVED";P1$;"IN PREVALENCE = ";QD
2290  REM  ESTABLISH DIRECTION CHANGE
2300  REM  IN TRUE PREVALENCE
2310  LET TD = T2 - T1
2320  IF TD > 0 THEN P1$ =  LEFT$ (M1$,6): LET F2 = 1: GOTO 2340
2330 P1$ =  RIGHT$ (M1$,6): LET F2 = 2: LET TD =  - TD
2340  LET TD =  FN TH(TD): LET RD = TD * 100.
2350  PRINT "    TRUE";P1$;"IN PREVALENCE = ";RD
2360  REM  PRINT A SUMMARY STATEMENT
2370  PRINT
2380  IF F1 <  > F2 THEN  GOTO 2460
2390  REM  PREVALENCE INCREASED IN BOTH
2400  IF RD > QD THEN P2$ =  LEFT$ (M2$,7): LET X = RD - QD: LET G1 = 1: GOT
     2420
2410 P2$ =  RIGHT$ (M2$,6):X = QD - RD:G1 = 2
2420  PRINT "WE";P2$;"REPORTED THE CHANGE"
2430  LET X =  FN TH(X)
2440  PRINT "IN PREVALENCE BY ";X;" POINTS ";
2442  REM  COMPUTE PERCENT CHANGE
2444 X5 = QD / RD:X5 =  FN TH(X5) * 100
2446  PRINT "(";X5;" %)"
2450  GOTO 2500
2460  REM  PREVALENCE DECREASED IN BOTH
```

```
2480  PRINT "WE OBSERVED A CHANGE IN THE OPPOSITE"
2490  PRINT "DIRECTION FROM THE TRUE CHANGE"
2500  REM  PREPARE RESULTS FOR FILE SAVE
2510  REM  CREATE A SINGLE STRING VARIABLE
2520  LET C5 = C5 + 1: REM  INCREMENT COUNTER
2530  LET A$(C5) = "": REM  START WITH NULL STRING
2540  LET S1$ =  STR$ (C5)
2550  LET A$(C5) = A$(C5) + S1$
2560  REM  BEGIN STRING CONVERSIONS
2570  LET S1$ =  STR$ (U1):WD = 4
2580  GOSUB 4200
2590  LET S1$ =  STR$ (V1):WD = 4
2600  GOSUB 4200
2610  LET S1$ =  STR$ (U2):WD = 4
2620  GOSUB 4200
2630  LET S1$ =  STR$ (V2):WD = 4
2640  GOSUB 4200
2650  LET S1$ =  STR$ (Q1):WD = 6
2660  GOSUB 4200
2670  LET S1$ =  STR$ (Q2):WD = 6
2680  GOSUB 4200
2690  IF F1 = 1 THEN S1$ =  LEFT$ (M1$,5): GOTO 2710
2700 S1$ =  MID$ (M1$,6,5)
2710  LET A$(C5) = A$(C5) + S1$
2720  LET S1$ =  STR$ (QD):WD = 5
2730  GOSUB 4200
2740  LET S1$ =  STR$ (R1):WD = 6
2750  GOSUB 4200
2760  LET S1$ =  STR$ (R2):WD = 6
2770  GOSUB 4200
2780  IF F2 = 1 THEN S1$ =  LEFT$ (M1$,5): GOTO 2800
2790 S1$ =  MID$ (M1$,6,5)
2800  LET A$(C5) = A$(C5) + S1$
2810  LET S1$ =  STR$ (RD):WD = 5
2820  GOSUB 4200
2830  ON G1 GOTO 2840,2850,2860
2840 S1$ =  LEFT$ (M2$,6): GOTO 2870
2850 S1$ =  RIGHT$ (M2$,6): GOTO 2870
2860 S1$ = " OPPOS"
2870  LET A$(C5) = A$(C5) + S1$
2880  IF G1 = 3 THEN  GOTO 2920
2890  LET S1$ =  STR$ (X):WD = 6
2900  GOSUB 4200
2902  LET S1$ =  STR$ (X5):WD = 6
2904  GOSUB 4200
2910  GOTO 2930
2920  LET A$(C5) = A$(C5) + BL$
2930  REM  IF LESS THAN 9 OPTIONS - GO ON
2940  IF C5 = 9 THEN  GOTO 2980
2950  GOSUB 4100
2960  GOTO 1060
2970  LET CE = CE + 1
2980  REM  WRITE ARRAYS INTO FILE
2990  REM  IF EXIT BEFORE 9, FILL ARRAYS
3000  IF C5 = 9 THEN  GOTO 3040
3010  LET C5 = C5 + 1
3020  LET A$(C5) = " "
3030  GOTO 3000
3040  REM  WRITE THE SEQUENTIAL FILE
3042  PRINT : PRINT "TO STORE IN A FILE"
3044  INPUT "ENTER FILE NAME (NO TO SKIP)";FF$
3046  IF FF$ = "NO" THEN  GOTO 5000
3050  LET D$ =  CHR$ (4)
3060  PRINT D$;"OPEN ";FF$
3070  PRINT D$;"WRITE ";FF$
```

```
3090  PRINT A$(I)
3100  NEXT I
3110  PRINT D$;"CLOSE ";F1$
3120  HOME
3130  PRINT "A FILE CONTAINING ALL OF THE RESULTS"
3140  PRINT "CALCULATED DURING THIS RUN HAS BEEN"
3150  PRINT "CREATED--TO SEE THESE RESULTS,"
3160  PRINT "RUN UR.PRNT.TABL"
3170  GOTO 5000
4000  HOME
4005  PRINT "                    INITIAL        FINAL"
4010  PRINT "SENSITIVITY =     ";U1,U2
4020  PRINT "SPECIFICITY =     ";V1,V2
4030  PRINT "OBS. PREV.  =     ";Q1,Q2
4040  RETURN
4100  PRINT
4110  PRINT "TO GO ON, TYPE C"
4120  GET Z$
4130  IF Z$ <  > "C" THEN 4120
4140  RETURN
4200  REM  SUBROUTINE TO PAD STRING
4201  REM  VARIABLES WITH BLANKS
4210  LET L1 =  LEN (S1$): LET L2 = WD - L1
4220  LET S2$ =  LEFT$ (BL$,L2)
4230  LET A$(C5) = A$(C5) + S2$ + S1$
4240  RETURN
5000  END
```

```
LOAD UR.PRNT.TABL
]LIST

100  REM  PROGRAM TO DISPLAY AN UNDERREPORTING
110  REM  TABLEAU, STORED ON DISK, ON THE
120  REM  SCREEN AND/OR THE PRINTER
130  REM  STORE UP TO 9 CONFIGURATIONS
140  DIM A$(9)
150  REM  REMINDER *** ACTIVATE 80-COL CARD
155  HOME
160  PRINT "IS 80-COLUMN CARD ACTIVATED (Y/N)? "
170  GET Z$
180  IF Z$ = "Y" GOTO 500
190  IF Z$ = "N" GOTO 210
200  GOTO 170
210  PRINT : PRINT "TYPE PR#3, RETURN"
220  PRINT "THEN TYPE RUN AGAIN"
230  GOTO 5000
500  REM  INITIALIZE HEADINGS
510  LET B$ = "#  U1  V1  U2  V2    O1    O2 OBS CHNGE    P1    P2 TRUE C
     HNG   RES   AMT     %"
520  LET C$ = "- --- --- --- --- ----- ----- --------- ----- ----- ------
     --- ----- ----- -----"
525  LET CR$ =  CHR$ (13)
530  REM  READ THE VALUES FROM THE DISK
540  INPUT "FILE NAME? ";F$
550  LET D$ =  CHR$ (4)
560  PRINT D$;"OPEN ";F$
570  PRINT D$;"READ";F$
580  REM  LOOP OVER EIGHT FIELDS
590  FOR I = 1 TO 9
600 : INPUT A$(I)
610  NEXT I
620  PRINT D$;"CLOSE";F$
630  REM  ASCERTAIN WHETHER TO PRINT ON SCREEN OR PRINTER
640  PRINT "TYPE S FOR SCREEN, P FOR PRINTER"
650  PRINT "      B FOR BOTH OR Q TO QUIT"
660  GET Z$
670  IF Z$ = "Q" THEN  GOTO 5000: REM  EXIT
680  IF Z$ = "B" THEN  LET W1 = 2: GOTO 730: REM  PRINT ON BOTH
690  IF Z$ = "P" THEN  GOTO 760: REM  PRINTER ONLY
700  IF Z$ = "S" THEN  LET W1 = 1: GOTO 730: REM  SCREEN ONLY
710  GOTO 660
730  REM  DISPLAY OUTPUT ON SCREEN
735  HOME
740  GOSUB 1000
750  IF W1 = 1 THEN  GOTO 5000
760  PRINT CR$;D$;"PR#1"
770  GOSUB 1000
780  PRINT D$;"PR#3"
790  GOTO 5000
1000  REM  PRINT SUBROUTINE
1010  PRINT B$: PRINT C$
1020  FOR I = 1 TO 9
1030 : PRINT A$(I)
1040  NEXT I
1050  RETURN
5000  END

]PR#3
```