62

CORNELL NUTRITIONAL SURVEILLANCE PROGRAM

Working Paper Series

Division of Nutritional Sciences
New York State Colleges of Human Ecology & Agriculture and Life Sciences
Statutory Colleges of the State University of New York
Cornell University
Ithaca, New York 14853

Cornell University is an equal opportunity, affirmative action employer and educator.

METHODS USED TO IDENTIFY PRIORITY MUNICIPALITIES FOR INTERVENTION BASED ON OPERATION TIMBANG AND SOCIO-ECONOMIC DATA

by

Hamid Tabatabai, John Mason, Marito Garcia, Terry Elliott and Janice Mitchell

September 1982 **No. 7**

The CNSP is supported by Cooperative Agreement # AID DSAN CA-0240 between the Office of Nutrition, Bureau for Science & Technology, USAID, and the Division of Nutritional Science, New York State Colleges of Human Ecology & Agriculture and Life Sciences, Cornell University, Ithaca, New York 14853.

A report of research of the Cornell University Agricultural Experiment Station

ABSTRACT

This paper reports on the analytical component of the initial assessment phase that was recently carried out in the Philippines in order to improve nutritional surveillance activities at the national and lower administrative levels. The major purpose of the analysis was to compile lists of nutritionally worst affected municipalities and to produce municipal and provincial level mappings of malnutrition for the entire country. Prevalence of second and third degree malnutrition was calculated for 815 municipalities using data from 1979 Operation Timbang. This information was then used along with certain socio-economic data to classify each of 683 other municipalities without nutritional data or with unreliable data into one of three nutritional categories: worst affected, moderately affected and least affected. 74 municipalities from a total of 1572 in the country were excluded from the entire analysis. Discriminant analysis was used for classification. Two listings of worst affected municipalities and a municipal level mapping of malnutrition were produced. A provincial level mapping was also produced where provinces were grouped into six categories depending on their proportion of worst affected and least affected municipalities. Least affected municipalities and provinces are located predominantly in the northern regions of the country whereas central regions include the bulk of the worst affected municipalities and provinces. Southern regions are about average.

TABLE OF CONTENTS

	Page
ABSTRACT	
1.0 INTRODUCTION	1
2.0 DATA AND METHODS	2
2.1 Nutritional Status in Philippines	2
2.2 Data used in the analysis	3
2.3 Data Cleaning	5
2.4 Methods of Analysis	7
3.0 RESULTS	10
3.1 Association between Variables	10
3.2 Effect of Substituting for Missing Socio-Economic Data	11
3.3 Listings of Worst-Affected Municipalities	12
3.4 Mappings of Malnutrition	15
4.0 DISCUSSION	17
4.1 Methods	18
4.2 The Future	18
REFERENCES	20

1.0 INTRODUCTION

In recent years a growing number of countries, developed and developing alike, have initiated nutritional surveillance systems of varying degrees of comprehensiveness. The principal function of these systems is to help improve decision making by providing the decision makers with timely and regular flow of information on nutritional situation and its determinants. It has been recommended that an initial assessment phase precede the actual establishment of a nutritional surveillance system (Mason et al. 1982). This assessment is intended to identify the relevant decisions that affect nutrition, the potential users of a surveillance system's output and their information needs and the sources of nutritional and relevant socio-economic data. It must specify in detail the objectives of the system, its design, resources required and, not least, a workplan that clearly delineates the responsibilities of all personnel and agencies involved. An initial assessment need not start only when the decision to establish a surveillance system has already been made. Indeed, such an assessment could reach the conclusion that a successful surveillance system need not be established at a given time. The success of a nutritional surveillance system hinges critically on a thorough and honest assessment of its feasibility.

An initial assessment was recently carried out in the Philippines in order to improve nutritional surveillance activities at national and lower administrative levels. A report was then issued (National Nutrition Council/Cornell Nutritional Surveillance Program 1981) that laid down in detail the purposes of a nationwide nutritional surveillance system, its sources of data, potential users of its outputs and their information needs, possible nutritional indicators, types of surveillance outputs and the responsibilities for data collection, processing, analysis and interpretation. The report suggested an organizational structure for the surveillance system at all administrative levels, identified resource requirements and training needs and suggested a detailed workplan for the first two years of its implementation. The initial assessment included a study of the nutritional situation in the Philippines in 1979. The major purpose of this study was to demonstrate certain approaches, using existing data, to producing useful information of the kinds that a surveillance system could regularly provide in the future. The major outputs of the analysis consisted of the identification of nutritionally worst affected municipalities and provinces and mappings of

nutritional status at municipal and provincial levels in 1979. The primary uses of these outputs are in identifying priority areas for targetting public programs and in providing baseline information that can serve in the future to evaluate the adequacy of new policies and programs. This paper reports the results of this analysis.

The outline of the paper is as follows. The sources of data, how data were cleaned and the methods of analysis are described first. The results described consist of a "short" and a "long" listing of worst affected municipalities which were produced at the request of the National Nutrition Council (NNC) for use by several public agencies for area targetting and two mappings at municipal and provincial levels which provide static pictures of the geographic distribution of malnutrition and can serve as baseline for detection of long term changes when similar mappings become available in the future. A discussion of the data, methods, findings and possibilities for future improvement wraps up the paper.

2.0 DATA AND METHODS

2.1 Nutritional Status Data in the Philippines

The principal source of data on nutritional status in the Philippines is a regular community weighing program for preschool children known as Operation Timbang (OPT). The main purpose of OPT is, as a screening mechanism, i.e. to identify malnourished preschool children at the local level for intervention. The data are also used for municipal planning of the Philippine Nutrition Program. OPT collects weight and age data on preschoolers on a yearly basis. The information is collected by rural health personnel (midwife, nurse), school principals or teachers, village nutrition workers (Barangay Nutrition Scholars), extension workers, and others. Tabulation of the information at the local level is done by rural health personnel, nutritionists or agricultural extension workers and at the regional level by the Regional Health Office personnel and by National Nutrition Council's Regional Offices.

Although intended to be a total enumeration, OPT actually covers an average around 30 percent of preschoolers nationwide. Many of the country's 1572 municipalities are either not covered at all, or only partially covered, or their data are not reported. Inadequate resources, non-compliance with

guidelines and problems with timely reporting hinder the use of OPT for planning purposes and create problems in the analysis of its data.

Since OPT is a primarily screening operation designed to identify the malnourished for direct individual intervention, possible deliberate selection of the malnourished may tend to result in overestimation of the actual prevalence of malnutrition. On the other hand lack of sufficient resources and similar reasons may mean that outlying areas may not be as adequately covered as more accessible areas. Since the less accessible areas are likely to have relatively higher malnutrition inadequate outreach tends to introduce a downward bias in the estimate of prevalence at the municipal level. Other biases probably exist as well. At any rate there appears to be an overall bias in OPT which overestimates the prevalence of malnutrition nationwide.

A carefully conducted nationwide nutrition sample survey covering 3274 randomly selected children revealed a national prevalence of malnutrition of 22 percent in 1978 (calculated from data in Food and Nutrition Research Institute, 1981, Table B 1.1, p. 69) whereas the OPT estimate of the prevalence for the same year is about 30 percent (NNC, 1979). Both sources define the prevalence of malnutrition as the percentage of preschool children with weight for age below 75 percent using Philippine standards. Hence OPT may be overestimating the actual prevalence of malnutrition by probably one third. If this bias were uniform across municipalities it would not be of much consequence when one is concerned with relative extent of malnutrition across municipalities as we are. The bias however is unlikely to be the same everywhere and our analysis will have to take this possibility into account.

2.2 Data Used in the Analysis

The OPT data used for this study were those for the year 1979 which were obtained from municipal nutrition action plans. OPT gives the breakdown of weighed preschoolers by nutritional category as normal (more than 90 percent of median weight for age using the Philippine Standards), mild (75-90 percent), moderate (60-75 percent) and severe (less than 60 percent). The prevalence of malnutrition is defined here as the percentage of moderate and severe cases in the total number of preschoolers weighed in a municipality. This prevalence is also referred to here as OPT and it should be clear from the text whether OPT is

referring to Operation Timbang or the prevalence of malnutrition at municipal level.

Table 1 provides the relevant information on variables used in the analysis. Since OPT data are sometimes unreliable or altogether lacking for a large number of municipalities, an attempt has been made to predict the nutritional status category of these municipalities by the use of information on some municipal-level demographic and socio-economic variables obtained from published sources. Demographic variables allow us to calculate the coverage of preschoolers by OPT. Coverage is defined as the percentage of preschoolers in a municipality who are weighed. It can be readily calculated if accurate information on the number of preschoolers were available. It is not and we will later describe how coverage was estimated for municipalities with OPT information. Socio-economic variables are needed to predict the nutritional status in municipalities without OPT or with unreliable OPT. By relating them to prevalence in municipalities with reliable OPT estimation of relationships which could then be used, with socio-economic information on other municipalities, to predict their likely nutritional situation was sought. Such an exercise depends critically on how strong the estimated relationships are, i.e. on the correlations between nutritional status and socio-economic characteristics.

Socio-economic information at municipal level is fairly extensive (see source) but only five which appeared a priori to be more useful: these are population density in number of persons per square kilometers, percent of rice area irrigated, percent of dwelling units with cogon and nipa roofing material indicating poor housing conditions, percent of barangays (villages) not served by local waterworks system and the number of physicians per 10,000 population (Table 1). The choice of only five is a concession to simplicity since the purpose of the analysis was at first primarily demonstrative. Similar analyses can be much improved with more and better socio-economic data.

The data set of 1498 municipalities excludes 74 (particularly large cities such as Metropolitan Manila) for a variety of reasons: too large a population density to fit the coded format, lack of information on any variable other than the identifying codes, etc. The 1498 cases in the set each contain at least some information but the number of missing observations for some variables is large.

2.3 Data Cleaning

Data "cleaning" was relatively straightforward. The mean, minimum and maximum values of each variable were examined and consistency checks involving more than one variable were made. Number of barangays (villages) in each municipality and the number of those not served by local waterworks system were adjusted in a few instances or considered missing so that the latter would in no case exceed the former (Table 1, Items 18 and 19).

The only major difficulty of cleaning the data arose in calculating the coverage of preschoolers by OPT, i.e. the percentage of preschoolers weighed in a municipality. Coverage is needed to assess the reliability of the reported figures for prevalence of malnutrition. Preschool population figures were reported from the field on the same document as were OPT-related data. In 525 municipalities out of 1498 in the data set both are missing. In 54 cases with OPT, preschool population is missing and, further, for one out of every six municipalities with preschool information, the reported figure was equal to, or came within one percent of, the number of weighed children in the municipality. This is believed to have resulted from a confusion in the field between the total number of preschoolers and those weighed. In some other cases the reported preschool population has no reasonable relationship to the population of the municipality. Hence preschool figures needed adjustment in many cases.

Total population figures come from census documents and are available for all but five municipalities. They refer to 1978 and are assumed to be reliable. It has further been assumed that the proportion of preschoolers in total population of each municipality lies within a range of 0.12 to 0.25 which appears to be reasonable from a demographic standpoint. From published data (United Nations 1980) the proportion of 0-6 year-old children in total population of the Philippines was calculated to be approximately 0.22 in 1976. The upper limit of 0.25 seemed high enough as the calculated proportion of preschoolers in total population exceeded this limit in only few municipalities. The lower limit of 0.12 appears to be the minimum reasonable. Many municipalities had preschool population figures which fell still less than this value as proportion of total population because they were underestimated in some cases by being confused with total weighed children and in others for reasons that are not very clear. At any rate, it has been assumed that preschool population is 20 percent of total

population whenever:

(a) The number of weighed children was within two percentage points of the reported preschool population;

or (b) The condition (a) did not apply and the reported preschool population figure was not between 12-25 percent of total population.

or (c) Preschool population figure was missing.

Note that condition (c) was really relevant only when a municipality did have OPT. In the absence of OPT, preschool population would not be needed. Hence the number of observations for this variable is 968 (same municipalities as with OPT except for five whose total population was missing) of which about half are adjusted (set to 20 percent of the population) because of conditions (a) - (c).

The reliability of OPT depends, among other things, on the coverage of the preschoolers by Operation Timbang. A coverage of more than 100% is clearly not possible (if coverage is accurately measured) and too low a coverage, as was noted earlier, increases the likelihood of serious bias in the OPT.

In order to reduce somewhat the potential effect of such bias, the OPT of a municipality was accepted as adequately reliable if the coverage was between 20 and 98 percent. The upper limit was set to accomodate a logical requirement that coverage was extremely unlikely to be just about perfect although a few cases with quite reliable OPT may have been inadvertently omitted because the calculation of coverage itself was not necessarily very precise. The lower limit of the coverage range is set at 20 percent for a different reason. Both the mean and standard deviation of OPT were calculated for groups of municipalities identified by coverage categories 0-10 percent, etc. (Figure 1). If coverage and OPT were unrelated, the to mean and standard deviation of OPT would be about the same in all ranges of coverage. On the other hand if, as we argued earlier, low coverage would tend to involve large biases in OPT (high variability) and an overall upward bias of OPT (high mean).

Figure 1 indicates that except in the two categories with lowest coverage where standard deviation of OPT is high, the variability is relatively the same. Similarly with increasing coverage the mean of OPT gradually declines, i.e. the upward bias is reduced with increasing coverage. All this appears to mean that at very low levels of coverage (below 20 percent) bias is likely to be too serious a problem to ignore. It would probably have been safer to choose a higher lower

limit for coverage but the cost of doing so is a reduction in the number of municipalities with "reliable" OPT. Several other limits were also tried in the analysis and a lower limit of 20 percent seemed like a reasonable compromise between reliability and quantity of data.

2.4 Methods of Analysis

A correlation matrix provided a preliminary understanding of associations among variables. Discriminant analysis (see e.g. Klecka 1980) is used next as the basic method of classifying municipalities without OPT, or with unreliable OPT, into one of three categories: worst affected, moderately affected and least affected. Alternative definitions of these categories are considered to produce a nationwide municipal-level mapping of malnutrition and to prepare two lists of worst affected municipalities, a "short" and a "long" one, which were of immediate interest to the NNC. Two prevalence cutoffs of 40 percent and 45 percent have been used, in conjunction with some socio-economic constraints, to identify worst affected municipalities.

The basic elements of applying the discriminant technique are as follows:

(a) Municipalities with accepted OPT data (i.e. those with coverage of 20-98 percent) have to be categorized by their nutritional status. Several alternatives were explored as to the number of categories and how they are delineated. Those reported here involve three categories of worst affected, moderately affected and least affected municipalities. Least affected municipalities are those with OPT of less than 20 percent. Worst affected municipalities are alternatively defined as those with more than 45 or 40 percent prevalence of malnutrition. These two cut offs are about what are needed to identify the desired number of worst affected municipalities as will be seen later.

(b) The discriminating variables are the five socio-economic variables noted earlier. The discriminant technique linearly combines these variables such that the discriminating functions derived are best able to tell apart the three prevalence categories involved.

All those cases with known prevalence category and complete socio-economic information are used in the analysis to establish these relationships which are then used to predict the prevalence categories of unknown cases given information on their socio-economic characteristics. All the observations with any missing information are ignored in the analytical stage when the discriminant model is developed. In the classification stage the missing values of socio-economic variables are set equal to the averages calculated from all available values to predict the prevalence category of unknown cases.

(c) A large number of cases with reliable OPT are excluded from the analytical stage for lack of complete data on socio-economic variables. To reduce this waste of information and, as it turns out, to improve the predictive ability of the discriminant model, at least insofar as the worst affected group is concerned, the cases with reliable OPT but with missing data on any socio-economic variable have had each of the missing values replaced with the average of available values of the same variable in their specific prevalence group. For example if the number of physicians is missing in a municipality that belongs to the least affected group, this missing value is replaced by the average of the known values of the number of physicians in the least affected group. This procedure is not of much consequence in the case of the first four socio-economic variables as their missing cases are few. For the number of physicians however there are several hundred missing values which are filled in by the constructed averages. This idea is a slight refinement of the procedure that is available in SPSS package (Nie et al. 1975) and results in a certain amount of improvement as will be discussed later.

(d) The SPSS subprogram DISCRIMINANT was used to perform the analysis. The stepwise selection method minimized Wilks' lambda with equal prior probalities (Nie et al. 1975). The stepwise inclusion of discriminating variables identifies the optimal set of variables which are entered one at a time depending on which variable not yet

entered has the lowest value for Wilks' lambda, an inverse statistic which declines as the differences between categories and the homogeneity within each increase. The probability that each municipality is classified into a given category by the model can be influenced by setting different prior probabilities making it more likely, a priori, for the observation to belong to one category rather than another. Prior probabilities for each of the 3 nutritional categories are assumed to be equal.

(e) The classification results of the estimated discriminant models are compared for their sensitivity, specificity and positive predictive value. The classification results of each discriminant model (occasionally modified as will be described below) are summarized in a 2 x 2 table of the following general form:

		Actually in category i? Yes	No
Predicted to belong to category i	Yes	a	b
	No	c	d

where, i = worst affected or least affected

Se = sensitivity = a/ (a + c) = percentage of cases in category i which are so predicted correctly.

Sp = specificity = d/ (b + d) = percentage of cases not in category i which are so predicted correctly.

V^+ = positive predictive value = a/ (a + b) = percentage of cases predicted to belong to category i that actually do belong to this category.

(f) The main models have a strong tendency to misclassify municipalities into the two extreme prevalence groups. This tendency is partly due to the fact that the missing socio-economic data on municipalities with reliable OPT are filled in (see (b) above). To counter this tendency and reduce the number of municipalities in the least and worst affected groups, certain modifications are necessary. The

definitions of worst and least affected municipalities could be made more restrictive by choosing, say, 50 and 15 percent as cutoffs. An alternative is to impose certain constraints on the classification results of the discriminent models to trim the number of municipalities predicted to belong to one or the other of the extreme groups. We will later describe why the latter approach is preferable. The constraints involve three of the five socio-economic variables and will be similarly explained later.

The trimmed lists of least affected and worst affected municipalities are used to produce a nationwide municipal-level mapping of nutritional status in the Philippines in 1979. A similar nationwide provincial mapping is also produced by classifying provinces into one of six grcups depending on the percentages of least affected and worst affected municipalities in each province.

3.0 RESULTS

3.1 Association between Variables

Table 2 shows the correlation matrix of the variables used in the analysis. Most of the coefficients are statistically significant ($p < 0.01$) but the largest magnitude does not exceed 0.30. The signs of significant coefficients seem to be correct given the usual expectation that high malnutrition (high OPT) tends to be associated with poor housing (high COGONIPA) and poor water supply system (high WATWKS). OPT is significantly correlated with only two of the five socio-economic variables, housing condition and number of physician per 10,000 population. This fact is somewhat unexpected and presages considerable difficulties in the analysis since the discriminating power of the analysis derives primarily from the strength of association between OPT and socio-economic variables. The correlation coefficient between coverage and OPT is significant but relatively small. The correlation coefficients between each pair of the socio-economic variables are always significant, but their values are not particularly high ranging from about 0.07 to 0.30 in magnitude. High percentages of rice area irrigated, population density and number of physicians are favorable factors associated positively with one another and each is

negatively associated with the percentage of poor housing and of barangays without waterworks.

3.2 Effect of Substituting for Missing Socio-Economic Data

There were 815 municipalities whose OPT data were accepted as reliable on account of their 20-98 percent coverage. The nutritional status category of the remaining 683 municipalities must be predicted. As the relevance of the results for targetting purposes was considered crucial our interest focusses more on the correct identification of the worst affected municipalities than of all municipalities. The latter however is needed for the purpose of mapping malnutrition.

Table 3 presents the classification results of two discriminant models. In both models worst affected municipalities refer to those with more than 45 percent prevalence, moderately - affected municipalities are those with 20-45 percent prevalence and least affected municipalities have less than 20 percent prevalence. Each of the 815 municipalities is placed into one of these categories depending on its OPT value.

The first model relies on 348 of 815 observations which have no missing socio-economic data. The second model uses all 815 cases after missing socio-economic information is "estimated" and replaced. The estimation involves 15 average values. These average values pertain to each socio-economic variable (of which there are 5) within each nutritional category (of which there are 3). For example, within the category of worst affected municipalities (those defined by a prevalence of, say, 45 percent or more) the average number of physicians is calculated using all non-missing values of this variable. All missing values for this variable in the worst affected category are then substituted for by this average. This has been done for all 15 variable-cum-category pairs. The idea of substituting these averages for missing values is a slight refinement of the usual practice of using overall averages for such missing data in the classification of ungrouped cases in package programs (see, e.g. Nie et al. 1975, p. 456, Option 2).

The classification results of these two models are presented in Table 3. Note that in Panel A (where only the 348 cases without any missing socio-economic information are used) the proportion of correctly classified cases is

only 36 percent vs 35 percent in Panel B (where all 815 cases with acceptable OPT have been used after substitution for missing socio-economic information). Hence neither of these two models appears to do a very good job of overall classification. Note however that the bulk of municipalities wrongly classified as worst or least affected come from the group which is actually only moderately affected not from the actually least or worst affected groups. Hence when a wrong prediction is made it is usually not as bad as it could be. This poor overall classification is a direct result of the small correlation between OPT on the one hand and socio-economic variables on the other. There is however a very noticeable difference between the results of the two models: the first correctly picks out one of every two worst affected municipalities (48%) whereas the second does the same three out of four times (75%). Since we are particularly interested in correctly identifying the worst affected municipalities, this difference is fairly important. The second model also has a slight advantage in correctly identifying the least affected municipalities (47% vs 42%). The sensitivity and positive predictive value of the second model also exceed those of the first (these quantities are derived from 2 x 2 tables produced from consolidation of moderately affected and least affected categories). The justification for substitution for missing information derives soley from the empirical superiority of the second model.

3.3 Listings of Worst Affected Municipalities

Because of the improvement that results from substitution for missing socio-economic information, henceforth only models employing such substitution will be discussed. The second model has a strong tendency to misclassify the moderately-affected municipalities into the two extreme groups. (A similar finding on this feature of the discriminant technique is reported in Arnold et al. 1981). This is of course of little import insofar as the 815 grouped cases are concerned: their actual breakdown is already available. It does indicate however that many of the municipalities predicted to be worst affected are, in fact, only moderately affected. Of the 683 unknown cases 246 were predicted to be worst affected which together with 66 known to be in this category constitute a list of 312 worst affected muncipalities. This tendency to enlarge the extreme groups is largely responsible for the low positive predictive value of the model (16%).

The unknown positive predictive value for these ungrouped cases is likely to be still less hence in the list of 312 municipalities there may be no more than (66 + 0.16 x 246 =) 105 worst affected cases.

The NNC requested a "short" and a "long" listing of the worst affected municipalities containing approximately 10 and 20 percent of all municipalities, i.e. about 150 and 300, respectively. For the latter, our listing of 312 actual and predicted worst affected cases is one option. But as we noted above, about two thirds of the cases in this list are probably misclassified by being predicted as worst affected.

Given the low association between socio-economic characteristics and nutritional status this result is largely inevitable but some additional improvements can be made in the light of the fact that nutritional situation in a municipality is only one consideration in directing development resources into an area. Others include the general socio-economic conditions of which the five characteristics selected in this study constitute a small portion. It would be preferable to also try to distinguish between cases predicted to be nutritionally worst affected by their state of socio-economic deprivation. Socio-economic characteristics are employed here to identify nutritionally worst affected municipalities but these characteristics can also be employed in their own right, not just as predictors of nutritional status. It is assumed that when two municipalities are both predicted to be worst affected, the one with inferior socio-economic conditions should, take precedence over the other for targetting of development programs. This implies that the list of predicted cases of worst affected municipalities can be trimmed down by comparing socio-economic characteristics. Hence an alternative procedure to come up with a long list of about 300 worst affected municipalities would be to produce a provisional list which is larger but then reduce it in size by retaining those that are relatively worse off from a socio-economic standpoint.

A provisional listing of a larger number of worst affected cases can be produced by reducing the cut off of 45 percent to, say, 40 percent. The resulting cases predicted to be worst affected can then be reduced by keeping those which are socio-economically relatively more deprived. Similarly, to produce a shorter list of about 10 percent of all municipalities (about 150) it is possible to begin with the list of 312 "worst affected" cases (already produced with a cut off of 45

percent) and reduce it by leaving out those which are relatively better off on socio-economic grounds.

This latter possibility was then considered by starting with the provisional listing of 312 municipalities and keeping those that are more disadvantaged to produce a "short" list of worst affected cases. Out of the five socio economic characteristics employed we focus on three which are more likely to better represent socio-economic conditions: housing condition, water system and the number of physicians. Three constraints are imposed but only on the predicted cases since it is among these that there are a large number of misclassified cases, cases which are in fact not nutritionally worst affected. A municipality was retained in the list of worst affected municipalities if it satisfied all of the following conditions:

(i) the percentage of houses with cogon and nipa roofing is more than average for all municipalities for which this variable is available;

(ii) the percentage of barangays without waterworks is more than an average calculated as in (i);

(iii) the number of physicians per 10,000 population is less than a similarly calculated average.

No municipality was excluded for lack of information on socio-economic variables. These constraints cut down the list of predicted worst affected cases from 246 to 104 which when added to the 66 known cases constitute a group of 170 worst affected municipalities.

Apart from the justification already advanced for employing socio-economic characteristics not just as predictors but also in their own right in order to produce a list of worst affected cases, we can also examine whether these constraints when applied to cases with known nutritional status, are likely to improve the selection of nutritionally worst affected cases as compared with the results of discriminant analysis. This aspect is examined in Table 4, Panels A and B by comparison of various indicators. Panel A is the collapsed version of Table 3, Panel B. The imposition of the constraints reduces sensitivity (by about half from 76 to 32 percent) and increases specificity (by a third from 66 to 88 percent) as it must, but the positive predictive value is raised somewhat from 16 to 19 percent. This improvement comes at the cost of having a good many worst affected cases excluded (because of low sensitivity) but this is inevitable. If a

list is to be short, a large number of truly worst affected cases will inevitably be excluded and, for a given size, the maximization of positive predictive value becomes the desired objective. Note that the primary consideration in screening the results of discriminant analysis through imposition of socio-economic constraints is the fact that by this approach more deprived municipalities can be identified. That the constraints also improve the proportion of nutritionally worst affected cases, if only slightly, is an added bonus but a secondary consideration.

Two options are available to come up with a "long" list of worst affected cases: first by keeping the 312 municipalities already identified or by reducing the cutoff from 45 to 40 percent and producing a larger listing but then trimming it down by imposing similar socio-economic constratins. The imposition of the three constraints mentioned above have basically the same effect regardless of the cutoff as can be seen from the comparison of indicators in Table 4, Panels A and B on the one hand and C and D on the other. The interest now, however, focusses on Panels A and D. With a cutoff of 45 percent and no constraints, the positive predictive value is only half as much as when the cutoff of 40 percent is supplemented with socio-economic constraints. Note also that the difference between the true and predicted proportions of worst affected cases in total is substantial in the former (A) and marginal in the latter (D). Another advantage of the second approach is that the municipalities known to have a prevalence of between 40-45 percent (i.e. only just below the former cut off of 45 percent) are certain to be included in the listing whereas with the first approach they are certain <u>not</u> to be included. This advantage is particularly desirable as the discriminant model does not have a very good predictive ability. If it did it would be more appropriate to define malnutrition more stringently (45 percent prevalence rather than 40 percent) so that those with high malnutrition would be more likely to be included in the list than those municipalities with 40-45% prevalence.

3.4 Mapping of Malnutrition

A list of worst affected municipalities is a useful starting point in deciding target areas for intervention. A list of least affected municipalities on the other hand can be useful in that it identifies areas where intervention may not be

presently needed, and allows a comparative study of possible reasons for nutritional differences between least and worst affected areas. Once these two lists are compiled and the rest of the municipalities in the country are identified as falling in between, i.e. moderately affected, a complete mapping of municipalities by nutritional status category can be had. A similar mapping can be produced at the provincial level by classifying provinces according to the proportions of worst and least affected municipalities they each have.

Least-affected municipalities are identified in a similar way that was employed to compile the "short" list of worst affected municipalities. Among the 815 muncipalities with reliable OPT, 99 are least affected, i.e. have a prevalence of second and third degree malnutrition below 20 percent. To find the municipalities predicted to be least affected the same model of discriminant analysis is used as the one which was employed to produce the list of 170 worst affected municipalities. Since the definition of a least affected municipality is the same in the two discriminant models (in which cut-offs of 45 and 40 percent were set to define worst affected cases) it matters little which model is adopted. This can be seen from the comparison of (A) and (C) in Table 5. The sensitivities, specificities and positive predictive values are about equal. It is also obvious that the model under discussion does a better job of predicting the worst affected cases than the least affected ones as the sensitivity in Table 4 (A) is considerably higher than that in Table 5 (A), 76 vs 46 percent. Other comparable statistics are about the same. The model identifies 249 of the 683 ungrouped cases to be least affected (Table 3 (B)). This number is cut down by eliminating any municipality that does not fail all three of the constraints used to trim the list of cases predicted to be worst affected. 170 cases survive which along with known cases comprise a list of 269 least affected municipalities. This elimination process, if applied to the 99 cases known to be least affected, reduces sensitivity and improves specificity (as it must) but improves the positive predictive value from 16 to 18 percent (Table 5 (A and B)).

A mapping of malnutrition by municipality is given in Figure 1 for the Philippines. Only the worst and least affected municipalities are marked on the map. The rest are moderately affected. Figure 2 provides a similar mapping at the provincial levels. There are six categories of provinces defined depending on

the proportion of least and worst affected cases in total number of municipalities in each province. These are as follows:

Province Group	Percentage of Municipalities: Least affected	Percentage of Municipalities: Worst affected	No. of Provinces
Least affected	≥ 25		24
Moderately least affected	$\geq 15, < 25$	< 15	6
Middle with more even dist.	< 15	< 15	24
Middle with less even dist.	$\geq 15, < 25$	$\geq 15, < 25$	2
Moderately worst affected	< 15	$\geq 15, < 25$	9
Worst affected		≥ 25	8
Total	**16**	**10**	**73**

Some conclusions stand out from these municipal and provincial maps.

(i) The bulk of the least affected municipalities and provinces are in the northern part of the country (Regions I-IV). There is only one worst affected province in this part and only about 6 percent of the municipalities in these four regions are worst affected. In general, these municipalities are on the eastern side of the area.

(ii) All provinces surrounding Manila and Manila Bay are least affected. More interesting is the fact that there are only two municipalities in these provinces that have been found to be worst affected whereas there are about 80 that are least affected.

(iii) Regions VI-IX in the central part of the country are unquestionably the worst-off. There are very few least affected municipalities or provinces in these regions. The former are also relatively scattered.

(iv) The southern part of the country (Regions X-XII) is a mixed bag. There are a few least affected provinces although their concentration is nothing like that in the northern part. There are few worst affected provinces in these regions. Most are moderately affected.

4.0 DISCUSSION

The analysis reported in this paper was designed to demonstrate some possible uses to which the existing data in the Philippines could be put. The principal use of such outputs is in identifying target areas for possible intervention but they also could serve as baseline data in depicting and

evaluating change in nutritional situation over time when similar outputs become available in the future. The major drawback of the specific results of this analysis however should not be ignored, namely, the insufficient confidence in their accuracy. While attempts have been made to minimize the strain on the data, the legitimacy of certain convenient assumptions which were made may be questioned. In the light of the need for increased accuracy and the usefulness of such outputs efforts must be directed towards producing more reliable raw data and creating the mechanisms that promote systematic production and use of similar outputs. In the following we will discuss methods and findings and suggest certain steps which may be taken to improve the generation of reliable and useful outputs.

4.1 Methods

The use of the discriminant technique was prompted by two characteristics of the available data: incompleteness of data in general and inadequate reliability of prevalence data in particular. In order to predict nutritional status in municipalities without such information a number of statistical techniques were available (e.g. regression methods) but the advantage of the discriminant analysis was that it was the least demanding as regards the accuracy of OPT data. Only when errors in OPT led to incorrect classification were the results affected. The exact value of OPT was irrelevant as long as measurement errors and biases were such that no misclassification resulted. Lack of high correlation between socio-economic variables and nutritional status diminishes considerably the confidence in final results but this is not an unusual situation. Improvement in modeling can only come from selection of more appropriate correlates of malnutrition but the scope for such improvement is fairly limited. Substantial increase in the reliability of these types of surveillance outputs can come only from more and better nutritional data. The modeling process is only the second best solution.

4.2 The Future

We have noted some of the sources of bias in the OPT data and suggested that instead of trying to remove these biases (essentially by increasing coverage) it may be more efficient to look elsewhere for data pertaining to malnutrition.

This would be feasible if the purpose is to find the extent of malnutrition rather than identify the malnourished as is the main objective of OPT. Hence the need for OPT as a screening device remains. However alternative indicators are needed to gauge the extent of malnutrition. It has been suggested that height-for-age (HA) of first graders may be an appropriate indicator. HA is an indicator of past nutritional conditions and can be collected for a large, appropriately selected sample of first graders. That more than 90 percent of eligible first graders go to school and that height, age, and other data can be conscientiously collected by teachers is an added incentive. Furthermore, as screening is no more the objective, a sample survey would do instead of a census. Concentrating on a sample also allows resources to be expended on covering as high a level of response as possible. If a properly designed and large enough sample is selected, statistical techniques can be employed to make inferences about the nutritional situation more confidently than would be possible with OPT. Although likely to be more accurate and less costly it should be noted that these advantages come at a cost. The cost is that no screening can be made of and hence no intervention can be undertaken to alleviate the conditions of individuals except insofar as they tend to constitute identifiable needy groups. Another important problem has to do with the anticipation of alternative policies available. HA data on a sample of first graders can provide very accurate measures of nutritional problem by region, province, school, barangay, etc. which can be of considerable value in targetting. There are however other considerations such as more focussed targetting and an understanding of the possible underlying causality which might help improve the selection of interventions. Hence grouping by socio-economic characteristics (functional classification) assumes particular importance. In the early 1982 the National Nutrition Council launched a nationwide survey of primary schools to collect information on sex, birth date, height, weight and father's occupation of first graders. The survey covers some 9500 schools (about one third of all primary schools in the country) from all of approximately 1900 districts and 1650 towns and cities. The total number of children sampled would exceed 200,000 with at least 100 in each district to allow statistically valid inferences to be made at school district level. (A school district is often the same administrative area as a municipality. Such data will provide in the near future far more accurate information on nutritional status in the philippines at moderate cost and with sufficient degree of disaggregation to permit planning at small administrative levels.

REFERENCES

Arnold, J.C., Engel, R.W., Aguillon, D.B., and Caedo, M.M. 1981. "Utilization of Family Characteristics in Nutritional Classification of Preschool Children. American Journal of Clinical Nutrition 34, pp. 2546-2550.

Food and Nutrition Research Institute. 1980. First Nationwide Nutrition Survey: Philippines, 1978. Manila: FNRI Publication No. GP-11, 2nd Revision.

Klecka, W.R. 1980. Discriminant Analysis. Sage University Paper series on Quantitative Application in the Social Sciences, 07-001. Beverely Hills and London: Sage Publications.

Mason J.B., Habicht, J-P., Tabatabai, H. and Valverde, V. 1982. Nutritional Surveillance. Cornell Nutritional Surveillance Program. Draft.

National Nutrition Council, The Philippines/Cornell Nutritional Surveillance Program. 1981. Philippine Nutrition Surveillance Project. Draft.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. 1975. Statistical Package for the Social Sciences, Second Edition. New York: McGraw-Hill.

United Nations, Department of International Economic and Social Affairs, Statistical Office. 1980. Demographic Yearbook 1979. New York: United Nations.

TABLE 1

VARIABLES AND SOURCES OF INFORMATION

	Variable	Label	Number of Observations
	Identification codes		
1.	Region	REG	
2.	Province	PROV	
3.	Municipality	MUNI	
	Demographic		
4.	Total population (1978)	TOTPOP	1493
5.	Total number of preschoolers	TOTPRE	919
6.	Proportion of preschoolers in population		919
7.	Total number of preschoolers (adjusted)[c]		968
	Nutritional status		
8.	Number of normal cases (90%)[d]	TOTNOM	973
9.	Number of mildly malnourished (75-90%)[d]	TOTMIL	973
10.	Number of moderately malnourished (60-75%)[d]	TOTMOD	973
11.	Number of severely malnourished (60%)[d]	TOTSEV	973
12.	Total number of weighed preschoolers	TOTWH	973
*13.	Prevalence of Malnutrition (%)	OPT	973
	Descriptive		
*14.	Coverage (%)	COVERAGE	919
*15.	Population density (persons/sq. km)	POP DENSITY	1492
*16.	Percent of rice area irrigated	IRR RICE	1461
*17.	Percent of dwellings with cogon and nipa roofing material	COGONIPA	1491
18.	Number of barangays (villages)	NUMBAR	1481
19.	Number of barangays not served by local waterworks		1232
*20.	Percent of barangays not served by local waterworks	WATWKS	1231
21.	Total number of physicians		610
*22.	Number of physicians per 10,000 population	PHYSICIANS	610

* Variables used in the analysis.

a Municipal Nutrition Action Plan.

b Ministry of Human Settlements, Settlement Profile: Final Edition, 1978 Update, 1980.

c For municipalities with OPT

d Weight for age as percent of Philippine standards

TABLE 2

CORRELATION MATRIX

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.	OPT	1.00						
2.	COVERAGE[a]	-0.17* (883)	1.00					
3.	POP. DENSITY	-0.01 (967)	-0.03 (882)	1.00				
4.	IRR. RICE	-0.02 (956)	0.04 (872)	0.07* (1460)	1.00			
5.	COGONIPA	0.14* (968)	-0.13* (883)	-0.30* (1490)	-0.10* (1459)	1.00		
6.	WATWKS	0.03 (823)	-0.02 (758)	-0.18* (1230)	-0.10* (1216)	0.18* (1230)	1.00	
7.	PHYSICIANS	-0.12* (413)	0.00 (378)	0.20* (610)	0.13* (609)	-0.20* (608)	-0.15* (587)	1.00

Note: Number of observations in parentheses.

[a] Observations with more than 100 percent coverage are excluded.

* $P < 0.01$

TABLE 3

RESULTS OF DISCRIMINANT ANALYSIS

(A) Without substitution for missing socio-economic data (N = 348)

Actual group membership	No. of cases	Predicted group membership			Summary Statistics
		Worst affected	Moderately affected	Least affected	
Worst affected	66	32 (48%)	18 (27%)	16 (24%)	C = 36%
Moderately affected	650	222 (34%)	219 (34%)	209 (32%)	Se= 48%
Least affected	99	27 (27%)	30 (30%)	42 (42%)	Sp= 67%
Ungrouped cases	683	216 (32%)	215 (31%)	252 (37%)	V+= 11%
					P = 8%
					p = 34%

(B) With substitution for missing socio-economic data (N = 815)

Actual group membership	No. of cases	Predicted group membership			Summary Statistics
		Worst affected	Moderately affected	Least affected	
Worst affected	66	50 (76%)	8 (12%)	8 (12%)	C = 35%
Moderately affected	650	236 (36%)	187 (29%)	227 (35%)	Se= 76%
Least affected	99	22 (22%)	31 (31%)	46 (47%)	Sp= 66%
Ungrouped cases	683	246 (36%)	188 (28%)	249 (36%)	V+= 16%
					P = 8%
					p = 38%

Note: Definition of Symbols:
C = proportion of grouped cases correctly classified;
Se= sensitivity;
Sp= specificity;
V+= positive predictive value;
P = actual prevalence for grouped cases;
p = estimated prevalence for all cases.

[a] These statistics , except C, are derived from a 2 x 2 table which consolidates moderately affected and least affected groups into a non-worst affected or "other" group. An example is Table 4, Panel A, which is based on Panel B above.

TABLE 4

COMPARISON OF ALTERNATIVE PROCEDURES FOR SELECTION OF WORST AFFECTED MUNICIPALITIES

(A) Worst affected defined as prevalence of > 45%; without constraints a/

		Actually worst affected?		
		Yes	No	Total
Predicted worst affected?	Yes	50	258	308
	No	16	491	507
	Total	66	749	815

Se = 76%

Sp = 66%

V+ = 16%

P = 8 %

p = 38%

(B) Worst affected defined as prevalence of > 45%; with constraints

		Actually worst affected?		
		Yes	No	Total
Predicted worst affected?	Yes	21	87	108
	No	45	662	707
	Total	66	749	815

Se = 32%

Sp = 88%

V+ = 19%

P = 8%

p = 13%

(C) Worst affected defined as prevalence of > 40%; without constraints

		Actually worst affected?		
		Yes	No	Total
Predicted worst affected?	Yes	107	257	364
	No	33	418	451
	Total	140	675	815

Se = 76%

Sp = 62%

V+ = 29%

P = 17%

p = 45%

(D) Worst affected defined as prevalence of > 40%; with constraints

		Actually worst affected?		
		Yes	No	Total
Predicted worst affected?	Yes	55	117	172
	No	85	558	643
	Total	140	675	815

Se = 39%

Sp = 83%

V+ = 32%

P = 17%

p = 21%

Note: Symbols as defined in Table 3.

a/ This is the consolidated form of Table 3, Panel B

TABLE 5

COMPARISON OF ALTERNATIVE PROCEDURE FOR SELECTION OF LEAST AFFECTED MUNICIPALITIES

(A) Worst affected defined as prevalence of > 45%; without constraints

		Actually least affected?		
		Yes	No	Total
Predicted least affected?	Yes	46	235	281
	No	53	481	534
	Total	99	716	815

Se = 46%

Sp = 67%

V+ = 16%

P = 12%

p = 34%

(B) Worst affected defined as prevalenc of > 45%; with constraints

		Actually least affected?		
		Yes	No	Total
Predicted least affected?	Yes	30	134	164
	No	69	582	651
	Total	99	716	815

Se = 30%

Sp = 81%

V+ = 18%

P = 12%

p = 20%

(C) Worst affected defined as prevalence of > 40%; without constraints

		Actually least affected?		
		Yes	No	Total
Predicted least affected?	Yes	46	245	291
	No	53	471	524
	Total	99	716	815

Se = 46%

Sp = 66%

V+ = 16%

P = 12%

p = 36%

Note: Symbols as defined in Table 3.

FIGURE 1

DISTRIBUTION OF STANDARD DEVIATIONS OF PREVALENCE
OF MODERATELY AND SEVERELY UNDERWEIGHT CHILDREN BY COVERAGE

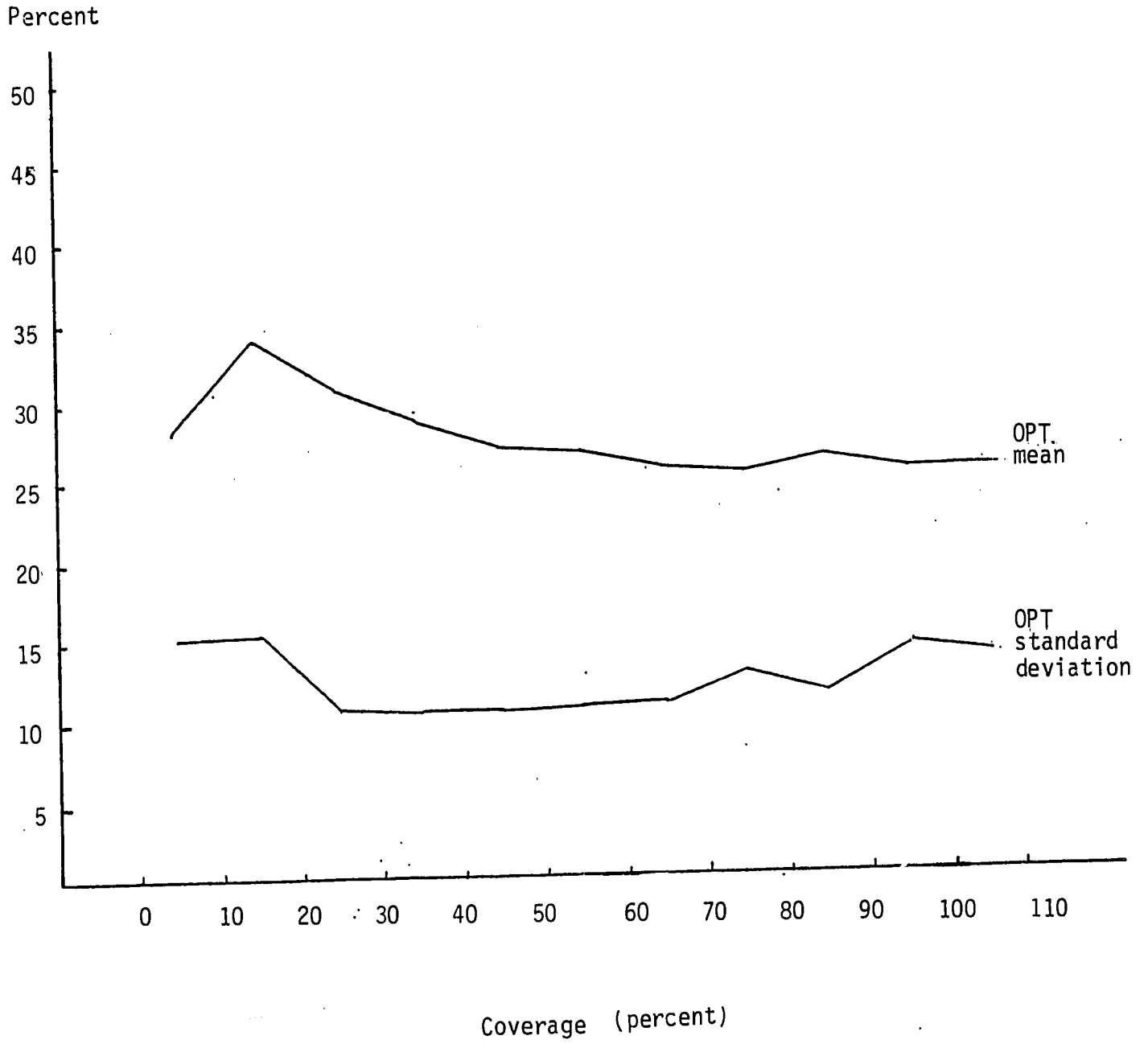
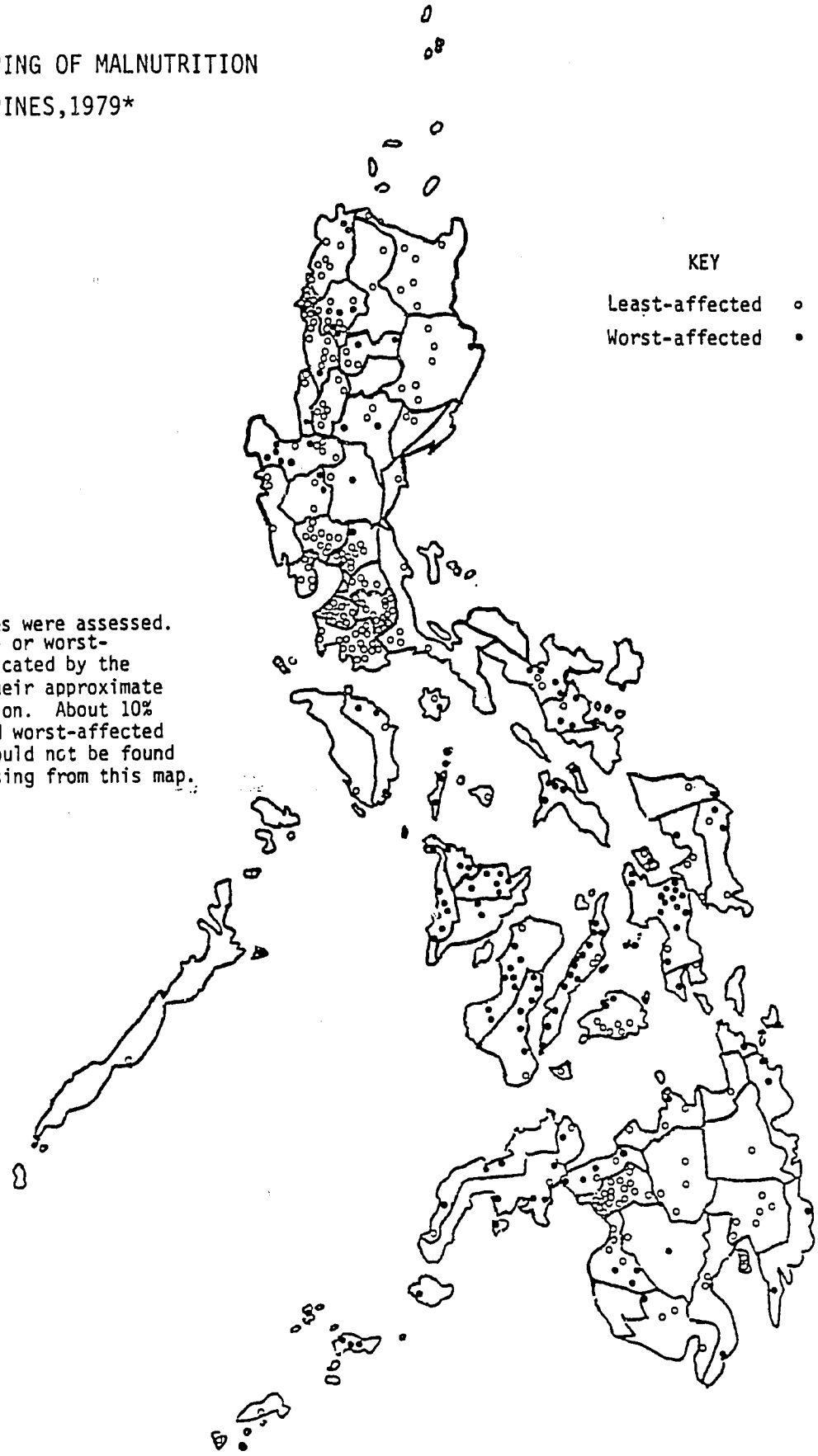


FIGURE 2

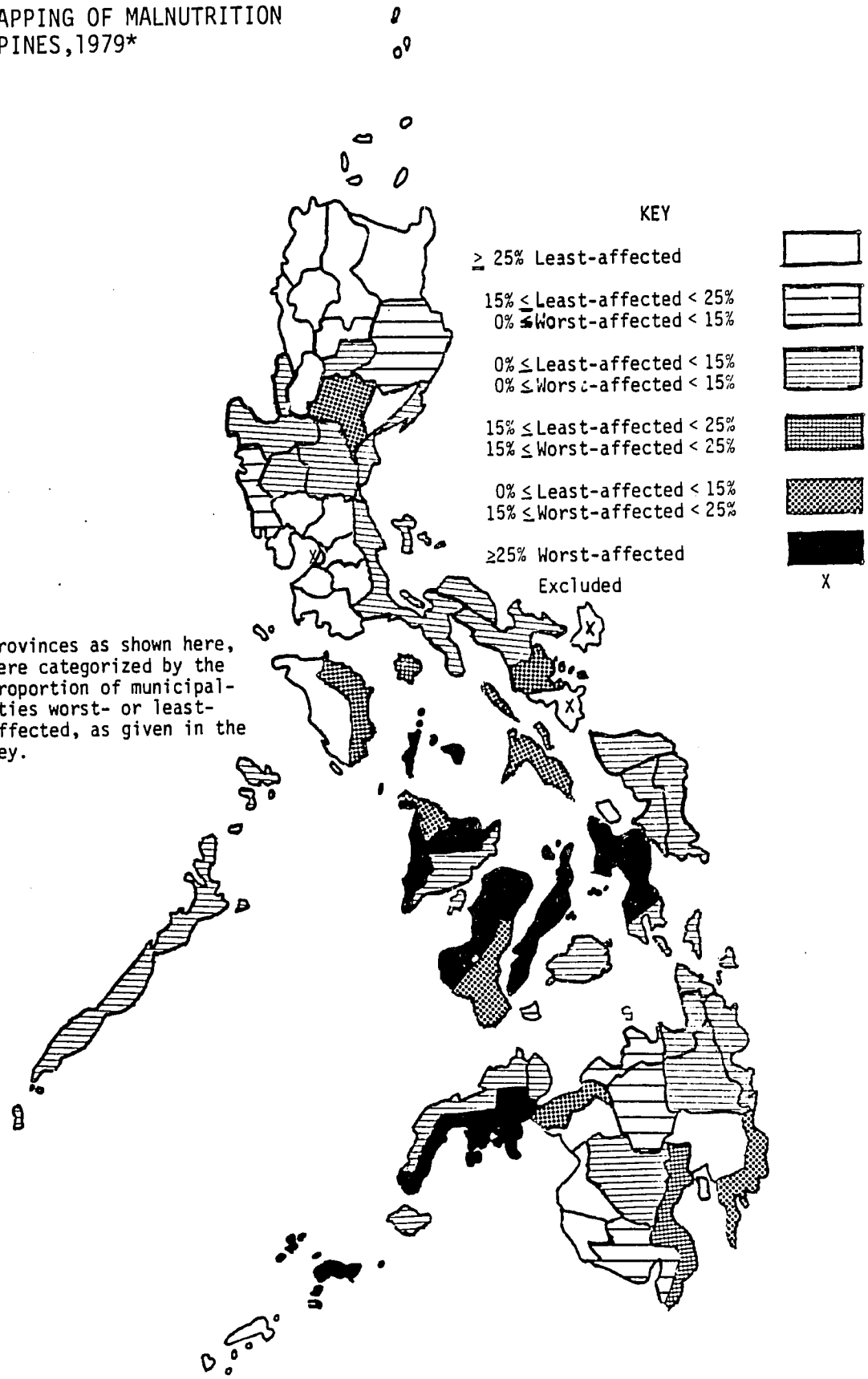
MUNICIPAL LEVEL MAPPING OF MALNUTRITION IN THE PHILIPPINES, 1979*



*All municipalities were assessed. Only those least- or worst-affected are indicated by the dots placed on their approximate geographic location. About 10% of the least- and worst-affected municipalities could not be found and are thus missing from this map.

FIGURE 3

PROVINCIAL LEVEL MAPPING OF MALNUTRITION IN THE PHILIPPINES, 1979*



*Provinces as shown here, were categorized by the proportion of municipalities worst- or least-affected, as given in the Key.

CNSP WORKING PAPER SERIES

Currently available:

1. Nutritional Surveillance: A synopsis. June 1981.
2. Nutritional Surveillance. August 1981.
3. Surveillance Summaries. January 1982.
4. Analytical Methods for Nutritional and Socio-economic Data. April 1983.
5. Principles for Evaluation of On-going Programs. May 1982.
6. An Outline Plan for Developing a Nutritional Surveillance System in New York State. May 1982.
7. Initial Assessment for Nutritional Surveillance in the Philippines. Forthcoming.
8. Preliminary Analysis of Nutritional and Agricultural Data, Eastern Province, Kenya. August 1982.
9. Static and Dynamic Poverty Indicators: An Indonesian Example. May 1983.
10. Reducing Malnutrition Through Rural Development: A test of Planning Methods in Haiti. June 1983.
12. Malnutrition by Socio-Economic Classification in Palawan, the Philippines. June 1983.
13. Changes in Nutritional Status in the Philippines from 1979 to 1980. Forthcoming.
14. National Health Information Systems and their use in Nutritional Surveillance. January 1983.
15. On Sampling Variance of the Sensitivity of an Indicator. January 1983.
16. Note on a Framework for Monitoring and Evaluation of UNICEF/WHO Nutrition Improvement Programmes. March 1983.

Additional Documents Available from CNSP:

Report of Workshop on Social and Nutritional Surveillance in Eastern & Southern Africa. May 17-19, 1982.

Background Papers for Workshop on Social and Nutritional Surveillance in Eastern & Southern Africa. May 17-19, 1982.

6/83