# SAMPLING STRATEGY FOR THE 1991 PHILIPPINES CENSUS OF AGRICULTURE

**by**

**I. P. David** [1]

## I. INTRODUCTION

The Philippines' latest Census of Population and Housing (CPH) was done in May 1990. It required a special appropriation of ₱ 215 million, which Congress approved readily the year before. In 1990 the National Statistics Office (NSO) again made a request for ₱ 207 million to conduct the Census of Agriculture and Fisheries (CAF) in early 1991, with 1990 calendar year as reference period. This time the request was turned down before first reading. Congress' main motivation could well be the growing public deficit which needed to be tamed, abetted perhaps by the knowledge that there is hardly little correlation between statistical activities and votes. There could be lessons here for statisticians.

First, setting the two censuses one year apart may not be a good idea. The large CPH budget must have been fresh still in the collective mind of Congress when it was presented again with the CAF request. Moreover, with their longer history (linked in fact with the origin of the word statistics), the reasons for doing population censuses are more widely known and accepted. Even politicians know that population counts are used in reapportioning congressional seats and redistricting, as well as in distributing public funds. On the other hand, it seems that the importance of reliable agricultural statistics in an agricultural economy like ours is not evident to most politicians -- perhaps understandably so since their background is rarely agrarian. Thus, it is interesting to speculate whether it would be better to do CAF first before CPH, or set them farther apart, say five years. The latter makes

---

[1] Senior Statistician, Asian Development Bank, Manila, January 1993. The views expressed here are mine and not those of ADB. I sincerely thank Administrator Tomas Africa and Josie Perez, Chief of the Agriculture and Fisheries Division, NSO, for the opportunity to work with them; Bruce Tolentino, who set the stage for my involvement in this work; and A. Nenette Mendoza for the first-rate computing assistance. Any errors are, of course, all mine.

sense also from a logistics standpoint, since present laws require the same agency (NSO) to conduct the censuses.

Until 1970 one reason for doing CPH first and CAF the following year was that complete lists of farming/fishing households were compiled as rider questionnaires during the CPH, which were then used as frames for the CAF. This practice was stopped in 1980 when separate listing operations were conducted to construct frames for the CAF; however, the two censuses were still scheduled one after the other following the same sequence.

Second, the original plan for the 1990 CAF called for a complete enumeration of all farming/fishing households in the country. This meant a large proposed outlay which might have attracted the attention of a Congress out to trim the national budget. To many, the word census is still synonymous with complete enumeration. The reason is partly historical inertia: complete enumeration censuses were being done long before the advent of modern scientific sampling and indeed of statistical science. For want of viable alternative methods of studying populations, complete enumeration censuses were a necessity. This may have been alright up to a time when populations were relatively small and societies were less complicated than now. However, populations exploded,[2] large sets of variables and complex models are required to just approximate population behavior, and a revolution in data processing technology has whetted the appetite for more and more timely information. Advances in Probability and Statistics, particularly in sampling theory and methods and inference from samples to populations, made possible the study of populations through observations from small samples. These parallel developments provide the means by which equally reliable and useful information can be obtained without resorting to resource intensive complete enumeration censuses.

Due to less support for CAF than CPH in general, the need for extensive use of sampling to

---

2

reduce costs is more acute in the former. The move is more compelling for the developing countries which have large agrarian sectors, very small farm sizes and hence vast numbers of farm households or census units, compared with developed countries where farm sizes are increasing but the number of farms or census units have been declining. The proportion of farm households to total number of households in developing countries is usually above 50 per cent, while 3-6 per cent is typical in developed countries. Hence, complete enumeration agricultural censuses will more likely continue to be popular in industrial countries. On the other hand, a point has or may soon be reached in the developing countries where sampling would be the only recourse to bring agricultural censuses within the realm of economic and political feasibility. (See also footnote 6).

In January 1991, a small group of data users and producers who were concerned about the consequences of the recent scrapping of the CAF item from the budget bill met after office hours to discuss probable alternatives.[3] One scenario was to try raising some funds from non-government sources; since such amount could not possibly be anywhere near the original request, the logical question was whether some "substitute" to the CAF can be done, whose results would still satisfy the major uses intended for the latter. An obvious answer was to use sampling. This prompted other questions: Will the results still be useful as sampling frame for the succeeding intercensal sample surveys? If you sample villages (barangays), will the sampling error of the town estimates be acceptably low? These questions and others that followed made it plain that, while on the one hand we were presented with a golden opportunity to move away from massive, resource intensive censuses, on the other hand some convincing based on hard, empirical evidence needed to be done before a sample census quite unlike any that had been tried before would stand a good chance of being accepted.

This paper is an account of some of the work that went into the search for an acceptable

[3] The group consisted of the Bruce Tolentino, then Undersecretary, Department of Agriculture; Administrator Tomas Africa, NSO; Mamerto Damasco, then Director (now retired), Bureau of Agricultural Statistics; and the author. Later meetings were joined by Gener de Guzman, former Director, BAS and Nelia Marquez, Deputy Administrator, NSO.

sampling strategy (sampling scheme and estimation procedure) for the Census of Agriculture, **the reference period of which had to be moved from 1990 to calendar year 1991 because of lack of funding.** *We say Census of Agriculture (CA) because our work was based on the assumption that the coverage would exclude fisheries, as revised plan called for a separate Census of Fisheries. However, a decision was made later to use the same sample to collect data on fisheries, hence the undertaking eventually came to be called 1991 CAF. This piece of information is important, so as not to pass judgment on the sample on the basis of the outcome of the fisheries part of the census.*

## II. BACKGROUND

### A. Use of Sampling in Past Censuses

To place what we are about to propose to do in proper perspective, we review briefly the sampling employed in previous CPHs and CAFs. There had always been complete headcounts in the CPHs; hence the households (census units) and villages had to be enumerated completely. A *short questionnaire* requiring names, sex, age and a few more characteristics of each household member was filled out for every household. Population counts (total, male, female) were published for every village (there were 41,293 villages or barangays including those in the Metropolitan Manila during the 1990 CPH). A *long questionnaire* covering more detailed demographic and housing characteristics was used on a sample of households chosen systematically from every village. The villages were still enumerated completely.

Censuses of agriculture were conducted in 1960, 1971 and 1980. The last two also had fisheries components. These were conducted in two phases. All households were interviewed and listed during phase 1 to group them into farm(fishing) and non-farm categories and to stratify the first group further into large and small farm households according to some definition. The large farm households were enumerated completely during the main census operation (phase 2); however, simple

systematic sampling in the small farm households group was resorted to in all the censuses, with the sampling rate declining continuously, namely 33 per cent in 1960, 20 per cent in 1971 and 10 per cent in 1980. (It is tempting to speculate that these reductions in sampling rates were consequences of increasingly difficult situations from financial and operational considerations. Since it will be difficult to go lower than 10% without the risk of missing "rare" crops or livestock for instance, sampling of the villages would seem to be the next natural course to pursue.)

Unlike in the CPHs where village level statistics were required and published, results of all the previous CAFs were tabulated and published down to the town level only. (There were 1584 towns tallied during the 1990 CPH.) Thus, theoretically it would have been feasible to draw probability samples of villages in each town, and still be able to produce independent statistics for each town. However, sampling had been limited to the strata of small farm households.

**B. Principal Uses of the 1991 CAF**

The original list of intended uses of the census was expansive, it being the product of consultations among major user departments in the central government (Agriculture, Agrarian Reform, Natural Resources, etc.), local governments and researchers. The list was also premised on an ample budget. Moreover, NSO resources were to be augmented by manpower from the main user departments, especially Agriculture. The plan, therefore, was equally ambitious. The draft questionnaire was definitely heftier than the final version. There was to be 100 per cent coverage of all farm (and fishing households), even relaxing the cutoff point between farm and non-farm households from the original 1,000 square meters to any positive number, i.e., any household engaged in some agricultural or fishing activity during the census reference year would be included regardless of the size of his plot of land. Results were to be tabulated down to the village level, not at the town level as before.

The plan had to be overhauled once it became clear that the final budget would fall short of

expectations. What remained of the original intended uses of the census can be grouped into four categories.

1. *Baseline estimates of structural variables of the agricultural sector*. Recent studies on the quality of the agricultural database from intercensal surveys have raised the possibility of serious inaccuracies; see eg. *David, Mendoza and Besa (1990a) and David and Mendoza (1990b*). These results drove home the need for more accurate benchmark information from the census, as well as the importance of doing a better census with demonstrably reliable results.

2. *Sampling frame/auxiliary information for redesigning the system of agricultural surveys*. There have been plans to redesign the country's agricultural surveys, including switching from individual commodity-based surveys to integrated survey systems. The same studies mentioned above stressed the importance of using the census results to attempt a more careful and extensive job of developing cost-efficient sampling strategies for the agricultural surveys of the 1990s.

3. *Continue the series of small area (town) statistics from the censuses*, at least for major structural variables such as land use, distribution of farm holdings, and inventory of farm resources. This is singularly important since the CAFs are the only sources of small area statistics on a national scale.

4. *The census is an international commitment,* e.g. it is part of FAOs World Census of Agriculture Program.

It is implicitly assumed also that the census will be useful to private sector needs, including scholarly research.

**C. Implications on the Sampling Plan**

In all likelihood, national level estimates will be good enough to satisfy item (4) above, i.e. as long as a census is conducted during the early part of the decade. Item (3) requires that each town be a

domain with separate estimates. The sampling plan should be efficient enough to cloak the town statistics with sufficiently low sampling errors. In addition, special studies -- preferably integrated with the census operations -- to estimate non-sampling errors should be part of the overall census plan in order to have estimates of total error of the town estimates. Finally, all those concerned need to be convinced that the data from the sample census can still fill the need for frame information to redesign the system of sample surveys. The difference is that, for sample surveys the domains are the provinces, not the towns. This means that a necessary but not sufficient condition for a census sampling strategy to provide suitable frames for the intercensal surveys is for the former to yield provincial estimates with acceptable sampling error levels, e.g. < 5 per cent. This is due to the fact that, whatever sampling scheme is used for the intercensal surveys, the outcome can always be viewed as a second phase sample -- with the census as the first phase sample. Consequently, the sampling variance of the estimates from the intercensal surveys will be a sum of two components -- one from each phase. Thus, item (2) is just as equally important as (1) and (3). All three should in fact be considered together in contemplating probable sampling strategies.

**D. Sample Census or Sample Survey: Tightening Terminology**

The suggestion to sample villages in the 1991 census prompted the question, Can you still call it a census? This writer's answer at the time was, " I don't see why not -- if it will cover all towns and provide the same small area (town) statistics as the previous censuses." In other words, if the previous ones were labelled as censuses, there would be no reason not to call this one by the same name.

That this is not a pedantic, but indeed an important matter, became evident to the writer at a recent international meeting on agriculture censuses.[4] Previously, the use of sampling in censuses had

[4] International Expert Consultation on Agriculture Censuses, sponsored by the State Statistical Bureau, PRC and FAO, 4-9 October 1991, Beijing.

been confined to the ultimate sampling units such as persons or households. Cochran (1977) describes sampling in the US population censuses, initially for purposes of (i) obtaining additional information by asking extra questions about occupations, parentage, fertility and the like on a sample of persons; (ii) speeding up publication of results, either by direct reduction of returns to be processed or by producing advance tabulations from a sample of the returns; and (iii) later, except for basic information such as population counts, the entire censuses were shifted to a sample basis, resulting in substantial savings and early publication of results. However, the sampling did not go higher than the level of persons or households. Yates (1981) considers it a mistake to confine the word **survey** to a sample survey or the word **census** to a complete enumeration. Thus, an extensive and in depth investigation of all families in a community may be aptly called a survey rather than a census, while samples with a very large area coverage may be called sample censuses. What is not clear is the extent of sampling, beyond or below which we should call the activity a sample census or a sample survey, respectively. FAO has been using **sample census** in its literature and in the process may have recommended its use to many countries. However, the intent may have been limited to sampling farm holdings or operators. In the above-mentioned meeting, not a few countries (Indonesia, Nepal and Philippines, for example), described as censuses what could only be referred to heretofore as nationwide sample surveys. Indonesia, for instance described a three stage sampling procedure, and Nepal a two-stage sampling scheme in conducting their agriculture "censuses". These point to a need to reexamine terminologies. This, however, is a subject better dealt with in another venue or paper.

## III. EXPLORING SAMPLING STRATEGIES

### A. Introduction

In addition to sampling, a number of other options were raised in view of the bleak financial prospects for the census. One was to do a complete enumeration but covering the top 20 or 30 rice and

corn producing provinces only. The disadvantages of this suggestion were that it could not be called a national census, there would be no baseline nor sampling frame information for the not-covered provinces, and lack of information on them may lead to planning and policy omission if not benign neglect.

There was overwhelming preference for complete enumeration of households (in sample barangays or psus) in view of persistent problems in previous censuses -- the 0.1000 hectare cut-off for classification to a farm household and be part of the sampled population excluded the poorest of the poor, such as landless workers, who are the targets of special programs like poverty alleviation. Moreover, misclassifications between farm and non-farm households groups, as well as between large and small farms had caused both sampling and non-sampling errors to rise and for the data processing and estimation to be more complicated than planned.

At the same time, reservations were expressed concerning the adequacy of sampling the barangays (even if accompanied by complete enumeration of the households within). It was mentioned that this had been tried a number of times in the Bureau of Agricultural Statistics, with the results having unacceptably high sampling errors even with a sampling rate as high as 50 per cent of the barangays. This underscored the need for empirical verification of the results of alternative sampling strategies using available background information prior to adoption in the census.

**B. Background Material**

We mention here the time interval spent in the conduct of the following empirical work leading to the choice of a sampling procedure and the selection of the sample. At the time that options were being discussed and debated in the face of a still uncertain budget for the census, (which was in February 1991), we were assuming that the field data collection would be done in August, or at the latest, September of the same year. Thus, there was only March to July to complete preparations,

starting with securing the background data required for the numerical work.

What actually transpired was that the field operations for the four pilot provinces -- Bukidnon, Isabela, Laguna and Marinduque -- were conducted in October 1991 and the rest of the provinces were covered in February 1992. This meant that the sampling scheme had to be chosen and the samples drawn by July 1991.

The natural source of background data suitable as sampling frame and for testing alternative sampling strategies would be the latest censuses. From the complete lists of households made for the 1980 CAF (i.e, from CAF Form 1), barangay totals for the following variables are available:

TFA = total farm area

PCA = permanent crops area planted

FCA = field crops area planted

RCA = rice crops area planted

CCA = corn crops area planted

PMA = permanent meadows area

We also used

NHH = number of households both from the 1980 CAF and 1990 CPH.

In theory, these barangay totals have zero sampling error.[5]

## C. Getting to Know the Target Population

Between 1980 and 1990 the number of towns increased by 26, from 1,584 to 1,690; (see Table 1). Since new towns are created by splitting existing ones into two or more, this means that at

[5] There are also barangay and town level data from the main census (i.e. from CAF Form 3). However, these are subject to sampling error inasmuch as only a 10% sample of the small farm households was included in the 1980 CAF. This sampling rate often resulted in very small samples particularly for specific crop areas, so that the main 1980 CAF barangay estimates are deemed unsuitable for the purpose at hand.

least 52 towns in the 1990 census will not have their "exact" matches in the 1980 census. This births of new units "spoil" correlations and diminish the utility of previous information for planning or improving future statistical operations.[6]

[6] Exact matches can be recreated by the tedious process of identifying the barangays which went into each new town, which had been tried before. However, this is not possible when it is the barangays that are split since in general, the new boundaries which are not based on smaller administrative subdivisions would be difficult to identify on the ground.

**Table 1. Changes Between 1980 CAF and 1990 CPH**

| Region | Number of Towns | | | Number of Barangays | | |
|---|---|---|---|---|---|---|
| | 1980 | 1990 | Change | 1980 | 1990 | Change |
| 1 | 176 | 176 | 0 | 3949 | 3963 | 14 |
| 2 | 115 | 118 | 3 | 2477 | 2740 | 263 |
| 3 | 121 | 122 | 1 | 2783 | 2870 | 87 |
| 4 | 218 | 221 | 3 | 5146 | 5300 | 154 |
| 5 | 115 | 115 | 0 | 3442 | 3456 | 14 |
| 6 | 130 | 131 | 1 | 4033 | 4045 | 12 |
| 7 | 131 | 132 | 1 | 2972 | 3002 | 30 |
| 8 | 142 | 143 | 1 | 4194 | 4388 | 194 |
| 9 | 97 | 104 | 7 | 2324 | 2699 | 375 |
| 10 | 121 | 124 | 3 | 2273 | 2481 | 208 |
| 11 | 84 | 86 | 2 | 1589 | 1764 | 175 |
| 12 | 104 | 108 | 4 | 3464 | 2896 | -568 |
| NCR | 30 | 30 | | 1700 | 1689 | -11 |
| Total | 1584 | 1610 | 26 | 40346 | 41293 | 947 |

Table 1 also shows that the number of barangays increased from 40,346 in 1980 to 41,293 in 1990, or a net increase of 947 barangays. The actual change was much more, since Region 12 registered 568 fewer barangays in 1990 compared to 1980! (It is generally recognized that part of the reason for this unusual phenomenon was that there must have been many ghost barangays earlier, for reasons more political than anything else, but this is another story.) This is quite discouraging for at least two reasons. First, the use of more sophisticated sample selection procedures such as probability proportional to size (pps) sampling -- assuming that it were desirable (which it is not for reasons to be explained more fully later) -- is effectively ruled out by the lack of useful measures of size for a large proportion of the sampling units. Second, the lack of relevant size measures and the diminished correlations between 1980 and later values lessen the gains in precision that could otherwise be expected

from the use of the earlier values during the estimation stage (e.g. using ratio- or regression-type estimators.)

The frequency distribution of the number of villages in a town is shown in Table 2. The average number of barangays per town is 26. However, some towns are very small: 21 have five or less and another 156 have between 6 to 10 barangays. These point to the severe limitation of using stratification within towns, as the small sizes do not leave much room to maneuver in constructing strata.

A barangay had an average of 276 households during the 1990 CPH. This would translate to about 10 million households nationwide by the time the 1991 CAF data collection was done. This would have been the number of census units listed, at least half of which will turn out to be farm or fishing households, had the original CAF plan been followed (complete coverage of all barangays and all households).[7]

**Table 2. Frequency Distribution of Number of Villages Per Town, 1990**

| No. of Villages in Town | Frequency of Towns |
|---|---|
| < = 5 : | 21 |
| 6 - 10 : | 156 |
| 11 - 15 : | 302 |
| 16 - 20 : | 285 |
| 21 - 25 : | 256 |
| 26 - 30 : | 184 |
| 31 - 35 : | 121 |
| 36 - 40 : | 76 |
| 41 - 45 : | 60 |
| 46 - 50 : | 42 |
| 51 - 55 : | 29 |
| 56 - 60 : | 17 |
| 61 - 70 : | 21 |
| 71 - 80 : | 13 |
| > 80 : | 27 |
| Total : | 1610 |

[7] Compare the figure with the 2.1 million farms enumerated in the 1987 U.S. (complete enumeration) census of agriculture. Compare also the capabilities and resources of the statistical services in the two countries -- and one cannot help but see the distorted proportions, and hence the force of the logic in proposing sample censuses in the Philippines and the rest of the developing world.

_____

**D. Sampling Exercises**

**Exercise 1**: **Cluster Sampling** -- In each town draw a proportion, f, of the barangays with equal probability and without replacement; and enumerate completely the households in the sample barangays. Try f = 0.25 and 0.50. An unbiased estimator of a town hotal, Y, is $\_{srs} = (N/n) \Sigma y_j$, where N is the number of barangays in the town, n=fN, and $y_j$ denotes the total in the jth sample barangay. The variance of the estimate is $Var(\_{srs}) = N^2(1-f)S^2_y/n$, where $S^2_y$ is the variance of the individual barangay totals. As illustration, the coefficients of variation

$$CV(\_{srs}) = \{ Var(\_{srs})/Y\}.100\% \qquad (1)$$

of town totals in one province -- Batangas -- are shown in Table 3 and summary results for one region -- Southern Tagalog -- are presented in Table 4. Following are some highlights of the findings:

(a) The majority of the CVs exceed 10% even with f = 0.50. Hence, the sampling strategy is unsatisfactory if, say, we require that three-fourths of the CVs should be within single digits. In retrospect, the claim that sampling even one-half of the barangays had been inadequate is correct (see section III.A, last para.); however, this was with simple random sampling, which is the scheme to use only if there is no background information available other than a list of the barangays.

**Table 3. CVs with SRS of Barangays and Complete Enumeration of Households: Batangas, 1980 (Per Cent)**

| Town Code | N | TFA CV25 | TFA CV50 | PCA CV25 | PCA CV50 | FCA CV25 | FCA CV50 | NHH CV25 | NHH CV50 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 21 | 29.1 | 16.8 | 40.3 | 23.3 | 29.2 | 16.8 | 22.3 | 12.9 |
| 2 | 19 | 19.9 | 11.5 | 50.4 | 29.1 | 20.0 | 11.5 | 17.7 | 10.1 |
| 3 | 48 | 23.0 | 13.3 | 24.2 | 14.0 | 23.0 | 13.2 | 9.8 | 5.7 |
| 4 | 13 | 35.1 | 20.3 | 37.8 | 21.9 | 35.0 | 20.2 | 25.3 | 14.6 |
| 5 | 105 | 35.1 | 20.3 | 37.8 | 21.9 | 35.0 | 20.2 | 25.3 | 14.6 |
| 6 | 40 | 19.8 | 11.4 | 27.0 | 15.6 | 19.7 | 11.3 | 22.1 | 12.8 |
| 7 | 40 | 20.4 | 11.8 | 36.3 | 21.0 | 20.8 | 12.0 | 14.6 | 8.4 |
| 8 | 25 | 21.7 | 12.5 | 28.6 | 16.5 | 22.5 | 13.0 | 14.1 | 8.1 |
| 9 | 21 | 31.9 | 18.4 | 43.1 | 24.9 | 31.8 | 18.4 | 20.8 | 12.0 |
| 10 | 25 | 15.3 | 8.8 | 34.6 | 20.0 | 15.4 | 8.9 | 17.6 | 10.2 |
| 11 | 21 | 30.2 | 17.5 | 36.3 | 20.9 | 29.2 | 16.9 | 19.7 | 11.3 |
| 12 | 46 | 29.8 | 17.2 | 31.8 | 18.4 | 30.0 | 17.3 | 9.3 | 5.4 |
| 13 | 19 | 25.5 | 14.7 | 28.5 | 16.5 | 25.3 | 14.6 | 14.3 | 8.2 |
| 14 | 72 | 15.0 | 8.6 | 19.5 | 11.3 | 15.0 | 8.7 | 11.9 | 6.9 |
| 15 | 27 | 26.4 | 15.3 | 32.7 | 18.9 | 26.3 | 15.9 | 12.0 | 6.9 |
| 16 | 34 | 21.1 | 12.2 | 18.6 | 10.7 | 20.5 | 11.9 | 12.3 | 7.1 |
| 17 | 16 | 26.3 | 15.2 | 27.1 | 15.6 | 26.1 | 15.1 | 30.1 | 17.4 |
| 18 | 16 | 35.5 | 20.5 | 35.6 | 20.6 | 35.8 | 20.6 | 16.8 | 9.7 |
| 19 | 42 | 27.5 | 15.9 | 36.2 | 20.9 | 28.0 | 16.2 | 12.7 | 7.3 |
| 20 | 18 | 19.1 | 11.0 | 32.0 | 18.5 | 19.1 | 11.0 | 17.5 | 10.1 |
| 21 | 48 | 22.0 | 12.7 | 28.9 | 16.7 | 16.4 | 9.5 | 9.5 | 5.5 |
| 22 | 32 | 25.4 | 14.7 | 26.9 | 15.5 | 25.1 | 14.5 | 14.5 | 8.4 |
| 23 | 42 | 20.5 | 11.8 | 24.7 | 14.3 | 20.4 | 11.8 | 13.3 | 7.7 |
| 24 | 31 | 25.1 | 14.5 | 22.8 | 13.1 | 23.8 | 13.8 | 17.0 | 9.8 |
| 25 | 16 | 25.4 | 14.7 | 29.3 | 16.9 | 25.3 | 14.6 | 25.1 | 14.5 |
| 26 | 29 | 15.8 | 9.1 | 31.9 | 18.4 | 15.4 | 8.9 | 27.9 | 16.1 |
| 27 | 15 | 34.6 | 20.0 | 39.5 | 22.8 | 34.6 | 20.0 | 22.9 | 13.2 |
| 28 | 30 | 25.7 | 14.8 | 30.9 | 17.8 | 25.7 | 14.8 | 16.3 | 9.4 |
| 29 | 42 | 29.0 | 16.7 | 21.8 | 12.6 | 28.4 | 16.4 | 17.3 | 10.0 |
| 30 | 22 | 41.6 | 24.0 | 51.2 | 29.6 | 44.0 | 25.4 | 25.9 | 15.0 |
| 31 | 48 | 15.9 | 9.2 | 23.4 | 13.5 | 15.6 | 9.0 | 13.3 | 7.6 |
| 32 | 20 | 26.5 | 15.3 | 35.6 | 20.5 | 22.0 | 12.7 | 10.4 | 6.0 |
| 33 | 15 | 35.1 | 20.3 | 36.2 | 20.9 | 31.2 | 18.0 | 16.6 | 9.6 |
| 34 | 22 | 22.6 | 13.0 | 22.5 | 13.0 | 22.6 | 13.0 | 13.6 | 7.8 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Batangas | 1080 | 4.7 | 2.7 | 6.2 | 3.6 | 4.5 | 2.6 | 3.3 | 1.9 |

Note: CV25 and CV50 are coefficients of variation when f=.25 and .50, respectively.

(b) Close to one-third of the provincial estimate CVs exceed 5% (Table 4). It would be desirable if most were under 5%. Some provinces, viz. Mindoro Occidental and Rizal, exhibit about twice as much variation as the others. These would need special handling, e.g. trace the offending units and assign them inclusion probabilities equal to one.

(c) The regional estimates have CVs below 2%, which should be acceptable. From those values, we can hazard a guess that the national level estimates will have negligible sampling error.

The low sampling errors of the regional and national estimates are reminders that what require more attention at these levels are the non-sampling errors. Since more objective measurement methods may not be practicable in a massive operation like a census, post-enumeration studies designed to estimate non-sampling errors should be part of the overall census program. Better still, though difficult it would not be a bad idea to attempt an error profile of the census, and then proceed to plan on estimating the magnitudes of the major sources.

**Table 4. Summary Results (CVs) with SRS of Barangays in Each Town and Complete Enumeration of Households: Southern Tagalog**
(Per Cent)

| Province | T | N | Ranges of Town CV50 TFA | PCA | FCA | NHH | Province CV50 TFA | PCA | FCA | NHH |
|---|---|---|---|---|---|---|---|---|---|---|
| Aurora | 8 | 147 | 10-41 | 13-88 | 10-30 | 8-22 | 7.0 | 7.8 | 5.7 | 5.5 |
| Batangas | 34 | 1080 | 9-24 | 11-30 | 8-25 | 5-17 | 2.7 | 3.6 | 2.6 | 1.9 |
| Cavite | 22 | 461 | 10-76 | 11-70 | 10-76 | 8-62 | 4.6 | 6.1 | 4.6 | 7.7 |
| Laguna | 30 | 670 | 9-38 | 9-48 | 9-38 | 8-44 | 3.8 | 4.7 | 3.8 | 4.4 |
| Marinduque | 6 | 218 | 9-17 | 9-20 | 8-18 | 7-12 | 4.9 | 5.6 | 4.8 | 3.7 |
| Mind. Occ. | 11 | 155 | 14-51 | 14-56 | 12-47 | 10-25 | 10.3 | 12.5 | 8.4 | 6.0 |
| Mind. Or. | 15 | 402 | 8-27 | 10-34 | 8-21 | 5-18 | 4.1 | 4.8 | 3.6 | 2.8 |
| Palawan | 20 | 402 | 9-60 | 12-44 | 9-39 | 8-27 | 7.1 | 4.9 | 4.1 | 3.9 |
| Quezon | 41 | 1225 | 12-42 | 7-45 | 7-42 | 6-43 | 2.7 | 2.6 | 2.5 | 2.4 |
| Rizal | 14 | 170 | 16-60 | 20-61 | 16-60 | 13-27 | 13.1 | 22.8 | 13.3 | 6.6 |
| Romblon | 16 | 209 | 8-24 | 8-25 | 8-21 | 8-17 | 4.6 | 4.4 | 4.1 | 3.0 |
| Region | 217 | 5139 | - | - | - | - | 1.9 | 1.5 | 1.3 | 1.5 |

**Exercise 2:** In each town, arrange the barangays in descending order of TFA, then choose one of the first two with equal probability and every other barangay from thereon. Enumerate all households within.

Computing the estimate of town total,

$$\bar{y}_{sy} = (N/n) \sum_{n} y_j \qquad (2)$$

and its mean square error,

$$MSE(\bar{Y}_{sy}) = (1/2)\{(\bar{Y}_{sy1} - Y)^2 + (\bar{Y}_{sy2} - Y)^2\} \qquad (3)$$

where Y is the actual town total (from the 1980 CAF) and $y_j$ is the total in the jth sample barangay would not require excessive computer time since the sample space consists of only two subsets which give the estimates $\bar{y}_{syi}$, i = 1, 2. The estimator $\bar{y}_{sy}$ is biased for N = odd number. Although the bias would be assuaged by the use of (N/n) instead of 2 as raising factor, it might not be negligible in towns where barangay totals are very skewed; e.g. the biggest value is much larger than the next one, and especially for smallish N. This needs to be investigated empirically so that, if need be, the sampling can be changed from simple systematic to circular systematic sampling. The latter leads to unbiased estimates, but it is more cumbersome to repeat over 1,600 times. Another way would be to select the largest unit with probability one whenever N = odd.

The biases (for N = odd) turned out to be less than one percent of the actual totals in all but two of the 217 towns (in Southern Tagalog where the computations were done). The reasons for the exceptional cases are easily identifiable and rectifiable; e.g. one was a town with 3 barangays (Kalayaan, Laguna), in which case we should simply enumerate all.

Numerical results from *Exercise 2* are given in Tables 5, 6 and 7. Note that CV ($\bar{y}_{sy}$ is defined as the ratio between the square root of Eq. (2) and the actual total Y, expressed in per cent. Some highlights of the findings follow:

(a) The gains in precision of *Exercise 2* over *Exercise 1* are substantial and in some cases remarkable; compare, for example, the values of CV ($\bar{y}_{sy}$ in Table 5 with the corresponding CV($\bar{y}_{srs}$) in Table 3 for the towns in Batangas province. Those in Table 5 are generally low and acceptable. For the few that are high, eg. > 20%, finding the source is a simple matter of going over the population of barangay values. Some of these are mentioned in the footnotes to the table; e.g., when non-zero observations are "rare events", an extreme case being only one barangay has non-zero value, which results in CV($\bar{y}_{sy}$) = 100%.

(b) The cumulative distribution function (cdf) of CV($\bar{y}_{sy}$) in the 217 Southern Tagalog towns are given in Table 6. Excluding CCA and PMA (estimates of which may be aggregated and presented at the provincial instead of town levels), more than half of the town estimates satisfy CV $\leq$ 10%; the proportion with CV $\leq$ 15% ranges from 62% in the case of RCA to 89% in the case of FCA. These are respectable figures especially since, with so many domains and variables, one should not hope to have every CV within, say even 30%; and as insinuated in (a) above and (d) below, something could and should be done about towns with unusually high CVs.

(c) Excluding CCA and PMA, the province level CVs are generally acceptable (table 7). There are isolated cases -- eg Mindoro Occidental and Rizal -- where the CVs exceed 5% which will therefore need special attention. The region level CVs are within 1%.

(d) Frequently, the CV is high because one barangay is unusually large, so that the systematic sample that has it will produce an overestimate of the town total, while its complement sample will yield an underestimate. Therefore, further gains in precision can be realized if that unit is drawn automatically, and a 50% systematic sample is drawn from the remaining ordered N' = N-1 barangays. (This will be *Exercise 3*, to be presented later.) We present three examples below.

In Table 8a -- which is a Batangas town with NSO assigned geocode number 30 -- the biggest of the 22 barangays is only moderately larger than the next one; i.e., compare barangays no. 7 and no. 1. The CVs obtained from using *Exercise 2* are shown on the second to the last row. Let CV($\bar{y}'$), the values of which are shown on the last row of the table, be the coefficient of variation when barangay no. 7 is drawn automatically and a 50% systematic sample is selected from the remaining ordered barangays (the derivations are presented in *Exercise 3* below). Notice that CV($\bar{y}'$) < CV($\bar{y}^{sy}$) in the four out of seven variables.

Table 8b presents a town -- code No.5, Mindoro Occidental province -- in which the total farm area (TFA) of the biggest barangay is more than 3 times that of the second biggest. Consequently, gains in precision in TFA and variables highly correlated with it (FCA, RCA) are dramatic, i.e. CV(_') << CV($_{sy}$). However, there is loss of precision in the case of PCA because, by coincidence, the two possible systematic samples resulted in a subset of large values in one, and of small units in the other. (This only reminds us of the expected: in a multi-purpose sample survey or census wherein the variables of interest are not always highly or positively correlated, one cannot hope to win all the time.)[8]

[8] In this connection, Dalenius ( ) raises the possibility of using some average instead of individual mean square error or CV values as a measure of accuracy for a group of small area estimates.

**Table 5. CV($\bar{y}_{sy}$) of Town Estimates: Batangas (Per Cent)**

| Towncode | N | TFA | PCA | FCA | RCA | CCA | PMA a/ | NHH |
|---|---|---|---|---|---|---|---|---|
| 1 | 21 | 5.1 | 30.7 | 0.9 | 17.4 | 16.0 | n.a. b/ | 11.1 |
| 2 | 19 | 5.5 | 9.2 | 5.1 | 6.9 | 14.8 | 95.1 | 16.1 |
| 3 | 48 | 3.0 | 6.5 | 3.2 | 13.9 | 97.7 | 1.7 | 0.4 |
| 4 | 13 | 1.5 | 0.2 | 1.6 | 40.3 | 38.8 | n.a. | 7.8 |
| 5 | 105 | 3.2 | 4.4 | 0.8 | 5.1 | 28.0 | 16.8 | 0.8 |
| 6 | 40 | 2.9 | 4.9 | 2.7 | 2.5 | 21.8 | 0.0 c/ | 3.2 |
| 7 | 40 | 3.5 | 0.9 | 3.7 | 8.4 | 10.1 | 100.0 d/ | 0.3 |
| 8 | 25 | 2.1 | 0.3 | 2.4 | 13.4 | 14.7 | 2.1 | 2.5 |
| 9 | 21 | 4.4 | 13.1 | 5.4 | 6.9 | 34.7 | 95.6 | 4.7 |
| 10 | 25 | 0.1 | 2.0 | 0.1 | 5.2 | 47.0 | 4.0 | 15.1 |
| 11 | 21 | 3.8 | 2.0 | 5.0 | 18.4 | 29.3 | 3.2 | 7.8 |
| 12 | 46 | 3.0 | 7.5 | 1.5 | 6.9 | 11.2 | 66.7 | 2.3 |
| 13 | 19 | 0.9 | 8.8 | 1.3 | 32.7 | 74.8 | 68.9 | 2.1 |
| 14 | 72 | 3.5 | 5.3 | 3.5 | 13.8 | 10.2 | 62.5 | 1.0 |
| 15 | 27 | 1.1 | 1.6 | 2.2 | 4.5 | 2.1 | 20.6 | 8.3 |
| 16 | 34 | 3.9 | 1.1 | 2.3 | 23.3 | 3.4 | 100.0 | 6.1 |
| 17 | 16 | 9.0 | 16.1 | 9.5 | 3.0 | 5.9 | 100.0 | 23.5 |
| 18 | 16 | 7.0 | 8.1 | 7.0 | 45.5 | 100.0 | 100.0 | 3.5 |
| 19 | 42 | 6.1 | 2.2 | 10.1 | 54.3 | 47.7 | 23.7 | 0.4 |
| 20 | 18 | 4.7 | 0.9 | 4.7 | 7.9 | 21.7 | n.a. | 12.8 |
| 21 | 48 | 4.0 | 3.1 | 0.3 | 0.1 | 2.4 | 23.6 | 0.0 |
| 22 | 32 | 7.3 | 14.8 | 6.4 | 0.3 | 86.0 | 33.3 | 5.1 |
| 23 | 42 | 5.4 | 12.8 | 5.3 | 4.6 | 75.5 | 48.0 | 0.2 |
| 24 | 31 | 0.2 | 0.4 | 0.4 | 6.5 | 30.8 | 23.3 | 7.4 |
| 25 | 16 | 9.0 | 14.3 | 8.4 | 3.0 | 26.2 | n.a. | 2.9 |
| 26 | 29 | 1.4 | 0.1 | 0.4 | 4.4 | 0.0 | 96.7 | 27.1 |
| 27 | 15 | 4.6 | 13.2 | 4.9 | 82.8 | 107.4 | n.a. | 2.0 |
| 28 | 30 | 4.8 | 14.9 | 4.5 | 31.1 | 35.2 | 12.5 | 11.5 |
| 29 | 42 | 13.3 | 3.6 | 12.2 | 22.8 | 47.6 | 0.0 | 5.9 |
| 30 | 22 | 9.8 | 25.9 | 14.8 | 7.1 | 1.5 | 23.4 | 7.0 |
| 31 | 48 | 5.1 | 7.4 | 3.8 | 8.3 | 21.6 | 77.8 | 2.0 |
| 32 | 20 | 9.3 | 29.2 | 6.3 | 14.6 | 21.0 | 46.4 | 1.4 |
| 33 | 15 | 2.9 | 3.5 | 13.0 | 35.9 | 46.3 | 21.1 | 9.3 |
| 34 | 22 | 6.9 | 7.8 | 6.8 | 95.6 | 100.0 | 81.8 | 4.2 |
| Batangas | 1080 | 1.0 | 1.9 | 1.0 | 3.2 | 7.3 | 11.4 | 1.1 |

a/ The high CVs for permanent meadows area (PMA) are mainly due to the fact that the variable is a "rare attribute", e.g. it is zero in most of the barangays. Perhaps, such variables are better aggregated and presented at the provincial levels.

b/ n.a. means all unit values are zero.

c/ A CV = 0.0 often is a result of only two barangays having non-zero but equal values.

d/ A CV = 100% results whenever only one barangay has a non-zero value.

**Table 6. Cumulative Frequency Distribution of CV($\bar{y}_{sy}$) in Southern Tagalog Towns a/ Systematic Sampling of Ordered Barangays**

| | T F A | | P C A | | F C A | | NHH | | R C A | | C C A b/ | | P M A b/ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CV Range | Freq | % | Freq | % | Freq | % | Freq | % | Freq | % | Freq | % | Freq | % |
| CV ≤ 5% | 97 | 45 | 80 | 37 | 105 | 48 | 75 | 35 | 58 | 27 | 19 | 9 | 18 | 10 |
| ≤ 10% | 166 | 76 | 124 | 57 | 167 | 77 | 138 | 64 | 108 | 51 | 32 | 16 | 27 | 16 |
| ≤ 15% | 189 | 87 | 154 | 71 | 194 | 89 | 172 | 79 | 133 | 62 | 55 | 27 | 36 | 21 |
| ≤ 20% | 201 | 93 | 170 | 79 | 205 | 94 | 195 | 90 | 145 | 68 | 67 | 33 | 47 | 27 |
| Over 20% | 217 | 100 | 216 | 100 | 217 | 100 | 217 | 100 | 214 | 100 | 204 | 100 | 174 | 100 |

a/ Total frequencies (last row) not equal since towns with zero values for all barangays are not counted.

b/ These values suggest that corn crops area (CCA) and permanent meadows area (PMA) would be more appropriately tabulated at the provincial or regional levels only.

**Table 7. CV($\bar{y}_{sy}$) of Provincial Estimates with Ordered Systematic Sampling**

| | Variable | | | | | | |
|---|---|---|---|---|---|---|---|
| Province | TFA | PCA | FCA | NHH | RCA | CCA | PMA |
| Aurora | | 4.3 | 6.1 | 2.6 | 4.1 | 4.4 | 22.9 | 31.0 |
| Batangas | | 1.0 | 1.9 | 1.0 | 1.1 | 3.2 | 7.3 | 11.4 |
| Cavite | | 2.2 | 4.2 | 2.2 | 7.4 | 2.1 | 8.5 | 37.6 |
| Laguna | | 1.9 | 1.4 | 1.6 | 5.4 | 3.2 | 22.8 | 24.3 |
| Marinduque | | 1.4 | 4.0 | 2.1 | 2.0 | 7.0 | 10.1 | 13.5 |
| Mindoro Occidental | | 6.7 | 9.4 | 7.2 | 4.6 | 5.6 | 11.2 | 10.7 |
| Mindoro Oriental | | 1.8 | 3.3 | 1.1 | 2.2 | 4.5 | 6.8 | 24.1 |
| Palawan | | 2.0 | 4.0 | 1.6 | 4.9 | 2.3 | 5.9 | 19.4 |
| Quezon | | 1.8 | 1.9 | 1.6 | 1.9 | 2.6 | 3.8 | 17.8 |
| Rizal | | 8.1 | 22.3 | 7.4 | 6.1 | 9.4 | 13.9 | 28.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Romblon | 2.4 | 2.1 | 1.9 | 2.7 | 4.3 | 11.7 | 15.4 |
| Southern Tagalog | 0.4 | 1.1 | 0.8 | 1.5 | 1.3 | 2.6 | 6.6 |

Table 8c -- a town in Palawan province, code suppressed -- is a reminder of the need for special frames or special handling of unusual units: the first barangay on the list had only one household, but which possessed a 12,800 hectare farm (a ranch, a corporate farm, or an error?). *This case also illustrates the conflict between the demand for local area statistics and the safeguarding of the confidentiality of individual respondent's responses. The household with 12,800 hectares can be identified easily!*

---

**Table 8a. Barangay Totals in Town No. 30, Batangas Province, 1980**

| BGYCODE | TFA | PCA | RCA | NHH | PMA | RCA | CCA |
|---|---|---|---|---|---|---|---|
| 7 | 436 | 259 | 412 | 277 | 6 | 150 | 2 |
| 1 | 317 | 66 | 269 | 486 | 3 | 126 | 71 |
| 19 | 199 | 95 | 184 | 156 | 1 | 61 | 33 |
| 4 | 178 | 75 | 132 | 251 | 1 | 49 | 4 |
| 3 | 152 | 119 | 140 | 120 | 0 | 18 | 4 |
| 20 | 140 | 24 | 112 | 191 | 6 | 79 | 8 |
| 24 | 109 | 18 | 92 | 192 | 11 | 43 | 34 |
| 10 | 104 | 65 | 102 | 86 | 0 | 29 | 3 |
| 21 | 103 | 22 | 82 | 72 | 0 | 43 | 8 |
| 23 | 92 | 41 | 64 | 103 | 18 | 17 | 2 |
| 25 | 71 | 28 | 69 | 114 | 0 | 29 | 10 |
| 9 | 43 | 21 | 36 | 103 | 1 | 10 | 4 |
| 8 | 39 | 10 | 29 | 145 | 0 | 16 | 2 |
| 6 | 38 | 16 | 28 | 132 | 0 | 9 | 2 |
| 11 | 32 | 17 | 24 | 87 | 0 | 4 | 2 |
| 12 | 22 | 15 | 19 | 113 | 0 | 3 | 1 |
| 13 | 12 | 4 | 10 | 90 | 0 | 4 | 2 |
| 18 | 12 | 7 | 11 | 89 | 0 | 1 | 1 |
| 14 | 10 | 4 | 8 | 99 | 0 | 2 | 1 |
| 15 | 10 | 8 | 9 | 21 | 0 | 0 | 0 |
| 17 | 9 | 5 | 9 | 75 | 0 | 2 | 1 |
| 16 | 6 | 4 | 4 | 67 | 0 | 0 | 0 |
| Total | 2134 | 923 | 1845 | 3069 | 47 | 695 | 195 |
| CV($\bar{y}_{sy}$) | 9.8 | 25.9 | 14.8 | 7.0 | 23.4 | 7.1 | 1.9 |
| CV($\bar{y}'$) | 6.8 | 1.3 | 3.8 | 11.7 | 32.1 | 10.8 | 5.2 |

_____
__

**Table 8b. Barangay Totals in Town No. 5, Occidental Mindoro Province, 1980**

| BGY | CODETFA | PCA | FCA | NHH | PMA | RCA | CCA |
|---|---|---|---|---|---|---|---|
| 10 | 1804 | 67 | 1693 | 935 | 26 | 1610 | 10 |
| 9 | 525 | 5 | 511 | 252 | 0 | 490 | 16 |
| 8 | 522 | 14 | 501 | 288 | 9 | 306 | 158 |
| 5 | 326 | 11 | 291 | 158 | 28 | 242 | 38 |
| 7 | 310 | 1 | 305 | 113 | 3 | 234 | 70 |
| 12 | 281 | 70 | 281 | 279 | 0 | 117 | 98 |
| 11 | 276 | 0 | 274 | 287 | 0 | 255 | 19 |
| 1 | 252 | 78 | 252 | 320 | 0 | 16 | 158 |
| 4 | 176 | 0 | 176 | 104 | 0 | 176 | 0 |
| 13 | 172 | 1 | 166 | 174 | 1 | 162 | 3 |
| 6 | 147 | 1 | 146 | 88 | 1 | 145 | 0 |
| 3 | 82 | 14 | 81 | 260 | 0 | 74 | 2 |
| Total | 4873 | 252 | 4677 | 3258 | 68 | 3827 | 572 |
| CV(_sy) | 32.7 | 34.1 | 32.3 | 11.4 | 14.7 | 42.5 | 10.1 |
| CV(_') | 1.5 | 54.7 | 2.0 | 10.9 | 18.1 | 5.7 | 3.0 |

**Table 8c. Barangay Totals in a Town in Palawan Province, 1980**

| BGY | CODETFA | PCA | FCA | NHH | PMA | RCA | CCA |
|---|---|---|---|---|---|---|---|
| 10 | 12800 | 15 | 1498 | 1 | 0 | 0 | 0 |
| 24 | 1651 | 797 | 828 | 268 | 0 | 147 | 5 |
| 21 | 1488 | 322 | 374 | 200 | 0 | 316 | 1 |
| 6 | 1041 | 194 | 604 | 227 | 0 | 204 | 0 |
| 20 | 895 | 217 | 218 | 140 | 0 | 201 | 0 |
| 3 | 800 | 370 | 424 | 357 | 0 | 81 | 83 |
| 5 | 788 | 208 | 286 | 98 | 0 | 62 | 0 |
| 9 | 358 | 132 | 186 | 93 | 0 | 54 | 0 |
| 7 | 318 | 72 | 83 | 60 | 0 | 37 | 4 |
| 1 | 234 | 32 | 118 | 60 | 0 | 42 | 0 |
| 22 | 185 | 111 | 111 | 31 | 0 | 0 | 0 |
| 12 | 170 | 17 | 42 | 32 | 0 | 25 | 0 |
| 19 | 99 | 14 | 37 | 29 | 0 | 15 | 0 |
| 11 | 47 | 10 | 24 | 1 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 320 | 0 | 0 | 0 |
| Total | 20874 | 2511 | 4833 | 1917 | 0 | 1194 | 93 |
| CV(_sy) | 58.8 | 23.6 | 7.9 | 41.7 | - | 6.6 | 89.2 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CV(_') | 2.5 | 17.7 | 18.6 | 35.3 | - | 13.3 | 83.1 |

**Exercise 3 (as described above):** In each town draw with certainty the barangay with the largest total farm area (TFA) in 1980, then use *Exercise 2* on the remaining N'= N-1 barangays. Like the sampling procedure, the estimators remain very simple. Let $y_c$ be the value observed from the certainty barangay, Y the total of the y's in the town (the target parameter), and n' the sample size from the N' barangays. The estimator of Y, and its mean square error and coefficient of variation are:

$$\_ = y_c + (N'/n') \sum_{n'} y_j , \qquad (4)$$

$$MSE(\_') = (1/2) \{ (\_'_1 - Y)^2 + (\_'_2 - Y)^2 \}, \qquad (5)$$

and

$$CV (\_) = \{ \quad MSE (\_')/Y \} \times 100\% \qquad (6)$$

where $\__1'$ and $\__2'$ are values of estimator (3) from the only two possible samples of size n'.

With this sampling procedure, valid estimates of sampling errors cannot be computed for the 1991 census. On the other hand, we can do so here using the complete barangay level data from CAF80. Results are summarized in Tables 9 and 10. Moreover, while the previous exercises were carried out on data from 217 towns in Southern Tagalog region only, *Exercise 3* is done nationwide (excluding the National Capital Region, NCR).

The cdfs of 1554 town CVs under *Exercise 2* and CV's ( ' for prime) under *Exercise 3* are compared in Table 9. The advantage of *Exercise 3* is clear: the CV' curves are above and rise faster than the CV counterparts in most cases. Except for the area under corn (CCA) and permanent meadows (PMA), at least half of the CV's were under 10%. Understandably, the precision is highest with TFA and similar global area variables like FCA and PCA, but lower for specific crop areas like corn because the latter can behave like rare events ( hence higher inherent variation) in some towns. Except for CCA and PMA, the proportion of CV' that are within 20% ranges from 73% to 99%.

The CV(_') values for all the provinces, regions and country (excluding NCR) are given on Table 10. Most of the province level values are below 5%, majority of the regional level values are under 2%, and the national level CV's are, for all intent and purposes, negligible. Thus, with few exceptions, these results show that with *Exercise 3* the sampling error is very much under control. Moreover, further improvements in the precision of estimates could be expected by exploring alternative estimators.

**Table 9. Comparison of 1554 Town CV($_{sy}$) and CV(') Using 1980 CAF Data, Philippines a/ (Per Cent)**

| Range | TFA CV($_{sy}$) | TFA CV(') | FCA CV($_{sy}$) | FCA CV(') | PCA CV($_{sy}$) | PCA CV(') | RCA CV($_{sy}$) | RCA CV(') | NOHH CV($_{sy}$) | NOHH CV(') | CCA CV($_{sy}$) | CCA CV(') | PMA CV($_{sy}$) | PMA CV(') |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ≤ 1% | 9.3 | 17.0 | 9.6 | 14.7 | 5.2 | 6.9 | 5.5 | 8.0 | 8.0 | 11.1 | 3.9 | 6.4 | 2.3 | 5.2 |
| ≤ 2% | 17.7 | 30.7 | 18.9 | 28.9 | 9.9 | 13.4 | 12.1 | 14.2 | 17.4 | 21.8 | 6.9 | 10.1 | 3.8 | 7.6 |
| ≤ 3% | 27.8 | 44.0 | 27.8 | 41.2 | 14.5 | 19.0 | 17.0 | 21.0 | 27.3 | 31.1 | 9.9 | 14.2 | 4.9 | 10.2 |
| ≤ 5% | 46.1 | 68.9 | 46.3 | 64.1 | 24.5 | 31.1 | 27.8 | 33.7 | 44.0 | 49.0 | 15.3 | 21.8 | 8.4 | 14.5 |
| ≤ 10% | 75.1 | 93.8 | 74.7 | 90.4 | 43.0 | 52.4 | 47.9 | 53.9 | 72.8 | 80.4 | 28.3 | 35.5 | 16.7 | 24.9 |
| ≤ 15% | 87.0 | 98.5 | 86.4 | 97.4 | 56.0 | 66.0 | 61.6 | 66.8 | 86.0 | 92.3 | 39.3 | 46.6 | 23.5 | 35.6 |
| ≤ 20% | 93.1 | 99.2 | 93.2 | 99.2 | 64.5 | 72.7 | 70.8 | 76.3 | 93.1 | 97.1 | 48.9 | 54.9 | 29.2 | 43.8 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $CV(\__{sy})/CV(\_')$ at ≤ 5% | 1.49 | 1.38 | 1.27 | 1.21 | 1.11 | 1.42 | 1.73 |
| $CV(\__{sy})/CV(\_')$ at ≤ 10% | 1.25 | 1.21 | 1.22 | 1.13 | 1.10 | 1.25 | 1.49 |

a/ Excluding National Capital Region

**Table 10. CV(_') of Provincial, Regional and National Estimates (Per Cent)**

| Province | TFA | PCA | FCA | RCA | NOHH | CCA | PNA |
|---|---|---|---|---|---|---|---|
| Abra | 1.2 | 8.8 | 1.6 | 1.7 | 1.8 | 2.6 | 8.2 |
| Benguet | 1.7 | 11.6 | 2.8 | 9.8 | 2.3 | 11.2 | 6.4 |
| Ilocos Norte | 0.7 | 7.7 | 1.1 | 1.5 | 0.9 | 4.7 | 5.5 |
| Ilocos Sur | 0.9 | 10.4 | 1.0 | 1.0 | 1.2 | 4.2 | 3.1 |
| La Union | 0.9 | 4.7 | 1.3 | 1.3 | 1.3 | 10.4 | 4.0 |
| Mt. Province | 4.4 | 48.2 | 4.8 | 4.9 | 3.0 | 2.9 | 12.4 |
| Pangasinan | 0.5 | 7.1 | 0.5 | 0.7 | 0.8 | 5.6 | 7.0 |
| Region I | 0.4 | 5.7 | 0.4 | 0.5 | 0.5 | 3.3 | 2.8 |
| Batanes | 5.0 | 3.6 | 3.7 | 3.6 | 7.7 | 13.9 | 9.7 |
| Cagayan | 1.0 | 7.9 | 1.3 | 2.1 | 1.6 | 3.2 | 12.3 |
| Ifugao | 2.8 | 14.0 | 8.1 | 6.9 | 2.9 | 14.7 | 8.5 |
| Isabela | 0.7 | 5.4 | 0.9 | 1.7 | 1.1 | 1.8 | 9.3 |
| Kalinga-Apayao | 2.5 | 4.4 | 3.8 | 4.4 | 3.0 | 8.0 | 13.2 |
| Nueva Vizcaya | 1.4 | 7.0 | 1.6 | 2.6 | 1.7 | 7.6 | 4.2 |
| Quirino | 2.2 | 25.5 | 2.1 | 3.0 | 2.9 | 6.1 | 19.3 |
| Region 2 | 0.6 | 3.1 | 0.8 | 1.1 | 0.8 | 1.5 | 4.4 |
| Bataan | 1.4 | 5.2 | 1.7 | 2.5 | 3.1 | 31.9 | 16.7 |
| Bulacan | 0.9 | 8.7 | 0.9 | 0.9 | 1.3 | 7.1 | 6.5 |
| Nueva Ecija | 0.6 | 20.9 | 0.6 | 0.9 | 1.1 | 13.5 | 9.6 |
| Pampanga | 0.9 | 17.9 | 0.9 | 1.6 | 1.2 | 20.1 | 15.5 |
| Tarlac | 0.9 | 9.2 | 0.9 | 1.4 | 1.4 | 7.6 | 8.4 |
| Zambales | 2.2 | 7.9 | 2.3 | 3.5 | 3.2 | 17.1 | 11.9 |
| Region 3 | 0.4 | 4.0 | 0.4 | 0.6 | 0.6 | 8.2 | 4.8 |
| Batangas | 0.7 | 1.6 | 0.8 | 3.0 | 1.0 | 6.8 | 4.5 |
| Cavite | 1.1 | 1.4 | 1.2 | 2.3 | 1.8 | 7.3 | 31.5 |
| Laguna | 1.1 | 2.1 | 1.1 | 2.2 | 2.6 | 20.9 | 9.4 |
| Marinduque | 1.5 | 2.1 | 1.0 | 7.0 | 2.1 | 12.5 | 13.6 |
| Mindoro Occ. | 2.9 | 8.9 | 2.5 | 3.5 | 6.5 | 10.3 | 8.8 |
| Mindoro Or. | 1.2 | 3.4 | 1.2 | 4.3 | 2.1 | 7.3 | 2.7 |
| Palawan | 0.9 | 2.4 | 2.0 | 5.2 | 2.2 | 7.5 | 10.9 |
| Quezon | 0.5 | 0.7 | 0.6 | 2.7 | 1.1 | 2.1 | 9.9 |
| Rizal | 3.8 | 4.8 | 4.2 | 5.2 | 2.8 | 15.7 | 21.8 |
| Romblon | 1.5 | 2.0 | 1.8 | 2.7 | 1.9 | 9.2 | 10.2 |
| Aurora | 1.5 | 4.2 | 1.6 | 4.1 | 2.2 | 26.9 | 5.1 |
| Region 4 | 0.4 | 0.6 | 0.4 | 1.4 | 0.6 | 2.5 | 4.4 |
| Albay | 1.0 | 2.0 | 1.0 | 3.5 | 1.4 | 4.8 | 12.8 |
| Camarines N. | 0.9 | 1.2 | 0.8 | 5.1 | 2.5 | 16.3 | 9.0 |
| Camarines Sur | 0.6 | 1.3 | 0.7 | 2.0 | 1.4 | 5.0 | 4.3 |
| Catanduanes | 1.2 | 2.2 | 1.0 | 3.5 | 2.0 | 10.2 | 5.9 |
| Masbate | 1.0 | 2.0 | 1.6 | 5.1 | 2.1 | 2.1 | 3.4 |
| Sorsogon | 0.7 | 1.5 | 0.9 | 3.3 | 1.5 | 11.7 | 14.1 |
| Region 5 | 0.4 | 0.7 | 0.5 | 1.5 | 0.7 | 1.8 | 2.5 |

**Table 10 (continued)**

| Province | TFA | PCA | FCA | RCA | NOHH | CCA | PNA |
|---|---|---|---|---|---|---|---|
| Aklan | 0.8 | 2.0 | 1.5 | 3.0 | 1.9 | 10.0 | 18.3 |
| Antique | 0.7 | 4.0 | 1.2 | 1.7 | 1.3 | 5.4 | 7.4 |
| Capiz | 1.1 | 4.4 | 1.3 | 1.5 | 1.6 | 8.0 | 8.8 |
| Iloilo | 0.6 | 2.7 | 0.7 | 1.1 | 0.8 | 2.3 | 3.5 |
| Negros Occ. | 1.6 | 6.2 | 1.6 | 3.3 | 3.2 | 4.5 | 10.8 |
| Region 6 | 0.7 | 2.2 | 0.7 | 0.9 | 0.8 | 2.6 | 3.5 |
| Bohol | 0.6 | 1.3 | 0.7 | 1.8 | 1.0 | 2.5 | 4.4 |
| Cebu | 0.6 | 2.3 | 0.7 | 8.7 | 1.4 | 1.2 | 5.7 |
| Negros Oriental | 1.1 | 3.9 | 1.5 | 6.1 | 1.9 | 2.5 | 12.0 |
| Siquijor | 0.9 | 5.4 | 2.3 | 10.3 | 1.9 | 4.7 | 13.5 |
| Region 7 | 0.5 | 1.5 | 0.6 | 1.9 | 0.8 | 1.2 | 7.8 |
| Eastern Samar | 1.0 | 1.7 | 1.2 | 3.7 | 2.0 | 20.5 | 15.1 |
| Leyte | 0.7 | 1.3 | 0.8 | 1.8 | 0.9 | 3.7 | 5.3 |
| Northern Samar | 0.9 | 1.3 | 1.0 | 4.3 | 1.4 | 13.1 | 11.9 |
| Western Samar | 0.8 | 1.3 | 1.1 | 4.0 | 1.6 | 3.8 | 9.0 |
| Southern Leyte | 0.9 | 1.0 | 0.9 | 4.3 | 1.5 | 14.7 | 9.3 |
| Region 8 | 0.4 | 0.6 | 0.4 | 1.4 | 0.6 | 2.9 | 3.9 |
| Basilan | 2.2 | 2.8 | 2.5 | 10.9 | 5.7 | 14.2 | 8.9 |
| Sulu | 1.6 | 1.4 | 1.6 | 9.4 | 1.7 | 18.8 | 6.1 |
| Tawi-tawi | 2.7 | 3.8 | 4.4 | 20.7 | 4.2 | 8.3 | 9.1 |
| Zam. del Norte | 0.9 | 2.2 | 1.6 | 6.0 | 1.5 | 3.7 | 7.9 |
| Zam. del Sur | 0.6 | 1.5 | 0.7 | 3.6 | 1.1 | 1.7 | 5.0 |
| Region 9 | 0.5 | 0.9 | 0.7 | 2.9 | 0.8 | 1.6 | 3.3 |
| Agusan d. Norte | 1.4 | 2.4 | 1.8 | 9.4 | 1.8 | 4.4 | 12.4 |
| Agusan del Sur | 1.6 | 6.3 | 1.1 | 5.5 | 2.0 | 5.2 | 21.6 |
| Bukidnon | 1.5 | 2.5 | 1.5 | 6.4 | 1.8 | 2.5 | 7.2 |
| Camiguin | 3.1 | 4.5 | 3.7 | 16.9 | 2.7 | 10.2 | 11.4 |
| Misamis Occ. | 0.7 | 1.9 | 1.1 | 9.7 | 2.0 | 6.5 | 14.2 |
| Misamis Or. | 1.0 | 2.2 | 1.3 | 5.3 | 1.2 | 2.7 | 31.4 |
| Surigao d. Norte | 1.3 | 2.3 | 1.7 | 3.9 | 1.8 | 12.0 | 11.0 |
| Region 10 | 0.7 | 1.0 | 0.7 | 3.0 | 0.7 | 1.8 | 6.6 |
| Davao | 1.3 | 2.5 | 1.3 | 7.7 | 2.4 | 5.2 | 6.8 |
| Davao del Sur | 1.1 | 2.4 | 2.4 | 10.3 | 1.8 | 6.6 | 7.3 |
| Davao Oriental | 1.5 | 2.5 | 2.4 | 7.4 | 2.1 | 4.9 | 12.4 |
| South Cotabato | 2.0 | 5.0 | 2.0 | 3.8 | 1.8 | 2.8 | 5.8 |
| Surigao del Sur | 1.7 | 2.5 | 1.8 | 2.6 | 2.7 | 7.3 | 12.8 |
| Region 11 | 0.7 | 1.3 | 0.9 | 3.0 | 1.0 | 2.2 | 3.9 |
| Lanao del Norte | 1.5 | 2.8 | 0.8 | 3.1 | 1.8 | 2.9 | 17.2 |
| Lanao del Sur | 1.0 | 2.5 | 0.9 | 1.1 | 0.9 | 2.0 | 4.5 |
| Maguindanao | 1.2 | 3.7 | 1.1 | 2.1 | 0.9 | 2.5 | 10.0 |
| North Cotabato | 1.1 | 4.5 | 1.3 | 2.7 | 1.3 | 3.4 | 5.2 |
| Sultan Kudarat | 2.4 | 9.0 | 1.3 | 2.9 | 1.7 | 3.4 | 11.9 |
| Region 12 | 0.6 | 1.8 | 0.5 | 1.1 | 0.6 | 1.4 | 4.7 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Philippines (excl. NCR) | 0.2 | 0.3 | 0.2 | 0.4 | 0.2 | 0.6 | 1.4 |

## IV. The 1991 CAF SAMPLING STRATEGY

### A. Sampling Procedure.

*Exercise 3* was adopted by NSO, except in the National Capital Region (NCR) where the sampling scheme used will be discussed in subsection IV.C below. The barangay frames were constructed by coupling the 1990 CPH barangay lists with the 1980 CAF data on total farm area (TFA). The barangays in each town were ranked by descending values of TFA. Those that did not have TFA values -- because either they were new, or they were no longer the same barangay in 1980 due to splitting or merging with other units -- were appended to the bottom of the ranked list either alphabetically or by increasing value of their numeric codes. These barangays without TFA values number 2,836 (see Table 11), or 7% of the total.[9] The barangay on top of the list is automatically part of the sample. Imagine the next two barangays to have labels {1} and {2}. The rest of the sample is chosen by selecting one of these with equal probability, and every other unit down the list. The sample sizes are given in Table 11 (last column). The 21,011 sample barangays represent a 52% sampling rate nationwide (excluding NCR).

[9] There were some well meaning suggestions to use probability proportional to size (pps) sampling or stratification of the barangays instead of systematic sampling, to further improve the precision of the estimates. We did not think these were practicable options since pps will require working out and assigning size measures to these 2,836 units, and stratification will require the same for the technique to effectively reduce the sampling error of estimates. It was mentioned earlier also that some towns have too few barangays to allow effective stratification. On the other hand, systematic sampling can be applied readily with an incomplete array of size measures. There are other reasons which will be discussed in Section V.

**Table 11. Barangays Without Size Measures (TFA) and Sample Sizes, By Region**

| Region | Barangays w/o TFA | Sample Size |
|---|---|---|
| 1 | 32 | 2,073 |
| 2 | 363 | 1,423 |
| 3 | 163 | 1,510 |
| 4 | 244 | 2,762 |
| 5 | 37 | 1.786 |
| 6 | 34 | 2,096 |
| 7 | 68 | 1,562 |
| 8 | 343 | 2,264 |
| 9 | 629 | 1,414 |
| 10 | 336 | 1,304 |
| 11 | 309 | 922 |
| 12 | 278 | 1,895 |
| Total, excl. NCR | 2,836 | 21,011 |
| Per Cent of Total | 7.0 | 51.9 |

There was complete coverage of all households in the sample barangays regardless of type (farm and non-farm) and size of landholding. Three forms were used during the February 1992 data collection by face-to-face interview. CAF Form 1, which is also called the Control Sheet, is the complete list of households, in which pertinent information is recorded which allows the identification of agriculture and fishing households or operators. CAF Form 2 is the questionnaire for the agriculture part of the census; i.e. it is used to record information on crops, livestock, poultry, farm equipment, etc. from every farm operator identified in CAF Form 1. CAF Form 3 is the questionnaire for fisheries. In addition, two more forms were used after the data collection. CAF Form 4, which is derived from CAF Form 2, is a worksheet produced for each sample barangay. It contains a list of farm operators together with selected summary farm characteristics such as total physical area, area planted to rice and to corn,

number of coconut and banana trees or hills, and inventory of major livestock and poultry species. CAF Form 5 is the corresponding worksheet from CAF Form 3, which contains selected operator level areas under fishpond, fishpen, fishcage, mussel culture, etc.

It should be mentioned also that separate lists of large "corporate-type" farms, i.e. those managed by cooperatives, partnerships, corporations and the like, were constructed independent of the implementation of CAF Form 1. These lists extend to the non-sample barangays. These were enumerated completely and merged with the information from the other CAF Forms for each relevant barangay; thus, sample barangay totals obtained from CAF Forms 4 and 5 for example, are supposed to reflect (but not count twice) the contribution from these types of farms.

**B. Estimation Procedure (for the Agriculture Part)**

The straightforward design estimator of a town total is Eq. (4) above. Higher aggregates -- province, region and national -- are obtained by simple summation of the town estimates.

Whenever feasible and desirable, potentially variance-reducing techniques should be explored during the estimation stage, like using ratio-type estimators. Let an auxiliary variable whose total X for the town, the value $x_c$ for the certainty barangay and $x_1$, ..., $x_{n'}$ for the sample barangays are known, where $x_1 + ... + x_{n'} > 0$. A ratio-type estimator of the town total Y is

$$\bar{Y}_R = y_c + \left(\sum_{n'} y_j \Big/ \sum_{n'} x_j\right)(X - x_c) \qquad (7)$$

The mean square error of this estimator is

$$MSE(\bar{Y}_R) = (1/2)\{(\bar{Y}_{R1} - Y)^2 + (\bar{Y}_{R2} - Y)^2\} \qquad (8)$$

where $\bar{Y}_{R1}$ and $\bar{Y}_{R2}$ are values of $\bar{Y}_R$ from the two possible samples. It is known that ratio-type estimators are more efficient for choice of auxiliary variable satisfying $2\rho > CVx/CVy$, where $\rho$ is the correlation coefficient between x and y. The logical sources of auxiliary variables are the 1980 CAF and

1990 CPH.

**C. Sampling Procedure in the National Capital Region**[10]

There are two reasons why one bothers with NCR in an agriculture census. One has to do with the so-called "household approach" of recording information in agriculture censuses (and surveys), i.e. information is credited to the barangay where the respondent household resides regardless of the physical location of the farm or activity. Thus, farms located outside NCR that are operated by households in NCR will be missed by the sample described above. Conversely, farming and fishing done inside NCR by households living outside NCR will be accounted for in the aforementioned sample (although in all likelihood these will be negligible). However, farming and fishing inside NCR by nearby residents (e.g. those garden plots near the international airport or the bay fishing and mussel harvesting in Paranaque) can only be accounted for by including NCR in the sampled population.

There are close to 1,700 barangays in NCR, which represent 4 per cent of the over 40,000 nationwide (see Table 1). In 1990, 1.57 million households, which comprise 14 per cent of 11.41 million Filipino households, were living in NCR. In 1980, however, only 0.8 per cent total farm area was operated by NCR residents.

The NCR is divided into four subregions, each of which is subdivided into districts. The sampling was done independently in each district in accordance with the following instructions: Include with certainty any barangay which satisfies any one of the following criteria based on its 1980 CAF data:

$\geq$ 100 hectares of farm area, or

$\geq$ 100 pigs, or

$\geq$ 500 heads of poultry, or

$\geq$ 50 households were engaged in municipal fishing, or

$\geq$ 1/3 of households were engaged in municipal fishing (regardless of the number).

Choose 1/3 of the remaining barangays using simple systematic sampling.

---

[10] I thank Gener de Guzman, who recommended the NCR sampling scheme to NSO, for verifying and supplying some of the details in this section.

This sampling procedure resulted in 496 sample barangays, 114 of which had inclusion probability equal to one. The allocation by district is shown in Table 12. To the extent of our knowledge, NSO is inclined to use simple expansion formulas to estimate subtotals for the sampled subsets of barangays in each district.

**Table 12. The NCR Sample (Number of Barangays)**

| Subregion-District | Sample Size a/ |
|---|---|
| I | |
| Ermita | 4 (1) |
| Malate | 15 (2) |
| Paco | 11 (0) |
| Pandacan | 10 (0) |
| Intramuros | 1 (0) |
| Port Area | 1 (0) |
| Quiapo | 4 (0) |
| Sampaloc | 55 (6) |
| San Miguel | 3 (1) |
| Binondo | 3 (0) |
| San Nicolas | 3 (0) |
| Sta. Cruz | 23 (3) |
| Sta. Ana | 25 (1) |
| Tondo | 67 (3) |
| II | |
| Mandaluyong | 8 (2) |
| Marikina | 5 (2) |
| Pasig | 13 (7) |
| Quezon City | 50 (21) |
| San Juan | 9 (4) |
| III | |
| Kalookan City | 55 (8) |
| Malabon | 13 (11) |
| Navotas | 8 (6) |
| Valenzuela | 17 (13) |
| IV | |
| Pasay City | 51 (2) |
| Las Pinas | 7 (3) |
| Makati | 13 (6) |
| Muntinlupa | 3 (1) |
| Paranaque | 7 (4) |
| Pateros | 2 (0) |
| Taguig | 10 (7) |
| TOTAL | 496 (114) |

a/ The first number represents the total sample size and the second in ( ) the certainty barangays; e.g. there were 4 sample barangays in Ermita, including I which was drawn with certainty.

# V. WHY NOT SOME OTHER SAMPLING STRATEGY? AND RELATED QUERIES

## A. Introduction

Doing a census involves execution of a number of steps from initial planning to final release of the results. Planning a sample census (or survey) is, to a large degree, a process of choosing for each required step one procedure from a number of competing alternatives: if there are k steps with $A_1,...,A_k$ alternatives, the resulting plan is one combination from the $A_1x$ ... $xA_k$ possibilities. In theory, the optimum plan is that combination which minimizes total cost for specified level(s) of accuracy of the estimate(s), or that which maximizes accuracy for fixed cost. In reality, the optimum plan either does not exist, or else the algorithm for finding it does not exist. Thus, at best we strive for the approximately optimum or "proximum" (Kish, 1965).

The fact that the process is dynamic adds to the complexity of the problem of census planning. The great bulk of the literature simplifies the problem by dealing with choosing from among alternatives in only one or two of the required steps -- ignoring what happens to the rest of the steps or holding them constant. In real world situations, however, decisions on any one step affect the properties as well as the number of options available for the other required steps. For example, electing to collect area data by interview instead of actual measurement will allow a bigger sample size (for fixed cost), thereby reducing sampling variance though not necessarily mean square error; the sample size required to produce a predetermined sampling error level will depend not only on the inherent variability of the variable(s) of interest and on the sampling procedure chosen, but also on the number of strata, which also depends on where to locate the stratum boundaries, which also depends on the allocation scheme to be used, etc.

It will be surprising if there would be unanimous agreement on the plan for a large undertaking like the agriculture census. Conflicting preferences are bound to surface, particularly on the sampling strategy, i.e. the sampling scheme and estimation procedure. Some of these are presented here for their informational and possible pedagogic value.

**B. PPS Sampling or Stratification of the Barangays in Each Town**

As an alternative to systematic sampling of the ordered barangays in each town, Megill (1991) and Sturdevant (1991), advisors from the US Bureau of the Census suggested using similar CAF 1980 information to draw the sample barangays with probability proportional to size (without replacement, presumably). They also proposed stratifying the barangays, including a certainty stratum, i.e. composed of the largest barangays which will be drawn with probability one. The justifications advanced were: (i) "this should improve the efficiency of the sample design considerably (in terms of higher precision for a lower cost)" and (ii) "Although these alternatives would result in differential weights by barangay, this should not present a problem at the estimation stage since the tabulations will be processed by computer."

Although improvement in precision may be achieved by the alternative proposals, this would not necessarily be at lower cost. Even if feasible, the construction of relevant size measures for 2,836 "new or altered" barangays (see Table 11) and for the towns similarly affected (see Table 1) will not be without extra cost and time.[11] Moreover, the advantage of stratifying very small towns (see Table 2) relative to sorting the units according to size and drawing a systematic sample is suspect. The former will also definitely lead to unequal weights for the sampling units.

What was missed in the proposals is that we aim for a simple sampling scheme not only to keep the estimation for the census simple, but more crucially, to ensure that the estimation theory remains tractable when using the barangay level census data as frame in redesigning the intercensal agricultural sample surveys. As mentioned previously, the domain in the sample surveys will be the province. In so far as the surveys are concerned, the barangays from the census will serve as a first-phase sample, which will be manipulated further in a variety of ways (e.g. multi-way stratification), before a second-phase sample is drawn (e.g. using pps sampling or replicated sampling with partial

[11] The author still recalls a project in the 1970s by the Research Department under Lilia Constantino in the (then) National Census and Statistics Office to reconcile town level data from past censuses. The process required going through congressional records to establish which old villages went into which new towns.

replacement between rounds). In this scheme of things, the estimation theory and method will become very complicated vis-a-vis the technical capability available in the country -- notwithstanding the presence of computers -- unless the first phase sample is kept simple. The unequal inclusion and joint inclusion probabilities of the barangays in the first-phase sample, and problems of adjusting these weights as the size measures change through the years, will have to be reckoned with in the estimation theory for the sample surveys. Additionally, the problem of mis-stratified units to begin with, and the movement of units across strata through time, will cause losses in precision and further complications in the estimation procedure (see e.g. David, 1985). Furthermore, to dismiss the problems brought about by these complications "since the tabulations will be processed by computer" overlooks the chronic lack of mathematical statisticians (as opposed to data processing personnel) in the government statistical service.

In contrast, the sampling scheme used in the agriculture census suggests a natural initial stratification of the first-phase sample in a province into two -- the certainty barangays (each with weight 1) and the sample barangays (each with weight 2). Alternatively, every barangay in the first stratum may be split into two more or less equal sized sampling units, in which case the entire first-phase units will have equal weights. These are just some of the possibilities, all leading to more tractable overall sampling strategies for the intercensal surveys.

The criticisms on the proposed sampling procedure seemed to stem also from the (mis)conception that the use of systematic sampling "guarantees equal representation of barangays with large and small agricultural operations", hence it cannot be much more efficient than simple random sampling. This is, of course, not true: ranking the units according to size and drawing a 50 percent systematic sample is comparable to constructing strata each of size 2 units that are most similar in terms of the same measure of size as that proposed for pps sampling and stratification (except that no attempt is made to construct size measures for the new or altered units), and choosing one unit in each stratum. One difference is that the relative position of the sample unit in the strata is the same with systematic sampling, whereas it is not in the case of stratified simple random sampling. It is true that the probability is positive that one sample may overestimate the target parameter and the other will underestimate it, as

pointed out by Sturdevant (1991). However, our sampling experiments show that, with few exceptions, this is not a serious problem; and special handling of the few cases where it could be a problem has been recommended. On the other hand, it is not evident, unless shown empirically, that stratification carried out less intensively than 2 units per stratum will be more efficient that systematic sampling as proposed here.

Last but not least, given a choice between using auxiliary information at the sampling stage to improve efficiency of estimates, but be stuck with unequal inclusion and joint inclusion probabilities on the one hand, and utilizing the same auxiliary information during the estimation stage to possibly gain comparable precision, and be left with a much simpler sampling plan on the other hand, this writer tends towards the latter in this particular application. Moreover, the latter offers further flexibility in two ways -- viz., (i) ratio-type estimates may be used only in characteristics where suitable auxiliary variables that reduce the sampling errors are found and (ii) the auxiliary variable chosen may vary for each main characteristic of interest.

**C. Use of Variable Sampling Rates**

Another proposal (Sturdevant, 1991) was to draw barangays with probability proportional to size, then vary the sampling rate of the households (proportional to the inverse of the selection probability of the barangay) to make the sample self-weighing. It was proposed further that the sampling be carried in three stages in some cases, by splitting the large sample barangays (drawn with pps) into segments, select sample segments, and select households (using variable sampling rates). The main objective is to improve efficiency of the census estimates; i.e. lower sampling errors for the same cost, or lower cost for the same sampling error levels as those demonstrated by the sampling scheme described in section IV.

This writer's reaction to the proposal is that it would perhaps be appropriate for household sample surveys, but not for a one-time massive census, much less an agriculture census. Our recollection of the NSO management response to the proposal was *a preference to be somewhat economically inefficient rather than complicate operations and data processing.* Hence, the large

sample barangays were indeed split into enumeration areas (EA) each consisting of no more than 300 contiguous households (enough for one interviewer to cover within the allotted 20 days for field data collection). However, all the EAs were enumerated, the main reason for their creation being to equalize workload among the field staff.

**D. Multiple Frames**

There was also a proposal to complement the barangay frames used for the sampling plan described above with a list frame of the largest agricultural estates and agribusiness operations, including large fishery establishments. This frame is to cover all such units, including those in the non-sample barangays. The units in this *certainty list* were to be enumerated completely. This was done, as discussed in Subsection IV.A.

It was proposed further that the *certainty list* include large household operated farms, large being defined either in terms of area (the example given was 100 hectares or more) or value of production (which is not easily obtainable). This was not done, because such farms located in the sample barangays will be captured by the census proper anyway, in which case design-based estimators like Eq. (4) will correctly (in the statistical sense) reflect the share of such farms in the estimates of totals. Hence, the author cannot comprehend the concern expressed by one advisor, that "As now planned, large operations included in selected barangays will be over-represented, while those in excluded barangays would be under-represented." (Much later, an effort was made at NSO to try to "reach the large farms in the non-sample barangays"; see next Section VI. It is not known to the writer whether this was an independent decision on the part of NSO or whether it was influenced by the above-quoted statement.

## VI. EPILOGUE

Subsequent to the drawing of the sample, some decisions or activities were made which, if known beforehand, could have influenced or even changed the recommendations concerning the sampling plan. Three of these are discussed here.

### A. Complete Coverage in Four Provinces

An **advance census** was to be conducted in Bukidnon, Isabela, Laguna and Marinduque. Hence, as mentioned in Subsection III.B, field operations covering the sample barangays in these provinces were conducted in October 1991. During the **main census** operations in February 1992, the complement samples -- i.e. the subsets of non-sample barangays -- in the four provinces were covered also. Thus, it can be said that complete enumeration censuses were conducted in these provinces if deviations caused by the difference in interview dates can be assumed away as negligible. Such an assumption might be tenable on the grounds that any changes in farm household populations and agricultural activities in a four month period would be small, particularly since the questionnaires do not include production and other seasonally sensitive characteristics, but more structurally stable ones such as areas planted, number of productive perennial trees, and inventories of livestock and farm equipment.

We are not privy to all the reasons and circumstances behind this fortuitous decision. The data from these four provinces can provide actual values of the sampling errors for some the main variables of interest, and for alternative estimation procedures.[12] This is uniquely valuable since, as mentioned previously, a single systematic sample cannot provide a straightforward and valid estimate of the sampling error.

[12] This will be the topic of a planned sequel to the present paper.

**B. Inclusion of Fisheries and Aquaculture Information**

Although the sample of barangays was chosen exclusively for agriculture, a decision was made later to collect fisheries and aquaculture data from the same sample. The consequence will be very high sampling error of estimates. The sampling errors, however, are not estimable from the main census. Again, the four provinces where full enumeration censuses were taken can be used to provide estimates and do numerical studies, including alternative estimation methods, to find out at what levels (province or region?) can estimates be shown to have acceptable sampling error levels, and hence publishable only at those levels.

Whenever feasible, the use of variance reducing estimators like $\__R$ given in Eq. (7) should be exlored. One problem would be the paucity of suitable auxiliary characteristic x. The 1980 CAF Form 1 contained barangay data on number of households engaged in fishing (municipal and marine) and it remains to be verified whether these still correlate highly with 1991 fisheries data. However, the 1980 CAF lists did not include items on aquaculture, e.g. seaweed or mussel gathering and farming, as well as prawn farming which became big business only in the 1980s. Moreover, since compared to farming, fishing and aquaculture tend to be rare and unevenly distributed activities, the probability of the event $'_1 + ... + '_{n'} = 0$ can be high.

**C. Second Thoughts About the Large Farms**

For reasons which have remained obscure to us, concern about the <u>non-coverage</u> of the large, household operated farms in the non-sample barangays kept coming to the surface even after the field data collection had been completed. For to sample is to choose only a part, deliberately missing the remainder from the whole; and Statistics' and statisticians' *raison dêtre* is to build bridges from the sample to the whole, <u>without having to observe the remainder</u>. This is the very essence of statistical inference, which is at the core of the modern scientific method.

The process can be discerned easily from the estimator _' in Eq. (4). The population of town barangays is split into two strata. The question of non-coverage in the first stratum (consisting of the largest barangay) does not arise, since the entire stratum is enumerated (drawn with probability one). The expression $\Sigma y_j$ from the n' barangays sampled from N' = N-1 in the second stratum does not "cover" those outside the sample; however, $(N'/n')\Sigma y_j$ does, in the sense that it estimates (infers about) the total of the N' barangays. In fact, were it not for the negligible bias brought about by systematic sampling when N' = odd, _' is otherwise design unbiased. Hence, there are no grounds to raise questions of non-coverage and over- or under-representation.

NSO eventually compiled a list of households with holdings 7 hectares from the 1980 CAF Form 1 data files and in October 1992 went to the non-sample barangays nationwide in an attempt to enumerate these "large" farms. Instead of resolving anything, this could raise more problems. First, one reason for sampling was to save money (the assumption being, there was not enough to do a complete enumeration census). Going back to the non-sample barangays effectively meant negating the earlier decision to sample, thus, perhaps wiping out whatever savings were realized. Second, how can the design-based estimates (4) and (7) be modified to make use of the additional information? The task can be daunting from a theoretical standpoint. And third, developing procedures for matching/merging of the data from the 1980 CAF-based lists and collected in October 1992 on the one hand, and the data from the main census obtained in February 1992 on the other hand, to arrive at 1991 estimates for large farms can also be challenging and interesting to watch.

## References

Cochran, W.G. (1977). *Sampling Techniques*, Third Edition, John Wiley, New York.

Dalenius, T.

David, I.P. (1985). Keeping Sample Survey Design and Analysis Simple. *Statistical Report Series*, No. 5, ADB, Manila.

David, I.P., A.N.C. Mendoza and D.C. Besa (1990a). Evaluation of Philippine Corn Statistics, *Journal of Philippine Development*, Vol. XVII, No. 1, 33-66.

Kish, L.

Mogill, D. (1991). Some Considerations for Improving the Efficiency of the Sample Design for the 1991 Census of Agriculture and Fisheries, June 19, 1991 (unpublished).

Sturdevant, T. (1991). Comments Concerning the Proposed Sample Design for the 1991 Census of Agriculture and Fisheries, June 20, 1991 and Observations Concerning Sampling Plans for the 1991 CAF, August 6, 1991 (unpublished).

Yates, F. (1981). *Sampling Methods for Censuses and Surveys*, Fourth Edition, Charles Griffin, London.