The Philippine Statistician
Vol. xxvi, Nos. 1 & 2 1977

# PRACTICAL PROBLEMS OF GENERATING DATA, WITH REFERENCE TO AN URBAN AND RURAL COMPARISON OF THE DEMAND FOR SOCIAL SERVICES IN THE SUKABUMI REGION OF WEST JAVA

by ROSS BULL*

## Introduction

The major purpose of this paper is to describe the decisions which have to be taken in designing a sample survey, and to discuss, in detail, the relatively simple and practical means by which raw data can be generated. The paper attempts to illuminate some of the practical problems involved in implementing such a design, and will discuss the practical reasons for necessarily deviating from theoretically accepted principles and standards. It is hoped that the main audience for this paper will be those social scientists who have an inherent mistrust of statistics and statistical methods. The paper hopes to demonstrate that collecting statistically reliable data is interesting, rather than tedious.

The survey was carried out, during September-November 1976, as an integral part of the Indonesia Development Studies (IDS) course at the University of Indonesia. This course, which concentrates on social problems of relevance to developing countries and, in particular, Indonesia, is a postgraduate training programme which is organized jointly by the University of Indonesia, and the Institute of Social Studies, at The Hague. A major component of the course is the design and implementation of an applied research project.

## Designing a Research Project

There are, in general, six stages involved in designing a research project.

* Professor, Institute of the Social Sciences, The Hague, Netherlands

First, it is necessary to identify a topic of interest. The IDS chose, as its topic for 1976, the delivery of social services. Although it is often difficult to identify a major topic of interest, when attempting to formulate constructive research proposals, the IDS felt that this particular subject was of extreme relevance, both to the course participants — of whom a number were planners — and to Indonesia, because, given the current, rapid expansion of population in Indonesia, increasing demands will be placed on all public sector services. Assuming future resource limitations, the services demanded will not necessarily be delivered.

Second, a topic as broad as that outlined above is insufficiently specific for any useful research to be carried out. It was necessary, therefore, to refine the topic, and to focus on one, or more, particular research objectives. A number of alternative objectives were considered for the project, although a final choice was made between two alternatives. The first objective considered was to examine the supply of social services. This would have necessitated an analysis of government, regional and local authority planning activities in the social services sector, with a subsequent assessment as to whether resources were being allocated in a useful way, in the light of given demand estimates.

This led to the consideration of the alternative objective — namely, the demand for social services. To our best knowledge, no current estimates of the demand for social services in Indonesia exist, although there are population and age-specific estimates, in addition to government policies and targets to provide specific groups in the community with a certain minimum standard of public services. It was felt, therefore, that the second, alternative objective — examining the demand for, rather than the supply of, social services — would be the most logical topic for research, because it would only be sensible to assess supply patterns, and the subsequent allocation of resources, if it was possible to turn to an independent assessment of what demand might actually be. In the absence of any estimates of demand, it seemed sensible to try and produce such estimates.

The third stage in designing a research proposal is to determine the method appropriate to meeting the research objective. It is frequently observed that an attempt is made to make the research objective fit the method, rather than choosing the appropriate method to realise a particular re-

search objective. It is the latter, rather than the former, approach which is correct. In the absence of any estimates of demand for social services, it was decided to undertake a sample survey, in order to try and estimate, i.e. to gain an impression of, what this demand might be. This paper describes the way in which this survey was designed.

It was decided to base the survey on a probability sample, to ensure that data was collected from a representative cross-section of the population, rather than concentrating on trying to find a number of "representative" villages where "in depth" analysis of demand could be carried out. The relative merits of these two methods of analysis, and estimation, will be discussed below.

Fourth, it is necessary to be certain, in advance, of the type of analysis necessary to attain the research objectives. It is frequently the case that researchers begin a study without being certain of the final analysis which they want to undertake. This leads, inevitably, to a waste of time, manpower and money — all substantial constraints in the particular survey under discussion — because it is likely that, at the end of the data collection stage, it will be found that the data available is inadequate to permit the later analysis.

Whilst it is difficult to think out, in advance, what data is required to meet the research objectives, prior thought will avoid, or reduce, the wastage of scarce resources. In the survey under discussion, it was decided that, in the absence of any demand estimates, it was of extreme importance to concentrate on determining facts about the population, with respect to its current and potential use of social services. It was decided that this was the major analysis that would be required, although it was appreciated that it would be desirable to try and establish relationships between variables through the examination of cross-tabulations.

The fifth stage is to determine which type of data is required, in order to carry out the analysis and to meet the research objectives. In this particular survey, the required analysis determined immediately the questions which would have to be included in the survey questionnaire. It is often the case that researchers, involved in survey work, will write the questionnaire in advance of thinking out what analysis is required, or what hypotheses should be tested. This approach is incorrect; it is likely to lead to the generation of

data which, whilst of interest in itself, might be of little use for specific hypotheses, cross-tabulations, or estimates.

The sixth, and final, stage in designing a research proposal is, perhaps, the most contentious: and that is that the research should be policy oriented, especially in developing countries. Given that resources are scarce, especially statistical (manpower) resources, any research should attempt, as far as possible, to produce data and results which are of maximum interest to planners and policy makers. The policy orientation of the survey under discussion was geared to producing demand estimates, where none existed, in order that working assumptions used by planners could be either endorsed, or challenged (with subsequent implications for any planned expenditures on the related social services).

**What is Meant by Demand?**

Having established that the research objective of the project was to examine the demand for social services, it became necessary to decide, a) which services were to be considered; b) which population was to be sampled; and, c) the exact definition of demand that would have to be used.

The first decision — which social services were to be considered — required concentrating on a limited number of possible social services. It was impossible to cover all services, because the questionnaire used was deliberately constrained to a maximum size. The major services eliminated from the enquiry were administrative and security, housing, transport and communications, and electricity, whilst those finally chosen for study were education, and health and sanitation (including family planning and nutrition).

The choice of the population to be sampled was an administrative decision. The survey was the first of its kind, within the IDS; as such, there was no prior knowledge or information about the difficulties that might be experienced in the field, and there was no knowledge about what could be achieved in the limited time available, given the budget and manpower constraints. Therefore, an area relatively close to Jakarta was chosen as the target area — namely, the Sukabumi region of West Java.

Given the research objective, however, it became clear that

the demand for social services was unlikely to be the same over the whole of the Sukabumi region. Whilst there would undoubtedly be intra-regional variations in demand, which could be accounted for in a highly complex survey design, a major hypothesis of the project was that there would be a significant difference in the urban and rural demands for social services; and it was on this anticipated difference between the urban and rural areas that the project was focussed.

A third necessity was to establish an acceptable, and meaningful, definition of demand. What exactly is meant by demand? Demand relates to both quantity and quality. It is relatively simple to interpret data about quantity demanded, but it is less easy to interpret data about quality demanded, especially in areas where (perhaps) quality means nothing if the population has not been exposed to a continuous, changing quality of goods and services.

Bearing in mind this difficulty in interpreting between quantity and quality demanded, four alternative interpretations of demand were considered. First, the level of satisfaction was considered as a possible way of reflecting demand. If the population said it was dissatisfied with a social service, this would tend to indicate that demand was in excess of supply. In this event, it would be necessary to identify by how much the supply (of a given, or alternative, quality) was deficient. It was felt likely that asking questions about satisfaction with the quantity, and the quality, of a service would lead to answers which could, justifiably, be interpreted as meaning one thing to one person, and another to a second. This interpretation was rejected.

The second interpretation, which was also rejected, concerned the concept of perceived need. Demand can be viewed as the requirement that a population has, as perceived either by itself, or on its behalf i.e. subjectively, by a third party, normally a government agency. The likelihood of a government agency ever being able to perceive realistically the needs of a population is small, especially when trying to differentiate between quantity and quality.

But, can a population perceive its quantitative needs, and especially its qualitative needs? It is natural to reply positively to this question. Whether a population really knows what it needs, however, depends on whether if understands fully the implications of saying that it does or does not need a certain

quantity, and/or quality, of service. Populations do not necessarily have this information.

A second danger in asking a population to state its perceived needs is that it will answer that it needs everything — both in increased quantity and quality. This is not helpful to the planner, given that resources are finite: he/she needs to know priorities and whilst, for this reason, the second interpretation of demand was rejected, it raised the issue of a third.

As a third possible interpretation of demand, it was considered that, as price is generally observed to be related to quantity demanded, the population should be asked to indicate what price it would be prepared to pay for a given change in the quantity and/or quality of a service. In other words, how much income (and/or harvest) would the population be prepared to forego, either in the form of taxation or private expenditure, in order to have a certain output provided. In addition, it would be useful to determine, as an indicator of demand, the relative placings of expenditure on alternative types of social services: for example, how important would the population conceive of expenditure on education *viz a viz* housing, and what reasons would be used to justify such a ranking. This interpretation of demand was rejected, because it was felt that the population in question — especially the rural population — would find it impossible to conceive of ranking priorities, or to understand what was meant by foregoing income to obtain a greater quantity, or better quality, of social services.

There is difficulty, therefore, in defining and interpreting quantity, and especially, quality demanded. To that extent, the survey confined itself to trying to assess the quantity, rather than the quality, demanded. Bearing this in mind, the fourth approach to identify demand was the one which was used: that was, to identify the current use of different types of social services; to identify reasons for their use, and non-use; to determine whether the population was better off as a result of using, or not using, the social services; and to determine what the implications for the population might be of using, or not using, these services further.

As the survey was a pilot study — see section 4 — it would have been sensible to construct a questionnaire which would have emphasized the four different approaches to demand. Unfortunately, due to time limitations, it was impos-

sible to do this completely.

### Pilot Study Versus Full Survey

There was never any doubt that, given the tight time schedule imposed on the IDS research project — see appendices 1A and 1B — the survey could only be interpreted as a pilot study which, in the event of undertaking a full survey, would be a (very) necessary condition for the successful completion of that full survey. A pilot study is absolutely vital, for several reasons: to test the questionnaire's reliability in obtaining the data which is required; to test the coding rules which will be used in analysing the data; to identifying major definitional problems; to obtain an approximate idea of the likely errors involved in the statistical estimates; and to illuminate the problems that are likely to be encountered, both in the field and in the office, when undertaking a full survey.

### The Sample Design — Practical Considerations

A discussion of the technical considerations for the sample design used in the survey is included in section 6; a complete, non-mathematical, yet technical, description of the sample design is given in appendix 3. This present section, however, will discuss some of the more general, practical problems that have to be dealt with in deciding upon a sample design, and the decisions that have to be taken with respect to that design.

At least seven major factors — all interrelated to varying degrees — have to be considered when designing a sample. The seven factors concern the sample size; the total manpower available; the time available for surveying the population, as well as completing the whole project; the desired coverage of the sample; the budget available; the desired statistical precision of any estimates; and the size of the questionnaire. An example of the interrelationship between these factors — the reader can consider other examples — is as follows. With a limited period for the whole research project, it is difficult to train interviewing staff adequately, unless the budget is of a substantial size: a limited budget, limited manpower, and a limited survey period will limit the size of the sample. This is likely to provide estimates which are of less than the desired statistical precision which, in turn, may mean reducing the coverage of the sample i.e. taking only a part of, rather

than the whole of, a population, or reducing the questionnaire size, (and, thus, the depth of the analysis), or both.

In the Sukabumi survey, the two basic constraints on the sample size were the fourteen days in which the data gathering had to be completed, and the availability of only 25 interviewers i.e. the 25 IDS course participants. Whilst it was desirable to have a coverage of households in the Sukabumi region which was as large as possible, the budget of approximately 5 million Rupiahs[1] did not, in the time available, permit the sample design to allow for more than an expected sample size of 1,000 households, based on an assumed questionnaire length of 1 1/2 hours.

Before explaining the derivation of this sample size, one further piece of information is required. As described, the survey objective was to compare the urban and rural demands for social services in the Sukabumi region. A decision had to be taken, therefore, as to how to allocate the sample size of 1,000 households between the urban and rural areas. Had the project been attempting to make an estimate of total regional demand, the sample of 1,000 households would have been allocated proportionally, according to the number of households in the urban and rural areas. As the objective of the project, however, was to make an urban and rural comparison, it was decided to allocate the sample equally between the two areas.

With this in mind, three assumptions had to be made in order to derive the expected sample size (of 1,000 households). First, because prior knowledge was available from the Indonesian Central Bureau of Statistics, it was known that there was an extremely low rate of non-response in surveys undertaken in Indonesia. As the sample was to be drawn on the basis of the 1971 Census data — see section 6 — it was expected that, due to a substantial population (and thus household) increase since 1971, the actual number of households from whom an interview would be sought would be in excess of 1,000. It was assumed that this excess would more than compensate for any non-response giving, if anything, an expected sample size in excess of 1,000.

Secondly, it was assumed that 100 questions could be asked, and answered, in 1 1/2 hours i.e. approximately the equivalent of one question and one answer per minute. It was decided

[1] Approximately US$12,077.00, at 1976 prices.

that a questionnaire of this size would permit an adequate, depth interview to take place. Whereas a shorter interview, of say 1 hour, would not permit such an adequate analysis, a longer interview, of say 2 hours, would have undoubtedly created too many data processing problems, in the short time available, prior to analysis and reporting.

Thirdly, it was assumed that the urban area interviews could be completed in 4 days, leaving 10 days for the completion of interviewing in the rural area. This substantial period of time was set aside to allow for any transportation difficulties in arriving at sampled villages, as well as allowing for any unforeseen contingencies.

The result of these three assumptions is summarised in table 1.

TABLE 1

| Expected sample size | 1,000 | | 2,000 | |
|---|---|---|---|---|
| Area | Urban | Rural | Urban | Rural |
| Sample size per area | 500 | 500 | 1,000 | 1,000 |
| Interviewers | 25 | 25 | 25 | 25 |
| Interviews per person | 20 | 20 | 40 | 40 |
| Days | 4 | 10 | 4 | 10 |
| Interviews per person per day | 5 | 2 | 10 | 4 |
| Interview hours per day (at 1 1/2 hours per interview) | 7½ | 3 | 15 | 6 |

Table 1 demonstrates the relationship between expected sample size and assumed questionnaire length.

Considering the three assumptions referred to above, it would have been an almost impossible task to collect more than 1,000 interviews if, in fact, the questionnaire was to last 1½ hours, as no interviewer could have been expected to interview for more than 8 hours per day. If the questionnaire had been reduced from 1½ hours to 1 hour, however, a sample of 1,500 households would have yielded exactly the same number of interview hours per day — and thus the same subsequent editing and data processing time — as a sample of 1,000 households based on a 1½ hour questionnaire. As the project was a pilot study, rather than a full survey, it seemed more sensible to have a longer questionnaire, and thus a smaller sample

size, because it permitted a larger and more varied set of questions to be tested in the field.[2]

One final practical consideration should be mentioned. Having decided that a sample size of 1,000 households should be allocated equally between the urban and rural areas, it was necessary to decide the regional coverage that could be expected from such a sample size. Whilst there was no problem in allocating the 500 households over the eight villages that constituted the urban area, it was slightly more complicated to allocate the remaining 500 households over the 145 villages that constituted the rural area. As indicated in section 3, a decision was taken to obtain the maximum geographical coverage. Table 2 helps to explain how that decision was implemented.

TABLE 2

| Number of Villages in Rural Area | Size of Interviewer Teams to visit Villages (to the Nearest Round Figure of Interviewers) |
|---|---|
| 1 | 25 |
| 2 | 12 |
| 3 | 8 |
| 4 | 6 |
| 5 | 5 |
| 6 | 4 |
| 8 | 3 |
| 12 | 2 |
| 25 | 1 |

It can be seen from table 2 that, given available interviewing manpower, it was possible, at one extreme, to send all 25 interviewers to one village; at the other extreme, it was possible to send an individual interviewer to a separate village, thereby permitting the survey to cover 25 villages.

For safety reasons, however, bearing in mind that an individual would be in the rural area on his/her own for a considerable period of time, it was felt unwise to send unac-

[2] But, being a pilot study, was it necessary to have as many as 1,000 households in the sample? One of the objectives of the survey was to expose the IDS participants to the problems that arise from a large-scale research project. These problems could only be experienced, and not simulated.

companied interviewers to the rural villages. It was decided, therefore, to send out 12 interview teams; (eleven of size 2 persons and one of size 3 persons i.e. 22 + 3 = 25 persons).

Given that each interviewer was required to obtain 20 interviews in the rural area — see table 1 — it became clear that the interviewer would be most unlikely to require (the allotted) 10 days in which the complete his/her task. Indeed, it was assumed that, if he/she could complete 10 interviews in 5 days in one village, and 10 interviews in 5 days in another village, the sample size would remain the same, whilst the rural coverage of villages would be increased by 100% i.e. from 12 to 24 villages.

This useful assumption was incorporated into the sample design.

**The Sample Design — Technical Considerations**

As mentioned above, a decision was made to undertake a probability sample, in order to collect data which was fully representative of the Sukabumi region. There are a number of advantages to probability sampling — where each member of a population has a known probability, i.e. chance of being selected in the sample — over— the complete enumeration of a population: probability sampling is cheaper, it is quicker and it is more reliable. At the same time, it has a large advantage over the in-depth analysis of a subjectively chosen "representative" section of a population. The probability sample is selected at random in which, if the sample is designed properly, it is possible to ensure that all extreme values will be incorporated in the data collected. This will yield data from a good cross-section of the population which, when aggregated and analysed, will yield estimates representative of that population.

The purpose of the Sukabumi survey was to compare estimates of the demand for social services in the urban and rural areas of the region. In the absence of knowledge about the actual, true demand for social services, it was unlikely that any estimate that was made would be equal to that unknown true value. (Even if it was equal, this fact would be unknown. This is true, whether probability sampling, or any other type of sampling e.g. quota sampling or in-depth analysis is used.)

The major objective in designing a probability sample, therefore, is to ensure that the difference between the estimate and the unknown, true value is as small as possible. Naturally, the larger the sample, the more likely it is that this will be the case. The gains from increasing the sample size, however, do not necessarily increase in proportion to an increase in that sample size. After a certain point the gain in the precision of the estimate, through increasing the sample size, is outweighed by the extra cost of collecting the additional data.

Whereas it is possible to design an optimum sample size such that, either variance in the estimated values is minimised for a given cost, or cost is minimised for a given variance in the estimated values, the survey under discussion did not approach the problem in this way. To have obtained an "acceptable" error in the estimates (s.g. a 5% error in the estimates with a 95% level of confidence), would have required a sample size of approximately 1,500 households. The project worked, however, under the constraints of manpower, time and a desired size of questionnaire which, together, as explained above, limited the sample size to 1,000 households.

The problem, therefore, was to design a sample of the highest precision which would yield estimates which were completely representative of the population. This meant drawing the sample from all areas in Sukabumi.

There are two major considerations in designing a sample. First, it is almost always of value to stratify a population before drawing the sample. The process of stratification means separating the members of a population into strata such that, in each stratum, the members of the population are as homogeneous as possible. It is possible to stratify a large number of times. For example, the Sukabumi population could be stratified into urban and rural strata. Taking the urban stratum, it is possible to stratify this further into two sub-strata — "Metropolitan Sukabumi" and "other urban areas". These two separate sub-strata can be further stratified by (for example) income, or social groups; and so on. In this way, it is possible to create strata where the elements in a particular stratum are, more or less, homogeneous, such that it is possible to obtain a completely representative sample of data by selecting the sample from all of the separate strata. Repeated stratification of a population, however, will not necessarily lead to substantial improvements in the precision of the estimates: the

gains to be had from further stratification have to be offset against the increased costs in varrying out that stratification.

In the Sukabumi survey, stratification was limited to stratifying the population into urban and rural strata, creating two sub-strata in the urban area — "Metropolitan Sukabumi", and "other urban areas" — and two sub-strata in the rural area — the capital villages of the 21 sub-regions in the Sukabumi region, and the remaining 124 villages in the rural area. There was (implicit) further stratification within the rural stratum, however. By listing the capital villages and the remaining villages, it was possible to select systematically the 24 villages which were to fall into the sample. But the order in which the capital villages and the remaining villages were listed would clearly affect which villages fell into the sample. As such, the capital villages were listed in terms of their distance from Metropolitan Sukabumi, with the capital of the Sukabumi sub-region clearly being closest to Metropolitan Sukabumi, whilst Ciricap — another sub-regional capital — was assumed to be the furthest away. Similarly, the remaining villages, within each of the 21 regions, were listed in order of their proximity to Metropolitan Sukabumi.

By systematically selecting from the list of capital villages and remaining villages, at a predetermined sampling interval, it was possible to guarantee that the sample would geographically cover the whole Sukabumi region (see Appendix 2). The advantage of this is clear: in addition to analysing whether the demand for social services was related to the distance from Metropolitan Sukabumi, it guaranteed that the sample would have complete, rather than partial, geographical coverage. If the capital villages and remaining villages had been listed haphazardly, and a simple random sample of 24 villages taken, it is conceivable that al (or most of) the villages included in the sample would have been concentrated in one particular area of Sukabumi e.g. around Metropolitan Sukabumi. This would have been undesirable, as the sample would not have included data from a substantial part of the remainder of the region.

Therefore, once the 24 villages had been selected, the first stage of the sample was completed in the rural area. The second stage of the sample was to select the actual households to be interviewed. Both the design and the drawing of the sample were completed in the office. But to understand how, requires a discussion of the second major consideration in de-

signing a sample.

It should be clear that undertaking a survey, on the scale discussed here, is a time consuming, expensive affair. One of the major sources of the expense is the interviewer costs, (which were zero in the Sukabumi study, as the interviewing was carried out by the IDS participants). A major task of the survey director is to ensure that the paid interviewer is doing what he/she is paid to do: that is, to interview rather than to spend a great deal of time travelling to visit households. In other words, it is much more sensible if the sampled households, for which an interviewer has responsibility, are close together, rather than far apart. In this way, interviewer travelling time (for the initial visit or return visits) can be, to some extent, minimised, whilst the time spent interviewing can be, to some extent, maximised. Therefore, it is sensible to try and formulate clusters of households, where the sample of households is selected on the basis of clusters, such that *every* household in a selected cluster is interviewed.

Whereas stratification requires that the elements of the population, in an individual stratum, should be as homogeneous as possible, drawing a sample on the basis of clusters means that each cluster should *ideally* be composed of households which are as heterogeneous as possible. In this way, interviewing every household in a cluster will yield a good cross-section of data. The problem is that, often, families which live in the same immediate area have a tendency to be of the same social class and background, to have the same educational level, to have similar type of work, and to hold similar views. Therefore, when trying to use interviewers efficiently, by reducing travelling time, given a budget restriction, there is a possibility that, whilst the sample is fully representative, the cross-section of views will not be as wide as might be the case if a simple random sample was taken, or a non-clustered design was used.

For this reason, it is desirable that the cluster size is relatively small. If households in the cluster are heterogeneous, only a small number of interviews need to be taken to ensure a good cross-section; and, if the households in the cluster are homogeneous, not too much of the sample will be wasted collecting closely duplicated views. If the cluster size is too small, however, the interviewer will spend too much time travelling between clusters which, as has been indicated, is expensive. It is a vicious circle in which, as with all aspects of designing a survey, it is a matter of experience as to what is the "right"

decision to take. In the Sukabumi survey, the "right" size of cluster used was approximately 10 households.

Therefore, having first stratified the population, the sample was chosen in the urban and rural areas on the basis of clusters although, in the rural area, the 24 villages included in the sample were selected first. This is known as a two stage, stratified sample. It was a complex sample design in which, because of the practical advantages of clustering, the households which fell into the sample were not chosen independently of one another. In statistical sampling theory, however, it is assumed that elements in a sample *should* be drawn independently of one another.

In order to carry out a probability sample, it is necessary to have a frame;; that is, a set of data which lists the target population, and from which the sample can be drawn. The frame might be complete or, where it is incomplete, it would have to be updated to cover the whole target population — the household head in the survey under discussion. If updating is impossible, the inferrences drawn from the sample would have to be restricted to that part of the population which was covered by the frame.

In the Sukabumi survey, the frame used was the 1971 National Census of Indonesia. The IDS project was fortunate to have the complete cooperation of the Indonesian Central Bureau of Statistics, which holds the detailed census data. At the time of the Census, the population was separated administratively into census blocks i.e. geographically related groups of households; there were, on average, 100 households to each census block, although some had more than 100 households, whilst some had less. In Metropolitan Sukabumi, however, the census block contained, on average, 50 households. It was assumed that, since 1971, the population increase and, therefore, the household increase, had been distributed equally over the whole countryside. Therefore, provided this assumption was correct, it was valid to use 5 year old data with which to draw the sample, in which every household in a selected cluster was to be interviewed.

Each census block can be divided into a number of clusters. Where the census block size was approximately 100 households, 10 clusters of approximately 10 households each were arbitrarily constructed. In Metropolitan Sukabumi, with approximately 50 households in each census block, 5 clusters

of approximately 10 households were arbitrarily constructed. An example of this is found in appendix 4. It should be made clear that where a census block had (say) 76 households (in the non-Metropolitan area), 10 clusters were still constructed in which, taking geographical considerations into account, the clusters were of an unequal size (of approximately 7 households each). The same was true for a Metropolitan census block which happened to have say (65) households; the clusters were of unequal size (of approximately 6 households each).

Each village is composed of a number of census blocks and, as such, is composed of a number of clusters. The Central Bureau of Statistics has a record of the number of census blocks for each village. For example, in the village of Tipar, in the Metropolitan Sukabumi sub-stratum, there are 69 census blocks, numbered 1 to 69. Appendix 4 is a Central Bureau of Statistics map of census block 52 in the village of Tipar, which was included in the sample. (The map — as with all Indonesian census block maps — is not drawn to scale; therefore, it is impossible to tell how much travelling the interviewer will have to do between households.) It will be seen that 5 (numbered) clusters have been arbitrarily constructed, although geographic locations have been taken into consideration.

In order to draw a sample, it is necessary to determine what is known as a measure of size (MOS), where size, in the Sukabumi survey, referred to the cluster of households. The MOS is critical, in that it will give an indication as to where the bulk of the target population is living. If, for example, a rural area is split into two regions, in which region A is 10 times larger than region B, in terms of household numbers, it seems reasonable that, in order to obtain a representative cross-section of households in the rural area, the sample should include approximately 10 times more households from region a than from region B. In principle, it is more sensible to sample where the population is, rather than where the sample is not. Therefore, the sample is (often) selected, though not always, with probability proportional to size. The chance of being included in the sample is, therefore, related to size.

In the Sukabumi project, the MOS that was used was the cluster of households. Where a village contained a large number of census blocks and, in consequence, a large number of clusters, the probability of that village being selected in the sample was greater than where a village had a small number of census blocks.

As discussed above, each interviewer was expected to complete approximately 20 household interviews in both the urban and rural areas. Due to illness, the number of interviewers available was reduced to 24, which meant that the expected sample size was reduced from 1,000 to 960 i.e. 24 x 20 (Rural) + 24 x 20 (Urban). However, as also mentioned above, the expected population increase since 1971 — and the consequent household increase — was expected to overcompensate for any non-response, and also to bring the sample size up to the desired value of 1,000 households.

For the purposes of comparing demands, the sample was allocated equally between the urban and rural areas i.e. an expected total of 480 households in both areas. Choosing cluster sizes of approvimately 10 households provided a convenient way of selecting the sample, in both the rural and urban areas. In the rural area, it was planned that the 12 teams of 2 interviewers would visit 2 villages each. As each interviewer was expected to collect 20 interviews, this was equivalent to interviewing 2 clusters. Therefore, each interviewer was expected to interview one cluster in each village, neccessitating that the rural sample should include 24 villages, 48 census blocks, and 48 clusters. To have an equally allocated sample, it was necessary to select 48 census blocks and 48 clusters from the urban stratum as a whole.

In the urban stratum, the sample of census blocks and clusters was drawn with probability proportional to size. In the rural stratum, however, in order to maximise coverage over the rural area, there were two stages involved in drawing the sample. First, the sample was allocated proportionally between the sub-strata — the capital villages, and the remaining villages — on the basis of the relative number of clusters in each sub-stratum. This led, through probability proportional to size, to the selection of 4 capital villages and 20 remaining villages in the sample. The second stage was to select the 2 census blocks and 2 clusters for each of the selected capital villages and remaining villages.

In order to gain a greater coverage of the rural area, whilst obtaining a sample of 1,000 households, more than 24 villages — say 48 villages — could have been selected. Each interviewer would, again, have interviewed one cluster in each village, provided the expected cluster size was reduced from 10 to 5 households. Whilst this would have yielded the same expected sample size in the rural stratum, (48 villages x 2

interviewers x 5 households per cluster = 480 households), it would have meant that the interviewer was spending too much time travelling. It was unlikely that, in the time available, all the interviews would have been collected, given the logistics of travelling in the Sukabumi region.

As it was, selecting 48 census blocks from the urban area gave excellent coverage of that stratum. In the rural area, the coverage was again satisfactory as the sample covered approximately 1 in 5 capital villages, and approximately 1 in 6 of the remaining villages. The sampled villages are shown in appendix 2.[3] Appendix 3 describes in detail how the sampled villages, the sampled census blocks and the sample clusters (of households) were selected for interview.

**Fieldwork** [4]

The fieldwork for the survey was undertaken in the Sukabumi region from October 11 until October 23, 1976. The 24 IDS participants acted as interviewers.

The following timetable was followed during the two weeks of the fieldwork. The timetable deviates from that assumed for the purposes of estimating the sample size, in that the first five days were reserved for interviewing in the urban area and for travelling to and from Metropolitan Sukabumi, whilst the remaining nine days were reserved for the rural area. Behind this timetable change was the expectation that, whilst interviewing in the urban area would raise fewer problems, because the census blocks were relatively close together, and the population spoke Bahasa Indonesia,[5] the lack of interviewing experience by the participants would require longer in the urban area than had been estimated.

The revised timetable for completing urban interviewing was still too optimistic. The first day was occupied with travelling to Sukabumi, and the subsequent activity of setting up survey headquarters. The morning of the second day was used to meet the local administrators. 3½ days remained, there-

[3] The majority of sampled villages are seen, in Appendix 2, to lie in the north eastern area of Sukabumi region. This is because the majority of the Sukabumi population is to be found in this quadrant, having been attracted towards the urban area by the main road connecting Jakarta and Sukabumi.

[4] Written in conjunction with Dr. J. Syatauw, Institute of Social Studies.

[5] The national rather than regional language.

fore, in which each participant was expected to collect approximately 20 urban interviews, i.e. two clusters. During the first two days, the pace of interviewing slowed due to rain and, after 5 days, a number of participants had failed to complete their urban interviews.

Having gained experience in the urban area, it was expected that interviewing in the rural area would be easier. It was assumed that this would permit the completion of interviewing in less than the 9 days available, thus enabling a return to the urban area to complete any interviews outstanding.

Whilst planning the survey, it was decided that interviewing in the rural area would be carried out in teams of two interviewers. It materialised that the geographical conditions did not always allow such a division: consequently, some interviewers worked alone whilst, in other cases, a team of three interviewers was assigned to three relatively close villages.

On arrival in Sukabumi, further information on the selected census blocks and clusters, transportation, regional geographical characteristics, accommodation and support from local census staff, both as guides and interpreters, was given by the Central Bureau of Statistics.

The identification of a selected census block and selected cluster number in the urban area was sufficiently easy that no local census staff was required as a guide, nor as an interpreter, (Bahasa Indonesia being spoken in Metropolitan Sukabumi). Interviewers were not given census block maps for the urban area, but were given information about the location of respective census blocks and clusters. A list of households to be found in the selected clusters was provided by the census staff.

Information on rural transportation and rural geography was provided, in addition to the names of the local census staff who were to accompany, and to assist, the interviewers in the rural area. The local census staff produced census block maps, as well as a list of sampled households in each selected cluster.

As a result of a field administration visit by the project staff, to identify and to eliminate any possible problems, preparations by the Sukabumi authorities and the Central Bureau

of Statistics proved unquestionably essential. Instructions had been given to local census staff that interviewers should be assisted, where possible. In preparation, the local census staff had been summoned to the Sukabumi headquarters of the Central Bureau of Statistics to receive explicit instructions. This facilitated greatly the work of the interviewers. Only rarely were preparations incomplete, e.g. a census block map was unavailable or incorrect, or the list of household members was obsolete because households had left a selected cluster.

Some practical problems were encountered during interviewing. In the urban area, most respondents were at work until approximately 2 p.m. The mornings were used, therefore, for completing an observation questionnaire, interviewing local administration officials, and making provisional appointments for household interviews. It was difficult to achieve the initial average target of 5 interviews each day,[6] necessitating a return after completion in the rural area.

It was optimistic to expect each interviewer to interview 5 times a day, for 90 minutes at each interview. Experience showed that some interviewers were rushed to achieve the daily target. Other interviewers proceeded without feeling the need to achieve the daily target. Competition developed between some interviewers, to achieve the greatest daily interviewer total. Ultimately, the idea that slow interviewers were bad interviewers gave way to a more realistic evaluation of the problems of interviewing.

Household mobility was probably greater in the urban area. Where listed households had left a selected cluster, some interviewers (correctly) replaced these households with the new households — if they existed — at a given address, or they (incorrectly) took replacement households from outside the sampled cluster.

In the rural areas, the majority of respondents, mostly farmers, were engaged in activities in, or near, their dwellings. This facilitated rapid contact. Moreover, they had often been warned beforehand by the village headman to expect the interviewer.

A local census official was attached to an interview team. If the team decided to interview separately, an alternative guide/

[6] 20 interviews ÷ 4 days = 5 interviews/day. This was adjusted to 20 interviews ÷ 3-1/2 days, i.e. approximately 6 interviews/day.

interpreter had to be found to accompany the single interviewer: this tended to be the village headman, or one of his assistants.

During interviews, the interviewer was sometimes accompanied by a number of village officials, in addition to the local census official. To avoid bias, the interviewer had to ask to be left alone, apart from the interpreter, although the local official was sometimes helpful when information was required that the respondent could not remember, or was unable to provide, e.g. the area of land owned.

Where no adequate census block map (showing a sampled cluster) was available, the interviewer had to decide, with or without the local census staff, how best to find substitute households to compensate for failing to interview the sampled cluster.

The distance between some sampled clusters was a problem. Before entering the field, the physically strong participants were allocated to the more difficult areas. But, geographical ignorance meant that some interviewers had to walk distance of 5-10 kilometres, in difficult terrain, to reach a village. Some interviewers gave up trying to visit, or did not attempt to visit, such clusters. On the other hand, some interviewers walked 10-15 kilometres to reach a selected cluster, whilst two interviewers were assigned (and reached) the distant village of Rambay, some 40 kilometres (i.e. one day's walk) from the capital village of that region.

Accommodation in the rural area was basic. In most cases, the village headman selected the best dwelling in the village to accomodate the interviewers. On the other hand, food was occasionally a problem: the interviewers — some, professional planners — were unaccustomed to the unvarying simple meal of rice and a little dried fish, three (or even two) times a day.

**The Questionnaire**

As discussed above, the Sukabumi survey could never have been conceived as anything more than a pilot study, given the short period of time available. This was glaringly demonstrated in respect of the questionnaire, which proved to be less than satisfactory for the purposes of a full survey. Nevertheless, that is what pilot studies are for: by undertaking pilot studies, it is hoped that all the likely problems to be

faced in a proper survey will be illuminated and eliminated, especially with respect to the questionnaire. It is of extreme importance to test, change, retest and probably again change the questionnaire, before it will usefully provide the data which is required to meet the research objectives.

The actual Indonesian questionnaire which was used was improperly pretested before going into the Sukabumi field, and this inadequate pretesting failed to illuminate the problems inherent in the questionnaire. Some questions were badly formulated and/or irrelevant to actual conditions in Sukabumi. This led to many unfilled answers and/or to many answers for which possible categories of answers had not been provided.

Some — but not all — of the obvious errors and omissions were eliminated after the limited pretesting. It is most probable that, if a full survey was now to be undertaken, the questionnaire would be substantially revised and shortened.

A further weakness in the questionnaire, given the short time scale of the project, was its length. There were 283 variables for which data was collected. It proved impossible, in the time available, to edit, to process and to verify all the data which had been collected and, as such, only a sub-sample of the data was processed.

**Conclusions**

This paper has described the process of — and the problems associated with — designing and implementing a sample survey, in which the purpose of the research was to collect data about which there was no positive prior knowledge. The paper has emphasised, at length, those important areas where theoretical assumptions have had to be modified to suit the economic and physical realities and constraints associated with the project.

The paper concludes by discussing certain interesting practical issues which have arisen from the project. It discusses additionally whether any positive lessons can be drawn from the method that was used to design and implement the project.

First, how reliable is the data that was collected? The sample design was based on the theory of probability, and resulted in the use of a two stage, stratified sample. There was

nothing wrong theoretically with the sample design although, with hindsight, the population could have been stratified further (by regional and/or economic characteristics). The cluster size — approximately 10 households per cluster — might be considered too large to be efficient. The Indonesian Central Bureau of Statistics generally constrains itself to using this size of cluster and, therefore, given the resources available to the project, it was probably generous to use a cluster size of 10.

The important question, however, concerns the reliability of the results, which are reported elsewhere. It has been discussed that, sometimes, interviewers did not interview the sampled cluster or all the households within a sampled cluster, or sometimes census block maps were unavailable.[7] Taking this into consideration, does this invalidate the results obtained? Whilst the probability sample was properly designed, it was improperly implemented in certain cases. A major lesson to be learned from projects of this type is that a large degree of tolerance must be allowed for when trying to collect data under the conditions described. The probability sample broke down at the last stage where, in a small number of cases, sampled households were not contacted, and where (rarely) visits were made to the wrong villages. These non-sampling errors, along with respondent and interviewer mistakes during interview, and coding and editing mistakes, certainly outweighed the sampling errors that arose from the sample design.[8]

The reliability of an estimate (as measured by the size of its total error) has to be balanced against the response rate. In the current survey, the response rate, i.e. the percentage of household heads sampled who answered the questionnaire, was in excess of 90%; this was remarkably high. (To some extent, especially in the rural areas, the sampled households were encouraged by the village headman to participate in the survey. This "pressure", or bias, could be seen as contributing to non-sampling error, although there is no reason to suppose that the answers given were anything but truthful.)

It is concluded, therefore, that such a high response rate

[7] Where the Central Bureau of Statistics undertakes a survey, and maps are missing, a complete census of the undocumented area is taken, maps are redrawn, and the procedure described in appendix 3 is carried out.

[8] It is known that, provided a sample is properly designed, the non-sampling errors will generally be greater than the sampling errors, and that the major component of the error attached to an estimate is made up of non-sampling errors.

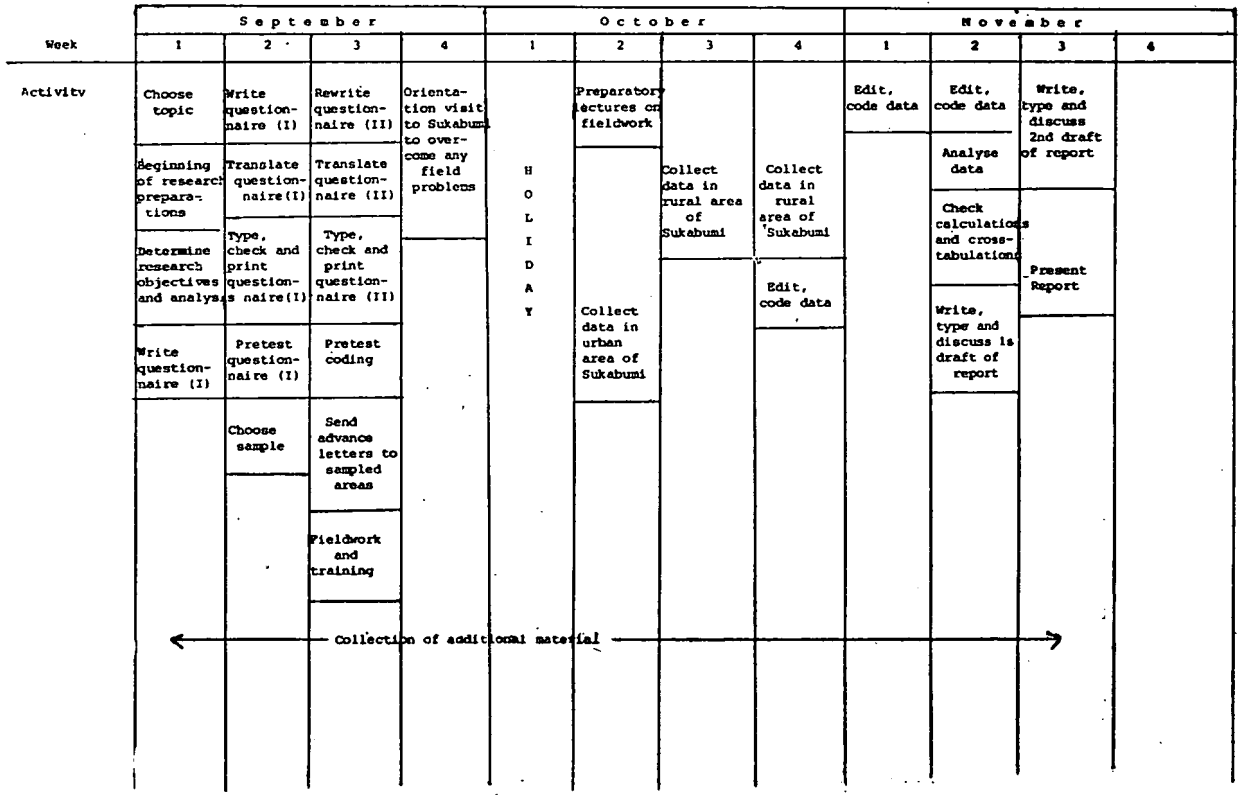
is most likely to outweigh any non-sampling errors incurred in failing to contact sampled households, and that the results are a representative and reliable estimate of the population as a whole, provided the administrative definitions of urban and rural populations — based on size — are considered acceptable. In other words, there would be no hesitation in applying tests of significance to, or constructing confidence intervals for, the data which has been collected.

Secondly, can any positive lessons be drawn from the survey? It is concluded that, as a teaching exercise, this type of project was extremely useful. The only helpful way to instruct how to use a technique is, first, to relate it to a practical problem and, secondly, to solve that problem with those who are being instructed.

Undertaking a large scale survey, based on probability sampling is not difficult; in fact, despite numerous problems, it is relatively simple. Much of the work can be carried out in the office. The problems arise in the field, where the major concern is the lack of adequately trainel interviewers. A second major problem is time: as appendices 1A and 1B show, there are a minimum number of activities which have to take place — in particular, the writing, testing, and rewriting of the questionnaire — before a survey will generate the results desired.

As with most surveys, the headache created by the Sukabumi pilot study concerned the questionnaire. First, the questionnaire was too large, containing too many variables. There is no point in collecting data unless it can all be analyse. With too many variables to code, edit and process, inadequate time was allowed (during the project period) for a proper analysis of the data. Secondly, it became evident during the preliminary analysis that the questionnaire was not providing the information that was required to meet the research objectives, because the questionnaire had not been properly pretested. It was also clear from the preliminary analysis that, with adequate pretesting, the final questionnaire could have deepened the analysis. Although, at the beginning of the project, the required analysis was determined, the preliminary analysis of the data — in effect, the analysis of the pretest of the second draft of the questionnaire — indicated that a more interesting analysis could have been undertaken than was at first thought possible.

## Appendix 1A
## Expected Timetable of Research Project

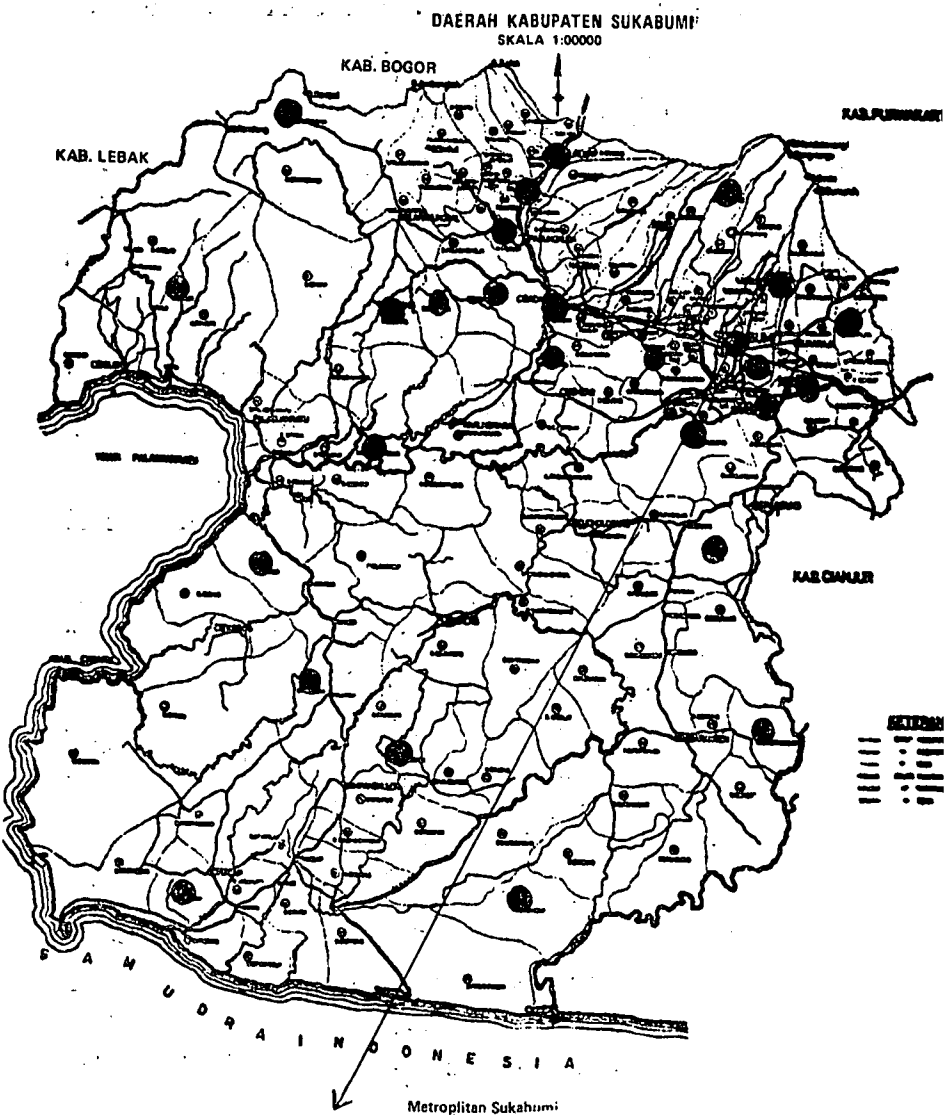


# Appendix 1B
## Actual Timetable of Research Project

| | September | | | | October | | | | November | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Week | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Activity | Choose topic | Write questionnaire (I) | Type, check and print questionnaire (I) | Orientation visit to Sukabumi to overcome any field problems | HOLIDAY | Preparatory lectures on field work | Collect data on rural area of Sukabumi | Collect data on rural area of Sukabumi | Edit, code data | Edit, code data | Analyse data | |
| | Beginning of research preparation | Translate questionnaire (I) | Choose sample | | | Rewrite, translate, type and check questionnaire (II) | | Code questionnaire (II) | | | Write, type report | |
| | Determine research objectives and analysis | | | Pretest questionnaire (I) | | Print questionnaire (II) | | | | | | |
| | Write questionnaire (I) | | | Send advance letters to sampled areas | | Collect data on urban area of Sukabumi | | | | | | |

Appendix 2

Geographical Distribution of Sampled Villages in the Sukabumi Region




Sampled village is represented as ●

Finally, it is concluded that the pilot survey described would be of substantial benefit to the Sukabumi authorities (or any other regional or national authority) if they wished to undertake a full survey on the demand for social services. Each stage of the survey has been thoroughly pretested. All the problems have been identified, and it is clear that they can be overcome.

## APPENDIX 3 Technical Description of Sample Selection

This appendix describes the mechanics of drawing the sample which was used in the Sukabumi Survey.

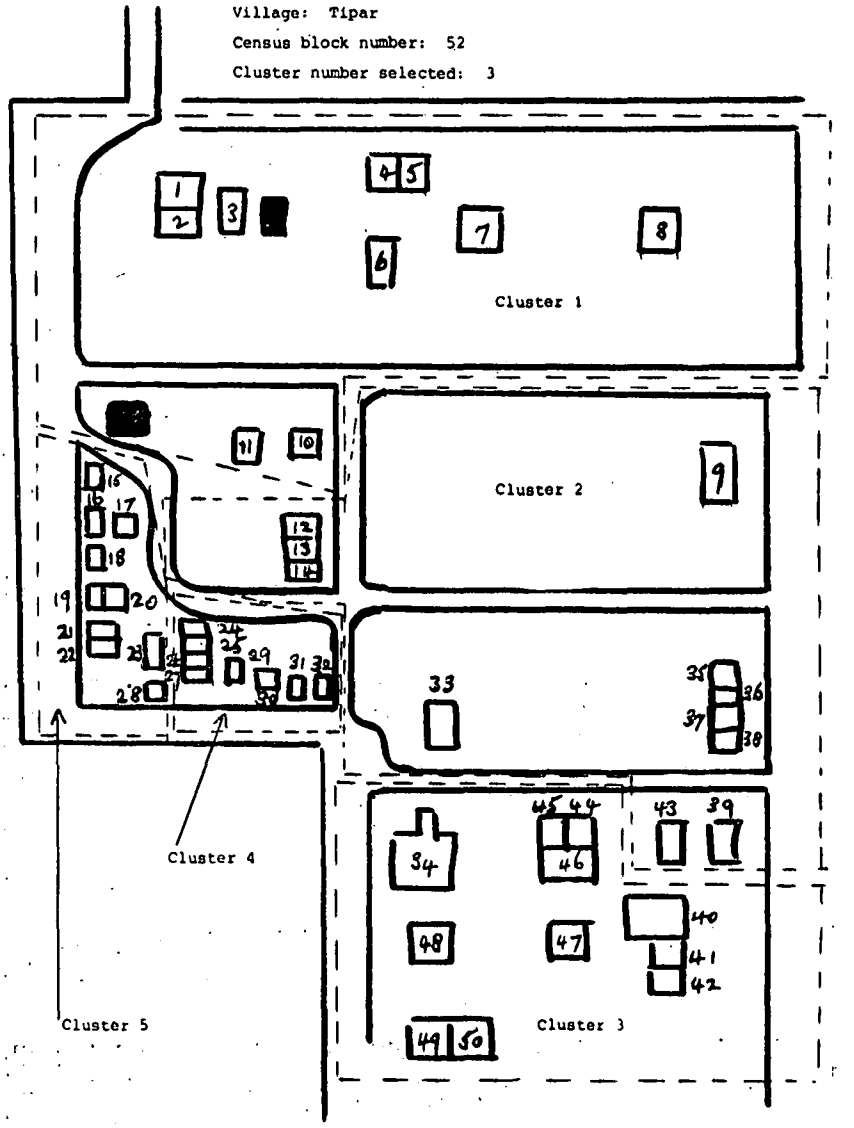
Having stratified the population into the urban and rural areas, it was decided to allocate the sample equally over the two strata.

Consider the two strata separately, dealing first with the rural stratum. This stratum was stratified further, creating two sub-strata — capital villages and the remaining villages. It was felt likely that the capital villages of the 21 subregions of Sukabumi were likely to be significantly different from the remaining rural villages, and that this would justify creating two sub-strata.

A decision had to be taken as to the number of capital villages and remaining villages to include in the sample. As explained in the paper, it was decided to include 24 villages (capital villages and the remaining villages) in the sample, in order to obtain the maximum geographical coverage of the Sukabumi region. By allocating the sample proportionally between the two sub-strata, on the basis of the total number of census blocks, 4 capital villages and 20 remaining villages were included in the sample. The total number of census blocks in the capital villages sub-stratum was 441, whereas, in the second sub-stratum, it was 2238. Therefore, for the first sub-stratum, the number of capital villages selected was $\frac{441}{2679} \times 24 = 3.95$. For the second sub-stratum, the number of remaining villages selected was $\frac{2238}{2679} \times 24 = 20.05$. As integers are required — partial village interviews cannot take place — the numbers were rounded to 4 and 20 respectively.

Appendix 4: Census Block Map

Stratum: Urban
Substratum: Metropolitan Sukabumi
Village: Tipar
Census block number: 52
Cluster number selected: 3



The next decision was to determine which villages would fall into the sample. Taking the capital villages sub-stratum, where the capital villages were listed in order of their distance away from Sukabumi, (the furthest coming last), the villages were selected with probability proportional to size. As can be seen from table 3, the measure of size was the cluster of households and, for each capital village, this was cumulated, giving a cumulated cluster total of 4410. In order to select the 4 villages, it was necessary to select a random number (from a table of random numbers), between 0 and the cumulative total of 4410, and add to it an apppropriate sampling interval which would, when applied systematically, ensure the choice of 4 villages. For example, the sampling interval was $\frac{4410}{4} = 1102.5$. The random under chosen, between 0 and 4410, was 2560. It can be seen from table 3 that Cikidang was included in the sample, because a village was included when the cumulated total appropriate for that village was greater than the random number selected. A village was not included if the cumulative total was smaller than the random number. Applying the sampling interval gave the second village of Cihaur; (2560 + 1102.5 = 3662.5). Applying again the sampling interval gave the third village of Baros; (3662.5 + 1102.5 = 4765.0, but 4756.0 — 4410.0 = 355.0). Finally, applying sampling interval again, Parungkuda was selected in the sample; (355.0 + 1102.5 = 1457.5).

Exactly the same procedure was used for the second sub-stratum of the remaining villages; although all the 124 remaining villages are not listed here, table 7 shows the 20 remaining villages selected. A random number between 0 and 22380 — the cumulative cluster total — was selected to give the first village. Following the procedure outlined above, the sampling interval of $\frac{22380}{20} = 119.0$ was systematically applied to the list of villages until the 19 remaining villages had been selected. Again, the villages were listed in order of their distance away from Sukabumi.

Having selected the capital villages and the remaining villages, it was necessary to determine which census block and cluster number should be included in the sample, for, as was stated in the paper, each household identified in a selected cluster was included in the sample. It was decided that, for

each village, two census blocks (and, thus, two clusters) would be visited. The selected census blocks and clusters for the two rural sub-strata are given in tables 6 and 7. Taking Baros as an example, for the capital village sub-stratum, it was known that there were 230 clusters in this particular village. Therefore, choosing a random number between 0 and 230 (077), and applying a sampling interval of $\frac{230}{2} = 115$, gave the following information: cluster number 7 and census block 20, cluster number 2 were selected for interview.

Taking one further example, in the second sub-stratum of the remaining villages, the procedure can be repeated for Cimanggu in the sub-region of Jampangkulon:

Census Blocks 14
Clusters 140

$$\text{S.I.} = \frac{140}{2} = 70$$

$RN_1 = 81$
$RN_{21} = 81 + 70 = 151$
$RN_{22} = 151 - 140 = 11$

Selected census blocks/cluster numbers = 2/1 and 9/1.

This procedure was applied to the entire stratum in order to determine (from the office) which census blocks and clusters, in a particular village, would be included in the sample.

Turning now to the urban stratum, the sample was selected in a similar manner. It was not necessary to select villages, however, as these had been determined previously on the basis of administrative definitions (i.e. population size) as to what was, and what was not, an urban area. Thus, the 6 villages that comprised Metropolitan Sukabumi, and the villages of Cibidak and Cicurug were incorporated automatically into the sample.

The census block and cluster numbers had to be selected, however, bearing in mind that, in order to obtain an equal allocation of the sample over the two strata, 48 census blocks had to be selected. Whereas it was possible to cumulate all

the cluster totals over the whole urban stratum to obtain the 48 census blocks, by selecting a random number and applying a sampling interval, this was not done. The Metropolitan Sukabumi statisticians have altered the 1971 census block definitions to suit their own purposes; as such, it was not possible to select systematically when sub-stratum a — Metropolitan Sukabumi — and sub-stratum b — other urban areas — were based on different census block definitions. Thus, the sample was selected differently for the two sub-strata.

What number of census blocks had to be selected from the two sub-strata which, together, would have given a sample of 48 census blocks? Taking the 1971 distribution of census blocks, it was possible to allocate the sample proportionally between the two sub-strata. There were 68 census blocks of size 100 households in Cibidak and Cicurug, and 388 census blocks of size 50 households in Metropolitan Sukabumi. Converting this latter figure into census blocks equivalent in size to those in Cibidak and Cicurug, Metropolitan Sukabumi had 169 census blocks. The distribution of the 48 census blocks over the two sub-strata showed, therefore, that 71% of the census blocks were in Metropolitan Sukabumi, whilst 29% were in Cibidak and Cicurug. This meant that 34 census blocks should have been selected from Metropolitan Sukabumi (.71 x 48 = 34.08 ÷ 34), and 14 census blocks should have been selected from the remaining urban areas (.29 x 48 = 13.9 ÷ 14). To correct for uneven distribution in the in the urban population increase since 1971, it was decided to increase to 36 the number of census blocks selected from Metropolitan Sukabumi and to decrease to 12 the number of census blocks selected from the remaining areas.

The census blocks and cluster numbers for the urban stratum are given in tables 4 and 5. For the Metropolitan Sukabumi sub-stratum, the 36 census blocks were chosen with probability proportional to size in which, choosing a random number (RN = 0299) between 0 and 1690, and adding systematically a sampling interval of $\frac{1690}{36} = 46.94$, yielded the desired sample.

The same procedure was used for the second urban sub-stratum. The selection was made with probability proportional to size, in which a random number (RN = 181) between 0

and 680 was chosen, and a sampling interval of $\frac{680}{12} = 56.67$ was added systematically.

Again it was possible to determine the census block and cluster numbers from these random numbers, as can be seen below.

Sampling interval (S.I.) = .115
Random Number$_1$ ($RN_1$) = .077
$RN_2$ = 077 + 115 = 192.

These two random numbers indicated something about the selected census blocks and clusters for Baros. Cumulating all the clusters for Baros, the following table was produced.

| *Census Block* | *Cumulated Clusters* | *Random Numbers* |
|---|---|---|
| 1 | 10 | |
| 2 | 20 | |
| 3 | 30 | |
| . | . | |
| . | . | |
| . | . | |
| 7 | 70 | |
| 8 | 80 | 77 |
| 9 | 90 | |
| . | . | |
| . | . | |
| . | . | |
| 19 | 190 | |
| 20 | 200 | 192 |
| . | . | |
| . | . | |
| . | . | |
| 23 | 230 | |

It was known that, in selecting the sample, the random number should always be less, not more, than the relevant cumulated total. Thus, for Baros, census blocks 8 and 20 were included in the sample. By taking the last digit of the random number, it was possible to derive the cluster number in the particular census block selected; (zero was interpreted as 10). Therefore, for Baros, all households which fell into census block 8,

| *Metropolitan Sukabumi* | | | *Other Urban Areas* | | |
|---|---|---|---|---|---|
| *Census Block* | *Cumulated Clusters* | *Random Number* | *Census Block* | *Cumulated Clusters* | *Random Number* |
| 1 | 5 | | 1 | 10 | |
| 2 | 10 | | 2 | 20 | 11 |
| 3 | 15 | | 3 | 30 | |
| 4 | 20 | 17 | 4 | 40 | |
| . | . | | . | . | |
| . | . | | . | . | |
| . | . | | . | . | |
| 8 | 40 | | . | . | |
| 9 | 45 | 41 | . | . | |
| . | . | | . | . | |
| . | . | | . | . | |
| . | . | | . | . | |
| 71 | 355 | | 43 | 430 | |

As before, the cumulated cluster total which exceeded the random number determined which census block was selected. Remembering that the census blocks had only 5 clusters in the Metropolitan Sukabumi sub-stratum, the last digit of the random number in sub-stratum a had a slightly different meaning to those used in sub-stratum b and in the rural stratum: 5 was subtracted from numbers 6 and above, whereas numbers below 6 were interpreted as they stood, although zero was always interpreted as 5.

Therefore, the first random number in sub-stratum a — 17 — meant that census block 4, cluster number 2 was to be interviewed in Gunung Puyuh whilst, in sub-tratum b, the first random number — 11 — meant that census block 2, cluster number 1 was to be interviewed in Cibadak; and so on.

The use of the random number and systematic sampling indicated which census blocks and clusters were to be contained in the sample. The 36 census blocks, however, had to be distributed over 6 villages in sub-stratum a, and 2 villages in sub-stratum b. How was this done? Examining table 1, and taking as an example random number 346 in sub-stratum a, this indicated that a census block in Selabatu was included in the sample. By subtracting the cumulated cluster total of 305 for Gunung Puyuhh from 346, the residual of 41 indicated that census block 9, cluster number 1 in Selabatu was selected

for interview. Similarly, taking random number 1191 in sub-stratum a meant subtracting 860 from 1191: the residual of 331 indicated that census block 67, cluster number 1 was selected in Nyomplong.

**Table 1.**

Stratum 1. Urban
Sub-stratum a. Metropolitan Sukabumi

| Village | Census Block | Number of Clusters | Cumulative Clusters | Random Number |
|---|---|---|---|---|
| 1. Gunung Puyuh | 61 | 305 | 305 | 299, 17, 64, 111, 158, 205, 252 |
| 2. Selabatu | 56 | 280 | 585 | 346, 393, 440, 487, 534, 581 |
| 3. Kebonjati | 55 | 275 | 860 | 628, 675, 722, 768, 815 |
| 4. Nyomplong | 71 | 355 | 1215 | 862, 909, 956, 1003, 1050, 1097, 1144, 1191 |
| 5. Tipar | 69 | 345 | 1560 | 1238, 1285, 1332, 1379, 1426, 1473, 1519 |
| 6. Nanggeleng | 26 | 130 | 1690 | 1566, 1613, 1660 |

Note: The Metropolitan Sukabumi census blocks no longer correspond with the 1971 census blocks. They have been changed for the convenience of the local statisticians.

**Table 2.**

Stratum 1. Urban
Sub-stratum b. Other urban.

| Village | Census Block | Number of Clusters | Cumulative Clusters | Random Number |
|---|---|---|---|---|
| 1. Cibadak | 43 | 430 | 430 | 181, 238, 294, 351, 408, 11, 68, 124 |
| 2. Cicurug | 25 | 250 | 680 | 464, 521, 578, 634 |

**Table 3.**

Stratum 2. Rural
Sub-stratum a. Capital villages

| Region/Capital Village | Census Blocks | Number of Clusters | Cumulative Clusters | Random Number |
|---|---|---|---|---|
| 1. Sukabumi/Sukabumi | — | — | — | |
| 2. Sukaraja/Pasirhalong | 24 | 240 | 240 | |
| 3. Baros/Baros | 23 | 230 | 470 | 355 |
| 4. Cisaat/Cisaat | 27 | 270 | 740 | |
| 5. Cikembar/Cikembar | 25 | 250 | 990 | |
| 6. Cibadak/Cibadak | — | — | — | |
| 7. Najiak/Najiak | 24 | 240 | 1230 | |
| 8. Nyalindung/Nyalindung | 12 | 120 | 1350 | |
| 9. Cicurug/Cicurug | — | — | — | |
| 10. Parungkuda/Parungkuda | 19 | 190 | 1540 | 1458 |
| 11. Warungkiara/Warungkiara | 43 | 430 | 1970 | |
| 12. Kalapa Nuggel/Kalapa Nuggel | 12 | 120 | 2090 | |
| 13. Jambag Tenbah/Jambag Tenbah | 26 | 260 | 2350 | |
| 14. Cikidang/Cikidang | 30 | 300 | 2650 | 2560 |
| 15. Lengkong/Lengkong | 11 | 110 | 2760 | |
| 16. Palabuhanratu/Palabuhanratu | 42 | 420 | 3180 | |
| 17. Cisolok/Cisolok | 24 | 240 | 3420 | |
| 18. Sagaranten/Sagaranten | 23 | 230 | 3650 | |
| 19. Ciemas/Cihaur | 17 | 170 | 3820 | 3663 |
| 20. Jampangkulon/Jampangkulon | 36 | 360 | 4180 | |
| 21. Ciracap/Jagamukti | 23 | 230 | 4410 | |

**Table 4.**

Stratum 1. Urban
Sub-stratum a. Metropolitan Sukabumi

| *Village* | *Census Block/ Cluster No. Selected* | *Number of Census Blocks Selected* |
|---|---|---|
| 1. Gunong Puyuh | 4/2, 13/4, 23/1, 32/3, 41/5, 51/2 and 60/4 | 7 |
| 2. Selabatu | 9/1, 18/3, 27/5,37/2, 46/4 and 56/1 | 6 |
| 3. Kebonjati | 9/3, 18/5, 28/2, 37/3 and 46/5 | 5 |
| 4. Nyomplong | 1/2, 10/4, 20/1, 29/3, 38/5, 48/2, 57/4, 67/1 | 8 |
| 5. Tipar | 5/3, 14/5, 24/2, 33/4, 43/1, 52/3 and 61/4 | 7 |
| 6. Nanggeleng | 1/1, 11/3 and 20/5 | 3 |
| | TOTAL | 36 |

**Table 5.**

Stratum 1. Urban
Sub-stratum b. Other Urban Areas

| *Village* | *Census Block/Cluster Number Selected* | *Number of Census Block Selected* |
|---|---|---|
| 1. Cibadak | 2/1, 7/8 13/4, 19/1, 24/8, 30/4, 36/1, and 41/8 | 8 |
| 2. Cicurug | 4/4, 10/1, 15/8 and 21/4 | 4 |
| | TOTAL | 12 |

Table 6

Stratum 2. Rural
Sub-stratum a. Capital Villages

| *Region/Village* | *Census Block Cluster Number Selected* | *Number of Census Block Selected* |
|---|---|---|
| 1. Baros/Baros | 8/7 and20/2 | 2 |
| 2. Parungkuda/Parungkuda | 6/5 and 15/10 | 2 |
| 3. Cikidang/Cikidang | 1/5 and 16/5 | 2 |
| 4. Ciemas/Cihaur | 3/5 and 8/9 | 2 |
| | TOTAL | 8 |

Stratum 2. Rural
Sub-stratum b. Remaining Villages

| *Village/Region* | *Census Block/ Cluster Number Selected* | *Number of Census Blocks Selected* |
|---|---|---|
| 1. Cisarua/Sukabumi | 12/8 and 28/8 | 2 |
| 2. Sudadjajagirang/* | 5/8 and 11/8 | 2 |
| 3. Semplak/Sukaraja | 3/9 and 10/9 | 2 |
| 4. Bodjong Sawah/** | 1/3 and 5/3 | 2 |
| 5. Cibatu/Cisaat | 2/4 and 12/4 | 2 |
| 6. Citaming/Cisaat | 4/7 and 17/7 | 2 |
| 7. Cimanggu/Cikembar | 4/10 and 13/5 | 2 |
| 8. Pamitajan/Cibadak | 10/6 and 25/6 | 2 |
| 9. Bojongkalong/Warungkiara | 17/2 and 34/2 | 2 |
| 10. Cisitu/Nyalindung | 7/10 and 18/5 | 2 |
| 11. Pondoklandak/Warungkiara | 2/9 and 8/9 | 2 |
| 12. Bantangading/*** | 12/8 and 24/8 | 2 |
| 13. Kebandungan/Kampung Nunggak | 7/4 and 17/4 | 2 |
| 14. Cicareuh/Cikidang | 3/7 and 10/2 | 2 |
| 15. Waluran/Lengkong | 6/8 and 13/8 | 2 |
| 16. Cimaja/Cisolok | 8/10 and 16/10 | 2 |
| 17. Curugkembar/Sagaranten | 8/8 and 22/8 | 2 |
| 18. Rambay/Sagaranten | 2/3 and 6/3 | 2 |
| | 2/1 and 9/1 | 2 |
| 20. Ciracap/Ciracap | 1/1 and 12/6 | 2 |
| | TOTAL | 40 |

/* Sukabumi
/** Baros
/*** Warungkiara