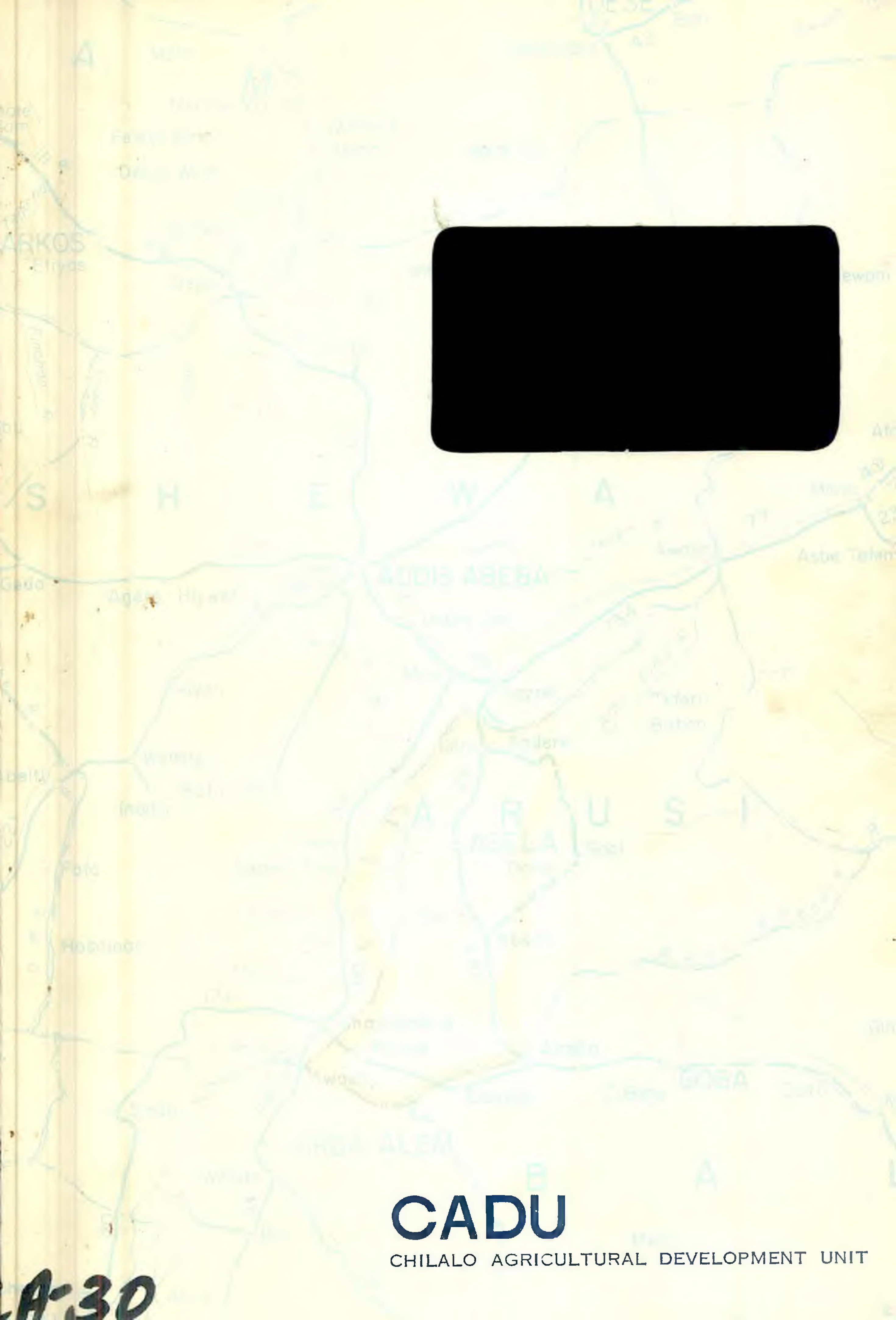


[redacted]

CADU

CHILALO AGRICULTURAL DEVELOPMENT UNIT

CADU EVALUATION STUDIES:
CROP SAMPLING 1968

by

Goran Nyberg
Asella, May 1969

CADU Publication No. 30

P R E F A C E

This report contains a rather detailed description of the methods used in CADU's crop sampling 1968. Because of the frequent changes of expatriate staff, typical of economic development work, a thorough written description of the methodology ought to be available. This is the only way to ensure not only repetition of the study with methods yielding commensurable results, but also a successive improvement in the precision of the methods and thereby more valuable results. By necessity, this makes the report rather technical. This is especially true for the statistical parts of it. To understand them, a background in elementary statistics will be necessary. Readers not familiar with that subject are referred to elementary textbooks (e.g. Cochran, W.G, Sampling Techniques, 2nd ed. New York 1963, and Freund, J.E. & Williams, F.J., Elementary Business Statistics, Englewood Cliffs, N.J. 1964). Readers mainly interested in the results of the study can find those in Chapter 4.

The study was made as a teamwork between several branches of CADU. The Planning and Evaluation Section has carried the main responsibility, but has drawn heavily on the experience and help of the Extension and Education and the Crop Production Departments, especially for the training and administration of field workders and for the laboratory part of the processing of results.

With the rather complex sampling procedure used, the choice of estimators for mean and variance turned out to be a complicated problem. For advice and help in this matter, the author is strongly indebted to Mr. Pushyendu S. Choudhury, United Nations Statistical Adviser to the Central Statistical Office, Addis Ababa.

Asella, May 1969

Goran Nyberg

Head, Planning and Evaluation Section

1. THE PROBLEM

1.1 Crop Sampling in Earlier Years

Each year from the project preparation period onwards, CADU has made attempts at estimating the average yield of different crops in the project area. The first two year's measurements, 1966 and 1967, were made as sample surveys of six and nine different crops respectively. The results have been published in two CADU Publications: "Crop Sampling in the Chilalo Awraja, Arussi Province 1966" and "Crop Sampling in the Chilalo Awraja, Arussi Province, 1967" respectively.

The methods employed in these two studies in both cases involved the use of judgment sampling in at least one step of the sampling procedure. It is, therefore, not possible to calculate the precision of the methods used nor to know anything about the representativity of the results. The bias introduced is likely to be one of over-representation of farmers with above average yields, but the amount of bias cannot be calculated.

1.2 Purpose of the Study

Against this background it was decided to make an attempt, for the first time since the start of the project, at getting an estimate of crop yields that would be statistically representative, using a method the precision of which could be calculated. The original objective was to get such estimates for each of the six extension areas, in which the project area was divided during 1968. Because of the limited resources available for the study, it was confined to four crops, viz. those for which yield estimates were most urgently needed: wheat, barley, beans and flax.

The crop sampling is part of the evaluation of CADU at project level, i.e. the evaluation of the aggregate effects of all the project activities. Clearly, the average yield level of different crops in the project area will be affected by a number of CADU project activities, such as training of model farmers, demonstration for other farmers, selling of farm supplies (especially seeds and fertilizer), marketing activities (to induce the farmers to increase their yields), etc. Other aspects of the overall economic development of the project area, such as changes in household incomes and consumption patterns, will be evaluated as well, as part of the evaluation at project level. The development of crop yields will then be compared to these and possibly other changes to give an overall picture of the aggregate effect of CADU.

Contrary to most other CADU evaluations, a control area is used in the crop sampling studies. The argument is that a number of factors outside CADU control influence the yield figures (above all the weather), which calls for an area with which each year's estimates can be compared. Ideally, such an area should have the same ecological characteristics as the project area. A second important requirement is the economic one: the control area should be within fairly easy reach from the project area in order not to make the study bill sky-rocket. Thirdly, the weather conditions should be as similar to those of the project area as possible. The area chosen actually consisted of two different parts, both adjacent to the project area. One, Sire wereda, lies to the north-east of the project area, the other one, the northern part of Lemu and Bilbillo wereda, to the south. Ecologically, they come reasonably close to the northern and southern parts of the project area respectively. From a weather point of view, the proximity should give conditions as similar to those of the project area as one can reasonably expect with Ethiopia's topography.

It was soon found, however, that the available resources would not allow for yield estimates to be made for each of six different subdivisions of the project area and for two different control areas. From a statistical point of view, namely, in principle equally big samples have to be taken from every area for which a yield estimate with a given error margin is desired. In other words, if yield estimates are wanted from each of six extension areas, the total sample size will have to be six times as big as if only one estimate for the whole project area is needed. Likewise, to use two control areas takes twice the sample of one. Therefore, it was decided to divide the project area in two parts, roughly corresponding to its main ecological zones, by drawing an east-west line through Asella. For each of the two resulting parts, the northern and the southern, separate yield estimates were wanted. The two parts of the control area had to be taken together to give one yield estimate only for the whole control area. A map showing the different areas included in the study can be found in Figure 1.

Thus, the purpose of the study can be summed up as follows:

To give statistically valid estimates of the average yields of wheat, barley, beans and flax for each of the following three areas:

a) the northern part of the project area (called "area N");

b) the southern part of the project area (called "area S");

c) a control area, consisting of Sire and part of Lemu and Bilbillo weredas (together called "area C").

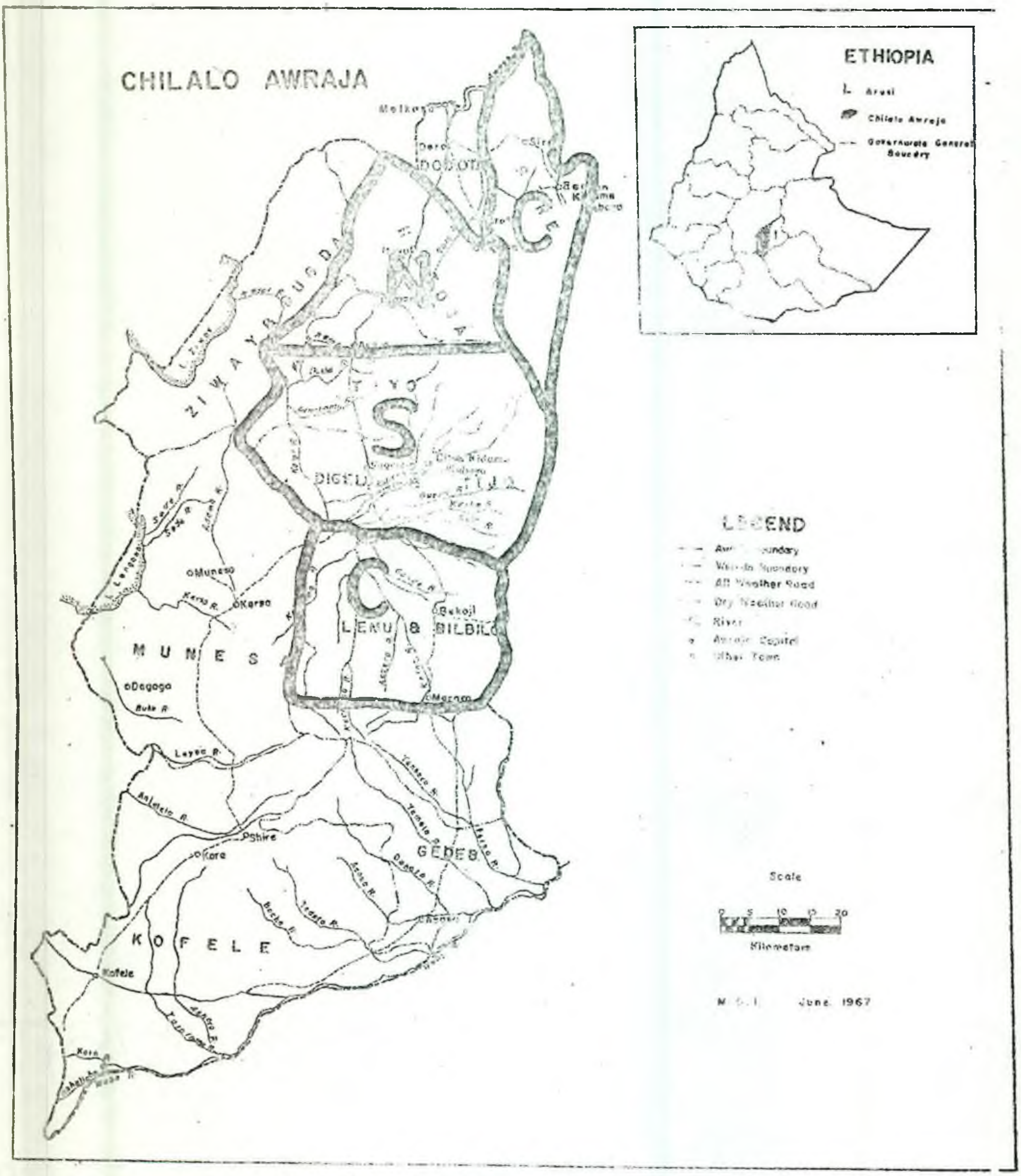
By "statistically valid" is in this context meant:

a) that the estimates should be based on unbiased samples;

b) that the precision of the methods used should be possible to calculate.

In addition to this study, a separate study of crop yields was carried out by the Extension and Education Department among the first group of model farmers, chosen in 1968. This study is separately reported in Chapter 5.

Figure 1. Map of Chilalo Awraja Indicating Areas Where Crop Yield Estimations Were Made.



2. STATISTICAL DESIGN.

2:1 Outline.

The basis for the statistical design has been the system used in Sweden for obtaining unbiased yield estimates. Because of the big difference between Swedish and Ethiopian conditions, however, the adjustment of the system has required a number of major changes. Just to mention some of the environmental differences that had to be observed by methodological adjustments, in Ethiopia no list of the farms nor of the farmers is available, from which the draw the sample; there is no advance information about farm sizes or crops grown; the farmer normally does not know the size of his fields (at least not in units of measurement that are precise enough to be used for sampling purposes); only very simple sampling procedures can be carried out by the field workers when they are at the farms, etc. It goes without saying that the precision of the methods used under these conditions will, ceteris paribus, be by far inferior to that of similar studies carried out under more favourable conditions. However, it is hoped that as time passes and succesively more information is collected, it will be possible to plan for increasingly higher precision every year.

Within each of the areas N,S, and C, the population (for each of the crops) can be defined as the total area, on which the crop is grown. Given the usual convention as to the way crop yields are indicated, kilogrammes (quintals, tons) per hectare, the square meter was chosen as the population element.

For evaluation purposes, one would desire estimates with rather narrow error margins. Especially for average crop yields, the changes to be expected as a result of the project activities are likely to be minute for a number of years to come (and will probably be difficult to tell from fluctuations due to weather differences between years). The situation, therefore, apparently calls for narrow confidence intervals around the estimates of average yields, if any project effects are to be detected. To start with, therefore, the sample sizes for each of the three areas of study were calculated on the basis of a desired precision (95-percent confidence intervals) of $\pm$ 5 percent around the means. The only advance variance estimates on which to base the calculations where those of last year's crop sampling. Furthermore, these were subject to an unknown amount of bias, more likely to make them smaller than bigger than the population variance. The sample sizes resulting from these calculations were quite unacceptable from a resource point of view. Instead, it was decided to accept twice as big error margins (i.e. 95-percent confidence intervals of $\pm$ 10 percent around the means), thus bringing down the sample size to a fourth of the original requirement (the size of the confidence interval is inversely proportional to the square root of the sample size). One argument behind this was the same as that used in other CADU evaluation studies: until more is know about prevailing condtions in the project area (especially as far as the variance of attributes to be studied and possible bases for stratification are concerned), it is uneconomic to try to attain high degrees of precision by simply increasing sample sizes. Rather, the increasing level of information likely to follow from the first few years project activities will make it possible to use more efficient research designs. Until then, it will be very expensive to avoid lower precision, since the only instrument for that will be increases of sample size.

As for the choice of sampling method, the use of multi-stage sampling seemed fairly obvious. Neither the available sampling frames nor cost considerations favoured any other arrangements. Also, there could be little doubt as to the steps of such a design: in the first stage, a number of farms were sampled, in the second fields within each farm, and in the third square meters within the fields. This is the same basic design as the one used in Sweden for similar studies. Ethiopian conditions made it attractive also for use in this study.

From the point of view of research efficiency, an important problem is that of allocating the available resources to the different sampling steps, i.e. to choose the number of units to be sampled in each stage. Unfortunately, very little advance information was available for a decision on this. The important factors to consider when allocating research resources would have been the relative costs of locating and observing one unit in each of the sampling stages, and the relationships between the within-field, within-farm and between-farm variances. The information available within CADU on the magnitude of these factors was virtually nil. Especially, the design of earlier years' corp sampling had not allowed for estimates of the different components of the variance. (The author is sorry to admit that by a mistake in the statistical design, there is still not very much more to report in this respect after the 1968 crop sampling). Without any particularly strong motives, it was decided to select one field per farm (for each crop) and two square meters per field. This, incidentally, is the same design as the one used in Sweden (although the author is aware that this fact is more likely to be an argument against than for the design chosen).

Each of the next three sections of this chapter are devoted to one of the stages in the sampling procedure.

2:2 The First Sampling Stage.

2:21 Sampling Frame.

As already indicated, there is no such thing as a list of all farms in the project area. The closest one can get to this is the official population census made in 1966 and supposed to list all households. Earlier CADU experience has shown, however, that a number of people have been missed when the census was taken. Since there is no satisfactory follow-up of population changes in Ethiopia (e.g. births, deaths and migrations), furthermore, some people listed in the census could not be found when CADU started to work in the area in 1967.

The second possibility for a sampling frame is the aerial photographs that have been made for mapping purposes in a bilateral United States aid project In principle, it should be possible to make area sampling from the photographs. When the possibilities for doing this were investigated, however, some important shortcomings became evident. Firstly, the photographing work was not completed for the entire project area when the planning of the crop sampling study was started. It was most doubtful whether photos for the whole area would be abailable when needed. Secondly, area sampling by this method presupposes quite some ability of the field workers to locate the sampled areas or spots on the ground with the help of the photographs. Under Ethiopian conditions, this means a heavy demand on personnel training and supervisory capacity, which was very hard to meet. The error risks likely to be incurrred had this method been chosen seemed rather prohibitive.

A closer look into the strengths and weaknesses of the population census from 1966 from the point of view of this study revealed that the main trouble is the bias likely to have been introduced by the failure to list some people. In one area south of Asella, Waji, a detailed census had been caried out recently by CADU personnel. In order to get an idea of the accuracy of the 1966 census for the planning of the crop sampling study, the two censuses were compared. It was found that the official 1966 census has missed 57 percent of the farm households (admittedly, this figure includes an unknown number of people who have migrated into the area after 1966). Still worse, the mistakes were systematic: 85 percent of the tenants had been missed but only 43 percent of the landowners. The number of households listed in the 1966 census but not found in 1968 was not investigated, since they would by definition not be of interest in the crop sampling study. Nothing is known about the representativity of the Waji area for the rest of the project area in providing a specimen collection of 1966 census errors. However, the census method used makes a concentration of errors to some areas in an unsystematic way likely. It can only be hoped that Waji was one of the areas with a high concentration of errors.

The paramount question for deciding whether to use the 1966 census as a sampling frame for the 1968 crop sampling, apparently is the effect on yield estimates of the main census error, the fact that a number of people were missed. The census method used was simple enough: census officials contacted low-level local authorities (balabats and golmasas), asking them to enumerate all people living within their areas of authority. Mistakes in these enumerations are likely to be mainly due to forgetfulness. As some local authorities are more likely than others to take the task ambitiously, the errors are, as already indicated, probably more concentrated to some areas than others. Within one area, on the other hand, the people most likely to be forgotten are the socially least important ones, i.e. the tenants. Hence the most probable bias. The important question for this study, therefore, is whether the conditions of cultivating the crops studied are likely to be different for landowners as compared to tenants.

From a general geographic-climatological point of view, this is most unlikely, since the enumeration areas were small and ecologically homoenous. Thus the farmers missed should have the same general agricultural conditions as those included. From the point of view of personal competence and ambition, on the other hand, one would at first glance be inclined to assume some difference between landowners and tenants. The information collected in various socio-anthropological studies made by CADU, however, does not give support for such a hypothesis. Instead, tenants have been found to have a strong interest to follow a "middle-of-the-road"-policy, i.e. to avoid getting both too low and too high yields. Since they all have various sorts of share-cropping arrangements, they run the risk of being evicted if their performance is unsatisfactory. On the other hand, if they produce too high yields, the landowner may think high of the potential of that land and evict the tenant in order to cultivate the land himself (thus getting all instead of, say, half the harvest). Furthermore, no other experience of CADU so far supports the view that landowners differ from tenants in their competence as farmers.

Against this background, it was decided to accept the 1966 census as a sampling frame for the crop yield estimation. The bias towards underrepresentation of tenants is likely to be there allright, but the disadvantage it implies seems clearly more acceptable than the ones that would follow from the use of aerial photographs. (For repetition of the study in the years to come, a better sampling frame will in all probability be available, viz. a register of all land holdings in the project area. It was discovered after the field work for this study had been completed and has been tested in a credit evaluation study during 1969.)

True, the census is a list of farmers, whereas the sampling units in the first stage of the sampling procedure were defined as farms. To cover the difference, a farm was defined as the land cultivated by a farmer in the census list. Under Ethiopian conditions, this definition is only likely to cause difficulties in two cases. One is if a farmer has died and the farm has been divided among the heirs. For this case, the instruction to the field workers was to make the second step of the sampling from the whole farm earlier owned by the farmer listed in the census (i.e. to treat it as if he were alive). The second case is when a farmer cultivates land at several different places. For practical reasons, the instruction to the field workers was to ask the sampled farmer which land he cultivated within the gasha* he was himself living in (including land in an adjacent gasha which is less than 400 meters from his house), and select the second sampling stage from that land. Since the number of farmers cultivating land in gashas in which they do not live could be expected to be small (an assumption which was confirmed during the field work), no bias was likely to be introduced by this practice. The main advantage of this definition of a farm is that it does not deal with land ownership conditions at all. These can be very complicated, with people owning land at distant locations, and cultivating part of it themselves and having tenants on part of it. There are also the problems of absentee landlords and ownership disputes. It is normally much easier to define what land a certain farmer actually cultivates, irrespectively of who owns or pretends to own it.

With the relationship between farmer and farm defined as per above, it is hoped that the reader will accept the changing terminology below, where sometimes farmers, sometimes farms are referred to as the first stage sampling units.

2:22 Sampling Method.

2:221 Areas N and S.

As earlier mentioned, the sample sizes were calculated on the basis of a precision requirement of 95-percent confidence intervals of ± 10 percent around the estimated mean yields and on the variances obtained in the 1967 crop sampling. For the sake of simplicity, the calculations were made on the assumption of simple random sampling, although other sampling methods were actually to be used. The obtained necessary numbers of observations per crop were the following:

	Area N	Area S
Wheat	69	95
Barley	71	94
Flax	44	77
Beans	72	68

Since the sampling fraction is negligible, the differences in sample size between areas N and S is entirely due to the higher variability of yields in area S (except for beans, where the situations is the reverse).

* A gasha is an administrative area, normally 40 hectares in size.

A systematic sample from the census had been made by CADU in the beginning of 1968 for another study (General Agricultural Survey of the Project Area, CADU Publications, Addis Ababa 1968). The sampled farmers had been interviewed about certain agricultural conditions at their farms, including total farm size and crops grown. Farmers included in the census but having moved (or died) later had been detected in that study and substracted from the sample. Therefore, and since the very sampling from the census lists was a rather troublesome procedure (the lists are in Amharic and kept at local offices), the 351 interviewed farmers were considered a convenient subsample for use in this study. To avoid bias, the not-at-homes were added to the interviewed farmers, making the subsample from the census totally 380. The remaining 6 non-response cases (apart from those not belonging to the population) were people with whom communication was for various reasons impossible. Their exclusion from the subsample cannot conceivably have created any bias.

The sampling fractions for the General Agricultural Survey had been chosen for each of the six extension areas separately and differed to some extent. To create an unbiased subsample of the census, therefore, certain corrections had to be made. Thus, the lowest of the sampling fractions used was divided by the sampling fraction of each of the extension areas. The result was a correction factor which was multiplied with the actual sample size for each area to make all sampling fractions equal. (Example: Huruta had a sample fraction of 0.05; the lowest of the fractions was that of Asella, 0.033; 0.033 divided by 0.05 is 0.667; 0.667 multiplied by 63, the sample size in Huruta extension area, is 42, which is the sample size that puts Huruta at par with the other areas as far as the sampling fraction is concerned). The surplus units of each extension area were removed by simple random sampling (with the use of a table of random numbers).

The result of this procedure was a representative subsample from the census, covering the entire project area and comprising 310 farmers. To give each population element (the square meters) equal probability of being selected, no matter whether they belonged to a big or a small farm, the farmers were now selected with probabilities proportional to the size of the farm.

There is, however, a difficulty with the traditional area measures used in Ethiopia, that is important from a sampling point of view and therefore has to be dealt with somewhat in detail. Most of the area measures used for indicating the size of a whole farm, such as tefer and gashas, can vary in size depending on the fertility of the land (they are bigger when the fertility is lower). For Arussi province, the Ministry of Agriculture has estimated the biggest units of one sort to be as much as 75 percent bigger than the smallest ones. On the other hand, the land in the project area is fairly homogenous from a fertility point of view: 84 percent is classified in the same fertility category (CADU Statistical Digest 1968). For this study the traditional measures were transformed to hectares according to standards earlier used both by CADU and other authorities (1 gasha = 40 hectares, 8 tefer = 1 gasha). In principle, this implies a faint bias towards overrepresentation of more fertile land. However, this bias can be taken care of in the procedure of estimation (by dividing each yield figure with the probability of selection for that particular farm) and should, therefore, be of no significance.

For another frequently used measure, the timad, the problem is a bit more complicated, since this is not an area measure but a measure of performance. It is defined as the area which can be ploughed in one day with one pair of oxen. Since this measure is more used for fields than for whole farms, the problem is more important for the second step in the sampling procedure, but it does affect also the first one. The size of a timad has been found to vary from 0.17 to 0.38 hectare in the project area alone, with a mean of 0.25 (no confidence intervals are known). Whenever the size of a farm was given in timads, it was transformed into hectares (using the mean of 0.25 timads per hectare). This means that a small bias has been introduced (also this one will be taken care of by the estimate chosen). The size of a timad varies with three main factors: the strength of the oxen, the diligency of the farmer and the easiness with which the soil can be ploughed. The lower a farmer and his farm score in these respects, the more timads is a given area divided into. The bias, therefore, is one of overrepresentation of less hard-working farmers with weak oxen and soils that are difficult to plough. Since only a minor share of the farms had their size indicated in this way, this disadvantage was accepted.

A number of farmers had not told the size of their farms at all. They were assigned values equal to the mean farm size of the extension area in question. There is no reason to believe these farmers to differ systematically from the others in any respect of importance for the study. Therefore, no bias is likely to result from this practice.

When these adjustments had been made, the required number of farmers for each of areas N and S were sampled with probabilities proportional to the farm size in hectares. The sampling was made as simple random sampling (using a table of random numbers) with replacement, after cumulation of the farm areas. The technique used was the one described in the earlier mentioned book by Cochran, p. 251. Replacement is necessary to keep the probabilities proportional to size, if the sampling fractions are not very samall. A number of farmers, consequently, were selected several times. The number of subsampled units from these farms was the same as for those only selected once, but the results were weighted with the number of times the farm was selected. This is the simplest of several possible methods having approximately the same precision, described in Cochran, p. 306. The required sample size was 72 in area N (the maximum number of observations necessary for any crop) and 95 in area S. Because a number of farmers in the sample were selected several times, the total number of farmers selected was only 55 and 59 for areas N and S respectively.

In addition to the sample, a reserve sample was drawn for each of the areas. It only contained 23 and 22 farmers for areas N and S respectively, but because of the weights assigned to some of the farmers, the sample plus the reserve sample would give a total of 166 and 113 observations in areas N and S respectively.

In the General Agricultural Survey information had been obtained on the crops each of the interviewed farmers grew in 1967. On the basis of this information the expected number of observations per crop studied in the crop sampling study could be calculated, assuming no changes in the proportions of farmers growing each of the four crops. If no major changes occurred between 1967 and 1968, the number of observations per crop would be enough to cover the sample size requirements earlier given, provided part of the reserve sample was used for some of the crops.

2:222 Area C.

As the control area (both parts of it) is not part of the project area, no advance information was available on conditions relevant for the planning of this study. For the same reasons as in areas N and S, the 1966 census was chosen as a sampling frame, although it had more disadvantages in area C. Thus, it had not been used by CADU at all, and therefore, the sample had to be big enough to allow for non-response from those no longer belonging to the population. Also, the addresses of the farmers in many cases were given in a rudimentary way, likely to make the farmers difficult to locate. For areas N and S, this trouble had been taken care of by the field workers of the General Agricultural Survey, so that the subsample of the census used for sampling in these areas carried the complete addresses of the farmers.

There was no information on farm sizes, crops grown or yield variances. On the assumption that the variances would be as big as in the part of the project area with the biggest variances (area S), therefore, it was decided to choose the same sample size as there, i.e. 95. Since the sample was to be divided between the two parts of area C, the number was rounded off and 50 farmers were sampled in each of the two parts. The southern part was defined as all balabat areas in Lemu and Bilbillo woreda north of an east-west line through Merarro (for practical reasons, this line had to follow the borders of the balabat areas, but they happened to form such a line pretty well). To allow for the high number of non-response to be expected, a reserve sample of 98 was also drawn. Since no information about crops grown was available, one could only wish that the number of observations per crop would be enough to give the desired error margins in the estimates, i.e. 95-percent confidence intervals of ± 10 percent around the mean yields.

Lacking advance information about the farmers, they were sampled with equal probabilities. Thus, a bias has been introduced: square meters belonging to farms of above average size are underrepresented. (Since the areas were not measured, for reasons given below, it was not possible to get an unbiased estimate by subsequent adjustments). As there was no alternative (within the resources available), this had to be accepted. In a coming year, however, the information collected from the control area should make possible the use of methods with increasingly high precision.

For drawing the sample from the census lists, systematic sampling was used. Practical considerations strongly favoured this method as compared to simple random sampling. The disposition of the census lists is not such that any bias can conceivably have been introduced by this method.

2:3 The Second Sampling Stage.

The sampling units in the second stage were defined (for each of the crops to be studied) as all the fields on which the crop was grown at a sampled farm. To give each square meter of, say, wheat grown at a farm the same probability of being selected, the fields ought to be sampled with probabilities proportional to size. Normally, however, a farmer could only give the size of each field in timads. To accept the timad measures as indications of size and to sample with probabilities proportional to them would give overrepresentation of soils that are difficult to plough. (The reader is reminded of the three main factors making for big timads: diligent farmers, strong oxen and soil that is easily ploughed. Now, on a given farm the first two factors will normally be constant over all fields, so that the differences in timad size will be almost exclusively accounted for by differences in "ploughability"). True, the differences in this factor should be rather small within the same farm, but certainly one cannot be sure of this. Rather, under the topographical conditions prevailing in the project area, the fields of a given farm can be very scattered and also have very different appearance and degree of slope. Thus, substantial variations in "ploughability" within the same farm are quite conceivable. Since this factor is likely to have something to do with the yield rate, furthermore, it was desired to avoid a bias in this respect.

The alternative that has no doubt been obvious to the reader for some time, to measure the fields in square meters and then do the sampling with probabilities proportional to size, unfortunately had to be ruled out entirely. To the average farmer in the project area, strangers with measuring devices appearing at the farm can only mean two things: expropriation or tax increase. Neither of the interpretations is likely to make the farmer cooperative and both may even be harmful to the good health of the field worker. It was decided, therefore, that time was not yet ripe for CADU employees to run the risk of creating frictions with the farmers. Technically, this sampling method would have been rather complicated and required some efforts as far as training and supervision of field personnel is concerned, (especially with the mostly very irregular shape of Ethiopian fields), to say nothing at all of the possibility of importing adequate equipment at a short notice. In the future, when CADU is better known locally and the field personnel better trained, changes can probably be made in this policy, with increased precision in the yield estimates as a consequence.

The only alternative to sampling with probabilities proportional to the size in timads, therefore, was sampling with equal probabilities. This, on the other hand, would give another bias: square meters belonging to smaller than average fields would be overrepresented and vice versa. After consultation with CADU:s specialists on crop production, it was decided that this bias was more acceptable than the other one. True, small fields may at times be located closer to the living house and thus be better supplied with manure. They could also sometimes lie on inferior soils, because a bigger field cannot be ploughed in, say, a ravine. On the whole, however, the connection between yield rate and field size seemed much less clear and systematic than between yield rate and "ploughability". One more advantage was attained by this decision. Sampling with probabilities proportional to size involves a much more complicated task for the field workers. Considering their lack of routine, this being the first time a statistically valid crop yield estimate was tried by CADU, the author was most happy to avoid adding new complications to their instruction.

For a given crop, the field to be studied was selected with simple random sampling, using a table of random numbers. The details of the instructions to the field workers can be found in Appendix 1 to this report.

2:4 The Third Sampling Stage.

The system used in Sweden to sample the square meters from a given field involves some rather sophisticated manipulations with coordinates and fields inscribed in rectangles. It was felt that a simpler method had to be designed for this study, especially considering the often very irregular shape of the fields in the project area.

The method used was the following. The samplers used a V-shaped measuring device, with 2 meters between its end-points (this distance could not be changed). They were instructed to start from a point at the edge of the sampled field (normally the point were the footpath from the house reached the field; it was called "the starting point"). Then two two-digit random numbers were chosen. The first one indicated the number of lengths of the measuring stick that they should go along the edge of the field (to the right if the number was odd, to the left if even). The other gave the number to go at a right angle into the field from the point to which the first number took them (called"the turning point"). At the point reached in the field (called "the sampling point"), a device was attached which worked as a pair of compasses giving a fixed area of one square meter (with its centre in the sampling point). All straws of the crop growing within that circle were harvested and collected.

For each field, a second square meter was also sampled. To do this, the samplers went back to the turning point (not to the starting point) at the edge of the field, took two new two-digit random numbers and repeated the procedure. The second sample was put in the same bag as the first one from the same field. (The statistical consequenses of this are dealt with in section 2:6). The average of the two samples from the same field is referred to below as "one sample" or "one observation" (also when the total number of observations made, n, is calculated).

This sampling procedure should give equal probabilities of selection for all square meters of a field, provided the field is not too big. As the intelligent reader has no doubt already concluded, the maximum circumference compatible with this requirement is 394 meters. If the circumference is bigger, the most remote parts will have a lower probability of being selected. Available information about field sizes, however, did not indicate any risk of getting too many such cases. The alternative would have been to use three-digit random numbers. For inexperienced field workers, however, they would have caused quite some difficulties and given room for a not negligible subjectivism, e.g. in deciding whether actually to measure a distance or to consider this unnecessary because the distance would take the sampler around the fields and past the starting point anyway. Also, there would definitely be much more running about, thus in general increasing the time per observation.

A main risk of bias with this sampling method is what happens when a random number gives a sampling point close to the edge of the field. To avoid too much trampling in the fields, the samplers were told not to actually measure the distance from the turning point into the field, if it obviously took them right across the field and out on the other side of it. On the other hand, this recommendation might lead to avoidance of sampling points close to the opposite edge of the field. Since crops often grow more sparsely close to the edge of the fields, this would have introduced a bias. The problem was dealt with by careful instructions and training, stressing the need of measuring the distances exactly when the sampling points appeared to be close to the edge of the fields.

2:5 Statistical Design Summed Up.

To recall briefly, the main characteristics of the statistical design were the following:

1. The desired precision was a 95-percent confidence interval around the estimated mean yields of not more than $\pm$ 10 percent.

2. In the first sampling stage, a number of farms were selected. Within areas N and S (the project area), they were selected with probabilities proportional to size. In area C, they were selected with equal probabilities.

Bias: overrepresentation of landowners, but this is not likely to affect the results. In area C there is an additional bias of overrepresentation of small farms, the influence of which on results is uncertain.

3. In the second stage, on each farm sampled one field per crop was selected with equal probability.

Bias: overrepresentation of small fields, which is not likely to have any systematic effect on yields.

4. In the third stage, two square meters were selected per sampled field with equal probabilities.

Bias: virtually none.

2:6 Estimators of Mean and Variance.

2:61 Type of Data Used.

When the field workers had harvested the two square meters of a selected field the two samples were put together in the same bag and brought to the laboratory. Thus, the available data allow for estimates of the average yield per farmer (per crop) only, but not for estimates of the within-field variance. For getting an estimate of the overall mean and variance of the sample, the within-field variance is not needed either, so from this point of view the procedure followed was OK. For efficient allocation of resources to the different sampling stages, however, information on the within-field variances is needed. In this respect, therefore, there was a deficiency in the data collection methods of the 1968 crop sampling. Unfortunately the next year's study shall have to be planned without this information, with a loss in efficiency to be expected, but it is hoped that this time more systematic measures will be taken to secure information pertinent to the planning of future studies. Especially, to get an idea of the within-farm variance (which is necessary for an efficient allocation of resources to the different sampling stages), at least two fields per farm should be sampled.

The samples were threshed and the water content determined for each sample. To make the results from different farms commensurable, the weight of each sample was then recalculated to give the dry matter weight at 10 percent moisture content. From a statistical point of view, this procedure is indifferent. The result was the weight in grammes for the yield of two square meters. This was divided by two to give each farmer's average yield per square meter (for each of the crops).

2:62 Areas N and S.

To get an unbiased estimate of the population mean and variance and the variance of the sample mean, the following procedure was followed. To simplify, the procedure is described for wheat only, but it will of course be the same for each of the other crops studied.

Notations (capital letters refer to the population, small letters to the sample):

Population:

Farmers in the population growing wheat

1, 2. ...... i, ... N'

Farmers in the population not growing wheat

N'+1, N'+2, ...N

Corresponding areas of farms in the population

$a_1, a_2, \ldots\ldots a_i \ldots\ldots a_{N'}, a_{N'+1}, a_{N'+2} \ldots\ldots a_N$

Total farm area in the population $= A = \sum_{i=1}^{N} a_i$

Respective probabilities of selection

$p_1, p_2, \ldots\ldots p_i \ldots\ldots p_{N'}, p_{N'+1}, p_{N'+2} \ldots\ldots p_N$

where $p_i = \frac{a_i}{A}$ and $\sum_{i=1}^{N} p_i = 1$.

Corresponding areas under wheat in the population

$t_1, t_2, \ldots\ldots t_i, \ldots\ldots t_{N'}, 0, 0, \ldots\ldots 0$ (0 for N-N' farmers not growing wheat)

Respective wheat yield rates based on all the wheat fields in the farm

$Y_1, Y_2, \ldots\ldots Y_i, \ldots\ldots Y_{N'}$

<u>Sample</u>:

Farmers in the sample growing wheat

$1, 2, \ldots\ldots i, \ldots\ldots n'$

Farmers in the sample not growing wheat

$n'+1, n'+2, \ldots\ldots n$

Areas under wheat in the corresponding sample farms

$t_1, t_2, \ldots\ldots t_i, \ldots\ldots t_{n'}, 0, 0, \ldots\ldots 0$ (0 for n-n' sample farms not growing wheat)

Average yield rates of the two sample plots in the respective sample farms growing wheat

$y_1, y_2, \ldots\ldots y_i, \ldots\ldots y_{n'}$

Two assumptions have to be made:

1. The yield rate of the two square meters sampled in a field is not correlated with the size of that field, or $E(y_i) = Y_i$. (E for " expected value of"). On the basis of the previous discussion of this problem, this assumption seems quite reasonable.

2. The total wheat area of a farm is not correlated with the wheat yield rate of that farm, or $Cov(t_i, y_i) = 0$. (Cov for "covariance of"). It should be noted that this is something else than to assume that the wheat yield is uncorrelated with the proportion of the area of a farm, on which wheat is grown. The latter may not hold, since a farmer who gets above average wheat yields may be inclined to allocate a larger share of his land to

wheat (and vice versa). But the assumption does not imply this. To exemplify what it does imply, suppose one has got the information that one farmer grows five hectares of wheat, and another only half a hectare. Assumption 2 implies that both are equally likely to have high (or low) wheat yields. That this is reasonable should be clear if one considers that with the information given in the example, one does not know whether the first farmer has a total farm area of, say, 50 or 5 hectares, or the second 10 or $\frac{1}{2}$. It may also be added that the assumption is supported by empirical evidence from several studies dealing with subsistence farmers, e.g. one FAO study from Arussi province itself, and several studies in India.

Then

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{t_i y_i}{p_i}\right] = \sum_{i=1}^{N} t_i Y_i$$

which is the total production of wheat for all the farms in the population. Since $t_i y_i$ is zero for the fields not growing wheat, this is equivalent to

$$E\left[\frac{1}{n}\sum_{i=1}^{n'}\frac{t_i y_i}{p_i}\right] = \sum_{i=1}^{N'} t_i Y_i$$

Similarly,

$$E\left[\frac{1}{n}\sum_{i=1}^{n'}\frac{t_i}{p_i}\right] = \sum_{i=1}^{N'} t_i$$

which is the total area under wheat for all the farms in the population.

Then the yield rate $\bar{Y}$ of wheat for all the farms in the population can be estimated by dividing the estimate of the total production of wheat by the estimate of the total area under wheat in the population.

Thus

$$\text{Est}(\bar{Y}) = \frac{\text{Est}\left(\sum_{i=1}^{N'} t_i Y_i\right)}{\text{Est}\left(\sum_{i=1}^{N'} t_i\right)} = \frac{\frac{1}{n}\sum_{i=1}^{n'}\frac{t_i y_i}{p_i}}{\frac{1}{n}\sum_{i=1}^{n'}\frac{t_i}{p_i}} = \frac{\sum_{i=1}^{n'}\frac{t_i y_i}{p_i}}{\sum_{i=1}^{n'}\frac{t_i}{p_i}}$$

This is a ratio estimate, and is thus subject to the usual limitations of such estimates.

Since the total areas t_i under wheat for the sample farms were not measured (true, they were measured in timads, but due to the weaknesses of that measure, it cannot be used in the formula), the above estimate cannot be used, unless assumption 2 above is also made. Then,

$$\sum_{i=1}^{N'} t_i Y_i = \bar{T}\sum_{i=1}^{N'} Y_i, \text{ where } \bar{T} = \frac{\sum_{i=1}^{N'} t_i}{N'}$$

$$\text{and } \frac{1}{N'}\sum_{i=1}^{N'} Y_i = \frac{\sum_{i=1}^{N'} t_i Y_i}{N'\bar{T}} = \frac{\sum_{i=1}^{N'} t_i Y_i}{\sum_{i=1}^{N'} t_i} = \bar{Y}_{\text{weighted}}$$

Thus $\bar{Y}_{\text{unweighted}} = \bar{Y}_{\text{weighted}}$ under these assumptions.

Hence one can act as if all the farmers growing wheat in the population had the same area under wheat $\bar{T}$. Then

$$\text{Est}(\bar{Y}) = \frac{\sum_{i=1}^{n'} \frac{t_i y_i}{p_i}}{\sum_{i=1}^{n'} \frac{t_i}{p_i}} = \frac{\bar{T} \sum_{i=1}^{n'} \frac{y_i}{p_i}}{\bar{T} \sum_{i=1}^{n'} \frac{1}{p_i}} = \frac{\sum_{i=1}^{n'} \frac{y_i}{p_i}}{\sum_{i=1}^{n'} \frac{1}{p_i}}$$

An unbiased estimate of the variance of $\text{Est}(\bar{Y})$, $\text{Est } V(\bar{Y})$, will be given by

$$\frac{n'}{\left(\sum_{i=1}^{n'} \frac{1}{p_i}\right)^2} \cdot \frac{1}{(n' - 1)} \cdot \sum_{i=1}^{n'} \frac{1}{p_i{}^2} \cdot (y_i - \bar{y})^2$$

Then, the sample variance (s^2) will be $s^2 = \text{Est } V(\bar{Y}) \cdot n'$.

2:63 Area C.

Since all units had been sampled with equal probabilities in area C,

there was no problem in choosing the estimators for $\bar{y}$, s^2 and $s^2_{\bar{y}}$.

On the same two assumptions as in areas N and S, the ordinary formulas were used.

3. THE FIELD WORK.

3:1 Administrative Arrangements.

A number of sampling teams were set up to carry out the field work. Each team consisted of two persons with the necessary equipment. The personnel was entirely Ethiopian, with at least one member per team who in addition to Amharic also spoke Gallinya (the prevailing languages in the studied areas are Amharic and Gallinya).

Before the field work started, a two-days training course was given to all sampling teams. The training included arguments to be forwarded to the farmers to make them co-operative, sampling procedures and harvesting techniques. Before any sampling team was sent out to the farmers, they had practiced the sampling procedures several times on CADU:s own farms. There did not seem to be any difficulty for the field staff to learn the techniques.

The sampling teams worked under the supervision of a field supervisor. His tasks were to collect information from CADU:s extension organization as to suitable times for harvesting in different areas and to direct the sampling teams accordingly, including the organization of transports. He was in his turn supervised by CADU:s Assistant Extension Supervisor.

The main practical problem was to avoid coming too late to the farmer, i.e. after he had harvested the crop. The field organization managed to create a most efficient system of communication, however, using a number of informal channels and thereby getting advance information about the farmers harvesting plans. The number of observations missed because the crop was already harvested was surprisingly small. Its implications are discussed in chapter 4 in connection with the results.

The field workers were instructed to fill in a form (called sampling form) for each farmer they visited. This was made partly for control purposes, partly to collect some additional information apart from the crop yields. The sampling form can be found as Appendix 2 to this report.

The field supervisor gave daily telephone reports to CADU:s office in Addis Ababa, stating code numbers of farmers visited and number of observations obtained per crop. This made it possible for the staff of the Planning and Evaluation Section to follow the work closely all the time (the field work took almost two months to complete) and to make necessary changes. Thus, it was soon found that the actual number of farmers who grew flax and beans was smaller than expected. Therefore, it was decided to include the entire reserve list for areas N and S in the sample, in order not to get too few observations for these crops.

In the control area, the practical problems were much bigger. The leaders of the sampling teams were not CADU personnel, but the ordinary Ministry of Agriculture extension agents. They had a much lower motivation than CADU:s own field workers and were less acquainted with the farmers. Great difficulties to find farmers were reported as well as a big number of observations missed because the farmers had already harvested. Still worse, the probability for a farmer to be missed obviously rose with increasing distance from the extension agent's office, which probably has introduced a bias. The number of samples obtained was extremely low for some of the crops. As will be seen in the chapter on results of the study, in some cases there were even too few observations to give any useful results at all.

On the whole, therefore, the results of the crop sampling in the control area are of only moderate value in themselves. The main use will be in providing a better ground for next year's crop sampling in these areas. The difficulties are now known and it should be possible to take the necessary precautions (above all by using CADU personnel) to avoid new failures.

4. RESULTS.

The mean yields per hectare of the sample and the data necessary to judge the reliability of them as estimates of the real yields can be found in Table 1. Although the actual measurements and the data processing work were carried out in grammes per square meter, the figures have been converted to quintals per hectare, in order to comply with current conventions as to how to indicate crop yields.

The fact that a number of observations were missed because the crop was already harvested is not likely to affect the result in a systematic way. There are three main causes for failures to get the crop sample in time:

a) The farmer could have sown an ordinary variety of the crop at an unusually early time, thus having to harvest it early. The main difference between his yield and that of farmers with ordinary sowing time would be due to weather differences during the growing period.

b) The farmer could have used an early variety of the crop. That variety might give higher or lower yields than the ordinary varieties.

c) The sampling team might for practical reasons have come too late to a farmer, although he harvested in normal time. The yields of such farmers would not be likely to differ systematically from those of the farmers from which samples were obtained.

Taken together, none of these three causes is likely to give a systematic bias in the results. To all intents and purposes, therefore, one can regard the results obtained for areas N and S as unbiased as far as this type of non-response is concerned.

In areas N and S, the number of observations missed for other reasons than "already harvested" was 12 and 17 percent of the farmers respectively (weights considered). In some of these cases, it was only a question of units not belonging to the population ("farmers" not being farmers at all but having other occupations and farms on which none of the crops studied were grown). In most of the cases, however, the reason was unability of the sampling teams to locate the farmers because they were not known in the area were they were supposed to be found. Some of these cases probably did not belong to the population either. Some of the farmers not found, however, may have been succeeded by other farmers as cultivators of a particular piece of land. These successors ought to have been included in the study, but it was not possible for the sampling teams to collect enough information for making such substitutions. No bias is likely to have been introduced by these cases.

In area C, the reporting from the field work was not accurate enough for the number of observations missed per crop to be known. The total number of farmers, from whom no samples were obtained, however, was 121 (out of a total of 198 farmers in the sample and reserve sample). Virtually no information is available as to the reasons why no samples were collected from these farmers. Clearly, this makes the data from area C very unreliable, especially as there is reason to believe the non-response farms to be located in more remote areas. As earlier indicated, therefore, the result of the 1968 crop sampling for the control area will be of more use as a basis for planning future studies in that area than as indicators of actual yield levels.

Table 1. Results of the 1968 Crop Sampling.

	Area N (north of Asella)					Area S (south of Asella)					Area C (the control area)			
Crop:	$\bar{y}$	error margin	s	n	No. of samples missed	$\bar{y}$	error margin	s	n	No. of samples missed	$\bar{y}$	error margin	s	n
Wheat	10.1	± 1.1	5.3	83	4	9.1	± 1.2	5.4	73	0	9.2	± 1.8	5.6	40
Barley	11.8	± 1.9	6.8	50	13	13.4	± 1.0	5.7	118	10	12.6	± 1.4	3.7	26
Flax	4.1	± 0.7	2.0	33	2	3.3	± 0.4	1.5	47	2	4.4	± 1.1	1.9	15
Beans	19.5	± 2.7	3.3	46	10	24.6	± 1.0	3.4	43	22	14.8	± 7.6	11.3	11

Explanations:

$\bar{y}$: the mean yield per hectare (in quintals).

Error margin: 95-percent confidence interval. This means that the probability that the true mean (i.e. the mean of the population) is covered by that interval is 0.95. The size of the interval depends on the variability of the results (which increases the interval) and on the number of observations (which decreases it). Quintals.

s: the standard deviation (a measure of the variability of the observations). Normally, two thirds of the observations lie within the mean ± the standard deviation. Quintals.

n: the number of observations (samples) obtained per crop (after weighting with the weights assigned to farmers selected more than once).

No. of samples missed: these samples were all missed because the farmers had already harvested when the sampling teams reached them. No samples missed for other reasons are included in these figures.

The desired precision, 95-percent confidence intervals of not more than ± 10 percent around the means, was attained in only two cases in the project area (barley and beans area S), and in no cases in the control area.The failure to reach the desired precision levels was mainly due to the insufficient advance information on variances and frequencies of growing different crops. With the increasing amount of relevant information available within CADU, it should be possible successively to reduce the number of such failures in the future.

The data are on an average lower than those obtained in earlier years' crop yield estimations made by CADU. This supports the earlier made suggestion of a bias in the earlier studies towards overrepresentation of farms with above average yields. As already indicated, the amount of bias in these studies is unknown and impossible to calculate, so a closer comparison between the 1968 and earlier data is meaningless.

Some interesting by-products were obtained from the study. Each farmer in the sample was asked about the total area cultivated by him and about the total area and number of fields for each of the crops studied. Also, his tenancy status (landowner or tenant) was investigated. This makes it possible to estimate the total cultivated area of the project area and its distribution on the four crops studied. Information on the total farm population was available earlier (CADU Statistical Digest 1968). The main results are given in Table 2.

Table 2. Certain Agricultural Data From the Project Area.

a) Average Size of Cultivated Area. Per Farm and Major Crops Timads. Notations as in Table 1.

	Area N			Area S		
	$\bar{y}$	error margin	s	$\bar{y}$	error margin	s
Cultivated area per farm	15.0	±2.2	10.4	16.3	±1.8	11.0
Wheat area (per farm growing the crop)	4.0	±0.7	3.2	3.4	±0.9	3.8
Barley area (ditto)	2.7	±0.1	1.9	7.2	±1.6	8.6
Flax area (ditto)	1.9	±0.8	2.3	2.7	±0.7	2.5
Beans area (ditto)	1.2	±0.1	0.8	1.0	±0.2	0.7

Cultivated area per farm, total project area (areas N + S): 15.8 timads.

Total cultivated area in the project area: 215,000 timads* ± 28,000 (approximate error margin). This is equivalent to 32,000 hectares at a conversion rate of 1 timad = 0.15 hectares; or to 54,000 hectares if 1 timad = 0.25 hectares.

b) Tenancy Status.	Area N	Area S
Percentage of farmers being tenants	46	29
" " " " landowners	54	71
	100	100

* This figure was calculated in the following way. Total no. of farm households: 13,600. (CADU Statistical Digest 1968 gives a total population in the project area of 108,300; from this was subtracted population in urban areas: Asella, Huruta, Dera, Sagure, Digellu, Gonde, Itaya; lacking population figures for the last four ones, they were estimated at 4,000 inhabitants together; this gave a net rural population figure of 85,900, which was divided by 6.3, the average household size in the project area according to the General Agricultural Survey.) 13,600 multiplied by 15.8 timads is 215,000 timads.

5. CROP SAMPLING AMONG MODEL FARMERS.

Contemporaneously with the main crop sampling study, the Extension and Education Department of CADU carried out a similar study among the first group of model farmers. (Non-CADU readers are informed that these are farmers used as innovation disseminators in a two-step dissemination process, used by CADU for spreading agricultural innovations. They are not necessarily better farmers, but should among other things have a more progressive attitude towards change and have more social contacts). 1968 being the first year for extension activities, the recommended innovations were demonstrated in small plots at the model farms. The farmers had, however, not yet had an opportunity to adopt the innovations.

The design of that study was exactly the same as that of the main study, except that all farmers were studied. Thus, it was a total survey as far as the farmers were concerned and a two-stage sampling survey as regards fields and square-meters. Also, two more crops were added: peas and teff.

This study was made by separate sampling teams, administered by the Extension and Education Department but having undergone the same training as the teams of the main study.

Since the sampling procedures used for fields and square meters were the same as those of the main study, the estimates should have the same sort of bias: overrepresentation of small fields. As earlier stated, this bias is not likely to affect the results systematically. The estimator used for mean and variance of the mean were the ordinary $\bar{y}$ and $s^2_{\bar{y}}$. The results can be found below with 95-percent confidence intervals.

Table 3. Results of the 1968 Crop Sampling among Model Farmers.

Notations as in Table 1.

Crop	$\bar{y}$	error margin	s	n
Wheat	9.2	± 1.8	5.5	36
Barley	12.9	± 3.2	8.4	28
Flax	4.1	± 1.0	2.1	18
Beans	17.4	± 4.3	8.3	17
Teff	11.8	± 7.6	4.8	4
Peas	13.5	± 7.5	8.8	8

If tables 1 and 3 are compared, it can be seen that for all the four crops to be found in both tables, there is overlapping between the error margins of the model farmer estimates and those of the two halves of the project area. Thus, there are no significant differences (on the 95 percent level) between model farmer yields and those of the other farmers.

REFERENCES:

CADU Statistical Digest. CADU Publications, Addis Ababa 1968 (mimeographed).

Cochran, W.G.: Sampling Techniques. Second ed., New York 1963.

Freund, J.E. & Williams, F.J.: Elementary Business Statistics. Englewood Cliffs, N.J. 1964.

General Agricultural Survey of the Project Area. CADU Publications, Addis Ababa 1968 (mimeographed).

CHILALO AGRICULTURAL DEVELOPMENT UNIT

Planning and Evaluation Section

Appendix 1

CROP SAMPLING 1968

INSTRUCTION FOR THE SAMPLERS

We make the crop sampling because we want to know the yields of some of the crops grown in the project area. For comparison, we also make crop sampling in two areas outside the project area (around Bokoji and Sire).

To measure the yields of every field in the project area would of course be an impossible task. Therefore, we use statistical techniques to calculate the yields of all fields from only a sample of fields. This is only possible if you follow these instructions very thoroughly. Otherwise the statistical techniques will not work, and we will get a wrong idea of the yields. The basic idea is that neither you nor we, but chance only should decide if a piece of land is to be sampled or not.

Before you go to a farm

The areas

We want to compare the yields of three different areas:

area N: the project area north of Asella;

area S: the project area south of Asella;

area C: the comparison area (Bokoji and Sire).

Therefore, the farmers to be visited are divided in these three groups. Each farmer has a code, for instance N H 24, showing the area (N), the extension area (H for Huruta) and the farmer's number within the area N. Make sure you use the correct code when you fill in the sampling form.

The lists

For each area, you have a sample list of all the farmers that you should visit. You may visit them in any order you like. It is very important that you visit every farmer in the list and that you make him co-operative enough to let you make the crop sampling. If you have troubles in finding some of the farmers home, you have to call back again. When you have visited a farmer's house three times without finding him home, you may give it up and make a note in the sampling list : "not at home".

If you cannot find the farmer at home, try to find somebody else who belongs to the same household (brother, father, grown up son, etc.) and can tell you where the fields are. If the person in the list is dead, but his sons (or daughters) have inherited his land, ask about all the land that the sons (and daughters) cultivate together.

In some cases you may not be able to find a farmer at all (he may be dead or have moved from the area). Some farmers may even refuse to co-operate, even if you try hard to persuade them. In all these cases, you make a note in the sample list, telling the reason why you could not sample the crops at his farm.

For every farmer in the sample list that you cannot get any crop samples from, you must take another farmer from the reserve list (for the same area). This is because the statistical techniques will not work if the number of farmers is too small. When you use the reserve list, it is very important that you take farmers from it in the same order as they have in the list. A farmer from the reserve list cannot be used unless all the farmers before him in the reserve list have been used first. For instance, if you take crop samples from twenty farmers of the sample list, and then find that you need two farmers from the reserve list, you take the first two from that list. You may visit them before you go on with the sample list, and you may also visit number two before number one, as long as you make sure you visit both. You may not, however, choose any other farmers from the reserve list than number one and two, when you take your first two names from that list.

The total number of farmers, from whom crop samples have to be made, is stated on top of the sample lists. Note that this number is different for each of the three areas.

Extra instruction for areas N and S.

Some of the farmers in the sample list have an extra number behind their code, for instance SA 56 (2), instead of only SA 56. This means that they have to be treated in a special way for statistical reasons. You don't have to worry about that. But there is one thing that you must remember about these farmers: if you can't make a crop sample from such a farmer (he may not be home, he may be dead or for some other reason), you must take not one farmer from the reserve list, but as many as the number within ( ) shows. For instance, if his code is SA 56 (2), and you cannot find him at home although you visit his house three times, you have to take two other farmers from the reserve list instead. If his extra number is (5), then you have to take 5 other farmers from the reserve list, and so on.

For statistical reasons we need a certain number of samples for each crop. Therefore it may happen that you have to visit some more farmers than those that are in the sample list now. We don't know about this, however, before you have visited all the farmers in the sample list (plus those from the reserve list that you need if you can't get crop samples from some in the sample list). You will be told about this later.

When you come to a farm

You start by explaining to the farmer why you have come. Then you start to fill in the Sampling Form. Suppose wheat is ready for harvesting when you come. Then you do the following:

1. Ask the farmer how many wheat fields he cultivates himself (or with the help of somebody) within the same gasha. Do not count fields in other gashas or fields that he has rented to other farmers. If he has fields in another gasha that is so close to his house that the fields there are not more than 400 meters away from the house, then you should include those fields also. It does not matter if the fields that he cultivates are his own or

only rented. If he cultivates for instance 4 fields of wheat, you call them no. 1, 2, 3 and 4 respectively, starting with the one closest to his house and ending with the one farthest away from the house.

2. Take the Table of Random Numbers, start at page 1, the upper left corner and go to the right along the first line (just as when you read in a book). The first time you find one of the numbers 01,02,03, or 04 (because the farmer has 4 fields), you make a ring around it with a pencil (not a pen). You draw a line across the numbers that you have passed by (also with a pencil). It will look like this:

~~59 71 09 78 91~~ (04) 09 26.......

The number with the ring around it shows you which of the wheat fields that you should collect the crop sample from. Notice that all the numbers in the Table of Random Numbers have two figures. Therefore, the numbers 1-9 have 0 before them, but that does not make any difference. As you understand, it is very important that you remember which of the wheat fields that you called number 1, 2 and so on. If the farmer has many fields, you had better make a sketch of them, showing the numbers, before you start using the Table of Random Numbers.

3. Go to the field that was chosen by the random number. Choose a starting point, that is a point at the edge of the field where you start the work. The best starting point is the point where the foot-path from the house reaches the field. If there is no such path, you choose the corner of the field that is closest to the house. If you cannot stand up and work at the corner (there may be a tree or something else), choose a point as close to it as possible.

4. Take the random number immediately after the one with the ring around it in the Table of Random Numbers. Make a ring around it too (a ring means that you have used the number). It shows you how many lengths of the

measuring stick that you should go along the edge of the field. If the number is odd (1, 3, 5, 7 and so on), you should go to the right from the starting point, if it is even (2, 4, 6, 8 and so on) you should go to the left. Don't forget to fill in the random numbers in the Sampling Form. When you go with the measuring stick, you go along the edge of the field, no matter if it is straight or bent. It may happen that the number is so high that it takes you round the field and back and past the starting point again. That is quite OK. You should always follow the random number, no matter if it seems silly to you. The point where you stop after having measured the length indicated by the random number is called the turning-point.

5. Take a new random number (which means that you take the one immediately following the last one with a ring). This shows you how many lengths of the measuring stick that you have to go from the turning-point into the field. You should go at right angles to the edge of the field. You find the right angle by standing on the edge of the field (on the turning-point) with your arms stretched out from your sides along the edge of the field. Your face should be towards the field. Put your hands together with your arms straight. If you look across your thumbs into the field you will have a line at right angles to the edge of the field. Before you start walking, find a point behind the field that you can walk towards (a tree, a house, or something else). As long as you go from the turning-point towards that thing, you will be on the line that is at right angles to the edge of the field.

6. Sometimes the random number may be so high that it takes you across the field and out on the other side. If this happens, you have to take another random number from the table. Sometimes you will have to try several numbers before you find one that takes you into the field but not out of it again. The point shown by this random number is called the sampling point. Make sure you draw a line across those numbers of the table that you have not used and a ring arond the one that you finally use. Don't forget to note the number that you really use in

the Sampling Form (you may also note those that you have tried but not used, if that is easier, but make sure you cross over them). If you get a number that is close to the other edge of the field, you must really measure the distance, so that you are sure whether the sampling point is inside or outside the field. This is because it is important that the parts of a field being closest to the edge are not missed in the sampling. Even if the sampling point is only a few centimeters inside the edge, you must use it. To be able to harvest one square meter, however, you have to move the point a bit into the field. Move it only just so far that the edge of the circle that you harvest touches the edge of the field. (On the other hand, if the sampling point is only a few centimeters outside the edge, you must take another random number and start measuring again. <u>You may absolutely not</u> go just inside the edge of the field and call that the sampling point if the point is really a bit outside the field.) If there is a tree or some other obstacle within the circle that you should harvest, choose another random number.

7. Exactly in the sampling point you put the centrum stick. Then you harvest as you have learned.

8. When you have harvested around the first sampling point, you go back to the turning-point. Use the turning-point as starting point for a new sampling point, that is take a new random number, go along the edge of the field to the right (if it is odd) or to the left (if it is even) to your second turning-point. Then take another random number and go into the same field again at a right angle from the second turning-point. When you have found your second sampling point, you harvest there.

9. Put both the crop samples from the field into the same bag. Write the sample code (see Sampling Form) on the bag. Now you have completed the sampling of this crop at this farm.

10. Make a sketch of the field in the Sampling Form, showing the starting, turning and sampling points.

After you have sampled one crop

For each of the other three crops, you make the whole procedure once again. If the farmer does not grow all the four crops, sample those of them that he does grow. They will probably not be ready for harvesting at the same time as the first crop. Then you have to come back to the farmer later, when each of them is ready to harvest. Try to have the farmer give a message to you when you can come back, so that you don't have to visit him too often. But make sure you don't miss his harvest! You may, however, make your harvesting some days before the farmer harvests the same field. If you do that, the crop has to be ripe before you collect your sample.

All the time you take new numbers from the Table of Random Numbers until you have used all of the table. If you still need more numbers, start from the beginning again, now using a pen instead of a pencil to make the lines and rings. It does not matter if you use a number that was used already the first time you went through the table.

If the reserve list does not have enough names in it, you stop when you have used all of them. We have no other names, so be sure you really try to find everybody at home.

Appendix 2

CHILALO AGRICULTURAL DEVELOPMENT UNIT
Planning and Evaluation Section

CROP SAMPLING 1968

Team leader: ....

Sampling Form

Code of farmer (from sample list): ....

Name of farmer: ....

Landowner ☐ Tenant ☐

Size of total area cultivated by the farmer: ....

Date: ....

Wheat

Wheat area: ....

No. of wheat fields: ....

RANDOM NUMBER 1: ....(= field chosen).

RANDOM NUMBER 2: .... is odd (right)/even(left).

RANDOM NUMBER 3: ....(into the field).

RANDOM NUMBER 4: .... is odd (right)/even (left).

RANDOM NUMBER 5: ....(into the field).

The code of this sample is "W........."(W + the code of the farmer from the first line of this page).

Sketch of the wheat field chosen (indicate starting points, turning points, sample points and living house):

Code of farmer: ....

Barley

Barley area: ....

No. of barley fields: ....

RANDOM NUMBER 6: ....(= field chosen).

RANDOM NUMBER 7: .... is odd (right)/even (left).

RANDOM NUMBER 8: ....(into the field).

RANDOM NUMBER 9: ....is odd (right)/even (left).

RANDOM NUMBER 10: ....(into the field).

The code of this sample is "BAR ...."

Sketch of the barley field chosen (indicate starting points, turning points, sample points and living house):

Flax

Flax area: ....

No. of flax fields: ....

RANDOM NUMBER 11: ....(= field chosen).

RANDOM NUMBER 12: .... is odd (right)/even (left).

RANDOM NUMBER 13: ....(into the field).

RANDOM NUMBER 14: .... is odd (right)/even (left).

RANDOM NUMBER 15: ....(into the field).

The code of this sample is "F ....".

Sketch of the flax field chosen (indicate starting points, turning points, sample points and living house):

Code of farmer: ....

Beans

Bean area: ....

No. of bean fields: ....

RANDOM NUMBER 16: ....(= field chosen).

RANDOM NUMBER 17: ....is odd (right)/even(left).

RANDOM NUMBER 18: ....(into the field).

RANDOM NUMBER 19: ....is odd (right)/even(left).

RANDOM NUMBER 20: ....(into the field).

The code of this sample is "BEA ......".

Sketch of the bean field chosen (indicate starting points, turning points, sample points and living house):

LIST OF CADU PUBLICATIONS

A. Project Preparation Period

1. Report No. I on the establishment of Regional development project in Ethiopia, October 1966
   Part I General Background
   Part II Project Outline
   Part III Appendices
   (A reprint of the Summary is also available)
2. Report No. II on the establishment of a regional development programme in Ethiopia, May 1967. (The building programme appears under separate cover)
3. Trials and demonstration plots at Kulumsa in 1966, July 1966
4. Reconnoitering survery of the water resources in Chilalo Awraja, March 1967.
5. Creation of a forestry administration in Arussi province, March 1967
6. Crop sampling in the Chilalo Awraja 1966, May 1967
7. Results of trials and observation plots at Kulumsa 1966/67, May 1967
8. Sagure, a market village, June 1967
9. Forest nursery and planning techniques, June 1967
10. Trials and demonstration plots at Kulumsa and Swedish Mission Asella in 1967, July 1967
11. Grain Marketing experiments 1967, August 1967

B. Implementation Period

1. Government Agreement on Plan of Operation
2. Some reflections on water erosion in Chilalo awraja, October 1967
3. The Taungya afforestation method, Novermber 1967
4. Grow better Bahr-Zaaf in Ethiopia, January 1968
5. CADU Semi-annual report 1967/68, January 1968

6. Census in Sagure-Yeloma 1967, February 1968
7. The changing rural society in Arussiland: Some findings from a field study 1966-67, March 1968
8. CADU (Pamphlet in English and Amharic)
9. CADU plan of work and budget 1968/69 (with preliminary estimates for 1969/70)
10. Cultivation practices and the weed, pest and disease situation in some parts of the Chilalo awraja, March 1968
11. Introductory agro-botanical investigations in grazed areas in the Chilalo awraja, June 1968
12. Results of trials and observations on fields forage crops at the Kulumsa farm and in Asella 1967/68, June 1968
13. Crop sampling in the Chilalo awraja, Arussi province 1967, June 1968
14. General agricultural survey, August 1968
15. CADU statistical digest, May 1968
16. Descriptions of agricultural demonstrations 1968
17. Field trials and observations 1968/69
18. Feasibility study on a farm for breeding of grade cattle at Gobe, Arussi province, September 1968
19. Feasibility study on the electrification of Sagure town, September 1968
20. CADU Annual report 1967/68, September 1968
21. Census in Dighelu village, May 1968
22. A case study of peasant farming in Dighelu and Yeloma areas, Chilalo awraja, Ethiopia, January 1969
23. CADU Semi-annual report 1968/69, February 1969
24. Results of demonstrations 1968/69
25. CADU Plan of work and budget 1969/70
26. Tentative CADU programme 1970/75, Addis Ababa, March 1969
27. Feasibility study on sunflower protein concentrate and fafa mixing plant, May 1969
28. Results of trials and observations 1968/69
29. CADU Evaluation studies, Health education (Base-line study) May 1969
30. CADU Evaluation studies, Crop sampling 1968, May 1969