

Rothamsted Repository Download

A - Papers appearing in refereed journals

Church, B. M. 1971. The place of sample survey in crop loss estimation.
FAO. pp. 2.2/1-2.2/8.

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/970w0/the-place-of-sample-survey-in-crop-loss-estimation>.

© Please contact library@rothamsted.ac.uk for copyright queries.

THE PLACE OF SAMPLE SURVEY IN CROP LOSS ESTIMATION

by

B. M. CHURCH

Rothamsted Experimental Station
Harpenden, Herts., England

1. Introduction

When a problem is intractable and there is no obvious method of tackling it, a frequent suggestion is to do a survey. This approach is no less common in crop-loss estimation work than elsewhere.

The aim of this chapter is to indicate the circumstances under which sample surveys may make a useful contribution, and the limitations on what may reasonably be attempted by survey methods, and to draw attention to the main requirements for planning successful surveys.

2. Requirements for successful agricultural surveys

Although the major points requiring attention cannot be fully or equally discussed here, it is worth stating them briefly, because they are closely interrelated and neglect of any one can waste much effort. A standard textbook on survey methodology (e.g. 1, 3, 4, 5) should be consulted for an adequate discussion and, in planning any survey, it is also best to seek advice from an agricultural statistician who has had survey experience under comparable conditions.

DEFINITION OF OBJECTIVES

The first requirement is to define objectives adequately. Only then can one usefully consider whether survey methods are appropriate and what techniques might be practicable and efficient.

Broadly, decisions are always needed on:

- (a) the region (or 'population') to be covered by the survey;
- (b) the characteristics to be estimated;
- (c) the sub-regions or sub-classes (varieties, soil types, etc.) for which estimates are needed, and the accuracy required.

These decisions determine the main features of the survey plan: (b) influences the techniques and type of people required and (c) the scale of the survey. Objectives may have to be modified if requirements are too exacting; crop-loss surveys can guide the allocation of resources but must not use too much of what they aim to conserve.

Survey planning is likely to be most effective when a draft of the required tables and a general outline of the report is prepared at an early stage, because this helps to clarify objectives.

THE SAMPLE

A survey should provide information from truly representative samples of the region under study, not just from farms convenient to visit or known to the extension services. The sample must be sufficient to provide estimates of required accuracy. Sampling is discussed more fully in Section 4.

METHODS OF OBSERVATION AND MEASUREMENT

Methods should be so defined that they are consistent and reproducible as used by the field staff, and as objective as possible. Methods must also be acceptable to the farming community.

RECORD FORMS

All forms should be readily understood, and convenient for use in the field and for office processing. Only information known to the respondents, and which they are prepared to give, should be sought. Good questionnaires are brief: if there are many questions, the quality of information suffers.

STAFF SELECTION, TRAINING AND SUPERVISION

The quality of staff must be appropriate to the work. All surveyors should know exactly what is expected of them and why. A training period will usually be essential. Good supervision is also needed to ensure that difficulties are resolved and a good standard of work maintained.

ESTABLISHING CONTACT—PUBLIC RELATIONS

Success depends on the collaboration of farmers and local officials; every effort must therefore be made to secure their trust and interest. This can sometimes be helped by advance publicity or by a letter of introduction; much also depends on the method of approach and personality of the surveyor.

DATA PROCESSING AND REPORTING

Before a survey is begun, a plan should be made for processing the data and ensuring the necessary clerical and computing facilities. Coding and computing should be carefully checked, because errors at this stage can invalidate much useful work.

Consideration should be given to reporting of the results back to local officials and participating farmers by regional reports or discussion meetings.

PILOT SURVEY AND PRE-TEST

A pilot survey, under similar conditions to those of the main survey, is essential, to check all aspects of the plan and obtain realistic estimates of the time and resources needed. The pilot survey should be done soon enough for necessary modifications and for retraining.

For small surveys (50 or so farms to be visited) a formal pilot survey is impracticable, but methods should be pre-tested on a few farms during training.

3. Crop-loss estimation surveys

We are mainly concerned in this Manual with methods for obtaining reliable estimates of the total or preventable crop-loss caused by specific pests.* However, three types of sample survey may, under appropriate circumstances, make a useful contribution to work on crop-loss:

- (a) General appraisal of the crop-loss situation.
- (b) Surveys of farm practice.
- (c) Reliable estimation of losses caused by a particular pest.

GENERAL APPRAISAL OF CROP-LOSSES

Some assessment of total crop-losses caused by different pests may be needed as a guide to policy. This would take account of any existing detailed work on specific pests, but might also call for indicative information on other possible factors about which little is known. Information for such an assessment may be obtained from *ad hoc* tours of a region by appropriate agricultural specialists. Although such tours can be valuable, by identifying some key factor, this is speculative and they will not be discussed further here.

More closely organized attempts at a general appraisal may also be useful, provided their limitations are appreciated. For example, a representative sample of fields might be visited at one or more specified dates or crop development stages and observations made on several types of damage or pest. The people used for such surveys, their training, the precise factors looked for and the methods and timing of observations should be closely defined, and fully described in any report so that reasonable (subjective) judgments can be made about the

* Here and subsequently the word 'pest' is used to indicate any harmful organism (insect, nematode, disease, weed etc.) capable of causing crop-loss.

reliability of the information, both when it is collected and later when there may be more known about the pests.

The value of this kind of survey is limited, because different pests are not equally obvious and the degree to which they are noticed may bear little relation to the damage caused; moreover the causes of obvious damage may not always be correctly identified. Nematodes, for example, may cause widespread damage which is not noticed, or which would be attributed to other factors (drought or nutrient deficiency) by all but the most skilled observers. Because objections of this kind can never be fully met, there is no advantage in doing these surveys on a large scale, as this can only give a spurious impression of reliability.

Results from a survey of foliar diseases of spring barley, based on a random sample of fields, in England and Wales in 1967 (6) illustrate the value and limitations of information from this kind of survey:

	% lamina affected
Mildew	11.0 ± 0.6
Brown rust	4.3 ± 0.4
Leaf blotch	1.7 ± 0.2
Yellow rust	1.0 ± 0.2
Halo spot	0.1 ± 0.05

The figures given are estimated percentages of leaf area affected by the diseases, based on examination of a sample of the two top leaves at growth stage 11.1.

This survey was done after much experimental work on both mildew and leaf blotch, and was repeated in later years. However, supposing the 1967 survey results were the only evidence, one might perhaps infer that mildew was the most serious foliar disease and could be given priority in future work. This would be a provisional opinion based on assumptions (i) that the diseases were correctly identified, and symptoms of individual diseases similarly assessed, (ii) that damage caused by a disease was roughly related to the leaf area affected *at the time of the survey*, and (iii) that the survey year was typical.

SURVEYS OF FARM PRACTICE

Surveys of farm practice can give background information that helps to assess a crop-loss situation. For example, a survey might show how the (known) total consumption of pesticides is divided between crops and regions, and whether dosage, methods and times of application conform with recommendations. Reliable information about varieties grown, planting and harvesting dates, cultivations and crop rotations, etc., is also useful if these factors are likely to affect pest incidence. Such information can help to direct work to areas and topics offering the best 'opportunities for gain', and guide the planning of further experiments so that they are relevant to current conditions. Problems in adopting recommended practices may also be discovered, leading to the revision of current advice. Repeat surveys can show how fast practices are improving.

ESTIMATION OF CROP-LOSSES CAUSED BY A PEST

We now turn to a fuller discussion of surveys to provide reliable estimates of the crop-losses caused by a pest. The aims of these surveys range from estimating the total loss caused (mainly useful in indicating the magnitude of a problem) to providing detailed information on the proportions of crops suffering different amounts of damage and their regional distribution, which may be needed for an economic assessment of possible control measures.

This section discusses the observations to be made at individual survey sites and the sort of accuracy to be attempted in crop-loss surveys; implications for the number of survey sites are discussed in Section 4.

When readily identifiable damage occurs soon before harvest (e.g. bird damage to ripening fruit), so that there is little opportunity for this damage to cause other harmful effects, or to be compensated by the increased growth of neighbouring healthy plants, it may be possible to estimate losses at a site directly in absolute terms or as a proportion of the total crop. More often some measure of pest incidence will be obtained, either as a rough indication of the

degree of damage, or to be used with a relationship between pest incidence and crop-loss established by experiment.

Damage caused by a pest should not be estimated by comparing yields of damaged crops with those on fields where the pest does not occur, because the pest may be associated with other pests and influenced, for example, by soil type and variety, the use of pesticides, or other farm practices that influence crop yields directly. This difficulty cannot be fully resolved by refinements of statistical analysis.

When survey estimates of pest incidence are to be used with the results of experiments to estimate crop-losses, the methods of measuring pest incidence should be the same as those in the experiments. The timing of the observations (referred to well-defined growth stages) is particularly important because, for example, presence of a pest may do little harm over much of the life of a plant, and pest population increases differ greatly between crops and seasons, so that assessments made at one time may bear little relationships to those at another, or to the damage caused.

Observations of pest incidence should be reproducible and as objective as possible, but it is usually less important that they should be closely related to the total pest population or that this relationship should be known.

When the survey results are obtained, it is necessary to consider how well the experiments that related the crop-loss to pest incidence covered the conditions of the survey. In particular, claims that attack by small populations over large areas cause important losses require precise experimentation at these populations, rather than interpolation from estimates of loss caused by large populations.

There are two stages in the measurement of pest incidence at a site

- (i) selection of a sample of plants, parts of plants or soil, etc.,
- (ii) examination of the sample.

The sample is usually made up from several small sampling units (e.g. individual plants, lengths of row or soil auger samples) taken in a rectangular grid pattern or, more commonly, at regular intervals along a diagonal of the field. A sample taken from one position is unlikely to represent the field adequately because pests tend to be aggregated, e.g. round the windward edges, or in more or less discrete patches within a field.

It is important to eliminate any element of personal selection in locating the units. Thus, the instruction might be to select the nearest plant to the right foot after walking a specified number of paces between sample points, and not to reject plants that are badly stunted.

When the estimates at each site are used only to estimate average pest incidence for a whole region, they need not be very accurate. Of course, sampling must be more intensive to get accurate estimates for individual fields, but such estimates are seldom needed, except as an advisory aid for the actual fields sampled. The numbers of sample units required for given accuracy can be estimated using information about sampling errors from previous work, or by taking preliminary samples from a few fields (see Section 4).

Individual soil auger units are bulked, mixed and sub-sampled for examination but, if information on within-field sampling errors is required, alternate units may be bulked separately, and sub-samples of the two bulks from each field examined separately.

With large plants or trees, the selected plants may have to be sub-sampled to decrease the bulk for examination. It is preferable to take a well-defined sample (e.g. top and bottom leaf from each plant) rather than to aim to get a completely random sample.

Selection and training of observers is particularly important when special skills are required (e.g. in identification) or when there is a subjective element in the assessment. For example, if plants are assigned to grades (slight, moderate, severe infestations) by referring to a 'key', group training will not only help to ensure that the 'key' is consistently interpreted but will give observers confidence. Because pest incidence must usually be determined at a particular date or growth stage, assessment may have to be made by local non-specialist staff, or samples may be sent for skilled examination in a central laboratory.

An exact count or measurement is preferable in principle to grading but may take too much time. Approximate counts or grades must be tested for reproducibility, and it may be more reliable, for example, to do exact counts on selected leaves than approximate counts on fewer whole plants.

Counting very large populations is always a problem. When, as frequently with insect pests, damage is related to the average value of $\log(n+1)$, rather than to the average value of an individual count n , proportionate errors matter: 5 is very different from 25 but 1,900 much

the same as 2,000. Large counts can therefore often be satisfactorily estimated by sub-sampling procedures.

4. Sample size and sampling methods

Once acceptable methods of observation and measurement are agreed, the sampling implications of survey objectives determine the amount of work. This aspect of planning is discussed below, with the minimum of technical detail.

Reliable estimates of pest incidence in a region must be based on a representative sample of fields, rather than fields selected from some sub-group that may not be typical. Farmers who are well known to the extension services and whose holdings are near to main roads are likely to be more progressive than average—they may grow better varieties, use more pest control measures and have different cash crop rotations from other farmers and their fields would therefore be a 'biased sample' of all fields. Although the investigator may guess that such differences have little effect on pest incidence, this opinion can be checked only when a representative sample is taken.

SIMPLE RANDOM SAMPLING

For a representative sample, each unit in the population has an equal (or known) chance of being selected. Suppose the incidence of a pest on barley is being studied in a region with 8,000 barley fields. In principle, each field might be listed and numbered serially; then (say) 100 numbers between 1 and 8,000 might be chosen using a random number table, and the corresponding fields visited. There are usually practical objections to simple random sampling, but it is convenient to discuss sample size in relation to a random sample before describing preferable methods.

SAMPLE SIZE

If the standard deviation of estimated pest incidence is \pm (S.D.) and the average is to be estimated with standard error \pm (S.E.), the number of independent random samples required is

$$n = (\text{S.D.})^2 / (\text{S.E.})^2 \quad (1)^*$$

Often the standard deviation is unknown and must be guessed, using information for comparable pests and situations. When there is no such information, a standard deviation $\pm 100\%$ of the average pest incidence might be assumed until there is better evidence.

The standard error of the proportion of fields infested, estimated from a random sample of n fields, is

$$\pm \sqrt{p(1-p)/n} \quad (2)$$

where p is the actual proportion infested.

As greater accuracy is required, the necessary sample size increases rapidly, so it is important to make realistic decisions about the degree of accuracy which is useful. In particular, if there are likely to be large year-to-year variations in pest incidence, it will be preferable to obtain approximate estimates for several years rather than an accurate estimate for one year only.

For this and other reasons, it will often suffice to obtain estimates with standard errors of about $\pm 10\%$, requiring (with a standard deviation of $\pm 100\%$) a random sample of 100 or so fields. When the only interest is an estimate of total crop-loss, the minimum acceptable amount of sampling at each site is likely to be small because, with a standard deviation of $\pm 100\%$ between fields, incidence at each site need only be estimated with a standard error of, say, $\pm 30\%$.† However, more than the minimum within-field sampling might be done when the cost is small compared with the cost of getting to the field.

More sample fields may be needed to estimate the proportion of the total crop area with some minimum pest incidence. For example, if about 5% of a crop is thought to need pest control, and this percentage is to be estimated with standard error ± 1 , at least $5 \times 95/1^2 = 475$ fields must be sampled. A sequential system of sampling within fields might ensure that fields close to the critical level of infestation were sampled more intensively.

* The proportion of all units in the sample is assumed small. The number of sample units needed is then *independent* of the total number in the population.

† Giving $100 \times 100^2 / (100^2 + 30^2) = 92\%$ of the maximum information about average pest incidence from a selected field.

Suppose now that a comparison of pest incidence on all varieties grown on 10% or more of fields is required. To estimate the average incidence on each variety with a standard error of $\pm 10\%$,* a random sample of about $100/0.10 = 1,000$ fields is needed, and even this would not be enough for valid comparisons if the varieties were not similarly distributed over the region and similarly treated. This shows that even the simplest type of comparative work greatly increases the necessary sample size, and that a large sample may not guarantee useful comparisons. Observations on series of controlled experiments on growers fields are usually a more reliable method of obtaining comparative information.

ACCURACY OF SURVEY AND EXPERIMENT ESTIMATES

Even when much experimental work has been done, the relationship established between pest incidence and crop-loss will usually be subject to large errors. Accurate estimates of loss cannot therefore be obtained even when average pest incidence is known exactly. For example, from the results of 25 replicated experiments covering four years, Large and Doling (2) estimated the percentage yield losses of Proctor barley caused by mildew as

$$L = (2.7 \pm 0.5) \sqrt{M} \quad (3)$$

where M is the estimated percentage of leaf area affected (top 4 leaves at growth stage 10.5). If the between-field standard deviation of \sqrt{M} is $\pm \sqrt{M}$ (i.e. $\pm 100\%$), the variance of the percentage loss estimated from pest incidence on a random sample of n -fields is

$$2.7^2 \left(\frac{M}{n} \right) + (\sqrt{M})^2 0.5^2 = M \left(\frac{7.3}{n} + 0.25 \right) \quad (4)$$

A random sample, for example, of 120 fields would estimate the average value of \sqrt{M} with standard error $\pm 0.09 \sqrt{M}$ and the percentage crop loss as

$$2.7 \sqrt{M} \pm 0.53 \sqrt{M}$$

Although a more accurate estimate may be preferred, little can be gained by sampling more fields, because the indicated standard error for percentage crop-loss estimated from equation (3) would not be less than $\pm 0.5 \sqrt{M}$ however large the sample. For greater accuracy it would be necessary to do more experimental work.

However, it may be sensible to accept crop-loss estimates with standard errors of the order of $\pm 20\%$ because of possible unknown biases:†

- (i) few crop-loss experimental techniques are free from all bias;
- (ii) average sampling conditions (particularly timing) will not be exactly the same for survey and experiments;
- (iii) there is usually too little evidence to establish the form of the relationship (e.g. equation (3)) beyond question.

As estimates of average pest incidence do not need to be very accurate to be useful, it may sometimes be unrealistic to insist on a sampling method free from bias if this adds greatly to the survey cost. However, before adopting a method that may be biased, it needs to be considered how far the possibility of bias may affect the value of information obtained. Equally, if a sample must inevitably be biased, it is wasteful to do so much work that the random sampling errors are trivial compared with likely biases.

STRATIFIED SAMPLING

Suppose that some measure of pest incidence has a standard deviation of $\pm 100\%$ for a random sample of fields within a region. If the average pest incidence is to be estimated with a standard error $\pm 10\%$ $(100/10)^2 = 100$ random fields should be sampled and, if the region is made up of four districts as in Table 1, the expected numbers of sampled fields in these districts would be as given for Sample I in the table. (The actual distribution of a random sample between the districts would of course be somewhat different.)

Now, if the districts have different average pest incidence, the standard deviation of pest incidence within districts will be less than for the region as a whole; suppose it is known to be

* Making a difference of 30% detectable with some confidence.

† A bias is an error in an estimate not affected by sample size. Biases may be caused by defective sampling or inappropriate estimation procedures.

Table 1. Comparison between simple random and stratified random samples

District	Total no. of fields	I Simple random sample Sampling Expected		II Stratified sample Sampling Expected		III Stratified sample with variable sampling fraction Sampling Expected		Estd. average pest incidence
		fraction	number	fraction	number	fraction	number	
A	800		10	1/133	6	1/28	28	3
B	1,600		20	1/133	13	1/57	28	6
C	1,600		20	1/133	13	1/57	28	13
D	4,000		50	1/133	32	1/142	28	16
	8,000	1/80	100	1/133	64		112	

$\pm 80\%$. Then a *stratified random sample* (with districts as strata) of $(80/10)^2 = 64$ fields, sampled at random within strata as shown for Sample II in the table, will give an estimate of the required accuracy for the region. In this example, stratification decreases the required sample size by 36% (or increases the accuracy obtained for given effort by 56%). The full advantages of stratification can be obtained only when the proportions of the population in the strata are known, and the sample size should be less than that indicated for a simple random sample only when the standard deviation within strata is known to be less than that for the region as a whole.

Suppose now that estimates are needed for each district with a standard error of $\pm 15\%$; then about $(80/15)^2 = 28$ fields should be sampled at random within each district, giving a total sample of 112 fields quite differently distributed between districts (Sample III). Sample III is described as a *stratified random sample with variable sampling fraction*. With such a sample, it is important to note, for example, that if the estimated average pest incidence in the 4 strata are as given in the last column of Table 1, the estimated average for the whole area is

$$(10 \times 3 + 20 \times 6 + 20 \times 13 + 50 \times 16)/100 = 12.1$$

not

$$28(3 + 6 + 13 + 16)/112 = 9.5$$

The standard error of this estimate is

$$\pm 80 \sqrt{\left\{ \left(\frac{10^2}{28} + \frac{20^2}{28} + \frac{20^2}{28} + \frac{50^2}{28} \right) / 100^2 \right\}} = \pm 8.8\%$$

or

$$\pm 12.1 \times 8.8/100 = \pm 1.1$$

Equal variability within the strata is assumed above. If the standard deviations for the individual strata differ and are known, or if estimates of different accuracy are needed for different strata, the sampling fractions of Sample III may be modified accordingly.

MULTI-STAGE SAMPLING

It is usually impracticable to take a simple random or stratified random sample, because such a sample would be so widely scattered that it would be costly to visit each holding. Moreover complete lists of holdings or of fields growing a particular crop, or large-scale maps giving equivalent information, are not often available.

To overcome these difficulties a *two-stage or multi-stage* sampling scheme should be adopted. For example, there may be a complete, or almost complete, list of villages. Instead of sampling 100 fields at random, 10 villages could be selected at random—or systematically from the list. In each selected village a list of holdings would be prepared and (say) 12 holdings selected systematically from the list; on each selected holding, one field would be chosen at random. Alternatively, a system might be defined for making a traverse of the village land

and selecting at regular intervals (e.g. every kilometre) the nearest field under the required crop.

When information is available for a random sample of n' villages, within each of which n'' fields have been sampled, the analysis of variance of the field records takes the form:

	d.f.	m.s.
Between villages	$n' - 1$	$n''s_2^2 + s_1^2$
Within villages	$n' (n'' - 1)$	s_1^2
Total	$n' (n'' - 1)$	

If the cost of visiting a sample village is c' and the extra cost of sampling each field is c'' , the optimum number of fields to be sampled per village* is estimated as

$$n = \frac{s_1}{s_2} \sqrt{\frac{c'}{c''}}$$

The total number of sampled fields should be increased by a factor $(n s_2^2 + s_1^2)/(s_1^2 + s_2^2)$ compared with that indicated for a simple random sample. The calculated optimum is not critical; moreover, s_1^2 and s_2^2 , are often poorly estimated, and discontinuities in work effort affect the actual costs. It is therefore sensible to decide on n so as to do one or more full days of work in each selected village.

When working with mobile teams, typical values may be about 6 for c'/c'' and about 1 for s_1/s_2 , and the required planning decision is typically whether to work for one or for two days at each centre.

In the above example, villages were selected with equal probability, and a fixed number of fields were sampled in each selected village. Obviously the sample gives more than proportionate representation to villages where little of the crop is grown. Possible bias might be avoided, if the crop areas were known, (a) by selecting the villages with probability proportional to crop area, or (b) by weighting the results for each village by crop area, or (c) by sampling numbers of fields in proportion to the village crop area (usually operationally inconvenient). If the crop areas are not known, bias may be decreased by sampling separately those districts where the crop is intensively grown; as indicated earlier, this stratification may be advantageous for other reasons.

REFERENCES

1. Hansen, M. H., Hurwitz, W. N., and Madow, W. G. 1953. *Sample survey methods and theory*. Wiley, New York. 332 pp.
2. Large, E. C., and Doling, D. A. 1962. The measurement of cereal mildew and its effect on yield. *Plant Pathol.* 11:47-57.
3. Sukhatme, P. V. 1954. *Sampling theory of surveys with applications*. Iowa State College Press, 1st ed., 491 pp.
4. Yates, F. 1960. *Sampling methods for censuses and surveys*. Griffin, London, 3rd ed., 440 pp.
5. Zarkovich, S. S. 1966. *Quality of statistical data*. FAO, Rome.
6. James, W. C. 1969. A survey of foliar diseases of spring barley in England and Wales in 1967. *Ann. appl. Biol.* 63:253-263.

* i.e. the number giving most information for unit cost.