Sampling Techniques for Surveys of Rubber Smallholdings I. Estimation of the Number of Trees in Tapping

COLIN BARLOW, LIM SOW CHING and P. O. THOMAS

Two sampling methods for estimating the number of trees in tapping on rubber smallholdings are discussed. They involve counting the number of trees in a series of measured sampling areas, and are less time-consuming than a complete count. One of these methods is best used in systematically planted holdings while the other is applicable to holdings with random growth and no regular planting system. Limits of accuracy can be established for the estimates obtained using both these methods, and the numbers of samples necessary to obtain estimates within the given limits are quoted for holdings of different acreages.

Lack of recorded data about individual rubber smallholdings in Malaya means that surveys designed to collect information of agro-economic interest have to rely mainly on estimates. Such estimates can often be improved by using sampling methods specially developed for smallholding conditions. Several methods have recently been evolved by the R.R.I.M., and this paper describes one designed for estimating the number of rubber trees in tapping per acre. Techniques which can be used to secure estimates of dry rubber yield from individual smallholdings will be outlined in a later paper in this series.

One way of determining the number of trees in tapping on a smallholding is to make a complete count. This is, however, far more laborious and time-consuming than taking a sample. Besides, keeping an accurate check on trees already counted is difficult. Sometimes it is also hard to establish the exact boundary of a holding, even though the precise acreage is known from land titles. Under most circumstances, a sampling technique is definitely advantageous provided the required accuracy can be secured.

Experience has shown that, for sampling purposes, rubber smallholdings are best classified in the following categories:

(a) *Holdings with a 'systematic' planting scheme* where trees are planted in distinct and orderly rows, either straight or along the contour; and

(b) *Holdings where tree growth is 'random'* without any definite planting system, or where self-sown trees with a girth of 8 inches or more, at 20 inches from the ground, are growing in between the planting rows.

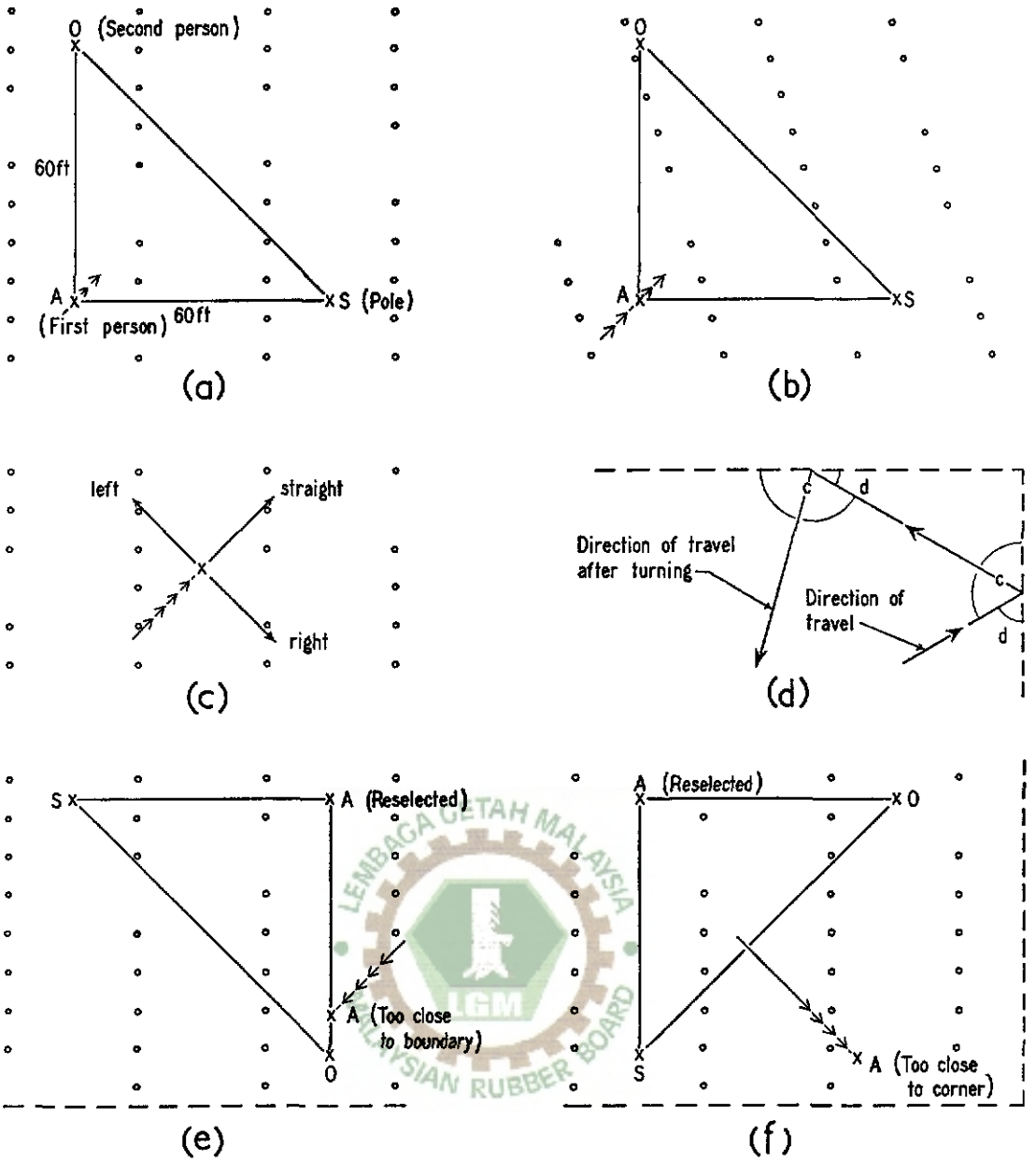
Smallholdings of both categories in Selangor were studied to determine the best type of sampling unit and the number of such units in a reliable sample. A procedure was also developed for selecting the location of each sampling unit. The results of this study are now discussed.

HOLDINGS WITH SYSTEMATIC PLANTING

Possible Types of Sampling Unit

During initial field trials with differently-shaped sampling units it became apparent that, where only two persons were available, the only manageable shape was a triangle or a combination of triangles. Other shapes did not enable adequate measurement. It was therefore decided to compare the suitability of two sampling units: a right-angled triangle, 200 sq. yards in area, with shorter sides each 60 ft long

COMMUNICATION 401



OS = Long side of triangular area across original direction of travel. Pole at S.
>>>> = Original direction of travel in proceeding to position A.
o = Rubber tree.
— — — = Boundary.

Figure 1. Selecting and measuring triangular sampling units.

(*Figure 1*); a square, 400 sq. yards in area, formed by four right-angled triangles with shorter sides each 43 ft long (*Figure 2*). The comparison was made by assessing the variability of the number of trees in tapping and judging how accurately the mean for each type of sampling unit predicted the actual average number of trees per unit. Variability within a smallholding can be ascribed not only to thinning, root disease and wind damage, but also to the fact that a holding may consist of two or more lots, each with a different planting arrangement.

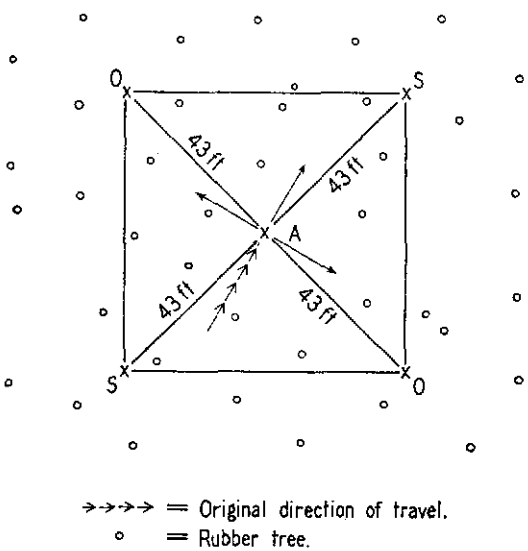


Figure 2. Selecting and measuring a square sampling unit.

The number of trees in tapping was therefore recorded for repeated samples of triangular as well as square units on a selected five-acre holding with a systematic planting system. The method of locating the units so as to produce an approximately random sample is described in the *Appendix*. A complete count of the trees in tapping was also made.

Data from this recording are presented in *Table 1*; it can be seen that the coefficient of variation of tree numbers is somewhat greater for the triangular units than for the square units (although this difference is nowhere near significant at the 5% probability level). On the other hand, the sample bias of estimates for triangular units is considerably lower. Records of the number of trees in tapping, using repeated triangular unit samples, were also made on other systematically planted holdings; a similar coefficient of variation with a bias in the range ±3% was obtained. Data for square units were not collected on any other holdings.

Number of Units in the Sample

The number of sampling units of a given type to be included in a sample depends on the degree of accuracy required in estimating the number of trees in tapping. It is considered that a resonable standard here is that the estimated number of trees in tapping should not be in error by more than 10% of the true value, excluding the possibility of a 1 in 20 chance. The latter reservation is necessary, otherwise

TABLE 1. COMPARISON OF TRIANGULAR AND SQUARE SAMPLING UNITS[a] ON A SYSTEMATICALLY PLANTED HOLDING

Type of unit	Mean no. of trees in tapping per unit		Coefficient of variation[b] (sample), %	Sample bias %
	sample	actual		
Triangle	7·0	6.8	16.5	+2.9
Square	13.0	13.6	14.0	−4.6

[a] Based on 30 triangular sampling units, 24 square sampling units, and a complete tree count on the selected 5-acre smallholding.

[b] See expression (4).

the required number of samples becomes unmanageably large.

Accuracy of estimation is conveniently expressed by the standard error of the mean of the sample observations, $S_{\bar{x}}$. This measure can be calculated by using the expression.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N}\right)} \qquad \text{.....} \quad (1)$$

where

S is the standard deviation of individual sample observations,

N is the total number of sampling units required to make up the full area of the holding being surveyed,

and n is the number of sampling units in the particular sample concerned.

A detailed explanation of these and other statistical terms used in this paper can be found in the work of COCHRAN (1953).

It will be noted that the limit of accuracy quoted above incorporates a probability statement, which can be linked to the standard error by the expression.

$$S_{\bar{x}} = \frac{d}{t} \qquad \text{.....} \quad (2)$$

where d is the stipulated maximum difference between the actual and the estimated number of trees in tapping, which should not be exceeded in more than a given number of cases, and t is a constant which depends on the number of samples and validates the latter probability statement. Substituting (2) in (1), the number of units required in a sample to satisfy a certain limit of accuracy can be expressed as

$$n = \frac{\left(\frac{tS}{d}\right)}{1+\frac{1}{N}\left(\frac{tS}{d}\right)^2} \qquad \text{.....} \quad (3)$$

A final expression, which will be of use in the calculations below, is

$$C = \frac{S}{\bar{x}} \times 100 \qquad \text{.....} (4)$$

where C is the coefficient of variation and $\bar{x}$ is the sample mean.

The number of triangular sampling units required to meet the quoted limit of accuracy can now be calculated using (3). It is considered reasonable to assume a value of 20% for C in this calculation. Although the coefficient of variation shown in *Table 1* is only 16.5%, experience with repeated samples using triangular units has also shown that small groups of samples may have values of 20% or even more for C. It is also considered that 6.8 is an appropriate mean number of trees in tapping, $\bar{x}$, to assume for each triangular sampling unit. This makes up to 165 trees in tapping per acre. Although the stand may often be less, it is necessary to allow for high numbers to preserve the limit of accuracy in most cases.

Re-arranging (4) to express S, the standard deviation of individual observation, in terms of the other variables, and substituting 20 for C and 6.8 for $\bar{x}$, a value for S of

$$\pm(6.8)(0.2) = \pm 1.36$$

is obtained. The value of d, assuming that the estimated number of trees in tapping should not be in error by more than 10% of the true value, is

$$\pm(0.1)(6.8) = \pm 0.68$$

and the appropriate value of t to hold good the 1 in 20 probability statement above is approximately 2. The value of N, the total number of sampling units to make up the full area of the holding, is 24.2 for a 200 sq. yard triangular sampling unit on a 1-acre holding, 48.4 on a 2-acre holding, etc.

Substituting these in (3), values of n for triangular sampling units (*Table 2*) are obtained.

With respect to appropriate values of n for 400 sq. yard square sampling unit, it is realistic enough to assume again that C is 20%. The expected value of S, taking the *Table 1* figure of 13.6 trees in tapping per 400 sq. yards (165 trees per acre) is thus

$$\pm(13.6)(0.2) = \pm 2.72$$

The appropriate value of d is

$$\pm(0.1)(13.6) = \pm 1.36$$

TABLE 2. NUMBER OF TRIANGULAR AND SQUARE SAMPLING UNITS REQUIRED ON SYSTEMATICALLY PLANTED HOLDINGS[a]

Area of smallholding, acres	Number of sampling units required	
	Triangular	Square
1	10	7
2	12	10
3	13	11
4	14	12
5	14	12

[a] To obtain an estimate with an error not greater than 10% of the true value, excluding the possibility of a 1 in 20 chance.

and t is again 2. The value of N is 12.1 for a 400 sq. unit on a 1-acre holding, 24.2 on a 2-acre holding, etc. If these figures are substituted in (3), values of n for square sampling units (also in *Table 2*) are obtained.

Comparison in *Table 2* of the relative numbers of triangular and square units required for various smallholding acreages denotes that slightly more triangular units are required to obtain the same level of accuracy. This disadvantage, however, is entirely offset by the fact that triangular units take only about half the time of square units to measure. In addition, estimates obtained using triangular units are less biased (*Table 1*). The triangular sampling units, using the numbers given in *Table 2*, are thus the most suitable.

One may naturally ask whether a sampling unit smaller than 200 sq. yards may be even more suitable. Experience shows, however, that smaller triangular areas are unsatisfactory in systematic plantings, because smaller triangles may fall in interrow spaces, resulting in trees being missed out entirely in the count. In such cases, the coefficient of variation would become much larger, requiring a greater number of samples to yield the required limit of accuracy.

Selecting the Location of Sampling Units

The best method of selecting the location of sampling units is that which consistently provides the most accurate estimate of the number of trees in tapping, giving consideration to the ease of operation. A description of the best method, selected after a period of trial, for locating triangular sampling units in systematically planted holdings is given in section (1) of the *Appendix*. This method has been used satisfactorily in 700 smallholdings in Malaya.

HOLDINGS WITH RANDOM TREE GROWTH

Possible Types of Sampling Unit

The 200 sq. yard triangular and the 400 sq. yard square sampling units again appeared to be the only types meriting serious comparison under conditions of random growth. They were hence compared by taking repeated samples on a selected $3\frac{3}{8}$-acre holding (*Table 3*).

The coefficient of variation of tree numbers in the triangular units is much higher than in the square units, and the percentage bias of the former is also higher. Data from sampling with square units on other holdings with random tree growth generally indicated a coefficient of variation about 10% less than that given in *Table 3*. Sampling with triangular units was not carried out on holdings other than the $3\frac{3}{8}$-acre lot.

Number of Units in the Sample

Table 3 shows that the coefficient of variation of tree numbers is much higher on holdings with random growth than on systematically planted holdings. It is hence considered appropriate to reduce the required accuracy of prediction below the limit used in the case of holdings with a systematic planting system.

TABLE 3. COMPARISON OF TRIANGULAR AND SQUARE SAMPLING UNITS[a] ON A HOLDING WITH RANDOM GROWTH

Type of unit	Mean no. of trees in tapping per unit		Coefficient of variation (sample), %	Sample bias, %
	sample	actual		
Triangle	5.7	7.7	74.6	−26.0
Square	14.1	15.3	58.5	− 7.9

[a] Based on 24 triangular sampling units, 30 square sampling units, and a complete tree count on the selected 3⅝-acre holding.

If not, the number of units to be included in the sample would become very large. The limit now taken is that the estimated number of trees in tapping should not be in error by more than 20% of the true value, excluding the possibility of a 1 in 20 chance. Users of these sampling systems may adopt other limits to suit their special requirements.

In using (3) to work out the appropriate values of n for the 200 sq. yard sampling unit, it is considered realistic to assume a value of 80% for the coefficient of variation C. The expected value of S, taking the *Table 3* figure of 7.7 trees in tapping per triangular unit (185 trees per acre) and using (4) re-arranged as before, is thus ± 6.16. The appropriate value of d is ± 1.54 and t is again 2. With respect to the 400 sq. yard units a value of 60% is assumed for C. Although this is very little above the figure of 58.5% quoted in *Table 3*, it has already been mentioned that a generally lower coefficient of variation was found on other holdings. The expected values of S and d for 400 sq. yard units are ± 9.18 and ± 3.06 respectively. Substituting these various figures in (3), the values of n given in *Table 4* are obtained for the two types of sampling unit on different smallholding acreages.

Table 4 shows that for the acreage groups given the number of triangular sampling units required is almost double the number of square units. The measurement of triangular units takes only about half the time required for measuring square units. However, the difficulty of moving from one sample location to another in holdings with random growth, which are frequently not maintained at all, gives the square units a distinct advantage.

TABLE 4. NUMBER OF TRIANGULAR AND SQUARE SAMPLING UNITS REQUIRED ON HOLDINGS WITH RANDOM GROWTH[b]

Area of smallholding, acres	Number of sampling units required	
	Triangular	Square
1	18	9
2	28	14
3	34	18
4	39	21
5	42	23

[b] To obtain an estimate with an error not greater than 20% of the true value, excluding the possibility of a 1 in 20 chance.

TABLE 5. NUMBER OF TRIANGULAR AND SQUARE SAMPLING UNITS REQUIRED TO ESTIMATE THE NUMBER OF TREES IN TAPPING

Area of smallholding, acres	Number of sampling units required	
	Triangular on systematically planted holdings[a]	Square on holdings with random growth[b]
0.5	7	5
1.0	10	9
1.5	11	12
2.0	12	14
2.5	13	16
3.0	13	18
3.5	14	19
4.0	14	21
4.5	14	22
5.0	14	23
5.5	14	23
6.0	14	24
6.5	15	25
7.0 and over	15	25

[a] To obtain an estimate with an error not greater than 10% of the true value, excluding the possibility of a 1 in 20 chance.

[b] To obtain an estimate with an error not greater than 20% of the true value, excluding the possibility of a 1 in 20 chance.

The estimate obtained using the square unit is very much less biased (*Table 3*). It is hence felt that the 400 sq. yard square sampling unit is more suitable for random growth holdings. Although larger units may be even more appropriate, they cannot be measured under random growth conditions because of sighting difficulties.

Selecting the Location of Sampling Units

The procedure for selecting the location of 400 sq. yard square sampling units on holdings with random tree growth is described in section (2) of the *Appendix*. This method has been used in 500 smallholdings.

DETAILED SAMPLE REQUIREMENTS

More detailed figures of the number of sampling units required to estimate the average number of trees in tapping per acre, according to the limits of accuracy laid down, are presented in *Table 5*. It can be seen that the number of sampling units required does not increase any more for holdings of over 7 acres.

Although the number of dry trees and those not yet in tapping can also be counted within each sampling unit, the average estimates obtained will not meet the limits of accuracy for trees in tapping. This is because the occurrence of both dry and immature trees is normally far more variable, requiring larger numbers of sampling units to achieve the same accuracy.

ACKNOWLEDGEMENT

The authors wish to acknowledge the help of Enche Abdul Rahman bin Ismail, Enche Wahab bin Mohd. Satir, Enche Leong Swee Chan, and Tuan Haji Mahfutz bin Haji Elias, all of the R.R.I.M. Smallholders' Advisory Service, who assisted in recording field data. Grateful thanks are also due to the smallholders on whose land the work was carried out.

REFERENCE

COCHRAN, W.G. (1953) *Sampling Techniques*. p. 55. Bombay: Asia Publishing House.

APPENDIX

(1) *Locating triangular sampling units in systematic holdings*

Two operators, equipped with a 60 ft measuring tape, a specially marked dice, a 6 ft pole with a white flag on one end, and a clip-board with recording sheet, are required for this method.

The boundaries of the holding are first established. Where the boundary is difficult to distinguish readily, it is marked with sticks planted at intervals.

The two operators proceed to the approximate centre of the holding and locate the mid-point of the rectangle made by the nearest four trees. This point, shown as *A* in *Figure 1(a)*, is the vertex point of the 200 sq. yard right-angled triangle required for measurement. The initial direction of travel should never be parallel to any of the rows of trees. If it turns out by chance that the direction taken towards the centre is, in fact, parallel to a row, it should be changed slightly.

Figure 1(a) shows the general lay-out of the triangular area, with the 'long side' OS always being across the original direction of travel in proceeding to *A*. The other two sides of the triangle, AO and AS, are, as far as possible, parallel to the rows of trees.

To measure the triangular area, one of the operators (referred to as the first person) remains at *A*, holding one end of the tape. The second person walks from *A* along a course parallel to the rows of trees, until he reaches *S*, 60 ft away, where he plants the pole. He similarly fixes the position of *0*, also 60 ft away from *A*. Where the rows of trees in the stand are at right-angles to one another, the right angle of the required triangle can usually be decided easily. In some stands of rubber, however, as illustrated in *Figure 1(b)*, it may only be possible for one of the shorter sides of the triangle to be parallel to the nearest rows of trees.

From his position at *0*, the second person recognises the lines OA and OS, and moves slowly back to *A* counting the number of trees in tapping within the triangular area. The number is duly recorded by the first person.

When the recording of one triangular area is completed, the first person tosses the dice to determine the direction of travel to the next vertex point. (The dice may have, for example, two faces coloured red, two blue and two black. This will enable a choice of three possible directions in relation to the colour of the top face obtained by the toss). These possible directions of travel—right, straight ahead and left —from *A* are shown in *Figure 1(c)*.

If the holding is 5 acres or less, the first person moves 100 paces (*D*) from *A* in the direction indicated by the dice tossed to arrive at the next vertex point. If the holding is larger, a good spread of the sampling units over the whole area may be obtained by determining the number of paces, (*D*), through using the equation $D=100+5x$, where x is the area in acres.

If the boundary is encountered before the required paces are completed, the first person turns along a line which bisects the angle made by his original direction of approach and the boundary as in *Figure 1(d)*. It should be remembered that the angle bisected is the larger angle c and not the smaller angle d. After completing the required number of paces, another vertex point *A* is reached. Then, the procedure for sampling units in the triangle area is repeated.

Sometimes the new vertex point may be too close to the boundary, or may even be the corner of a smallholding; it is then necessary to 'retreat' from the boundary along the same line until reaching another point (reselected *A*) from which a triangle can be formed. The procedure involved is illustrated in *Figure 1(e)* and (*f*). When the measurement of this triangle is completed, the direction then taken from the

vertex point should still be the same as that originally followed in reaching the point from which the retreat was made.

(2) *Locating square sampling units in holdings with random growth*

Position *A* is determined in the same manner as that described above, except that it is not possible in random growth holdings to take the mid-point between four trees. An illustration of the procedure is given in *Figure 2*, which shows that *A* is the centre of a square formed by four adjacent right-angled triangles, whose smaller arms are each 43 ft. long. These triangles are measured according to the same procedure followed in the case of the 200 sq. yard units.

There is no question of the sides of the triangle being parallel to any row of trees, and the second person can proceed in any direction he chooses to measure the first triangle. The original direction of approach to *A* must, however, be used in determining the direction in which to proceed to the next location. It will be apparent, from *Figure 2*, that the same position *S* can be used for determining the location of two adjacent triangles.

Economics and Planning Division
Rubber Research Institute of Malaya
Kuala Lumpur *July 1966*