

## THE THEORY OF AMOEBIC SURVEYS

BY H. O. LANCASTER

*School of Public Health and Tropical Medicine, Sydney, Australia*

(With 1 Figure in the Text)

1. *Introductory.* The problem of the interpretation of amoebic surveys is so important that it seems worth while to discuss it in full. It is proposed, therefore, to set out here the theory, but not to review the literature systematically.

The problem may be expressed in terms of a familiar analogy: Suppose we are given a large number of urns in which there are black and/or white balls in unknown proportions; can we, by drawing a single ball from each and replacing it, make any useful statements about the total number of urns which contain at least one black ball? Further, how many times do we need to draw in order to obtain a sufficiently accurate answer, and can we assign any adequate upper or lower limits to the number of urns containing black balls? There are numerous other important applications of this problem. Perhaps the commonest is in the interpretation of pathological tests used to determine whether an individual patient is likely to carry a certain parasite which our techniques are unable infallibly to detect in a single examination. The related problem of surveys is to estimate, when we have done one or more examinations on each, the true proportion of persons in a population who carry the parasite. A somewhat similar problem arises in demography in the study of fecundity.

Here only the special case of amoebic surveys is studied in detail. A measure, 'the demonstrability', which has some frequency distribution is defined. Certain consequences are quite independent of the form of this frequency distribution, and from them we derive an inequality and a criterion for the adequacy of the lower limit assigned to the carrier rate in the populations studied. It is shown by life-table methods that most estimates given in the literature are so low as to be inconsistent with the data on which they are founded. Mathematical models of the problem are set up, the most convenient being one in which the frequency distribution is of the beta-function form. We have given several examples to show that extrapolation from two or three examinations by means of a logistic or similar curve leads to estimates of the carrier rate that are too low. We have then suggested certain points essential for the proper understanding of the reason why the demonstrability is not always unity by any given technique.

We conclude that carrier rates of different populations can be compared only with the greatest reserve unless the techniques are exhaustive.

2.1. *The problem stated.* For the purpose of this article a 'carrier' is defined as any person who at the time of examination carries the parasite we are interested in (whether it is detected or not), and regardless of whether he has symptoms or not. That is, our term includes both classes covered by the clinical terms 'carrier'

and 'case'. In attempting to estimate the number of carriers in a given population who carry a particular parasite, say the intestinal parasite, *Entamoeba histolytica*, one of the following procedures may be carried out:

(i) A group of  $N$  persons is examined by some laboratory test, usually an examination of the stools by microscopy. Not all carriers are detected at the first examination, but only a number  $X_1$ . These  $X_1$  will not be re-examined, but under favourable conditions the remainder,  $(N - X_1)$  in number, will be, and some number,  $X_2$ , will be found positive at this second examination. This process is continued and we shall have  $X_i$  carriers found positive for the first time at the  $i$ th examination. After an indefinite number of examinations it is assumed that all carriers will have been detected and the total number of carriers will be given by  $X$ , where

$$X = X_1 + X_2 + X_3 + \dots \quad (1)$$

The process assumes that no one originally a carrier ceases to be so during the examination.

(ii) Occasionally the more informative process of examining the whole  $N$  persons a fixed number of times has been adopted; and the number of persons found positive once, twice, thrice and so on recorded. This is most informative if 'controlled' by alternative methods of diagnosis such as culture and examination following a purge, which are more efficient in detecting the organism.

From data collected in this latter way a figure known as the 'demonstrability' can be derived. 'Demonstrability' is here defined as the ratio of the number of positive examinations (made from the same specimen) to the total number of examinations, where the total number of examinations is indefinitely large; that is, it is the probability of a carrier being found positive at a single examination, and we shall denote it by the letter  $p$ . So defined, the demonstrability may be determined sufficiently accurately at one point of time, since an indefinitely large number of examinations can be carried out on any given specimen, the usual examination for amoebae requiring only a very small fraction of the total stool. A number of specimens may be examined over a period of days or over a longer period, and thus the individual's 'total demonstrability' may be defined as the arithmetic mean of the demonstrability measured at each stool. This latter concept is useful theoretically. *It has usually been considered that this demonstrability is characteristic of the parasitic species and constant over the members of the host population, but we shall show that there is heterogeneity even among such groups of persons as asylum patients (Svensson's data). This heterogeneity would be greater in mixed groups containing diverse classes, such as healthy persons, institutional groups, asylum patients and so on.* It will be useful also to define  $p_i$  to be the average demonstrability of those carriers still outstanding, i.e. not yet found positive, after  $i$  examinations. In the various discussions that follow we shall, in general, assume  $N$ , the number of persons, to be so large that sampling errors can be neglected, so that we are only considering expectations.

3.1. *The mean and variance of the demonstrability,  $p$ .* In the early days of microscopy, the possibility of not finding parasites in known carriers was overlooked or

at least ignored. As we should say nowadays, the demonstrability is not unity. But all carriers do not have the same demonstrability,  $p$ , so that we are led to consider also the spread or variance of the  $p$ 's about their mean,  $p$ . It is convenient, therefore, briefly at this stage to consider some well-known work, that of Dobell (1917) and Svensson (1934).

3.2. *The mean demonstrability.* Dobell (1917) was probably the first to study this problem in any of the common fields of microscopic work in pathology. His conclusions for *E. histolytica* are set out on p. 50 of his paper:

(1) Both positive and negative examinations are commonly made on untreated cases infected with *Entamoeba histolytica*—just as in *Lambliia* and *E. coli* infections.

(2) The available data—which are meagre—indicate that the expectation for the average infected case is that it will be found positive twice in every five examinations; or the chances are 3:2 against the infection being detected at any one examination. (There is reason to suppose that 2:1, or an even higher ratio, may be nearer the truth.)

(3) On analogy with the results obtained for *Lambliia* and *E. coli*, it seems probable that in a series examined three times per case, not more than one-half to two-thirds of the infected cases will have been detected.

(4) Since three negative examinations per case has been the arbitrary standard of non-infection usually adopted, the greatest number of examinations made on any case before infection was actually detected is very low; though by analogy it may be expected to be in reality very high. (This is supported by the fact that one case has been found positive only once in 23 examinations.) The number of examinations per case necessary to detect all infected members of a series is therefore unknown, but probably—as with *E. coli* and *Lambliia*—very high.

It must be noted that the selection of cases for repeated examination was rather haphazard. Dobell noticed that the very criterion for selection—namely, at least one positive finding—led to bias if there were but a few examinations made on each case. At a later date, Svensson (1934) tried to overcome this difficulty by an independent method of selection, the results of search after a purge and of culture methods. It seems that a larger scale research along lines, set out by her, would yield valuable information.

3.3. *The heterogeneity of the demonstrability.* Dobell's findings suggested, moreover, that the demonstrability may vary very considerably from person to person. We show some of his results in Table 1, to which we have applied the  $\chi^2$  test for homogeneity.

It is clear that such a result or one more extreme would be expected to occur with excessive rarity on the hypothesis of homogeneity. We must set aside, therefore, the hypothesis of homogeneity, which of course was Dobell's conclusion also. Dobell and other authors, however, have failed to note that such heterogeneity within populations would render comparisons between populations most hazardous, since we could expect even greater heterogeneity in the demonstrability between populations of widely different hygienic habits.

In a survey of an asylum population by means of examination after saline purge, Svensson (1934) was able to find thirteen male carriers and fifteen female carriers who had had at least six examinations and one male who had had five examinations by saline preparations but to whom we have assigned a sixth positive examination (no. 9v75) by means of random numbers.

Table 1. *Examinations of untreated cases, Entamoeba histolytica.*  
(Dobell's Table V)

	No. of examinations made		
	Positive	Negative	Total
	5	17	22
	12	1	13
	8	0	8
	3	17	20
	1	22	23
	3	11	14
	11	4	15
	6	15	21
	22	21	43
Total	71	108	179

$\chi^2 = 59.44$  for 8 D.F.,  $P < 0.001$ .

The data for each of the sexes separately are shown in Tables 2A and 2B.

In Tables 2A and 2B, + indicates that the amoeba has been found at that particular examination, 0 that it has not.

Table 2A. *Data of Svensson (1934). Examinations of known cases of male carriers of Entamoeba histolytica*

Case no.	Results of successive examinations for <i>E. histolytica</i>						Total no. of positive examinations
9v73	0	+	0	0	0	0	1
9v74	0	0	0	+	0	0	1
9v75	+	0	+	0	+	+	4
9v76	0	0	+	0	0	0	1
v237	0	0	0	+	0	0	1
v282	+	+	0	0	+	0	3
v291	0	0	0	0	0	0	0
v304	0	0	0	0	0	0	0
v307	0	0	0	0	0	0	0
v323	0	0	0	0	0	0	0
v346	0	+	+	0	+	+	4
v349	+	+	+	+	+	+	6
v354	0	0	+	0	0	0	1
v381	0	0	0	0	0	+	1
Total	3	4	5	3	4	4	23

\* 14 carriers have been examined, each six times, except 9v75, to whom we have assigned a value to the sixth result by random numbers.

$\chi^2$  for row totals, between individuals = 37.90 for 13 D.F.,  $P < 0.001$ .

$\chi^2$  for column totals, between successive examinations = 1.02 for 5 D.F.,  $P > 0.95$ .

In Table 2A the first  $\chi^2$  value is significantly high, the second is significantly low. The demonstrability of individual carriers is therefore heterogeneous, i.e. the means of the individuals differ from the grand mean of the group by more than sampling errors. Similar results hold for Table 2B.

Table 2B. *Data of Svensson (1934). Examinations of known cases of female carriers of Entamoeba histolytica*

Case no.	Results of successive examinations						Total no. of positive examinations
9v5	0	0	0	0	0	0	0
9v9	0	0	+	0	0	0	1
9v12	0	0	0	0	0	0	0
9v15	+	+	+	+	+	0	5
9v19	0	+	0	+	+	0	3
9v20	0	0	0	0	0	0	0
9v24	0	0	0	0	+	+	2
9v31	0	0	0	0	0	0	0
9v70	+	0	0	+	+	0	3
9v71	0	+	0	+	+	+	4
9v72	+	0	0	0	+	+	3
v12	0	0	0	0	0	0	0
v16	+	+	+	+	+	0	5
v67	0	0	0	+	0	0	1
v107	0	0	+	0	0	0	1
Total	4	4	4	6	7	3	28

$\chi^2$  for rows, between individuals = 37.12 for 14 D.F.,  $P < 0.001$ .

$\chi^2$  for columns, between examinations = 3.53 for 5 D.F.,  $P = 0.60$  (approx.).

If we combine the results for the two sexes, we obtain the following results:

Between individuals, but within sexes,  $\chi^2 = 75.02$  for 27 D.F.,  $P < 0.001$ .

Between examinations, but within sexes,  $\chi^2 = 4.55$  for 10 D.F.,  $P > 0.90$ .

Linders (1934) has assumed that the results shown in Tables 2A and 2B can be treated by supposing that there is a  $p$  common to all the persons of either sex. We may note that the row totals are a Lexian case and the columns a Poissonian case. Since a group of patients of the same sex in an asylum is heterogeneous, we may confidently expect pooling of different groups to result in even greater heterogeneity. It is clear therefore that Svensson & Linders (1934) are justified in believing that their groups are not homogeneous with that of Dobell (1917). This has a bearing on the so-called 'standard error' of the carrier rate, which we shall discuss in a later section.

4.1. *Theoretical considerations.* In this section we follow out the consequences of supposing the demonstrability to have certain properties, starting from the simplest case. In §§ 4.2-4.8 we consider the type of results which would be expected to occur in surveys carried out by method (i) of § 2.1, that is, each carrier is re-examined until the first 'positive' result is obtained. In § 4.9 we consider method (ii) of § 2.1, in which the whole  $N$  persons are repeatedly examined. Then in § 4.10 we discuss several common methods of assessing the mean demonstrability of a group. In § 4.11 is set out a discussion of the standard errors. § 4.12 deals briefly with some uses of the mathematical models set up—chiefly their value in exemplifying the dangers of extrapolation.

4.2. *Hypothesis (i).* A constant demonstrability common to every member of the population. Here every member of the population carries the parasite and has a demonstrability, independent of the time. We easily find from the definition that

$X_1$ , ( $=Np$ ), is the expected number detected at the first examination. There remain  $N(1-p)$  to be examined a second time, so  $X_2$  the expectation of the number detected at the second examination is  $N(1-p)p$ , and in general

$$X_i = N(1-p)^{i-1}p. \quad (2)$$

After the  $i$ th examination there will remain  $N(1-p)^i$  persons, who have not yet yielded a positive result at any examination, but this number rapidly approaches zero, if  $p$  be not very small.

4.3. *Hypothesis (ii). Equal demonstrability of all carriers, but not all the members of the population are carriers.* This is a slight extension of the previous case. Here we assume that only  $X$  are carriers in a total population,  $N$ , and that each of the  $X$  persons has the same demonstrability,  $p$ .  $X$  is less than  $N$ . This assumption of an equal demonstrability was made by Sawitz & Karpinos (1942) in their study of the chances of detecting *Enterobius* by repeated examinations, but they did not realize that their problem can be completely solved if this assumption is made (see also Sawitz & Faust, 1942). A similar assumption was made by Gini (1924) in his study of 'fecondabilit ' of women. It appears, however, from a study of the data given by him that there is definite heterogeneity in the probability of conception per month for women. Moreover, it is unlikely that equal demonstrability ever holds in any actual population. The arithmetic is similar to the previous case since the sub-population of carriers in the second case is equivalent to the whole population of the first. In either of these cases we may estimate  $p$  from

$$1-p = X_i/X_1 - 1. \quad (3)$$

This result may also be derived by the method of maximum likelihood.

4.4. *Hypothesis (iii). There is heterogeneity. Demonstrability is constant in a given person over the time considered but differs from person to person.* This hypothesis assumes that there is some frequency function,  $f(p)$ , such that

$$\int_0^1 f(p) dp = 1. \quad (4)$$

By the definitions  $\bar{p}$ ,  $\bar{p}_i$  are given by equation (2) as follows:

$$\bar{p} = \int_0^1 pf(p) dp, \quad (5)$$

$$\bar{p}_i = \int_0^1 p(1-p)^i f(p) dp / \int_0^1 (1-p)^i f(p) dp, \quad (6)$$

where  $\bar{p}_i$  is the average demonstrability of those persons left after the  $i$ th examination. It will be shown below that  $\bar{p}_i$  decreases from examination to examination (equations (14) and (43)). We have

$$X\bar{p} = X_1 = X \int_0^1 pf(p) dp, \quad (7)$$

$$X_i = X \int_0^1 p(1-p)^{i-1} f(p) dp. \quad (8)$$

Further we note that, for all values of an arbitrary constant  $k$ ,

$$\int_0^1 \{k(1-p) - 1\}^2 pf(p) dp \geq 0, \quad (9)$$

so that by expansion and integration we have for arbitrary  $k$

$$k^2 X_3 - 2k X_2 + X_1 \geq 0. \tag{10}$$

By the elementary theory of quadratic equations, the roots of the equation

$$k^2 X_3 - 2k X_2 + X_1 = 0 \tag{11}$$

must be equal or imaginary, and so there follows

$$X_2^2 \leq X_1 X_3, \tag{12}$$

and so

$$X_1/X_2 \geq X_2/X_3. \tag{13}$$

The equality sign can hold only if all the  $p$ 's are equal, as in the hypotheses of §§ 4.2 and 4.3. This inequality is easily extended to

$$X_{i-1}/X_i \geq X_i/X_{i+1}, \text{ where } i \text{ is any positive integer,} \tag{14}$$

and then it follows that

$$X_1/X_2 \geq (X_1 + X_2 + X_3 + \dots)/(X_2 + X_3 + X_4 + \dots), \tag{15}$$

i.e.

$$X_1/X_2 \geq X/(X - X_1), \tag{16}$$

and so

$$X \geq X_1^2/(X_1 - X_2) = X_1 + X_2 + X_2^2/(X_1 - X_2). \tag{17}$$

No higher value can be assigned to the lower limit of the total number of carriers, in the general case, if the estimate is made after two examinations, because the equality (17) holds in the case of equal  $p$ 's. No inequality can be found for the upper limit of  $X$ , except for the trivial statement that  $X$  is less than the total  $N$ . For if we had a number of persons with a very low demonstrability, then in our relatively few examinations we might obtain no indication of their presence. *Indeed, the possibility of a great number of carriers with a low demonstrability in any population may render the results of all survey work in endemic areas illusory. The whole population may be infected and the surveys may be merely measuring demonstrability and not carrier rates at all.* This is almost certainly the case in malarial surveys in hyperendemic areas. After a few examinations have been made we may obtain a hint of the presence of these carriers of low demonstrability if we note an initial comparatively rapid fall in the ratio  $X_i/X_{i+1}$  and then a slow decline towards 1. Or again, with the small numbers necessarily involved in any practical survey, we may note a small number of positives appearing sporadically over an indefinitely large number of examinations. Such a result was obtained by Marsden (1946) in the survey conducted by him and discussed in an appendix by Fairfield Smith (1946).

4.5. *Hypothesis (iv). This is the most general case of heterogeneity. Demonstrability varies from one person to another and over time in one person.* There appears to be evidence for changes of  $p$  in the individual persons over a period of time. This 'periodicity' has not been systematically studied and is probably due largely to fluctuations in the number of amoebae present in the intestine, although there are no doubt effects due to changes in the consistency of the stools and other causes. The inequality (17) still holds if the function  $f(p)$  is the same at all examinations, but not in those cases where there is a *systematic* relation between the results of successive examinations. For example, it will not hold if a negative in the first

case makes a positive in the second examination more probable. There is no reason to suppose this happens in amoebic surveys, but such relations may hold in other forms of pathological examination, for instance, in the performance of daily blood examinations for benign tertian malaria. In practice, in amoebic surveys, there is usually a quite variable time between examinations. We are entitled therefore to neglect the possibility of *these systematic* relations, as it does not appear likely that there could be any natural mechanism to bring them about. In the absence of systematic relations between the demonstrabilities at successive examinations we may proceed as follows. Consider those carriers whose average demonstrability lies between  $p$  and  $p + dp$ . At the first examination, by definition a proportion  $p$  will be detected. The remainder on re-examination will have an average demonstrability  $p$ , since we have assumed no systematic relation between successive examinations. This group of carriers will behave, so far as expectations only are considered, as if they all had a constant demonstrability  $p$ . We may therefore neglect the effect of variation of demonstrability with time if it be *unsystematic*.

4.6. *An alternative derivation of the inequality (17).* If we begin with  $X$  carriers in a population of  $N$  persons and obtain a series of numbers of positives  $X_1, X_2, X_3, \dots$ , then we should expect that the mean demonstrability of the  $(X - X_1)$  remaining after the first examination is at most not greater than that of the original  $X$ . Thus

$$X_1/X \geq X_2/(X - X_1) \quad (18)$$

and

$$X \geq X_1^2/(X_1 - X_2). \quad (19)$$

Similar considerations will give

$$X \geq X_1 + X_2 + \dots + X_{i-1} + X_i^2/(X_i - X_{i+1}), \quad (20)$$

and the estimates of  $X$  can be shown to be continually increasing as we perform further tests. This inequality is likely to be of little direct practical importance after the first three or four examinations, since the numbers  $X_i$  and  $X_{i+1}$  will be relatively small in any survey and the difference,  $(X_i - X_{i+1})$ , will have a relatively high sampling error. However, the inequality will often show that the estimates of  $X$ , given in particular cases, are inconsistent with the data. A third derivation of the inequality (17) follows from (44) below, since the variance of  $p$  must be positive or zero.

4.7. *Some theoretical distributions of the demonstrability,  $p$ .*

(i) *First distribution of  $p$ .*  $f(p)$  is such that every person has the same  $p$ . This case has already been considered in § 4.2.

We have

$$f(p) = 0, \text{ if } p \text{ does not equal some fixed value } p', \text{ say.} \quad (21)$$

There follow easily

$$X_1 = Xp, \quad (22)$$

$$X_i = Xp(1-p)^{i-1}, \quad (23)$$

$$X_i/X_{i-1} = 1-p, \quad \text{for } i > 1, \quad (24)$$

$$p = (X_{i-1} - X_i)/X_{i-1}, \quad \text{for } i > 1, \quad (25)$$

$$(1-p)^{i-1}X = X_i^2/(X_i - X_{i+1}), \quad (26)$$

$$\bar{p}_i = \bar{p} = p. \quad (27)$$



(ii) A second distribution of  $p$ .

$$f(p) = \text{constant}, \quad 0 < p \leq 1. \tag{28}$$

There follow

$$X_i = X \int_0^1 p(1-p)^{i-1} dp = X\{1/i - 1/(i+1)\}, \tag{29}$$

$$X_1 = \frac{1}{2}X, \tag{30}$$

$$X_{i-1}/X_i = (i+1)/(i-1). \tag{31}$$

It is clear that this ratio falls to its limiting value 1 as  $i \rightarrow \infty$ . The distribution of  $p$  after  $i$  examinations has the form

$$f_i(p) = (1-p)^i \times \text{constant}. \tag{32}$$

Here  $f_i(p)$  is defined as the frequency function of the demonstrability of those carriers still outstanding after  $i$  examinations.

(iii) The beta-distribution. A generalization of these preceding cases is given by the beta-distribution

$$f(p) = p^{r-1}(1-p)^{s-1}/B(r, s) \quad (r > 0, s > 0). \tag{33}$$

From (32) and (33),

$$f_i(p) = p^{r-1}(1-p)^{s+i-1}/B(r, s+i). \tag{34}$$

The distribution  $f_i(p)$  will often tend to approach a curve of the form

$$f_i(p) = \text{constant} \times p^{r-1} \exp\{-p(s+i-1)\}, \tag{35}$$

and practically no person is left with a value of  $p$  not small, say for our purpose not less than  $\frac{1}{20}$ , when the number of examinations is indefinitely increased (say 100). Curve fitting may well prove of small value in estimating the total carrier rate, but the beta-functions give easily integrable solutions which are of use in illustrating the problems arising in practice and detecting fallacious methods of procedure.

Thus, if the frequency function is given by (33) and the total number of carriers is  $X$ , then

$$X_1 = Xr/(r+s), \tag{36}$$

$$X_2 = Xrs/(r+s)(r+s+1), \tag{37}$$

$$X_i = Xrs(s+1)\dots(s+i-2)/(r+s)(r+s+1)\dots(r+s+i-1), \tag{38}$$

and the inequality (17) becomes

$$X \geq Xr(r+s+1)/(r+1)(r+s), \tag{39}$$

which is, of course, reasonable.

The demonstrability of the  $i$ th examination

$$\begin{aligned} &= \int_0^1 p^r (1-p)^{s+i-2} dp \bigg/ \int_0^1 p^{r-1} (1-p)^{s+i-2} dp \\ &= B(r+1, s+i-1)/B(r, s+i-1) \\ &= r/(r+s+i-1). \end{aligned} \tag{40}$$

In Fig. 1 we show graphically the form of certain beta-function distributions.

We may fit beta-functions by estimating  $r$  and  $s$  from equations (36) and (37) using the observed values for  $X_1, X_2, X$ .

4.8. *The average demonstrability.* We are easily able to find the average demonstrability of those positives outstanding before the  $i$ th examination if we know the frequency function of the demonstrability. It is given by

$$\bar{p}_{i-1} = \int_0^1 p f_{i-1}(p) dp. \tag{41}$$

However, this quantity is not very useful, as we do not as a rule know  $f(p)$  or  $f_{i-1}(p)$ . We can, however, use our inequalities to obtain a result. By (17) the mean demonstrability of carriers at the first examination is given by

$$\bar{p}_i = X_1/X \leq (X_1 - X_2)/X_1 = 1 - X_2/X_1, \tag{42}$$

and in general

$$\bar{p}_i \leq 1 - X_{i+2}/X_{i+1}. \tag{43}$$

This is a function decreasing with increase of  $i$  by the inequalities (14).

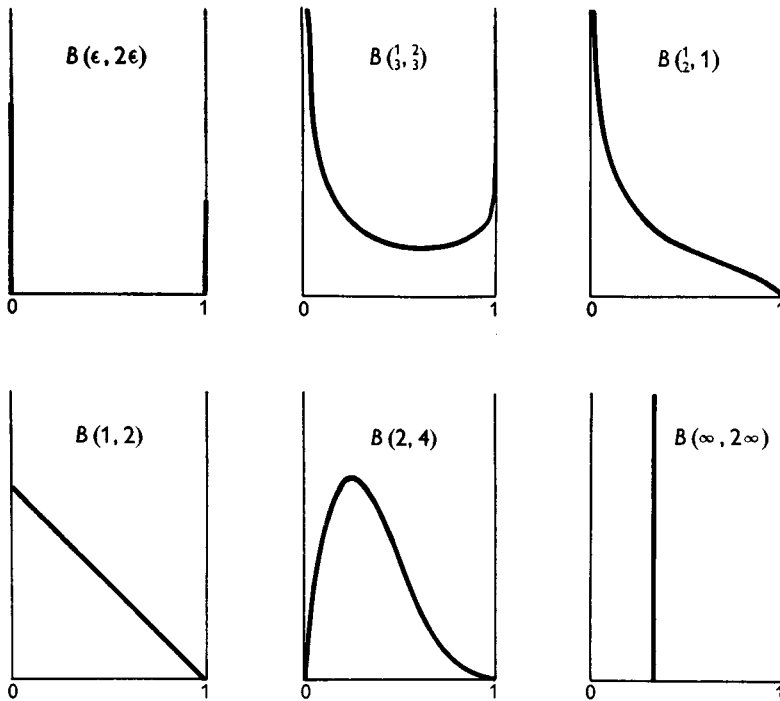


Fig. 1. The beta-distribution for certain values of  $r$  and  $s$ . The mean of the distribution, that is, the demonstrability of our problem, is  $\frac{1}{3}$  in each case, but the variance is different for each distribution.  $\infty$  is used for 'an infinitely large number' and  $\epsilon$  for an 'indefinitely small positive number'.

4.9. *The exhaustive examination of a group of carriers.* It is of some interest to investigate the theoretical implications of method (ii) of § 2.1. To avoid confusion we shall suppose in this section that there are  $Y$  carriers each examined  $n$  times, and that  $Y_i$  of them yield positive findings at  $i$  separate examinations. As before,

we shall suppose that  $y$  is large and that the demonstrability of the carriers has a frequency function  $f(p)$ . Then it follows that

$$Y_i = Y \int_0^1 {}^n C_i p^i (1-p)^{n-i} f(p) dp, \tag{44}$$

$$\bar{p} = \sum_i i Y_i / n Y, \tag{45}$$

$$\begin{aligned} \text{var } p &= E\{(p - \bar{p})^2\} \\ &= E(p^2) - \bar{p}^2 \\ &= E\{p^2(p + (1-p))^{n-2}\} - \bar{p}^2 \\ &= \left\{ \sum_{i=2}^n {}^{n-2} C_{i-2} Y_i / n C_i Y \right\} - \bar{p}^2 \\ &= \left\{ \sum_{i=0}^n i(i-1) Y_i / n(n-1) Y \right\} - \left\{ \sum_{i=0}^n i Y_i / n Y \right\}^2, \end{aligned} \tag{46}$$

and we can immediately derive an inequality for  $Y$  since  $\text{var } p$  must be positive

$$Y \geq (n-1) \left\{ \sum_{i=0}^n i Y_i \right\}^2 / n \sum_{i=0}^n i(i-1) Y_i. \tag{47}$$

We may apply this inequality (47) to Svensson's data. We summarize the data of Table 2A and 2B, and illustrate the necessary computations, in Table 3. For males and females, the inequality (47) leads to the results that  $Y \geq 8$  and  $Y \geq 10$ , for males and females respectively, if we substitute observed for expected values on the right-hand side. Both of these results are of small value in the present case, but would be of value in larger samples.  $Y$  was, of course, known *a priori* here, since Svensson had chosen a definite group of known carriers.

Table 3

No. of positive findings ( $i$ )	Frequency of positive findings ( $Y_i$ )		$i(i-1)$	$i Y_i$		$i(i-1) Y_i$	
	Men	Women		Men	Women	Men	Women
0	4	5	0	0	0	0	0
1	6	3	0	6	3	0	0
2	0	1	2	0	2	0	2
3	1	3	6	3	9	6	18
4	2	1	12	8	4	24	12
5	0	2	20	0	10	0	40
6	1	0	30	6	0	30	0
Total	14	15	—	23	28	60	72

4.10. *The average demonstrability estimated by certain authors.* It has been the custom to estimate an average demonstrability,  $\bar{p}$ , for the amoebic species and then to multiply the numbers found at a single examination by the reciprocal of the figure,  $1/\bar{p}$ . A widely used factor is that of Dobell (1917), namely, 3. We show later that this figure, the reciprocal of the average demonstrability, is often too low because of bias entering into the original estimate. Most observers, including Dobell, had a relatively high number of easily demonstrable persons, just because the criterion of positiveness was provided by the actual observations themselves. Every person was positive at least once in the limited number of observations, for otherwise he could not have been included in the known 'carrier' class. We may

illustrate this by the theoretical example treated above in § 4·7, ‘Distribution (ii)’, where every value for  $p$  is equally likely. The real average demonstrability,  $\bar{p}$ , is  $\frac{1}{2}$ , but if it is estimated merely from the data themselves after one, two, three,... examinations the estimates of the demonstrability will on the average be 1,  $3/4$ ,  $4/6$ ,  $5/8$ ,  $6/10$ ,  $7/12$ , etc., for the total number of positive findings after  $i$  examinations will be  $\frac{1}{2}iX$  and the number of known carriers will be  $Xi/(i + 1)$ . Therefore the total number of positive findings is  $\frac{1}{2}iX$  and these will be presumed to come from only  $Xi/(i + 1)$  carriers who have each been examined  $i$  times, and the mean demonstrability will be estimated to be  $(i + 1)/2i$ . As the number of examinations is increased this figure will fall towards its true value  $\frac{1}{2}$ . This result is also obvious from practical considerations, since if the whole  $N$  persons are examined each time then the expectation of the number found positive will be a constant for each examination, but the expression in the denominator will be constantly increasing. This point was well realized by Svensson (1934). The amended data in our own Tables 2A and 2B show 51 positive examinations out of a total number 174, a mean demonstrability of 0·27 approximately. Yet if we were to use the data alone to tell us which are the carriers we should say that there are only 20 carriers each of whom was examined six times and who in total gave 51 positive results, and we should then estimate the mean demonstrability as  $51/120$  or 0·42, clearly a biased result.

Some authors have made an even more deceptive calculation of the demonstrability when using method (i) of § 2·1. In our notation they would calculate as follows:  $X_i$  persons have been examined  $i$  times, and so after  $k$  examinations the number of positive examinations obtained is  $\sum_{i=1}^k X_i$ , and the number of examinations made on these cases is  $\sum_{i=1}^k iX_i$ , so that the ‘demonstrability’ is given by

$$\frac{\sum_{i=1}^k X_i}{\sum_{i=1}^k iX_i}.$$

This estimate is usually too high because  $k$ , the number of examinations, is taken to be 3 or 6. It could be made too small, on the other hand, if there were several carriers with very low demonstrability in the population studied, and they were examined a great number of times, thus giving undue weight to their  $p$ . The mean of the demonstrabilities so obtained is the harmonic mean and approaches zero as the number of examinations is increased in the case of the second distribution of § 4·7.

4·11. *The standard error of the observed carrier rate. Method (ii). Notation.* Let  $c$  be the true carrier rate,  $\gamma$  the observed carrier rate. Then

$$E(\gamma) = cE(p) = c\bar{p}. \tag{48}$$

In the usual binomial sampling of Bernoullian type it is customary to attach a standard error  $p'(1 - p')/N$  to an observed proportion  $p'$ . If Bernoullian conditions do not hold then the use of this standard error is not valid. We shall now discuss the standard error in two main cases. (1) The same  $N$  persons are repeatedly examined and  $X$  of them are carriers.  $X'$  are discovered in one examination.

(2)  $N$  persons are drawn from some population at each trial. In each class of cases we may have different results depending on the values of  $c$  and the distribution of  $p$ . In this section 'var' will mean 'sampling variance of'.

(1)  $N$  persons are repeatedly examined (the so-called Poisson sampling).  $X' = Nc'p'$ , where  $c'$  and  $p'$  may differ by sampling errors from the true  $c$  and  $p$ , but we are not concerned with the variation of  $c'$  which is regarded as fixed.

(i)  $c = 1$ , and therefore  $c' = 1$ , so that  $X = N$ .

(iA)  $p = 1$ , and therefore  $p' = 1$ . This is trivial. There is a sampling error of zero.

(iB)  $p$  is not 1 or 0, but some constant. In this case

$$\text{var } \gamma = \text{var } p' = p(1-p)/N. \tag{49}$$

(iC)  $p$  has some frequency distribution with mean  $\bar{p}$  and non-zero variance.

Then 
$$\text{var } \gamma = \text{var } p' = \sum p(1-p)/N^2. \tag{50}$$

(ii)  $0 < c < 1$ , so that  $E(X) = Nc$ , but  $X$  in any given sample is  $Nc'$ .

(iiA)  $p = 1$ , and therefore  $p' = 1$ . The sampling error is still zero, but the usual Bernoullian expression would give

$$\text{var } \gamma = c'(1-c')/N. \tag{51}$$

An expression of this form is valid in case (2) below.

(iiB)  $p$  is a constant for all carriers.

$$\text{var } \gamma = c'^2 \text{var } p' = c'^2 p(1-p)/X \simeq \gamma(c' - \gamma)/c'N. \tag{52}$$

(iiC)  $p$  has some frequency distribution with mean  $\bar{p}$  and non-zero variance  $V$ .

$$\text{var } \gamma = c'^2 \{\bar{p}(1-\bar{p}) - V\}/X \simeq \gamma(c' - \gamma)/c'N - c'V/N. \tag{53}$$

(2) If  $N$  persons are picked at random from an infinite population, then the probability that any single examination will give a positive result is  $c\bar{p}$ . The sampling variance of  $\gamma$  is

$$\text{var } \gamma = c\bar{p}(1-c\bar{p})/N. \tag{54}$$

When  $\bar{p} = 1$ , this is similar in form to (51). It is clear that we may use the ordinary Bernoullian standard error to compare two different populations if all we are interested in is the observed carrier rate at a single examination, that is,  $X_1/N$ . But we can get no significance test to compare the real carrier rates,  $c$ , unless we know  $\bar{p}$ . As an exhaustive examination of each population to determine  $\bar{p}$  is necessary, it is more logical to compare the results of the exhaustive examinations.

4.12. We may use the mathematical model in several ways, chiefly to illustrate certain difficulties of extrapolation. Table 4 shows the results for a number of beta-function distributions all with the same mean  $\frac{1}{3}$ , but with different variances. The most favourable case is that distribution  $B(\infty, 2\infty)$  where all the  $p$ 's are equal to  $\frac{1}{3}$ , the least favourable is that in which the  $p$ 's are either close to unity or zero,  $B(\epsilon, 2\epsilon)$ . When all the demonstrabilities are equal to  $\frac{1}{3}$ , there are only 8.78% outstanding after six examinations, whereas in the population with the same average demonstrability, but with the greatest possible variance,  $B(\epsilon, 2\epsilon)$ , two-thirds are still outstanding after six examinations (or, indeed, any number). In this distribution  $p = 0$  with probability  $\frac{2}{3}$  and  $p = 1$  with probability  $\frac{1}{3}$ , which

means that  $\frac{2}{3}$  of the carriers carry so rarely as to be undetectable and the remainder are bound to be detected. This example brings out clearly the difficulty inherent in the whole subject; no method of examination will distinguish between non-carriers and those who carry very rarely.

Two populations may contain quite widely differing proportions of carriers and yet give the same number of positives at the first, second and third examinations. Thus great caution is necessary in extrapolating from the results of the first few examinations. In Table 4 we have illustrated the point that with the same mean demonstrability in each population, the outstanding cases not yet detected will be greatest in the populations with the greatest variance or scatter of the demonstrability about the mean demonstrability.

Table 4. *A table of the expectation of the proportion of outstanding carriers after varying numbers of examinations*

No. of examinations	Population sampled							
	$B(\epsilon, 2\epsilon)^*$	$B(\frac{1}{3}, \frac{2}{3})$	$B(1, 2)$	$B(2, 4)$	$B(3, 6)$	$B(4, 8)$	$B(10, 20)$	$B(\infty, 2\infty)$
1	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667	0.6667
2	0.6667	0.5556	0.5000	0.4762	0.4667	0.4615	0.4516	0.4444
3	0.6667	0.4938	0.4000	0.3571	0.3394	0.3297	0.3105	0.2963
4	0.6667	0.4527	0.3333	0.2778	0.2545	0.2418	0.2164	0.1975
5	0.6667	0.4225	0.2857	0.2222	0.1958	0.1813	0.1528	0.1317
6	0.6667	0.3990	0.2500	0.1818	0.1538	0.1387	0.1091	0.0878
$\bar{p}$	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
var $p$	0.2222	0.1111	0.0556	0.0317	0.0222	0.0171	0.0072	0.0000

\*  $\epsilon$  is an arbitrarily small positive number.

Linders graduated Svensson's data by means of a logistic curve so that in our notation

$$\sum_{i=1}^x X_i = X/(1 + e^{B-Cx}), \tag{55}$$

where  $C$  is positive. It is of interest to note that a logistic curve cannot satisfy inequality (17) since it gives the following:

$$\begin{aligned} X_i &= X/(1 + e^{B-Ci}) - X/(1 + e^{B-Ci+C}) \\ &= Xe^{B-Ci}(e^C - 1)/(1 + e^{B-Ci})(1 + e^{B-Ci+C}). \end{aligned} \tag{56}$$

The number of carriers outstanding before the  $i$ th examination is given by

$$X - \sum_{j=1}^{i-1} X_j = Xe^{B-Ci+C}/(1 + e^{B-Ci+C}). \tag{57}$$

The mean demonstrability at the  $i$ th examination is

$$\bar{p}_{i-1} = (1 - e^{-C})/(1 + e^{B-Ci}), \tag{58}$$

and this is an increasing function of  $i$ , asymptotically approaching  $(1 - e^{-C})$ , as  $i \rightarrow \infty$ .

5. *The treatment of survey material by life-table methods.*

5.1. *Introductory.* In practice it is found to be impossible to examine by any test a large body of persons a number of times without withdrawals. These withdrawals we must hope to be independent of the probability of detecting amoebae.

We should obtain gross bias, for example, if the later examinations were carried out on predominantly institutional patients who are known to have relatively high carrier rates. The results given below of analyses of the data of several well known surveys may be regarded as illustrations of the methods appropriate.

5.2. *Application to published data.* McCoy (1936) examined the employees of certain Chicago hotels after the famous dysentery epidemic in that city. Only in hotel X were enough persons examined to make it worthwhile to analyse the findings by our methods. Table 5 gives some of McCoy's data in a slightly amended form.

Table 5. *The proportion of stool specimens from employees of hotel X*  
(McCoy's page 38, Table 10.)

Examination	No. examined	No. positive	Proportions positive
First	1176	203	0.172619
Second	867	123	0.141869
Third	670	56	0.083582
Fourth	558	33	0.059140
Fifth	454	32	0.070485
Sixth	202	8	0.039604
Seventh	24	2	0.083333
Total specimens examined	3951	457	—

Table 6

(McCoy's data treated by life-table methods.)

No. of examinations	No. of persons to have examination	No. of positives at this examination	Presumed no. of outstanding carriers	Mean demonstrability
1	100,000	17,262	51,786 (45,351)	0.33 (0.38)
2	82,738	11,738	34,524 (28,089)	0.34 (0.42)
3	71,000	5,934	22,786 (16,351)	0.26 (0.36)
4	65,066	3,848	16,852 (10,417)	0.23 (0.37)
5	61,218	4,315	13,004 (6,569)	0.33 (0.66)
6	56,903	2,254	8,689 (2,254)	0.26 (1.00)
Total	—	45,351	—	—

We computed the proportion of persons, not previously positive at any examination, positive at the  $i$ th examination ( $i = 1, 2, \dots, 6$ ). The results agreed with McCoy's. Life-table methods were then used and the numbers that would be found positive at each examination out of an initial 100,000 persons were obtained. The common practice of estimating the total number of carriers by (i) three times the number of positives from the first examination ( $X = 3X_1$ ) or (ii) the total number of positives from the first six examinations, was then followed. The number of carriers outstanding at any examination (that is, the number of carriers not yet proved positive) was obtained by subtraction, and the demonstrability at any stage then followed straightforwardly by division. The results are given in Table 6. The demonstrability calculated by (i) is shown in the final column with that given by (ii) in brackets. In this example the first estimate appears adequate. The demonstrability continues to fall as we should expect on theoretical grounds with the exception of the fifth value which may be due to sampling errors in the original data.

We rarely have any idea of the real magnitude of the average demonstrability. The estimate of the lower limit for  $X$  by our rule is  $(17,262)^2/5524$  or 53,924, which is close to the estimate of  $3X_1$  or  $3 \times 17,262$ , i.e. 51,786.

5.3. *Application to published data, continued. Data of Boeck & Stiles (1923).* Similar conclusions may be drawn from the data of Boeck & Stiles (1923) with regard to the *Entamoeba coli* findings, but both estimates are hopelessly low in the cases of *E. histolytica*.

Table 7. *A table to show the number of isolations of Entamoeba histolytica and E. coli*

(Table 2, p. 20, of Boeck & Stiles (1923).)

No. of examinations	No. of cases examined	New cases of <i>E. histolytica</i>	New cases of <i>E. coli</i>
1	8029	234	1269
2	1441	34	155
3	1050	27	73
4	912	13	44
5	791	13	27
6	623	9	13
Total	—	330	1586

Table 8. *The data of Table 7 treated by life-table methods (Entamoeba coli cases)*

No. of examinations	No. of persons for examination	No. of positives at this examination	Presumed no. of carriers outstanding	Mean demonstrability
1	100,000	15,805	47,415 (37,070)	0.33 (0.43)
2	84,195	9,056	31,610 (21,265)	0.29 (0.43)
3	75,139	5,224	22,554 (12,209)	0.23 (0.43)
4	69,915	3,373	17,330 (6,985)	0.19 (0.48)
5	66,542	2,271	13,957 (3,612)	0.16 (0.63)
6	64,271	1,341	11,686 (1,341)	0.11 (1.00)
Total		37,070		

Table 9. *The findings of Table 7 with regard to Entamoeba histolytica treated by life-table methods*

No. of examinations	No. of persons for examination	No. of positives at this examination	Presumed no. of carriers outstanding	Mean demonstrability
1	100,000	2,914	8,742 (11,749)	0.33 (0.25)
2	97,086	2,291	5,828 (8,835)	0.39 (0.26)
3	94,795	2,438	3,537 (6,544)	0.69 (0.37)
4	92,357	1,316	1,099 (4,106)	— (0.32)
5	91,041	1,496	— (2,790)	— (0.54)
6	89,545	1,294	— (1,294)	— (1.00)
Total		11,749		

A slight difficulty arises here, since some of those persons examined the second time had already been found positive. We cannot sort out these, but this group is probably not important in the case of *E. histolytica* in which we are most interested,



since positives for this organism would tend to be treated immediately after its discovery. We conclude that the first estimate is consistent with the data but that the second gives a result obviously too low. The rule gives

$$X \geq (15,805)^2/6749 = 37,012 \text{ for } E. coli,$$

and  $X \geq (2914)^2/623 = 13,630 \text{ for } E. histolytica.$

‘Demonstrability’ has little meaning in the latter case. The first estimate is too low, in fact it is exceeded by the number of positives in the first four examinations. The second estimate gives a continuously rising figure for the mean ‘demonstrability’ of the carriers still outstanding before each examination and is thus also too low.

5.4. *Application to published data, continued. Marsden’s data (1946).* Fairfield Smith comments on the approximate equality of the numbers at the fourth, fifth and sixth examinations ‘apparently suggesting an indefinite continuation of positive findings if the series were to be extended’. With this statement our theory would agree. The small number of positives at each examination would be from a large reservoir of carriers each with low probability of being detected at any given examination. Our rule applied after two examinations gives an estimate of a minimum of carriers 26,296 but later estimates will be much higher. If one accepts the view that the parasite once acquired is probably never lost one can only conclude that in an endemic population the whole population may really be infected, that ‘carrier rates’ are mere artifacts, due to the low numbers of amoebae present and the relative inefficiency of the methods employed, and that our surveys merely measure mean demonstrability.

Table 10. *The data of Marsden (1946) treated by life-table methods (Entamoeba histolytica)*

No. of examinations	No. of persons for examination	No. of positives at this examination	No. of carriers outstanding	Mean demonstrability
1	100,000	11,034	33,102 (30,246)	0.33 (0.36)
2	88,966	6,404	22,068 (19,212)	0.29 (0.33)
3	82,562	4,335	15,664 (12,808)	0.28 (0.34)
4	78,227	2,660	11,329 (8,473)	0.23 (0.31)
5	75,567	2,956	8,669 (5,813)	0.34 (0.51)
6	72,611	2,857	5,713 (2,857)	0.50 (1.00)
Total		30,246		

5.5. *Application to published data, continued. The data of Carter, Mackinnon, Matthews & Malins Smith (1917).* The numbers of persons in their fifth and sixth examinations were small, as was the number of positives, so that they are not reproduced here. The mean demonstrability is not falling, so it is very probable that the estimate of  $3 \times 5896$  is too small (see Table 11).

Table 11. *The detection of E. histolytica*  
(Data of Carter *et al.* (1917).)

No. of examination	No. of persons for examination	No. of positives at this examination	No. of carriers outstanding	Mean demonstrability
1	100,000	5,896	17,688 (= $3X_1$ )	0.33
2	94,104	2,594	11,792	0.22
3	91,510	2,428	9,198	0.26
4	89,516	1,950	6,770	0.29

5.6. *Conclusions and summary.* We summarize our conclusions in Table 12.

Table 12. *The minimum estimate permissible for several surveys of the carrier rate for Entamoeba histolytica*

Authors	Minimum estimate of carrier rate of <i>E. histolytica</i> %
McCoy (1936), hotel X	52
Boeck & Stiles (1923)	12
Carter <i>et al.</i> (1917)	20
Marsden (1946)	33

6. *Lines of investigation which should be followed.*

6.1. *The investigation of the changes in demonstrability with time.* No detailed data appear to be available on the actual numbers of amoebae in the stools, although Porter (1916) has given figures for *Giardia lamblia*. No doubt an 'analysis of variance' approach to the problem would reveal significant differences between stools on different days. Detailed studies could then be made on the counts over successive days for a considerable period of time in known carriers.

6.2. *The demonstrability at a given time.* It seems to have been the custom in some investigations such as that of Andrews (1934) to examine a stool with a knowledge of previous results. This type of investigation certainly shows that amoebae actually present may often be missed and can sometimes be found by further searching, but is really not otherwise helpful. A knowledge of the counts would help in the discussion of this problem. In general the following method would be satisfactory. The microscopist, not knowing the previous outcome of examinations on the particular person examined, examines a fixed quantity of stool under standard conditions (saline preparation). At the same time permanent stained films are made. He then finds whether he has missed any amoebae previously recorded in this person; if so he will again search another preparation and also the stained films to find if the amoebae were present in the stools. In the event of finding an amoeba not previously recorded for the particular person he will search stained films made from the stool at the previous examinations. These data would help to explain why the demonstrability is not unity. Counts would be made on each stool at every examination once the amoebae had been recorded at any examination.

6.3. *The best system of examinations as to time spacing.* This problem will be nearer solution when the data of an investigation of the kind set out in § 6.1 are available. It is probable that with the patients in hospital other considerations than the mere number of examinations will be of importance in assessing the best system. In surveys, however, a time interval of weeks may be the most efficient method.

6.4. *The question, 'How many examinations must be done to assure that a person is not a carrier?'* The work described in earlier sections suggests that a number of persons may have a demonstrability of, say, 0.1 or less. It is clear that many examinations would have to be made in order to reduce the probability of missing such a person to a reasonably low figure. If in a group of carriers the demonstrability is  $p$ ,  $k$  is the number of examinations and  $\alpha$  is the proportion of the group of actual carriers we think we can afford to miss, then  $k$  must be such that

$$(1-p)^k < \alpha, \quad (59)$$

or 
$$k > \log \alpha / \log(1-p). \quad (60)$$

7. *Summary.* (i) The theory of amoebic surveys and of related types of pathological surveys is discussed.

(ii) An inequality is derived,

$$X \geq X_1^2 / (X_1 - X_2),$$

where  $X$  is the total number of carriers and  $X_1, X_2$  are the new carriers found at the first and second examinations.

(iii) Demonstrability is defined and several theoretical distributions of it are considered.

(iv) Demonstrability is shown not to be a constant in the populations considered.

(v) The danger of extrapolation is noted.

(vi) Life-table methods are introduced to treat some well-known published series.

(vii) A criterion for the adequacy of any estimate of the total carrier rate is given. The mean demonstrability of the outstanding carriers after successive examinations must form a decreasing series.

(viii) This criterion usually shows that the method of estimating the carrier rate for *Entamoeba histolytica* as three times the proportion of positives at a single examination, or as the total number of positives picked up in six examinations, both give too low a carrier rate.

(ix) It is suggested that the investigation of the 'carrier' rate may in certain cases be merely a measure of 'demonstrability'.

(x) Possible future lines of investigation of the problem are noted.

The work was partly completed while the author was a Rockefeller Fellow in Medicine. He would like to thank Professor A. Bradford Hill for the facilities of his department at the London School of Hygiene and Tropical Medicine, and Dr J. O. Irwin and Mr P. Armitage for assistance in clearing up doubtful points. An abbreviated form of this paper was read by the author to the meeting of the Biometric Society on 28 April 1949 in London.

## REFERENCES

- ANDREWS, J. (1934). The diagnosis of intestinal protozoa from purged and normally passed stools. *J. Parasit.* **20**, 252.
- BOECK, W. C. & STILES, C. W. (1923). Studies on various intestinal parasites (especially amoebæ) of man. *Hygienic Laboratory Bull. U.S. Treasury Dep.* No. 133.
- CARTER, H. F., MACKINNON, D. L., MATTHEWS, J. R. & SMITH, A. MALINS. (1917). Protozoological investigation of cases of dysentery conducted at the Liverpool School of Tropical Medicine (Second Report). *Ann. trop. Med. Parasit.* **11**, 27.
- DOBELL, C. (1917). Amoebic dysentery and the protozoological investigation of cases and carriers. *Spec. Rep. Ser. med. Res. Comm., Lond.*, no. 4.
- FAIRFIELD SMITH, H. (1946). Appendix to Marsden (1946).
- GINI, C. (1924). Prime ricerche sulla 'fecondabilità' della donna. *Atti Ist. veneto*, **83**, 315.
- MCCOY, G. (1936). Epidemic amebic dysentery. The Chicago outbreak of 1933. *Nat. Inst. Health Bull. U.S. Treasury Dep.* No. 166.
- MARSDEN, A. T. H. (1946). The detection of the cysts of *Entamoeba histolytica* in the faeces by microscopic examination. *Med. J. Aust.* **33** (1), 915.
- PORTER, A. (1916). An enumerative study of the cysts of *Giardia (lamblia) intestinalis* in human dysenteric faeces. *Lancet*, **1**, 1166.
- SAWITZ, W. G. & FAUST, E. C. (1942). The probability of detecting intestinal protozoa by successive stool examinations. *Amer. J. trop. Med.* **22**, 130.
- SAWITZ, W. & KARPINOS, B. D. (1942). Statistical problems involved in the application of the N.I.H. swab for the diagnosis of oxyuriasis. *Amer. J. Hyg.* **35**, 15.
- SVENSSON, R. (with the co-operation of LINDERS, F. J.). (1934). The chances of detecting infections with intestinal protozoa. *Acta med. scand.* **81**, 267.

(MS. received for publication 4. IV. 50.)