J. Hyg., Camb. (1969), **67**, 533
Printed in Great Britain

Tables of the upper limit to the estimate of the density of contaminating particles in a medium

BY S. PETO AND B. J. MAIDMENT
Microbiological Research Establishment, Porton, Salisbury, Wilts

(*Received* 1 *March* 1969)

The expression 'contaminating particles' is used in the broadest sense: micro-organisms, virus particles and mutants of a parent organism are examples from microbiology, but other sciences face a problem similar to that dealt with in this paper. According to the case, the 'medium' can be air, a culture medium, a suspension or any other material containing 'foreign' bodies. It is assumed that the contaminating particles are randomly distributed throughout the medium (i.e. that the probability of finding any individual particle in any one unit volume is constant and that the sampling procedure itself does not interfere with this assumption).

Theoretically, the number of particles found in a random sample is an unbiased estimate of the density of the contaminant in the medium; but this is of little practical value, particularly where the sample drawn proves to be free from contaminants. The determination of an upper limit to the estimate of the density is often required and this upper limit is defined by a probability level previously fixed by the investigator. If, for instance, the 5 % upper limit of contamination of a medium is found to be six particles per litre (this is usually called the 5 % fiducial or confidence upper limit) the investigator concludes that the density of contamination is not higher, unless he is the victim of a 1 in 20 mischance of sampling. (Obviously the upper limit caters for sampling fluctuations only and the assumption of random distribution must hold good.)

Four cases emerge and will be treated separately: (A) an uncontaminated sample drawn from an infinitely large medium; (B) an uncontaminated sample drawn from a medium of finite size; (C) a contaminated sample drawn from an infinitely large medium; (D) a contaminated sample drawn from a medium of finite size.

(A) *An uncontaminated sample drawn from an infinitely large medium*

'Infinitely large' in this context means that the sample volume is negligible compared with the volume of the medium, e.g. samples of sterilized air. The numbers of contaminating particles in the separate units of volume will follow the Poisson distribution and therefore the upper limit n of its expectation is given by

$$n = -\log_e P = (-\log_{10} P) \times 2{\cdot}3026, \quad (1)$$

where P stands for the selected probability level. This gives for instance $n = 3$ for $P = 5\,\%$ and $n = 4{\cdot}6$ for $P = 1\,\%$.

It follows that the question 'How large (assuming it will turn out to be sterile) must the sample volume v be to ensure a density of not more than n contaminants, the risk of error being P?' is answered by

$$v = \frac{-\log_e P}{n}. \tag{2}$$

Examples

To test its efficiency, a sample of 100 ft.^{3} of effluent air from an air sterilization plant was examined and no contaminants were found. Using (1) it can be concluded that not more than three contaminants per 100 ft.^{3} (on average) will in future defy the sterilization operation, where the probability of being misled by the sample is 5 %, (assuming, naturally, that the conditions of running the plant remain constant). If the test sample were increased to 10^4 or even 10^6 ft.^{3} and still no contaminants were found, this would only show that the upper limit of density of contamination had fallen to three particles per 10^4 or three particles per 10^6 ft.^{3} respectively. Obviously 100 % efficiency cannot be proved by mere sampling.

The same logic applies to a wide variety of cases: e.g. drugs intended for humans are tested with guinea-pigs or a number of flights carried out to demonstrate the safety of an aircraft; if 1000 animals have shown no ill effect or if 1000 sorties have been flown without accident, then the 5 % upper limit is three adverse results per 1000 and no increase in sample size will ever support a claim of absolute safety.

If, however, the air sterilization plant (see above) had been improved so that, when tested, it produced an uncontaminated sample of 100 ft.^{3}, and if it had formerly returned on average five contaminants per 100 ft.^{3}, this would be good evidence for the superiority of the new technique, because the probability of obtaining an uncontaminated sample if the expectation was 5 particles would be small. [Using (1): $P = e^{-5} = 0{\cdot}7$ %.]

(B) *An uncontaminated sample drawn from a medium of finite size*

Let V be the volume of the medium, let n be the maximum acceptable number of contaminating particles per f units of volume, let P be the probability that n is in fact exceeded, let p be the fraction of V to be sampled, let q be equal to $(1-p)$, let Z be equal to $q \log q$, then, provided the sample is found to be uncontaminated,

$$q^{Vqn/f} = P, \tag{3}$$

i.e.

$$q \log q = \frac{f \log P}{Vn} = Z. \tag{4}$$

This equation was solved with the help of a computer for values of Z from $-0{\cdot}004$ ($0{\cdot}001$) to $-0{\cdot}15$, which correspond to p values in the range 0·9–59·9 %. Table 1 records these pairs of values.

Examples

(*a*) How large a sample must be drawn from a culture vessel of 4 l. to state that the rest of the vessel does not contain more than three infective units (of a certain virus) per l., assuming that the sample will be sterile and that the risk of being led into error by the sample is not greater than 5 %?

https://doi.org/10.1017/S0022172400041966 Published online by Cambridge University Press

Here: $V = 4$, $n = 3$, $P = 0{\cdot}05$ ($\log P = -1{\cdot}3$), $f = 1$. Therefore [using (4)] $Z = 1 \times (-1{\cdot}3)/(4 \times 3) = -0{\cdot}1083$ and entering Table 1 at this value for Z gives 29·9 %.

Answer: 29·9 %; that is, 1·2 l. must be sampled.

Table 1. *Percentage of the medium to be sampled*

Z	0	1	2	3	4	5	6	7	8	9
−0·00					0·9	1·2	1·4	1·6	1·9	2·1
1	2·3	2·6	2·8	3·0	3·3	3·5	3·8	4·0	4·2	4·5
2	4·7	5·0	5·2	5·4	5·7	5·9	6·2	6·4	6·7	6·9
3	7·2	7·4	7·7	7·9	8·2	8·4	8·7	8·9	9·2	9·4
4	9·7	10·0	10·2	10·5	10·7	11·0	11·2	11·5	11·8	12·0
−0·05	12·3	12·6	12·8	13·1	13·4	13·6	13·9	14·2	14·5	14·7
6	15·0	15·3	15·6	15·8	16·1	16·4	16·7	17·0	17·2	17·5
7	17·8	18·1	18·4	18·7	19·0	19·3	19·5	19·8	20·1	20·4
8	20·7	21·0	21·3	21·6	21·9	22·3	22·6	22·9	23·2	23·5
9	23·8	24·1	24·5	24·8	25·1	25·4	25·7	26·1	26·4	26·7
−0·10	27·1	27·4	27·8	28·1	28·4	28·8	29·1	29·5	29·8	30·2
1	30·6	30·9	31·3	31·7	32·0	32·4	32·8	33·2	33·6	34·0
2	34·4	34·8	35·2	35·6	36·0	36·4	36·8	37·3	37·7	38·1
3	38·6	39·0	39·5	39·9	40·4	40·9	41·4	41·9	42·4	42·9
4	43·5	44·0	44·5	45·1	45·7	46·3	46·9	47·6	48·2	48·9
−0·15	49·6	50·4	51·2	52·0	52·9	53·9	54·9	56·2	57·6	59·5

(*b*) If in the previous example the acceptable upper limit was six infective units per 100 ml. and the vessel contained 30 l.: $V = 30$, $n = 6$, $P = 0{\cdot}05$, $f = 0{\cdot}1$; then $Z = -0{\cdot}0007$; this Z value is smaller than the table provides for and a Poisson distribution is a fair approximation, formula (2) supplying the answer. [3/6 unit volumes = $0{\cdot}5 \times f$ = 0·05 l.]

(*c*) A sample of 0·5 l. from a 4 l. vessel was found to be uncontaminated; 1 % fiducial upper limit to the density in the remaining part of the culture is required. Starting from the sample size which is 12·5 % (and taking as unit volume 100 ml.) Table 1 gives the corresponding Z value of −0·0506 and changing the subject of formula (4):

$$n = \frac{f \log P}{VZ} = 0{\cdot}1 \times (-2)/[4 \times (-0{\cdot}0506)] = 0{\cdot}99.$$

Answer: There should not be more than one infective unit per 100 ml. in the vessel unless there has been a one in hundred mischance in sampling.

(*d*) If in the previous example the question had been to find the probability that the unknown density should not be more than 0·5 per 100 ml. the subject of (4) is again changed:

$$\log P = \frac{ZnV}{f} = -0{\cdot}0506 \times 0{\cdot}5 \times 4/0{\cdot}1 = -1{\cdot}0133 = \bar{2}{\cdot}9867 \text{ and } P = 0{\cdot}09699.$$

Answer: The probability of the density exceeding 0·5 infective units per 100 ml. is approx. 9·7 %.

(C) *A contaminated sample drawn from an infinitely large medium*

In most cases the investigator trying to assess the density of contaminants expects only a small number of them and hopes to find none in his sample. If, however, the sample turns out to be contaminated he might want to assign an upper limit to the most probable density determined by k, the number of contaminants observed in the sample. The problem is to find the expectation n of a Poisson distribution so that the chance of obtaining a sample containing k or less is equal to the desired probability level or, in mathematical form, the solution for n of the equation

$$\sum_{k=0}^{k} \frac{e^{-n}k^n}{k!} = P. \tag{5}$$

This equation can be solved by trial and error (for instance), preferably with the help of a computer. Table 2 may prove useful. (The figures are given to the nearest first place of decimals that would keep P below the quoted value.) Other tables exist giving confidence or fiducial limits for the expectation of the Poisson distribution, e. g. Fisher & Yates, *Statistical Tables*, table VIII, 1, and *Biometrika Tables*, table 40.

Table 2. *The probability* P *that the density in an infinitely large medium will exceed the indicated number, if* k *particles are found in the sample*

k ...	1	2	3	4	5	6	7	8	9	10
10 %P	3·9	5·4	6·7	8·0	9·3	10·6	11·8	13·0	14·3	15·5
5 %P	4·8	6·3	7·8	9·2	10·6	11·9	13·2	14·5	15·8	17·0
1 %P	6·7	8·5	10·1	11·7	13·2	14·6	16·1	17·5	18·8	20·2
0·1 %P	9·3	11·3	13·1	14·9	16·5	18·1	19·7	21·2	22·7	24·2
0·01 %P	11·9	14·0	16·0	17·9	19·7	21·4	23·1	24·7	26·3	27·9

Example

A yearly average of 12 cases of a disease used to be reported in a certain country. After measures claiming to be of prophylactic value had been introduced the incidence dropped in the first year to two cases. Is this convincing evidence for the efficacy of these measures? Since it seems reasonable to assume a Poisson distribution, Table 2 is entered at $k = 2$ and 12·0 is found to lie between $P = 0{\cdot}1\,\%$ and 0·01 %, which makes it extremely unlikely that the drop in the number of cases is the result of mere sampling fluctuation. Answer: Yes (naturally provided that any other cause of the decrease in incidence of the disease can be ruled out).

(D) *A contaminated sample drawn from a medium of finite size*

Let p be the proportion of the medium to be sampled, let q be equal to $(1-p)$, let k be the number of contaminating particles observed in the sample, let N be the upper limit of the number of particles in the remaining medium, let P be the probability level defining the upper limit, let m be equal to $(N+k)$, then equating the tail of the appropriate binomial distribution to P:

$$\sum_{r=0}^{k} q^{m-r}\, p^r = P, \tag{6}$$

https://doi.org/10.1017/S0022172400041966 Published online by Cambridge University Press

Table 3. 5% (1%) *fiducial upper limit to the estimate of the remaining number of contaminating particles in the medium after a sample of* p % *of it contained* k *particles*

p \ *k*…	0	1	2	3	4	5	6	7	8	9	10
1	299 (459)	472 (661)	626 (836)	770 (999)	909 (1153)	1044 (1302)	1176 (1447)	1305 (1589)	1433 (1728)	1559 (1965)	1683 (2000)
2	149 (228)	235 (329)	311 (416)	383 (496)	452 (573)	518 (647)	584 (719)	648 (789)	711 (858)	773 (926)	835 (993)
3	99 (152)	156 (218)	206 (275)	254 (329)	299 (380)	343 (428)	386 (476)	429 (522)	470 (568)	512 (613)	553 (657)
4	74 (113)	116 (163)	154 (205)	189 (245)	223 (283)	256 (319)	288 (354)	319 (389)	350 (423)	381 (456)	411 (489)
5	59 (90)	92 (129)	122 (163)	150 (195)	177 (225)	203 (254)	228 (282)	253 (309)	278 (336)	302 (362)	326 (388)
6	49 (75)	79 (107)	101 (135)	124 (161)	146 (186)	168 (210)	189 (233)	210 (256)	230 (278)	250 (300)	270 (321)
7	42 (64)	65 (91)	86 (115)	106 (137)	125 (158)	143 (179)	161 (198)	178 (218)	195 (236)	212 (255)	230 (273)
8	36 (56)	57 (80)	75 (100)	92 (119)	108 (138)	124 (155)	140 (172)	155 (189)	170 (205)	184 (221)	199 (237)
9	32 (49)	50 (70)	66 (89)	81 (106)	96 (122)	110 (137)	123 (152)	136 (167)	150 (181)	173 (195)	176 (209)
10	29 (44)	45 (63)	59 (79)	73 (94)	85 (109)	98 (122)	110 (136)	122 (149)	134 (162)	145 (174)	157 (187)
11	26 (40)	41 (57)	54 (72)	66 (85)	77 (98)	88 (110)	99 (123)	110 (134)	120 (146)	131 (157)	141 (169)
12	24 (37)	37 (52)	49 (66)	60 (78)	70 (90)	80 (101)	90 (112)	100 (122)	109 (133)	119 (143)	128 (153)
13	22 (34)	34 (48)	45 (60)	55 (71)	64 (82)	74 (92)	83 (102)	91 (112)	100 (122)	109 (131)	117 (140)
14	20 (31)	31 (44)	41 (55)	50 (66)	59 (75)	68 (85)	76 (94)	84 (103)	92 (112)	100 (121)	108 (129)
15	19 (29)	29 (41)	38 (51)	47 (61)	55 (70)	63 (79)	70 (87)	78 (96)	85 (104)	93 (112)	100 (120)
16	18 (27)	27 (38)	36 (48)	44 (57)	51 (65)	58 (73)	65 (81)	72 (89)	79 (96)	86 (104)	93 (111)
17	17 (25)	25 (36)	33 (45)	41 (53)	48 (61)	54 (68)	61 (76)	68 (83)	74 (90)	80 (97)	87 (104)
18	16 (24)	24 (33)	31 (42)	38 (50)	45 (57)	51 (64)	57 (71)	63 (78)	69 (84)	75 (91)	81 (97)
19	15 (22)	23 (32)	29 (39)	36 (47)	42 (54)	48 (60)	54 (67)	60 (73)	65 (79)	71 (85)	76 (91)
20	14 (21)	21 (30)	28 (37)	34 (44)	40 (51)	45 (57)	51 (63)	56 (69)	61 (75)	67 (80)	72 (86)
21	13 (20)	20 (28)	26 (35)	32 (42)	37 (48)	43 (54)	48 (59)	53 (65)	58 (70)	63 (76)	68 (81)
22	13 (19)	19 (27)	25 (33)	30 (39)	35 (45)	40 (51)	45 (56)	50 (62)	55 (67)	59 (72)	64 (77)
23	12 (18)	18 (25)	24 (32)	29 (37)	34 (43)	38 (48)	43 (53)	47 (58)	52 (63)	56 (68)	61 (73)
24	11 (17)	17 (24)	22 (30)	27 (36)	32 (41)	36 (46)	41 (51)	45 (55)	49 (60)	53 (65)	58 (69)
25	11 (17)	17 (23)	21 (29)	26 (34)	30 (39)	35 (44)	39 (48)	43 (53)	47 (57)	51 (61)	55 (66)
26	10 (16)	16 (22)	20 (27)	25 (32)	29 (37)	33 (42)	37 (46)	41 (50)	45 (54)	48 (59)	52 (63)
27	10 (15)	15 (21)	20 (26)	24 (31)	28 (35)	32 (40)	35 (44)	39 (48)	43 (52)	46 (56)	50 (60)
28	10 (15)	14 (20)	19 (25)	23 (30)	26 (34)	30 (38)	34 (42)	37 (46)	41 (50)	44 (53)	47 (57)
29	9 (14)	14 (19)	18 (24)	22 (28)	25 (32)	29 (36)	32 (40)	36 (44)	39 (47)	42 (51)	45 (55)
30	9 (13)	13 (19)	17 (23)	21 (27)	24 (31)	28 (35)	31 (38)	34 (42)	37 (45)	40 (49)	43 (52)
31	9 (13)	13 (18)	17 (22)	20 (26)	23 (30)	26 (33)	30 (37)	33 (40)	36 (44)	39 (47)	41 (50)
32	8 (12)	12 (17)	16 (21)	19 (25)	22 (29)	25 (32)	28 (35)	31 (39)	34 (42)	37 (45)	40 (48)
33	8 (12)	12 (17)	15 (21)	18 (24)	21 (28)	24 (31)	27 (34)	30 (37)	33 (40)	35 (43)	38 (46)
34	8 (12)	11 (16)	15 (20)	18 (23)	21 (27)	23 (30)	26 (33)	29 (36)	31 (39)	34 (41)	37 (44)
35	7 (11)	11 (15)	14 (19)	17 (22)	20 (26)	23 (29)	25 (31)	28 (34)	30 (37)	33 (40)	35 (43)
36	7 (11)	11 (15)	14 (18)	16 (22)	19 (25)	22 (28)	24 (30)	27 (33)	29 (36)	31 (38)	34 (41)
37	7 (10)	10 (14)	13 (18)	16 (21)	18 (24)	21 (27)	23 (29)	26 (32)	28 (34)	30 (37)	33 (39)
38	7 (10)	10 (14)	13 (17)	15 (20)	18 (23)	20 (26)	22 (28)	25 (31)	27 (33)	29 (36)	31 (38)
39	7 (10)	10 (13)	12 (17)	15 (19)	17 (22)	19 (25)	22 (27)	24 (30)	26 (32)	28 (34)	30 (37)
40	6 (10)	9 (13)	12 (16)	14 (19)	17 (21)	19 (24)	21 (26)	23 (29)	25 (31)	27 (33)	29 (35)

and solving for m (by computer) establishes the N values in Table 3 for $P = 5\%$ and $P = 1\%$ [m being the nearest integer not to exceed P].

Example

Taking the data of example (a) page 534: If the sample of 29·9 % of the vessel had returned a count of two infective units, Table 3 should have been consulted at $p = 30$ and $k = 2$, giving N (the 5 % fiducial upper limit of infective units in the remaining 2·8 l. of the culture) as 17, amounting to an upper limit of density of 6·1 infective units per litre instead of the density of three which had been fixed before the sample was taken, assuming it would turn out to be uncontaminated.

Table 3 can readily be extended to any value of P and k using the normal as an approximation to the binomial distribution, involving the solution of the following quadratic equation in m;

$$m^2p^2 - m(2pq + t^2pq) + k^2 = 0, \tag{7}$$

where t stands for the normal equivalent deviate of P (bearing in mind that P refers to a single tail and the numerically greater solution applies).

The authors hope that the present paper, particularly the tables, will prove to be useful in many fields besides microbiology.

SUMMARY

The theory of assigning an upper limit to the estimate of the degree of contamination of a medium is briefly explained. Tables to save computational labour are presented and their use elucidated by examples.

REFERENCES

FISHER, R. A. & YATES, F. *Statistical Tables for Biological, Agricultural and Medical Research.* London: Oliver & Boyd.

Biometrika Tables for Statisticians. Ed. E. S. Pearson and H. O. Hartley. Cambridge University Press.