Quest Journals
Journal of Research in Applied Mathematics
Volume 8 ~ Issue 5 (2022) pp: 04-10
ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Research Paper

Animal or plant pathologies: Number of samples to be collected to know with a given precision if a farm is positive

Jean-Philippe Terreaux

ABSTRACT: *The objective of this document is to determine the minimum number of measurement points (or the cardinal of the sample to be collected) necessary to know, with a given precision, if a structure (field crop farming, arboriculture, stockbreeding, forest property) is affected or not by a pathology. We explain the related demonstration, then we display the results obtained in the form of tables and graphs. This allows the reader to understand how these results are obtained, to understand their scope and limits, to check if they correspond to the problem he has to deal with, and possibly to recalculate them taking into account the particularities he is facing.*

KEYWORDS: *Pathology, disease, animal, plant, livestock, agriculture, arboriculture, forestry, sampling, probability, infectiousness, prevalence, accuracy, cost, benefit*

R*eceived 01 May, 2022;* ***R****evised 10 May, 2022;* ***A****ccepted 12 May, 2022 © The author(s) 2022.*
P*ublished with open access at* ***www.questjournals.org***

I. INTRODUCTION

Animal and plant pathologies have an increasing impact on the production value chain. To give just one example concerning France, the avian flu has led to the elimination of millions of farm animals over the last two winters (2020-2021 and 2021-2022). We can also mention many other threats of epidemics on livestock in Europe and in the world, such as the bovine tuberculosis, the swine fever or various pathologies of bees. Concerning plants, we can mention the pathologies affecting field crops, vineyards and arboriculture (for example, the evolution in European regulations concerning sharka on stone fruit trees may lead to a change in the consequences of this pathology), or the numerous diseases affecting forest trees (ash dieback, bark beetles on spruce trees or the threat of pinewood nematode).

Compared to those weighing directly on agricultural and forestry activities, the impacts on the production chains can be of a different order of magnitude[1]; those on the world food balance can be counted in human lives.

Different means of control exist and are implemented with more or less success. Basically, there is a problem of estimating the presence or absence of the pathology in a farm (animal production, field crops, arboriculture, forestry) so as to be able to integrate this result in the implementation of a control strategy.

This estimation is often expensive (in terms of time spent, laboratory costs, or simply because the measurement is destructive). Also, it is important to know the number of measurement points (or in other words, the sample size) needed to have a given accuracy of the result.

This document has for only objective to indicate this number of points by explaining the corresponding demonstration; then the results are given in the form of tables and graphs. This allows the reader to understand how these results are obtained, to understand their scope and limits, and to check if they correspond to the problem he has to deal with, and possibly to repeat these computations taking into account the particularities he faces. Some results, which might seem paradoxical, can also be explained, for example the fact that an increase in prevalence here leads to a decrease in sampling for a given precision, i.e. the more potentially impacted a farm is (in the sense of an increase in prevalence), the lower the number of samples needed to confirm or refute it; we comment on this in section 4.

[1] For example, for BSE, SARS, H5N1 and H1N1, the costs have been estimated by the World Bank at US$ 20 billion in direct losses and US$ 200 billion in indirect losses, over the first decade of the 21st century. Source: World Bank (2010).

II. DEMONSTRATION ON AN EXEMPLE

Take the example of a farm with *y* animals. The problem we are addressing here is whether the farm is infected with a disease, i.e. whether there is at least one sick animal on the farm.

Assumption: It is assumed that if the farm is affected by the disease, at least x = 5% of the animals are 'sick' or affected. Here we will name this value of x the prevalence. It is possible, for example, to assume that the disease is sufficiently contagious for this threshold to be reached, if the disease is present; another possibility is the simultaneous contamination of several animals by the same external source. We want to demonstrate this positivity with a level of reliability of p = 95% by revealing at least one positive animal.

Question: How many specimens must be tested in order to obtain this accuracy? It is assumed that each specimen allows without error to know if the animal is affected (there is no false positive nor false negative).

Solution: The number of animals in the farm is called *y*. The population is then fictitiously separated into two sets: A, the disease-free animals; and B, the disease-carrying animals (see Figure 1).

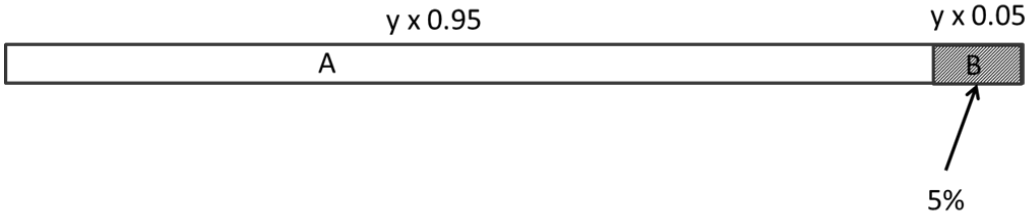

Figure 1 : Total population *y* and number of animals affected by the disease 0.05y

Let N be the number of samples. N must be chosen in such a way that the random drawing of N animals from the whole population (A+B) will only lead to the drawing of these animals in part A with a probability lower than 1-p = 5%.

We then calculate, for a number N, the total number of possible drawings only in part A, and the total number of possible drawings (in the set A+B). The ratio between these two numbers must be less than 5% (100% - 95%, the desired level of reliability).

N.B.: the number of possible drawings of k elements in a population of n (with n > k) is

$$C_n^k = \frac{n!}{k!(n-k)!} \qquad (1)$$

We must have here :

$$\frac{C_{y.0.95}^N}{C_y^N} \leq 0.05 \qquad (2)$$

The figure represents the worst case, in the sense that B contains exactly 5% of the animals (i.e. 0.05 *y*). If B contains more than 5% of the animals (i.e. if more than 5% of the animals are affected by the disease), and if we note u the proportion of affected animals (with 0.05 < u < 1) then we notice that the accuracy of the measurement is at least preserved; it is generally improved:

$$\text{If } \frac{C_{y.0.95}^N}{C_y^N} \leq 0.05 \text{ , then } \frac{C_{y.(1-u)}^N}{C_y^N} < 0.05$$

Thanks to the definition (1), the equation (2) is expressed as :

$$\frac{\frac{(y.0.95)!}{N!\,(y.0.95-N)!}}{\frac{(y)!}{N!\,(y-N)!}} \leq 0.05$$

Or :

$$\frac{\frac{(y.0.95)!}{(y.0.95-N)!}}{\frac{(y)!}{(y-N)!}} \leq 0.05$$

That is :

$$\frac{(y.0.95).(y.0.95-1).\ldots.(y.0.95-N+1)}{(y).(y-1).\ldots.(y-N+1)} \leq 0.05$$

Or else :

$$0.95.\left(\frac{y.0.95-1}{y-1}\right).\left(\frac{y.0.95-2}{y-2}\right).\ldots.\left(\frac{y.0.95-N+1}{y-N+1}\right) \leq 0.05$$

It remains to calculate numerically N. We establish a computer program that starts from the value N = 1, and increases N by 1 at each step, until the above inequation is verified (see the numerical results in Figure 2).

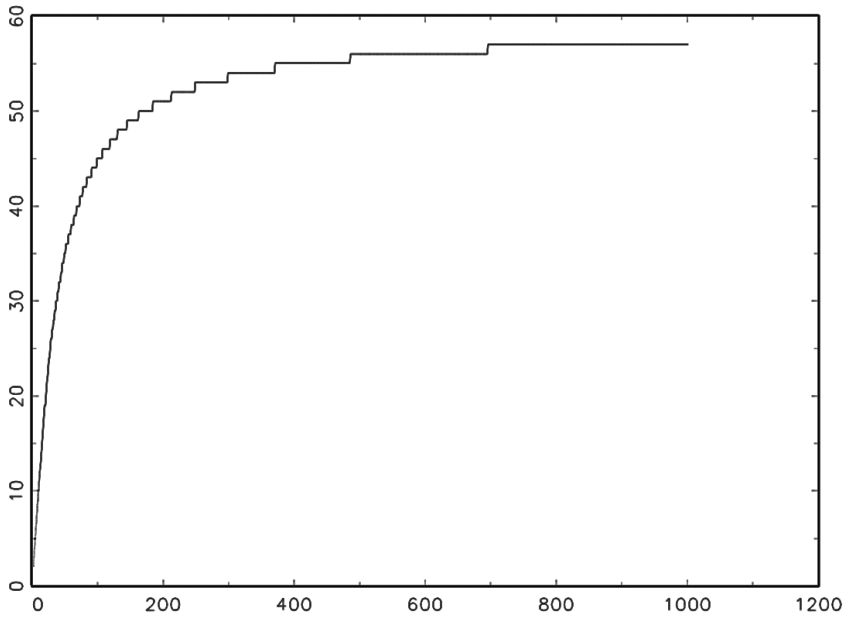

Figure 2: Number N of samples to be taken (y-axis)
as a function of the number *y* of animals on the farm (x-axis). Accuracy obtained: 95%;
If a farm is positive, it is assumed that at least 5% of the animals are affected.
Each sample allows an analysis without false positive or false negative.

The generalization of this demonstration to any population of *y* individuals, of which we want to know with a precision of p% (e.g. 95%) if it is affected by the pathology, knowing that if it is affected, at least x% of the animals are positive, is immediate. In Section 3 we give different results in the form of tables and graphs.

III. RESULTS

3.1. p = 95 % (precision); x = 5 % (prevalence)

number *y* of animals		number N	number *y* of animals		number N
from	to	of samples	from	to	of samples
1	1	1	39	40	31
2	2	2	41	43	32
3	3	3	44	45	33
4	4	4	46	48	34
5	5	5	49	51	35
6	6	6	52	55	36
7	7	7	56	59	37
8	8	8	60	63	38
9	9	9	64	67	39
10	10	10	68	72	40
11	11	11	73	77	41
12	12	12	78	83	42
13	13	13	84	90	43
14	14	14	91	98	44
15	15	15	99	107	45
16	16	16	108	118	46
17	17	17	119	130	47
18	18	18	131	144	48
19	20	19	145	162	49
21	21	20	163	184	50
22	22	21	185	212	51
23	24	22	213	248	52
25	25	23	249	298	53
26	27	24	299	370	54
28	28	25	371	485	55
29	30	26	486	695	56
31	32	27	696	1205	57
33	34	28	1206	4237	58
35	36	29	4238	100000	59
37	38	30			

Table 1 : First and second columns: *y*, the number of animals in the farm.
Third column: N, the number of samples to be taken.
Here p = 95 % ; x = 5 %

3.2. p = 99 % (precision) ; x = 5 % (prevalence)

number *y* of animals		number N	number *y* of animals		number N	number *y* of animals		number N
from	to	of samples	from	to	of samples	from	to	of samples
1	1	1	33	33	31	105	109	61
2	2	2	34	34	32	110	114	62
3	3	3	35	36	33	115	120	63
4	4	4	37	37	34	121	126	64
5	5	5	38	39	35	127	132	65
6	6	6	40	40	36	133	139	66
7	7	7	41	42	37	140	147	67
8	8	8	43	43	38	148	155	68
9	9	9	44	45	39	156	164	69
10	10	10	46	47	40	165	174	70
11	11	11	48	49	41	175	186	71
12	12	12	50	51	42	187	198	72
13	13	13	52	53	43	199	212	73
14	14	14	54	55	44	213	227	74
15	15	15	56	57	45	228	245	75
16	16	16	58	59	46	246	265	76
17	17	17	60	61	47	266	288	77
18	18	18	62	64	48	289	315	78
19	19	19	65	66	49	316	348	79
20	20	20	67	69	50	349	386	80
21	21	21	70	71	51	387	434	81
22	22	22	72	74	52	435	494	82
23	23	23	75	77	53	495	572	83
24	24	24	78	81	54	573	676	84
25	25	25	82	84	55	677	824	85
26	27	26	85	88	56	825	1050	86
28	28	27	89	91	57	1051	1439	87
29	29	28	92	95	58	1440	2265	88
30	30	29	96	100	59	2266	5204	89
31	32	30	101	104	60	5205	100000	90

Table 2 : First and second columns: *y*, the number of individuals in the farm.
Third columns: N, the number of samples to be taken.
Here p = 99% and x = 5%.

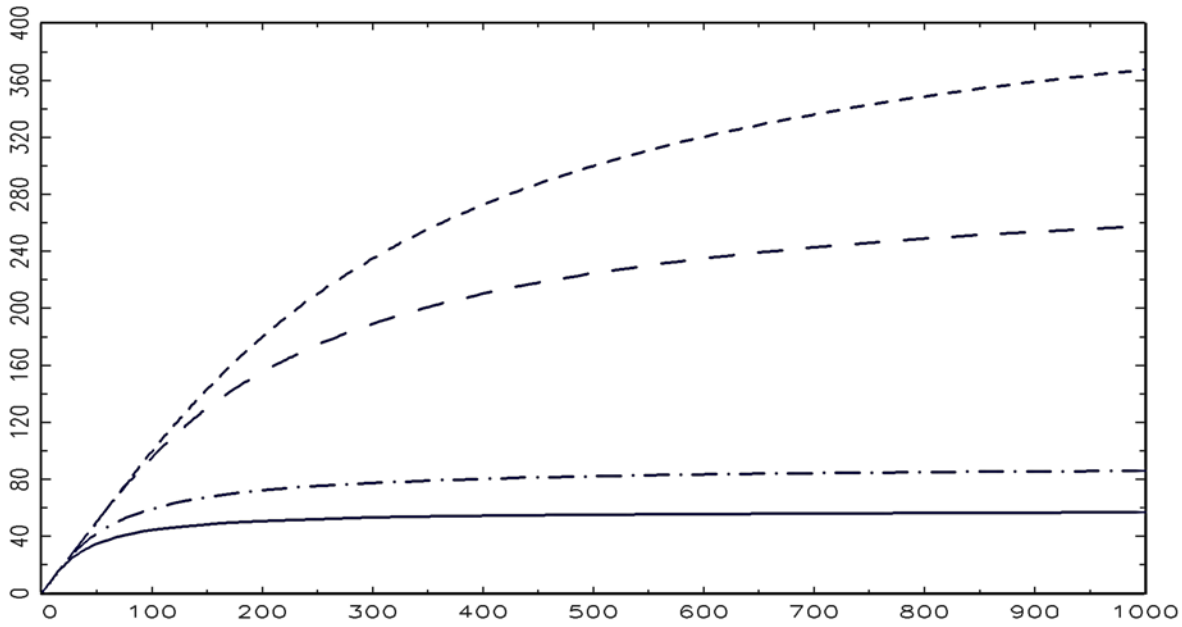

Figure 3: Number of samples to be taken based on prevalence definition and desired precision;
y-axis: number N of samples; From top to bottom :
Small dashed curve: 1% prevalence; 99% accuracy.
Large dashed curve: 1% prevalence; 95% accuracy
Dashed and dotted curve: 5 % prevalence; 99 % accuracy
Continuous curve: 5 % prevalence; 95 % accuracy.
x-axis: Number *y* of individuals, between 1 and 1000

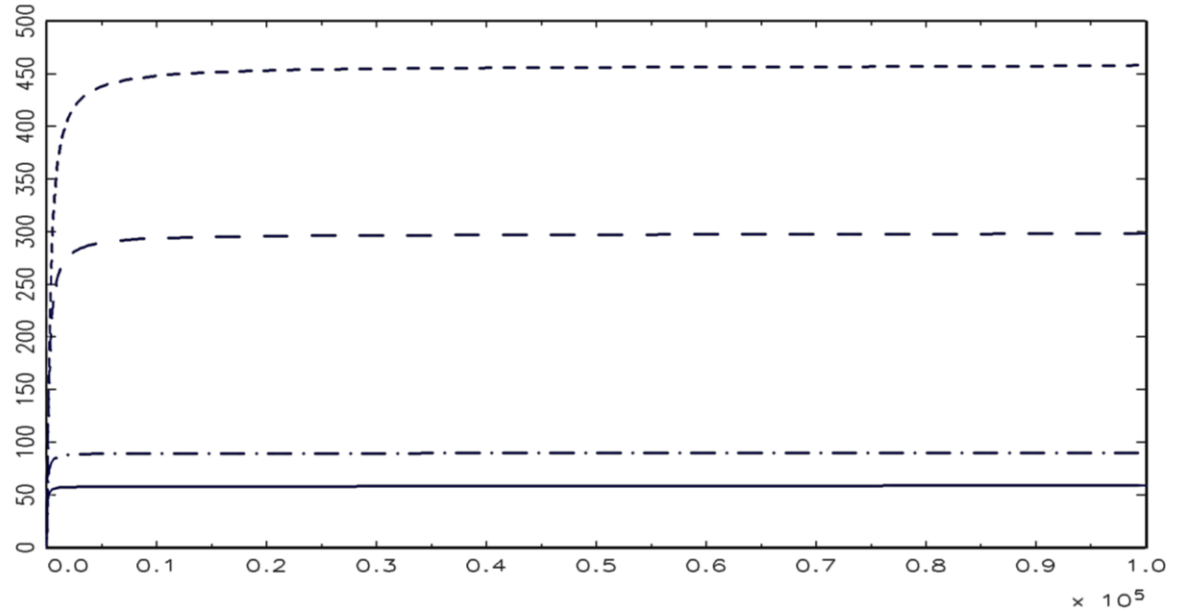

Figure 4: Same curves as Figure 3, with *y* between 1 and 100,000

IV. COMMENTS

It can be seen that the number of necessary samples is more sensitive to the prevalence of the pathology (here 5% or 1%) than to the desired precision (here 95% or 99%): the difference between the curves for prevalence 1% and 5% is greater than the difference between the curves for precision 95% and 99%.

Let us return to the paradox stated in the introduction: the higher the prevalence, i.e., all other things being equal, the more contagious the disease seems to be (i.e., we assume, for example, that if a herd is affected, then systematically at least 5% of the animals are affected; in comparison with another disease for which only 1% of the animals would be infected), the less sampling is necessary; this might seem surprising. But if a disease is particularly contagious, it may be useful to detect it at a very low threshold of infected individuals (i.e., at 1% of the infected population, without waiting for 5%). This leads to an increase in the sampling rate (i.e., to place oneself on one of the two upper curves of Figures 3 and 4, depending on the desired precision). In doing so, we return to the original intuition that the more contagious a disease is, the more samples should be taken.

In addition, however, it should be noted that contagiousness is only one aspect leading to the impacts of the disease: for example, a disease may be relatively uncontagious, but its presence may lead to the closure of borders to export. Therefore, it may be important to detect it at a low threshold (1%), which leads to more sampling (top two curves in Figures 3 and 4).

The final choice of the number of samples depends ultimately on their costs (cost of sampling; time spent; impact on the animal or plant sampled: is the sampling harmful? cost of analysis, etc.) and the economic impact of the imprecision of the measurement, which results in a risk of not detecting a pathology that is present.

ACKNOWLEDGEMENTS

This publication was supported by the MoDerRiSC (Moving towards Deregulation: Risks in Sharka Control) project of the INRAE Metaprogram SuMCrop (Sustainable management of Crop Health, 2021).

REFERENCES

[1]. Mann P.S., 2010, Introductory statistics, 7th edition, Wiley, 750 p.
[2]. Terreaux J. P., 2017, Epizooties et efficacité des processus de décision : un exemple en apiculture, Revue Française d'Economie, 32, 2, 160-197.
[3]. The World Bank, 2010, People, pathogens and our planet, volume 1 : Toward a one health approach for controlling zoonotic diseases, rapport n° 50883, 74 p.
[4]. Wanacott T.H, R.J Wanacott, 1999, Statistique : Economie - Gestion – Sciences - Médecine, 4e edition, Economica, 910 p.
[5]. Weiss N.A., 2012, Introductory statistics, 9th edition, Pearson, 912 p.