Statistical aspects of health monitoring of laboratory animal colonies

by *Axel Kornerup Hansen*
Animal Department, Panum Institute, University of Copenhagen, Copenhagen, Denmark.

Basic principles

Health monitoring of laboratory rodent colonies is based on three principles:

1. A few animals can be sampled for examination, but the results can be used to describe the whole colony.
2. If one animal is found to be infected with a certain organism the whole colony is considered infected with that particular organism.
3. If no animals are found to be infected with a certain organism the whole colony is considered free of that particular organism.

If these principles are to have any validity statistical evaluation of the results must be a demand.

Random sampling

The basic principle No 1 presupposes independence between the animals sampled. One can define the group to sample from, e.g. if a certain change looked for in histopathology is only found in animals of a certain age, all animals sampled should be of that age. Sensu strictu, the results will only be valid for animals of that age, but we assume that the change is not found in animals of the specified age they are not found at all. When one or several of such criteria have been made out, it is of vital importance to sample among the animals fulfilling these criteria in a way that avoids the influence of other criteria, e.g. animals must not be sampled from the same cage, from the same end of the unit etc. If such independence claims are not fulfilled, the conclusion cannot be extended to cover all animals within the unit.

Choise of method and sample size

Prevalence

The fraction of animals in a colony infected at a certain moment is termed the *instaneous prevalence rate* or simply the *prevalence (p)* (*Schwalbe et al.* 1977). The prevalence that a certain infection reaches depends on many factors, e.g. the contact between the animals, the resistance of the animals etc. However, characteristics of the agent itself plays a major role. It is the experience in laboratory animal epidemiology that the observed prevalences of a certain agent normally range within some limits (*Hansen et al.* 1990, *Hansen et al.* 1992, *Hansen* 1992). In this article, it is assumed that this range is independent of the population size. More animals in the colony simply also means more infected animals. So, independent of the colony size we should be able to make a fair guess of the expected prevalence.

A simple calculation of the sample size

If all the infected animals, and only the infected animals, in a population react positively in a given test system, then the risk of reaching a false positive by sampling one animal is

(1) $$1-p$$

This means that if we will only accept this risk to be C we will have

(2) $$C = (1-p)^s$$

where S is the sample size. This gives the formula for the sample size normally used for health monitoring in colonies of laboratory animals (*ILAR* 1976, *Hsu* 1980):

(3) $$S \geq \frac{\log C}{\log (1-p)}$$

If we want our examination to have more than 95 % probability of being correct (*the confidence limit*) C is 0.05. It should be noticed that if certain criteria are lined out prior to sampling, as described under "random sampling", then the prevalence to be used is the prevalence in the group sampled from, and not the over all prevalence in the colony.

False positives and negatives

Formula 3 is a simplified formula, because using the term prevalence it is assumed that this is *the real prevalence,* which means the number of animals actually infected with the organisms. However, in reality is does not matter how many animals are infected, only how many we can expect to react positively in the test system used. E.g. if no infected animals react positively in the system, then the chance of reaching a false diagnosis would be 100 %, and we should find some other test system.

Table 1. Definition of the terms (a) true positives, (b) false negatives, (c) false positives, (d) true negatives.

	Test result +	Test result −
Infected	a	b
Not infected	c	d

Based on table 1 the following formulas can be made (*Wulff* 1976, *Martin* 1977):

(4) *Nosografic sensitivity (N_1)* = $\frac{a}{a+b}$ Fraction of infected animals reacting positively in the test.

(5) *Nosografic specificity (N_2)* = $\frac{d}{c+d}$ Fraction of non-infected animals reacting negatively in the test.

(6) *Diagnostic sensitivity (D_1)* = $\frac{d}{b+d}$ Fraction of negative results caused by non-infected animals.

(7) *Diagnostic specificity (D_2)* = $\frac{a}{a+c}$ Fraction of positive results caused by infected animals.

Sensitivity

Let us use the disease XX as a result of infection with YY as an example. From studying the litterature we know that in infected colonies 20–40 % of the animals will have antibodies. So the lowest observed prevalence is 0.2. Only 25 % of the animals with antibodies will have the disease XX characterized by histopathological changes in the lungs. If we assume that the nosographic sensitivity of serology is 1, which means that every infected animal has antibodies, then the nosographic sensitivity of histopathology of the lungs is 0.25. Now we raise the question: "Is the colony infected with YY?" If we do serology we will have the sample size $S \geq 13.4$ from formula 3. However, if we do histopathology, we will first have to calculate the addendum that the low nosographic sensitivity adds to the probability of reaching a false negative diagnosis, as described by *Adler & Wiggins* (1973):

(8) $$(1-N_1)p$$

which gives a more precise calculation of the probability of reaching a false negative diagnosis than formula 2, as

(9) $$C = [((1-N_1)p) + (1-p)]^s$$

This is the formula to be used for evaluation of health monitoring results. C values above 0.05 should lead to the rejection of a negative result. The sample size needed to achieve C-values less than 0.05 must be calculated as

(10) $$S \geq \frac{\log C}{\log [((1-N_1)p) + (1-p)]}$$

which in this case means that we have a sample size $S \geq 58.4$. The nosographic sensitivity being lower for histopathology means that we will have to sample more animals. However, if we do so, we shall still

have exactly the same confidence limit. By calculating from table 1 we will find, that the diagnostic sensitivity is low, too, which means that we cannot really conclude that an animal without any changes in the lungs was not infected with YY. However, we were only asked the question whether the infection was found in the colony, and that has been answered.

Specificity

We assumed that all animals infected with YY made antibodies. Let us assume that YY cross-react with ZZ, an apathogen bacteria that 20 % of the animals in the colony harbour and make antibodies to. If we fill in table 1 we will find that the nosographic or the diagnostic sensitivity is not changed by this new fact. However, the nosographic specificity is only 0.80, as only 80 % of the animals not infected also react negatively, when tested by serology. The diagnostic specificity is only 0.56, which means, that if we conclude that a seropositive animal is infected with YY, our diagnosis is only correct in 56 % of the cases. If we assume that our colony is only infected with ZZ, and not with YY, we would incorrectly diagnose the presence of YY. So, mathematically, it could be argued, that the low specificity reduces the probability of reaching a false negative diagnosis, which can be expressed by changing formula 9 into (*Adler & Wiggins* 1973):

$$(11) \qquad C = [((1-N_1)p) + N_2(1-p)]^s$$

which means that mathematically the sample size is reduced at a low specificity. However, even though we would reach a positive diagnosis by the lower sample size, the probability that it was correct would also be low, because the diagnostic specificity would be low. So, while the problem of a low nosographic sensitivity can be solved by sampling more animals, a low nosographic specificity is unacceptable under all circumstances. In the example given above we could try to solve the problem by combining e.g. immunofluorescence assay (IFA) and enzyme-linked immunosorbent assay (ELISA). If we assume that ELISA does not cross-react with ZZ, but with QQ, which is also harboured by 20 % of the animals all making antibodies to it, and we say that a positive diagnosis is only made if we find antibodies in both IFA and ELISA, we will find that the nosographic specificity has been raised to 0.95 and the diagnostic specificity has been raised to 0.87. In this case, it still is not good enough. We cannot accept that there is still a possibility of 13 % for a positive result being caused by a non-infected animal. We want the diagnostic specificity to be at least 0.95. So, let us assume that 80 % of the animals infected with YY in IFA have titers above 1:80, while the cross-reacting microorganism ZZ at maximum gives rise to a titer of 1:40. If we raise the cut-off value from the normally used 1:20 to 1:80, we will have a diagnostic and nosographic specificity of 1, while we have reduced the nosographic sensitivity to 0.8. This is, however, acceptable, because this problem can be solved by sampling more animals.

Correlation between specificity and prevalence

Nosographic specificity and sensitivity are independent of the prevalence. By calculation one will find that a given nosographic sensitivity and specificity the diagnostic specificity is raised if the prevalence is raised, which means that in the example given above, the combination of IFA and ELISA had been acceptable, if the prevalence had been higher.

Sampling frequency

A sample taken vizualizes the status at the moment of sampling. As soon as it has been taken, it becomes historical. Curiosity only will dictate when to take the next sample. By comparing two samplings taken at a certain interval the last sampling vill visualize whether changes have occurred between the two samplings.

One could argue that if we must sample 50 animals to reach a confidence limit above 95 %, and we sample 1 animal every week we have reached the confident sample size after a year. If we assume that the infection in question persists when first having entered an animal unit, the calculation is correct, but when we have the result we can only be 95 % sure on the status a year ago, because if the colony was not infected before after a half year of sampling had passed, we shall still have to wait another half year until we have reached the sample size of 50.

Conclusion

Conclusions on the health status of laboratory animal colonies must be based upon a number of independent observations. Negative results should only be accepted if C in formula 9 is less than 0.05. This means that a low nosographic sensitivity is acceptable, if the sample size is raised accordingly. Positive results should only be accepted if the diagnostic specificity is more than 0.95. The results are purely historical.

Summary

Sample size, sampling frequency and the importance of random sampling in health monitoring of colonies of laboratory animals are discussed and the terms *nosographic sensitivity (N_1), nosographic specificity (N_2), diagnostic sensitivity (D_1),* and *diagnostic specificity (D_2)* are explained. It is concluded that test systems with a diagnostic specificity above 0.95 should be chosen, while a low nosographic sensitivity can be accepted, if the sample size (S) is calculated from the formula

$$S \geq \frac{\log 0.05}{\log [((1-N_1)p) + (1-p)]}$$

in which p is the prevalence.

Resumé

Stikprøvestørrelse, prøvetagningsfrekvens og vigtigheden af tilfældig prøveudtagelse ved udførelse af sundhedskontrol i forsøgsdyrskolonier diskuteres, ligesom begreberne *nosografisk sensitivitet (N_1), nosografisk specificitet (N_2), diagnostisk sensitivitet (D_1)* og *diagnostisk specificitet (D_2)* forklares. Det konkluderes, at testsystemer bør have en diagnostisk specificitet over 0.95, mens en lav nosografisk sensitivitet kan accepteres, hvis stikprøvestørrelsen (S) beregnes som

$$S \geq \frac{\log 0.05}{\log [((1-N_1)p) + (1-p)]}$$

hvor p angiver infektionens forventede prævalens.

References

Adler HE & AD Wiggins: Interpretation of serologic tests for Mycoplasma gallisepticum. World Poult. Sci. J. 1973, *29,* 345–353.

Hansen AK, O Svendsen & KE Møllegaard-Hansen: Epidemiological studies of Bacillus piliformis infection and Tyzzer's disease in laboratory rats. Z. Versuchstierkd. 1990, *33,* 163–169.

Hansen AK: The aerobic bacterial flora of laboratory rats from a danish breeding centre. Scand. J. Lab. Anim. Sci. 1992, *19* (2), 59–68.

Hansen AK, F Dagnæs-Hansen & KE Møllegaard-Hansen: Correlation between Megaloileitis and Antibodies to Tyzzer's Disease in Laboratory Rat Colonies. Lab. Anim. Sci. 1992, *42* (5), 449–453.

Hsu CK, AE New & JK Mayo: Quality assurance of rodent models. In *Spiegel A, Erichsen S & Solleveld HA:* Animal Quality and Models in Biomedical Research. Gustav Fischer Verlag, Stuttgart/New York 1980, pp 17–28.

Institute of Laboratory Animal Resources: A report of the committee on long term holding of laboratory rodents. Assembly of life sciences: National Research Council 1976.

Martin SW: The evaluation of tests. Can. J. Comp. Med. 1977, *41,* 19–25.

Schwalbe CW, HP Riemann, CE Franti: Epidemiology in Veterinary Practice. Lea & Fibiger, Philadelphia 1977.

Wulff HR: Rational diagnosis and treatment. Blackwell Scientific Publications. Oxford, London, Edinburgh, Melbourne 1976.