J. Parasitol., 86(2), 2000 p. 228–232

QUANTIFYING PARASITES IN SAMPLES OF HOSTS

Lajos Rózsa, Jenö Reiczigel*, and Gábor Majoros†
Department of Zoology, University of Veszprém, H-8201 Veszprém, P.O. Box 158, Hungary

ABSTRACT: Whereas terminological recommendations require authors to use mean intensity or mean abundance to quantify parasites in a sample of hosts, awkward statistical limitations also force them to use either the median or the geometric mean of these measures when making comparisons across different samples. Here, we propose to reconsider this inconsistent practice by giving priority to biological realism in the interpretation of different statistical descriptors and choosing the statistical tools appropriate to our decisions. Prevalence, mean intensity, and indices of parasite distribution (such as median intensity) are suitable descriptors to quantify parasites in a sample of hosts. These measures have different biological interpretations and need different statistical methods to be compared between samples.

Intensity and abundance (=relative density) are among the most important descriptors one must use when quantifying parasite numbers in a host sample or population. Intensity is defined as the number of conspecific parasites living in (or on) an infected host, and abundance is defined as the number of conspecific parasites living in (or on) any host (intensity > 0, abundance ≥ 0). Recommendations that aim to establish a consistent terminology of ecological parasitology and also to provide a conceptual basis for this discipline strongly recommend the use of mean intensity and mean abundance to characterize samples of hosts (Margolis et al., 1982; Papp, 1987; Bush et al., 1997). Mean intensity is the arithmetic mean of the number of individuals of a particular parasite species per infected host in a sample. Mean abundance is the arithmetic mean of the number of individuals of a particular parasite species per host examined.

To compare mean intensity or mean abundance of parasites obtained from 2 or more different samples, one would use Student's *t*-test or other parametric procedures like ANOVA, if parasite distributions were not known to be aggregated (Crofton, 1971). Thus, parasitologists often prefer to use nonparametric tests like Mann–Whitney's *U*-test or Kruskal–Wallis test that have the advantage of being distribution free but actually compare other characteristics of the distributions instead of means. Consequently, parasitologists speaking about sample means according to terminological recommendations and actually comparing other characteristics of samples may get surprising results. Let us imagine 2 samples of hosts each containing 10 infected individuals:

Sample A: 1, 1, 1, 1, 1, 1, 1, 1, 2, 50;
Sample B: 1, 1, 2, 2, 2, 2, 3, 3, 4, 10.

There is a striking inconsistency between reporting and analyzing the above data, i.e., intensity is 2 times higher in sample A than in sample B (means: 6, 3), whereas a statistical comparison (Mann–Whitney *U*-statistic) will indicate that intensity is significantly lower in sample A than in sample B. Should we change our terminology or the usual statistical analysis? Bush et al. (1997) recommended the use of mean intensity and mean abundance like previous authors (Margolis et al., 1982) but also maintained that ''in some cases, median intensity or modal intensity will be appropriate substitutes for mean intensity'' and they made a similar statement for abundance too. In what sense can median substitute mean?

As an alternative, log transformation may sometimes help to normalize data, though it fails when dealing with highly aggregated parasites. Typically, authors log transform ($\log[x + 1]$) raw values of either intensity or abundance, calculate the mean of the transformed data, and then back transform the mean to obtain the so-called geometric mean. The reason for this procedure is that it hopefully fixes the skewness of the parasite distribution. One should keep in mind that it works perfectly only if the frequency distribution of parasites is log normal, so that it can be normalized by log transformation.

In the present paper, we aim to summarize the properties of the different measures and to argue that their biological interpretation is markedly different. Then, we give basic recommendations on how to provide quantitative data on the occurrence of parasites and how to compare these data. The recommendations outlined below are consistent with Margolis et al. (1982), Papp (1987), and Bush et al. (1997).

PROPERTIES AND INTERPRETATION OF DIFFERENT MEASURES

Imagine that one aims to publish the quantitative results of a faunistic survey on parasites. Definitely, host sample size (N) and prevalence (%) must be provided. Now, there are 6 basic choices for describing parasite quantities, i.e., mean intensity, mean abundance, median intensity, median abundance, geometric mean intensity, and geometric mean abundance. Most parasitologists prefer to use mean intensity and typically they also provide $\pm$SD.

In general, it seems reasonable to prefer intensity to abundance. A sample can be split into 2 parts, uninfected and infected. Prevalence provides information about the relative sizes of these 2 parts, and uninfected hosts cannot be further characterized. Infected hosts can be characterized by intensity that therefore provides information logically independent of prevalence. Abundance, on the other hand, carries information partly in common with that of the prevalence. Additionally, the distribution of intensity may be a little bit less skewed than that of the abundance; therefore, its confidence interval can be a little more precise and informative (see below).

An advantage of preferring means to medians is that given the prevalence, mean abundance can easily be calculated from mean intensity or vice versa. A further advantage is that the

Received 11 January 1999; revised 14 June 1999 and 31 August 1999; accepted 31 August 1999.
* Department of Biomathematics and Informatics, University of Veterinary Science, H-1400 Budapest, P.O. Box 2, Hungary.
† Central Veterinary Institute, H-1581 Budapest 146, P.O. Box 2, Hungary.

expected value of the mean does not depend on sample size (Fulford, 1994) (Fig. 1A, D). Furthermore, mean intensity varies independently of prevalence (Fig. 1G). Thus, providing sample size, prevalence, and mean intensity does not cause redundancy in the information. Mean abundance carries the same information as mean intensity does, but it correlates with prevalence (Fig. 1J). A disadvantage of the mean is that its value is highly dependent on a few, extremely high, individual intensities (the large infrapopulations). Because of the aggregated nature of parasite distributions, a high proportion of parasites is concentrated on a few host individuals, and this may result in a wide confidence interval.

In contrast, preferring medians instead of means may make it necessary to present measures of both intensity and abundance, because median intensity does not predict median abundance and vice versa. Both sample size and prevalence affect median intensity (Fig. 1B, H) as well as median abundance (Fig. 1E, K). Especially, prevalence below 50% implies that median abundance is 0 (Fig. 1K). Thus, it may happen that median intensity or median abundance does not add much if any to the information provided by sample size and prevalence. Although the comparisons of medians are distribution free, medians themselves are not. This is because they are indices that characterize the parasite distribution; thus their values will definitely correlate with other measures of this distribution.

The geometric mean, like the median, is not sensitive to the effects of a few large infrapopulations. It appears to correlate, however, with sample size (Fig. 1C, F) and prevalence (Fig. 1I, L) (Fulford, 1994).

We suggest that the biological interpretation of mean, median, and geometric mean measures should be clearly distinguished. Mean simply refers to the quantity of parasites. Provided that host sample size (N) and prevalence (%) are known, which is a basic requirement, either mean intensity or mean abundance refers exactly to the total number of parasite individuals in the whole sample. On the other hand, the median of either intensity or abundance signifies typical levels of infection. The geometric mean of either intensity or abundance has no clear biological interpretation. It does not give us the total number of parasites nor the most typical intensity class. It is a statistical construction aimed to produce a statistically better distribution but has not much to do with biological reality.

Basically, the most informative way to quantify the occurrence of parasites in a sample of hosts is to describe the frequency distribution of parasites (this is simply a histogram of abundance or intensity). It can either be tabulated or illustrated graphically. If it is not possible to provide it, one has to consider what kind of information is preserved and what is lost when reporting either the mean, or the median, or the geometric mean of intensity or abundance. It may well be reasonable to include several different measures that carry different information about the hosts and the parasites found in the sample and that can be used for different purposes later.

RECOMMENDED WAY OF PRESENTING QUANTITATIVE DATA

Always report host sample size (N) and prevalence (%). Optionally, also give a confidence interval for the prevalence based on the binomial distribution (Bush et al., 1997). Report the mean intensity to give information on the total quantity of parasites in the sample. Providing standard deviation (±SD) of the mean is useless for aggregated distributions exhibited by parasites. A confidence interval for the mean intensity can be constructed by normal theory only if the number of infested individuals is large enough (≥30) in the sample. Otherwise, the bootstrap confidence interval (BC_a) of Efron and Tibshirani (1993) should be used (Appendix I). However, one cannot expect an accurate estimate if the majority of the parasite population is from a very few hosts. Report the frequency distribution of parasites, i.e., the histogram of intensity or abundance. If prevalence is low, a histogram of abundance is less informative. If it is not possible to include histograms, provide a box-and-whiskers plot, or quartiles or percentiles of the distribution, or at least the median intensity to give some information about the distribution. In this case, also give a confidence interval for median intensity (see Appendix II). Geometric mean intensity and its confidence interval can sometimes substitute median as an alternative measure of the frequency distribution of parasites. Provided that the normality assumption holds for the log-transformed data, the confidence interval of the geometric mean can be calculated by standard statistical procedures and then back transformed to the original scale. Also provide a measure to characterize the skewness of the distribution, such as the variance to mean ratio, the exponent k of the negative binomial model (Krebs, 1989), or the discrepancy index as defined by Poulin (1993).

STATISTICAL COMPARISONS

Compare prevalences by chi-square test or, preferably, by Fisher's exact test. Keep in mind that this comparison does not refer to the quantities of parasites, rather it shows whether the proportions of infested hosts are significantly different between the samples.

Compare mean intensities. This comparison may indicate whether the total numbers of parasites are significantly different between the parasitized parts of the samples. Comparisons based on normal theory such as Student's t-test or ANOVA are usually not applicable for parasites because of the skewness of their distribution. Bootstrap tests are recommended (Appendix III), though the power of the test is rather low when the majority of parasites is concentrated on very few hosts.

Compare the frequency distributions of parasite intensities. This comparison indicates whether the typical levels of infection tend to be different between the parasitized parts of the samples. Distributions are compared by chi-square test, or by a generalized version of Fisher's exact test, or by rank-based procedures like Mann–Whitney's U-test and the Kruskal–Wallis test. Randomization tests may also offer useful tools to test differences between distributions (see e.g., Thomas and Poulin, 1997). If medians themselves are of particular interest, Mood's median test (Appendix IV) can be applied for comparisons.

In general, publishing confidence intervals for all measures is advisable because they give information about the uncertainty of the estimates, thus enabling a comparison at a glance. Confidence intervals should preferably refer to the confidence level ≥ 95%, but other levels (say, confidence level ≥ 90%) may be reasonable too.

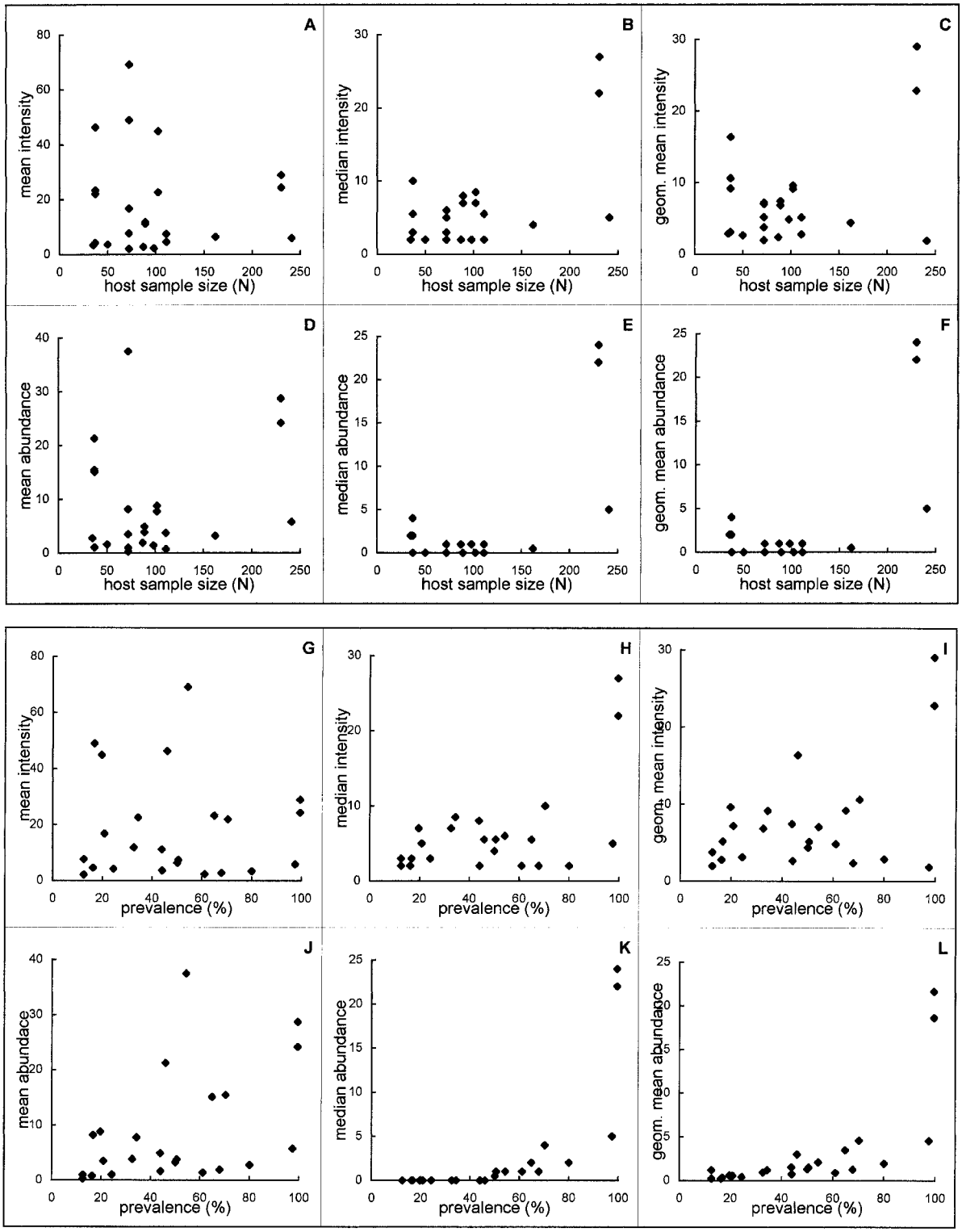

FIGURE 1. Different measures of louse numbers as functions of host sample size (N) and prevalence (%) are illustrated here using 23 samples of avian lice (Phthiraptera: Amblycera, Ischnocera) harbored by 12 samples of birds (Aves). Mean intensity (**A**) and mean abundance (**D**) do not correlate with sample size (linear regression: $P = 0.804$, $P = 0.271$, respectively), while measures of the median (**B**) and geometric mean of intensity (**C**), and the median (**E**) and geometric mean of abundance (**F**) are affected by sample size (linear regression, $P = 0.001$, $P = 0.033$, $P < 0.001$, $P < 0.001$, respectively). Mean intensity (**G**) does not correlate with prevalence (linear regression: $P = 0.936$), while the median (**H**) and the geometric mean of intensity (**I**), and the mean (**J**), the median (**K**) and the geometric mean of abundance (**L**) are affected by prevalence (linear regression: $P = 0.005$, $P = 0.020$, $P = 0.024$, $P < 0.001$, $P < 0.001$, respectively). Data were obtained from Fowler and Miller (1984), Fowler and Hodson (1988), Fowler and Shaw (1989), Fowler and Hodson (1991), Clark et al. (1994), Lee and Clayton (1995), and Rékási et al. (1997).

DISCUSSION

Which sample of hosts is more parasitized? Because the above 3 kinds of comparisons focus on 3 different measures of parasitism that have markedly different biological interpretations, they may yield contradictory answers to the above question. One may say that sample A is more parasitized than sample B if all 3 comparisons correspond with this conclusion. On the other hand, if different comparisons yield contradictory answers, one cannot tell which sample is more parasitized; rather a more specific question and a more complex interpretation of data are needed to describe the difference between levels of parasitism in different samples of hosts.

We believe that the current practice of presenting and comparing quantitative results of parasitological surveys could be improved by considering the biological interpretations of statistical descriptors and comparisons. Recent developments in biostatistics, like randomization tests and bootstrap, offer powerful tools enabling us to choose the ones appropriate to our biological decisions. A computer program for IBM PC to perform the computations described in the appendices is available from the authors.

ACKNOWLEDGMENTS

The first author is supported by a Széchenyi Professorship and by the Hungarian National Scientific Research Fund (F 016792).

LITERATURE CITED

ARNOLD, B. C., N. BALAKRISHNAN, AND H. N. NAGARAJA. 1992. A first course in order statistics. John Wiley & Sons, New York, New York, 279 p.

BUSH, A. O., K. D. LAFFERTY, J. M. LOTZ, AND A. W. SHOSTAK. 1997. Parasitology meets ecology on its own terms: Margolis et al. revisited. Journal of Parasitology **83:** 575–583.

CLARK, F., J. FARREL, AND L. A. HILL. 1994. A study of a population of the house martin (*Delichon urbica* (L.)) feather louse *Brüelia gracilis* (Mallophaga: Ischnocera) in Lincolnshire, UK. Entomologist **113:** 198–204.

CROFTON, H. D. 1971. A quantitative approach to parasitism. Parasitology **62:** 179–193.

EFRON, B., AND R. TIBSHIRANI. 1993. An introduction to the bootstrap. Chapman & Hall, London, U.K., 436 p.

FOWLER, J. A., AND D. HODSON. 1988. The Mallophaga of Leach's Petrels *Oceanodroma leucorhoa* from North Rona, Scotland. Seabird **11:** 47–49.

———, AND ———. 1991. Feather lice (Mallophaga) of the Dipper *Cinclus cinclus* in Central Wales. Ringing & Migration **12:** 43–45.

———, AND C. J. MILLER. 1984. Non-haematophagous ectoparasite populations of Procellariiform birds in Shetland, Scotland. Seabird **7:** 23–30.

———, AND G. J. SHAW. 1989. The Mallophaga of Manx Shearwaters *Puffinus p. puffinus* from Ynys Enlli, Wales. Seabird **12:** 14–19.

FULFORD, A. J. C. 1994. Dispersion and bias: Can we trust geometric means? Parasitology Today **10:** 446–448.

KREBS, C. J. 1989. Ecological methodology. Harper and Row, New York, New York, 654 p.

LEE, P. L. M., AND D. H. CLAYTON. 1995. Population biology of swift (*Apus apus*) ectoparasites in relation to host reproductive success. Ecological Entomology **20:** 43–50.

MARGOLIS, L., G. W. ESCH, J. C. HOLMES, A. M. KURIS, AND G. A. SHAD. 1982. The use of ecological terms in parasitology (report of an ad hoc committee of the American Society of Parasitologists). Journal of Parasitology **68:** 131–133.

PAPP, L. 1987. On certain ecological concepts of parasitology. Parasitologia Hungarica **20:** 17–31. [In Hungarian.]

POULIN, R. 1993. The disparity between observed and uniform distributions: A new look at parasite aggregation. International Journal for Parasitology **23:** 937–944.

RÉKÁSI, J., L. RÓZSA, AND J. B. KISS. 1997. Patterns in the distribution of avian lice (Phthiraptera: Amblycera, Ischnocera). Journal of Avian Biology **28:** 150–156.

SEN, P. K. 1998. Multivariate median and rank sum tests. *In* Encyclopedia of Biostatistics, Vol. IV, P. Armitage and T. Colton (eds.). John Wiley & Sons, Chichester, U.K., p. 2887–2900.

THOMAS, F., AND R. POULIN. 1997. Using randomization techniques to analyse fluctuating asymmetry data. Animal Behaviour **54:** 1027–1029.

APPENDIX I: BOOTSTRAP CONFIDENCE INTERVAL FOR THE MEAN

Confidence intervals based on normal theory perform poorly for skewed distributions, in particular if the sample is small, but BC_a (bias-corrected and accelerated) bootstrap confidence intervals proposed by Efron and Tibshirani (1993) offer a solution even in such cases. Let x denote the sample mean and F its bootstrap cumulative distribution function (c.d.f.) estimated by resampling. The α level endpoints of the BC_a confidence interval for the population mean (e.g., for 95% confidence level, $\alpha = 0.025$ and 0.975 should be used) are calculated by the following formula

$$l^{(\alpha)} = F^{-1}\left\{\Phi\left[z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a(z_0 + z^{(\alpha)})}\right]\right\}$$

where Φ is the standard normal c.d.f., $z^{(\alpha)}$ is its α level critical value, $z_0 = \Phi^{-1}\{F(\bar{x})\}$, and a is calculated from the jackknife values of the sample mean according to

$$a = \frac{\sum_{i=1}^{n} (l_{(.)} - l_{(i)})^3}{6\left\{\sum_{i=1}^{n} [l_{(.)} - l_{(i)}]^2\right\}^{3/2}}$$

where $l_{(i)}$ denotes the jackknife value with the ith sample point deleted, and $l_{(.)}$ denotes the average of all $l_{(i)}$ values.

APPENDIX II: DISTRIBUTION-FREE CONFIDENCE INTERVAL FOR THE MEDIAN

Let $x_1^*, x_2^*, \ldots x_n^*$ denote the ordered sample. A distribution-free confidence interval for the median can be constructed by simply choosing 2 elements x_i^* and x_j^* of the ordered sample. The confidence level for (x_i^*, x_j^*) can be obtained as $p_i + p_{i+1} + \ldots + p_{j-1}$ where the p_k's are the binomial probabilities with parameters n and $p = 0.5$, i.e., $p_k = \binom{n}{k}0.5^n$ (Arnold et al., 1992). A strategy to choose 1 of these possible confidence intervals is to start with the broadest, i.e., with (x_1^*, x_n^*), and then make it narrower and narrower, leaving out points from the 2 ends until we reach the required confidence level (e.g., 95%). As the confidence level decreases in steps when we leave out a sample element (or more elements if there are ties), it is not always possible to construct a confidence interval with exactly the required level, say, 95%.

To illustrate all this, let us have an example with a sample of size 12 (Fig. 2). At each step of narrowing the interval, 1 element is removed from that side where it gains a greater decrease of interval width. For a confidence level of 95%, one

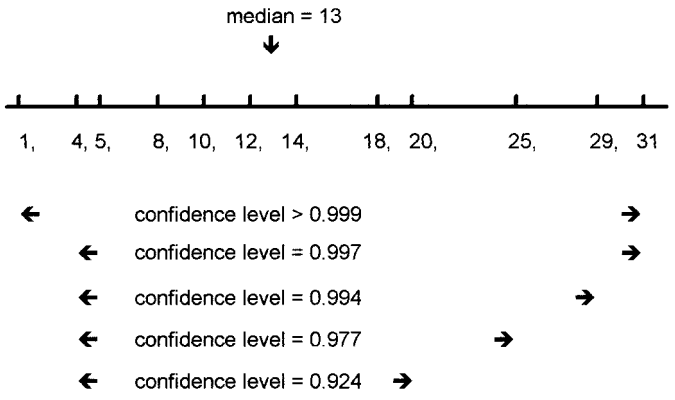

FIGURE 2. Confidence intervals for the median intensity of a sample of 12 infested hosts. At each step of narrowing the interval, 1 element is removed from that side where it gains a greater decrease of interval width. For the confidence level ≥ 95%, one has to outline the 4–25 interval.

has to outline the 4–25 interval, i.e., the confidence interval of the nearest confidence level above 95%.

APPENDIX III: DISTRIBUTION-FREE COMPARISON OF MEAN INTENSITIES

Let us have 2 samples, $x_1 \ldots x_n$, and $y_1 \ldots y_m$ with sample means $\bar{x}$ and $\bar{y}$ and sample variances s_1^2 and s_2^2. Measure the difference between the mean values by the studentized statistic

$$t_{\text{observed}} = \frac{\bar{x} - \bar{y}}{\sqrt{s_1^2/n + s_2^2/m}}.$$

Transform the 2 samples to have a common mean by, e.g., $\dot{x}_i = x_i - \bar{x}$ and $\dot{y}_j = y_j - \bar{y}$. Now we have 2 samples that have equal mean values, so we can use bootstrap resampling to determine the null distribution of the above statistic t by drawing samples of n and m with replacement from the $\dot{x}_i$ and $\dot{y}_j$ values (Efron and Tibshirani, 1993). Let $x_1^{(B)}, x_2^{(B)} \ldots x_n^{(B)}$ and $y_1^{(B)}, y_2^{(B)} \ldots y_m^{(B)}$ denote these samples, $\bar{x}^{(B)}$ and $\bar{y}^{(B)}$ their sample means, and $s_1^{2(B)}$ and $s_2^{2(B)}$ their sample variances. Calculate the above test statistic for the bootstrap samples

$$t_{\text{bootstrap}} = \frac{\bar{x}^{(B)} - \bar{y}^{(B)}}{\sqrt{s_1^{2(B)}/n + s_2^{2(B)}/m}}.$$

The bootstrap P-value is defined as

$$P = \frac{\text{number of bootstrap samples with } t_{\text{bootstrap}} \geq t_{\text{observed}}}{\text{total number of bootstrap samples}}.$$

The use of 1,000 or more samples is recommended.

APPENDIX IV: DISTRIBUTION-FREE COMPARISON OF MEDIAN INTENSITIES BY MOOD'S MEDIAN TEST

Let us have 2 samples again, $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$ and let M denote the median of the combined sample. We are going to test "H_0: the samples come from distributions with equal medians" against "H_1: the samples come from distributions with unequal medians." Note that here, unlike the Mann–Whitney test, we do not need to assume that the 2 distributions are of identical shape. Let us construct the following 2 × 2 table:

.

	≤M	>M	Total
Sample 1	na	$n2 = n - n1$	n
Sample 2	$m1$	$m2 = m - m1$	m

Under H_0, the rows of this table are proportional (except random errors), which does not hold under H_1. The proportionality of rows can be tested by applying a standard chi-square test or preferably Fisher's exact test. The test assumes that the distributions are continuous, i.e., if there are many ties at the median value, it will not work well. The extension to a multisample situation is straightforward. In case of k samples, it leads to a k*2 table where homogeneity of the rows can be tested similarly (Sen, 1998).