Trends in Parasitology

CellPress REVIEWS

Forum

Biostatistics for Parasitologists – A Primer to Quantitative Parasitology

Jenő Reiczigel,[1] Marco Marozzi,[2] Ibolya Fábián,[1] and Lajos Rózsa[3,4,*]

The aggregated distributions of host–parasite systems require several different infection parameters to characterize them. We advise readers how to choose infection indices with clear and distinct biological interpretations, and recommend statistical tests to compare them across samples. A user-friendly and free software is available online to overcome technical difficulties.

> Women were frequent visitors: Humboldt counted the lice in their plaited hair.
>
> Bonpland . . . wanted to know what statistics about lice were good for.
>
> One wanted to know, said Humboldt, because one wanted to know.
>
> Daniel Kehlmann: *Measuring the World*, 2005

The Nature of Host–Parasite Distributions

When collecting a sample of parasites, host individuals typically act as natural sampling units. Consequently, collection is a two-step procedure: first we collect hosts, and then we collect parasite individuals from them. Thus, parasites are practically collected in groups, so-called infrapopulations [1], where group size, expressed as the number of parasite individuals, may be 0 or a positive integer. The occurrence of parasites across members of a host sample (or the whole host population) exhibits a complex pattern that cannot be adequately described by a single measurement or index of infection, but different indices need to be applied that capture more-or-less different aspects of infection.

To describe the distribution of conspecific parasites across host individuals, it is traditional (i) to create infection classes: such as the categories of hosts with 0 parasite, those with one parasite, those with two parasites etc., (ii) to classify each host individual into one of these categories, and then (iii) to draw a histogram to represent either the number or the proportion of hosts belonging to each of these classes. Such frequency distributions do not approximate a normal distribution, but they generally exhibit an aggregated distribution. This means that most hosts have no, or just a few, parasites, and a few hosts have many [2] (except for some strictly controlled experimental infections under laboratory conditions).

This results in two problems. First, unlike normal distributions that can be described by two easy-to-understand parameters representing location and spread [(mean ± standard deviation (SD)], aggregated distributions are characterized by less familiar statistical measures, thus their biological interpretation, and a clear understanding of their properties, require several different measurements. Second, as parasite distributions seriously violate the so-called normality assumption, that the simplest and most commonplace statistical tests [like Student *t* test, analysis of variance (ANOVA), etc.] are based on, their application is inappropriate. Our purpose here is to describe indices recommended to characterize and compare parasitic infections, and to provide a free and user-friendly software to solve the most commonplace biostatistical tasks with parasites.

Apply Only Indices Having Clear and Distinct Biological Interpretations

Some indices used to quantify parasite infection just make no sense. Taking the usual scheme 'mean ± SD' as an example, it results in paradoxical values like '10 ± 15', suggesting that mean intensity can well have a negative value, which is nonsense. This is because 'mean ± SD' is meaningful only for symmetrical distributions, but not for the aggregated ones so characteristic of parasites. Asymmetry of distribution implies that spread differs left and right from the mean, thus one single number cannot adequately characterize it.

Another problematic statistical measure is the geometric mean of intensity or abundance, as it depends on both the total amount and the variability. Thus it may happen that the total amount increases, yet the geometric mean decreases (for instance, geometric mean of 10, 10, 10 is 10 but that of 1, 12, 50 is 8.43). As a consequence, the difference between geometric means has no simple interpretation (either the totals differ, or the variabilities, or both). Sometimes means of log-transformed data are compared by a *t* test or ANOVA and, based on this, a conclusion is drawn for the means of the original data. This procedure is completely invalid, as comparison of means after log-transformation is equivalent with comparison of geometric means, which may well be in a reversed relationship compared to the original means. In general, using parameters that mix several aspects (or those that more or less predict each other) may cause a confusion or, at least, a redundancy of information.

When statistically describing a sample, first, we need to provide sample size. Most authors report the number of hosts as sample size, reflecting the fact that, in most studies, the hosts are

UNCORRECTED PROOF

Q9 Table 1. The Most Important Infection Indices Recommended for the Statistical Description of a Sample of Hosts

Index	Definition	Notes	The uncertainty of the sample value	Refs
Prevalence (also called extensity, especially in the Russian literature)	The proportion of infected individuals within the host sample or population. Expressed as a percentage (0–100%) or as a probability (that a randomly chosen individual is infected, 0–1 range).	We usually get direct information referring to the sample, called a sample prevalence.	Provide a 95% confidence interval (CI) to express the uncertainty of sample prevalence as an estimate of population prevalence. The shortest exact CIs are obtained by Sterne's or Blaker's method.	[3]
Mean intensity	Intensity is the number of parasites found in an infected host. Sample mean intensity is the arithmetic mean or average of intensity values calculated for a sample, excluding the zeroes of uninfected hosts.	Given the aggregated nature of distributions, it does not characterize a 'typical' level of infection, rather it is highly dependent on the presence or absence of a few highly infected individuals. Provided that sample size and prevalence are known, mean intensity defines the total number of parasites in the sample.	95% CI is used to extrapolate sample mean intensity as an estimate of true population mean intensity. Apply bias-corrected and accelerated (BCa) bootstrap CI.	[4]
Median intensity	Median intensity is the median of intensity values calculated for a sample, excluding the zeroes of uninfected hosts.	Sample median intensity is suitable to describe the 'typical' level of infection within a sample.	Due to the discreteness of data, it is often impossible to construct exactly 95% confidence limits. In such cases, report the shortest interval that reaches the desired confidence level.	[5]
Abundance, mean and median abundance	Abundance is the number of parasites found in any host, either infected or noninfected.	Mean abundance combines, thus predicts to a certain degree, two of the former measures: prevalence and mean intensity. It is preferable to provide prevalence and mean intensity separately; abundance is rarely useful (unless we have to express parasite quantity with a single parameter). If prevalence (defined as a probability) is <0.5, median abundance must be 0.	CI calculations for mean and median abundance require the same statistical methods as those used for intensity.	
Crowding, mean crowding	The size of the infrapopulation, to which an individual parasite belongs. It equals intensity; however, while intensity is defined as a host character, crowding is a parasite character. Therefore, mean crowding is obtained by averaging the intensity values over the parasite (rather than host) individuals. Example: if 3 hosts have intensities 1,2,3 (mean intensity = 6/3 = 2) then the 6 parasites have crowding values 1,2,2,3,3,3 (mean crowding = 14/6 = 2.33).	Crowding is a meaningful measure when studying density-dependent parasite characters.	By definition, there are dependencies (ties) between the crowding values. All parasites infecting the same host have the same value and, therefore, all values change simultaneously whenever a parasite is added or removed. This implies that most statistical methods are not applicable. Create 95% CI for mean crowding by the BCa bootstrap method.	[4,6]
Aggregation indices: variance-to-mean ratio, exponent k of the negative binomial, Index of discrepancy	The variance-to-mean ratio of abundance is the simplest measure. It is traditional to approximate the sample distribution of abundances by a negative binomial model. If the model fits acceptably, the exponent k of the negative binomial distribution serves as an index. The index of discrepancy is a modified version of the Gini-coefficient.	These different indices can all quantify aggregation levels. Their interpretations are identical and they more-or-less predict each other.	Even though it is not a widespread praxis, it is advisable to provide a 95% CI for aggregation indices whenever possible (for the negative binomial exponent by maximum likelihood, for the variance-to-mean-ratio and Index of Discrepancy by bootstrap BCa).	[2,7]
Parasite sex-ratio	Expressed as the proportion of males within the sample of adult (sexed) parasites.	It is worth testing whether it differs from equality (0.5) by the exact binomial test. Its correlation with intensity can be expressed using Spearman's rank correlation.	95% CI is constructed in the same way as for prevalence.	

Table 1. (continued)

Index	Definition	Notes	The uncertainty of the sample value	Refs
Parasite species richness	Sample species richness is the number of parasite species found in a sample of hosts.	Sample species richness is likely to be affected by a sample size bias.	Several methods can extrapolate sample values to the true parasite species richness harbored by the whole host population, including the Chao2 estimator. A large sample (300 or more) is needed to obtain a reliable estimate.	[8]

Table 2. The Most Important Statistical Tests for Comparing Infection Indices between Two or More Samples of Hosts

Distribution character (parameter or index) to be compared.	Statistical tests	Refs
Prevalences	Chi-squared test is acceptable for large samples. Fisher's Exact Test is a better choice either for two or more samples. For two samples, a more powerful alternative is an Unconditional Exact Test. It is more sensitive in detecting differences, particularly in the case of small samples (n1, n2 <100).	[9]
Mean intensities or mean abundances	Do not apply Student *t* test or ANOVA (as it often occurs in the literature) because these are based on the normality assumption that is violated by parasite distributions. Log-transformation often fails to normalize data, but even if it does the job, comparison of means on the log-scale would be hard to interpret. Mean intensities or mean abundances can be validly compared by a Bootstrap two-sample *t* test, or by bootstrap ANOVA for >2 samples.	[4]
Median intensities or median abundances	A commonplace method is the nonparametric Wilcoxon–Mann–Whitney *U* test (WMW). Unfortunately, it does not work without imposing additional assumptions on the distributions (same variability, same shape). In general, it tests for some differences between the distributions, so it may happen that the two medians are exactly equal and WMW detects a significant difference (between the distributions), or the other way round, the medians are markedly different and WMW does not notice any difference. Therefore, if differences between medians are of interest, the best choice is Mood's median test.	[10]
Stochastic equality of intensities or abundances of distributions	The bootstrap test for stochastic equality of distributions is a variant of the WMW. It compares pairs of values taken from the two samples and tests whether the probability of getting higher values from one sample than from the other is the same (50–50%) or not. Here, we ask only 'how often' a value taken from one sample is higher than that from the other, but we do not ask 'how much higher'. Therefore, if this test shows that infections in one sample tend to exceed those in the other, it does not necessarily mean that the latter sample hosts fewer parasites.	[11]
Shape of the frequency distributions of intensities or abundances	Intensity or abundance frequency distributions also can be reliably compared by Lepage's Location-scale test. This test is sensitive to any location or scale difference, such as differences between the means, medians, variances, etc.	[12]
Mean crowding	The nonindependence of data makes statistical analysis difficult because one must control for the dependencies between sample values. As bootstrap CIs for mean crowding do that job, tests can be based on them. First, 97.5% CIs are generated for both samples. If these CIs overlap, the difference between the two samples is nonsignificant at the prescribed level of 0.05, that is, $P > 0.05$. The power of this testing method is rather low.	[6]
Aggregation indices	As in the case of mean crowding, these comparisons are also based on testing the potential overlap between 97.5% CIs.	

sampled. However, if both prevalence and mean intensity are provided (see below), the number of parasite individuals can be reproduced. Then, optimally, we should choose indices for describing levels of infection that have clear and distinct biological interpretations, as related to the purpose of our study. The most important ones are summarized in Table 1. The statistical tests recommended to compare these descriptive indices of infection across two or more samples are summarized in Table 2.

Quantitative Parasitology on the Web (QPweb)

To overcome methodological problems, we have published a brief overview of the suitable biostatistical tools together with some new methods proposed by ourselves [13]. That paper was accompanied by a freely distributed software, called Quantitative Parasitology (QP), to make the recommended statistical procedures easily accessible. Subsequent software versions, QP2.0, and QP3.0, followed with an increasing number of functions. These were downloadable software that ran on Windows PCs. They were capable of handling only one type of parasite per host sample, thus multispecies infections or sex-ratios could not be analyzed. Finally, we introduced Quantitative Parasitology on the Web (QPweb) in 2013, that is, an R-based interactive web service

capable of communicating with computers via an internet browser, independently of their operating system. Contrary to former versions, this one is already capable of representing different types of parasite (different species, different sexes, etc.) co-occurring in the same host sample, opening new possibilities for analyzing parasite communities.

Parallel to the introduction of subsequent software versions, we also published new biostatistical procedures potentially useful in characterizing the infection level of a sample or comparing infection parameters across samples of hosts [3,6,9,11]. All of these new procedures became incorporated into the newer software versions. The latest version of QPweb (presently v1.0.13; http://www2.univet.hu/qpweb/qp10/index.php) is freely available on the web for carrying out most of the procedures mentioned above, coming together with a simple user's guide to get through potential technical difficulties.

Quantitative Measures Other than the Number of Individuals

Most authors quantify parasites per host individual by computing the number of parasite individuals. Here, 'number' is often not meant in a strict sense since only adult parasites are typically considered. For example, the number of *Ascaris* worms per human being is usually meant as the number of adult worms per host, excluding the number of eggs and larvae. Alternatively, several authors consider the number of dispersive stages (spores, eggs, etc.) per gram of feces as a proxy of infection intensity, or other quantitative measures related to infection, such as parasite biomass per host individual. Any numerical variable describing infection intensity can be analyzed by QP. Furthermore, QP can also be used to analyze any binary variable indicating the presence of infection.

Avoid Over-interpretation Pitfalls

Ecologists often claim to quantify 'parasite pressure' or 'pathogen pressure' exerted upon host populations, even without clarifying what type of 'pressure' (e.g., a metabolic or a selective pressure) is meant. Unfortunately, none of the above indices can, in itself, reliably indicate any 'pressure' or 'burden'. Low prevalence – to take it as an illustrative example – may occur due to several different causes. For instance, infected hosts may be rare either because infections rarely happen at all or, alternatively, because infections are so highly lethal that infected hosts cannot survive long. Other things being equal, selection pressure is lower in the former, but higher in the latter case. Therefore, low prevalence, in itself, should not be taken as an indication of weak selection pressure upon the hosts.

This example signifies a recurrent threat of over-interpreting quantitative results. Taking ornithologists as an example, they often ask us to tell them which bird species is more 'parasitized' than the other. This is a wrong question, of course; thus, we often have to say 'I do not know' – unless all meaningful measurements and indices show the same direction of difference. What we can tell, however, is that prevalence of parasite species A is highest in host species 1, and median intensity of parasite species B is highest in host species 2, and parasite species richness is highest in host species 3, etc. If our ornithologist colleagues are not yet fully satisfied, we can still ask them 'could you please tell me which forest is more birdized?'

Acknowledgments

L.R. was supported by grant GINOP2.3.2-15-2016-00057.

Disclaimer Statement

The authors declare that they have no conflicts of interest.

[1]Department of Biomathematics and Informatics, University of Veterinary Medicine, Budapest, Hungary
[2]Department of Environmental Sciences, Informatics and Statistics, University of Venice, Venice, Italy
[3]Evolutionary Systems Research Group, Centre for Ecological Research, Hungarian Academy of Sciences, 8237 Tihany, Hungary
[4]MTA-ELTE-MTM Ecology Research Group, Hungarian Academy of Sciences, Budapest, Hungary

*Correspondence: lajos.rozsa@gmail.com (L. Rózsa).
https://doi.org/10.1016/j.pt.2019.01.003

© 2019 Elsevier Ltd. All rights reserved.

References

1. Bush, A.O. *et al.* (1997) Parasitology meets ecology on its own terms: Margolis *et al.* revisited. *J. Parasitol.* 83, 575–583
2. Crofton, H.D. (1971) Quantitative approach to parasitism. *Parasitology* 62, 179–193
3. Reiczigel, J. (2003) Confidence intervals for the binomial parameter: some new considerations. *Stat. Med.* 22, 611–621
4. Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, Chapman & Hall
5. Arnold, B.C. *et al.* (2008) *A First Course in Order Statistics*, SIAM-Society for Industrial and Applied Mathematics
6. Reiczigel, J. *et al.* (2005) Properties of crowding indices and statistical tools to analyze crowding data. *J. Parasitol.* 91, 245–252
7. Poulin, R. (1993) The disparity between observed and uniform distributions: a new look at parasite aggregation. *Int. J. Parasitol.* 23, 937–944
8. Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43, 783–791
9. Reiczigel, J. *et al.* (2008) An exact confidence set for two binomial proportions and exact unconditional confidence intervals for the difference and ratio of proportions. *Comput. Stat. Data An.* 52, 5046–5053
10. Sen, P.K. (1998) Multivariate median and rank sum tests. In *Encyclopedia of Biostatistics*, vol. IV (Armitage, P. and Colton, T. eds), pp. 2887–2900, John Wiley
11. Reiczigel, J. *et al.* (2005) A bootstrap test of stochastic equality of two populations. *Am. Stat.* 59, 156–161
12. Neuhäuser, M. *et al.* (2010) The comparison of mean crowding between two groups. *J. Parasitol.* 96, 477–481
13. Rózsa, L. *et al.* (2000) Quantifying parasites in samples of hosts. *J. Parasitol.* 86, 228–232

UNCORRECTED PROOF