*J. Parasitol.*, 96(3), 2010, pp. 477–481

# THE COMPARISON OF MEAN CROWDING BETWEEN TWO GROUPS

**Markus Neuhäuser, Jana Kotzmann, Maja Walier*, and Robert Poulin†**
Department of Mathematics and Technique, RheinAhrCampus, Koblenz University of Applied Sciences, Südallee 2, D-53424 Remagen, Germany.
*e-mail: neuhaeuser@rheinahrcampus.de*

ABSTRACT: The infrapopulation size, or crowding, is an important measure in parasitology because of its impact on many facets of parasite biology. However, crowding values are difficult to handle statistically because of dependencies. Therefore, a bootstrap procedure was proposed in order to compare crowding between 2 groups or samples of hosts. Here, we propose Lepage's location-scale test based on intensities as an alternative to test whether there is a difference in mean crowding. Simulation results indicate that Lepage's test is superior to the bootstrap, even when sample sizes are large as needed for the bootstrap procedure. Lepage's test, however, can also be applied with moderate or small sample sizes.

In most animal species, individuals occur in groups of different sizes, either because of social behavior or limited habitat availability. The same is true for parasites, for which individual hosts usually act as natural sampling units, each one increasing from a few to many individual parasites. The number of parasites per host is called intensity of infection, and the parasites within a host form an infrapopulation. For the parasite, the infrapopulation size serves as a measure of crowding, i.e., the size of the pool of conspecifics with whom an individual parasite can interact. Crowding is a factor with important implications for several aspects of parasite ecology. For instance, the existence of density dependent regulation may depend on threshold infrapopulation sizes (Poulin, 2007). Moreover, parasite growth and mating probabilities are highly dependent on crowding (Poulin, 2007).

To illustrate crowding, consider sample B of Reiczigel et al. (2005). There are 3 hosts with 1, 4, and 7 parasites, respectively. Mean intensity is (1 + 4 + 7)/3 = 4. To calculate mean crowding, the numerator is no longer the sum of 3 values. Instead it is the sum of 12 values because there are 12 parasites in total. And the denominator of mean crowding is equal to the number of parasites rather than the number of hosts (Lloyd, 1967):

$$\text{mean crowding} = (1\times1+4\times4+7\times7)/(1+4+7) = 5.5.$$

Crowding data consist of non-independent values. In other words, the crowding experienced by a worm inside its host is the same as that experienced by its conspecifics inside the same host. By definition, all worms in the same infrapopulation have the same number of neighbors, and thus their crowding values are not statistically independent of one another. Therefore, classical statistical methods, assuming that data consist of independent events, cannot be applied (Reiczigel et al., 2005, 2008). To apply these classical methods to crowding data where replicates are not statistically independent would be pseudoreplication according to the definition of Hurlbert (1984).

Reiczigel et al. (2005) introduced a bootstrap method in order to calculate a confidence interval for the mean crowding of a group. Based on these confidence intervals, they developed a test to compare the mean crowding of 2 groups. The null hypothesis that crowding is equal in the 2 groups will be rejected if the $(1 - \alpha/2)$-confidence intervals do not overlap.

This bootstrap method should not be applied for sample sizes of fewer than 100 hosts per group (Reiczigel et al., 2005). Thus, an alternative method is needed for the common situation of small, or medium, sample sizes. Moreover, the principle of calculating confidence intervals and checking for overlap usually creates extremely conservative comparisons (Payton et al., 2003). Therefore, the bootstrap method may have low power even in cases of large sample sizes.

Neuhäuser (2009) proposed a procedure that is applicable when individuals can be considered as independent. This test was applied for colony sizes of penguins where every pair seems to choose independently from all other pairs one of the available breeding locations. However, the situation is different for parasite infrapopulations. When, for example, an adult female wasp (*Cotesia* sp.) deposits between 50 and a few hundred eggs into a single host larva (Thompson et al., 2005), several young will live on the same host and might stay there for their entire life. In this case, there is a strong dependence between individual parasites on the same host. Similarly, when several larval parasitic worms inside an invertebrate host are simultaneously transmitted to the same predatory definitive host, their co-occurrence in the same host is not the outcome of independent infection events.

Here, we propose an alternative method to test for differences in mean crowding between 2 groups of hosts that does not need independence between individual parasites. Furthermore, the decision based on the new method does not depend on confidence intervals, and the method can be applied in cases with small samples. Following the description of the proposed method, it is compared with the bootstrap method in a simulation study.

## THE PROPOSED TEST

The population mean crowding $E(C)$ is a function of the population mean $E(I)$ and variance $\mathrm{Var}(I)$ of the intensity:

$$E(C) = E(I) + \frac{\mathrm{Var}(I)}{E(I)}$$

(Lloyd, 1967; Reiczigel et al., 2005). Hence, when comparing 2 groups and finding that there is a difference in $E(I)$, or a difference in $\mathrm{Var}(I)$, or both, one can conclude that there is also a difference in mean crowding, $E(C)$, between the 2 groups. Therefore, one can perform a so-called location-scale test based on the intensity values. In contrast to the crowding values, the intensities can be regarded as independent. This assumption is justified whenever independence between the different hosts exists. Independent intensities are also usually assumed when testing for differences in

---

Received 28 May 2009; revised 21 July 2009, 17 January 2010; accepted 1 February 2010.

* Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany.

† Department of Zoology, University of Otago, Dunedin, New Zealand.

**DOI: 10.1645/GE-2177.1**

intensity (e.g., Rózsa et al., 2000; Neuhäuser and Poulin, 2004). A location-scale test can be applied since it is sensitive to both differences in mean, or differences in variability, or both. Therefore, a location-scale test can be significant if there is, say, a difference between the 2 variances only. Thus, when the location-scale test is significant, there is evidence for a difference in mean intensity, or in variability of intensity, or both. Because of the formula given above, one can additionally conclude that there is a difference in mean crowding. This conclusion is justified because a situation in which there are differences in both mean and variance of intensity, but no difference in mean crowding between 2 groups, is extremely unlikely and can be ignored.

Since parasite intensities are usually not normal in their distribution (Neuhäuser and Poulin, 2004), a nonparametric location-scale test such as Lepage's (1971) test should be chosen. Lepage's test combines the Wilcoxon rank sum test with the Ansari–Bradley test (Hollander and Wolfe, 1999).

Let $n_1$ and $n_2$ be the sample sizes (of observed intensity values) in groups 1 and 2, respectively, $n_1 + n_2 = N$. Hence, $n_1$ and $n_2$ are the numbers of hosts in the 2 groups. The Wilcoxon rank sum is defined as the sum of ranks in group 1, or, more formally, $W = \sum_{i=1}^{N} iV_i$, where $V_i = 1$ if the $i$-th smallest of the $N$ observed intensities is in group 1, and $V_i = 0$ otherwise.

Under the null hypothesis of no difference between the 2 groups, we have $E(W) = (1/2)n_1(N + 1)$ and $\mathrm{Var}(W) = (1/12)n_1n_2(N + 1)$ (Hollander and Wolfe, 1999). Note that $E(W)$ is the expected sum of ranks for group 1 under the assumption that there is no difference between the 2 groups, and $\mathrm{Var}(W)$ is the variance of the sum of ranks under this assumption. In the presence of ties, the variance changes (see Hollander and Wolfe [1999] for a formula that is valid in case of ties).

The test statistic of the Ansari–Bradley test is defined as follows:

$$AB = \frac{1}{2}n_1(N+1) - \sum_{i=1}^{N} \left| i - \frac{1}{2}(N+1) \right| \cdot V_i,$$

under the null hypothesis of no difference between the 2 groups we have (Hollander and Wolfe, 1999)

$$E(AB) = \begin{cases} \frac{1}{4}n_1(N+2) & \text{if } N \text{ is even} \\ \frac{1}{4}n_1(N+1)^2/N & \text{if } N \text{ is odd} \end{cases}$$

and

$$\mathrm{Var}(AB) = \begin{cases} n_1n_2(N^2-4)/(48(N-1)) & \text{if } N \text{ is even} \\ n_1n_2(N+1)(N^2+3)/(48N^2) & \text{if } N \text{ is odd.} \end{cases}$$

Again, the variance changes in the presence of ties (see Hollander and Wolfe, 1999).

The test statistic of the location-scale test according to Lepage (1971) is

$$L = \frac{(W - E(W))^2}{\mathrm{Var}(W)} + \frac{(AB - E(AB))^2}{\mathrm{Var}(AB)}.$$

Under the null hypothesis, $L$ is asymptotically $\chi^2$-distributed with df $= 2$, because $W$ and $AB$ are uncorrelated under the null hypothesis (Lepage, 1971; see also Randles and Wolfe, 1979). This test based on an asymptotic distribution can be performed when sample sizes are moderate or large. When sample sizes are small, an exact permutation test (Manly, 2007) can be applied with Lepage's statistic in order to avoid any asymptotic assumption. Thus, the Lepage test can also be applied in case of small sample sizes. Please note that exact critical values for small sample sizes are tabulated by Lepage (1973) and Hollander and Wolfe (1999).

The location-scale test proposed above can be applied because it uses intensity values, i.e., there is 1 value only per host. Moreover, host individuals can usually be regarded as independent. Lepage's test can be carried out using an R program available at www.rheinahrcampus.de/fileadmin/fb_mathematik_technik/neuhaeuser/Lepage_Test_R.pdf.

There are other situations in which a location-scale test is more appropriate than testing for a difference in location only. This can be the case when homogeneous experimental units are randomly assigned to different treatments or groups (Neuhäuser, 2002). Furthermore, Lepage's test has been used, for instance, to compare the body sizes of males and females of the euphausiid *Pseudeuphausia latifrons* (Hanamura et al., 2004). Lepage's test has also been widely used to analyze data about rainfall and water level (Yonetani, 1993; Matsuyama et al., 2002; Chen et al., 2009).

## COMPARISON OF METHODS

In a simulation study, we compared Reiczigel's bootstrap procedure with Lepage's test. Data from negative binomial distributions were simulated because the distribution of parasite numbers among individual hosts can be adequately described by negative binomial distributions (Crofton, 1971). Formally, a negative binomial distribution arises when the number of failures is modeled as the sequence of Bernoulli trials needed to get $m$ successes, with probability $k$ of success on each trial. We used the same negative binomial distributions as Reiczigel et al. (2005). Hence, $m$ is chosen as 2 and 20, respectively, and $k$ ranges from 0.05 to 0.2. The number of hosts per group ($n_1$ and $n_2$, respectively) ranges from 100 to 500. For each scenario, 1,000 simulation runs were carried out using SAS (SAS, Cary, North Carolina). We simulated independent intensity values.

Figures 1–3 display the simulation results for negative binominal distributions with $m = 2$. Results for $m = 20$ are very similar (not shown, but available from the corresponding author). In all investigated scenarios, the Lepage test is less conservative and more powerful than the bootstrap. The reduced power of the bootstrap procedure may partly be a consequence of the greater conservatism. To investigate this, we simulated Lepage's test with a nominal significance level equal to the actual level of the bootstrap procedure. For example, in case of the situation displayed in Figure 1C, the simulated actual significance level of the bootstrap procedure is 0.014. When we use this value as the nominal significance level of Lepage's test (instead of 0.05), the power of the test decreases, with a shift of 0.05 from 0.54 to 0.36. However, this value is still larger than the bootstrap's power of 0.29.

## DISCUSSION

The infrapopulation size, or crowding, is an important measure in parasitology, but it is difficult to handle statistically because of dependencies. In many helminths, the larger the infrapopulation size,

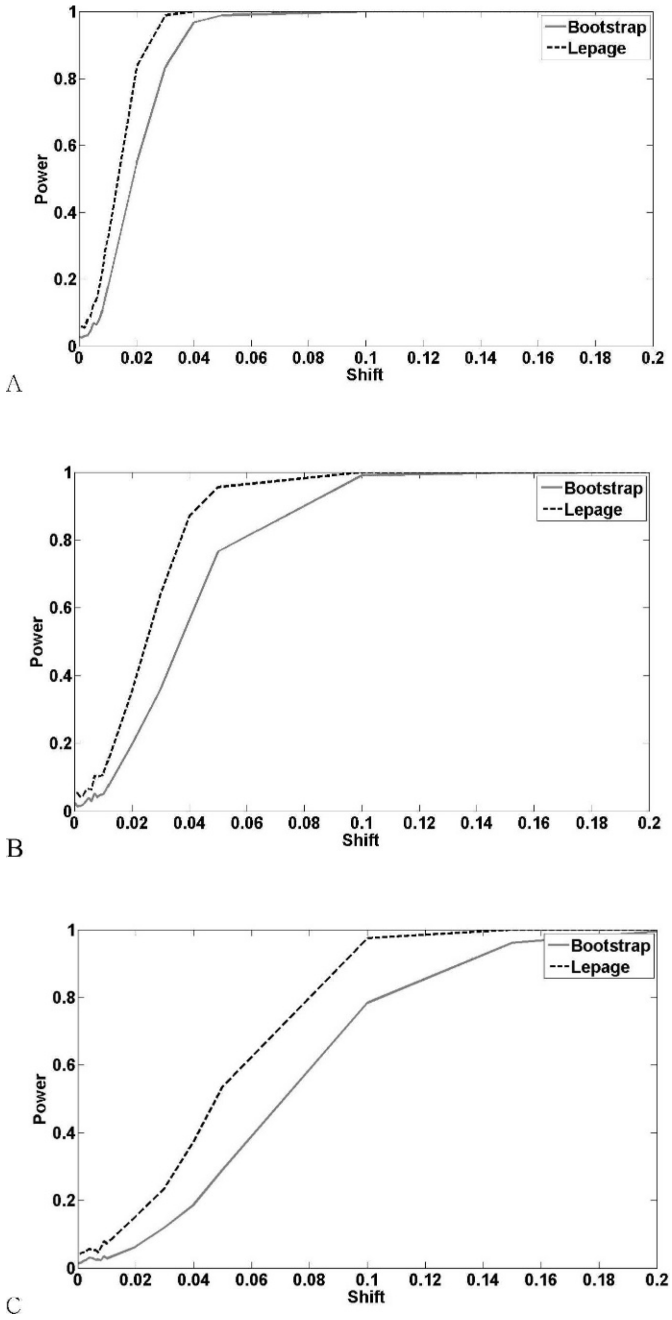

FIGURE 1. Simulated powers of the bootstrap procedure and Lepage's test for different negatively binomially distributed data, $m$ is the number of successes needed in a sequence of Bernoulli trials, with probability $k$ of success on each trial, this probability is shifted by the amount specified on the x-axis, nominal significance level $\alpha = 0.05$, number of hosts per group $= 100$. (**A**) $m = 2$, $k = 0.05$, (**B**) $m = 2$, $k = 0.10$, (**C**) $m = 2$, $k = 0.20$.

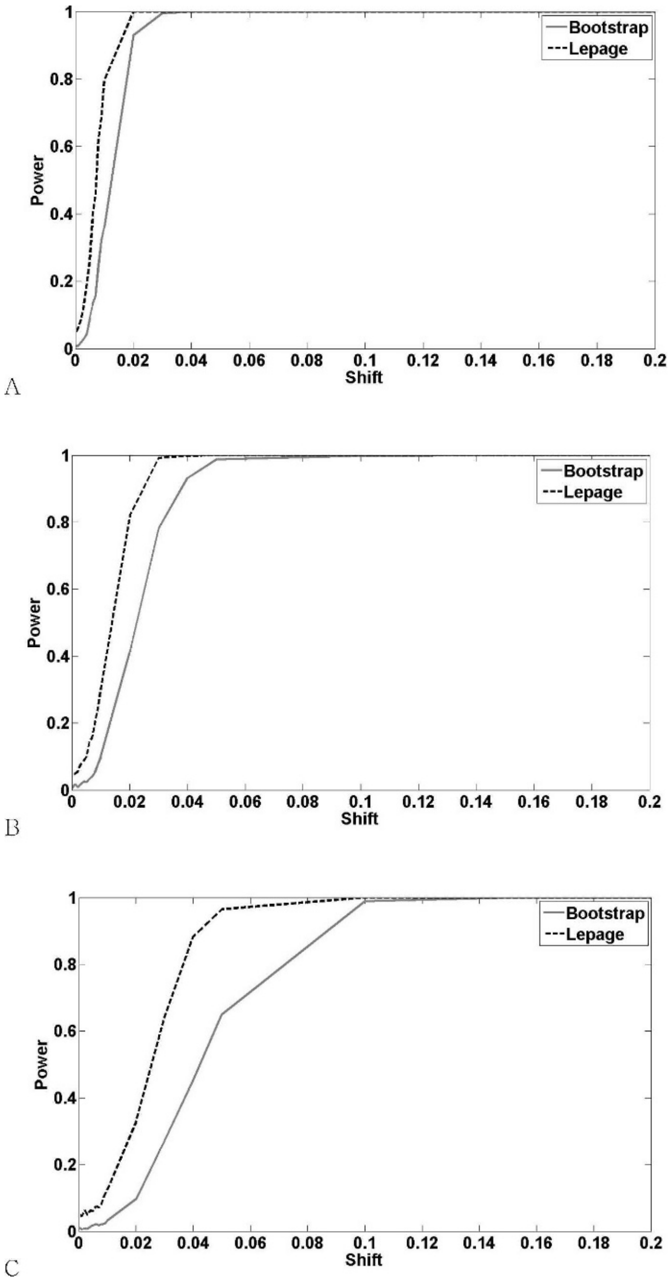

FIGURE 2. Simulated powers as in Figure 1; number of hosts per group $= 300$. (**A**) $m = 2$, $k = 0.05$, (**B**) $m = 2$, $k = 0.10$, (**C**) $m = 2$, $k = 0.20$.

the smaller the average size of individual worms. This is well documented in cestodes within the gut of their definitive host (Read, 1951; Heins et al., 2002; Fichet-Calvet et al., 2003) but occurs also in other helminth groups (Szalai and Dick, 1989; Poulin, 2007). Crowding effects on growth are also evident in larval helminths within their intermediate hosts (Keymer, 1981; Rosen and Dick, 1983; Fredensborg and Poulin, 2005; Saldanha et al., 2009). Because worms in crowded infrapopulations achieve smaller sizes, they also produce fewer eggs on average than conspecifics in smaller infrapopulations (Poulin, 2007). Moreover, crowding can reduce the effective size of the parasite population, i.e., the number of individuals contributing genes to the next generation, by causing a large proportion of worms to remain very small (Dobson, 1986). Crowding resulting from large infrapopulation sizes can, therefore, influence the population dynamics and genetics of parasite populations. In order to assess the efficacy of antihelminthics or other parasite control treatments or the influence of basic ecological conditions such as the presence of other parasite species or differences in host characteristics on parasite infrapopulation sizes, we need robust statistical tests that consider the non-independence of individual data points.

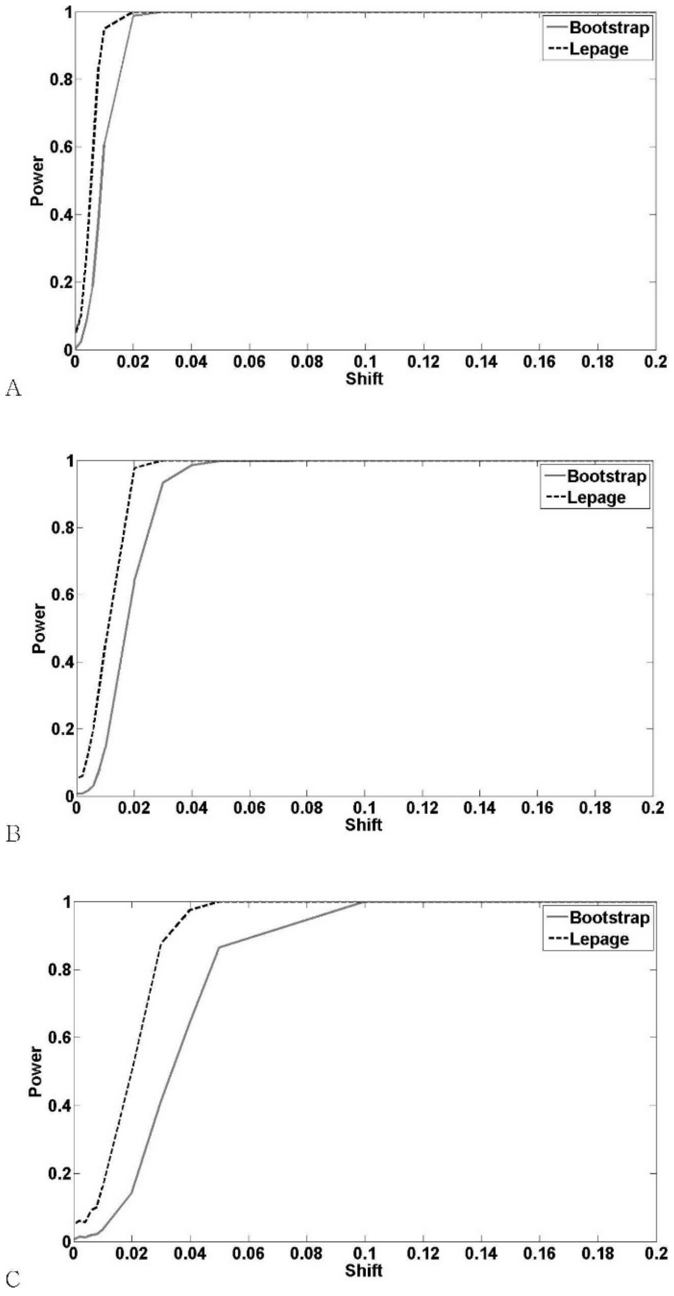

FIGURE 3. Simulated powers as in Figure 1; number of hosts per group = 500. (**A**) $m = 2$, $k = 0.05$, (**B**) $m = 2$, $k = 0.10$, (**C**) $m = 2$, $k = 0.20$.

For the comparison of crowding between 2 groups or samples of hosts, a bootstrap procedure has been proposed (Reiczigel et al., 2005). Because mean crowding depends on the mean and variance of intensity, we propose a location-scale test based on intensities in order to test whether there is a difference in mean crowding. Simulation results indicate that the location-scale test introduced by Lepage (1971) is superior to the bootstrap test. Lepage's test has the additional advantage that it can also be applied in case of moderate or small sample sizes, in contrast to Reiczigel's bootstrap procedure.

Other location-scale tests could be applied; see Goria (1982) for a review of 2-sample location-scale tests. A variety of more recent location-scale tests exists, see, e.g., Neuhäuser and Hothorn (2000), Büning (2002), Murakami (2007), or Marozzi (2008), but Lepage's test seems to be still the standard nonparametric 2-sample location-scale test (Büning, 2002). Lepage's test is also effective when the data do not follow a negative binomial distribution, see, for example, the simulation results presented by Neuhäuser (2002), or more recently by Murakami (2007) and Marozzi (2008).

A location-scale test can also be applied when more than 2 host samples need to be compared. The multi-sample Lepage statistic, as well as a new multi-sample nonparametric location-scale test, are given in Murakami (2008).

We offer a free R program to carry out Lepage's test, as a permutation test in cases with small sample sizes, or asymptotically based on the $\chi^2$ distribution. That program can also be used beyond the comparison of mean crowding values. As mentioned above, a location-scale test can be more appropriate than testing for a difference in location only. Moreover, a location-scale test may be used as a first step before testing separately for differences in location and variability. In such a 2-step procedure, both a location test and a scale test could be additionally performed (without any adjustment of the $\alpha$-level) if the location-scale test in step 1 proved to be significant (Neuhäuser and Hothorn, 2000). However, it should be noted that such a so-called closed test procedure is coherent; that is, a non-significant location-scale test implies the non-rejection of both hypotheses tested in step 2. Therefore, if there are differences in location or variability in conjunction with a non-significant location-scale test, these differences cannot be claimed to be statistically significant with the 2-step procedure.

## ACKNOWLEDGMENTS

The first author gratefully acknowledges support of this work by the Ministry for Education, Science, Youth, and Culture of Rhineland-Palatinate for a Competence Center in Biomathematics. We also thank Andreas Schulz for writing the R program.

## LITERATURE CITED

Büning, H. 2002. Robustness and power of modified Lepage, Kolmogorov-Smirnov and Cramer-von Mises two-sample tests. Journal of Applied Statistics **29:** 907–924.

Chen Y. D., Q. Zhang, C.-Y. Xu, T. Yang, X. Chen, and T. Jiang. 2009. Change-point alterations of extreme water levels and underlying causes in the Pearl River delta, China. River Research and Applications **25:** 1153–1168.

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

Dobson, A. P. 1986. Inequalities in the individual reproductive success of parasites. Parasitology **92:** 675–682.

Fichet-Calvet, E., J. Wang, I. Jomâa, R. B. Ismail, and R. W. Ashford. 2003. Patterns of the tapeworm *Raillietina trapezoides* infection in the fat sand rat *Psammomys obesus* in Tunisia: Season, climatic conditions, host age and crowding effects. Parasitology **126:** 481–492.

Fredensborg, B. L., and R. Poulin. 2005. Larval helminths in intermediate hosts: Does competition early in life determine the fitness of adult parasites? International Journal for Parasitology **35:** 1061–1070.

Goria, M. N. 1982. A survey of two-sample location-scale problem, asymptotic relative efficiencies of some rank tests. Statistica Neerlandica **36:** 3–13.

Hanamura, Y., N. Saito, and K.-I. Hayashi. 2004. Shore stranding on the neritic euphausiid *Pseudeuphausia latifrons* (G.O. Sars, 1883) in western Japan. Crustaceana **76:** 1147–1152.

Heins, D. C., J. A. Baker, and H. C. Martin. 2002. The "crowding effect" in the cestode *Schistocephalus solidus:* Density-dependent effects on plerocercoid size and infectivity. Journal of Parasitology **88:** 302–307.

Hollander, M., and D. A. Wolfe. 1999. Nonparametric statistical methods, 2nd ed. Wiley, New York, New York, 787 p.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs **54:** 187–211.

Keymer, A. E. 1981. Population dynamics of *Hymenolepis diminuta* in the intermediate host. Journal of Animal Ecology **50:** 941–950.

Lepage, Y. 1971. A combination of Wilcoxon's and Ansari-Bradley's statistics. Biometrika **58:** 213–217.

———. 1973. A table for a combined Wilcoxon-Ansari-Bradley statistic. Biometrika **60:** 113–116.

Lloyd, M. 1967. Mean crowding. Journal of Animal Ecology **36:** 1–30.

Manly, B. F. J. 2007. Randomization, bootstrap and Monte Carlo methods in biology, 3rd ed. Chapman and Hall, Boca Raton, Florida, 455 p.

Marozzi, M. 2008. The Lepage location-scale test revisited. Far East Journal of Theoretical Statistics **24:** 137–155.

Matsuyama, H., J. A. Marengo, G. O. Obregon, and C. A. Nobre. 2002. Spatial and temporal variability of rainfall in tropical South America as derived from climate prediction center merged analysis of precipitation. International Journal of Climatology **22:** 175–195.

Murakami, H. 2007. Lepage type statistic based on the modified Baumgartner statistic. Computational Statistics and Data Analysis **51:** 5061–5067.

———. 2008. A multisample rank test for location-scale parameters. Communications in Statistics—Simulation and Computation **37:** 1347–1355.

Neuhäuser, M. 2002. Two-sample tests when variances are unequal. Animal Behaviour **63:** 823–825.

———. 2009. The importance of the biological system underlying the data when choosing a statistical test: Why penguins need to be treated differently to parasites. Animal Behaviour **77:** e1–e3.

———, and L. A. Hothorn. 2000. Parametric location-scale and scale trend tests based on Levene's transformation. Computational Statistics and Data Analysis **33:** 189–200.

———, and R. Poulin. 2004. Comparing parasite numbers between samples of hosts. Journal of Parasitology **90:** 689–691.

Payton, M. E., M. H. Greenstone, and N. Schenker. 2003. Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? Journal of Insect Science **3:** 34.

Poulin, R. 2007. Evolutionary ecology of parasites, 2nd ed. Princeton University Press, Princeton, New Jersey, 332 p.

Randles, R. H., and D. A. Wolfe. 1979. Introduction to the theory of nonparametric statistics. Wiley, New York, New York, 450 p.

Read, C. P. 1951. The "crowding effect" in tapeworm infections. Journal of Parasitology **37:** 174–178.

Reiczigel, J., Z. Zang, L. Rózsa, and B. Tóthmérész. 2005. Properties of crowding indices and statistical tools to analyze parasite crowding data. Journal of Parasitology **91:** 245–252.

———, ———, ———, and ———. 2008. Measures of sociality: Two different views of group size. Animal Behaviour **75:** 715–721.

Rosen, R., and T. A. Dick. 1983. Development and infectivity of the procercoid of *Triaenophorus crassus* Forel and mortality of the first intermediate host. Canadian Journal of Zoology **61:** 2120–2128.

Rózsa, L., J. Reiczigel, and G. Majoros. 2000. Quantifying parasites in samples of hosts. Journal of Parasitology **86:** 228–232.

Saldanha, I., T. L. F. Leung, and R. Poulin. 2009. Causes of intraspecific variation in body size among trematode metacercariae. Journal of Helminthology **83:** 289–293.

Szalai, A. J., and T. A. Dick. 1989. Differences in numbers and inequalities in mass and fecundity during the egg-producing period for *Raphidascaris acus* (Nematoda: Anisakidae). Parasitology **98:** 489–495.

Thompson, S. N., R. A. Redak, and L. W. Wang. 2005. Host nutrition determines blood nutrient composition and mediates parasite developmental success: *Manduca sexta* L. parasitized by *Cotesia congregate*. Journal of Experimental Biology **208:** 625–635.

Yonetani, T. 1993. Detection of long term trend, cyclic variation and step-like change by the Lepage test. Journal of the Meteorological Society of Japan **71:** 415–418.