DOI 10.1515/helm-2017-0054

HELMINTHOLOGIA, 55, 1: 52 – 59, 2018

Optimum sample size to estimate mean parasite abundance in fish parasite surveys

S. SHVYDKA[1], V. SARABEEV[2]*, V. D. ESTRUCH[3], C. CADARSO-SUÁREZ[4]

[1]Department of Mathematics, Zaporizhzhia National University, Zhukovskogo 66, 69063 Zaporizhzhia, Ukraine; [2]Department of Biology, Zaporizhzhia National University, Zhukovskogo 66, 69063 Zaporizhzhia, Ukraine, *E-mail: *vosa@ext.uv.es*, *volodimir.sarabeev@gmail.com*; [3]Research Institute for Integrated Management of Coastal Zones, Department of Applied Mathematics, Polytechnic University of Valencia, Paranimf, 1, 46730 Grau de Gandia, Spain; [4]Unit of Biostatistics, Center for Research in Molecular Medicine and Chronic Diseases (CIMUS), University of Santiago de Compostela, Av. Barcelona, 15782 Santiago de Compostela, Spain

Article info

Received February 09, 2017
Accepted September 26, 2017

Summary

To reach ethically and scientifically valid mean abundance values in parasitological and epidemiological studies this paper considers analytic and simulation approaches for sample size determination. The sample size estimation was carried out by applying mathematical formula with predetermined precision level and parameter of the negative binomial distribution estimated from the empirical data. A simulation approach to optimum sample size determination aimed at the estimation of true value of the mean abundance and its confidence interval (*CI*) was based on the Bag of Little Bootstraps (BLB). The abundance of two species of monogenean parasites *Ligophorus cephali* and *L. mediterraneus* from *Mugil cephalus* across the Azov-Black Seas localities were subjected to the analysis. The dispersion pattern of both helminth species could be characterized as a highly aggregated distribution with the variance being substantially larger than the mean abundance. The holistic approach applied here offers a wide range of appropriate methods in searching for the optimum sample size and the understanding about the expected precision level of the mean. Given the superior performance of the BLB relative to formulae with its few assumptions, the bootstrap procedure is the preferred method. Two important assessments were performed in the present study: i) based on *CI*s width a reasonable precision level for the mean abundance in parasitological surveys of *Ligophorus* spp. could be chosen between 0.8 and 0.5 with 1.6 and 1x mean of the *CI*s width, and ii) the sample size equal 80 or more host individuals allows accurate and precise estimation of mean abundance. Meanwhile for the host sample size in range between 25 and 40 individuals, the median estimates showed minimal bias but the sampling distribution skewed to the low values; a sample size of 10 host individuals yielded to unreliable estimates.

Keywords: fish; *Ligophorus* spp.; mean abundance; optimum sample size; precision; Bag of Little Bootstraps

Introduction

The mean abundance is the most common epidemiological index that quantifies parasites in host samples (Rózsa *et al*., 2000). In many cases, it is useful to know the optimum sample size in order to obtain the population's actual characteristics and their confidence intervals (*CI*s). The estimated parasitological indices are often based on small sample sizes due to high time or monetary constrains, logistical problems associated with host capture or low abundance of some host populations. On the other hand, if the sample size is too large, researchers' additional time, money and other resources might be a wasted effort for minimal gain. Statisti-

* – corresponding author

cal descriptors of fish parasites were calculated based on samples from three individuals to over 1000 (Poiani, 1992; Belghyti *et al*., 1994; Ismen & Bingel, 1999), but it was commonly done without special attention to the effect of sample size on these estimates.

In case of natural infections, parasites typically exhibit an aggregated distribution pattern, with most host individuals harbouring low numbers of parasites and a few individuals hosting too many (Anderson & Gordon, 1982; Shaw & Dobson, 1995; Poulin, 2013). In most cases, the aggregated distribution of the parasites can be fitted to the negative binomial distribution (NBD), where smaller values of the dispersion parameter k ($k < 1$) indicate a highly aggregated distribution attributed to the most macroparasites of wildlife hosts (Shaw & Dobson, 1995). The aggregated nature of parasite distribution affects the mean abundance and the width of its *CI* (Rózsa *et al*., 2000). Therefore, in the determination of the optimum sample size, a compromise between representativeness of the parasitological data from small samples and unnecessary costs for collection excess data should be found. In addition, the determination of the threshold for minimum sample size depends from level of precision that should be well understood and chosen by the researchers. A few studies have focused on the effect of sample size on parasitological parameters (Gregory & Woolhouse, 1993, Jovani & Tella, 2006, Marques & Cabral, 2007). Gregory and Woolhouse (1993) claimed that if parasite sampling is not correctly selected, it may result in artefactual patterns for epidemiologic and aggregation indices. Jovani and Tella (2006) argued that a sample size around 15 fish specimens is enough to get statistically acceptable data for estimating the actual prevalence within a population. Marques and Cabral (2007) examined the effects of sample size on estimates of infection indices and demonstrated that even though samples with less than 40 individuals do not substantially influence parasite prevalence, whereas the mean intensity and mean abundance may be underestimated. Thus, sample size determination is a common problem when dealing with parasitological data.

Karandinos (1976) presented a formula for sample size calculation with precision as a fixed proportion of the mean. The formula has been subsequently developed by Ruesink (1980) for data with distribution patterns ranging from highly clumped to uniform. However, these formulae are based on several assumptions that are often not fulfilled by actual data. An alternative approach to estimating the sample size is to use Monte Carlo simulations and bootstrapping techniques (Efron & Tibshirani, 1993).

The purpose of the present study is to explore the best method to determine the optimum sample size in parasitological surveys to obtain the true values of the mean abundance and their *CI*s. Marques and Cabral (2007) showed the power of Monte Carlo simulation and bootstrap procedures in determining the minimum sample size to estimate the mean abundance, mean intensity and prevalence. However, their work does not address the issue of parasite aggregation in the hosts. Moreover, the precision level and *CI*s for the studied indices were not considered, whereas these aspects are deliberately addressed here. The *CI* can provide information about estimation accuracy of the parasitological indices, where its width is used as a measure of estimation uncertainty and is largely determined by sample size. Marques and Cabral (2007) constrained their work to demersal flatfish and involved two species of cestodes, one acanthocephalan and one copepod. The present study focuses on monogenean parasites of *Ligophorus* spp. from the pelagic flathead grey mullet, *Mugil cephalus* L.

We considered two approaches, analytic and simulation, to estimate an adequate sample size to obtain ethically and scientifically valid values of mean abundance. To achieve this aim, the study is organized as follows. On the first stage, the required sample size was determined using a formula with predetermined precision levels. The second stage involved a simulation study based on applying the Bag of Little Bootstraps (BLB) (Kleiner *et al*., 2014) to empirical data sets and randomly generated parasite distributions in order to assess the optimal sample size as a balance between suitable estimates of the mean abundance and acceptable level of uncertainty in these estimates. Biases and *CI*s were used as measures of accuracy for the simulation modelling. The objective of the present paper is threefold: i) to estimate optimal sample size for parasitological surveys of *Ligophorus* spp.; ii) to evaluate the precision level and *CI*s in samples with different elements, and iii) to test the reliability of each approach to determine the optimal sample size for parasitological studies.

Materials and Methods

Study area, fish sampling and parasite collection

This study is based on 205 dissected individuals *M. cephalus* from three localities, the Kerch Strait, the Sivash Lake and the Balaklava Bay, in the Azov-Black Seas in the period of 2001 – 2013 (Sarabeev, 2015) and one extra sample with 19 fish individuals from the Sivash Lake collected in 2014. The sample sizes ranged between 15 and 35 specimens. Only two-years old, and older fish within the size range of 24 – 65 cm (total length) were used in the analyses. Nine samples were studied across all localities, years and seasons. Collected fish were measured and surveyed for parasites within the day of capture or after freezing. Gills were carefully examined under a stereomicroscope for ectoparasites. All monogeneans were identified and counted. Taxonomic identification was attempted to the species level. Identification of *Ligophorus* spp. followed Sarabeev *et al.* (2013). The present study considers two species of *Ligophorus* from *M. cephalus* across the Azov-Black Sea localities, *L. cephali* Rubtsova, Balbuena, Sarabeev, Blasco-Costa & Euzet, 2006 and *L. mediterraneus* Sarabeev, Balbuena & Euzet, 2005. For each parasite species, samples with more than 6 infected hosts were considered to avoid inadequate estimation of mean abundance due to very low prevalence (Poulin, 2013). Therefore, the data sets included 197 and 192 fish individuals of which 132 and 96 were infected by *L. cephali* and *L. mediterraneus*, respectively.

Data analysis

The mean abundance was calculated according to Bush *et al*. (1997). The distribution pattern of parasite data was characterized by two parameters, using values of parameter *k* of the NBD and parameter *b* of the Taylor's power law (Taylor, 1961) $s^2 = am^b$ in which s^2 is the sample variance, *m* is the sample mean and *a* is a scaling factor related with the sample size. The dispersion parameter *k* was estimated by using the maximum likelihood method (Bliss & Fisher, 1953; Davis, 1994; Young & Young, 1998). The chi-square statistic (Bliss & Fisher, 1953) was used to test goodness-of-fit of the NBD for empirical data.

Through use of a ln transformation the coefficient *a* and the exponent *b* are estimated by the *y*-intercept antilog and the slope, respectively, of the least square regression line of $\ln s^2$ on $\ln m$ as:

$$\ln s^2 = \ln a + b \ln m \,, \qquad (1)$$

using empirical sample means and variances. The values of *a* and *b* were tested for departure from 0 and 1, respectively, by using a two tailed *t*-test (Snedecor & Cochran, 1980). The coefficients of determination R^2 were calculated, to characterize the fit of Taylor's model. For each parasite species 10 mean-variance pairs (9 samples and one aggregated data set) were obtained from empirical samples.

The analytic approach to determine an optimum sample size (*n*) for mean parasite abundance estimation is based on the general formula (Karandinos, 1976):

$$n = \left(\frac{Z_{\alpha/2}}{D}\right)^2 \frac{s^2}{m^2} \,, \qquad (2)$$

in which $Z_{\alpha/2}$ is the standard normal deviate such that $P(Z > Z_{\alpha/2}) = \alpha/2$; *D* is a level of precision and is used to define half-width of the *CI* as a fixed proportion of the mean ($CI/2 = Dm$ (Wilson, 1985). For a 95 % *CI*, $\alpha = 0.05$, then $Z_{\alpha/2}$ equals 1.96.

In the present study the optimum sample size was determined for two precision levels: $D = 0.5$ and $D = 0.8$. These levels are reasonable for practical applications, and are acceptable in most sampling research (Cyr *et al*., 1992; Mouillot *et al*., 1999, Opit *et al*., 2009).

If the dispersion pattern of the target population is well described by the NBD, Karandinos' equation (2) can be rewritten as:

$$n = \left(\frac{Z_{\alpha/2}}{D}\right)^2 \left(\frac{1}{m} + \frac{1}{k}\right) \,. \qquad (3)$$

Incorporating Taylor's power law into Karandinos' equation (2) the sample size model becomes:

$$n = \left(\frac{Z_{\alpha/2}}{D}\right)^2 am^{b-2} \,. \qquad (4)$$

The BLB using the R statistical data analysis software (version 3.3.3, R Development Core Team, 2017) was applied here in a simulation study to determine the optimal sample size. The effect of sample size on *CI* width was tested for: i) two empirical data sets of *L. cephali* and *L. mediterraneus*; and ii) five simulated data sets with fixed mean and variable *k*. The random 1000-dimension samples with the NBD were generated, parameterized by the fixed value of mean abundance equal to 5.55 and exponent *k* of the NBD ranging between 0.1 and 0.9. This range covers most variation of *k* values found for *Ligophorus* spp. (our unpublished data). The random simulation procedure was implemented using the R function *rnegbin()* in the MASS package (Ripley *et al*., 2017). To examine the effect of sample size on mean abundance, a bootstrapping method was applied to generate 95 % *CI*s for given parasite data set: *n* elements from each data set were randomly selected 10000 times, and then we performed a bootstrap with 5000 iterations and computed the mean for each *n*-dimension sample occasion, based on samples from 10 to 100 elements in steps of 5. The 95 % *CI*s for bootstrap were defined using the values that mark the upper and lower 2.5 % of the bootstrap distribution. Bias significance was evaluated through *t*-test. The difference between the estimates of the mean abundance obtained based on different sample sizes was examined by Dunnett's Modified Tukey-Kramer Pairwise Multiple Comparison Test (DTK) from package "DTK" (Lau, 2015) after a logarithmic transformation of the data. The DTK test allows to conduct a pairwise multiple comparison test for mean differences with no assumption of equal population variances. A significance level of 0.05 was used for all test procedures.

According to the purpose of the study, several criteria were used to determine the appropriate sample size: the fit of empirical data to the theoretical distribution; the desired precision (*CI*s width for mean abundance); the achievement of minimal bias and the comparison of mean abundance differences based on different sample sizes. All criteria had to be met in order to accept a given *n* as the minimum sample size needed for estimation of the parasite mean abundance.

Table 1. Summary data for samples of *Mugil cephalus* surveyed from the Azov-Black Seas with information on abundance, variance and aggregation indices of two helminth species (ln *a*: *y*-intercept, *b*: slope, *SE*: standard error, *t*: *t*-test result, R^2: coefficient of determination, *k*: negative binomial parameter, χ^2: chi-square statistic).

	Mean abundance	Variance	ln *a* (SE, *t*-value)	*b* (SE, *t*-value)	R^2	*k* (SE, χ^2)
Ligophorus cephali	15.65	1908.92	-0.16 (1.11, -0.14*)	2.64 (0.43, 6.1)	0.84	0.25 (0.03, 26.02)
L. mediterraneus	5.55	221.25	0.98 (0.67, 1.47*)	2.33 (0.42, 5.59)	0.82	0.21 (0.03, 2.48**)

*No significantly differs from 0 ($P>0.05$). **The NBD model fits to data ($P>0.05$)

Results

The dispersion pattern of both helminth species could be characterized as a highly aggregated distribution with the variance being substantially larger than the mean values. Obtained values of *k* were lower than 1, also indicating on a highly aggregated distribution of these species in the host (Table 1). The chi-square test revealed that *L. cephali* data set does not fit to the NBD, thus not allowing to determinate the optimal sample size using parameter *k* of the NBD in formula (3). For both parasite species, the ordinary least square regression showed a very strong relationship between the means and variances (R^2 =0.82 and 0.84, P<0.0001) with values of b>2 that also indicates a high degree of aggregation. Because the slope b exceeded 2, formula (4) based on Taylor's power law could not be used (Shelton & Trumble, 1991). For *L. mediterraneus*, minimum sample sizes needed to reach the predetermined precisions D=0.5 and D=0.8 based on formula (3) are 76 and 30, respectively.

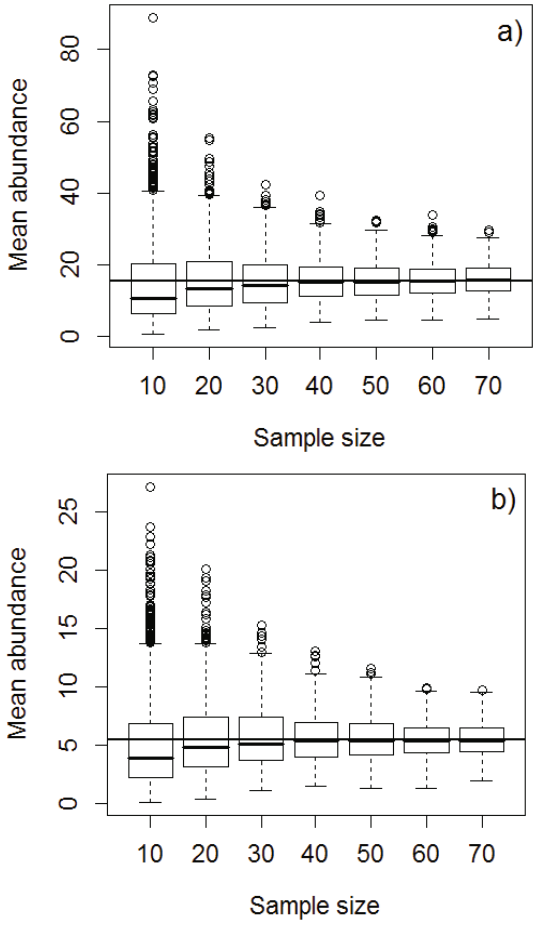

Fig. 1. Distribution of mean abundance obtained from empirical parasite data sets by BLB for different sample sizes for *Ligophorus cephali* (a) and *L. mediterraneus* (b). The box spans the first and third quartiles; the median is marked inside the box by thick horizontal line; minimum and maximum values excluding outliers (whiskers) and outliers (circles); the straight line is the empirical mean abundance.

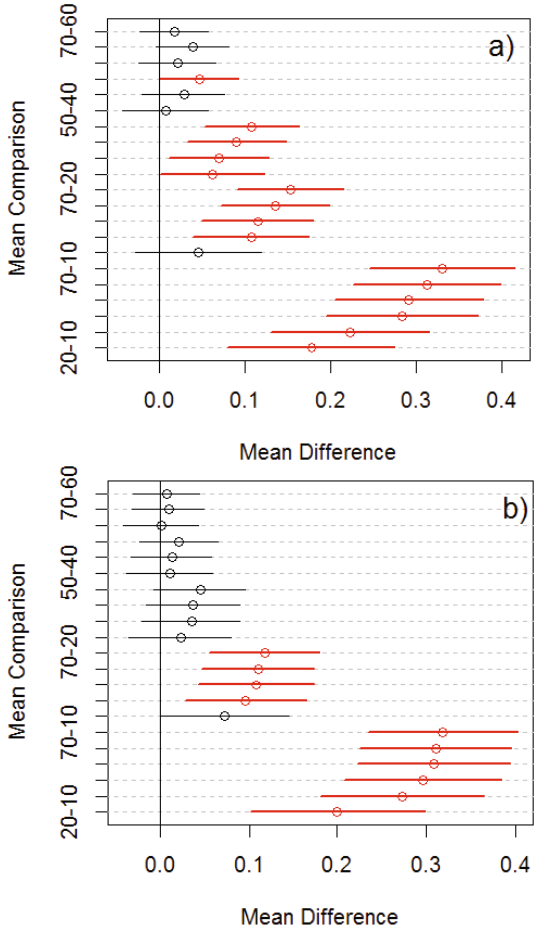

Fig. 2. Simultaneous confidence intervals for all pairwise comparisons of group means. Intervals were computed by the Dunnett's Modified Tukey-Kramer Pairwise Multiple Comparison Test for the mean abundance data of *Ligophorus cephali* (a) and *L. mediterraneus* (b) across different sample sizes. If the interval does not include a zero, the corresponding means are significantly different.

For both parasite species, the mean abundance values obtained by the BLB were close to empirical values of the mean, and no significant biases were found in the estimates. The distribution of mean abundance estimates obtained by simulations was highly right-skewed and the median values were always under-estimating the empirical value at low sample size (Fig.1). The results showed overlapping between the medians of bootstrapping means and the empirical mean abundance beginning from the sample with 40 elements for both examined species. The pairwise statistical comparison between mean abundance values across sample sizes using the 95 % *CI*s is represented in Figure 2. Following the TDK test the estimates related to sample sizes up to 30 and 20 specimens were convincingly different from all others, while there were moderate differences between samples over 40 and 30 fish specimens for *L. cephali* and *L. mediterraneus*, respectively.

The bootstraped 95 % *CI*s were non-symmetric, which correspond to the asymmetry of the underlying mean parasite abundance distributions, and became narrower as sample size increased for both

parasite species (Fig. 3). The effect of sample size on *CI* width as a fixed proportion of the mean for variable values of *k* is shown in Figure 4. The simulation results revealed that as *k* decreases *CI* width was more strongly affected by sample size. The width of the 95 % *CI* is not markedly narrowed with increases in sample size for samples above 25 for *k*=0.9, 30 for *k*=0.5, 40 for *k*=0.2, 45 for *k*=0.15 and 50 for *k*=0.1 elements. For empirical data sets of both species studied here, the largest decrease in *CI* width (exponential phase) was found for sample sizes below 40 individuals, while the further increase of samples resulted in a slow linear decrease in *CI* width (linear phase). This means that the *CI* markedly decreases with increasing sample size up to ca. 70 – 80 specimens. However, further increase of sample size did not really narrow the *CI*. The width of the 95 % *CI*s was decreased from values (1.6 x mean) for sample size with 35 elements to (1 x mean) for sample size with 70 elements in both model species. For *L. cephali* and *L. mediterraneus*, which have close values of the parameter *k* (0.25 versus 0.21) and different values of the mean abundance (15.65 versus 5.55), the variation of the *CI* width was either small or negligible.

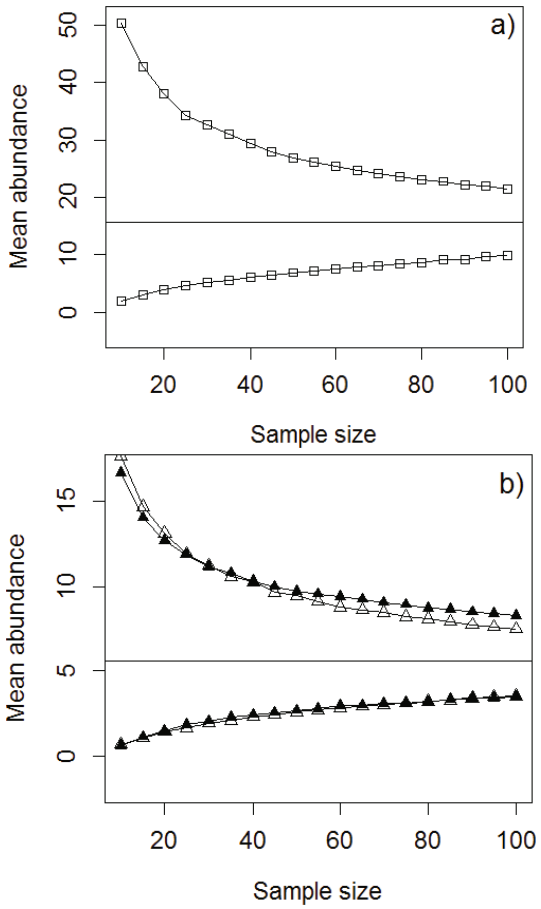

Fig. 3. Mean abundance and its 95 % *CI*s calculated by BLB for different sample sizes for *Ligophorus cephali* (a) and *L. mediterraneus* (b); bootstrap *CI*s based on empirical data set for *L. cephali* (open square); bootstrap *CI*s based on randomly generated and empirical data sets for *L. mediterraneus* (filled and open triangle, respectively); the straight line is the empirical mean abundance.

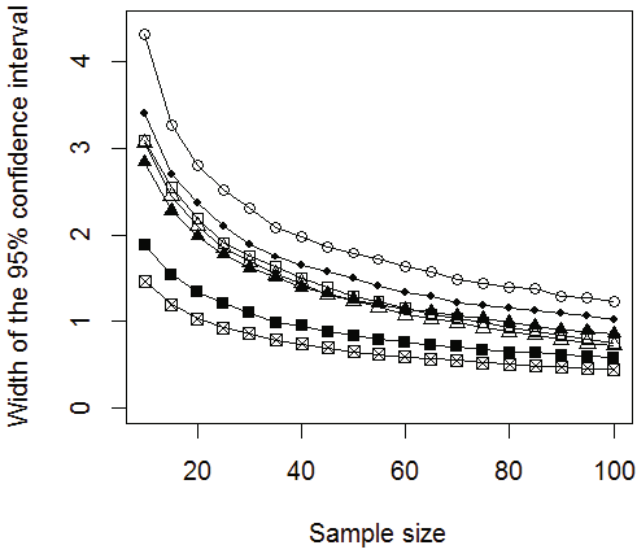

Fig. 4. Width of the 95% CIs as a fixed proportion of the mean abundance determined by BLB from the empirical parasite data sets for *Ligophorus mediterraneus* and *L. cephali* (open triangle and open square, respectively) and randomly generated data with fixed mean abundance of 5.55 and variable *k* of 0.1, 0.15, 0.2, 0.5 and 0.9 (open point, filled point, filled triangle, filled square and square cross, respectively).

Discussion

In the present study, the BLB analysis showed that the minimum required sample size depends greatly on the actual aggregation of the parasite population. The higher degree of variability in the size of parasite infrapopulation, the larger sample size needs to be examined in order to obtain the true value of the mean abundance (Wilson *et al*., 2002). On the other hand, the measure of aggregation will tend to underestimate true aggregation in small samples. This is because heavily infected hosts are rarely found in wild populations and therefore, most likely the probability to be observed in small sample sizes is low (Poulin, 2013). Similarly, the mean abundance calculated from low sample sizes will be underestimated if we do not account for the distribution tail (Marques & Cabral, 2007).

The mean abundance estimates should be reported along with *CI*s, which will allow researchers to assess the biological significance of presented findings (Steidl *et al*., 1997). From a practical point of view, the level of precision is the dominant factor in determining the sample size. Following Buntin (1994), one of the ways to determine the precision is to express it as a confidence interval such that the estimate of the mean should be within a certain value of the true mean with a given probability. Most investigators prefer narrow *CI*s that require large sample sizes for aggregated populations. It stimulates researchers to look for the balance between the limitations of the time and effort required for sample collection, on the one hand, and the essential degree of precision of parasitological indices on the other hand. Because of heterogeneity in parasite infection, it is difficult to apply a theoretical approach for

this purpose. Therefore, the simulation bootstrap procedure based on an empirical data set is a much more robust tool. For small samples, the 95 % bootstrap *CI*s for estimates of the mean abundance are typically very large and skewed upwards. The exponential decrease in the *CI* width as sample size increases indicates the rapidly decrease in the level of uncertainty, and in this way, sample size becomes reasonable for estimation of mean abundance. The further slow linear decreases in the *CI* with sample size increases could be explained by a high total number of non-zero values in such samples. In the example of *L. mediterraneus* and *L. cephali* it was shown that the *CI* becomes more precise and less skewed upwards when sample size is between 35 and 70 fish individuals. Possibly for the reason that such samples are less variable in a number of parasite individuals per host.

Depending on the study aims, researchers may seek higher confidence with a wider interval. For *Ligophorus* species, the reasonable precision level could be chosen between 0.8 and 0.5. If the purpose is to get a general idea about the population abundance, the sample size required could be reduced by lowering the level of precision to *D*=0.8 (Opit *et al*., 2009). If this lower level of precision is used, the sample size with 35 fish specimens is recommended for the estimation of mean abundance for both studied here *Ligophorus* spp. However, for highly aggregated populations, the sample size needs to be sufficiently large to provide a statistically acceptable data for estimation of less abundant parasites (Fig. 4). According to Rózsa *et al*. (2000), the mean abundance is strongly dependent on a few heavily infected individuals; therefore, more specimens may be needed to improve the *CI*s.

Obtained minimum sample size did not substantially depended on the mean abundance of the studied monogeneans, although the difference was about three times fold. The similarity in the sample size required for studied model species could be related to their congeneric relationships that could have the effect on the parasite dispersion pattern. The distributions of both examined species are characterized as a highly aggregated with close values of parameters *k* and *b*. Metazoan gill parasites of fish form non-saturated, multispecies and rich infracommunities in which aggregation ensures cross-fertilization and was found to be an important factor determining the distribution on the gills (Rohde *et al*., 1995; Bagge *et al*., 2005). Monogeneans tend to be more aggregated at lower abundances, what happen because more aggregation is needed as the distance to a potential mate increases with decreasing number of conspecifics (Bagge *et al*., 2005). Our results for the required minimum sample size are in accordance with those of Marques and Cabral (2007) obtained for a system of flatfishes and their parasites.

The values of sample size obtained by the analytical formula show a good correlation with estimates based on the simulation technique. However, the *CI*s based on normal theory are less accurate for skewed distributions, in particular for cases where sample sizes are small (Rózsa *et al*., 2000). Therefore, by utilization of the BLB method a more precise bootstrap *CI*s can be obtained.

The results from this study allowed a direct comparison of sample size estimation by two approaches, analytic and simulation. The advantage of using formulae is the possibility to analyze effects of the precision level, mean abundance, parameter *k* of NBD and parameters of Taylor's model on sample size. The most apparent weakness of the analytic approach is the requirement of the fit of sample data to the theoretical distribution. The application of the formula (3) with parameter *k* of the NBD requires that the *k* value was estimated accurately. For highly aggregated parasite populations, according to the formula (3), the sample size strongly depends on the dispersion of value *k* and the precision of *D*, while the mean abundance >1 does not significantly affect the sample size. Although the Taylor's power law has been widely used due to its statistical stability, the formula (4) for sample size calculation based on the Taylor's model is useful only when $b<2$. This is a significant limitation (Shelton & Trumble, 1991) because the aggregation leads to an increase of coefficient *b* to the critical value 2, or in some cases more than 2. Since the analytic approach is often impossible to apply the non-parametric BLB method is preferable for optimum sample size determination. The primary advantage of bootstrapping is that no assumptions are made on the distribution of the initial data set. Researchers need to assume only that the sample data are independent and representative of the population. The accuracy of estimates obtained by bootstrapping depends on the number of observations in the original sample and the number of resamples. Obviously, large samples are likely to be more representative than small samples.

Conclusions

The holistic approach applied here offers a wide range of appropriate methods to sample size computation and to understand the expected precision level for the mean. While the formulae for sample size estimation may not be very meaningful in practice, their value is that they can provide some strategy in sampling plan before a study. Monte Carlo simulations and bootstrap procedures are powerful techniques for sample size determination. In case of small samples, bootstrapping methods are especially useful to compute the descriptive statistics with associated *CI*s. Such approach is reasonable when dealing with critically endangered species for which low sample sizes are often unavoidable. Regarding sample sizes for parasite data sets with a highly aggregated dispersion pattern, sample size equal 80 or more host individuals allows accurate and precise estimation of mean abundance, whereas for the host sample size in range between 25 and 40 individuals, the median estimates showed minimal bias but the sampling distribution skewed to low values. A sample size of 10 host individuals yields to unreliable estimates, particularly for highly aggregated parasite data sets. These findings will help guide prospective design of sampling plan and will aid researchers in understanding the precision level for the estimated mean abundance in parasitological surveys. At the same time, for the studies aimed to com-

pare epidemiological parameters the question about the optimum sample size remains open. Therefore, the next studies should be focused on the investigating the optimum sample size for comparative studies in parasitology and epidemiology.

Acknowledgements

SS and VS were supported by MEDEA project fellowships, Erasmus Mundus Action 2. CC-S was funded by project #MTM2014-52975-C2-1-R:"Inference in Structured Additive Regression (STAR) Models with Extensions to Multivariate Responses. Applications in Biomedicine", cofinanced by the Ministry of Economy and Competitiveness (SPAIN) and by the European Regional Development Fund (FEDER). This study is partially supported by Ministry of Education and Science of Ukraine, project #1/17.

References

Anderson, R.M., Gordon, D.M. (1982): Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology*, 85: 373 – 398. DOI: 10.1017/S0031182000055347

Bagge, A.M., Sasal, P., Valtonen E.T., Karyonen, A. (2005): Infracommunity level aggregation in the monogenean communities of crucian carp (*Carassius carassius*). *Parasitology*, 131: 367 – 372. DOI: 10.1017/S0031182005007626

Belghyti, D., Berrada-Rkhami, O., Boy, V., Aguesse, P., Gabrion, C. (1994): Population biology of two helminth parasites of flatfishes from the Atlantic coast of Morocco. *J. Fish Biol.*, 44: 1005 – 1021. DOI: 10.1111/j.1095-8649.1994.tb01272.x

Bliss, C.I., Fisher, R.A. (1953): Fitting the negative binomial distribution to biological data. *Biometries*, 9: 176 – 200. DOI: 10.2307/3001850

Buntin, G.D. (1994): Developing a primary sampling program. In: Pedigo, L. P. & Buntin, G. D. (Eds) *Handbook of sampling methods for arthropods in agriculture*. Boca Ratón, Florida, CRC Press, pp. 99 – 115

Bush, A.O., Lafferty, K.D., Lotz, J.M., Shostak, A.W. (1997): Parasitology meets ecology on its own terms: Margolis *et al*. revisited. *J. Parasitol.*, 83: 575 – 583. DOI:10.7939/R3J38KV04

Cyr, H., Downing, J.A., Lalonde, S., Baines, S.B., Pace, M.L. (1992): Sampling larval fish populations: choice of sample number and size. *Trans. Am. Fish. Soc.*, 121: 356 – 368. DOI: 10.1577/1548-8659(1992)121<0356:SLFPCO>2.3.CO;2

Davis, P.M. (1994): Statistics for describing populations. In: Pedigo, L.P., Buntin, G.D. (Eds) *Handbook of sampling methods for arthropods in agriculture*. Boca Ratón, Florida, CRC Press, pp. 33 – 54

Efron, B., Tibshirani, R. (1993): *An Introduction to the Bootstrap*. Chapman & Hall, London, 436 pp.

Gregory, R.D., Woolhouse, M.E.J. (1993): Quantification of parasite aggregation: a simulation study. *Acta Trop.*, 54: 131 – 139. DOI: 10.1016/0001-706X(93)90059-K

Ismen, A., Bingel, F. (1999): Nematode Infection in the whiting *Merlangius merlangus euxinus* off Turkish Coast of the Black Sea. *Fish Res.*, 42: 183 – 189

Jovani, R., Tella, J.L. (2006): Parasite prevalence and sample size: misconceptions and solutions. *Trends Parasitol.*, 22(5): 214 – 218. DOI: 10.1016/j.pt.2006.02.011

Karandinos, M.G. (1976): Optimum sample size and comments on some published formulae. *Bull. Entomol. Soc. Am.*, 22: 417 – 421. DOI: 10.1093/besa/22.4.417

Kleiner, A., Talwalkar, A., Sarkar, P., Jordan, M. (2014): A scalable bootstrap for massive data. *J. R. Stat. Soc. B.*, 76(4): 795 – 816. DOI: 10.1111/rssb.12050

Lau, M.K. (2015): *Package 'DTK'*. https://cran.r-project.org/web/packages/DTK/

Marques, J.F., Cabral, H.N. (2007): Effects of sample size on fish parasite prevalence, mean abundance and mean intensity estimates. *J. Appl. Ichthyol.*, 23: 158 – 162. DOI: 10.1111/j.1439-0426.2006.00823.x

Mouillot, D., Culioli, J.-M., Lepretre, A., Tomasini, J.-A. (1999): Dispersion statistics and sample size estimates for three fish species (*Symphodus ocellatus, Serranus scriba* and *Diplodus annularis*) in the Lavezzi Islands Marine Reserve (South Corsica, Mediterranean Sea). *Mar. Ecol.*, 20(1): 19 – 34. DOI: 10.1046/j.1439-0485.1999.00064.x

Opit, G.P., Throne, J.E., Flinn, P.W. (2009): Sampling plan for the Psocids *Liposcelis entomophila* and *Liposcelis decolour* (Psocoptera: Liposcelididae) in Steel Bins Containing Wheat. *J. Econ. Entomol.*, 102(4): 1714 – 1722. DOI: 10.1603/029.102.0365

Poiani, A. (1992): Ectoparasitism as a possible cost of social life: a comparative analysis using Australian passerines (Passeriformes). *Oecologia*, 92: 429 – 441. DOI: 10.1007/BF00317470

Poulin, R. (2013): Explaining variability in parasite aggregation levels among host samples. *Parasitology*, 140: 541 – 546. DOI: 10.1017/S0031182012002053

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D. (2017): *Package 'MASS'*. https://cran.r-project.org/web/packages/MASS/MASS.pdf

Rohde, K., Hayward, C., Heap, M. (1995): Aspects of the ecology of metazoan ectoparasites of marine fishes. *Int. J. Parasitol.*, 25: 945-970. DOI: 10.1016/0020-7519(95)00015-T

Rózsa, L., Reiczigel, J., Majoros,G. (2000): Quantifying parasites in samples of hosts. *J. Parasitol.*, 86: 228 – 232. DOI: 10.1645/0022-3395(2000)086[0228:QPISOH]2.0.CO;2

Ruesink, W.G. (1980): Introduction to Sampling Theory. In: Kogan, M., Herzog, D.C. (Eds) *Sampling Methods in Soybean Entomology*. Springer-Verlag, New York, pp. 61 – 78

Sarabeev, V.L. (2015): Helminth species richness of introduced and native grey mullets (Teleostei: Mugilidae). *Parasitol. Int.*, 64: 6 – 17. DOI: 10.1016/j.parint.2015.01.001

Sarabeev, V.L., Rubtsova, N.Yu., Yang, T., Balbuena, J.A. (2013): Taxonomic revision of the Atlantic and Pacific species of *Ligophorus* Euzet and Suriano, 1977 (Monogenea: Dactylogyridae) from

mullets (Teleostei: Mugilidae) with proposal of a new genus and description of four new species. *Vestn. Zool.*, 28 (Suppl.): 1 – 112
Shaw, D.J., Dobson, A.P. (1995): Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology*, 111 (Suppl.): S111 – S133. DOI: 10.1017/S0031182000075855
Shelton, A.M., Trumble, J.T. (1991): Monitoring insect populations. In: Pimentel, D. (Ed), *Handbook of Pest Management in Agriculture*. 2nd Edition, CRC Press, Boca Raton, pp. 45 – 55
Snedecor, G.W., Cochran, W.G. (1980): *Statistical Methods*. 7th Edition, Ames: Iowa State University Press, 476 pp.
Steidl, R.J., Hayes, J.P., & Schauber, D.E. (1997): Statistical power analysis in wildlife research. *J. Wildl. Manage.*, 61(2): 270 – 279. DOI: 10.2307/3802582
Taylor, L.R. (1961): Aggregation, variance and the mean. *Nature*, 189: 732 – 735. DOI: 10.1038/189732a0
Wilson, L.T. (1985): Estimating the abundance and impact of arthropod natural enemies in IPM systems. In: Hoy, M.A., Herzog, D.C. (Eds) *Biological Control in Agricultural IPM Systems*. Academic Press, Orlando, pp. 303 – 322
Wilson, K., Bjørnstad, O.N., Dobson, A.P., Merler, S., Poglayen, G., Randolph, S.E., Read, A.F., Skorping, A. (2002): Heterogeneities in macroparasite infections: patterns and processes. In: Hudson, P.J., Rizzoli, A., Grenfell, B.T., Heesterbeek, H., Dobson, A.P. (Eds) *The Ecology of Wildlife Diseases*. Chapter 2. Oxford University Press, pp. 6 – 44
Young, L.J., Young, J.H. (1998): *Statistical ecology: a population perspective*. Kluwer Academic Publishers, Norwell, MA, 565 pp.