Szent István University
Faculty of Veterinary Science
Doctoral School

Statistical and Probability Theoretical Modelling and Analysis of Spatial Structure of Epidemiological and Ecological Indices

Ph.D. Thesis

Zsolt Lang

2014

Supervisor and Consultants:

.....
Prof. Dr. Jenő Reiczigel
Szent István University,
Dept. for Biomathematics and Informatics
supervisor

Prof. Dr. Lajos Rózsa
Animal Ecology Research Group
MTA-MTM
consultant

Dr. Andrea Harnos
Szent István University,
Dept. for Biomathematics and Informatics
consultant

Dr. Franz Rubel
Vet. Med. Univ. Wien,
Dept. für Naturwiss.
consultant

……
Zsolt Lang

Table of Contents

1 Introduction....4

2 Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity....5

2.1 Methods....6

2.2 Results....8

2.2.1 Application: the prevalence of *Trichomonas gallinae* in house finches (*Carpodacus mexicanus*)....9

2.3 Discussion....10

3 Relationship between seropositivity and prevalence on the basis of epidemiological analysis of data of BHV-1 eradication program in Hungary....12

3.1 Material and methods....13

3.1.1 Statistical methods....13

3.2 Results....15

3.3 Discussion....17

4 Crowding and diversity....19

5 Probability theoretic modelling of individuals and their groups....23

6 New scientific achievements....24

7 List of publications related to the present thesis...35

8 Acknowledgements....36

1 Introduction

Prevalence of a disease or a risk factor plays a central role in epidemiology. Perfect diagnosis of a disease is frequently not possible; diagnoses may end up with false positive or false negative results. In this study my first goal is to construct a new confidence interval for true prevalence taking into account that sensitivity and specificity of the diagnostic test have been estimated from outcomes of previous trials. My second theme involves application of exact statistical methods for estimating true prevalence based on data of the national BHV-1 eradication program in Hungary, supposing known sensitivity and specificity.

Aggregated distribution of the target population may influence spatial and temporal changes of prevalence. Individuals are often organised in groups (e.g. herds, territorial groups, parasites on hosts, etc.). Aggregation and crowding is characterised by the abundances of individuals in groups. I discovered that crowding indices are formally analogous to biological diversity indices. My third goal is to unveil this similarity. I pointed out also that crowding and diversity are closely connected if the

members of the groups belong to different species. Furthermore, I developed a general probability structure for individuals forming groups, comparing group-level and individual level distributions and regression models of group characteristics. I applied my new theoretical mathematical methods to louse parasite abundance data on rook and hooded crow and to gender frequencies of deer ked parasites harbouring on red deer.

2 Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity

Prevalence of a disease is usually assessed by diagnostic tests that may produce false results. Rogan and Gladen (1978) described a method to estimate the true prevalence correcting for sensitivity and specificity of the diagnostic procedure, and also deduced an asymptotic formula for the variance of the prevalence estimate. Reiczigel et al. (2010) provided exact confidence intervals for the true prevalence assuming sensitivity and specificity were known. We constructed approximate confidence intervals for the true prevalence when sensitivity and specificity are estimated from independent samples (Lang and Reiczigel 2014).

2.1 Methods

We assume that we have three independent samples:

- the first one consists of patients who are known to have the disease (this will be used to estimate the sensitivity of the diagnostic test),
- the second one consists of patients who don't have the disease (this will be used to estimate the specificity of the diagnostic test),
- the third one is a random sample from the target population, examined by the diagnostic test (true disease statuses of the patients are unknown, just their test results − positive or negative − are known).

Rogan and Gladen (1978) provided an estimate of true prevalence and its variance based on these samples. Furthermore, Wald's well known confidence interval can be calculated for the true prevalence. We modified the method of Rogan and Gladen similarly to the method of Agresti and Coull (1998). We increased the frequencies of both proper and false diagnoses by 1 in the samples of sensitivity and specificity, and by $z_{crit}^2/2$ in the sample from the target population (z_{crit} is the critical value of the standard normal distribution corresponding to the

prescribed confidence level). A new confidence interval was constructed from the adjusted point estimates of true prevalence and its variance (Lang and Reiczigel 2014).

We evaluated the coverage probability of the confidence interval with and without the proposed adjustment by simulation under the following scenarios.

− True sensitivity and specificity were assumed to be 1, .99, .95, .90, .70.

− Sample sizes for estimation of sensitivity, specificity, and prevalence were 30, 100, 300, 1000, 3000.

− Nominal confidence level was set to 90%, 95%, and 99%.

− Coverage was determined for true prevalence .005, .01, .02, .03, .05, .10, .20, .30, .50.

For each combination of the above listed features 20000 random samples (in fact 20000 triplets of random samples) were generated and the coverage probability was calculated.

We also investigated the loss of precision due to application of confidence interval procedures assuming known sensitivity and specificity in such cases when they were estimated. We carried out simulations with confidence intervals using the Blaker and the Clopper-Pearson methods (Reiczigel et al., 2010; Blaker, 2000; Clopper and Pearson, 1934), with sample sizes for sensitivity and specificity equal, two-fold, 3-fold, 5-fold, and 10-fold of the sample size for prevalence estimate. Coverage probability was calculated based on 10000 random samples for various sample sizes and nominal levels, assuming true sensitivity and specificity .99, .95, and .90.

2.2 Results

The coverage probabilities of the confidence interval based on normal approximation and the Rogan-Gladen variance estimate without any adjustment were too low, in many cases less than 90% at nominal 95%, and less than 80% at nominal 90%, in particular in cases when sensitivity and/or specificity was high. Application of the proposed method considerably improved the coverage; it became quite acceptable for all examined parameter

configurations and nominal confidence levels. At nominal 95% the minimum coverage fell seldom below 94% and never below 93%. At nominal 99% the minimum coverage was never less than 98% and at nominal 90% it was never less than 88%.

There is considerable loss of precision due to application of confidence interval procedures assuming known sensitivity and specificity in such cases when they are estimated. We found that loss of precision is negligible only if sample sizes for sensitivity and specificity are five to ten times greater than that for estimating prevalence, but this also depends on sample size for prevalence estimation and on true sensitivity and specificity.

2.2.1 Application: the prevalence of *Trichomonas gallinae* in house finches (*Carpodacus mexicanus*).

Anderson et al. (2009) studied the occurrence of trichomonad protozoa in free ranging songbirds. As an example we will use the prevalence of Trichomonas gallinae in house finches (*Carpodacus mexicanus*). Out of 2971 birds, 51 had the parasite, which resulted in an apparent prevalence of .017. Sensitivity and specificity of the diagnostic method was determined by culturing of

the parasites. Sensitivity and specificity were found to be .97 (32/33), and 1 (20/20), respectively. Using these values, the point estimate for the true prevalence is .018.

Assuming that .97 and 1 are known values for sensitivity and specificity, the 95% CI is .013 to .023 by the method of Reiczigel et al. (2010) using Blaker's CI. Taking into account that sensitivity and specificity are estimated from as small samples as 33 and 20, respectively, the 95% CI for the true prevalence is 0 to .053. This CI is much wider than that assuming known sensitivity and specificity.

2.3 Discussion

We found that uncertainty in sensitivity and specificity assessment may considerably influence the precision of prevalence estimates. As expected, if sample sizes for estimating sensitivity and specificity are much higher (5-10-fold) than that for prevalence estimation, the actual coverage is near the nominal level (>.93 at nominal .95). However, if sample sizes for estimating sensitivity and specificity are comparable to that for prevalence estimation (3-fold or less), the actual coverage of CIs

may be unacceptably low relative to the nominal level (even <.90 at nominal .95).

Our proposed CI for the true prevalence provides good coverage for a wide range of prevalence, sensitivity and specificity values, including prevalence close to 0 and 1 and sensitivity and specificity close to 1. As a consequence of the applied adjustments it is somewhat conservative when prevalence is close to 0 or 1. The accuracy of coverage was preserved even when the three sample sizes differed considerably.

There is a widespread opinion that the 95% Wald confidence interval for population proportion given by the formula $\bar{p} \pm 1.96 \cdot var(\bar{p})^{\frac{1}{2}}$ is reasonably accurate provided the interval $\bar{p} \pm 3 \cdot var(\bar{p})^{\frac{1}{2}}$ contains neither 0 nor 1 and the sample size is “sufficiently large”. We checked the accuracy of this method and found that the coverage was even worse than that of the original Wald-Rogan-Gladen method.

3 Relationship between seropositivity and prevalence on the basis of epidemiological analysis of data of BHV-1 eradication program in Hungary

The national BHV-I eradication program in Hungary, using gE negative marker váccines, has been going on since 2002. We estimated true prevalence of BHV-1 infection from serological results, by re-analysing the data obtained until 2006, considering also the proportion of false positive and false negative reactions attributable to sensitivity and specificity of the applied tests. This information allows more precise evaluation of the progress of eradication as well as the statistical assessment of heterogeneity of infection among farms, age groups and geographical areas and the risk of spreading. We suggest a new ecological parameter the crowding index to characterize the risk of spreading. The statistical methods presented in this chapter are suitable tools to follow up the progress of eradication programs of any infectious disease, and their use may facilitate the in-time decisions for the modification of the programs.

3.1 Material and methods

In the 2002-2006 time period of the eradication program samples were obtained from 155 farms. Farms involved in screening and monitoring were partly different (Pálfi et al. 2007). In the new analysis only that test results were included in which the type of the test, sensitivity, specificity and the diagnoses were known for each sample.

3.1.1 Statistical methods

True prevalence was estimated by the method of Rogan and Gladen (1978). The two-sided 95% confidence interval for true prevalence was constructed by the method of Reiczigel based on Blaker's exact test (Blaker 2000, Reiczigel et al. 2010).

Prevalence was estimated for age groups, counties, groups of counties and farms. The significance of differences of prevalence between screening and monitoring for each age group was tested by one-sided 97.5% Clopper-Pearson confidence intervals (Clopper and Pearson 1934, Reiczigel et al. 2010).

We assessed variances of prevalence within and between counties for each age group. The total variance was decomposed into between county and within county variances (Reiczigel, Harnos, Solymosi 2007). Standard deviations were calculated as the square roots of variances (see also Lang et al. 2013, www.univet.hu/users/zslang/korrekciok.pdf).

We estimated average prevalence and its within-group variances for three disjoint geographically connected and homogeneous groups of counties (Dunántúl without Komárom-Esztergom county, Alföld, Northern counties: Komárom-Esztergom, Pest, Nógrád, Heves, Borsod-Abaúj-Zemplén).

We analysed the relationship between the number of animals and prevalence in the 115 farms where the number of animals were available. We hypothesized that in farms having more animals the number of contacts between animals was large, the spread of infection was faster, and therefore prevalence might reach higher level. In order to test this hypothesis we first sorted the farms in increasing order according to the number of animals they possessed, and then we classified them

into quintiles including 23-23 farms. BHV-1 prevalence was estimated for each quintile both at screening and at monitoring. Statistical analysis was based on simultaneous confidence intervals using the method of Bonferroni.

The number of animals has large variance between farms. Consequently, average number of animals does not properly express the number of possible contacts between animals. Crowding index was applied to measure the average contacts as one of the indicators of infection spread. (Lloyd 1967, Reiczigel et al. 2005, 2008).

3.2 Results

True prevalence was smaller at monitoring than at screening both in separate age groups and pooled together as well. True prevalence was higher at screening and lower at monitoring than seropositivity. Consequently, the eradication was actually more successful than the observed difference in seropositivity might suggest during the study period (from 2002 to 2006).

Counties showed large heterogeneity of prevalence both at screening and at monitoring. Prevalence was reduced considerably in most of the counties. No relevant differences occurred in average prevalence and its variance between the three geographically homogeneous regions at screening. Average prevalence became reduced more effectively in Dunántúl and Northern counties than in Alföld at monitoring. Heterogeneity was reduced, too, especially in the Northern counties.

We analysed the relationship between the number of animals and prevalence using frequency quintiles in the 115 farms where the number of animals were known. We observed that BHV-1 prevalence had positive association with the number of animals at screening. We showed using simultaneous confidence intervals that the prevalence at screening was significantly greater in farms having more than 450 animals than in smaller farms. At monitoring the trend was reversed, prevalence was smaller in larger farms. Positive association between the number of animals and prevalence at screening was also shown for calves and adult cows. No association was observed for heifers and pregnant

heifers. At monitoring no significant association was recognised in age groups.

Crowding index was 867 at screening and 817 at monitoring. The average possible contact number was greater by 50–80% than the average number of animals in farms.

3.3 Discussion

True prevalence proved to be larger than seropositivity at screening and smaller than seropositivity at monitoring. The explanation of this paradox is that eradication reduces seropositivity and according to the Rogan-Gladen formula if seropositivity is large then true prevalence is even larger, if seropositivity is small then true prevalence is even smaller.

What kind of diagnostic test should we choose? Before eradication, when prevalence is large false negative samples may distort the precision of true prevalence. Therefore the chosen test should have as high sensitivity as possible. Following successful eradication the prevalence is small and false positive samples may increase the standard error of prevalence estimates.

Consequently, at monitoring after eradication the selected test should have as high specificity as available, the level of sensitivity is less important.

The large crowding index together with heterogeneous prevalence may result in quick spread of the infection. During the inspected period the crowding index was high and the success of eradication was very heterogeneous in counties and groups of counties in Hungary. The eradication was poor especially in farms with small number of animals. Consequently, there is a considerable risk that large eradicated farms will be infected again by the populations of infected animals living nearby in smaller farms.

At screening there is a high level of BHV-1 prevalence in large farms, hence without vaccination the large crowding index and contact number will increase the prevalence of the disease.

The methods introduced can be applied most effectively to countrywide analyses or studies involving many farms. In such analyses precise estimates of prevalence in age groups and territorial and temporal differences are available.

4 Crowding and diversity

Crowding or average individual group size is defined as the monotone increasing scaled abundance averaged over all the individuals living in groups in a community (Lloyd 1967, Reiczigel et al. 2005, 2008). I discovered tight analogy between the notions of crowding and the diversity indices of species. According to Patil and Taille (1982) biological diversity (D) is the rarity of species averaged over the individuals in the community studied. The rarity of a species $R(p)$ is a monotone decreasing function of the proportion p of the species in the community that satisfies $R(1) = 0$. Crowding and diversity can be expressed by each other through fixing the scale functions. For any fixed c constant $c - R(p)$ is a monotone increasing function of the abundance of the species and the corresponding crowding index is $c - D$. The most important corresponding pairs of crowding and diversity indices are for linear scale functions the Lloyd and Reiczigel crowding indices and Simpson's diversity index, for logarithmic scale function the logarithmic crowding index and the Shannon diversity index, for hyperbolic scale function the hyperbolic crowding index

and species richness (i.e. number of species minus one).

I introduced the notion of effective number of groups as the crowding theory counterpart of the notion of effective number of species (more frequently named as numbers equivalent) used in the theory of diversity indices. Effective number of groups is the number of groups in a community consisting of groups with equal number of individuals and having the same number of individuals and the same crowding index as the investigated community. Based on this notion I generalised Bez's aggregation index (Bez 2000) as the ratio of the actual and effective number of groups in the community.

I showed that there is an immediate connection between crowding and diversity when the individuals in the groups of the investigated community belong to different species. Overall crowding of individuals can be decomposed into the sum of average crowding indices over species and the local, within group diversities averaged over groups. Besides this additive relationship there is a multiplicative one, too. It is known that the so called true gamma diversity (or gamma effective number

of species) of communities containing individuals that belong to several species can be decomposed into the product of average within group diversity called true alpha diversity and true beta diversity measuring the heterogeneity of species distributions between groups (Whittaker 1960, 1972, Hill 1973, Jost 2006, 2007, Tuomisto 2010a). In crowding theory effective number of groups or the index of aggregation correspond to the effective number of species. I introduced gamma aggregation being the aggregation of the whole community, alpha aggregation as the average aggregation over species and beta aggregation as the ratio of alpha and gamma aggregations. Here beta aggregation measures heterogeneity of aggregation patterns between species.

I applied the theoretical results obtained for crowding and diversity measures to the comparison of louse parasite infections of rooks (*Corvus frugilegus*) and hooded crows (*Corvus cornix*). Parasite crowding is larger in rooks and the percentage of the subjective diversity component is also larger in them. It can be explained by the fact that hooded crows are infected

more frequently (50%) with only one louse species thus having 0 subjective diversities compared to rooks (21%).

True alpha and gamma diversities of louse species are somewhat larger in rooks than in hooded crows. True beta diversities are approximately the same, being between 1.5 and 2.5. The average alpha aggregation of louse species is greater in hooded crows than in rooks. This means that louse species harbouring on hooded crows have more aggregated distribution among hosts than the corresponding louse species harbouring on rooks. Beta aggregations are between 1 and 2 in both rooks and hooded crows; therefore I could divide the louse genuses in rooks into more or less disjoint groups, the Brueelia and the Myrsidea and Menacanthus genuses using block clustering algorithm and correspondence analysis. In hooded crows the two dominant louse genuses the Myrsidea and the Menacanthus form disjoint groups separating their hosts. It is interesting that the Myrsidea and Menacanthus genuses will not separate in rooks and will separate in hooded crows.

5 Probability theoretic modelling of individuals and their groups

I introduced a general probability theoretic model for individuals forming disjoint groups. In this model distributions of certain characteristics are compared when the observation unit is an individual and when it is a group. I derived transformations and formulas connecting the two approaches. I compiled a few sufficient and a necessary and sufficient condition for a general regression model to have exactly the same form according to the two points of view. One of them is that the number of individuals in the groups should be included in the set of explanatory variables of the regression model. The other versions are mathematical refinements of this simple condition.

I applied the described probability theoretic model to compare gender proportion and parasite intensity of deer ked (*Lipoptena cervi*) parasites harbouring on red deer (*Cervus elaphus*). Mixed effects logistic regression was fitted to the data. Parasite intensity was included in the covariates of the regression ensuring that the model

had the same form both in the point of view of deer ked parasites and their hosts.

I showed that the proportion of male deer ked was significantly lower in more intensely parasitized red deer ($p = 0.0001$). This result is valid for both red deer as independent entities and for micro populations of own hosts of selected deer ked. Based on the results of the regression model I showed that the proportion of deer ked gender was 1:1 in a typical red deer and the proportion of deer ked gender was shifted significantly towards females on the host of a typical deer ked. The explanation of this paradox is that a typical deer ked harbours on such a red deer host that is more parasitized than a typical red deer.

6 New scientific achievements

1. We introduced a new method to construct a confidence interval for true prevalence when sensititivity and specificity of the diagnostic test are estimated from independent samples.

2. We proved by computer simulation that the new confidence interval preserves more precisely the

prescribed confidence level than the currently accepted Wald-Rogan-Gladen interval, even for sample sizes as small as 30; the coverage is at least 88%, 93% and 98% if the prescribed level is 90%, 95% and 99%, respectively.

3. We proved theoretically that the Rogan-Gladen point estimate of true prevalence falls within the new confidence interval provided the estimated sensitivity and specificity of the diagnostic procedure is at least 0.5, the sample size for sensitivity is ≥26, the sample size for specificity is ≥26 and the size of the sample taken from the target population is ≥16.

4. We illustrated the advances of the new confidence interval by re-analyzing real-life published applications.

5. We conducted simulations to assess the loss of precision if Blaker and Clopper-Pearson confidence intervals were constructed based on sensitivity and specificity estimated from samples. (These methods assume sensitivity and specificity are known.) We found that loss of

precision was negligible only if sample sizes for sensitivity and specificity were five to ten times greater than that for estimating prevalence, but this also depended on sample size for prevalence estimation and on true sensitivity and specificity.

6. We proved by computer simulation that the widely published and applied candidate interval $\hat{P} \pm 3 \cdot var\left(\hat{P}\right)^{\frac{1}{2}}$ did not improve the precision of the Wald-Rogan-Gladen confidence interval and frequently it would even produce no confidence interval at all.

7. We estimated true prevalence of BHV-1 infection from seropositivity in Hungarian cattle farms. We showed using exact confidence intervals that true prevalence was smaller at monitoring than at screening both in the whole population and in each age group. We also concluded that true prevalence was greater than seropositivity at screening and it was less than seropositivity at monitoring. Consequently, the eradication was actually more successful in the investigated

period (2002-2006) than it would appear from the observed seropositivity proportions.

8. We investigated the fluctuations of true prevalence between and within counties. We concluded that there was great heterogeneity in prevalence between cattle farms regardless of their counties, both at screening and at monitoring.

9. We divided the country into three geographically homogeneous regions and investigated the mean and variance of prevalence in them. We proved that at screening there was no relevant difference between these regions with respect to mean and variance of prevalence. Average prevalence was reduced more effectively in Dunántúl and Northern counties than in Alföld at monitoring. Heterogeneity was reduced, too, especially in the Northern counties.

10. We studied the relationship between the number of animals and BHV-1 prevalence in farms, applying simultaneous confidence intervals and Bonferroni's method. We showed that the

prevalence at screening was significantly greater in farms having more than 450 animals than in smaller farms. At monitoring the trend was reversed, prevalence was smaller in larger farms. The significance level was set to 5%.

11. Positive association between number of animals and prevalence at screening was also shown for calves and adult cows. No association was observed for heifers and pregnant heifers. At monitoring no significant association was recognised in separate age groups. The significance level was set to 5%.

12. We showed that the crowding index was 867 at screening and 817 at monitoring. The average possible contact number was greater by 50–80% than the average number of animals in farms.

13. I discovered an analogy between the notions of crowding and biological diversity of species. Patil and Taillie (1982) defined biological diversity of species as the scaled rarity averaged over the individuals. I showed that having set fixed the scale function crowding and diversity can be

expressed by each other. The most important corresponding pairs of crowding and diversity indices are for linear scale functions the Lloyd and Reiczigel crowding indices and Simpson's diversity index, for logarithmic scale function the logarithmic crowding index and the Shannon diversity index, for hyperbolic scale function the hyperbolic crowding index and species richness (i.e. number of species minus one).

14. I introduced the notion of effective number of groups as the crowding theory counterpart of the notion of effective number of species used in the theory of diversity indices. Effective number of groups is the number of groups in a community containing groups with equal number of individuals and having the same number of individuals and the same crowding index as the investigated community. Based on this notion I generalised Bez's aggregation index as the ratio of the actual and effective number of groups in the community.

15. Under general conditions I proved that the aggregation index is ≥1 and it equals 1 only if the distribution of abundances of individuals in the groups are perfectly even (i.e. the abundances are equal).

16. I showed that there is an immediate connection between crowding and diversity when the individuals in the groups of the investigated community belong to different species. Overall crowding of individuals can be decomposed into the sum of average crowding indices over species and the local, within group diversities averaged over groups.

17. It is known that the so called true gamma diversity of communities containing individuals that belong to several species can be decomposed into the product of average within group diversity called true alpha diversity and true beta diversity measuring the heterogeneity of species distributions between groups. In crowding theory effective number of groups or the index of aggregation correspond to the

effective number of species. I introduced gamma aggregation being the aggregation of the whole community, alpha aggregation as the average aggregation over species and beta aggregation as the ratio of alpha and gamma aggregations. Here beta aggregation measures heterogeneity of aggregation patterns between species.

18. I applied the theoretical results obtained for crowding and diversity measures to the comparison of louse parasite infections of rooks (*Corvus frugilegus*) and hooded crows (*Corvus cornix*). Parasite crowding is larger in rooks and the percentage of the subjective diversity component is also larger in them. It can be explained by the fact that hooded crows are infected more frequently (50%) with only one louse species thus having 0 subjective diversities compared to rooks (21%).

19. True alpha and gamma diversities of louse species are somewhat larger in rooks than in hooded crows. True beta diversities are approximately the same, being between 1.5 and

2.5. The average alpha aggregation of louse species is greater in hooded crows than in rooks. This means that louse species harbouring on hooded crows have more aggregated distribution among hosts than the corresponding louse species harbouring on rooks.

20. Beta aggregations are between 1 and 2 in both rooks and hooded crows; therefore I could divide the louse genuses in rooks into more or less disjoint groups, the Brueelia and the Myrsidea and Menacanthus genuses using block clustering algorithm and correspondence analysis. In hooded crows the two dominant louse genuses the Myrsidea and the Menacanthus form disjoint groups separating their hosts. It is interesting that the Myrsidea and Menacanthus genuses will not separate in rooks and will separate in hooded crows.

21. I introduced a general probability theoretic model for individuals forming disjoint groups. In this model distributions of certain characteristics are compared when the observation unit is an

individual and when it is a group. I derived transformations and formulas connecting the two approaches.

22. Within the framework of the introduced model I gave a probability theoretic interpretation of crowding indices having general type scale function.

23. I compiled a few sufficient and a necessary and sufficient condition for a general regression model to have the same form according to the two points of view. One of them is that the number of individuals in the groups should be included in the explanatory variables of the regression model. The other versions are mathematical refinements of this simple condition.

24. I applied the described probability theoretic model to relate gender proportion to parasite intensity of deer ked (*Lipoptena cervi*) parasites harbouring on red deer (*Cervus elaphus*). Mixed effects logistic regression was fitted to the data. Parasite intensity was included in the covariates

of the regression ensuring that the model had the same form both in the point of view of deer ked parasites and their hosts.

25. I showed that the proportion of male deer ked is significantly lower in more intensely parasitized red deer ($p = 0.0001$). This result is valid for both red deer as independent entities and for micro populations of own hosts of selected deer ked. Based on the results of the regression model I showed that the proportion of deer ked gender is 1:1 in a typical red deer and the proportion of deer ked gender is shifted significantly towards females on the host of a typical deer ked.

7 List of publications related to the present thesis

Lang Zs., Földi J., Ózsvári L., Reiczigel J.: Szeropozitivitás és prevalencia összefüggései hazai BHV-1-mentesítés adatainak járványtani elemzése alapján. Magy. Állatorv. Lapja 135, 525–534. 2013. IF_{2012}: 0.146

Lang Z., Reiczigel J.: Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. Preventive Veterinary Medicine 113, 13–22. 2014. IF_{2012}: 2.389

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Properties of crowding indices and statistical tools to analyse crowding data. Journal of Parasitology 91, 245–252. 2005. IF_{2011}: 1.320

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Measures of sociality: Two different views of group size. Animal Behaviour 75: 715–721. 2008. IF_{2012}: 3.068

Vágó E., Lang Zs., Kemény S.: Overdispersion at the Binomial and Multinomial Distribution. Periodica Polytechnica 55, 17–20. 2011. IF_{2012}: 0.269

8 Acknowledgements

I would like to express my gratitude to my supervisor Dr. Jenő Reiczigel and Dr. Lajos Rózsa, for that they involved me in their fields of research and supported my studies with great competence. I am grateful to Dr. József Földi and Dr. László Ózsvári for our joint research and their enthusiastic support. My gratitude goes to Dr. Sándor Kemény for his restless encouragement and also to Dr. Emese Vágó for our joint research. I am grateful to my opponents Dr. Andrea Harnos and Dr. Júlia Singer as they have read through my dissertation and improved its level by their comments. I thank Dr. Szilvia Pásztory-Kovács for that she shared her experiences related to the doctoral procedure. I thank Péter Fehérvári and Dr. Zoltán Vas for their helpful comments and suggestions.

I am grateful to my wife Anna and my son Péter for their love and patience. I thank my parents that they have always been supporting me.