

FROM DESIGN TO ANALYSIS: EFFECTIVE STATISTICAL APPROACHES FOR HOST RANGE TESTING

Thomas S. HOFFMEISTER

Institute of Ecology and Evolutionary Biology
University of Bremen
28359 Bremen, Germany

hoffmeister@uni-bremen.de

ABSTRACT

The major goal of host range testing in biological control is to minimize the probability that released biological control agents have unwanted effects on populations of non-target hosts. This leads to a non-trivial problem in statistical hypothesis testing, since the standard approach in statistical tests is to ask whether or not an effect – in this case acceptance of a non-target host – exists and to attribute a precise probability to err only with rejecting the null hypothesis that assumes no effect. The problem is that it is difficult to assign a probability with accepting the null hypothesis of no effect, i.e., that the biological control agent does not include a given non-target insect into its host range. Yet, this piece of information is exactly what we need for high precision and confidence. Confidence in this respect increases with sample size and the statistical effect size, i.e., the difference from the null hypothesis that is considered biologically meaningful. However, sample size is often limited due to limitations in test subjects, research money, and space for testing arenas. Consequently, there is a high premium on using a very good experimental design and employing the most powerful statistical approach available. This paper discusses common problems with experimental designs, emphasizes the necessity to decide on the statistical effect size that is biologically meaningful, points towards the need to determine the statistical power of the host range test employed, and provides an overview about powerful statistical approaches for analyzing experiments on the host range of potential biological control agents.

INTRODUCTION

Over the last two decades ecologists have become increasingly aware of novel and powerful statistical approaches. This trend can be witnessed by a number of recent textbooks on design and statistical approaches in the life sciences (e.g., Crawley 1993; Crawley 2002; Grafen and Hails 2002; Hilborn and Mangel 1997; Quinn and Keough 2002; Ruxton and Colegrave 2003) and changes in approaches used in more recent publications. This reflects both the increased awareness that conclusions in ecological studies need to be drawn in a quantitative manner with high precision and confidence, and that, for a number of reasons, large sample sizes are often difficult to obtain. This is especially so for studies on the host range of agents for bio-

logical control, since these animals have to be tested on a number of non-target hosts. Thus, the need for powerful statistical tools that allow precise analysis from limited sample sizes is especially evident in this field of research. Formerly, the statistical analysis of data in ecological investigations has been fraught with the difficulty that many if not most of the data sampled in these cases are not normally distributed and are thus not suitable for the parametric 'standard' approaches of Analysis of Variance (ANOVA) and Student *t*-tests. Instead, non-parametric statistics like, e.g. Kruskal-Wallis and Mann-Whitney U-Tests have been used that are known to be less powerful. In theory, the lack of power of non-parametric statistics may be compensated by larger sample sizes. However, an increase in sample size is often not feasible for agricultural entomologists who are usually limited by the time that can be invested, the money that can be spent on experiments, and/or the number of replicates that can be obtained through a shortage of either experimental fields or insects to work with.

In this paper, I want to make 4 points: Firstly, that in many experiments of host range testing it becomes most interesting when we do not find a statistical effect, e.g., no effect on non-target hosts, a situation that is inherently difficult to interpret in statistical testing. Secondly, and following from the first point, that it is generally important to determine and to report on the precision with which we can conclude that no effect exists when no statistically significant effect has been found, i.e., the Power of the statistical test. Thirdly, that it is usually advisable to carefully consider the distribution of the data and find the most powerful means of analyzing them. And fourthly, that as yet, not all research questions in insect host range testing can be analyzed with easily accessible powerful statistical methods and that further progress in this field is clearly needed.

Throughout, I will use verbal examples or computer generated (fake) data sets to elucidate my arguments.

β -ERRORS AND THEIR IMPORTANCE FOR INSECT HOST RANGE TESTING

The very basis of statistical testing is that, by performing an experiment, it remains impossible to prove, for example, that a natural enemy will never attack a non-target host or prey. Using a sound experimental design, we can only aim at achieving high accuracy and precision in what we conclude from the sample that we tested. Yet, using standard statistical procedures, there is always some possibility that our interpretation of the data is wrong. This is due to the fact that all the measurement variables we are interested in are usually subject to random variation (i.e., variation between sample units that cannot account for a treatment factor considered) and that our conclusion is based on a sample rather than the entire population. In general there are two ways to err: 1) based on test results we may either conclude that there is an effect when in fact there is none, or 2) we may conclude that there is no effect when in fact there is an effect (Fig. 1). Standard statistical testing is much concerned with the first kind of error, the so called α -error or Type I error, which is returned as *P*-value in test results. However, in insect host range testing, it is often much more important to know the probability of committing a β -error: let us assume that we have tested the mortality of non-target hosts in field cages with and without the presence of a biological control agent, have found 10 and 17 % mortality in control and treatment cages, and have not found a statistically significant deviation from the null hypothesis that states in our case that no difference exists in mortality

of the non-target prey in control cages without and treatment cages with the biological control agent present. Assume further that our statistical test returns a P -value of $P = 0.167$. Is it safe to conclude that we cannot reject the null-hypothesis? In this case we would usually state – using words rather than statistical jargon – that in our test the biological control agent did not cause significant mortality of the non-target prey. However, we do not know the β -error (that an effect exists that we did not detect). If we decide to release an exotic natural enemy for biological control based on such results, and if in fact we committed a β -error, i.e. the natural enemy in fact causes mortality of the non-target prey, unwanted non-target effects may be the consequence. This seems much more problematic than committing an α -error, i.e. rejecting a natural enemy for biological control based on tests that falsely led to the conclusion that the biological control agent would cause mortality of non-target prey. Therefore, in non-target testing, it seems fundamental to obtain information about the β -error. This is where power analysis comes into play.

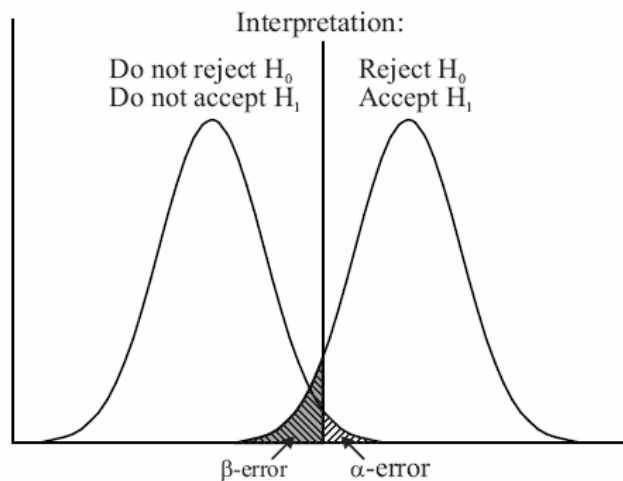


Figure 1. Graphical representation of α -error (area hatched in white and black) and β -error (area hatched in grey and black) probabilities, using a one-sided t -test, comparing, e.g., encounter rates of biological control agents with non-target hosts. The curves on the left (for the null hypothesis) and right (for a specified alternative hypothesis) represent the probability sampling distribution of the statistical test done. Note that usually, the alternative hypothesis is not specified, i.e. H_1 is just different from H_0 , and the probability distribution of the statistical test done for H_1 is unknown (modified from Quinn and Keough 2002).

REPLICATE NUMBER, EFFECT SIZE, AND POWER OF STATISTICAL TESTS

While the β -error is defined by the probability of not finding an effect when in fact there is an effect, statistical power is the probability of a given statistical test finding an effect (rejecting the null hypothesis) when in fact there is an effect. Hence, $\text{power} = 1 - \beta$. For any particular test, power is dependent on the α -level, the sample size, the sampling variance and the so called effect size. The effect size can be regarded as the magnitude of the departure from the null hypothesis (observed effect size) or the difference between the values considered in the null and the alternative hypothesis (Fig. 1). Sample size is positively related to power, i.e., with increasing sample size does the power of a statistical test increase. However, this rela-

tionship is not linear, thus a twofold increase in power requires more than a twofold increase in sample size. Power analysis can follow three different routes, it might: 1) be used *a-priori* to define the sample size necessary to detect an effect with a predefined precision, 2) be used *a-posteriori* to calculate the Power of a test that has not detected a significant effect, or 3) to find compromise levels for α - and β -errors when sample size is fixed. The latter is a consequence of the fact that α - and β -errors are closely related. As can be seen in Fig. 1 a decrease in the α -error leads to an increase in the β -error and vice versa (e.g., imagine to shift the interpretation borderline in Fig. 1 between not rejecting H_0 and accepting H_1 to the left; the shaded areas for α - and β -errors would increase and decrease, respectively). Thus, if sample size cannot be increased, and β -errors are of concern one may compromise the α -error in the interpretation of test results, e.g., stating that a significant effect exists up to a P -value of 0.2, to use a somewhat extreme example. If sample size can be increased, i.e. before an experiment is carried out, *a-priori* power analysis can be used to define the necessary sample size. However, the effect of size needs to be determined in advance. While there are conventions for small, medium, or large effect size for different tests (Cohen 1998), in non-target tests, one may simply use the deviance from the null hypothesis of no effect as being biologically meaningful.

Let us use the above mentioned example of a field cage test on non-target effects of a biological control agent. If we would consider a mortality of 5 % induced by the biological control agent as the maximum that is acceptable and we know that in such experiments we have a background mortality rate of 10 % with a known standard deviation, we can use the arcsine-transformed proportional values (to allow for parametric tests like t -tests) to calculate the effect size. With transformed means of 0.322 and 0.398 and a standard deviation of 0.22, the effect size is 0.341 and thus falls between the values of 0.2 for small and 0.5 for medium effects that are conventionally considered. An *a-priori* power analysis for a one-tailed t -test (we are not interested whether mortality in the treatment is lower than in the control) for an α -error of 0.05 and power of 0.8 (note that this allows a β -error of 20 %) suggests a required sample size of 216, a replicate number that is often unachievable in host range testing. Allowing for 10 % mortality induced by the biological control agent would increase the effect size to 0.645 and reduce the total sample size needed to 78.

While it is usually advisable to conduct *a-priori* power analyses before conducting experiments, often some needed values like the variation around means or, in our example background mortality rates are unknown. Thus, in many cases, power analysis only comes into play, after researchers have not found a statistically significant effect and need to know the confidence with which they can decide not to reject the null hypothesis of no effect. For those *a-posteriori* power analyses, a critical parameter is the effect size assumed. Generally there are two possibilities to determine the effect size. First, the effect size may be computed from the data. However, this does not add new information about the data (see Thomas 1997 for a valuable discussion why this is so). Rather, the effect size should be either determined by using conventions or should – and I would consider this more sensible – be calculated from a biological meaningful effect that we wish to detect.

If, for example, one would have carried out the above mentioned experiments with 10 field cages each for control and treatment and would have found on average 10 % mortality in the control cages and 17 % mortality in the cages with biological control agent and non-target prey, and we would have found no significant effect of the biological control agent on the

mortality of non-target prey ($P = 0.167$), the power would be 0.277 if the effect size would reflect that we accept a maximum of 10 % mortality induced by the biological control agent. This value is unacceptably low.

Programmes to conduct power analyses are either available for free in the internet or are increasingly often included as modules in current statistical software packages (see Thomas and Krebs 1997 for a list of programs and comprehensive review on this topic and Hoffmeister *et al.* 2006 for a recent discussion and alternative ways of achieving power estimates). However, not all tests are covered yet. For example, to my knowledge, no power analysis is as yet available for Generalized Linear Models (see below).

PSEUDOREPLICATION AND DATA INTERDEPENDENCE, A CLASSICAL ISSUE, UNFORTUNATELY

One of the central assumptions of almost all statistical tests is that data points are independent from each other (one exception is planned dependencies in paired data designs). This said, we might wonder why this assumption is so often violated in experiments (see e.g., Hurlbert 1984). One of the most frequent reasons for data interdependence is pseudoreplication. It occurs whenever inferential statistics are used to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent (Hurlbert 1984). Statistical independence means that each individual data point might positively or negatively deviate from the population average due to random variation not related to the deviation of another point. Although the awareness of researchers to avoid pseudoreplication has increased and fewer studies contain analyses with pseudoreplicated samples (Heffner *et al.* 1996), an alarmingly 46% of 105 studies were found to be pseudoreplicated in a recent study on pseudoreplication in experiments on the olfactory response of insects (Ramirez *et al.* 2000) Thus, pseudoreplication still is an issue in the design of experiments, and much care has to be taken to avoid any spatial or temporal segregation of samples from different treatments. For example, when testing the host range of biological control agents, it is essential that insects for the tests on non-target hosts do not come from one rearing container or incubator and control animals (for the test on target hosts) come from another, or that non-target hosts are always tested in the same container or field cage or on the same plant and target hosts are tested in another cage or on another plant. Equally, positions of experimental units within an experimental chamber or on a field plot need to be switched between treatments to avoid confounding effects of differences in temperature and light conditions etc. In the same manner, the full set of trials on non-target hosts should not be conducted before tests with target hosts are carried out. Randomization of testing order or random assignment to plants or test cages assures that pseudoreplication can be avoided. For further reading, I encourage the reader to take a look at the section on pseudoreplication in Ruxton and Colegrave (2003).

GENERALIZED LINEAR MODELS, POWERFUL STATISTICAL APPROACHES FOR INSECT HOST-RANGE TESTING

Many of the traits to be analysed in biological investigations do not follow a Gaussian (also called “Normal”) distribution, and thus standard *t*-tests, analyses of variance (ANOVA) or regression analyses cannot be used to statistically test the effect of a treatment. All these different “classical” methods assume that the distribution of residuals around the fitted model (i.e., the error distribution) is normal (Gaussian). Thus data need to be transformed to achieve a Gaussian distribution or different approaches have to be used. While transformation is often possible, it changes the relationships between parameters in the model. For example, log-transformation of data would make the relationship between parameters in the statistical model multiplicative that has been additive for untransformed values. Thus approaches should be favoured that do not make it necessary to transform values to achieve a Gaussian distribution of data. While non-parametric tests like Mann-Whitney U tests or Kruskal-Wallis tests lack statistical power, Generalized Linear Models can be used to predict responses both for dependent variables that are not normally distributed and for dependent variables which are nonlinearly related to the predictors. They are a generalization of general linear models that underlie classical statistical tests like ANOVA and regression. While in general linear models, the data distribution is Gaussian and the link function is identity, various types of data distribution and link functions (see McCullagh and Nelder 1989) can be chosen, depending on the assumed distribution of the *y* variable values. Table 1 gives the list for the four main generalized linear models that can be used in experiments done to estimate host range of biological control agents.

To give an example, imagine a large arena choice test as suggested in van Lenteren *et al.* 2006. Three different treatments are used, with 10 field cages each: (1) with the target prey (or host which is used synonymously here) and non-target prey present in the same field cage together with the natural enemy, (2) with only the non-target prey and the natural enemy in the same field cage, and (3) with only the target prey and the natural enemy in the same field cage. We are interested in whether the target prey is killed at a higher rate than the non-target prey and whether the mortality of the non-target prey depends upon the fact whether the target prey is available to the natural enemy or not. To achieve independent data, one should not compare whether mortality rates of target and non-target prey are equal within a single treatment. Rather, one should test whether the mortality of non-target prey in treatment (1) is equal to the mortality of non-target prey in treatments (2) and equal to the target prey in treatment (3) (this is our null hypothesis). Again, I use computer-generated data. Given the mortality rates found were 4.1 %, 10.6 % and 50.5 % in (1), (2) and (3), respectively, a Generalized Linear Model with binomial distribution and logit link finds a significant effect overall and also between treatments (Table 2). Thus, in this example, the non-target prey is attacked at relatively low rate and even less so, when target prey are available. This result is visible from the estimates in Table 2, where the estimate for mortality is positive and thus higher in treatment (2) than in treatment (1), and much higher (more than 3 times higher) in treatment (3) than in treatment (1).

Table 1. List of the main generalized linear models that can be used in experiments done to estimate the host range of biological control agents. Link functions indicated are the most frequently used ones. Other can be used in particular cases (see McCullagh and Nelder 1989, for an exhaustive description).

Distribution	Model description	Appropriate link function	Example for data type
Gaussian	General linear model	identity: $f(y) = y$	Morphological data
Binomial	Logistic regression	logit: $f(y) = \log\{y/(1-y)\}$	Proportions like parasitism
Poisson	Log-linear model	log: $f(y) = \log(y)$	Counts like egg load or number of prey consumed
Gamma	Gamma model	inverse: $f(y) = 1/y$	Time durations like survivorship

Table 2. Results of a Generalized Linear Model on computer-generated data for the mortality rates of target and non-target prey in large arena choice tests (for details, see text).

Parameter	Treatment	Estimate	DF	χ^2	Pr > ChiSq
Intercept		-3.2591	1	378.47	<.0001
Target host	(3)	3.2511	1	329.63	<.0001
Non-target prey in no-choice test	(2)	1.0839	1	30.14	<.0001
Non-target prey in choice test *	(1)	0	0	0.0000	

* In the SAS statistics package, which was used here, the last treatment [in this case (1)] is set to zero by convention and the difference between the last and all other treatments [(2) and (3)] is tested.

A special case of Generalized Linear Models exists if measurements are taken repeatedly. If, for example one plans to monitor the mortality induced by the natural enemy on the target and non-target host across a time period after the release of the natural enemy, several data points from the same treatments will be taken. In this case, A GEE model can be specified with the Generalized Linear Model (see e.g., Quinn and Keough 2002) that adequately deals with such data.

TIME DURATIONS AND CENSORED DATA

Time duration data like survival times or latency until attack usually follow an exponential distribution, because the probability λ to die or to become attacked in each time unit is constant. While generally such data can be analysed with Generalized Linear Models with gamma distribution and inverse link function, they cannot if data points are censored, i.e., when we were unable to measure a quantifiable value. Right-censored data origin, for example, from host range experiments in which we measure the latency until attack of target and non-target prey in small arenas with behavioural observation, when a predator did not attacked the prey until the end of the observation (in this case we just know that the latency is larger than the

time of observation, but cannot quantify it properly). If we just ignore those censored values, the interpretation of the test might be wrong. A Cox regression model (= proportional hazards model) can adequately deal with censored time duration data (Cox 1972). Recently, a plethora of different studies have used such an statistical analysis for ecological investigations on insects (e.g., van Alphen *et al.* 2003). Besides using this sort of analysis to study changes in survival time, a Cox survival analysis can also be used when it comes, for example, to testing residence times or giving up times of natural enemies on patches with target and non-target prey, or when testing the latency until a natural enemy attacks a host or prey.

A POWERFUL STATISTIC FOR EVERY PROBLEM? – UNFORTUNATELY NOT

Recent advancements in statistical methods may give the impression that almost every biological problem imaginable in insect host range testing could be analysed with one of the powerful methods described above. Unfortunately this is not so. Besides the banality that good statistics cannot cure poor experimental designs, some of the research questions one will often address in insect host range testing cannot be easily analyzed with powerful statistical methods. For example imagine a no choice test with a natural enemy on target and non-target host. It is statistically not problematic to test the null hypothesis that acceptance of target and non-target prey does not differ. However, this test is not the most interesting research question we might have in mind. If we are to decide whether or not to introduce an exotic natural enemy, we need to know whether the natural enemy will accept the non-target host at all. One approach would be to assume that host acceptance does not vary and, given that we have found in say, 10 replicates on non-target hosts, that they are not accepted while the target host has invariably been accepted. No statistical test would be needed in this case. However, host acceptance usually is variable. Host acceptance experiments with biological control agents of different degrees of host deprivation clearly show increasing acceptance rates with increasing host deprivation (Withers and Mansfield 2005, this issue).

One possibility to solve the problem using statistical methods would be to decide on a threshold of acceptance that can be tolerated, and given one has found no acceptance of non-target hosts in n replicate trials, one can compute the probability to obtain a series of n host rejections given the threshold level (see Porter *et al.* 1995 for a published example). Alternatively, we might use an exact test based on a binomial distribution. Here, we need to define a null hypothesis (H_0) about the likelihood that a biological control agent accepts the non-target host and an alternative hypothesis (H_A) about a threshold level of this probability that we believe would be crucial to detect. For example, let us assume the H_0 that the biological control agent would have an inherent probability of $\lambda_0 = 0.01$ to accept the non-target host (thus on average, 1 out of 1000 parasitoids would accept the non-target host). Let us further assume that we wish to detect if the true acceptance rate of the parasitoid, our H_A , is $\lambda_A = 0.05$ (the dotted line in Fig. 2). In this case, we would need 32 replicates to obtain a power of > 80 % (Fig. 2). Critical values to detect a significant deviation ($P < 0.05$) from the null hypothesis of 0.1 % acceptance rate of non-target hosts are detected if at least r non-target hosts are accepted ($r = 1$ for sample sizes of $1 \leq n \leq 51$ and $r = 2$ for $52 \leq n \leq 100$).

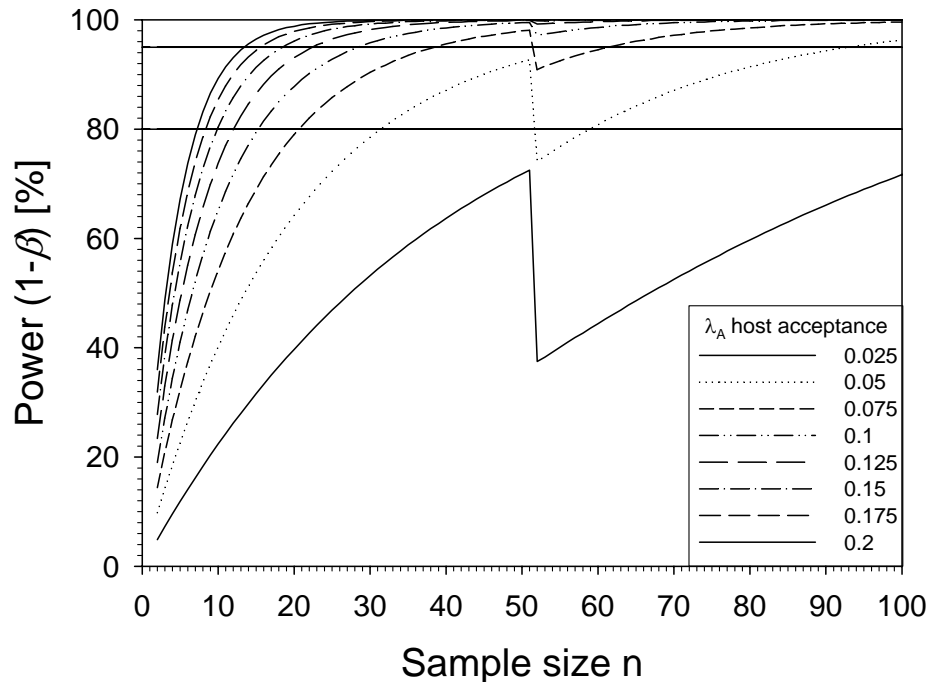


Figure 2. Statistical power for a non-target test based upon an exact binomial test under the null hypothesis H_0 of an acceptance rate of non-target hosts of $\lambda_0 = 0.001$. The tests specifies the Power, given one does not accept ($P = 0.05$) the alternative hypothesis H_A that assumes an acceptance rate of λ_A given in the figure legend, and given fewer than r host were accepted by the parasitoid, with $r = 1$ for $n < 52$ and $r = 2$ for $n \leq 52$. Horizontal lines mark 80 and 95 % power. See text for details.

680

A second issue that cannot be solved statistically are optimal designs for choice or no-choice tests. As Withers and Mansfield (2005) point out, there are different benefits associated with no-choice and choice tests. If we think about the statistical analysis of such tests, choice tests can be problematic. Since the same animal will be confronted with target and non-target hosts, we usually wish to obtain more than a single data point for each animal, i.e. for example acceptance rates of target *and* non-target hosts. Thus, some sort of repeated measurement design has to be used in this case (alternatively, only the acceptance rate of non-target hosts is analyzed; see the above example). While analysis of such dependent data is generally possible (see, e.g., GEE models in Generalized Linear Models), an additional problem exists, if target and non-target hosts or prey are exposed to the natural enemy simultaneously. The acceptance of non-target hosts or prey may well depend upon the frequency of target and non-target hosts within the experimental arena. If this is so, every target prey that is removed or every target host that is accepted and that is not replaced alters the experimental conditions of the experiment, and the acceptance of any given host or prey may depend on the current availability of alternative hosts or prey. If exploited hosts or prey cannot be replaced immediately, simultaneous choice test may become almost impossible to interpret. Thus, from a statistical point of view, sequential no-choice tests may be favourable (see Singer 1986 for a discussion), where all effects like the sequence of species presented, the motivational status of the tested insect can be statistically controlled for. Yet, these two tests may lead to very different outcomes biologically (Withers and Mansfield 2005) and thus both tests have their merit, despite the problems associates with simultaneous choice tests.

CONCLUSIONS

In the past, decisions to use or reject a species as biological control agent were more often based on gut feeling than exact scientific methods. Today, sound host range tests are a prerequisite in the evaluation of biological control agents. However, despite great advances in the field (Van Driesche and Murray 2004; van Lenteren *et al.* 2006; Withers and Mansfield 2005), some issues on the interpretation of data are still unsolved. This paper advocates for a rigorous use of Power analyses to obtain a measure of confidence if one does not find significant deviations from the null hypothesis of no effect. Further, the most powerful statistical methods should be used when sample sizes are a limiting factor in insect host range studies. Despite the introduction of a number of new statistical tools, some of the basic statistical problems in host range testing are still unresolved. For example, no standard test is available to calculate a measure of confidence for an experiment where one has not found acceptance of the non-target host in n replicates (but see above for a possible method). Until now, researchers working in biological control are largely dependent on educated guesses with respect to how many replicates would be necessary to decide that an insect does not accept a given non-target host (D. Sands, J. van Lenteren, pers. comm.). Thus, further advances in statistical techniques are clearly needed.

ACKNOWLEDGEMENTS

I am grateful to W. Wosniok, Institute of Statistics, University of Bremen for the SAS code of the exact binomial test used in this chapter. I wish to thank P. Mason for the invitation to provide this chapter and his comments on an earlier version of this manuscript.

681

REFERENCES

- Cohen, J. 1998. "Statistical Power Analysis for the Behavioural Sciences", 2nd Edition. Lawrence Erlbaum, Hillsdale, New Jersey, USA.
- Cox, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society Series B* 34, 187-220.
- Crawley, M. J. 1993. "GLIM for Ecologists". Blackwell Scientific Publications, Oxford, U.K.
- Crawley, M. J. 2002. "Statistical Computing: An Introduction to Data Analysis Using S-Plus". John Wiley and Sons Ltd.
- Grafen, A., and Hails, R. 2002. "Modern Statistics for the Life Sciences". Oxford University Press, Oxford, U.K.
- Heffner, R. A., Butler, M. J., and Reilly, C. K. 1996. Pseudoreplication revisited. *Ecology* 77, 2558-2562.
- Hilborn, R., and Mangel, M. 1997. "The Ecological Detective. Confronting Models With Data". Princeton University Press, Princeton, New Jersey, U.S.A.

- Hoffmeister, T. S., Babendreier, D., and Wajnberg, E. 2006. Statistical Tools to Improve the Quality of Experiments for Assessing Non-Target Effects. In "Environmental Impact of Invertebrates for Biological Control of Arthropods: Methods and Risk Assessment" (F. Bigler, D. Babendreier, and U. Kuhlmann, Eds.). CABI Publishing, Wallingford, Oxon (in press).
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**, 187-211.
- McCullagh, P., and Nelder, J. 1989. "Generalized Linear Models", 2nd Edition. Chapman and Hall, New York, New York, U.S.A.
- Porter, S. D., Fowler, H. G., Campiolo, S., and Pesquero, M. A. 1995. Host specificity of several *Pseudacteon* (Diptera: Phoridae) parasites of fire ants (Hymenoptera: Formicidae) in South America. *Florida Entomologist* **78**, 70-75.
- Quinn, G. P., and Keough, M. J. 2002. "Experimental Design and Data Analysis for Biologists." Cambridge University Press, Cambridge, Massachusetts, U.S.A.
- Ramirez, C. C., Fuentes-Contreras, E., Rodriguez, L. C., and Niemeyer, H. M. 2000. Pseudoreplication and its frequency in olfactometric laboratory studies. *Journal of Chemical Ecology* **26**, 1423-1431.
- Ruxton, G. D., and Colegrave, N. 2003. "Experimental Design for the Life Sciences". Oxford University Press, Oxford, U.K.
- 682 Singer, M. C. 1986. The Definition and Measurement of Oviposition Preference in Plant-Feeding Insects. In "Insect-Plant Interactions" (J. R. Miller, and T. A. Miller, Eds.), pp. 65-94. Springer-Verlag, New York, New York, U.S.A.
- Thomas, L. 1997. Retrospective power analysis. *Conservation Biology* **11**, 276-280.
- Thomas, L., and Krebs, C. J. 1997. A review of statistical power analysis software. *Bulletin of the Ecological Society of America* **78**, 126-139.
- van Alphen, J. J. M., Bernstein, C., and Driessen, G. 2003. Information acquisition and time allocation in insect parasitoids. *Trends in Ecology and Evolution* **18**, 81-87.
- Van Driesche, R. G., and Murray, T. J. 2004. Overview of Testing Schemes and Designs Used to Estimate Host Ranges. In "Assessing Host Ranges for Parasitoids and Predators used for Classical Biological Control: A Guide to Best Practice" (R. G. Van Driesche, and R. Reardon, Eds.), pp. 68-89. United States Department of Agriculture, Forest Health Technology Enterprise Team, Morgantown, West Virginia, U.S.A.
- van Lenteren, J. C., Cock, M. J. W., Hoffmeister, T. S., and Sands, D.P.A. 2006. Host Specificity in Arthropod Biological Control: Methods for Testing and Interpretation of the Data. In "Environmental Impact of Invertebrates for Biological Control of Arthropods: Methods and Risk Assessment" (F. Bigler, D. Babendreier, and U. Kuhlmann Eds.). CABI Publishing, Wallingford, Oxon, U.K. (in press).