Penultimate version. If citing, please refer instead to the published version in Journal of Applied Entomology.

Common pitfalls when testing additivity of treatment mixtures with χ^2

Philip Pallmann[1]*, Frank Schaarschmidt[1]

[1] Institute of Biostatistics, Leibniz University Hannover, Germany

* Correspondence: Philip Pallmann (corresponding author), Institute of Biostatistics, Leibniz University Hannover, Herrenhäuser Straße 2, 30419 Hannover, Germany. E-mail: pallmann@biostat.uni-hannover.de

Abstract

Studying interactions of multiple pesticides applied simultaneously in a mixture is a common task in phytopathology. Statistical methods are employed to test whether the treatment components influence each other's efficacy in a promotive or inhibitory way (synergistic or antagonistic interaction) or rather act independent of one another (additivity). The trouble is that widely used procedures based on χ^2 tests are often seriously flawed, either because people apply them in a preposterous way or because the method simply does not fit the problem at hand. Browsing recent volumes of entomological journals, we found that numerous researchers have (in all likelihood unwittingly) analyzed their data as if they had had a sample size of 100 or, equally bad, a sample size of one! We show how to avoid such poor practices and further argue that χ^2 testing is, even if applied correctly (meaning that no technical errors are made), a limited-purpose tool for assessing treatment interactions.

Keywords: interaction, synergism, antagonism, integrated pest management, statistics

Introduction

Applying mixtures of several plant protection measures is a popular strategy for integrated pest management. A common research goal is to assess the interactions of treatment combinations, which are conventionally classified as additive, synergistic, or antagonistic. Treatments showing a synergistic interaction enhance each other's effect when applied together, hence they are particularly interesting for pest control purposes. However, the opposite may occur as well i.e., treatments impeding each other so that the compound treatment performs worse than expected under additivity of effects.

The whole idea of investigating (departures from) additivity first of all hinges on a reasonable definition of the term "additive". Two prevalent reference models are:

- Bliss independence (Bliss 1939): the components of the mixture have different modes of action and therefore do not interact.
- Loewe additivity (Loewe 1953): the components of the mixture have a shared mode of action and differ only in their potency.

Both models come with assumptions and implications that are often debatable and hard to verify in practice; see the review articles of Goldoni and Johansson (2007) and Cedergreen *et al.* (2008). The concept of Loewe additivity cannot be applied to single-dose experiments, which are common (though not necessarily recommendable) in phytopathology.

Various approaches for analyzing treatment interactions have been circulating in the entomological literature for decades. They were often developed before the advent of modern computation, and from a statistical point of view, many of them are at best clumsy and at worst erroneous, as will be illustrated by several examples from entomological publications. We will have a focus on χ^2 testing, which appears to be the standard procedure in phytopathology. The goal of this article is a) to point out and correct prevalent mistakes

occurring with χ^2 analyses, b) to clarify the underlying notion of additivity, c) to raise awareness for the severe limitations of χ^2 tests on additivity, and d) to suggest alternatives.

The Method

When investigating the interaction of compounds applied simultaneously in a pest management trial, the experimental setup typically comprises three treatments: a single substance A, a single substance B, and their 1:1 mixture, denoted by AB (1:1 means that both single doses are added together i.e., a 1+1 mixture to be precise). Each of them is applied to a number of individuals (usually insects, etc.), and these sample sizes are denoted by n_A, n_B, and n_{AB}. After some exposure time each individual is classified into either of two mutually exclusive categories such as dead and alive. The observed numbers of dead individuals in each treatment group are denoted by x_A, x_B, and x_{AB} i.e., x_A is a whole number that can take values between 0 and n_A. From these numbers we can compute estimates of the mortality in each treatment group (p_A, p_B, and p_{AB}), expressed as the proportion of dead insects i.e., as a number between 0 and 1. Nonetheless, our responses are inherently counts, quite unlike outcomes that occur naturally as percentages (e.g., relative activity of an enzyme).

Following the Bliss-type notion of additivity used by Finney (1952), one can find the expected mortality in the 1:1 mixture AB from the mortality proportions of the two single treatments as

$$p_{AB_{\text{exp}}} = p_A + (1 - p_A)p_B . \qquad (1)$$

In practice we plug in the estimators $\hat{p}_A = \frac{x_A}{n_A}$ and $\hat{p}_B = \frac{x_B}{n_B}$ for p_A and p_B as proposed by Finney (1952).

Given that there were n_{AB} individuals under observation in the AB treatment, we would expect $x_{AB_{\text{exp}}} = n_{AB} p_{AB_{\text{exp}}}$ deaths if substances A and B were acting additively. Finney (1952) proposes

then to use the χ^2 test (Pearson 1900) with one degree of freedom to assess whether the observed number x_{AB} deviates from the expected $x_{AB_{\exp}}$ by such an extent that the deviation can be deemed significant. The test statistic is computed as

$$\chi^2 = \sum_{i=1,2} \frac{(x_i^{obs} - x_i^{\exp})^2}{x_i^{\exp}} \qquad (2)$$

where $x_1^{obs} = x_{AB}$ and $x_2^{obs} = n_{AB} - x_{AB}$ denote the observed numbers of dead and surviving insects in the AB treatment, and $x_1^{\exp} = x_{AB_{\exp}}$ and $x_2^{\exp} = n_{AB} - x_{AB_{\exp}}$ the corresponding expected numbers under the null hypothesis of additivity. If the χ^2 test statistic is larger than the critical value $\chi^2_{0.95,df=1} = 3.8415$, we may conclude that the deviation from additivity is significant at the 5% level.

It should be noted that the test statistic in Eqn. (2) follows a χ^2 distribution if and only if it is computed with counted and expected *numbers* of insects (x_{AB} and $x_{AB_{\exp}}$) under the null hypothesis of additivity, and if $p_{AB_{\exp}}$ is based on theoretically expected proportions and not sample estimates. Its form is motivated by the assumption that for counts the variance is a function of the mean (see Appendix A). This relation between mean and variance will not be the same if we express mortalities as *proportions* or *percentages*. Thus, if one plugs proportions or percentages into Eqn. (2) instead of the counted and expected *numbers* of insects, the resulting statistic does not follow a χ^2 distribution anymore. Comparing such a test statistic with a critical value from the χ^2 distribution is likely to produce far too many or too few rejections of the null hypothesis of additivity, depending on the circumstances.

One keystone with the analysis of treatment interactions is a justifiable notion of the term *additive* when dealing with mortalities. Finney's formula for $p_{AB_{\exp}}$ (Eqn. 1) rests upon *one* reasonable definition of additivity (but certainly *not the only* possible definition): the assumption that both agents, A and B, act independently in the sense of Bliss (1939). That is, for a single individual, the probability to die due to agent B does not depend on whether it

dies from or survives application of agent A, and *vice versa*. In other words, the proportion of insects killed by A is the same in the subgroup of insects that survive B as it would have been in the subgroup that have already been killed by B, and *vice versa*. Under this assumption of independence, the probability to survive both A and B is simply the product of the probability to survive A (which is $1-p_A$) and the probability to survive B (which is $1-p_B$). This becomes obvious when rewriting Eqn. (1) as

$$p_{AB_{\text{exp}}} = 1-(1-p_A)(1-p_B)\,.$$

Another intuitive explanation is that we expect the single mortalities to add up ($p_A + p_B$), but have to subtract that proportion of individuals that would die from both components A and B under the assumption of independent action, which is simply the product of the two single mortalities, $p_A p_B$:

$$p_{AB_{\text{exp}}} = p_A + p_B - p_A p_B\,.$$

Loosely speaking, if you die for two reasons you are dead only once. A vivid illustration of the idea behind independent action is given in Berenbaum (1981).

This clarifies the following: Eqn. (1) makes sense only for mortalities expressed as *probabilities* of dying, or equivalently, as *proportions* of dead insects i.e., as values between 0 and 1. It makes no sense to plug *percentage* mortalities or observed *numbers* of dead individuals into the formula. We cannot easily interchange the probability of surviving (1 – mortality) with the probability of dying (mortality) in order to compute the expected mortality under additivity. The definition ensures that $p_{AB_{\text{exp}}}$ cannot exceed one, as we would never expect more than 100% mortality without interference of supernatural events.

Understanding the background of Finney's definition of additivity is also useful when it comes to extending the approach to mixtures of more than two single treatments. Using the same notion of independent action in a mixture of three components A, B, and C, we would expect the mortality to be

$$p_{ABC_{\exp}} = 1 - (1 - p_A)(1 - p_B)(1 - p_C)\,.$$

Flaws in Applications of Finney's Test on Additivity

Motivated by a statistical consultancy (Otieno *et al.* 2015), we discovered a number of entomological publications attempting to evaluate the independence of treatments applied as mixtures. We found dozens of papers where the description of methods and the presentation of results were at least highly ambiguous and at worst completely flawed. This aroused our curiosity as to where all these misconceptions had their origins. A closer look revealed that the authors of all these publications refer to Finney's seminal classic *Probit Analysis* (1952) and state that they conducted χ^2 tests "modified" by McVay *et al.* (1977). So what does their "modification" entail?

First, they propose to calculate the expected *percentage* mortality (under additivity) of the mixture, $P_{AB_{\exp}}$, from the observed mortality *percentages* of the single treatments, $P_A = 100 \times \frac{x_A}{n_A}$ and $P_B = 100 \times \frac{x_B}{n_B}$, as

$$P_{AB_{\exp}} = P_A + P_B - P_A P_B\,.$$

Applying this formula to the data examples in McVay *et al.* (1977) yields expected mortalities of –440 and –1063, which is obviously absurd. This indicates that one main problem in McVay *et al.* (1977) is obscure notation and ambiguous use of symbols; they do not clearly distinguish between *numbers* of dead insects, mortality *proportions*, and *percentage* mortalities.

The second part of their "modification" is to calculate a χ^2 statistic via

$$\chi^2 = \frac{(P_{AB} - P_{AB_{\exp}})^2}{P_{AB_{\exp}}}$$

where P_{AB} is the observed "mortality" with the mixture, and compare it to the 95% quantile of a χ^2 distribution with one degree of freedom. Here McVay *et al.* withhold the information whether they use mortality proportions or percentages. Indeed, both is wrong: plugging proportions into the χ^2 test implies a sample size of one, whereas percentages suggest there were 100 replications per treatment group!

Another frequently cited reference for the "modified" χ^2 tests is Salama *et al.* (1984); this paper basically reproduces the opaque descriptions or actual errors committed by McVay *et al.* and adds another misguided remark: the authors claim that a mixture of two treatments each of which leads to 25% mortality should be expected to result in 50% mortality. This notion of additivity is obviously inconsistent with Finney's definition (and with common biological sense). It cannot even be applied to the plausible range of single treatment mortalities. What if each single treatment led to 60% mortality? We would by no means expect 120% mortality for the mixture, not even 100%, but rather something between 60 and 100%. According to Finney's formula, mixing two treatments with 25% mortality each should lead to 43.75% but certainly not 50% mortality in the combination.

Even though the instructions of McVay *et al.* (1977) and Salama *et al.* (1984) are ambiguous and at the end of the day unfeasible, researchers have not been discouraged from following them down to the present day. A (non-systematic) review of recent papers in entomology brings some substantial errors to light: the most prevalent mistake (e.g., in Ansari *et al.* 2008; Koppenhöfer and Fuzy 2008; Gosselin *et al.* 2009; Baloyi *et al.* 2012; Ma *et al.* 2013; Zhou *et al.* 2013) is to compute the χ^2 statistic from the mortality *percentages*, leading to overoptimistic results whenever the sample size is less than 100 and to unnecessary pessimistic conclusions otherwise.

Jazzar and Hammad (2004), Hammad and McAuslane (2006), and Kullik *et al.* (2011) insert *proportional* mortalities to the formula of the χ^2 test statistic as if their total sample size were one! They yield χ^2 values so tiny that there is almost no way for them to exceed the 95%

reference quantile of 3.8415. Hammad and McAuslane and Kullik *et al.* settle for the conclusion of no interaction although their data clearly indicate synergism (they have mortality proportions of e.g., $\hat{p}_A + (1 - \hat{p}_A)\hat{p}_B = 0.26$ versus $\hat{p}_{AB} = 0.62$, and $\hat{p}_A + (1 - \hat{p}_A)\hat{p}_B = 0.66$ versus $\hat{p}_{AB} = 0.97$). Jazzar and Hammad (2004), probably intuiting their fallacy, try to smooth it out by using the 5% χ^2 quantile (0.0039) as a reference value, which makes things all the worse.

This little survey of recent publications is alarming but by far not exhaustive; not least because other authors (e.g., Kazemi-Dinan *et al.* 2014) entirely fail to present any numerical results so that reproducing their analysis is made impossible.

Additional Problems with the χ^2 Test on Additivity

Even without the calamitous "modifications", the χ^2 procedure described up to here is far from being a silver bullet. In fact, it is a rather limited technique that comes with restrictions and downsides:

- As already noted by Finney (1952, p. 145), the expected number of deaths under additivity is computed from the *estimated* mortalities of the single treatments. Hence this estimate is subject to uncertainty, and this is not accounted for in the above formulation of the χ^2 test, so it is at best an approximate test. This is fundamentally different from the χ^2 test as applied in genetics, where the expected values are *given* by theory e.g., by Mendel's laws. Improved tests on additivity that account for the uncertainty of the estimated expectation under additivity, $\hat{p}_{AB_{\text{exp}}}$, may be formulated in generalized linear models (GLMs) for binomial data (McCullagh and Nelder 1989), as illustrated in Appendix B and the supplementary material.

- The above calculation rule for $p_{AB_{exp}}$ (Eqn. 1) is only justifiable if the doses of the single treatments are added for the mixture (i.e., the 1:1 is, properly speaking, a 1+1 mixture). The method cannot (without further modifications) deal with experiments involving combinations that are composed of *fractions* or *multiples* of the single doses. In comparison, the solution in a GLM framework is straightforward (see Appendix B).
- Many experimental setups are too complex to be reflected adequately by simple χ^2 tests e.g., when they involve blocks, subsampling, technical replications, additional covariates, etc. Some *ad hoc* strategies distort the problem so that it "fits" the χ^2 solution by ignoring randomization structures, leaving out covariates, and so on. Again, proper solutions for such problems may be found by formulating appropriate GLMs.
- Ideally, experiments should involve replicated observations for the same treatments. In the case of entomological trials, the observed numbers of dead insects in replicated experimental units subjected to the same treatment may show more variation than expected under simplistic assumptions like the binomial or Poisson distribution for such count data. This phenomenon is known as overdispersion (McCullagh and Nelder 1989) and can be accounted for in GLMs (see the supplementary material for an illustration). If an experiment exhibits overdispersion, it is inappropriate to sum up dead and surviving individuals over replications of the same treatment and then plug the "simplified" data into the χ^2 formula. Such an analysis will underestimate the variance of the estimated mortalities and thus tend to overstate the importance of observed deviations from additivity.
- Moreover, GLMs allow for assessing (lack of) additivity in treatment mixtures also for other types of count data, like the number of offspring or eggs, based on a quasi-

Poisson assumption. An example of such an analysis where the theoretical upper limit of counts remains unknown is provided in Otieno *et al.* (2015).

- When assessing dose-response relationships it is common (and advisable!) to consider multiple dose levels. For assessing additivity of several dosages for single treatments and several mixtures under the assumption of nonlinear dose-response relationships, Ritz and Streibig (2014) provide a comprehensible overview as well as free software.
- χ^2 tests produce p-values as a measure of significance, but confidence limits on a biologically interpretable scale would often be preferable (Gardner and Altman 1986). We show how to obtain and interpret them in the supplementary material.
- The method is unsuitable for "verifying" additivity, as alluded by the phrase "absence of evidence is not evidence of absence" (Altman and Bland 1995). A large p-value simply means that the data do not contain enough evidence (or sample size) to reject the null hypothesis at significance level α, but this is not in the least a "proof" of additivity! If the aim is to demonstrate additivity, we refer to equivalence tests as suggested in Stork *et al.* (2006).

Discussion

When faced with the task of exploring interactions of pest control agents applied simultaneously as mixtures, many biologists consider the χ^2 test as a panacea. We suspect its frequent use in similar publications has made the method appear trustworthy, or at least citable. Acting in opposition to this common belief, we have accumulated evidence that this practice is often inadequate, either because the test is carried out in a faulty manner or because it is outright unsuitable to solve the research question.

We discourage from referring to papers like McVay *et al.* (1977) and Salama *et al.* (1984) as (intermediate) references for Finney's test on additivity as they are likely to lead astray, with three main consequences:

1. Plugging mortality *proportions* into the χ^2 formula for binomial counts creates a procedure that is extremely unlikely to show significance, even in the presence of crystal clear synergism or antagonism.
2. Plugging in *percentages* will lead to the significance being overstated if the actual sample size is below 100, and understated if it is (much) larger than 100.
3. Finney's test is applied on the scale of observed and expected mortality counts and will therefore overrate significance because the variance due to the estimated mortality being plugged in is ignored (Finney 1952, p. 145), and also possibility of overdispersion that is typically observed in real-world experiments (McCullagh and Nelder 1989).

However, the high number of citations of McVay *et al.* and Salama *et al.* (despite their incorrect or ambiguous descriptions of the method) suggests that there is a severe lack of accessible, comprehensible, and statistically sound texts on the topic.

Most real-world experiments cannot be adequately analyzed using Finney's χ^2 test anyway. It is simple at first sight, but this alleged virtue can quickly turn into a weakness when the data do not „fit". In particular it cannot account for 1) the uncertainty of the estimated single treatment mortalities, 2) potential overdispersion, 3) additional experimental structures like blocked replications or covariates, and 4) several dosages of the single treatments of interest as well as several mixtures (involving different fractions of the individual treatments). GLMs with a (quasi-)binomial assumption can account for these four problems. It should be noted, however, that various definitions of additivity are possible in GLMs, depending on the link function chosen (log, logit, etc.). We outline a GLM-based analysis in Appendix B and

provide R code as supplementary material. For a general introduction to binomial GLMs we recommend the textbooks of Faraway (2006, chapter 2) and Dobson and Barnett (2008).

Acknowledgements

The authors are grateful to the editor and two anonymous referees whose comments have led to a clearer presentation of several arguments, and sound mathematical notation.

References

Altman DG, Bland JM, 1995. Absence of evidence is not evidence of absence. Br. Med. J. 311, 485.

Ansari MA, Shah FA, Butt TM, 2008. Combined use of entomopathogenic nematodes and *Metarhizium anisopliae* as a new approach for black vine weevil, *Otiorhynchus sulcatus*, control. Entomol. Exp. Appl. 129, 340-347.

Baloyi MA, Laing MD, Yobo KS, 2012. Use of mixed cultures of biocontrol agents to control sheep nematodes. Vet. Parasitol. 184, 367-370.

Berenbaum MC, 1981. Criteria for analyzing interactions between biologically active agents. Adv. Cancer Res. 35, 269-335.

Bliss CI, 1939. The toxicity of poisons applied jointly. Ann. Appl. Biol. 26, 585-615.

Cedergreen N, Christensen AM, Kamper A, Kudsk P, Mathiassen SK, Streibig JC, Sørensen H, 2008. A review of independent action compared to concentration addition as reference models for mixtures of compounds with different molecular target sites. Environ. Toxicol. Chem. 27, 1621-1632.

Dobson AJ, Barnett AG, 2008. An introduction to generalized linear models. 3rd edition. Chapman & Hall/CRC, Boca Raton, FL.

Faraway JJ, 2006. Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models. Chapman & Hall/CRC, Boca Raton, FL.

Finney DJ, 1952. Probit analysis. 2nd edition. Cambridge University Press, London, UK.

Gardner MJ, Altman DG, 1986. Confidence intervals rather than p values: estimation rather than hypothesis testing. Br. Med. J. 292, 746-750.

Goldoni M, Johansson C, 2007. A mathematical approach to study combined effects of toxicants *in vitro*: evalutation of the Bliss independence criterion and the Loewe additivity model. Toxicol. In Vitro 21, 759-769.

Gosselin ME, Bélair G, Simard L, Brodeur J, 2009. Toxicity of spinosad and *Beauveria bassiana* to the black cutworm, and the additivity of subletal doses. Biocontrol Sci. Techn. 19, 201-217.

Hammad EAF, McAuslane HJ, 2006. Effect of *Melia azedarach* L. extract on *Bemisia argentifolii* (Hemiptera: Aleyrodidae) and its biocontrol agent *Eretmocerus rui* (Hymenoptera: Aphelinidae). Environ. Entomol. 35, 740-745.

Hothorn T, Bretz F, Westfall P, 2008. Simultaneous Inference in general parametric models. Biom. J. 50, 346-363.

Jazzar C, Hammad EAF, 2004. Efficacy of multiple biocontrol agents against the sweet potato whitefly *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae) on tomato. J. Appl. Entomol. 128, 188-194.

Kazemi-Dinan A, Thomaschky S, Stein RJ, Krämer U, Müller C, 2014. Zinc and cadmium hyperaccumulation act as deterrents towards specialist herbivores and impede the performance of a generalist herbivore. New Phytol. 202, 628-639.

Koppenhöfer AM, Fuzy EM, 2008. Early timing and new combinations to increase the efficacy of neonicotinoid-entomopathogenic nematode (Rhabditida: Heterorhabditidae) combinations against white grubs (Coleoptera: Scarabaeidae). Pest Manag. Sci. 64, 725-735.

Kullik SA, Sears MK, Schaafsma AW, 2011. Sublethal effects of Cry 1F Bt corn and clothianidin on black cutworm (Lepidoptera: Noctuidae) larval development. J. Econ. Entomol. 104, 484-493.

Loewe S, 1953. The problem of synergism and antagonism of combined drugs. Arzneim. Forsch. 3, 285-290.

Ma Y, Zhang Y, Chen RR, Ren XL, Wan PJ, Mu LL, Li GQ, 2013. Combined effects of three crystalline toxins from *Bacillus thuringiensis* with seven proteinase inhibitors on beet armyworm, *Spodoptera exigua* Hübner (Lepidoptera: Noctuidae). Pest. Biochem. Physiol. 105, 169-176.

McCullagh P, Nelder JA, 1989. Generalized linear models. 2nd edition. Chapman & Hall/CRC, Boca Raton, FL.

McVay JR, Gudauskas RT, Harper JD, 1977. Effects of *Bacillus thuringiensis* nuclear-polyhedrosis virus mixtures on *Trichoplusia ni* larvae. J. Invertebr. Pathol. 29, 367-372.

Otieno JA, Pallmann P, Poehling HM, 2015. Combination of soil-applied azadirachtin with entomopathogens for integrated management of western flower thrips. Accepted for publication in J. Appl. Entomol., doi: 10.1111/jen.12242.

Pearson K, 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine 50, 157-175.

Ritz C, Streibig JC, 2014. From additivity to synergism: a modelling perspective. Synergy 1, 22-29.

Salama HS, Foda MS, Zaki FN, Moawad S, 1984. Potency of combinations of *Bacillus thuringiensis* and chemical insecticide on *Spodoptera littoralis* (Lepidoptera: Noctuidae). J. Econ. Entomol. 77, 885-890.

Stork LG, Gennings C, Carchman RA, Carter WH, Pounds J, Mumtaz M, 2006. Testing for additivity at select mixture groups of interest based on statistical equivalence testing methods. Risk Anal. 26, 1601-1612.

Zhou LT, Jia S, Wan PJ, Kong Y, Guo WC, Ahmat T, Li GQ, 2013. RNA interference of a putative S-adenosyl-L-homocysteine hydrolase gene affects larval performance in *Leptinotarsa decemlineata* (Say). J. Insect Physiol. 59, 1049-1056.

Appendix A

We can rearrange Eqn. (2) in a way that its motivation from the binomial distribution becomes explicit. Plugging $x_1^{obs} = x_{AB}$ and $x_2^{obs} = n_{AB} - x_{AB}$ as well as $x_1^{\exp} = x_{AB_{\exp}}$ and $x_2^{\exp} = n_{AB} - x_{AB_{\exp}}$ into the formula of the test statistic yields

$$\chi^2 = \frac{(x_{AB} - x_{AB_{\exp}})^2}{x_{AB_{\exp}}} + \frac{((n_{AB} - x_{AB}) - (n_{AB} - x_{AB_{\exp}}))^2}{n_{AB} - x_{AB_{\exp}}}$$

which can be simplified to

$$\chi^2 = \frac{(x_{AB} - x_{AB_{\exp}})^2}{x_{AB_{\exp}}} + \frac{(x_{AB} - x_{AB_{\exp}})^2}{n_{AB} - x_{AB_{\exp}}}.$$

Reduction to a common denominator gives us

$$\chi^2 = \frac{(n_{AB} - x_{AB_{\exp}})(x_{AB} - x_{AB_{\exp}})^2 + x_{AB_{\exp}}(x_{AB} - x_{AB_{\exp}})^2}{x_{AB_{\exp}}(n_{AB} - x_{AB_{\exp}})}$$

and with a simplified numerator we have

$$\chi^2 = \frac{n_{AB}(x_{AB} - x_{AB_{\exp}})^2}{x_{AB_{\exp}}(n_{AB} - x_{AB_{\exp}})}.$$

Rewriting $x_{AB_{\exp}}$ in the denominator as $n_{AB}p_{AB_{\exp}}$ and subsequently placing n_{AB} outside the brackets leaves us with

$$\chi^2 = \frac{n_{AB}(x_{AB} - x_{AB_{\exp}})^2}{n_{AB} p_{AB_{\exp}}(1 - p_{AB_{\exp}}) n_{AB}}$$

where n_{AB} cancels out so that we end up with

$$\chi^2 = \frac{(x_{AB} - x_{AB_{\exp}})^2}{n_{AB} p_{AB_{\exp}}(1 - p_{AB_{\exp}})}.$$

Now it is obvious that the form of the denominator resembles the variance $Var(x) = np(1-p)$ of a counted number x that follows the binomial distribution $x \sim Bin(n, p)$.

Appendix B

Yet another reformulation of Eqn. (1) as

$$1 - p_{AB\exp} = (1 - p_A)(1 - p_B)$$

enables us to carry out (large-sample) tests and construct related confidence intervals that

- account for the uncertainty due to estimation of p_A and p_B,
- allow for an interpretation of the biological relevance of the deviation from additivity in terms the ratios of the proportion of survivors, and
- can be easily extended to the application in GLMs that account for overdispersion or additional effects in the model.

Under the null hypothesis of Bliss additivity, the proportions of surviving insects would be related as

$$1 - p_{AB} = (1 - p_A)(1 - p_B),$$

which corresponds to

$$H_0 : \frac{(1 - p_{AB})}{(1 - p_A)(1 - p_B)} = 1,$$

or, on the log-scale

$$H_0 : \log(1 - p_{AB}) - (\log(1 - p_A) + \log(1 - p_B)) = 0.$$

That is, one can use a GLM with logarithmic link for the proportions of survivors and assumption of a binomial distribution to obtain estimates and standard errors for all treatments involved, and then asymptotically test the above hypothesis on the log scale using a linear combination of the parameters of this model. For the simple single-dose design that was described above, the model is

$$(n_i - x_i) \sim Bin(n_i, 1 - p_i)$$

where $n_i - x_i$ is the observed number of surviving insects in the *i*th treatment group, and 1 – p_i is the unknown proportion of surviving insects in treatment group *i*, with index $i = A, B, AB$. Fitting the model yields estimates for the log-proportion of the *i*th treatment, b_i, and the corresponding estimated standard errors. The linear combination of interest is defined by the coefficients c_i: $c_A = -1$, $c_B = -1$, and $c_{AB} = 1$ so that the hypothesis above can be written as

$$H_0 : \sum_i c_i b_i = 0$$

where $\log(1 - p_i) = b_i$. Even if the mixture AB is not a 1+1 combination of the single dosages, it is straightforward to adapt the coefficients c_i. Suppose the mixture consists of 50% the single dose of A and 30% the single dose of B, then we use $c_A = -0.5$ and $c_B = -0.3$, and c_{AB} remains unchanged.

For the computational details of this hypothesis test and the compatible confidence intervals based on the estimates obtained from the GLM, we refer to Hothorn *et al.* (2008). R code that applies this method to a toy example (also in the presence of overdispersion) is provided as supplementary material.