MedCrave
Step into the World of Research

Biometrics & Biostatistics International Journal

Research Article

Open Access

CrossMark

Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion

Abstract

Many bioassays that assess toxicity or mutagenicity give rise to clustered binomial data with relatively few replicated experimental units per treatment group. Confidence intervals for the risk ratio to control can then be used to interpret the relevance of effect size or test hypotheses of superiority, non-inferiority or equivalence. A frequently observed property of clustered binomial data is overdispersion. So far, the available large-sample confidence interval methods for ratios of proportions in presence of overdispersion have been validated for use in epidemiological settings with high numbers of clusters per exposure group.

In this paper, the coverage probability and symmetry of non-coverage of several available methods is investigated in an extensive Monte-Carlo simulation study, for the small number of replications that are typical for a number of bioassays. An additional method is proposed that combines profile deviance intervals with the method of variance recovery. So far available confidence intervals have far too low coverage probabilities in the simulated settings. Their performance can be improved by restricting estimators of dispersion not to fall below the binomial variance and by using pooled dispersion estimators. The newly proposed method outperforms the so far available methods by showing coverage probabilities closest to the nominal level. All discussed methods are made available in an add-on package for the R software.

Keywords: ratio of proportions, extra-binomial variance, beta-binomial, monte carlo simulation, coverage probability, bioassay, small sample performance

Volume 2 Issue 7 - 2015

Frank Schaarschmidt
Leibniz Universität Hannover, Germany

Correspondence: Frank Schaarschmidt, Institute of Biostatistics, Leibniz Universität Hannover, Herrenhäuser Straße 2, D-30419 Hannover, Germany, Tel +49 511 762 5821, Fax: +49 511 762 4966, Email schaarschmidt@biostat.uni-hannover.de

Received: August 7, 2015 | **Published:** October 26, 2015

Abbreviations: FN, fieller-type intervals applied on the observed proportions; BM, binomial MOVER-R-Wilson method after summing-up observed counts across replications; LOD, asymptotic method on the log-scale; accounting for overdispersion via residual variation; FOD, fieller-bailey-type interval accounting for over dispersion via residual variance; LBB, delta method on the log-scale under the beta-binomial assumption; FBB, fieller-bailey-type interval under beta-binomial assumption; QBM, MOVER-R for quasibinomial profile deviance intervals

Introduction

Various bioassays give rise to replicated binomial count data. For example, in ecotoxicological assays, fish larvae or daphnids in several tanks are exposed to different dosages of a substance and the number of dead or immobile animals per tank is used to assess the hazardousness of the substance. Usually, a small number of replicated tanks are used for each dosage under consideration. If some experimental conditions differ between tanks such that the proportion of dead or immobile daphnids is effected, counted numbers may show higher variance than expected under the binomial assumption, i.e., extra-binomial variability or overdispersion. Similar situations arise in the *in-vivo* micronucleus assay: the number of cells showing micronuclei is counted for a given number of exposed cells for each (randomized) animal, with the aim to assess the substances' potential to cause cytogenetic damage. Also here, a limited number of replications per dosage is performed, such that differences between animals in the *in-vivo* micronucleus animals may cause overdispersion. In summary, bioassays that lead to binomial data, often contain clustered replication, and thus make it possible to account for overdispersion in the data. However, the number of replications or clusters per treatment group that allows to assess the extend of overdispersion, is rather limited.

In the statistical analysis of such bioassays, major interest is usually in comparisons to the untreated control group. While a test on significance for the overall effect of dosage of the substance may be of preliminary interest, usually more detailed interpretation for the single dosages is required: Confidence intervals for the effect of given dosages compared to the untreated control or a positive control are required to interpret the toxicological relevance of the observed effect size. Tests of non-inferiority (or equivalence) for given dosages compared to an untreated control[1] may be more important than an overall test on significant change in the event rates: In toxicological assessment, confidence is needed primarily when claiming no effect. Both approaches require an interpretable definition of the change of the rate of the detrimental event (death or immobility, presence of micronculei, malformations, etc.) compared to the control treatment: for judging relevance of an effect size or for the definition of a particular non-inferiority margin.[1]

For this reason, this paper is focused on the ratio of proportions (risk ratio). Compared to the plenty of publications considering the construction of confidence intervals for a single binomial proportion, as well as for differences, ratios or odds-ratios of binomial proportions, the construction confidence intervals for risk ratios of overdispersed binomial data has received only little attention.[2,3] The available methods are all asymptotic methods. Their construction and their evaluation is usually motivated by their application to epidemiological studies, where the absence or presence of a disease is counted for a given number of individuals in clusters. In this context, clustering

of individuals may arise from humans being clustered in families or locations or from repeated measurements within a given animal, or repeated animals within a given farm when veterinary epidemiology is concerned. In these settings, there is usually a relatively large number of clusters available (many families or locations, many farms or animals), such that simple asymptotic methods work well and the estimation of the overdispersion parameters is rather precise. This is not the case in bioassays that are the focus of this paper. In epidemiological settings it is also rather improbable to observe no single disease case across all clusters in one of the groups to be compared, or, conversely, to observe only disease cases across all clusters in one group. Such outcomes, however, plausibly occur in the untreated control groups of bioassays and may cause problems with Wald-type test and related confidence intervals.[4] Finally, in epidemiological studies, major interest is usually in estimation, such that the statistical evaluation of confidence interval methods for the risk ratio does plausibly focus on (two-sided) coverage probability and interval width. In the case of bioassays, estimation as well as related hypothesis tests are of interest. In testing, one-sided hypotheses are most relevant and may involve margins of non-inferiority.

In consequence, previous recommendations of confidence intervals methods for risk ratios of overdispersed data, motivated by epidemiological applications, cannot directly transferred to their application to bioassays. In this paper, previously proposed confidence intervals are thus investigated with special focus on small to very small number of clusters (i.e. replications) and the possibility that all observations in one group may show the same event. Due to the need of estimation as well as one-sided non-inferiority tests, two-sided as well as one-sided coverage probabilities are investigated for a wide range of parameter combinations. Further, a new method based on a straightforward combination of profile deviance intervals, e.g. [5,6] and the method of variance recovery[7] is proposed and is shown to clearly outperform existing approaches under the assumption of a common level of overdispersion.

Material and methods

Parameter and hypotheses of interest

Assume an experimental setup, where for each treatment i, there are J_i replicated experimental units (tanks, cages, animals, petri dishes, etc.) with index $j = 1, \ldots, J_i$. In each experimental unit there is a number of n_{ij} biological units under observation, and the number of events of interest, x_{ij}, is counted in unit ij. These counted events may be death (or survival) of animals, mobility (or immobility) of daphnids, presence or absence of micronuclei, etc. Consider the comparison of one dose group, $i = 1$, to the untreated control, $i = 0$. Denote the unknown probability of events in the two groups by i, and parameter of interest is the risk ratio $\rho = \frac{\pi_1}{\pi_0}$. Beside estimating ρ and displaying the uncertainty of this estimate in terms of a 95% confidence interval, decisions concerning one-sided hypotheses on tests on non-inferiority or superiority may be of interest. The particular choice of non-inferiority margins, ρ_0, may be fixed by convention, compare[1] suggesting $\rho_0 = 0.75$ or $\rho_0 = 0.8$ for certain applications. In other situations, it might be a matter of debate. Although general focus is in valid two-sided confidence intervals (i.e., with coverage probability close to the nominal level), it will be further investigated whether confidence intervals do also provide valid upper and lower confidence limits and can thus be used to perform (approximate) level α test for one-sided hypotheses.

Overdispersed binomial data

There are two well-known approaches to model overdispersion in binomial data.[8] The quasibinomial approach models overdispersion by assuming the variance-mean-dependency

$$V^{QB}\left(X_{ij}\right) = \phi n_{ij}\pi_{ij}\left(1-\pi_{ij}\right),$$

where ϕ is the overdispersion parameter that inflates the binomial variance term by a common fold, independent of the sample size n_{ij}. In this parameterization, the binomial assumption,

$$V^{B}\left(X_{ij}\right) = n_{ij}\pi_{ij}\left(1-\pi_{ij}\right), \text{ is met for } \phi = 1.$$

The beta-binomial distribution derives from a beta mixture of binomial distributions, i.e.,

$$\pi_{ij} \sim Beta(\mathrm{a}_i, \mathrm{b}_i), \text{ and } x_{ij} \sim Binomial(\mathrm{n}_{ij}, \pi_{ij}) \qquad (1)$$

where $E(\pi_{ij}) = \frac{a_i}{(a_i+b_i)}$, $E(\mathrm{x}_{ij}) = \frac{n_{ii}a_i}{(a_i+b_i)}$.[9] Denote the sum of the two parameters of the beta-parameters by $a_i^* = a_i + b_i$ The variance of beta-binomial counts, $V^{BB}(\mathrm{X}_{ij})$ is a function of n_{ij}, π_{ij}, a_i^*.[5,9] When a_i^* approaches ∞, the variance of the beta-binomial counts approaches that of binomial counts. Here, the overdispersion relative to the binomial variance, $v^{BB}(\mathrm{X}_{ij})$, is denoted by $\phi^{BB} = \frac{V^{BB}(\mathrm{X}_{ij})}{V^{B}(\mathrm{X}_{ij})}$ and is a function of n_{ij} and a_i^*. This

allows to choose a_i^* given n_{ij} such that the over dispersion ϕ^{BB} is a constant factor, namely $a_i^* = \frac{\phi_i^{BB} - n_{ij}}{(1-\phi_i^{BB})}$.

Main interest here is in the performance of confidence interval methods in highly controlled laboratory settings. Under such conditions it can be assumed that n_{ij} is equal for all experimental units, ij. Under such conditions, the quasibinomial assumption on the variance mean dependency coincides with the variance-mean-dependency under the beta-binomial distribution.[5,9] Hence, methods are considered that are explicitly constructed for the beta-binomial distribution, as well as methods that account for overdispersion under the quasibinomial assumption.

Confidence interval methods

Fieller-type intervals applied on the observed proportions (FN): Naively, the t-test for ratios, with a common variance estimator and the assumption of normal distributed residuals, may be used to test the above hypotheses, treating the observed proportions as the variable of interest, $y_{ij} = \hat{\pi}_{ij}$. The corresponding Fieller-type confidence interval can be obtained by analytically inverting the t-test statistic for ratios.[10] The method assumes normal distribution and variance homogeneity for the observed proportions, which is clearly not the case in this application. This interval is referred to as FN.

Binomial MOVER-R-Wilson method after summing-up observed counts across replications (BM): It may be tempting to sum up the counts over experimental units within each treatment group, $x_{i.} = \sum_{j=1}^{J_i} x_{ij}$, $n_{i.} = \sum_{j=1}^{J_i} n_{ij}$ and apply a confidence interval for risk ratios under the assumption of binomial distribution, i.e., ignoring possible extra-binomial variation. As a place holder for the many available options, here the MOVER-R method proposed by[7] is used; this is computationally simple and was among the best methods in a

Citation: Schaarschmidt F. Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion. *Biom Biostat Int J.* 2015;2(7):200–208. DOI: 10.15406/bbij.2015.02.00050

recent comparative study under the binomial assumption by.[11] It is referred to as BM.

The two above methods are merely included here to illustrate the effects of either ignoring the mean-variance-relation and skewness implied by binomial distribution (FN) or the effect of applying binomial methods (BM) when data are indeed overdispersed binomial.

Asymptotic method on the log-scale, accounting for overdispersion via residual variation (LOD): Among other methods,[3] investigate and recommend a method based on the delta method applied for the log risk ratios (called MR1 therein). It can be computed from:

$$n_{i.} = \sum_{j=1}^{J_i} n_{ij}, x_{i.} = \sum_{j=1}^{J_i} x_{ij}, \hat{\pi}_i = \frac{x_{i.}+0.5}{n_{i.}+1}, \hat{\rho} = \frac{\hat{\pi}_1}{\hat{\pi}_0}$$

$$\hat{v}_i = \left(\frac{J_i}{J_i-1}\right)\sum_{j=1}^{J_i}\frac{(\mathrm{x}_{ij}-\hat{\pi}_i \mathrm{n}_{ij})^2}{n_{i.}^2} \qquad (2)$$

Where the variance is estimated from the residuals on the scale of the original observations. The interval is then given by:

$$\hat{\rho}\exp\left(\pm \mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\sum_{i=0}^{I}\frac{\hat{v}_1}{\hat{\pi}_i^2}}\right) \qquad (3)$$

By plugging-in the observed residual variance per treatment group *i*, this method does not assume a particular mean-variance relation and accounts for overdispersion in a more general way. However, if the number of replications per treatment, *Ji*, is small, these variance estimates might be unstable. Zaihra and Paul[3] additionally consider a closely related method with a sandwich-type variance estimator, which performs worse in their simulation study, and is thus ignored here.

Fieller-Bailey-type interval accounting for overdispersion via residual variance (FOD): Using the estimators above,[3] follow the approach of[10] and[12] that accounts for the skewed distribution of the original Fieller statistic

$$Z = \frac{\hat{\pi}_1 - p\hat{\pi}_0}{\sqrt{\hat{v}_1+\rho^2\hat{v}_0}}$$

By considering the solutions of a cubic equation, with

$$A = \hat{\pi}_0^{2/3} - z_{\alpha/2}^2\frac{\hat{v}_0}{9\hat{\pi}_0^{4/3}}, B = (\hat{\pi}_1\hat{\pi}_0)^{1/3}, C = \hat{\pi}_1^{2/3} - z_{\alpha/2}^2\frac{\hat{v}_1}{9\hat{\pi}_1^{4/3}}$$

The interval (referred to as MR4 by [3]) can then be computed by

$$\left[\max\left(\left(\frac{B-\sqrt{B^2-AC}}{A}\right),0\right)^3;\left(\frac{B-\sqrt{B^2-AC}}{A}\right)^3\right]$$

If $A > 0$ and $B^2 - AC > 0$. If these two restrictions are not met, the interval has unbounded or disjoint solutions which do not provide meaningful interpretations of ρ. In the simulation study below, the interval $[0,\infty]$ is returned in such cases. Zaihra and Paul[3] again consider a closely related method that uses sandwich estimator for the variance of $\hat{\pi}_i$ instead. It is not considered here.

Delta method on the log-scale under the beta-binomial assumption (LBB): Lui et al.,[2] propose methods that are constructed under the assumption of the beta-binomial distribution. Theoretically, the variance of $\hat{\pi}_i$ under this assumption is $\hat{\pi}_i\left(1-\pi_i\right)\phi^{BB}\left(\mathbf{n}_i,c_i\right)/n_i$ where $\phi^{BB}\left(\mathbf{n}_i,c_i\right)$ is the beta-binomial overdispersion factor, expressed as a function of the number under risk in each replication *ij* of treatment *i*, $\mathbf{n}_i = (n_{i1}, n_{i2}, \ldots n_{iJi})$ and of c_i, the intraclass correlation coefficient. Under beta-binomial sampling, the intraclass correlation depends on a_i^* via $c_i = 1/(a_1^*+1)$.[2]

Lui et al.,[2] estimate the intraclass correlation using

$$\hat{c}_i = \frac{BMS_i - WMS_i}{BMS_i + \left(n_i^*-1\right)WMS_i}, \text{ with } n_i^* = \frac{n_{i.}^2 - \Sigma_{j=1}^{J_i} n_{ij=1}^2}{(J_i-1)n_{i.}}$$

Based on the between and within mean squared error of the observations,

$$BMS_i = \frac{\sum_{j=1}^{J_i}\frac{x_{ij}^2}{n_{ij}} - \frac{x_{i.}^2}{n_{i.}}}{J_i-1} \text{ and } BMS_i = \frac{\sum_{j=1}^{J_i}\frac{x_{ij}^2}{n_{ij}} - \frac{x_{i.}^2}{n_{i.}}}{J_i-1}$$

This leads to estimators for the beta-binomial overdispersion factor, and the related variance of the proportion estimator under the beta-binomial assumption, for each treatment group *i* separately:[2]

$$\hat{\phi}^{BB}\left(\mathbf{n}_i,\hat{c}_i\right) = \sum_{j=1}^{J_i} n_{ij}\left(1+\left(n_{ij}-1\right)\hat{c}_i\right)/n_{i.} \qquad (5)$$

$$\hat{v}_i^{BB} = \frac{\left(1-\hat{\pi}_i\right)\hat{\phi}^{BB}\left(\mathbf{n}_i\hat{c}_i\right)}{n_i\hat{\pi}_i} \qquad (6)$$

The asymptotic interval relying on the delta method applied for the log risk ratio[2] can be constructed by:

$$\hat{\rho}\exp\left(\pm z_{1-\alpha/2}\sqrt{\sum_{i=0}^{1}\hat{v}_i^{BB}}\right) \qquad (7)$$

Fieller-Bailey-type interval under beta-binomial assumption: FBB: Lui et al.,[2] consider a Fieller-type interval and its modification according to[12] under the beta-binomial distribution:

$$A = \hat{\pi}_0^{2/3} - z_{\alpha/2}^2\left(1-\hat{\pi}_0\right)\phi^{BB}\left(\mathbf{n}_0,\hat{c}_0\right)/\left(9n_{0.}\hat{\pi}_0^{1/3}\right), B = \left(\hat{\pi}_1\hat{\pi}_0\right)^{1/3} \text{ and}$$

$$C = \hat{\pi}_1^{2/3} - z_{\alpha/2}^2\left(1-\hat{\pi}_1\right)\phi^{BB}\left(\mathbf{n}_1,\hat{c}_1\right)/\left(9n_{1.}\hat{\pi}_1^{1/3}\right)$$

As in Eq. (4), a meaningful interval can be calculated if $A > 0$ and $B2 - AC > 0$:

$$\left[\max\left(\left(\frac{B-\sqrt{B^2-AC}}{A}\right)^3,0\right);\left(\frac{B-\sqrt{B^2-AC}}{A}\right)^3\right] \qquad (8)$$

In the simulation study below, the interval $[0,\infty]$ is returned if $A > 0$ and $B^2 - AC > 0$. Following,[2] in case of the extreme events $\hat{\pi}_i = 0$ or $\hat{\pi}_i = 1$, is replaced by $\tilde{\pi}_i = \frac{x_{i.}+0.5}{n_i+1}$ in the LBB and FBB method and their subsequent modifications.

Modifications of LOD, FOD, LBB, FBB by pooling and restricting the variance estimates: Lui et al.,[2] state, based on theoretical considerations and in the context of estimation problems that the intraclass correlation c_i and the overdispersion $\phi^{BB}\left(\mathbf{n}_i,c_i\right)$ cannot fall

below 0 and 1, respectively. However, their estimates may fall below the boundaries imposed by the binomial assumption. For example, the event $x_{i1} = x_{i2} = ... = x_{iJi}$ may lead to the unreasonable estimates $\hat{c}_i < 0$ and $\hat{\phi}_{BB}(\mathbf{n}_i, \hat{c}_i) = 0$. Moreover, interest here is in small sample laboratory experiments and estimating c_i and $\phi^{BB}(n_i, c_i)$ separately for each treatment from very few replications J_i may result in over fitting. Then, the assumption of a common beta-binomial overdispersion parameter $\phi^{BB}(\mathbf{n}_i, \hat{c}_i)$ may lead to a more stable estimation with small sample sizes. Therefore, the methods LBB and FBB as well as LOD an FOD are simulated with the following additional restrictions and pooling of variance estimates:

I LBB1, FBB1 refer to methods LBB, FBB with the beta-binomial overdispersion factor restricted to be at least 1, i.e., using $\max(1, \phi^{BB}(\mathbf{n}_i, \hat{c}_i))$ instead of $\phi^{BB}(\mathbf{n}_i, \hat{c}_i)$ in the equations (5,6) ff.

II LBBp, FBBp refer to methods LBB, FBB with the intraclass correlation estimator using pooled observations across the groups, $\hat{C}_{\cdot} = \frac{\sum_{i=0}^{2} \hat{c}_i n_i}{\sum_{i=0}^{2} n_i}$ and $\phi^{BB}(\mathbf{n}_i, \hat{c}_{\cdot})$, instead of $\phi^{BB}(\mathbf{n}_i, \hat{c}_{\cdot})$ in equations (5, 6)ff.

III LBB1p, FBB1p combine the two approaches by using the group wise variance estimators with the pooled intraclass correlation estimator and restriction of the over dispersion parameter to be ≥ 1, that is, using $\max(1, \phi^{BB}(\mathbf{n}_i, \hat{c}_i))$ instead of $\phi^{BB}(\mathbf{n}_i, \hat{c}_i)$ in equations (5,6)ff. This procedure has been already suggested in the example evaluation of.[2]

IV LOD1, FOD1: refer to methods LOD, FOD, but the group wise variance estimators are restricted to be greater than or equal to the binomial variance estimate: using $\hat{v}_i = \max\left(\frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_{i\cdot}}, \left(\frac{J_i}{J_i - 1}\right)\sum_{j=1}^{J_i}\frac{(x_{ij} - \hat{\pi}_{i\cdot} n_{ij})^2}{n_{i\cdot}^2}\right)$ in Eq.(2)

MOVER-R for quasibinomial profile deviance intervals (QBM): An alternative option to obtain intervals for the risk ratio would be to fit a generalized linear model (GLM) under the quasibinomial assumption $V_{QB}(X_{ij}) = \phi n_{ij} \pi_{ij}(1-\pi_{ij})$, using a logit-link, $\eta_{ij} = \log(\pi_{ij}), \eta_{ij} = \beta_i$. Then the risk ratio can be estimated via $\exp(\beta_1 - \beta_0)$ and intervals for the difference $(\beta_1 - \beta_0)$ can be computed by the signed root profile deviance method [6,5]. However, with the current user-level implementations in R, fitting this model (glm, stats) and obtaining profile-deviance intervals (profile, confint, package MASS) suffers from numerical difficulties if at least one of the groups shows estimated success probabilities close or equal to 1 or equal to 0.

As a numerically stable work-around, the QBM method is proposed: A GLM with the quasibinomial assumption and logit link is fitted $\eta_{ij} = \log(\pi_{ij} / (1 - \pi_{ij})) = \beta_i$. If the estimated dispersion parameter in the model fit falls below $1, \hat{\phi}^{(QB)} < 1$, a binomial model is used instead (i.e., assuming $\phi = 1$). For the β_i, (*1-α*)--signed root profile deviance intervals can be computed, with limits denoted $[l_{\hat{\beta}0}, u_{\hat{\beta}0}], [l_{\hat{\beta}1}, u_{\hat{\beta}1}]$ and estimates denoted $\hat{\beta}_i$. In R, these computations can be done in several packages, e.g. package MASS,[5] or the add-on package mcprofile [13]. Again, in extreme cases, the automatic search of values for the grid of parameter values for the deviance profile may fail in both packages. The signed root deviance is then computed over a pre-specified grid of parameter values,

$$\beta_i^* = (-10, -9.5, ..., -5, -4.8, -4.6, ..., 4.8, 5, 5.5, ..., 10)$$

with elements β_{ik}^*, k = 1,…, K. For each i and each k,

$$t_{ik} = sign(\beta_{ik} - \hat{\beta}_i)\sqrt{\frac{d(\beta_{ik}, \beta_{i'}) - d(\hat{\beta}_i, \hat{\beta}_{i'})}{\hat{\phi}^{QB}}}$$

is computed, where $\beta_{ik} - \beta_i$ retains the sign of the difference $\beta_{ik}^* - \hat{\beta}_i$, $d(\beta_{ik}^*, \hat{\beta}_{i'})$ is the deviance when replacing $\hat{\beta}_i$ by β_{ik}^* while leaving all other parameters at their ML estimates, $\beta_{i'}$, $d(\hat{\beta}_i, \hat{\beta}_{i'})$ is the deviance at the ML estimates, and $\hat{\phi}$ is the dispersion estimate with all parameters at their ML estimates. For each parameter *i*, a cubic spline is fitted for t_{ik} depending on β_{ik}^* and the cut points of the spline with quantiles of the t-distribution, $t_{\alpha/2, df=dfr}, t_{1-\alpha/2, df=dfr}, dfr = \sum_{i=1}^{I}(J_i - 1)$ is determined by linear interpolation between fitted values. When the binomial model is used, t_{ik} is replaced by

$$z_{ik} = sign\left((\beta_{ik}^* - \hat{\beta}_i)\right)\sqrt{d(\hat{\beta}_{ik}, \hat{\beta}_{i'}) - d(\hat{\beta}_i, \hat{\beta}_{i'})}$$

And the quantiles of the standard normal distribution $z_{\alpha/2}, z_{1-\alpha/2}$, are used instead.

The interval bounds and ML estimates are transformed to the proportion scale using the inverse link, $[l_i, u_i] = \left[\frac{\exp(l_{\hat{\beta}i})}{1+\exp(l_{\hat{\beta}i})}, \frac{\exp(u_{\hat{\beta}i})}{1+\exp(u_{\hat{\beta}i})}\right], \hat{\pi}_i = \exp(\hat{\beta}_i) / (1+\exp(\hat{\beta}_i))$

These estimators and confidence limits are then used to compute intervals for ρ by the MOVER-R method.[7] Eq. (9) of[7] is recalled in the following as:

$$\left[\frac{\hat{\pi}_1\hat{\pi}_0 - \sqrt{(\hat{\pi}_1\hat{\pi}_0)^2 - l_1u_0(2\hat{\pi}_1 - l_1)(2\hat{\pi}_0 - u_0)}}{u_0(2\hat{\pi}_0 - u_0)}; \frac{\hat{\pi}_1\hat{\pi}_0 + \sqrt{(\hat{\pi}_1\hat{\pi}_0)^2 - u_1l_0(2\hat{\pi}_1 - u_1)(2\hat{\pi}_0 - l_0)}}{l_0(2\hat{\pi}_0 - l_0)}\right]$$

Like the FOD and the FBB method, this method may yield (partially) unbounded intervals, particularly when $\hat{\pi} = 0$, the upper limit for the risk ratio is naturally ∞.

Simulation study

The beta-binomial distribution is chosen to simulate overdispersed data such that the resulting data are in line with the quasibinomial assumption, i.e., the assumption of the QBM method is met.

In the simulation, the number of experimental units per treatment group, J_i is chosen balanced, *(J_0, J_1) = (3, 3), (5, 5)* and *(10, 10)*. The number of biological units under risk in each unit, n_{ij}, is chosen balanced n_{ij} = *10* or *20* for all *i, j*. Overdispersion is set at levels $\phi^{BB} = 1.25$ or $\phi^{BB} = 2$, that is, for given n_{ij}, a_i^* is chosen according to Section 4.2 to achieve the specified overdispersion: for each set of π_0, π_1, $a_i = a_i^*\pi_i$ and $b_i = a_i^*(1-\pi_i)$ in the beta distribution. The distribution of the counts x_{ij}, as well as that of the estimators of the proportions are skewed to different extent, depending on π_i,

Citation: Schaarschmidt F. Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion. *Biom Biostat Int J.* 2015;2(7):200–208. DOI: 10.15406/bbij.2015.02.00050

especially if π_i is close to the border of the parameter space. Thus, also the distribution of the estimator of ρ or log (ρ) can be skewed if π_1 and π_0 differ. Thus, the performance of large sample methods may severely depend on particular choices of π_0, π_1. To investigate these potential dependencies, the simulations have been run for a grid of all combinations of $\pi_0 = (0.02, 0.04, ..., 0.96, 0.98)$ and $\pi_1 = (0.02, 0.04, ..., 0.96, 0.98)$ that imply odds-ratio between 0.1 and 10.

Simulations have been performed with 10000 runs for all methods except the QBM method. Due to high computation times, only 5000 simulation runs are used for QBM, such that the standard error of the estimated coverage probabilities for QBM is by factor $\sqrt{2}$ higher than for the remaining methods. The simulation study has been performed in R 3.1.2,[14] the implementation of all confidence interval methods is available in the R-package pairwiseCI, version > 0.1-25 [15] FBB, LBB and related methods are implemented in the function Betabin.ratio, FOD, LOD and related methods are implemented in the function ODbin.ratio, and the QBM method is implemented in function Quasibin.ratio.

Results and discussion

Coverage probability of two-sided 95% confidence intervals

Figure 1 shows the simulated coverage probabilities of all 13 methods under comparison: the NF method can either show too low or too high coverage probability, irrespective of increasing sample size, because the relation of mean and variance in the binomial data is ignored. However, its average coverage probability is closer to the nominal level than some of those methods which explicitly account for over dispersed binomial data (LBB, FBB, LBBp, FBBp, LOD and FOD). Simply ignoring the possibility of overdispersion and assuming the binomial distribution results in too low coverage probability even for moderate (1.25-fold) overdispersion: BM method has too low coverage probabilities in nearly all settings. For very small numbers of replications (J_i=3, 5), the far too low coverage probability of the asymptotic methods for overdispersed binomial data (LBB, FBB, LOD, FBB) can be improved slightly by using a pooled variance estimator (LBBp, FBBp) and can be largely improved by setting a lower limit to their variance estimators: If we replace variance estimates suggesting under dispersion by the corresponding estimates under the binomial assumptions, the coverage probabilities of these methods are much closer to the nominal level. The QBM method is always very close to nominal coverage probability but can have slightly too high average coverage probability when overdispersion is moderate and the number of replications is small.

Figures 2 and Figure 3 shows a detailed view on the coverage probabilities in dependence on the true underlying proportions, π_0 and π_1. This detailed view is restricted to those 6 methods which have an average coverage probability close to 0.95. Figure 2 shows the more difficult case with substantial overdispersion ($\phi = 2$), and only n_{ij} = 10 biological units in each replication. The QBM methods is close to the nominal confidence levels for a wide range of proportions, but can have too low coverage probabilities if at least one of the proportions is close to 0 or 1. The LBB1p and LOD1 methods need more replications to have coverage probabilities close to 0.95 for a similar range of π_i, and still are slightly liberal for almost all π_i. LBB1p and LOD1 booth have too high coverage probability for π_i close to 0 and too low coverage probabilities when π_i close to 1. That is test decisions based on these two methods may be conservative if hypotheses are formulated in terms of mortalities which should be low in the control group, but will be liberal when a similar hypothesis is formulated in terms of the proportion of survivors. In this simulation setting, the Fieller-Bailey-type intervals FBB1p and FOD1 have lower coverage probabilities for almost all parameter combinations considered as compared to the LBB1p and LOD1 method, respectively.

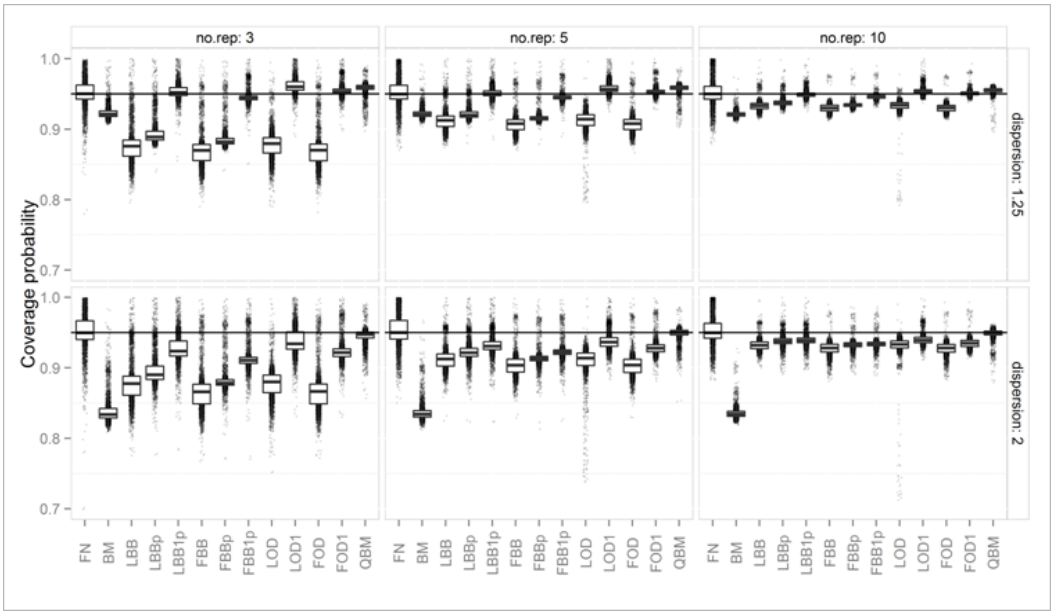

Figure 1 Boxplots of simulated coverage probabilities of nominal two-sided 95% intervals, for different numbers of replications (no.rep) and two different levels of overdispersion.

Citation: Schaarschmidt F. Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion. *Biom Biostat Int J.* 2015;2(7):200–208. DOI: 10.15406/bbij.2015.02.00050

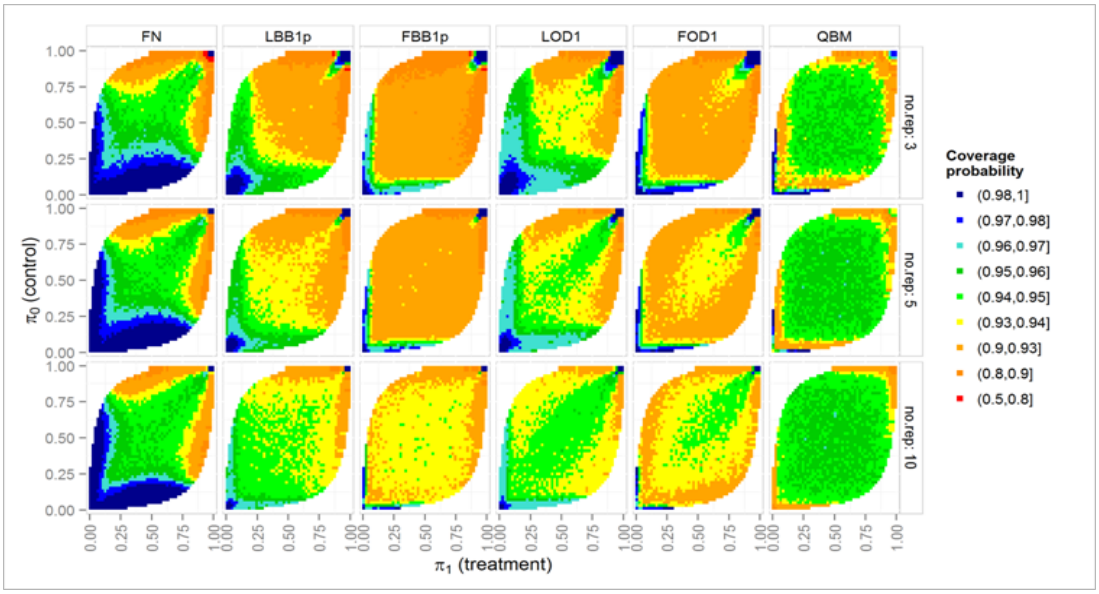

Figure 2 Simulated coverage probabilities (color scale) of nominal two-sided 95% intervals over a grid of true proportions π_0, π_1, for two-fold overdispersion (ϕ = 2) and n_{ij} = *10* biological units in each experimental unit.

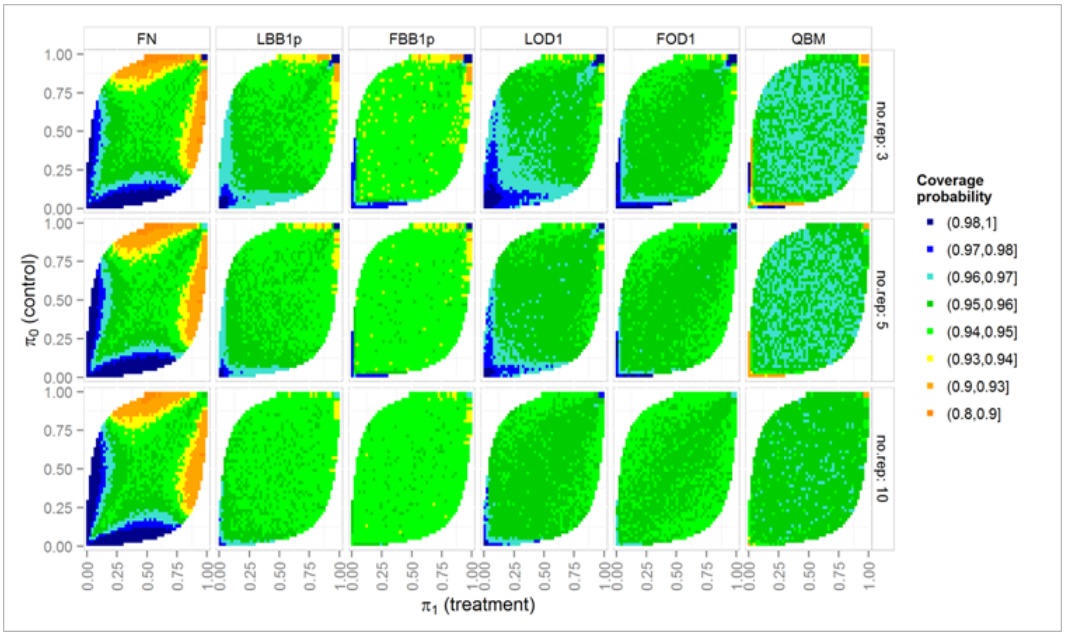

Figure 3 Simulated coverage probabilities (color scale) of nominal two-sided 95% intervals over a grid of true proportions π_0, π_1, for moderate overdispersion (ϕ = *1.25*) and n_{ij} = *20* biological units in each experimental unit.

Figure 3 shows results for the less problematic case of moderate over dispersion (ϕ = *1.25*) and n_{ij} = 20 biological units in each experimental unit. The coverage probability of QBM rarely falls below 0.94, but is slightly too large (between 0.96 and 0.97) if there are only 3 or 5 experimental units per treatment. The LBB1p and FOD1 method have again slightly too low coverage probability if any π_i is close to 1, and slightly too high coverage probability if any π_i is close to 0. The two Fieller-Bailey-type intervals have slightly lower coverage probabilities than their counterparts based on the log-delta-method.

Symmetry of non-coverage

Figure 4 shows the simulated proportion of cases, where the true parameter was excluded by the lower bound, relative to all cases where the parameter was excluded by the interval. For valid one-sided decisions, methods are preferable that exclude the true parameter equally likely by the lower and upper bound, i.e., with probability $\alpha/2$ for each limit. For brevity, only the challenging setting with n_{ij} = *10* and marked over dispersion (ϕ = *2*) is shown, while conclusions for the remaining simulation settings are similar: The Fieller-Bailey-type methods (FBB1p and FOD1) show a wider range of parameter

Citation: Schaarschmidt F. Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion. *Biom Biostat Int J.* 2015;2(7):200–208. DOI: 10.15406/bbij.2015.02.00050

settings where probability of parameter exclusion is equal between lower and upper bounds, as compared to the corresponding methods based on the delta methods on the log scale, LBB1p and LOD1. The QBM method shows asymmetric non-coverage for similar parameter settings as do the Fieller-Bailey-type intervals, i.e. when any of the true proportions is close to 0, but is clearly more symmetric than the FBB1p and FOD1 for a wide range of parameter settings where at least one π_i is close to 1.

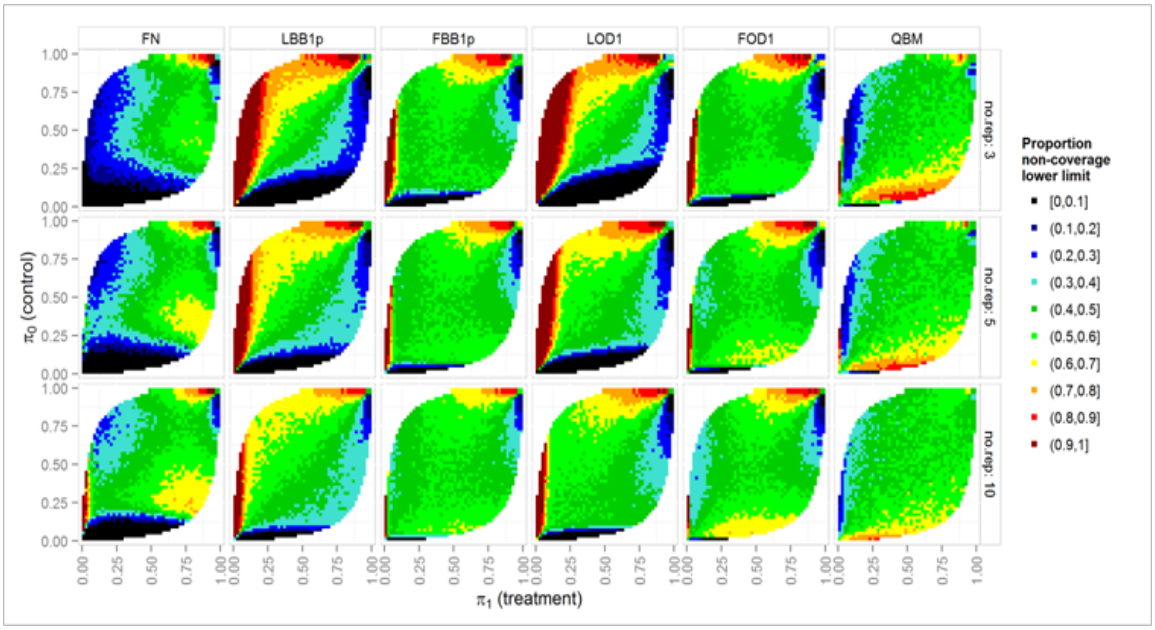

Figure 4 Asymmetry of non-coverage: color scale shows the proportion of cases where the true parameter is excluded by the lower limit, relative to all cas-eswhere the parameter is excluded by nominal two-sided 95% intervals over a grid of true prop proportions π_0, π_1, for clear overdispersion ($\phi = 2$) and $n_{ij} = 10$ biological units in each experimental unit.

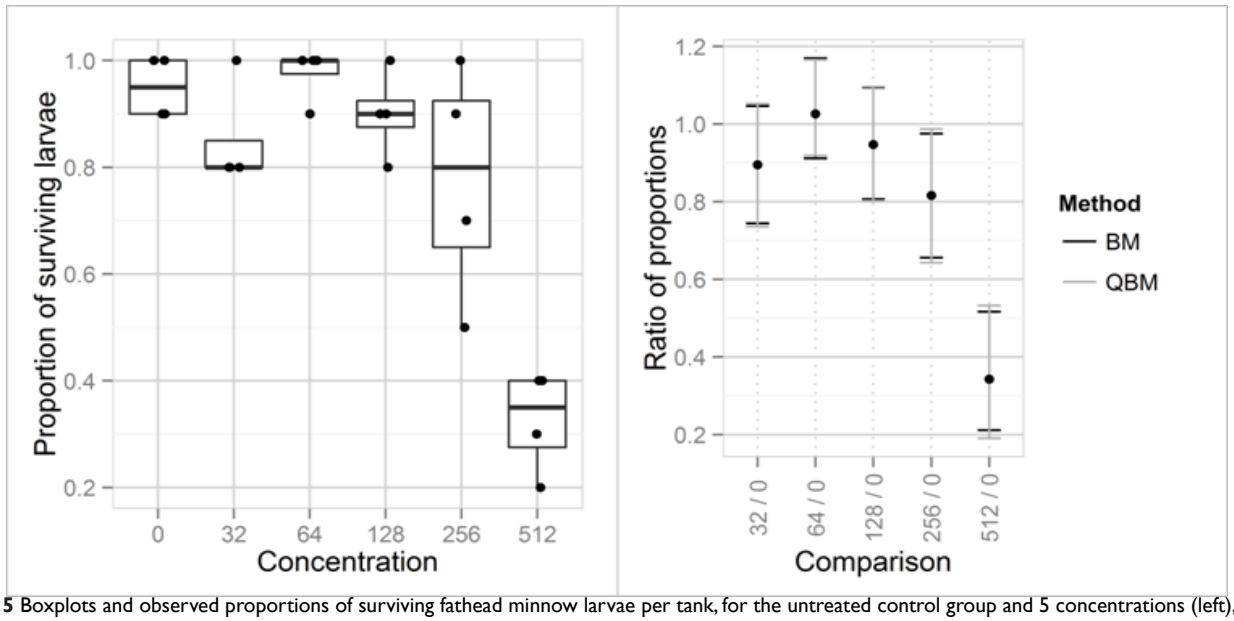

Figure 5 Boxplots and observed proportions of surviving fathead minnow larvae per tank, for the untreated control group and 5 concentrations (left), and 95% confidence intervals of the proportions of surviving larvae in the treatment groups relative to the control group (right).

Examples

Extreme cases: Table 1 shows four extreme cases: Cases 1 and 2 represent cases where proportions are close to 0 in the control treatment as could result from testing the ratio of mortality or immobility proportions. Cases 3 and 4 show data that could arise from test systems that assume proportions close to 1 in the control, for example, survival proportions as in the fathead minnow data below.

Citation: Schaarschmidt F. Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion. *Biom Biostat Int J.* 2015;2(7):200–208. DOI: 10.15406/bbij.2015.02.00050

Table 1 95% confidence intervals of selected methods for four extreme cases, assuming all n_{ij} = 10, and J_i = 4 for two treatment groups, i = 0, 1

	Case 1	**Case 2**	**Case 3**	**Case 4**
Method	x_{0j} = (0,0,0,2) x_{1j} = (1,2,5,6) $\hat{\rho}$ =0.35/0.05=7	x_{0j} = (0,0,1,1) x_{1j} = (1,2,2,4) $\hat{\rho}$ =0.225/0.05=4.5	x_{0j} = (8,10,10,10) x_{1j} = (6,7,9,9) $\hat{\rho}$=0.775/0.95=0.816	x_{0j}=(10,10,10,10) x_{1j} = (9,9,9,10) $\hat{\rho}$ =0.925/1=0.925
BM	(1.909, 26.2)	(1.15, 17.5)	(0.656, 0.975)	(0.801, 1.028)
FBB	(1.331, 151.9)	(1.45, 19.8)	(0.657, 1.005)	(0.879, 0.997)
FOD	(1.269, 182.6)	(1.41, 21.0)	(0.654, 1.009)	(0.885, 0.990)
FBB1p	(1.290, 190.0)	(1.24, 28.9)	(0.643, 1.022)	(0.851, 1.028)
FOD1	(1.269, 182.6)	(1.14, 36.8)	(0.654, 1.009)	(0.846, 1.033)
QBM	(0.902, 574.1)	(1.24, 27.6)	(0.558, 1.083)	(0.817, 0.998)

In cases 1 and 3, data lead to variance estimates exceeding that of the binomial distribution. Then, the BM method leads to shorter intervals than all other methods. In both cases, the QBM as at least slightly wider confidence intervals than the FBB1p and FOD1 method, which might correspond to the observation that these two have too low coverage probability for small samples. In cases 2 and 4, data show a variance below that of the binomial variance. Then the methods without restriction of variance estimates to that assumed by the binomial distribution (FOD and FBB) yield considerably shorter intervals, than methods which assume that under dispersion is implausible like BM, FBB1p, FOD1 and QBM.

Fathead minnow data: The toxicity of a compound to fathead minnow larvae was investigated using an untreated control group and 5 concentrations of a compound.[16] The experiment comprised 24 tanks, 4 tanks in each treatment group, each tank contained 10 larvae. The observed proportions of surviving larvae are shown in (Figure 5, left side). Analyzing the data in a generalized linear model with quasibinomial assumption, logit link shows that there are highly significant differences between the mean proportion of surviving larvae between the treatments ($p < 0.0001$; F-test in analysis of deviance). An estimated dispersion parameter of 1.082 suggests that the observations are at most slightly overdispersed, i.e., the data are at least roughly in line with the binomial assumption. The right side of Figure 5 shows two-sided 95% confidence intervals for ratios of the proportion of surviving larvae in the treatment groups relative to that in concentration 0. In this case, confidence limits based on the quasibinomial assumption (QBM) based on the full data including all 6 treatment groups and confidence intervals under the binomial assumption (MOVER-R method for Wilson-Score intervals, BM) do hardly differ.

Conclusion

Asymptotic methods based on the delta method applied on the log-scale or Fieller-Bailey type intervals have too low coverage probabilities when applied in small sample settings that are typical for many bioassays, i.e., they cover the true ratio of proportions less often than claimed by their nominal confidence level. Violation of the nominal level is most severe for small numbers of replications, low number of biological units in each replications and extreme proportions. Even for as much as 10 replications (i.e., clusters), coverage probabilities are considerably below the nominal level for wide ranges of proportions. When these intervals are then used for decisions in hypothesis tests for equivalence or non-inferiority, erroneous conclusions of equivalence or non-inferiority will occur more often than claimed by the nominal level α of such tests.

For the small number of replications and the small number of biological units per replication that are typical for some bioassays, restricting the variance estimates to that of the binomial variance (i.e., setting the dispersion parameter to 1 if under dispersion is estimated) leads to major improvements of the coverage probabilities. Further improvements can be achieved by combining the MOVER-R method with a profile deviance approach leads to intervals with better coverage probabilities for low sample sizes. However, this approach also shows too low coverage probabilities for small sample sizes and cases where one proportion is very close to 0 or 1. However, the simulation results shown here rely on the simplifying assumption, that there is a common overdispersion factor for the treatments in the experiments. Based on single data sets with few replications per treatment it will be hard to assess whether this assumption is appropriate, or whether different overdispersion factors per treatment group (as are used in the FBB, LBB method, for example) would be more appropriate. For given, highly standardized bioassay, however, available collections of historical data sets could be used to assess the plausibility of the different assumptions concerning distribution and mean-variance dependency and homogeneity or heterogeneity of overdispersion factors among treatment groups.

Bioassays usually involve several dosages. Depending on the global hypotheses to be tested, adjustments for multiple comparisons may be needed, see, e.g.[17] The methods for confidence intervals discussed here can be extended to construct approximate simultaneous confidence intervals, using approaches as described in.[17-20] However, such extensions require additional investigation as some approaches involve additional approximations. This is subject to further research.

Acknowledgement

I wanted to thank L. A. Hothorn as a person, thanks have no special relation to the paper cited by.[17]

Conflict of interest

None.

Citation: Schaarschmidt F. Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion. *Biom Biostat Int J.* 2015;2(7):200–208. DOI: 10.15406/bbij.2015.02.00050

References

1. Denton DL, Diamond J, Zheng L. Test of significant toxicity: a statistical application for assessing whether an effluent or site water is truly toxic. *Environ Toxicol Chem*. 2011;30(5):1117–1126.
2. Lui KL, Mayer JA, Eckhardt L. Confidence intervals for the risk ratio under cluster sampling based on the beta-binomial model. *Stat Med*. 2000;19(21):2933–2942.
3. Zaihra T, Paul S. Interval Estimation of Some Epidemiological Measures of Association. *Int J Biostat*. 2010;6(1):35.
4. Hauck WW, Donner A. Wald's Test as Applied to Hypotheses in Logit Analysis. *J Amer Statist Assoc*. 1977;72(360):851–853.
5. Venables WN, Ripley BD. Modern applied statistics with S. 4th edn. New York, USA. Springer-Verlag Inc; 2002:197–210.
6. Chen JS, Jennrich RI. The Signed Root Deviance Profile and Confidence Intervals in Maximum Likelihood Analysis. *J Amer Statist Assoc*. 1996;91(435):993–998.
7. Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation. *Stat Methods Med Res*. 2012;21(4):347–359.
8. McCullagh P, Nelder JA. Generalized linear models. 2nd edn. USA. Chapman & Hall/CRC;1989:124–135.
9. Johnson NL, Kotz S, Kemp AW. Univariate discrete distributions. 2nd edn. New York, USA. John Wiley & Sons; 1993:239–284.
10. Fieller EC. Some problems in interval estimation. J Roy Statist Soc Ser B-Stat *Methodol*. 1954;16(2):175–185.
11. Fagerland MW, Newcombe RG. Confidence intervals for odds ratio and relative risk based on the inverse hyperbolic sine transformation. *Stat Med*. 2013;32(16):2823–2836.
12. Bailey BJR. Confidence limits to the risk ratio. *Biometrics*. 1987;43(1):201–205.
13. Gerhard D. mcprofile: Multiple Contrast Profiles. R package version 0.1-5. 2013.
14. R Core Team. R: A language and environment for statistical computing. Vienna, Austria; 2014.
15. Schaarschmidt F, Gerhard D. pairwiseCI: Confidence Intervals for Two Sample Comparisons. R package version 0.1-25. 2015.
16. Anonymous. Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms. 5th edn. U.S. Washington, DC, USA. Environmental Protection Agency; 2002.
17. Hothorn LA, Hasler M. Proof of hazard and proof of safety in toxicological studies using simultaneous confidence intervals for differences and ratios to control. *J Biopharm Statist*. 2008;18(5): 915–933.
18. Hothorn T, Bretz F, Westfall P. Simultaneous Inference in General Parametric Models. *Biom J*. 2008;50(3): 346–363.
19. Lauzon C, Caffo B. Easy Multiplicity Control in Equivalence Testing Using Two Onesided Tests. *Am Stat*. 2009;63(2):147–154.
20. Dilba G, Bretz F, Guiard V. Simultaneous confidence sets and confidence intervals for multiple ratios. *J Statist Plann Inference*. 2006;136(8):2640–2658.

Citation: Schaarschmidt F. Confidence intervals for the risk ratio when analyzing bioassays in the presence of over dispersion. *Biom Biostat Int J*. 2015;2(7):200–208. DOI: 10.15406/bbij.2015.02.00050