Leibniz
Universität Hannover

Reports of the Institute of Biostatistics

No 10 / 2007

Leibniz University of Hannover

Natural Sciences Faculty

Titel: CI for Odds Ratios accounting for Extra Variation between Replicated Experiments

Authors: Daniel Gerhard

1 2 Sample Comparisons

1.1 Study Design

Let us assume a one-way layout in a completely randomized design with 2 factor levels. For each treatment $i = 1, 2$, we have n_i proportions as a response, with mean p_i. The proportions x_{ij} for each treatment with $j = 1, \ldots, n_i$ are assumed to be Beta distributed with density

$$Beta(x, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

for $a > 0, b > 0$ and $0 \leq x \leq 1$. The mean is calculated by $\frac{a}{a+b}$ with variance $\frac{ab}{(a+b)^2(a+b+1)}$. In the following a dispersion parameter is defined as $\phi = \frac{1}{a+b}$, where the extra variation between the proportions increases with increasing ϕ. Further, we assume that each x_{ij} is the proportion of binomial distributed successes $y \sim B(N, x)$ divided by the size N of each experiment. For the assumption of no occurrence of extra variation, x_{ij} is taken as a fixed value with $n_i = 1$.

1.2 Construction of Confidence Intervals

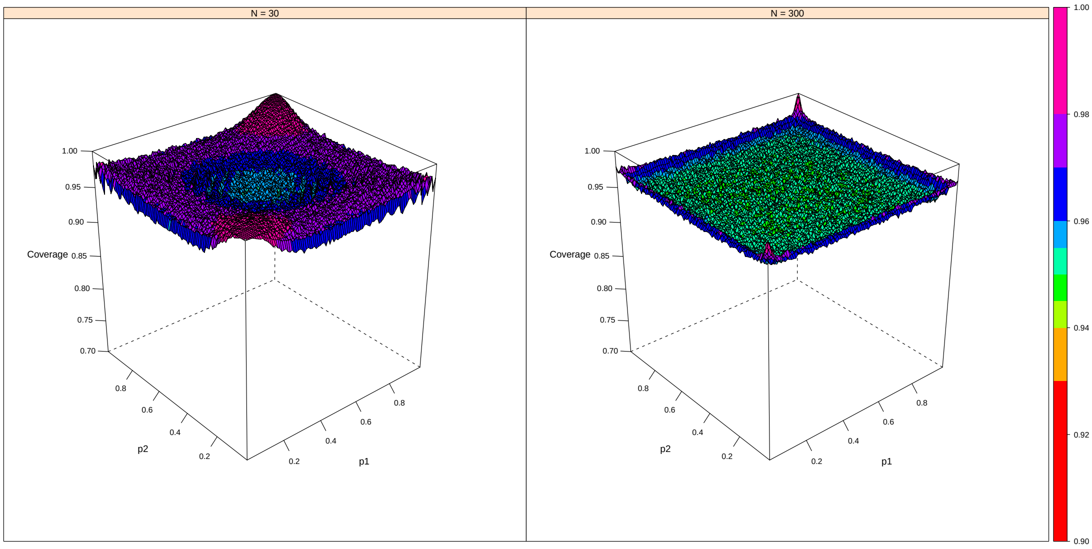
With the numbers of successes y and the number of failures $N - y$ in a $2 \times k$ table with $k = 2n_i$, a generalized linear model (GLM) is fitted with Binomial, Quasibinomial, Betabinomial, or Beta family on the logit link. Approximate confidence intervals for the odds ratio are calculated by

$$\frac{p_2(1-p_1)}{p_1(1-p_2)} \in \exp\left[\hat{p}_{2-1} \pm z_{0.975}\sqrt{\hat{\sigma}^2_{p_2-p_1}}\right]$$

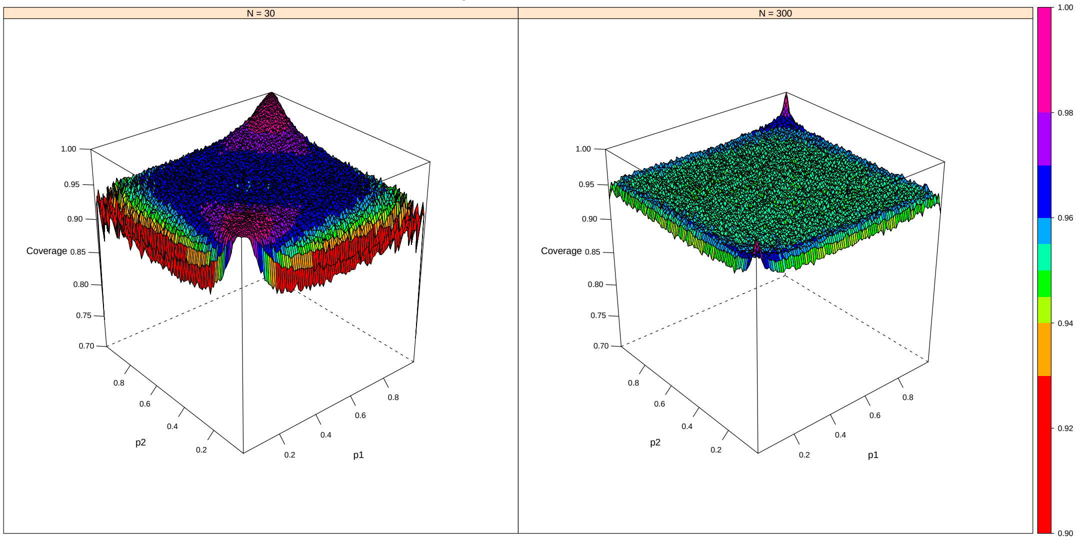
with $z_{0.975}$ being the $(1-\frac{\alpha}{2})$ quantile of the standard normal distribution.

1.3 Results

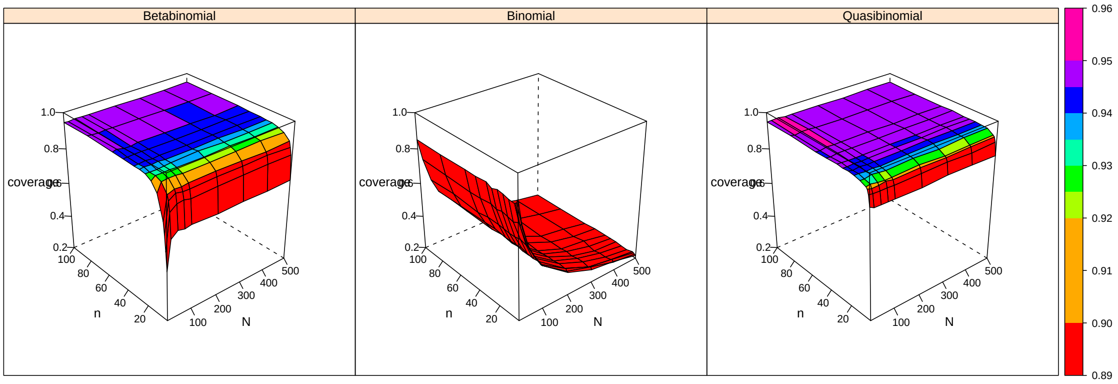
n = 1, N = 30,300, with no extra variation



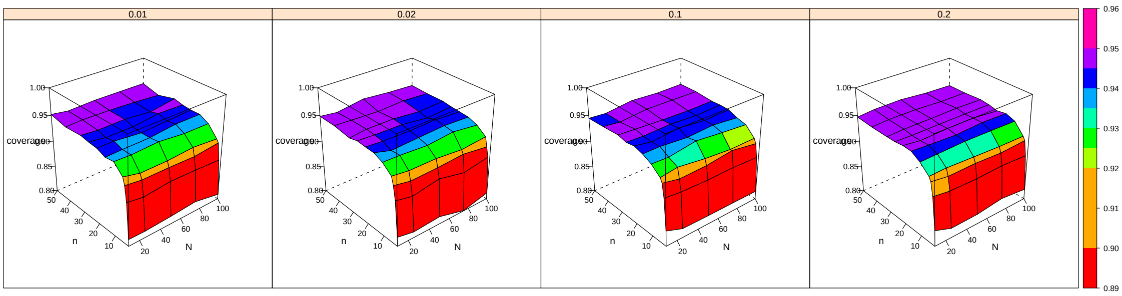
Adding 0.5 to each cell of the 2 x k table



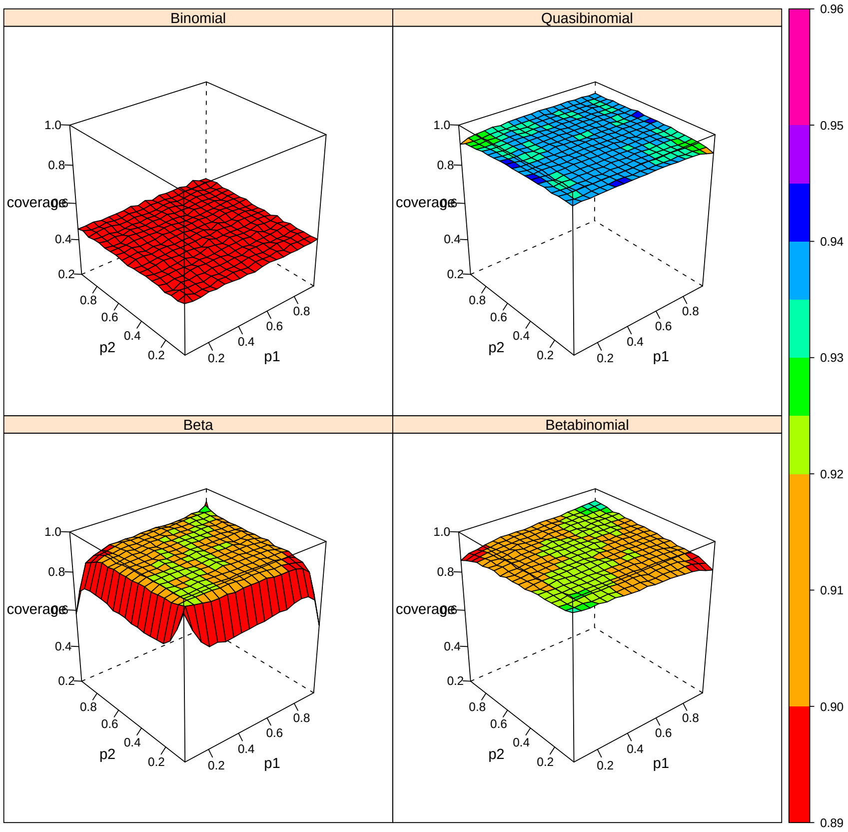
Varying n and N at $\phi = 0.1$, $p_1 = 0.5$, $p_2 = 0.5$



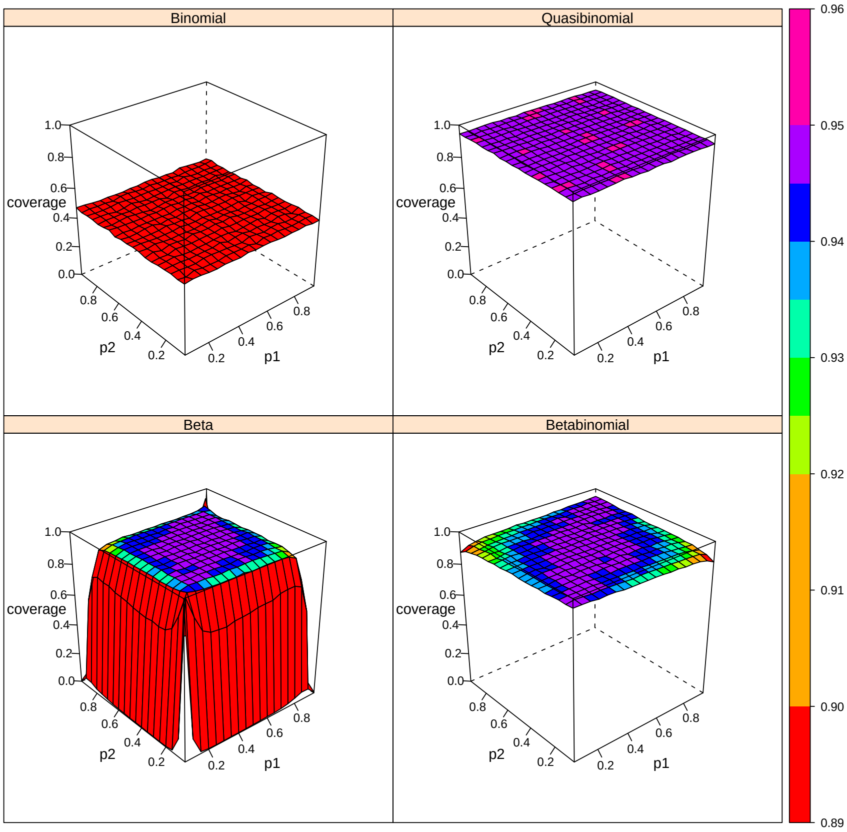
Varying n and N for quasibinomial models at different dispersion parameters ϕ, $p_1 = 0.5$, $p_2 = 0.5$



n = 10, N = 100, at $\phi = 0.1$



n = 100, N = 100, at $\phi = 0.1$



1.4 Discussion

Without overdispersion parameter estimates from a binomial model result in reasonable confidence intervals with quite good properties. For small sample sizes (N = 30) these confidence intervals get conservative, especially when both proportions are located near the lower or the upper limit of the parameter space. At large sample sizes (N=300) the intervals hold the nominal level for a wider range of parameter combinations. The intervals can be adjusted by adding 0.5 to each cell of the $2 \times k$ table to maintain also reasonable results at the occurrence of proportions of 0 and 1. These intervals have coverage probabilities near 0.95 for a larger number of parameter settings, but at the boundary of the parameter space, comparing a large proportion with a small one, the intervals are getting liberal.

At occurrence of overdispersion the variance will be underestimated by the simple binomial model; therefore the corresponding confidence intervals will be more liberal with increasing extra variation. By assuming directly a betabinomial likelihood or assuming only a variance-mean relationship in a quasibinomial model the variation between the single proportions x_{ij} can be represented. Here, the confidence intervals for parameters of the quasibinomial models show nice properties, even at relative small sample sizes and small proportions. The Betabinomial model CI get a bit liberal at large distances between the proportions.

2 Multiple Comparisons

2.1 Study Design

Let us now assume a $2 \times k$ table, where $k = gn_i$ with g equals the number of groups to be compared. Then, after the fit of a logistic model, simultaneous confidence intervals for m contrasts $\boldsymbol{C}_{m \times g}$ of proportions on the logit link are constructed by

$$\exp\left[\boldsymbol{C}\hat{\boldsymbol{p}} \pm z^{2sided}_{0.95,\hat{\boldsymbol{R}}}\,\hat{\boldsymbol{\sigma}}_{\boldsymbol{C}\boldsymbol{p}}\right].$$

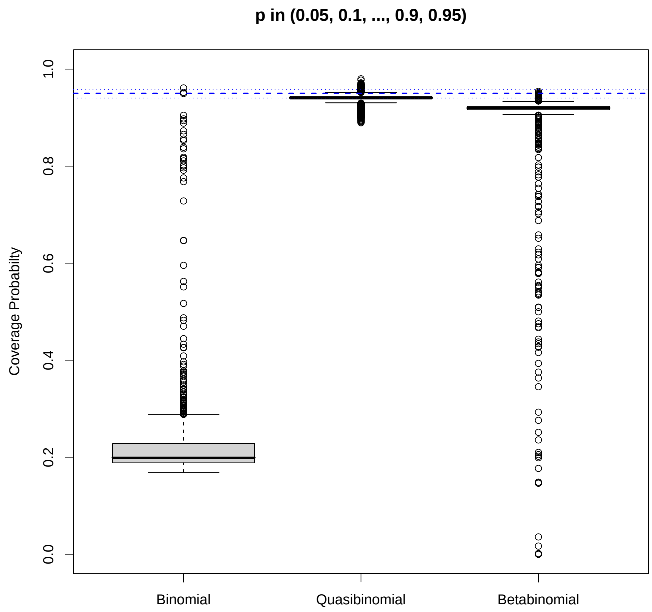
The critical value $z^{2sided}_{0.95,\hat{\boldsymbol{R}}}$ is a two-sided quantile of the multivariate normal distribution, assuming a correlation of $\hat{\boldsymbol{R}}$, which is obtained by standardizing the variance covariance matrix of the linear combinations of $\hat{\boldsymbol{p}}$. $\hat{\boldsymbol{\sigma}}_{\boldsymbol{C}\boldsymbol{p}}$ is the standard error of the linear combinations of $\hat{\boldsymbol{p}}$.

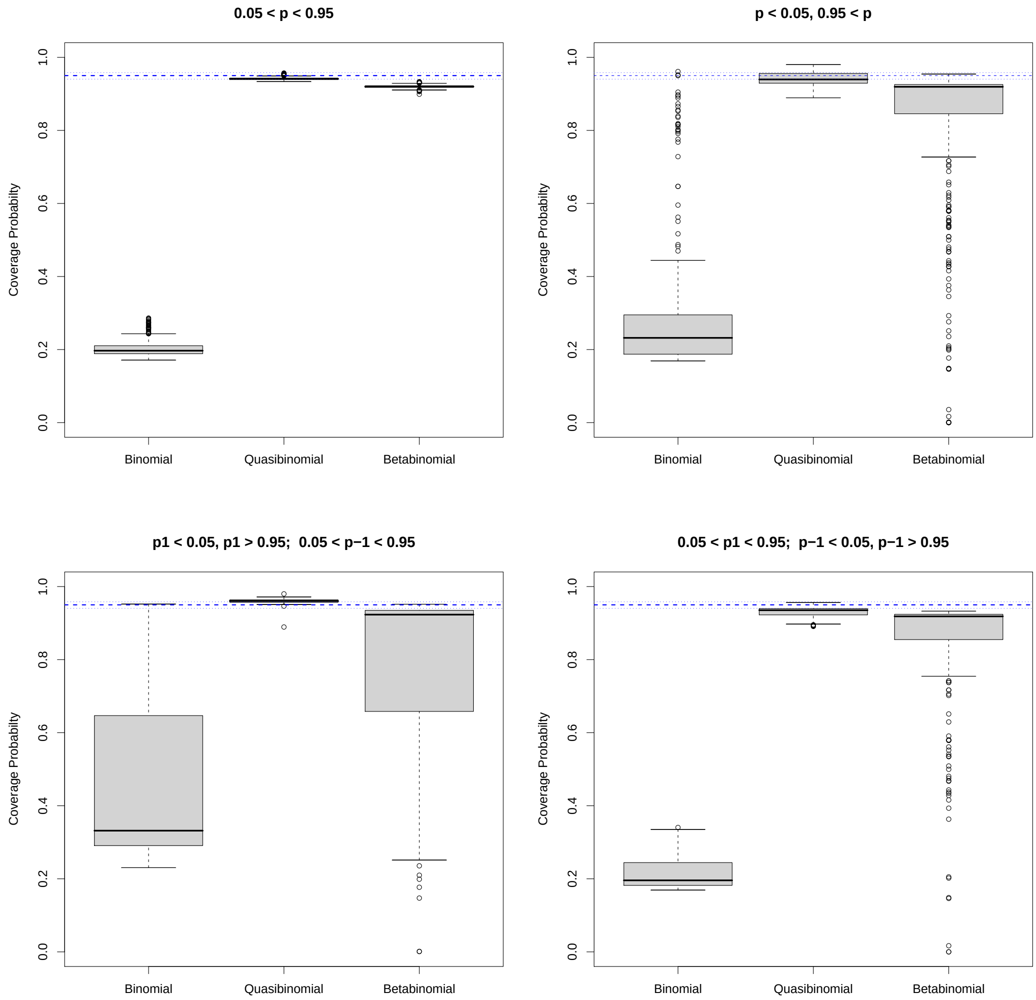
2.2 Parameter Settings

As an exemplary setting $g = 4$ groups are chosen. 1,000 combinations of 4 proportions are generated by sampling from a uniform distribution. For each set of them 10,000 simulation runs were performed. Many-to-one comparisons were performed by definition of $\boldsymbol{C}_{m \times g}$ as Dunnett contrast.

2.3 Results

n = 10, N = 100, at $\phi = 0.1$





2.4 Discussion

The quasibinomial model shows again good properties. The confidence intervals are only a bit liberal, looking at every parameter setting under investigation. For the betabinomial model there are some more parameter combinations, where the confidence intervals get far liberal. If the control proportion is located at the border of the parameter space and the others being rather in the center, the quasibinomial CIs are conservative; the other way round, with control proportions around 0.5 and the remaining ones at the border of the parameter space, the intervals get liberal.