Title

A comparison of two varying coefficient meta-analysis methods for an average risk difference

Journal

Journal of Statistical Computation and Simulation, 85(12)

ISSN

0094-9655

Authors

Price, Robert M
Bonett, Douglas G

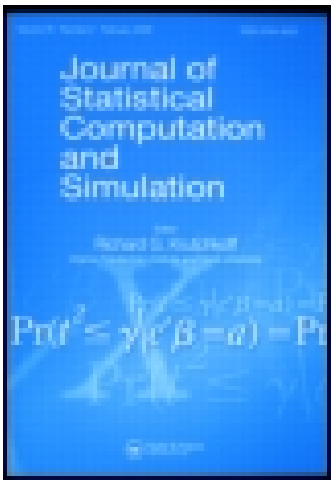
Publication Date

2015-08-13

DOI

10.1080/00949655.2014.929132

Peer reviewed



Journal of Statistical Computation and Simulation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

A comparison of two varying coefficient meta-analysis methods for an average risk difference

Robert M. Price Jr[a] & Douglas G. Bonett[b]

[a] Department of Mathematics & Statistics, East Tennessee State University, Johnson City, TN 37604-0663, USA

[b] Department of Psychology, University of California, Santa Cruz, CA 95064, USA

Published online: 18 Jun 2014.

To cite this article: Robert M. Price Jr & Douglas G. Bonett (2014): A comparison of two varying coefficient meta-analysis methods for an average risk difference, Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2014.929132

To link to this article: http://dx.doi.org/10.1080/00949655.2014.929132

Journal of Statistical Computation and Simulation, 2014
http://dx.doi.org/10.1080/00949655.2014.929132

Taylor & Francis
Taylor & Francis Group

A comparison of two varying coefficient meta-analysis methods for an average risk difference

Robert M. Price, Jr[a]* and Douglas G. Bonett[b]

[a]*Department of Mathematics & Statistics, East Tennessee State University, Johnson City, TN 37604-0663, USA;* [b]*Department of Psychology, University of California, Santa Cruz, CA 95064, USA*

(*Received 10 January 2014; accepted 26 May 2014*)

Two interval estimation methods for a general linear function of binomial proportions have been proposed. One method [Zou GY, Huang W, Zhang X. A note on confidence interval estimation for a linear function of binomial proportions. Comput Statist Data Anal. 2009;53:1080–1085] combines Wilson interval estimates of individual proportions, and the other method [Price RM, Bonett DG. An improved confidence interval for a linear function of binomial proportions. Comput Statist Data Anal. 2004;45:449–456] uses an adjusted Wald interval. Both methods are appropriate in varying coefficient meta-analysis models where the risk differences are allowed to vary across studies. The two methods were compared in a simulation study under realistic meta-analysis conditions and the adjusted Wald method was found to have the best performance characteristics.

Keywords: meta analysis; adjusted Wald interval; risk difference

1. Introduction

The difference between two proportions, referred to as a risk difference, is a useful measure of effect size in 2-group studies where the response variable is dichotomous. Unless the sample size in each group is large, a confidence interval for the population risk difference can be unacceptably wide. A more precise interval estimate of the risk difference can be obtained by combining results from two or more studies that have each estimated a risk difference for the same response variable. Methods for combining results from two or more studies, also known as meta-analysis, are used extensively in public health research.

The classic meta-analysis methods [1] are based on the constant coefficient meta-analysis model (also known as the fixed-effect model) and the random coefficient meta-analysis model. The constant coefficient model requires population effect sizes to be identical across studies and the classic inverse–variance estimator of the common effect size is inconsistent if the homogeneous effect size assumption is violated. It has been argued that the homogeneous effect size assumption is unrealistic and that the constant coefficient meta-analysis methods should no longer be used.[2]

The random coefficient meta-analysis model does not require effect size homogeneity. However, the random coefficient methods assume that the effect sizes from the selected studies represent a random sample from some definable superpopulation of effect sizes, which is an assumption

*Corresponding author. Email: pricejr@etsu.edu

that is difficult to justify.[3, p.41] Furthermore, the variance component of the random coefficient model is a fundamental parameter to estimate, but interval estimates of the variance component are hypersensitive to minor and difficult to detect violations of the superpopulation normality assumption.[4]

Meta-analysis methods for combining risk differences based on a varying coefficient model [5] do not require the population risk differences to be identical across studies nor do they assume that risk differences from the selected studies are a random sample from a superpopulation of effect sizes. The varying coefficient meta-analysis model is attractive in applications where the assumptions of the constant coefficient and random coefficient models are difficult to justify. Let $\Delta_j = p_{1j} - p_{2j}$ denote the population risk difference estimated in study $j (j = 1, 2, \ldots, m)$, where p_{ij} is the population proportion under condition $i (i = 1, 2)$. A parameter of interest in the varying coefficient model is the average risk difference $\Delta = \sum_{j=1}^{m} \Delta_j / m$, which can also be expressed as a linear contrast of the population proportions $\Delta = \mathbf{c}'\mathbf{p}$ where $\mathbf{c}' = \mathbf{1}' \otimes [1 - 1]/m, \mathbf{1}$ is a $m \times 1$ vector of ones, and $\mathbf{p}$ is a $2m \times 1$ vector of population proportions.

Until recently, confidence intervals for linear functions of binomial proportions were obtained using the Wald method.[6] The traditional Wald confidence interval for a linear function of binomial proportions has been shown to have a coverage probability that can be far below the nominal level, and an adjusted Wald method referred to here as the PB method, has been shown to perform substantially better than the traditional Wald method.[7]

Let f_{ij} denote the frequency count for the specified dichotomous trait in condition i and study j. A $100(1 - \alpha)\%$ PB confidence interval for Δ [5] is given below

$$\bar{\Delta} \pm z_{\alpha/2} \sqrt{\widehat{\text{var}}(\bar{\Delta})}, \tag{1}$$

where $\bar{\Delta} = m^{-1} \sum_{j=1}^{m} \hat{\Delta}_j$, $\widehat{\text{var}}(\hat{\Delta}) = m^{-2} \sum_{j=1}^{m} \widehat{\text{var}}(\hat{\Delta}_j)$, $\hat{\Delta}_j = \hat{p}_{1j} - \hat{p}_{2j}$, $\hat{p}_{ij} = (f_{ij} + 1/m)/(n_{ij} + 2/m)$, and $\widehat{\text{var}}(\hat{\Delta}_j) = \hat{p}_{1j}(1 - \hat{p}_{1j})/(n_{1j} + 2/m) + \hat{p}_{2j}(1 - \hat{p}_{2j})/(n_{2j} + 2/m)$.

A confidence interval proposed by Zou et al. [8], and referred to here as the ZHZ method, combines Wilson confidence intervals for each individual proportion. Let p_k be the kth element of the $2m \times 1$ vector $\mathbf{p}$, and c_k be the kth element of the $2m \times 1$ vector $\mathbf{c}$. A $100(1 - \alpha)\%$ ZHZ confidence interval for Δ may be expressed as

$$\begin{aligned} L &= \sum_{k=1}^{2m} c_k \hat{p}_k - \sqrt{\sum_{k=1}^{2m} [c_k \hat{p}_k - \min(c_k l_k, c_k u_k)]^2}, \\ U &= \sum_{k=1}^{2m} c_k \hat{p}_k + \sqrt{\sum_{k=1}^{2m} [c_k \hat{p}_k - \max(c_k l_k, c_k u_k)]^2}, \end{aligned} \tag{2}$$

where $\hat{p}_k = f_k / n_k$ and

$$l_k, u_k = \frac{\hat{p}_k + z^2_{\alpha/2}/(2n_k) \pm z_{\alpha/2} \sqrt{[\hat{p}_k(1 - \hat{p}_k) + z^2_{\alpha/2}/(4n_k)]/n_k}}{1 + z^2_{\alpha/2}/n_k}.$$

It is interesting to note that the PB confidence interval is a generalization of the Agresti-Caffo confidence interval for Δ_j,[9] and the ZHZ method is a generalization of the Newcombe (method 10) confidence interval for Δ_j.[10] Both the Agresti-Caffo and Newcombe confidence intervals have been shown to have excellent small-sample performance characteristics and both methods are options in SAS PROC FREQ.

The ZHZ method has been compared with the PB method in 3-group and 4-group designs for several different linear contrasts with very small sample sizes ($n_j = 5$ in at least one group). The two methods were found to be similar in performance.[8] Some methods described by Tebbs and Roths [11] were not included in the simulation study of Grizzle et al. [8] because of their poor worst-case coverage probabilities.

Although the ZHZ and PB methods have similar performance characteristics, the relative performance of the ZHZ and PB methods appears to depend on the specific type of linear contrast and the number of proportions. The purpose of the simulation study described in the following section is to compare the ZHZ and PB confidence intervals in meta-analysis designs for risk differences where the total number of proportions can be large and the contrast coefficients are of the form $\mathbf{c}' = \mathbf{1}' \otimes [1 \ -1]/m$.

2. Simulation study

The PB (1) and ZHZ (2) methods were compared under 4500 different patterns of sample sizes and p_{ij} values for $m = 2, 5, 15, 30$, and 50. The results are summarized in Table 1. For each row of Table 1, 500 sets of p_{1j} and p_{2j} values were randomly generated from a uniform distribution

Table 1. Comparison of two 95% varying coefficient confidence intervals for average risk difference.

		Average coverage		Minimum coverage		Average width	
p_{1j}	p_{2j}	ZHZ	PB	ZHZ	PB	ZHZ	PB
$m = 2$							
.02–.05	.02–.05	.989	.983	.969	.962	.166	.131
.05–.15	.05–.15	.966	.961	.953	.951	.199	.184
.45–.55	.45–.55	.945	.947	.935	.939	.285	.296
.45–.55	.05–.15	.950	.949	.939	.943	.243	.245
.85–.95	.05–.15	.936	.956	.901	.945	.189	.187
.05–.25	.05–.25	.959	.955	.950	.950	.226	.218
.40–.60	.40–.60	.945	.947	.938	.939	.286	.298
.40–.60	.05–.25	.948	.949	.935	.943	.260	.265
.05–.95	.05–.95	.949	.950	.932	.940	.251	.256
$m = 5$							
.02–.05	.02–.05	.992	.969	.975	.955	.106	.076
.05–.15	.05–.15	.969	.953	.956	.949	.128	.115
.45–.55	.45–.55	.942	.946	.930	.938	.182	.191
.45–.55	.05–.15	.947	.947	.940	.939	.157	.159
.85–.95	.05–.15	.924	.950	.903	.945	.122	.118
.05–.25	.05–.25	.960	.950	.948	.943	.144	.137
.40–.60	.40–.60	.942	.946	.930	.938	.181	.191
.40–.60	.05–.25	.945	.947	.936	.941	.161	.164
.05–.95	.05–.95	.948	.947	.922	.937	.160	.163
$m = 15$							
.02–.05	.02–.05	.995	.955	.983	.950	.062	.042
.05–.15	.05–.15	.970	.948	.960	.944	.076	.068
.45–.55	.45–.55	.937	.945	.929	.941	.105	.111
.45–.55	.05–.15	.944	.946	.938	.942	.091	.092
.85–.95	.05–.15	.913	.947	.894	.943	.071	.067
.05–.25	.05–.25	.959	.947	.950	.942	.083	.079
.40–.60	.40–.60	.937	.945	.928	.940	.105	.112
.40–.60	.05–.25	.941	.945	.934	.940	.095	.097
.05–.95	.05–.95	.946	.946	.934	.941	.095	.096

(*Continued*)

Table 1. Continued.

		Average coverage		Minimum coverage		Average width	
p_{1j}	p_{2j}	ZHZ	PB	ZHZ	PB	ZHZ	PB
$m = 30$							
.02–.05	.02–.05	.996	.950	.989	.948	.044	.030
.05–.15	.05–.15	.970	.947	.963	.943	.054	.048
.45–.55	.45–.55	.935	.945	.927	.942	.075	.079
.45–.55	.05–.15	.942	.945	.935	.942	.065	.065
.85–.95	.05–.15	.908	.946	.895	.943	.051	.048
.05–.25	.05–.25	.959	.946	.952	.942	.059	.056
.40–.60	.40–.60	.935	.945	.930	.942	.075	.079
.40–.60	.05–.25	.940	.945	.933	.942	.067	.068
.05–.95	.05–.95	.945	.945	.934	.941	.067	.068
$m = 50$							
.02–.05	.02–.05	.996	.948	.991	.945	.034	.023
.05–.15	.05–.15	.971	.946	.965	.942	.042	.037
.45–.55	.45–.55	.934	.945	.927	.942	.058	.062
.45–.55	.05–.15	.941	.945	.936	.942	.050	.051
.85–.95	.05–.15	.905	.946	.894	.943	.039	.037
.05–.25	.05–.25	.958	.946	.952	.942	.046	.043
.40–.60	.40–.60	.934	.945	.927	.942	.058	.062
.40–.60	.05–.25	.939	.945	.934	.942	.052	.053
.05–.95	.05–.95	.944	.945	.936	.942	.052	.053

[a] $n_{1j} = n_{2j} = 10$ to 100.
[b] The entries are based on 500 conditions with 75,000 trials per condition.
[c] ZHZ, Zou–Huang–Zhang method; PB, Price–Bonett method.

for the specified range of p_{ij} values, and 500 $n_{1j} = n_{2j}$ values were randomly generated from a uniform distribution with a range of 10–100. For each of the 500 conditions within each row of Table 1, 75,000 sets of $2m$ binomial random variates were randomly generated. Within each of the 500 conditions, the coverage probability and mean confidence interval width were estimated for 95% PB and ZHZ confidence intervals. Within each row of Table 1, the mean and minimum coverage probabilities across the 500 conditions are reported along with the average of the mean confidence interval widths. The results in Table 1 show that the PB method has an average coverage probability that is closer to .95 than the ZHZ method in almost every condition. More importantly, the minimum coverage probability of the ZHZ method within each set of 500 conditions can be far below the nominal .95 level. The worst-case coverage probability for the ZHZ method was .894 while the worst-case coverage probability for the PB method was .937. Although the minimum coverage probability is substantially larger for the PB method than the ZHZ method, the average confidence interval widths of the two methods are similar. In meta-analysis applications where the sample sizes typically differ across studies and the p_{ij} values are unknown, it is important that a confidence interval method perform properly regardless of the patterns of sample sizes and unknown population proportions. For the wide range of conditions examined in this study, the PB method is clearly superior to the ZHZ method. We found a similar pattern of results for 90% and 99% confidence levels.

3. Example

The computation of the PB and ZHZ confidence intervals will be illustrated in a meta-analysis of $m = 11$ studies of antibiotic treatment on urinary tract infection described by Kulinskaya et al.[12] The sample data are reported in Table 2 with sample sizes that range from 19 to 50. The

Table 2. Eleven studies of antibiotic treatment to prevent recurrent urinary tract infection.

Trial	Drug	Placebo	95% Agresti-Caffo CI
1	1/23	17/22	[−0.8733, −0.4667]
2	8/21	17/19	[−0.7152, −0.2165]
3	2/15	4/13	[−0.4564, 0.1427]
4	1/20	8/21	[−0.5332, −0.0676]
5	0/11	10/13	[−0.9229, −0.3898]
6	4/18	13/17	[−0.7611, −0.2126]
7	1/13	5/6	[−0.9625, −0.2708]
8	3/25	15/25	[−0.6731, −0.2157]
9	1/20	13/23	[−0.6978, −0.2404]
10	1/13	5/7	[−0.8861, −0.1806]
11	2/16	9/11	[−0.8891, −0.3160]

95% PB confidence interval for these data is [−0.6335, −.4703] indicating that the population proportion of patients who would develop a urinary tract infection would be .4703 to .6335 smaller under the antibiotic treatment compared to a placebo. Note that this confidence interval is substantially narrower than any of the single-study confidence intervals in Table 2. The 95% ZHZ confidence interval is [−0.6159, −0.4561]. However, the results in Table 1 suggest that a 95% ZHZ confidence interval could have a coverage probability that is considerably less than .95 and so the PB confidence interval is a more trustworthy result.

4. Concluding remarks

Meta-analysis has been used in literally thousands of studies over the last 30 years and the frequency of its use has increased dramatically in recent years. Virtually all of the published meta-analyses have used statistical methods based on either the constant coefficient model or random coefficient model. Both models require restrictive assumptions and statistical methods based on the varying coefficient model are useful alternatives to the classic methods. Previous simulation studies with 3-group and 4-group designs suggest that the PB and ZHZ methods have similar performance characteristics. The results of our simulation study for the average risk difference with $2m$ groups ($m = 2, 5, 15, 30$, and 50), which represents a typical meta-analysis design, showed that the PB method performs considerably better than the ZHZ method. When reporting a meta-analysis of risk differences, it is informative to also report confidence intervals for the risk difference in each study and display the results in a forest plot. Methods 10 or 11 of Newcombe [10] or the Agresti-Caffo confidence interval have good performance characteristics and are recommended alternatives to the traditional Wald confidence interval for a single-study risk difference.

References

[1] Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. New York: Wiley; 2009.
[2] National Research Council. Combining information: statistical issues and opportunities for research. Washington, DC: National Academy Press; 1992.
[3] Schulze R. Meta-analysis: a comparison of approaches. Toronto: Hogrefe and Huber; 2004.
[4] Kraemer KA, Bonett DG. Confidence intervals for the amount of heterogeneity in meta-analysis: small-sample properties and robustness to nonnormality. Paper presented at The meeting of the American Statistical Association, Vancouver, British Columbia, Canada; 2010.
[5] Bonett DG, Price RM. Meta-analysis methods for risk differences. Br J Math Statist Psych. 2014. doi:10.1111/bmsp.12024.

[6] Grizzle JE, Starmer CF, Koch GG. Analysis of categorical data by linear models. Biometrics. 1969;25:489–504.
[7] Price RM, Bonett DG. An improved confidence interval for a linear function of binomial proportions. Comput Statist Data Anal. 2004;45:449–456.
[8] Zou GY, Huang W, Zhang X. A note on confidence interval estimation for a linear function of binomial proportions. Comput Statist Data Anal. 2009;53:1080–1085.
[9] Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. Amer Statist. 2000;54:280–288.
[10] Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Statist Med. 1998;17:873–890.
[11] Tebbs JM, Roths SA. New large-sample confidence intervals for a linear combination of binomial proportions. J Statist Plan Infer. 2008;138:1884–1893.
[12] Kulinskaya E, Morgenthaler S, Staudte RG. Variance stabilizing the difference of two binomial proportions. Amer Statist. 2010;64:350–356.