Statistical Papers 46, 541-554 (2005)

Statistical Papers

The failure of meta-analytic asymptotics for the seemingly efficient estimator of the common risk difference

Ronny Kuhnert and Dankmar Böhning

Charité - University Medicine Berlin, Centre for Humanities and Health Sciences,
Institute for International Health, Biometry and Epidemiology,
Haus 562, Fabeckstraße 60-62, 14195 Berlin, Germany

Received: October 20, 2003; revised version: May 3, 2004

Abstract We consider the case of a multicenter trial in which the center specific sample sizes are potentially small. Under homogeneity, the conventional procedure is to pool information using a weighted estimator where the weights used are inverse estimated center-specific variances. Whereas this procedure is efficient for conventional asymptotics (e. g. center-specific sample sizes become large, number of center fixed), it is commonly believed that the efficiency of this estimator holds true also for meta-analytic asymptotics (e. g. center-specific sample size bounded, potentially small, and number of centers large). In this contribution we demonstrate that this estimator fails to be efficient. In fact, it shows a persistent bias with increasing number of centers showing that it is *not* meta-consistent. In addition, we show that the Cochran and Mantel-Haenszel weighted estimators are meta-consistent and, in more generality, provide conditions on the weights such that the associated weighted estimator is meta-consistent.

Key words Bias in conventional estimator; Cochran's estimator; Heterogeneity in baseline risk; Mantel-Haenszel estimator; Risk difference.

Offprint requests to: R. Kuhnert

1 Introduction

The interest in a clinical trial lies in estimating treatment effect. There are different effect measures, see Hutton (2000). It is common to estimate the odds ratio (OR) because of the statistical properties of the OR. Health care professionals and managers might prefer the difference of success rates of treatment group versus control group, sometimes called the risk difference or treatment difference. The risk difference is the inverse of the NNT ("Number Nedded to Treat"). The NNT-measure has gained much attention in recent years as a useful way of reporting the results of randomized controlled trials with a binary outcome. Laupacis et al. (1988) argue that the NNT is a more meaningful summary of results than the statistics like the odds ratio or the relative risk. In this paper emphasis is on estimating the risk difference which provides an estimate of the NNT as a by-product. First we take a look at the following typical setting. Consider k centers in which a treatment group and a control group are compared and the individual outcome measures are binary. Let n_i^T be the number at risk in the treatment arm in the i-th center, whereas n_i^C denotes the number at risk in the control arm. Let x_i^T be the number of events in the treatment arm for the i-th center, similarly for x_i^C. p_i^T is the probability for a positive response in the i-th center, analogical for p_i^C. In this paper, we assume a common risk difference ϑ and later, we will compare it with the expectation of estimates $\hat{\vartheta}$. The p_i^C may be different in each center, that is, we allow for baseline heterogeneity. Consequently p_i^T is given by $p_i^T = p_i^C + \vartheta$. $\hat{\vartheta}_i = \frac{x_i^T}{n_i^T} - \frac{x_i^C}{n_i^C}$ is the estimate of risk difference in center i, $i = 1, 2, ..., k$. Now, we can estimate $\hat{\vartheta}$ by

$$\hat{\vartheta} = \frac{\sum_{i=1}^{k} w_i \hat{\vartheta}_i}{\sum_{i=1}^{k} w_i}, \tag{1}$$

with suitable, non-negative weights $w_1, w_2, ..., w_k$. There are several suggestions for the weights. We will consider the weighted least squares (WLS) estimator $\hat{\vartheta}_{WLS}$ which uses

$$w_i^{WLS} = \frac{1}{\frac{p_i^T(1-p_i^T)}{n_i^T} + \frac{p_i^C(1-p_i^C)}{n_i^C}}, \tag{2}$$

the Cochran's estimator $\hat{\vartheta}_{Co}$ (Cochran, 1954), which uses

$$w_i^{Co} = \frac{n_i^T n_i^C}{n_i^T + n_i^C}, \tag{3}$$

and the Mantel-Haenszel estimator $\hat{\vartheta}_{MH}$ (Mantel and Haenszel, 1959), which uses

$$w_i^{MH} = n_i^T n_i^C. \tag{4}$$

To illustrate, let us look at the following example in Table 1 of a multicenter study with high sparsity in the data. We return to the data considered previously by Lipsitz et al.(1998). The data are from the Cancer and Leukemia

Group B (CALGB) randomized clinical trial comparing two chemotherapy treatments with respect to survival (lived/died by the end of the study) in patients with multiple myeloma (Cooper et al., 1993). A total of 156 eligible patients was accrued in the 21 centers. This example shows a typical setting of a multicenter study. The sample size is very small in the two treatment arms.

Table 1 Available data for each center in the CALGB study

Center	x_i^T	n_i^T	x_i^C	n_i^C	$\hat{\vartheta}_i = \frac{x_i^T}{n_i^T} - \frac{x_i^C}{n_i^C}$
1	1	3	3	4	-0.4167
2	8	11	3	4	-0.0227
3	2	3	2	2	-0.3333
4	2	2	2	2	0
5	0	3	2	2	-1
6	2	3	1	3	0.3333
7	2	3	2	2	-0.3333
8	4	4	1	5	0.8
9	2	3	2	2	-0.3333
10	2	3	0	2	0.6667
11	3	3	3	3	0
12	0	2	2	2	-1
13	1	5	1	4	-0.05
14	2	4	2	3	-0.1667
15	4	6	2	4	0.1667
16	3	9	4	12	0
17	2	3	1	2	0.1667
18	1	4	3	3	-0.75
19	2	3	1	4	0.4167
20	0	2	0	3	0
21	1	5	2	4	-0.3

An important factor is the efficiency of the estimated risk difference. In the conventional asymptotics, where the center-specific sample size become large and the number of centers fixed, the procedure (1) is always consistent. Another type of asymptotics is the meta-analytic framework, where the center-specific sample size is bounded, potentially small and number of centers large. This will be considered in the next section.

2 Meta-Consistent Estimators

In the following theorem, we show that all estimators of $\hat{\vartheta}$ with non-random and bounded weights are meta-consistent. A meta-consistent estimator is defined such that its mean squared error (MSE) becomes small for a large

numbers of studies, e.g.

$$\lim_{k\longrightarrow\infty} MSE(\hat{\vartheta}) = \lim_{k\longrightarrow\infty} \left(Bias(\hat{\vartheta})^2 + Var(\hat{\vartheta})\right) = 0.$$

Here, $Bias(\hat{\vartheta}) = E(\hat{\vartheta}) - \vartheta$ and $Var(\hat{\vartheta}) = E(\hat{\vartheta} - E(\hat{\vartheta}))^2$. The Bias of $\hat{\vartheta}$ is equal to zero in the case of non-random weights, because

$$\begin{aligned} E(\hat{\vartheta}) &= \frac{\sum_{i=1}^{k} w_i E(\hat{\vartheta}_i)}{\sum_{i=1}^{k} w_i} = \frac{\sum_{i=1}^{k} w_i \left(E\left(\hat{p}_i^T\right) - E\left(\hat{p}_i^C\right)\right)}{\sum_{i=1}^{k} w_i} \\ &= \frac{\sum_{i=1}^{k} w_i \left(p_i^T - p_i^C\right)}{\sum_{i=1}^{k} w_i} = \vartheta. \end{aligned}$$

Thus, we show in the following proof only that $\lim_{k\longrightarrow\infty} Var(\hat{\vartheta}) = 0$.

Theorem 1 *If a risk difference estimator* $\hat{\vartheta}$ *is defined as* $\hat{\vartheta} = \frac{\sum w_i \hat{\vartheta}_i}{\sum w_i}$ *with* $p_i^T, p_i^C \in (0,1)$ *and*

$$0 < w \leq w_i \leq W < \infty$$

for $i = 1, ..., k$ *and where* w *and* W *are real numbers, then* $\hat{\vartheta}$ *is MSE-consistent.*

Proof We will use at various stages that $p(1-p) \leq \frac{1}{4}$ *for any* $p \in [0,1]$ *and that for integers* $n, m \geq 1 : \left(\frac{1}{n} + \frac{1}{m}\right) = \frac{n+m}{nm} \leq 2$

$$\begin{aligned} Var(\hat{\vartheta}) &= \frac{\sum w_i^2 Var(\hat{\vartheta}_i)}{\left(\sum w_i\right)^2} = \frac{\sum w_i^2 \left(\frac{p_i^T(1-p_i^T)}{n_i^T} + \frac{p_i^C(1-p_i^C)}{n_i^C}\right)}{\left(\sum w_i\right)^2} \\ &\leq \frac{\frac{1}{4}\sum w_i^2 \left(\frac{1}{n_i^T} + \frac{1}{n_i^C}\right)}{\left(\sum w_i\right)^2} \leq \frac{1}{2}\frac{\sum w_i^2}{\left(\sum w_i\right)^2} \\ &\leq \frac{1}{2}\frac{kW^2}{k^2w^2} = \frac{1}{2}\frac{W^2}{kw^2} \underset{k\to\infty}{\longrightarrow} 0 \end{aligned}$$

In the next theorem, we show that for the Cochran's and WLS weights the upper bound is not required.

Theorem 2 *If a risk difference estimator* $\hat{\vartheta}$ *is defined as* $\hat{\vartheta} = \frac{\sum w_i \hat{\vartheta}_i}{\sum w_i}$ *with* $p_i^T, p_i^C \in (0,1)$ *and*

a) Cochran's weights $w_i = \left(\frac{1}{n_i^T} + \frac{1}{n_i^C}\right)^{-1}$ *or*

b) WLS weights $w_i = \frac{1}{\frac{p_i^T(1-p_i^T)}{n_i^T} + \frac{p_i^C(1-p_i^C)}{n_i^C}}$,

then $\hat{\vartheta}$ *is MSE-consistent.*

Proof a)

$$Var(\hat{\vartheta}) = \frac{\sum w_i^2 \left(\frac{p_i^T(1-p_i^T)}{n_i^T} + \frac{p_i^C(1-p_i^C)}{n_i^C} \right)}{\left(\sum w_i\right)^2}$$

$$\leq \frac{\sum w_i^2 \left(\frac{n_i^T + n_i^C}{n_i^T n_i^C} \right)}{4\left(\sum w_i\right)^2} = \frac{1}{4\sum w_i} = \frac{1}{4\sum \left(\frac{1}{n_i^T} + \frac{1}{n_i^C} \right)^{-1}}$$

$$\leq \frac{1}{2k} \underset{k\to\infty}{\to} 0$$

b)

$$Var(\hat{\vartheta}) = \frac{\sum w_i^2 \left(\frac{p_i^T(1-p_i^T)}{n_i^T} + \frac{p_i^C(1-p_i^C)}{n_i^C} \right)}{\left(\sum w_i\right)^2} = \frac{\sum w_i}{\left(\sum w_i\right)^2}$$

$$= \frac{1}{\sum w_i} = \frac{1}{\sum \left(\frac{p_i^T(1-p_i^T)}{n_i^T} + \frac{p_i^C(1-p_i^C)}{n_i^C} \right)^{-1}}$$

$$\leq \frac{1}{4\sum \left(\frac{1}{n_i^T} + \frac{1}{n_i^C} \right)^{-1}} \leq \frac{1}{2k} \underset{k\to\infty}{\to} 0$$

3 Weighted Least Squares Estimator

However, the WLS-estimator (2) cannot be used in practice since p_i^T and p_i^C are unknown. In several textbooks of epidemiology (see Kleinbaum, Kupper, and Morgenstern, 1982) and in textbooks of meta-analysis (see Petitti, 1994), it is suggested to replace p_i^T and p_i^C in (2) by their sample estimators $\frac{x_i^T}{n_i^T}$ and $\frac{x_i^C}{n_i^C}$. This leads to $\hat{\vartheta}_W$, which uses

$$w_i^W = \frac{1}{\frac{x_i^T(n_i^T - x_i^T)}{(n_i^T)^3} + \frac{x_i^C(n_i^C - x_i^C)}{(n_i^C)^3}}. \tag{5}$$

One could presume that $\hat{\vartheta}_W$ is a consistent estimator, but this estimator is generally no longer unbiased. Note that (5) is not defined, if $x_i^T = 0$ or $x_i^T = n_i^T$ in combination with $x_i^C = 0$ or $x_i^C = n_i^C$. In our example (Table 1) this case appears in the centers 4, 5, 11, 12, 20. There are different possibilities in dealing with these cases. Lipsitz et al. (1998) remove that center from the pool where such a case has occurred. This would imply that five centers are removed in our example. Consequently, 24 patients (15.4%) are lost in the analysis through the statistical procedure. Greenland and Robins (1985) added to each cell $\frac{1}{2}$ and mentioned that they tried several other constants, including adding $\frac{1}{2}$ only in calculating of the weights w^W, but none of these produced a better performance. O'Gorman (1994)

used the same approach as Greenland and Robins (1985). Based upon a comparative investigation of these three methods (see appendix), we prefer in our considerations the approach to add $\frac{1}{2}$ only in calculating the weights w^W. Consequently, the weights become

$$w_i^W = \frac{1}{\frac{(x_i^T+0.5)(n_i^T-x_i^T+0.5)}{(n_i^T+1)^2 n_i^T} + \frac{(x_i^C+0.5)(n_i^C-x_i^C+0.5)}{(n_i^C+1)^2 n_i^C}} \tag{6}$$

The estimated risk differences in our example are $\hat{\vartheta}_{Co} = -0.0572$, and $\hat{\vartheta}_{MH} = -0.0198864$, and $\hat{\vartheta}_W = -0.0710186$. All three estimators show that treatment 1 (in our case treatment group) is better than treatment 2.

4 Exact Computation of Bias, Variance and Mean Squared Error of Risk Difference $\hat{\vartheta}$

Let $f(\vartheta')$ denote the density of the risk difference $\hat{\vartheta}$ in the population of all possible outcomes of the clinical trial. Then the moments of $\hat{\vartheta}$ can be found with respect to $f(\vartheta')$, using that x_i^T and x_i^C are binomial variates. Let x^T denote the k-vector $(x_1^T, x_2^T, ..., x_k^T)^T$ and x^C the k-vector $(x_1^C, x_2^C, ..., x_k^C)^T$. In this case, the density of (x^T, x^C) is given as

$$f(x^T, x^C) = \prod_{i=1}^{k} \binom{n_i^T}{x_i^T} (p_i^T)^{x_i^T} (1-p_i^T)^{n_i^T - x_i^T} \times \binom{n_i^C}{x_i^C} (p_i^C)^{x_i^C} (1-p_i^C)^{n_i^C - x_i^C} \tag{7}$$

Formula (7) can be used to compute the quantities of interest, exactly. The expectation of $\hat{\vartheta}$ is obtained by

$$E(\hat{\vartheta}) = \sum_{\text{all possible } x^T, x^C} f(x^T, x^C)\hat{\vartheta}(x^T, x^C) \tag{8}$$

$$\text{with } \hat{\vartheta}(x^T, x^C) = \frac{\sum_{i=1}^k w_i \left(\frac{x_i^T}{n_i^T} - \frac{x_i^C}{n_i^C}\right)}{\sum_{i=1}^k w_i}. \tag{9}$$

The variance is given by

$$Var(\hat{\vartheta}) = E(\hat{\vartheta}^2) - E(\hat{\vartheta})^2 \tag{10}$$

$$\text{with } E(\hat{\vartheta}^2) = \sum_{\text{all possible } x^T, x^C} f(x^T, x^C)\hat{\vartheta}(x^T, x^C)^2 \tag{11}$$

and the MSE is

$$MSE(\hat{\vartheta}) = Var(\hat{\vartheta}) + (E(\hat{\vartheta}) - \vartheta)^2. \tag{12}$$

Now, to compare the several estimators with respect to bias, variance and MSE, we assume fixed values of $p_1^C, ..., p_k^C$. The risk difference is set to a predetermined value, and consequently $p_i^T = p_i^C + \vartheta$. Furthermore, we

assume common values of n^T and n^C for each center. Consequently, the Cochran's and Mantel-Haenszel weights provide identical solutions, so only one of them is considered. For the example, we set ϑ to 0.2, n_i^T to 4 and n_i^C to 3 for each center and $k = 1, ..., 6$. The outcomes are shown in Table 2. The p_i^C-values are taken from a uniform distribution on 0 to 0.5. In Table 2, we can see that Cochran's estimator is unbiased. If k equal 1, then bias and variance and mean squared error are identical in all estimators, but otherwise the variance of the estimator using Cochran's weights is less than the variance of the WLS-estimator, in all cases. The same is true for the mean squared error.

Table 2 Bias and Variance and M.S.E. of $\hat{\vartheta}$ as a function of number of centers k for given values of baseline values and of the common risk difference with $p_1^C = 0.22$, $p_2^C = 0.124$, $p_3^C = 0.32$, $p_4^C = 0.376$, $p_5^C = 0.263$, $p_6^C = 0.241$ and $\vartheta = 0.2$

k	Bias		Variance		M.S.E	
	w^{WLS}	w^{Co}	w^{WLS}	w^{Co}	w^{WLS}	w^{Co}
1	.000000	.000000	.076517	.076517	.076517	.076517
2	.003801	.000000	.038467	.037056	.038481	.037056
3	.009067	.000000	.031268	.028439	.031350	.028439
4	.013242	.000000	.025876	.022533	.026051	.022533
5	.014618	.000000	.021014	.018047	.021228	.018047
6	.006197	.000000	.014892	.013633	.014930	.013633

4.1 Property of the Bias of WLS-Estimator

The bias depends on diverse factors. One factor is the true risk difference itself and a second factor is the baseline rate p^C. In the next step, we will investigate these properties. For this purpose, we set n_i^T and n_i^C to 4 for $i = 1, ..., 5$. The respective results for various values of ϑ (=0,0.1,0.2,0.3,0.4) in combination with various values of p^C (=0.001,0.01,0.020.05,0.1,0.15,0.2, 0.25,0.3), constant over all centers, are shown in Table 3. An interesting point in Table 3 is that, if ϑ equal zero, then the WLS-estimator is unbiased. Next, we can see that the larger p^C becomes, the larger is the amount of bias for $\hat{\vartheta}$.

Another important influential factor is the difference between n^T and n^C. For that, we will set ϑ to 0 , k to 4 and p^C to 0.3. The bias for different n^T, n^C values is presented in Table 4. In this table, it can be seen again, that when ϑ equal zero and $n^T = n^C$, then the WLS-estimator is unbiased. If n^T is less than n^C and greater than 1, then the bias of $\hat{\vartheta}$ is negative. On the other hand, if n^C is less than n^T and greater than 1, then the bias of $\hat{\vartheta}$ is positive. Note the quasi-symmetries in Table 4.

We can only compute the exact form of bias, variance and MSE for small k, n^T and n^C values. For k larger than six there are too many possibilities.

Table 3 Bias as a function of p^C and $\hat{\vartheta}$, with $n_i^T = n_i^C = 4$ for $i = 1, ..., k$ and $k = 5$.

$p^C \backslash \vartheta$	0.0	0.1	0.2	0.3	0.4
.001	.00000000	-.02659156	-.03725382	-.03368700	-.01945577
.01	.00000000	-.02517564	-.03457782	-.03008781	-.01538668
.02	.00000000	-.02361120	-.03164446	-.02617322	-.01099683
.05	.00000000	-.01899706	-.02313324	-.01499920	.00131722
.10	.00000000	-.01170146	-.01010149	.00154691	.01887077
.15	.00000000	-.00509560	.00122232	.01526837	.03258072
.20	.00000000	.00065347	.01063313	.02601115	.04238060
.25	.00000000	.00543737	.01801346	.03370386	.04825662
.30	.00000000	.00919283	.02330501	.03832191	.05021397

Table 4 Bias of $\hat{\vartheta}$ for several sample sizes, with $p^C = 0.3$ and $\vartheta = 0$.

$n^T \backslash n^C$	1	2	3	4	5	6
1	.000000	.012212	.010853	.008370	.006425	.005027
2	-.012212	.000000	-.003548	-.009063	-.013688	-.017230
3	-.010853	.003548	.000000	-.006743	-.012836	-.017734
4	-.008370	.009063	.006743	.000000	-.006541	-.011990
5	-.006425	.013688	.012836	.006541	.000000	-.005600
6	-.005027	.017230	.017734	.011990	.005600	.000000

Note that the total of all possibilities is $\left((n^T + 1)(n^C + 1)\right)^k$. When $k = 7$ and $n^T = 4$ and $n^C = 3$, we have 1.28×10^9 possibilities. We used a Pentium III processor with 700 MHz and 64 MB Ram. This computer can calculate 4597 possibilities per second. With $k = 7$ centers, 77 hours are needed. To calculate the quantities of interest for larger k values within a reasonable time, simulation techniques are required.

5 A Simulation Study

In this section, we are interested in studying bias, variance and MSE of the suggested estimators when considering larger k -values. We assume that $p_1^C, ..., p_k^C$ arise from a uniform distribution on 0 to 0.5 and the risk difference is set to 0.3. x^T and x^C are generated from a binomial distribution, with parameters n_i^T, n_i^C and p_i^T, p_i^C for $i = 1, ..., k$. To mimic variation in the sample size, n_i^T and n_i^C were generated from a Poisson distribution with parameter $\lambda = 5$ for $i = 1, ..., k$. Then the three estimators $\hat{\vartheta}_W$, $\hat{\vartheta}_{Co}$ and $\hat{\vartheta}_{MH}$ were computed for $k = 1, 10, 20, ..., 2000$. The procedure was replicated $rep = 1,000$ times. From these replicates, bias, variance and MSE were calculated. The values of $\hat{\vartheta}_{Co}$ and $\hat{\vartheta}_{MH}$ are nearly identical, so that we only consider one of them. The result from this simulation is shown in the Figures

1, 2 and 3, where the parameter of interest is compared with the $\hat{\vartheta}_W$ (fine line) and $\hat{\vartheta}_{Co}$ (thick line), respectively. In Figure 1, the bias of $\hat{\vartheta}_{Co}$ oscillated around the zero line, but the graph of $\hat{\vartheta}_W$ showed a mean bias of 0.02. Note that, for the bias of $\hat{\vartheta}_W$, we cannot observe a persistent trend towards zero, if the number of centers becomes large. Consequently, we cannot expect a persistent trend for the MSE (Figure 3) towards zero. For k as large as 2000 a persistent bias is evident. An interesting point is that the variance of $\hat{\vartheta}_W$ (Figure 2) is larger than the variance of $\hat{\vartheta}_{Co}$ and $\hat{\vartheta}_{MH}$, in all cases. Thus, the weights of the estimator of $\hat{\vartheta}_W$ are not variance minimizing as erroneously assumed in the literature, see for example Shadish and Haddock (1994) or Kleinbaum, Kupper, and Morgenstern (1982). Shadish described the variance minimizing weights w_i^{WLS}, but in the example the true weights are replaced by the sample estimates, without any comments.

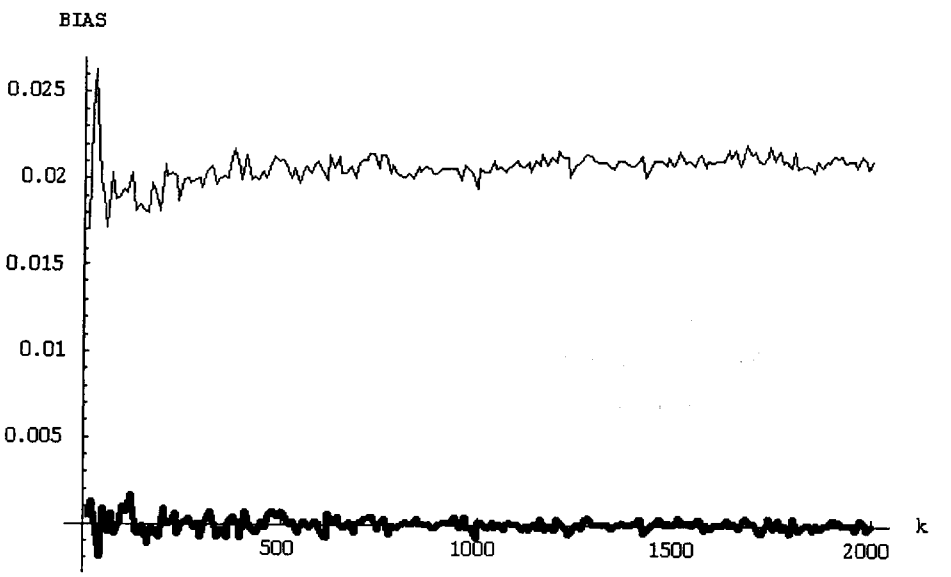

Fig. 1 Bias of $\hat{\vartheta}_W$ (fine line) and $\hat{\vartheta}_{Co}$ (thick line) as function of number of centers $k = 10, 20, ..., 2000$, $\vartheta = 0.3$, $p^C \sim U(0, 0.5)$

6 Discussion

The results show that the commonly believed efficient estimator $\hat{\vartheta}_W$ is only really efficient under specific conditions. If these conditions include the situation that the true risk difference is zero (we do not have an effect) and the sample size of treatment and control are identical (balanced trial), then the estimator $\hat{\vartheta}_W$ is unbiased. In practice it might be that the balanced design

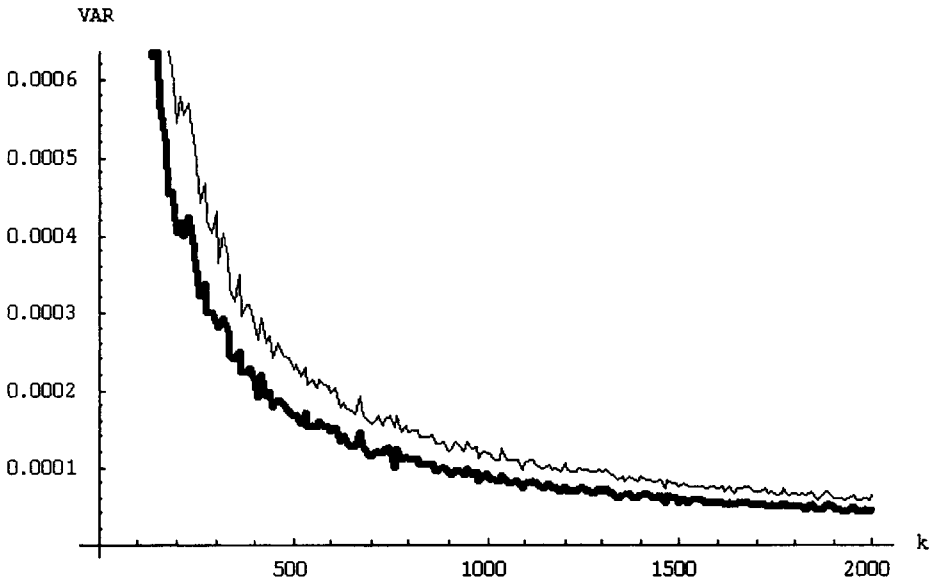

Fig. 2 Variance of $\hat{\vartheta}_W$ (fine line) and $\hat{\vartheta}_{Co}$ (thick line) as function of number of centers $k = 10, 20, ..., 2000$, $\vartheta = 0.3$, $p^C \sim U(0, 0.5)$

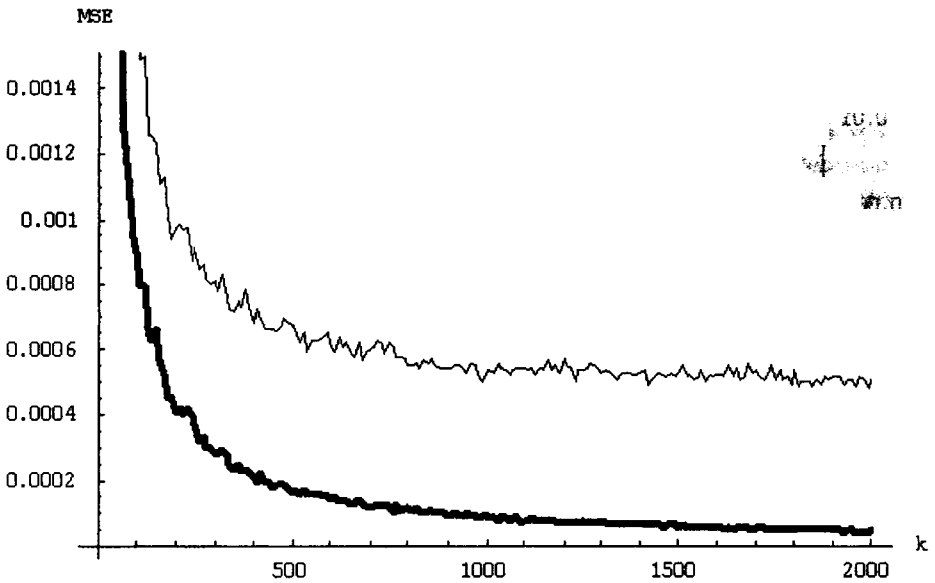

Fig. 3 m.s.e. of $\hat{\vartheta}_W$ (fine line) and $\hat{\vartheta}_{Co}$ (thick line) as a function of number of centers $k = 10, 20, ..., 2000$, $\vartheta = 0.3$, $p^C \sim U(0, 0.5)$

of a randomized trial is met in majority, but it is less likely that the true risk difference is exactly or close to zero for most cases. We have shown, by means of an exact computation for small number of centers and by means of a simulation study for larger number of centers, that the weighted least squares estimator appears to behave unsatisfactory, in general. Furthermore, we have shown that the bias of $\hat{\vartheta}_W$ is dependent on the true risk difference itself, on the baseline rate, and on the difference of sample size between treatment arm and control arm. Another important point is that a persistent bias for $\hat{\vartheta}_W$ exists, also when the number of centers becomes large. Consequently, the weighted least squares estimator is not meta-consistent. The source of the bias for $\hat{\vartheta}_W$ is the replacement of the unknown p_i^T and p_i^C with $\frac{x_i^T}{n_i^T}$ and $\frac{x_i^C}{n_i^C}$. In conclusion, the WLS-estimator with estimated weights should be avoided. Instead, the Cochran's or the Mantel-Haenszel estimator provide meta-consistent alternatives, which were also in special cases efficient.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Ekkehart Dietz for his helpful comments. This research is supported by the German Research Foundation.

APPENDIX

Here, we compare the three different published possibilities in dealing with the occurrence of zeros, when we use the WLS-estimator. The first approach comes from Lipsitz et al. (w_{Lip}). They remove the center if $\frac{x_i^T(n_i^T - x_i^T)}{(n_i^T)^3} + \frac{x_i^C(n_i^C - x_i^C)}{(n_i^C)^3}$ is equal to zero. The second approach used, due to Greenland and Robins (w_{GR1}), is to add $\frac{1}{2}$ to each cell. The third approach is mentioned also by Greenland and Robins (w_{GR2}). In this case $\frac{1}{2}$ is added only in the weights and not for calculating of $\hat{\vartheta}_i$. In Figures 4, 5 and 6 we compare these three approaches with the following settings: $\vartheta = 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4$, and 0.45, $p^C \sim U(0, 0.5)$, and $n^T, n^C \sim Poi(5)$, and $k = 1000$. In Figure 4, we can see that in Lipsitz's approach a nearly constant bias of $\hat{\vartheta}_W = -0.032$ is generated. The other two approaches are unbiased when ϑ is zero. If ϑ is unequal to zero, then the absolute value of bias $\hat{\vartheta}_W$, when using the weights w_{GR2} is always lower than the with using the weights w_{GR1}. Particularly, when looking at the MSE, we can see that in the third approach using w_{GR2} the value of the MSE is lower than in the second approach using w_{GR1}. This is the main reason why we used the third approach for our comparative analysis. When looking at the variance (Figure 5), the approach using w_{GR1} has always the lowest value.

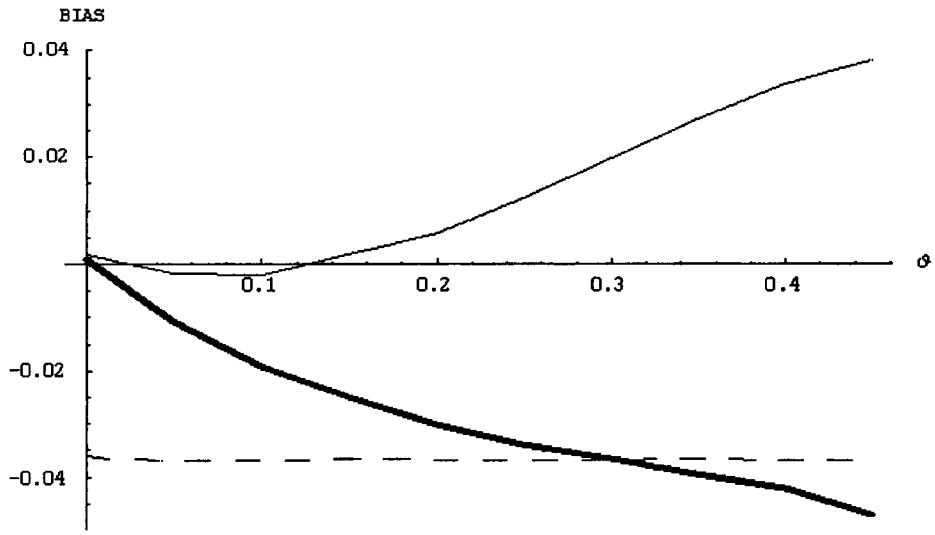

Fig. 4 Bias of $\hat{\vartheta}_W$ with w_{Lip} (dashed line), w_{GR1} (thick line) and w_{GR2} (fine line) as function of ϑ

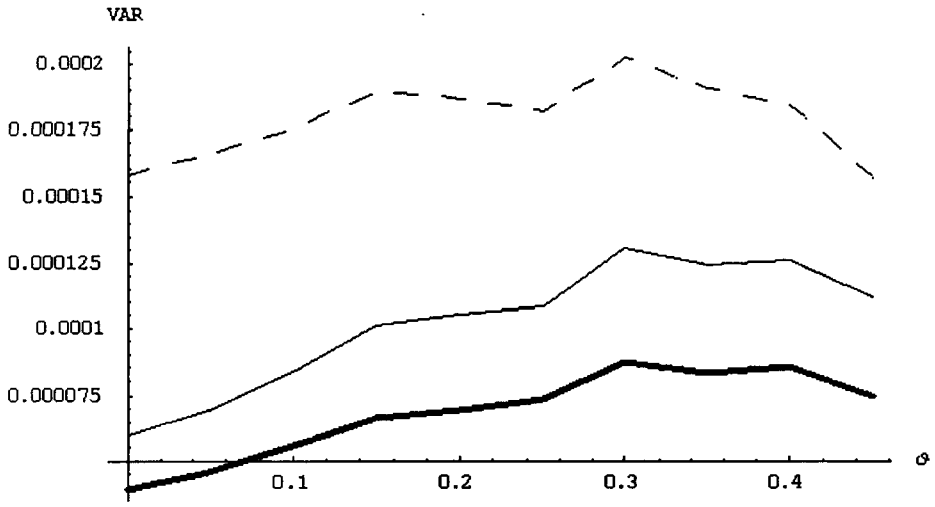

Fig. 5 Variance of $\hat{\vartheta}_W$ with w_{Lip} (dashed line), w_{GR1} (thick line) and w_{GR2} (fine line) as function of ϑ

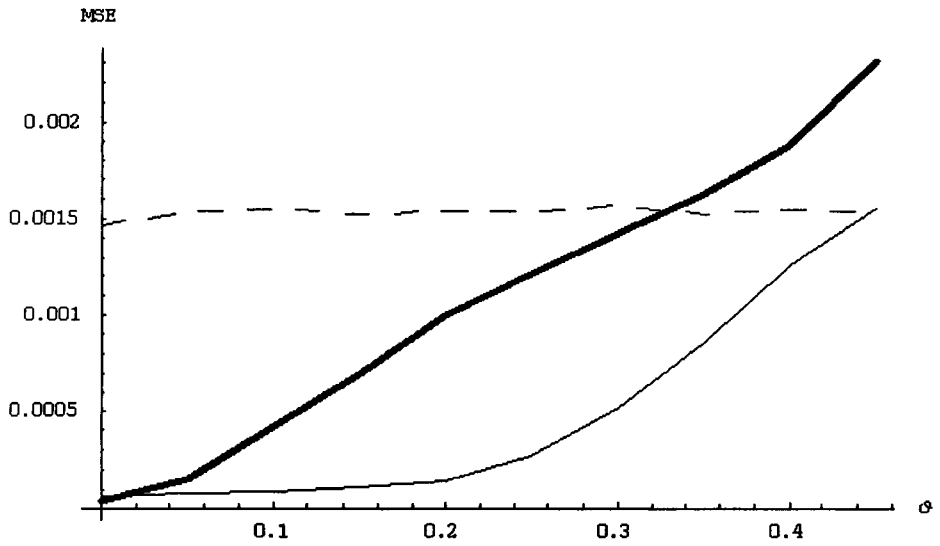

Fig. 6 m.s.e. of $\hat{\vartheta}_W$ with w_{Lip} (dashed line), w_{GR1} (thick line) and w_{GR2} (fine line) as function of ϑ

REFERENCES

1. Cochran WG (1954) Some methods of strengthening the common χ^2 tests. Biometrics **10,** 417-451
2. Cooper MR, Dear KBG, McIntyre OR, Ozer H, Ellerton J, Cannellos G, Duggan B, and Schiffer C (1993) A randomized clinical trial comparing Melphalan/Prednisone with and without α-2b interferon in newly-diagnosed patients with multiple myeloma: A cancer and leukemia group B study. Journal of Clinical Oncology **11**, 155-160
3. Greenland S, Robins JM (1985) Estimation of a common effect parameter from sparse follow-up data. Biometrics **41**, 55-68
4. Hutton JL (2000) Number needed to treat: properties and problems. Journal of the Royal Statistical Society A **163**, 403-19
5. Kleinbaum DG, Kupper LL, and Morgenstern H (1982) Epidemiologic Research. New York, Van Nostrant Reinhold
6. Laupacis A, Sackett DL, and Roberts RS (1988) An assessment of clinically useful measures of the consequences of treatment. New England journal of medicine **318**, 1728-1733
7. Lipsitz SR, Dear KBG, Laird NM, Molenberghs G (1998) Tests for homogeneity of the risk difference when data are sparse. Biometrics **54**, 148-160
8. Mantel N and Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst **22**, 719-748,

9. O'Gorman TW, Woolson RF, Jones MP (1994) A comparison of two methods of estimating a common risk difference in a stratified analysis of a multicenter clinical trial. Controlled Clinical Trials **15**, 135-153
10. Petitti DB (1994) Meta-analysis, decision analysis and cost-effectiveness analysis: methods for quantitative synthesis in medicine. Oxford University Press
11. Shadish WR, and Haddock CK (1994) Combining estimates of effect size. In: Cooper H, and Hedges LV (editors) The Handbook of Research Synthesis. New York, Russell Sage Foundation, 261-84.