A comparison of risk difference estimators in multi-center studies under baseline-risk heterogeneity

Chukiat Viwatwongkasem\textsuperscript{a,}*\textsuperscript{3}, Walailuck Böhning\textsuperscript{b}

\textsuperscript{a}Department of Biostatistics, Faculty of Public Health, Mahidol University, 420/1 Rajvithi Road, Phayatai District, Bangkok 10400, Thailand
\textsuperscript{b}Department of Medicine, Phramongkutklao Hospital, Rajvithi Road, Phayatai District, Bangkok 10400, Thailand

Received 1 February 2002; received in revised form 1 March 2002

Abstract

The risk difference is frequently used as a measure of the actual gain in the success rate between two treatments within a center (i.e. hospital). Interest is devoted to combining the risk difference across several centers under homogeneity but allowing for baseline-risk heterogeneity in each of the treatment arms. The purpose is to compare the efficiency of six estimators for the common risk difference. The six estimators consist of the Pooling method ignoring the stratification of centers, several popular sets of different weights, and a new estimator. A simulation study was done to compare bias, variance and mean-square error. The sample sizes in each center varied as 4, 8, 16, 32, 64 and the number of centers as 4, 8, 16, 32, 64. The major result is that the new estimate is an attractive compromise when choosing between the estimators of the set of the center-specific sample size weights and the estimators of the set of the inverse-variance weights. It is not an optimal strategy, but it widely extends to cover heterogeneity cases. For small sample size ($n \leq 8$), the Cochran and the Mantel–Haenszel estimators are most efficient because of their smallest mean square errors. Cochran and Mantel–Haenszel estimates are also unbiased and consistent with respect to both sample size and center size. For large sample size ($n \geq 32$), Lipsitz et al. and Rothman–Boice estimates whose weights are the inverses of variances are the most appropriate. Lipsitz et al. and Rothman–Boice estimates are considerably biased (even if asymptotically unbiased with respect to the sample size). The Pooling estimate is very close and similar to Cochran’s estimate under homogeneity of equal risk difference across centers. We recommend to use Cochran, Mantel–Haenszel, or the Pooling estimators when $n \leq 8$, to use

\*Corresponding author.
\textsuperscript{3}E-mail address: phcvw@mahidol.ac.th (C. Viwatwongkasem).

0167-9473/03/$ - $ see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-9473(02)00175-5
Lipsitz et al. and Rothman–Boice estimators when \( n \geq 32 \), and to use the new estimator when strong baseline heterogeneity occurs. © 2002 Elsevier Science B.V. All rights reserved.

**Keywords:** DerSimonian–Laird; Homogeneity; Heterogeneity adjusted estimator; Meta-analysis; Pooling of sparse clinical trials; Random effect model; Risk difference

1. Introduction

In many clinical trials, patients are randomized to one of two treatment groups within a center (i.e. hospital), with treatment allocation (approximately) balanced within centers. The risk difference is certainly one of the most important epidemiologic indices to measure actual gain of success rate between two treatment groups. The interest is in estimating the common risk difference or combining risk difference of several centers in the summary estimator under homogeneity of equal risk difference across centers but allowing for baseline-risk heterogeneity in each of the treatment arms.

Considering \( k \) centers in which two treatments are compared and the outcome measures are binary, the *risk difference* is defined by \( \delta_i = q_i - p_i \) for \( i = 1, 2, \ldots, k \) where \( p_i \) be the probability for positive response in treatment 1, \( q_i \) be the probability for positive response in treatment 2. The maximum likelihood estimate of \( \delta_i \) is provided by

\[
\hat{\delta}_i = \hat{q}_i - \hat{p}_i = \frac{Y_i}{m_i} - \frac{X_i}{n_i}.
\]

The \( X_i \), following the binomial distribution, are the number of positive responses out of the sample size \( n_i \) that received treatment 1 (the control arm), and \( Y_i \) are the number of positive responses out of the sample size \( m_i \) that received treatment 2. We know that \( E(\hat{\delta}_i) = \delta_i = \delta \), and, since \( X_i \) and \( Y_i \) are independent, we have

\[
\text{var}(\hat{\delta}_i) = \text{var}\left( \frac{Y_i}{m_i} - \frac{X_i}{n_i} \right) = \frac{q_i(1-q_i)}{m_i} + \frac{p_i(1-p_i)}{n_i}.
\]

The parameter of interest (the true unknown overall treatment effect under homogeneity of equal risk difference across centers) is \( \delta \). In case there is *effect heterogeneity* \( \delta \) is defined as the population average of the effect distribution \( \delta = \int \delta_i f(\delta_i) \, d\delta_i \) though this situation is not considered here. Summary estimates are given by

\[
\hat{\delta}_w = \frac{\sum_{i=1}^{k} w_i \hat{\delta}_i}{\sum_{i=1}^{k} w_i} \quad \text{where} \quad w_i \geq 0.
\]

Many choices of \( w_i \) are possible and the question arises which is the best one.

The purpose of this research is to compare the efficiency of six estimators for the common risk difference \( \delta \). Six estimators comprise of the pooling method, two popular sets of different weights and a new estimator as will be described in detail in the following sections.
2. Common weighting strategies

2.1. Cochran’s weights

Among the first papers dealing with combining the study-specific risk differences is Cochran (1954). The Cochran weighted estimator is widely used as a standard estimator for common risk difference estimation; it is weighted by the harmonic means of the center-specific sample sizes. Cochran proposed the weighted risk difference estimate as

$$\hat{\delta}_{\text{COC}} = \frac{\sum_{i=1}^{k} w_i \hat{\delta}_i}{\sum_{i=1}^{k} w_i}, \text{ where } w_i = n_i m_i/(n_i + m_i).$$

We notice that the centers with a large number of patients get more weight than centers with a smaller number of patients. There is no need to estimate any nuisance parameters since the weights are only a function of the sample sizes. The non-random (fixed) weights occur optimally when there is complete homogeneity. (e.g. $p_i = q_j$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, k$). In addition, Cochran also provided more generally in the case that the risk difference might be considered as a constant ($\delta_i = \delta = c$), for varying levels of $p_i$, the choice of $w_i$ could also be close to the optimum. The variance of $\hat{\delta}_{\text{COC}}$ is readily available as

$$\text{var}(\hat{\delta}_{\text{COC}}) = \frac{\sum_{i=1}^{k} w_i^2 \text{var}(\hat{\delta}_i)}{(\sum_{i=1}^{k} w_i)^2} = \frac{1}{(\sum_{i=1}^{k} w_i)^2} \sum_{i=1}^{k} w_i^2 \left( \frac{p_i(1 - p_i)}{n_i} + \frac{q_i(1 - q_i)}{m_i} \right).$$

Under complete homogeneity, the estimated variance can be easily derived as

$$\text{var}(\hat{\delta}_{\text{COC}}) = \frac{\sum_{i=1}^{k} w_i \tilde{p}_i(1 - \tilde{p}_i)}{(\sum_{i=1}^{k} w_i)^2} , \text{ where } \tilde{p}_i = (X_i + Y_i)/(n_i + m_i).$$

It is straightforward to show that $\hat{\delta}_{\text{COC}}$ is an unbiased estimator under the assumption of homogeneity of the risk difference across centers.

2.2. Böhning–Sarol estimator as a Mantel–Haenszel type

Böhning and Sarol (2000) proposed an estimator of the form

$$\hat{\delta}_{\text{MH}} = \frac{\sum_{i=1}^{k} w_i \hat{\delta}_i}{\sum_{i=1}^{k} w_i} \text{ with } w_i = n_i m_i.$$

This estimator is in line with the Mantel–Haenszel’s. Let us consider again $\hat{\delta}_i = Y_i/m_i - X_i/n_i$ which evidently can be written as

$$\hat{\delta}_i = \frac{Y_i n_i - X_i m_i}{n_i m_i}.$$
Now, instead of taking \( (1/k) \sum_{i=1}^{k} \hat{\delta}_i \) we consider the ratio of sums
\[
\hat{\delta}_{\text{MH}} = \frac{\sum_{i=1}^{k} (Y_i n_i - X_i m_i)}{\sum_{i=1}^{k} n_i m_i}.
\]

Note that \( \hat{\delta}_{\text{MH}} \) is a weighted average of the \( \hat{\delta}_i \)'s, these weights \( w_i \) are non-random. Consequently, \( \hat{\delta}_{\text{MH}} \) is unbiased. In addition, its variance is readily available as
\[
\text{var}(\hat{\delta}_{\text{MH}}) = \frac{\sum_{i=1}^{k} \left[ n_i^2 m_i q_i (1 - q_i) + m_i^2 n_i p_i (1 - p_i) \right]}{(\sum_{i=1}^{k} m_i n_i)^2}
\]
from which an estimated variance can be easily derived as
\[
\text{var}(\hat{\delta}_{\text{MH}}) = \frac{\sum_{i=1}^{k} \left[ n_i^2 Y_i (m_i - Y_i)/m_i + m_i^2 X_i (n_i - X_i)/n_i \right]}{(\sum_{i=1}^{k} m_i n_i)^2}.
\]

### 2.3. Weighted least-square estimators

Lipsitz et al. (1998) were mainly interested in the issue of developing and evaluating (heterogeneity) tests for the null hypothesis (of homogeneity):

- **H0**: \( \delta_i = \delta \) for every \( i = 1, 2, \ldots, k \)
- **H1**: \( \delta_i \neq \delta_j \) for some \( i \neq j \).

They considered linear, unbiased estimates of the form
\[
\hat{\delta}_{\text{LS}} = \hat{\delta}_w = \frac{\sum_{i=1}^{k} w_i \hat{\delta}_i}{\sum_{i=1}^{k} w_i}
\]
with non-random, non-negative constant weights \( w_i \), defined by the reciprocals of the variances as
\[
w_i = \frac{1}{\text{var}(\hat{\delta}_i)} = \frac{1}{p_i (1 - p_i)/n_i + q_i (1 - q_i)/m_i}.
\]

Radhakrishna (1965) showed that under homogeneity of risk difference across several centers (\( \hat{\delta}_i \) are constant) the inverse-variance weights \( w_i \) minimize the variance of the summary estimator (1). Note that in this case the variance of the summary estimator (1) is just given by
\[
\text{var}(\hat{\delta}_{\text{LS}}) = \frac{1}{\sum_{i=1}^{k} w_i}.
\]

This implies that there is no other estimator of the form
\[
\frac{\sum_{i=1}^{k} w_i^* \hat{\delta}_i}{\sum_{i=1}^{k} w_i^*}
\]
with smaller variance. However, the summary estimator (1) cannot be used in practice since \(p_i\) and \(q_i\) are unknown. Therefore, it has become common practice to replace them by their sample estimators \(\hat{p}_i = \frac{X_i}{n_i}\) and \(\hat{q}_i = \frac{Y_i}{m_i}\) leading to

\[
\hat{\delta}_{LS} = \frac{\sum_{i=1}^{k} \hat{w}_i \hat{\delta}_i}{\sum_{i=1}^{k} \hat{w}_i}
\]

with

\[
\hat{w}_i^{-1} = \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i} + \frac{\hat{q}_i(1 - \hat{q}_i)}{m_i}
\]

or

\[
\hat{w}_i^{-1} = \frac{X_i(n_i - X_i)}{n_i^3} + \frac{Y_i(m_i - Y_i)}{m_i^3}.
\]

This estimator is suggested in several textbooks of epidemiology such as Kleinbaum et al. (1982, p. 359) or in textbooks of meta-analysis such as Petitti (1994, p. 103). The work of Lipsitz et al. was later critically discussed and extended by Lui and Kelly (1999). The problem with this estimator is that a weight in (2) is not defined in the occurrence of any of the four cases: \(X_i = 0\) or \(n_i\) in combination with \(Y_i = 0\) or \(m_i\). Lipsitz et al. (1998) remove that center from the computation of \(\hat{\delta}_{LS}\) for which such a case has occurred.

2.4. The Rothman and Boice estimator

We know that a weight in (2) is not defined in any of the four cases: \(X_i = 0\) or \(n_i\) in combination with \(Y_i = 0\) or \(m_i\). Rothman and Boice (1979) proposed a modified least-squares estimator (in order to avoid division by 0) by adding a constant \((c = \frac{1}{2})\) to each cell, and they also replace \(w_i\) with \(w_i = n_i m_i N_i / [t_i (N_i - t_i)]\). The estimator is defined by

\[
\hat{\delta}_{ROTH} = \frac{\sum_{i=1}^{k} \hat{w}_i \hat{\delta}_i}{\sum_{i=1}^{k} \hat{w}_i},
\]

where \(\hat{w}_i = n_i m_i N_i / [t_i (N_i - t_i)]\), \(N_i = n_i + m_i\), \(t_i = X_i + Y_i\).

2.5. The pooling estimator under homogeneity

For the pooling method, the success rate for each treatment is calculated by simply dividing the total number of successes by the corresponding total sample size, ignoring the stratification of centers. This yields

\[
\hat{p} = \frac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{k} n_i} \quad \text{and} \quad \hat{q} = \frac{\sum_{i=1}^{k} Y_i}{\sum_{i=1}^{k} m_i}.
\]

Under homogeneity for each treatment arm, we have that \(p_i = p\) and \(q_i = q\) for all \(i = 1, \ldots, k\). This implies that

\[
\hat{w}_i^{-1} = \text{var}(\hat{\delta}_i) = p_i(1 - p_i)/n_i + q_i(1 - q_i)/m_i = p(1 - p)/n_i + q(1 - q)/m_i.
\]

\(^1\) A “\(^\sim\)” on the index of the estimator indicates that the non-random weight is replaced by estimated weight.
Hence, in a pooled manner we get
\[ \hat{\delta}_{\text{pool}} = \frac{\sum_{i=1}^{k} \hat{\delta}_i}{\sum_{i=1}^{k} \hat{w}_i}, \]
where \( \hat{w}_i^{-1} = \hat{p}(1 - \hat{p})/n_i + \hat{q}(1 - \hat{q})/m_i. \)

3. A new estimator under baseline-risk heterogeneity

DerSimonian and Laird (1986) applied the idea of two-stage random effect models to obtain an overall stratified-adjusted assessment of the log-odds ratio of the treatment effect from a meta-analysis of many studies where there is unobserved heterogeneity, or extra-variation or over-dispersion, among studies.

Let us consider again (1), namely \( \hat{\delta}_{\text{LS}} = \frac{\hat{\delta}}{\sum_{i=1}^{k} \hat{w}_i} = \frac{\sum_{i=1}^{k} \hat{w}_i \hat{\delta}_i}{\sum_{i=1}^{k} \hat{w}_i} = \frac{\sum_{i=1}^{k} \hat{w}_i}{\sum_{i=1}^{k} \hat{w}_i}, \)
where \( \hat{w}_i^{-1} = \text{var}(\hat{\delta}_i) = p_i(1 - p_i)/n_i + q_i(1 - q_i)/m_i. \)

Now, if we consider a more general \( w_i \), namely
\[ w_i^{-1} = \text{var}(\hat{\delta}_i) = p_i(1 - p_i)/n_i + \tau^2_p + q_i(1 - q_i)/m_i + \tau^2_q. \]

Note that if \( \tau^2_p = \tau^2_q = 0 \) in (3), the weight \( w_i \) in (3) is the same as in (1). How do these weights develop? The idea is that the variance in each of the treatment arms is a partition into two terms: the first term \( p_i(1 - p_i)/n_i \) is a variation within centers and the next term \( \tau^2_p \) is a variation between centers (heterogeneity variance across centers).

Let us give some theoretical motivations here and consider the two-stage model in only the control arm, for the time being. We now assume that the true value of success rate \( p_i \) varies from center to center. Thus, at the first stage of the model we assume that
\[ \hat{p}_i = p_i + e_i, \]
for \( i = 1, 2, \ldots, k, \)
where \( \hat{p}_i = X_i/n_i \) is the observed probability of success in the control arm and \( e_i \) is a random error. Conditionally upon the random effect \( p_i \), \( E(\hat{p}_i | p_i) = p_i \) and \( V(\hat{p}_i | p_i) = p_i(1 - p_i)/n_i. \)
At the second stage, \( p_i \) is assumed to be distributed with \( E(p_i) = p \) and \( V(p_i) = \tau^2_p. \)

The unconditional (marginal) mean and variance is given as
\[
E(\hat{p}_i) = E(E(\hat{p}_i | p_i)) = E(p_i) = p
\]
\[
V(\hat{p}_i) = V(E(\hat{p}_i | p_i)) + E(V(\hat{p}_i | p_i)) = V(p_i) + E \left( \frac{p_i(1 - p_i)}{n_i} \right)
\]
\[
= \tau^2_p + E \left( \frac{p_i}{n_i} \right) - E \left( \frac{p_i^2}{n_i} \right) = \tau^2_p + p_i - \left( \frac{\tau^2_p + p^2}{n_i} \right)
\]
\[
\frac{1}{n_i} V(X_i) = \tau^2_p \left( \frac{n_i - 1}{n_i} \right) + p(1 - p)
\]
Consequently,
\[
\tau^2_p = \frac{\text{var}(X_i)}{n_i(n_i - 1)} - \frac{p(1 - p)}{(n_i - 1)}
\]
Replacing again $E(X_i) = n_i \hat{p}$ by $\bar{X} = n_i \hat{p}$ and $\text{var}(X_i)$ by

$$S^2 = \frac{\sum_{i=1}^{k} (X_i - n_i \hat{p})^2}{k - 1}$$

where $\hat{p} = \sum_{i=1}^{k} X_i / \sum_{i=1}^{k} n_i$, we are led to

$$\hat{\tau}_p^2 = \frac{1}{k - 1} \sum_{i=1}^{k} \frac{(X_i - n_i \hat{p})^2}{n_i(n_i - 1)} - \frac{1}{k} \hat{p}(1 - \hat{p}) \sum_{i=1}^{k} \frac{1}{n_i - 1}.$$  \hspace{1cm} (4)

In fact, this simple, non-iterative estimator is suggested in Böhning and Viwatwongkasem (1998) and in Böhning (1999, Chapter 6). Similarly, the estimator $\hat{\tau}_q^2$ is obtained for the other treatment arm. Note that $\hat{\tau}_p^2$ and $\hat{\tau}_q^2$ in (3) have the nice feature of leading to positive and well-defined weights, unless both of the heterogeneity variances $\hat{\tau}_p^2$ and $\hat{\tau}_q^2$ are estimated to be 0. Finally, the new estimator can be obtained as

$$\hat{\delta}_{\text{New}} = \frac{\sum_{i=1}^{k} \hat{w}_i \hat{\delta}_i}{\sum_{i=1}^{k} \hat{w}_i}$$

with

$$\hat{w}_i^{-1} = X_i(n_i - X_i)/n_i^3 + \hat{\tau}_p^2 + Y_i(m_i - Y_i)/m_i^3 + \hat{\tau}_q^2,$$

where $\hat{\tau}_p^2$ and $\hat{\tau}_q^2$ are defined according to (4).

4. A simulation study

We propose a simulation study to compare these six estimators following the design of Lipsitz et al. (1998). Under homogeneity of equal risk difference across centers and allowing for baseline-risk heterogeneity in each treatment arm, the baseline-heterogeneity risks $p_1, p_2, \ldots, p_k$ are generated from a uniform distribution on 0–0.8, and $q_i = p_i + \delta = p_i + 0.1$. To mimic variation in the sample sizes, $n_i$ and $m_i$ are generated from a Poisson distribution with parameter $n$ and $m$ for $i = 1, \ldots, k$. Binomial variates $X_i$ with parameters $n_i$, $p_i$ and binomial variates $Y_i$ with parameters $m_i$, $q_i$ are generated for each center $i$, $i = 1, \ldots, k$. All proposed estimates are then computed. Ten thousand replications are considered. From these replicates bias, variance, and mean-square error are computed. The sample sizes in the centers are designed to vary as $n = 4, 8, 16, 32, 64$ and the number of centers equal as $k = 4, 8, 16, 32, 64$. A total of 25 constellations is studied. Consequently, the results should be analyzed with respect to different asymptotic behaviors, namely if

- the sample size of each center $n_i$ becomes large
- the number of centers $k$ becomes large.
5. Results

Under homogeneity of the risk difference across centers \( \delta_i = \delta = 0.1 \), six risk difference estimates \( \hat{\delta}_{\text{COC}}, \hat{\delta}_{\text{MH}}, \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{New}}, \hat{\delta}_{\text{Pool}}, \text{ and } \hat{\delta}_{\text{ROTH}} \) of \( \delta \) are compared the efficiency. The results are presented with respect to bias, variance, mean square error, and consistency.

5.1. Bias, variance, and consistency

Unbiasedness \( (E \hat{\delta}_{n,k}(X) = \delta \text{ for all } \delta) \) is a criterion of a good estimator. The bias of \( \hat{\delta}_{n,k} \) is \( b_{n,k}(\delta) = E \hat{\delta}_{n,k}(X) - \delta \).

From Table 1, it is reconfirmed that the estimators \( \hat{\delta}_{\text{COC}} \) and \( \hat{\delta}_{\text{MH}} \) are unbiased. The estimator \( \hat{\delta}_{\text{Pool}} \) is not unbiased, but its bias goes to 0 as \( n \) increases and \( k \) is fixed (\( \lim_{n \to \infty} b_{n,k}(\delta) = 0 \) for all \( \delta \)); hence, \( \hat{\delta}_{\text{Pool}} \) is an asymptotically unbiased with respect to \( n \) with fixed \( k \). Similarly, \( \hat{\delta}_{\text{Pool}} \) is an asymptotically unbiased with respect to \( k \) while fixing \( n \) (\( \lim_{k \to \infty} b_{n,k}(\delta) = 0 \) for all \( \delta \)).

The three estimators \( \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{New}}, \text{ and } \hat{\delta}_{\text{ROTH}} \) have considerable bias with respect to \( n \) while fixing \( k \). In particular, both of \( \hat{\delta}_{\text{LS}} \) and \( \hat{\delta}_{\text{New}} \) have considerable bias for small sample size \( (n \leq 8) \) as \( k \) is fixed. The Rothman–Boice estimate \( \hat{\delta}_{\text{ROTH}} \) is considerably biased for small to moderate sample size \( (n \leq 16) \), regardless of \( k \). However, for large sample sizes, all three estimates are asymptotic unbiased with respect to \( n \), regardless of \( k \).

The bias of \( \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{New}}, \text{ and } \hat{\delta}_{\text{ROTH}} \) is persistent even if \( k \) becomes large while calculating at each level of the mean of \( n_i \) \( (E(n_i) = n) \). This indicates that estimators \( \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{New}}, \text{ and } \hat{\delta}_{\text{ROTH}} \) are not asymptotically unbiased with respect to \( k \) while fixing \( n \) \( (\text{the mean of } n_i) \).

In summary, \( \hat{\delta}_{\text{COC}} \), \( \hat{\delta}_{\text{MH}}, \text{ and } \hat{\delta}_{\text{Pool}} \) are unbiased or asymptotically unbiased with respect to \( n \) and \( k \). Moreover, \( \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{New}}, \text{ and } \hat{\delta}_{\text{ROTH}} \) are asymptotically unbiased with respect to \( n \), but not to \( k \).

Table 2 exhibits some points related to the variance of the estimators. With fixing \( k \), the variance of each estimate decreases and goes to 0 with respect to \( n \). Likewise, the variance of each estimate decreases and converges to 0 with respect to \( k \) while fixing \( n \) \( (\text{the mean of } n_i) \). Thus, the variance of all six estimates converges to 0 with respect to \( n \) and \( k \).

For small sample size \( (n \leq 8) \), regardless of \( k \), the variance of \( \hat{\delta}_{\text{COC}} \) and \( \hat{\delta}_{\text{MH}} \) is smaller than that of \( \hat{\delta}_{\text{LS}} \) and \( \hat{\delta}_{\text{ROTH}} \). In contrast, for large sample size \( (n \geq 32) \), regardless of \( k \), the variance of \( \hat{\delta}_{\text{LS}} \) and \( \hat{\delta}_{\text{ROTH}} \) is smaller than that of \( \hat{\delta}_{\text{COC}} \) and \( \hat{\delta}_{\text{MH}} \). Thus, the estimates \( (\hat{\delta}_{\text{LS}} \text{ and } \hat{\delta}_{\text{ROTH}}) \) involve weights calculated by the inverses of variances have more precision than the estimates in the set of the weights calculating by the center-specific sample sizes \( (\hat{\delta}_{\text{COC}} \text{ and } \hat{\delta}_{\text{MH}}) \). This makes sense, since with large sample sizes the variability in estimating the variances in the centers decreases.
The variance of \( \hat{\delta}_{\text{New}} \) lies in between the variance of \( \{ \hat{\delta}_{\text{COC}}, \hat{\delta}_{\text{MH}} \} \) and the variance of \( \{ \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{ROTH}} \} \). This indicates that this estimator is not an optimal strategy, but widely extends to cover heterogeneity cases.

The variance of \( \hat{\delta}_{\text{Pool}} \) is very close to that of \( \hat{\delta}_{\text{COC}} \); we notice that \( \hat{\delta}_{\text{Pool}} \) and \( \hat{\delta}_{\text{COC}} \) behave very similar under the null hypothesis of equal risk difference over centers.

Considering for the consistency, we know that \( \hat{\delta}_{\text{MH}}, \hat{\delta}_{\text{COC}}, \) and \( \hat{\delta}_{\text{Pool}} \) are unbiased or asymptotically unbiased with respect to both \( n \) and \( k \). Moreover, their variances converge to 0 with the increase of \( n \) and \( k \). Therefore, \( \hat{\delta}_{\text{MH}}, \hat{\delta}_{\text{COC}}, \) and \( \hat{\delta}_{\text{Pool}} \) are consistent with respect to both \( n \) and \( k \).

The three estimators \( \{ \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{New}}, \hat{\delta}_{\text{ROTH}} \} \) are asymptotically unbiased with respect to \( n \) conditional on fixing \( k \). Moreover, their variances converge to 0 with the increase of \( n \) at each level of \( k \). Consequently, \( \hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{New}}, \) and \( \hat{\delta}_{\text{ROTH}} \) are consistent with respect.
Table 2
Variance as a function of the number of centers ($k$)

<table>
<thead>
<tr>
<th>Number of centers, $k$</th>
<th>Sample size, $n$</th>
<th>Variance of $\delta_{\text{COC}}$</th>
<th>Variance of $\delta_{\text{MH}}$</th>
<th>Variance of $\delta_{\text{LS}}$</th>
<th>Variance of $\delta_{\text{Ne}}$</th>
<th>Variance of $\delta_{\text{Pool}}$</th>
<th>Variance of $\delta_{\text{ROT}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>0.02595621</td>
<td>0.02754803</td>
<td>0.04070091</td>
<td>0.03364240</td>
<td>0.02607480</td>
<td>0.03308874</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.01286737</td>
<td>0.01336687</td>
<td>0.01488179</td>
<td>0.01438822</td>
<td>0.01289915</td>
<td>0.01257657</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00631677</td>
<td>0.00644713</td>
<td>0.00646271</td>
<td>0.00643434</td>
<td>0.00632455</td>
<td>0.00575114</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00306191</td>
<td>0.00308907</td>
<td>0.00297368</td>
<td>0.00308482</td>
<td>0.00306356</td>
<td>0.00281824</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00148991</td>
<td>0.00150201</td>
<td>0.00146633</td>
<td>0.00148620</td>
<td>0.00149010</td>
<td>0.00137649</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0.01298790</td>
<td>0.01394162</td>
<td>0.02175236</td>
<td>0.01728864</td>
<td>0.01300453</td>
<td>0.01550428</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00658504</td>
<td>0.00689022</td>
<td>0.00901935</td>
<td>0.00736173</td>
<td>0.00658988</td>
<td>0.00613822</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00313804</td>
<td>0.00319727</td>
<td>0.00319005</td>
<td>0.00322249</td>
<td>0.00313888</td>
<td>0.00319311</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00076573</td>
<td>0.00077069</td>
<td>0.00070888</td>
<td>0.00076548</td>
<td>0.00075676</td>
<td>0.00069238</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0.00655234</td>
<td>0.00705191</td>
<td>0.01172933</td>
<td>0.00915763</td>
<td>0.00655244</td>
<td>0.00756487</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00326516</td>
<td>0.00342838</td>
<td>0.00448526</td>
<td>0.00360873</td>
<td>0.00326426</td>
<td>0.00292205</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00156118</td>
<td>0.00159795</td>
<td>0.00158906</td>
<td>0.00159323</td>
<td>0.00156095</td>
<td>0.00134433</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00076736</td>
<td>0.00078208</td>
<td>0.00074546</td>
<td>0.00076664</td>
<td>0.00076731</td>
<td>0.00069679</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00037434</td>
<td>0.00037826</td>
<td>0.00034787</td>
<td>0.00037330</td>
<td>0.00037436</td>
<td>0.00033866</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>0.00322462</td>
<td>0.00353130</td>
<td>0.00578904</td>
<td>0.00416012</td>
<td>0.00322412</td>
<td>0.00365043</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00162764</td>
<td>0.00171147</td>
<td>0.00231745</td>
<td>0.00179549</td>
<td>0.00162767</td>
<td>0.00144322</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00076352</td>
<td>0.00078655</td>
<td>0.00079026</td>
<td>0.00077520</td>
<td>0.00076354</td>
<td>0.00065526</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00037662</td>
<td>0.00038231</td>
<td>0.00037756</td>
<td>0.00037856</td>
<td>0.00037662</td>
<td>0.00035014</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00019271</td>
<td>0.00019451</td>
<td>0.00018308</td>
<td>0.00019248</td>
<td>0.00019271</td>
<td>0.00017826</td>
</tr>
<tr>
<td>64</td>
<td>4</td>
<td>0.00161849</td>
<td>0.00175951</td>
<td>0.00295342</td>
<td>0.00209807</td>
<td>0.00161824</td>
<td>0.00179685</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00080841</td>
<td>0.00085013</td>
<td>0.00114075</td>
<td>0.00089516</td>
<td>0.00080850</td>
<td>0.00071344</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00038389</td>
<td>0.00039581</td>
<td>0.00039864</td>
<td>0.00039095</td>
<td>0.00038383</td>
<td>0.00032327</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00019346</td>
<td>0.00019736</td>
<td>0.00019122</td>
<td>0.00019306</td>
<td>0.00019344</td>
<td>0.00017685</td>
</tr>
</tbody>
</table>
|                       | 64              | 0.00009494      | 0.00009561      | 0.00009131      | 0.00009497      | 0.00009494      | 0.00008904      

to $n$ by fixing $k$; see also Lloyd (1999, p. 50). He said that the weighted least squares estimator $\hat{\delta}_{\text{LS}}$ was consistent as the $n$ increased for fixed $k$.

Whereas, $\hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{Ne}},$ and $\hat{\delta}_{\text{ROT}}$ are not asymptotically unbiased with respect to $k$, while calculating at each level of $n$ (the mean of $n_i$); however, their variances converge to 0 with the increase of $k$ at each level of $n$. Hence, they are inconsistent with respect to $k$ by fixing $n$ (Fig 1).

5.2. Mean-square errors

The mean-square error (MSE) of an estimate is equal to the variance of the estimate plus the square of its bias, so it involves both its precision and its validity. Table 3 shows some points of MSE. With fixing $k$, MSE of every estimate decreases and goes to 0 with respect to $n$ (the mean of $n_i$). As the estimators of $\hat{\delta}_{\text{LS}}, \hat{\delta}_{\text{Ne}},$ and $\hat{\delta}_{\text{ROT}}$ are not asymptotically unbiased with respect to $k$ while fixing $n$, we know that the MSE
Fig. 1. Bias and Mean square Error in dependence of sample size and number of centers.
Table 3
Mean-square error as a function of the number of centers ($k$)

<table>
<thead>
<tr>
<th>Number of centers, $k$</th>
<th>Sample size, $n$</th>
<th>MSE of $\delta_{COC}$</th>
<th>MSE of $\delta_{MH}$</th>
<th>MSE of $\delta_{LS}$</th>
<th>MSE of $\delta_{New}$</th>
<th>MSE of $\delta_{Pool}$</th>
<th>MSE of $\delta_{ROTH}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>0.02595621</td>
<td>0.02754803</td>
<td>0.04085584</td>
<td>0.03666032</td>
<td>0.02607487</td>
<td>0.03319554</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.01286737</td>
<td>0.01336687</td>
<td>0.01690435</td>
<td>0.01439246</td>
<td>0.01289915</td>
<td>0.01257673</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00631694</td>
<td>0.00644720</td>
<td>0.00646389</td>
<td>0.00643598</td>
<td>0.00632472</td>
<td>0.00576068</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00306197</td>
<td>0.00308915</td>
<td>0.00297368</td>
<td>0.00308500</td>
<td>0.00306361</td>
<td>0.00282268</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00148993</td>
<td>0.00150201</td>
<td>0.00140664</td>
<td>0.00148627</td>
<td>0.00149011</td>
<td>0.00137754</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0.01298827</td>
<td>0.01394168</td>
<td>0.02189678</td>
<td>0.01729478</td>
<td>0.01300489</td>
<td>0.01556898</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00658535</td>
<td>0.00689063</td>
<td>0.00904998</td>
<td>0.00736804</td>
<td>0.00659020</td>
<td>0.00614438</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00313828</td>
<td>0.00319753</td>
<td>0.00319021</td>
<td>0.00322388</td>
<td>0.00313913</td>
<td>0.00275808</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00154253</td>
<td>0.00156442</td>
<td>0.00147857</td>
<td>0.00154712</td>
<td>0.00154252</td>
<td>0.00140736</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00076589</td>
<td>0.00077086</td>
<td>0.00070986</td>
<td>0.00076769</td>
<td>0.00076592</td>
<td>0.00069760</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0.00655432</td>
<td>0.00705379</td>
<td>0.01188269</td>
<td>0.009186106</td>
<td>0.00655440</td>
<td>0.00760276</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00326517</td>
<td>0.00342838</td>
<td>0.00450334</td>
<td>0.00361006</td>
<td>0.00326429</td>
<td>0.00293487</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00156122</td>
<td>0.00159798</td>
<td>0.00159148</td>
<td>0.00159334</td>
<td>0.00156098</td>
<td>0.00138414</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00076737</td>
<td>0.00078208</td>
<td>0.00074948</td>
<td>0.00076669</td>
<td>0.00076732</td>
<td>0.00071889</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00037436</td>
<td>0.00037828</td>
<td>0.00034871</td>
<td>0.00037334</td>
<td>0.00037439</td>
<td>0.00034396</td>
</tr>
<tr>
<td>32</td>
<td>4</td>
<td>0.00322472</td>
<td>0.00353155</td>
<td>0.00600490</td>
<td>0.00416763</td>
<td>0.00322422</td>
<td>0.00370833</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00162785</td>
<td>0.00171158</td>
<td>0.00234146</td>
<td>0.00179967</td>
<td>0.00162788</td>
<td>0.00145645</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00076355</td>
<td>0.00078658</td>
<td>0.00079537</td>
<td>0.00075727</td>
<td>0.00076358</td>
<td>0.00070383</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00037671</td>
<td>0.00038239</td>
<td>0.00038490</td>
<td>0.00037859</td>
<td>0.00037671</td>
<td>0.00038030</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00019272</td>
<td>0.00019452</td>
<td>0.00019455</td>
<td>0.00019249</td>
<td>0.00019273</td>
<td>0.00018528</td>
</tr>
<tr>
<td>64</td>
<td>4</td>
<td>0.00161864</td>
<td>0.00175955</td>
<td>0.00313817</td>
<td>0.00210092</td>
<td>0.00161838</td>
<td>0.00183945</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.00080841</td>
<td>0.00085013</td>
<td>0.00115465</td>
<td>0.00089676</td>
<td>0.00080850</td>
<td>0.00073156</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.00038391</td>
<td>0.00039582</td>
<td>0.00040555</td>
<td>0.00039105</td>
<td>0.00038385</td>
<td>0.00037451</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.00019348</td>
<td>0.00019737</td>
<td>0.00019846</td>
<td>0.00019313</td>
<td>0.00019346</td>
<td>0.00020707</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.00009494</td>
<td>0.00009561</td>
<td>0.00009308</td>
<td>0.00009497</td>
<td>0.00009494</td>
<td>0.00009684</td>
</tr>
</tbody>
</table>

of $\hat{\delta}_{LS}$, $\hat{\delta}_{New}$, and $\hat{\delta}_{ROTH}$ cannot go to 0 (their MSE should converge to certain small values) with the increase of $k$ on the condition of fixing $n$.

For small sample size ($n \leq 8$), regardless of $k$, MSE of $\hat{\delta}_{COC}$ and $\hat{\delta}_{MH}$ is smaller than that of $\hat{\delta}_{LS}$ and $\hat{\delta}_{ROTH}$. MSE of $\hat{\delta}_{Pool}$ is very close to that of $\hat{\delta}_{COC}$, since $\hat{\delta}_{Pool}$ and $\hat{\delta}_{COC}$ behave very similar under the null hypothesis of equal risk difference over centers. Thus, for small sample size we recommend the use of $\hat{\delta}_{COC}$, $\hat{\delta}_{MH}$, and $\hat{\delta}_{Pool}$.

In contrast, for large sample size ($n \geq 32$), regardless of $k$, MSE of $\hat{\delta}_{LS}$ and $\hat{\delta}_{ROTH}$ is smaller than that of $\hat{\delta}_{COC}$ and $\hat{\delta}_{MH}$. Therefore, for large sample size we suggest the use of $\hat{\delta}_{LS}$ and $\hat{\delta}_{ROTH}$.

The MSE of $\hat{\delta}_{New}$ is in between the MSE of $\{\hat{\delta}_{COC}, \hat{\delta}_{MH}\}$ and the MSE of $\{\hat{\delta}_{LS}, \hat{\delta}_{ROTH}\}$. If we don’t know which estimate is good to be selected from two popular sets, this new estimate is a considerable alternative. We also recommend the use of $\hat{\delta}_{New}$ under strong baseline-risk heterogeneity across several centers.
Finally, we note that the MSE of $\hat{\delta}_{\text{COC}}$ is uniformly smaller than that of $\hat{\delta}_{\text{MH}}$, regardless of $n$ and $k$.

6. Conclusions and discussions

The new proposed estimator, adjusting heterogeneity by the means of a random effects model, is an attractive compromise when choosing between the estimators of the set of the center-specific sample size weights and the estimators of the set of the inverse-variance weights. This new estimator is not of optimal weight, but it can widely extend to cover heterogeneity cases, and it is more appropriate when the sample size is greater or equal to 16 regardless of the number of centers. To obtain an overall risk difference from multi-center clinical trials where there is some heterogeneity among several centers, the idea of two-stage random effect models is initiated by DerSimonian and Laird (1986), extended by Böhning (1999, Chapter 6) and presented in a mutual project of Böhning and Viwatwongkasem (1998).

For large sample sizes ($n \geq 32$), the weighted least square and the Rothman–Boice estimators whose weights are the inverses of variances are the best choice. Their mean square errors are smaller than that of other estimates when the sample size is large. This result is related to Mehrotra and Railkar (2000). They stated that the inverse-variance weights yield more power than the center-specific sample size weights. In fact, the work of Mehrotra and Railkar (2000) was mainly interested in the finding a minimum risk test for the null hypothesis of homogeneity of risk difference.

For small sample size ($n \leq 8$), the Cochran and the Mantel–Haenszel estimators are the most efficient because of their smallest mean square errors. The pooling estimate behaves very close to Cochran’s estimator under homogeneity of equal risk difference over centers.

We recommend to use Cochran, Mantel–Haenszel, and the pooling estimators when $n \leq 8$, to use Lipsitz et al. and Rothman–Boice estimators when $n \geq 32$, and to use the new estimator when heterogeneity occurs. However, we also see that all estimates are very close, similar, and appropriate when the sample size is greater or equal to 16 regardless of the number of centers.

Acknowledgements

We are grateful for the support provided by NRCT (The National Research Council of Thailand) grant E (GE) 5/43 in cooperation with the DFG (German Research Foundation) grant DFG Bo 865/6-1/6-2. This work would have not been possible without the help of our co-worker Dankmar Böhning. We are also grateful to the editor, associated editor, and the referees for their helpful comments.

References