Open Journal of Statistics, **2012, 2, 48-59**
doi:10.4236/ojs.2012.21006 Published Online January 2012 (http://www.SciRP.org/journal/ojs)

Minimum MSE Weights of Adjusted Summary Estimator of Risk Difference in Multi-Center Studies

Chukiat Viwatwongkasem[1*], Jirawan Jitthavech[2], Dankmar Böhning[3], Vichit Lorchirachoonkul[2]
[1]Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok, Thailand
[2]School of Applied Statistics, National Institute of Development Administration, Bangkok, Thailand
[3]Applied Statistics, School of Biological Sciences, University of Reading, Reading, UK
Email: *phcvw@mahidol.ac.th

Received October 14, 2011; received November 18, 2011; accepted November 30, 2011

ABSTRACT

The simple adjusted estimator of risk difference in each center is easy constructed by adding a value *c* on the number of successes and on the number of failures in each arm of the proportion estimator. Assessing a treatment effect in multi-center studies, we propose minimum MSE (mean square error) weights of an adjusted summary estimate of risk difference under the assumption of a constant of common risk difference over all centers. To evaluate the performance of the proposed weights, we compare not only in terms of estimation based on bias, variance, and MSE with two other conventional weights, such as the Cochran-Mantel-Haenszel weights and the inverse variance (weighted least square) weights, but also we compare the potential tests based on the type I error probability and the power of test in a variety of situations. The results illustrate that the proposed weights in terms of point estimation and hypothesis testing perform well and should be recommended to use as an alternative choice. Finally, two applications are illustrated for the practical use.

Keywords: Minimum MSE Weights; Optimal Weights; Cochran-Mantel-Haenszel Weights; Inverse Variance Weights; Multi-Center Studies; Risk Difference

1. Introduction

It is widely known that the conventional proportion estimator, $\hat{p} = X/n$, is a maximum likelihood estimator (MLE) and an uniformly minimum variance unbiased estimator (UMVUE) for the binomial parameter p where the binomial random variable X is the number of successes out of the number of patients n. However, Agresti and Coull [1], Agresti and Caffo [2], Ghosh [3], and Newcombe [4,5] highlighted the point that $\hat{p}$ might not be a good choice for p when the assumption of $n\hat{p} \geq 5$ and $n(1-\hat{p}) \geq 5$ was violated; this violation often occurs when the sample size n is small, or the estimated probability $\hat{p}$ is close to 0 or 1 (close to the boundaries of parameter space), leading to the problem of the zero estimate of the variance of $\hat{p}$. The estimated variance of $\hat{p}$, provided by $\hat{V}(\hat{p}) = \hat{p}(1-\hat{p})/n$, is *zero* in the occurrence of any case: $X = 0$ or $X = n$. Böhning and Viwatwongkasem [6] proposed the simple adjusted proportion estimator by adding a value c on the number of successes and the number of failures; consequently, $\hat{p}_c = (X + c)/(n + 2c)$ is their proposed estimate of p with the non-zero variance estimate $\hat{V}(\hat{p}_c) = n\,\hat{p}_c(1-\hat{p}_c)/(n+2c)^2$. They concluded that the estimator $(X+1)/(n+2)$ minimizes the Bayes risk (the average MSE of $\hat{p}_c$) in the class of all estimators of the form $(X+c)/(n+2c)$ with respect to uniform prior on [0,1] and Euclidean loss function; furthermore, the estimator $(X+1)/(n+2)$ has smaller MSE than X/n in the approximate interval $[0.15, 0.85]$ of p. For another argumentation in the Bayesian approach, Casella and Berger [7] showed that $(X+\alpha)/(n+\alpha+\beta)$ is a Bayes estimator of p under the conditional binomial sampling $X \mid p \sim binomial(n, p)$ and the prior beta distribution $p \sim beta(\alpha, \beta)$. Note that in case of $\alpha = \beta = 1$ the beta distribution has a special case as the uniform distribution over [0,1]. Consequently, the estimator $(X+c)/(n+2c)$ derived from the Bayesian approach and the Bayes risk approach under the above mentioned criteria provides the same result at $c = 1$.

With the idea of $\hat{p}_c = (X+c)/(n+2c)$, the extension leads to $\hat{\theta}_c = \hat{p}_{c1} - \hat{p}_{c2}$, the adjusted risk difference estimator between two independent binomial proportions, for estimating a common risk difference θ where $\hat{p}_{c1} = (X_1 + c_1)/(n_1 + 2c_1)$ and $\hat{p}_{c2} = (X_2 + c_2)/(n_2 + 2c_2)$ are proportion estimators for treatment and control arms. In a multi-center study of size k, the parameter of in-

*Corresponding author.

terest is also a common risk difference θ that is assumed to be a constant across centers. We concern about a combination of several adjusted risk difference estimators $\hat{\theta}_{cj} = \hat{p}_{c1j} - \hat{p}_{c2j}$ from the j^{th} center $(j = 1,2,\cdots,k)$ into the adjusted summary estimator of risk difference of the form $\hat{\theta}_{cw} = \sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}$ where f_{cj} are the weights subject to the condition that $\sum_{j=1}^{k} f_{cj} = 1$. In this study, we would propose the optimal weights f_{cj} as an alternative choice based on minimizing the MSE of $\hat{\theta}_{cw}$ in Section 2, then state the well-known candidates such as the Cochran-Mantel-Haenszel (CMH) weights and the inverse variance (INV) weights in Section 3. A simulation plan for comparing the performance among weights in terms of estimation and hypothesis testing is presented in Section 4. The results of the comparison among the potential estimators based on bias, variance, and MSE and also the evaluations among tests related the mentioned weights through the type I error probability and the power criteria lie on Section 5. Some numerical examples are applied in Section 6. Finally, conclusion and discussion are presented in Section 7.

2. Deriving Minimum MSE Weights of Adjusted Summary Estimator

Under the assumption of a constant of common risk difference θ across k centers, we combine several adjusted risk difference estimators $\hat{\theta}_{cj} = \hat{p}_{c1j} - \hat{p}_{c2j}$ in which $\hat{p}_{c1j} = \left(X_{1j} + c_1\right)/\left(n_{1j} + 2c_1\right)$ and $\hat{p}_{c2j} = \left(X_{2j} + c_2\right)/\left(n_{2j} + 2c_2\right)$ from the j^{th} center $(j = 1,2,\cdots,k)$ arrive at an adjusted summary estimator of risk difference of the form $\hat{\theta}_{cw} = \sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}$ where f_{cj} are non-random weights subject to the constraint that $\sum_{j=1}^{k} f_{cj} = 1$. Please observe that for a single center $(k = 1)$ the adjusted summary estimator $\hat{\theta}_{cw} = \sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}$ subject to $\sum_{j=1}^{k} f_{cj} = 1$ is a shrinkage estimator of a simple adjusted estimator $\hat{\theta}_c = \hat{p}_{c1} - \hat{p}_{c2}$. Our minimum MSE weights f_{cj} of the adjusted summary estimator $\hat{\theta}_{cw}$ were derived by following Lagrange's method [8] under the assumption of a constant of common risk difference over all centers with the pooling point estimator to estimate θ. Lui and Chang [9] proposed the optimal weights proportional to the reciprocal of the variance with the Mantel-Haenszel point estimator under the assumption of noncompliance. It was observed that both of optimal weights provided the different formulae because of different assumptions even though they were derived from the same method of Lagrange. Now, we wish to present the proposed weights minimizing the MSE of $\hat{\theta}_{cw}$ as follows:

$$Q = MSE\left(\hat{\theta}_{cw}\right) = E\left(\hat{\theta}_{cw} - \theta\right)^2 = E\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj} - \theta\right)^2$$

To obtain the minimum Q subject to a constraint $\sum_{j=1}^{k} f_{cj} = 1$, we form the auxiliary function ϕ to seek f_{cj} that minimize

$$\phi = E\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj} - \theta\right)^2 + \lambda\left(\sum_{j=1}^{k} f_{cj} - 1\right)$$

where λ is a Lagrange multiplier. The weights f_{cj} and λ are derived by solving the following equations simultaneously: $\frac{\partial\phi}{\partial\lambda} = 0$, $\frac{\partial\phi}{\partial f_{cj}} = 0$, $j = 1,2,\cdots,k$. The details are presented in Appendix. The result of the weighted estimate for the j^{th} center yields

$$\hat{f}_{cj} = \left(\frac{\hat{V}_j^{-1}\left(1 + \hat{\tau}_j\hat{\theta}_{pool}\right)}{\hat{a}}\right) - \left(\frac{\hat{V}_j^{-1}\hat{\tau}_j}{\hat{a} + \sum_{m=1}^{k}\hat{\tau}_m\hat{E}_m\hat{V}_m^{-1}}\right)\left(\frac{\sum_{m=1}^{k}\hat{V}_m^{-1}\hat{E}_m\left(1 + \hat{\tau}_m\hat{\theta}_{pool}\right)}{\hat{a}}\right)$$

where $\hat{E}_j = \frac{n_{1j}\hat{p}_{c1j} + c_1}{n_{1j} + 2c_1} - \frac{n_{2j}\hat{p}_{c2j} + c_2}{n_{2j} + 2c_2}$

$$\hat{V}_j = \frac{n_{1j}\hat{p}_{c1j}\left(1 - \hat{p}_{c1j}\right)}{\left(n_{1j} + 2c_1\right)^2} + \frac{n_{2j}\hat{p}_{c2j}\left(1 - \hat{p}_{c2j}\right)}{\left(n_{2j} + 2c_2\right)^2}, \quad \hat{\tau}_j = \hat{a}\hat{E}_j - \hat{b}$$

$$\hat{a} = \sum_{j=1}^{k}\frac{1}{\hat{V}_j} = \sum_{j=1}^{k}\hat{V}_j^{-1}, \quad \hat{b} = \sum_{j=1}^{k}\frac{\hat{E}_j}{\hat{V}_j}, \quad \hat{\theta}_{pool} = \hat{p}_1 - \hat{p}_2$$

$$\hat{p}_1 = \frac{\sum_{j=1}^{k} n_{1j}\hat{p}_{1j}}{\sum_{j=1}^{k} n_{1j}} = \frac{\sum_{j=1}^{k} X_{1j}}{\sum_{j=1}^{k} n_{1j}},$$

$$\hat{p}_2 = \frac{\sum_{j=1}^{k} n_{2j}\hat{p}_{2j}}{\sum_{j=1}^{k} n_{2j}} = \frac{\sum_{j=1}^{k} X_{2j}}{\sum_{j=1}^{k} n_{2j}}$$

In the particular case of $c_1 = c_2 = 0$, our estimator $\hat{\theta}_{cw} = \sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}$ has a shrinkage estimator to be the popular inverse-variance weighted estimator. Under a common risk difference θ over all centers, the variance of $\hat{\theta}_{cw}$ in the case of non-random weights f_{cj} are obtained by

$$V\left(\hat{\theta}_{cw}\right) = \sum_{j=1}^{k} f_{cj}^2 V\left(\hat{\theta}_{cj}\right) = \sum_{j=1}^{k} f_{cj}^2\left(\frac{n_{1j}p_{1j}\left(1 - p_{1j}\right)}{\left(n_{1j} + 2c_1\right)^2} + \frac{n_{2j}p_{2j}\left(1 - p_{2j}\right)}{\left(n_{2j} + 2c_2\right)^2}\right)$$

Suppose that a normal approximation is reliable, the asymptotic distribution is

$$\frac{\hat{\theta}_{cw} - \theta}{\sqrt{\hat{V}\left(\hat{\theta}_{cw}\right)}} = \frac{\sum_{j=1}^{k}\hat{f}_{cj}\hat{\theta}_{cj} - \theta}{\sqrt{\sum_{j=1}^{k}\hat{f}_{cj}^2\hat{V}\left(\hat{\theta}_{cj}\right)}} \to N(0,1)$$

for testing $H_0:\theta=\theta_0$ we have the normal approximate test

$$Z_{cw}=\frac{\sum_{j=1}^{k}\hat{f}_{cj}\hat{\theta}_{cj}-\theta_0}{\sqrt{\sum_{j=1}^{k}\hat{f}_{cj}^2\hat{V}\left(\hat{\theta}_{cj}\middle|H_0\right)}}$$

We will reject H_0 at α level for two-sided test if $\left|Z_{cw}\right|>Z_{\alpha/2}$ where $Z_{\alpha/2}$ is the upper $100\left(\alpha/2^{th}\right)$ percentile of the standard normal distribution. Alternatively, H_0 is rejected when the p-value (p) is less than or equal to α $\left(p\le\alpha\right)$ where $p=2\left[1-\Phi\left(\left|Z_{cw}\right|\right)\right]$ and $\Phi\left(Z\right)$ is the standard cumulative normal distribution function.

3. Other Well-Known Weights

3.1. Cochran-Mantel-Haenszel (CMH) Weights

Cochran [10,11] proposed a weighted estimator of center-specific sample sizes for a common risk difference based on the unconditional binomial likelihood as

$$\hat{\theta}_{CMH}=\frac{\sum_{j=1}^{k}w_j\hat{\theta}_j}{\sum_{j=1}^{k}w_j}$$

where $w_j=\left(\frac{1}{n_{1j}}+\frac{1}{n_{2j}}\right)^{-1}=n_{1j}n_{2j}\big/\left(n_{1j}+n_{2j}\right)$ and $\hat{\theta}_j=\hat{p}_{1j}-\hat{p}_{2j}=\frac{X_{1j}}{n_{1j}}-\frac{X_{2j}}{n_{2j}}$. Cochran's weight w_j is widely used as a standard non-random weight derived by the harmonic means of the center-specific sample sizes. Note that $f_j=w_j\big/\sum_{j=1}^{k}w_j$ is also Cochran's weight subject to the condition that $\sum_{j=1}^{k}f_j=1$. A straightforward derivation illustrates that $\hat{\theta}_{CMH}$ is an unbiased estimate of θ and the variance of $\hat{\theta}_{CMH}$ is readily available as

$$V\left(\hat{\theta}_{CMH}\right)=\frac{\sum_{j=1}^{k}w_j^2V\left(\hat{\theta}_j\right)}{\left(\sum_{j=1}^{k}w_j\right)^2}$$

where $V\left(\hat{\theta}_j\right)=p_{1j}\left(1-p_{1j}\right)\big/n_{1j}+p_{2j}\left(1-p_{2j}\right)\big/n_{2j}$. Assuming that a normal approximation is reliable, the Cochran's Z-statistic for testing $H_0:\theta=\theta_0$ is provided as

$$Z_{CMH}=\frac{\left(\sum_{j=1}^{k}w_j\hat{\theta}_j\big/\sum_{j=1}^{k}w_j\right)-\theta_0}{\sqrt{\left(\sum_{j=1}^{k}w_j^2\hat{V}\left(\hat{\theta}_j\middle|H_0\right)\right)\Big/\left(\sum_{j=1}^{k}w_j\right)^2}}$$

where $\hat{V}\left(\hat{\theta}_j\middle|H_0\right)=\hat{p}_{1j}\left(1-\hat{p}_{1j}\right)\big/n_{1j}+\hat{p}_{2j}\left(1-\hat{p}_{2j}\right)\big/n_{2j}$. The rejection rule of H_0 follows the same as the previous standard normal test.

Alternatively, Mantel and Haenszel [12] suggested the test based on the conditional hypergeometric likelihood for a common odds ratio among the set of k tables under the null hypothesis of $H_0:OR=1$ $\left(\theta=0\right)$. With the null criterion, Mantel-Haenszel's weight stated by Sanchez-Meca and Marin-Martine [13] was equivalent to $w_j=n_{1j}n_{2j}\big/\left(n_{1j}+n_{2j}-1\right)$. Since the minor difference between the conditional Mantel-Haenszel weight and the unconditional Cochran weight is in the denominators, thus the two are often referred to interchangeably as the Cochran-Mantel-Haenszel weight. In this study, we use $w_j=n_{1j}n_{2j}\big/\left(n_{1j}+n_{2j}\right)$.

3.2. Inverse Variance (INV) or Weighted Least Square (WLS) Weights

Fleiss [14] and Lipsitz *et al.* [15] showed that the inverse-variance weighted (INV) estimator or the weighted-least-square (WLS) estimator for θ was in the summary estimator of the weighted mean (linear, unbiased estimator) of the form

$$\hat{\theta}_{INV}=\sum_{j=1}^{k}w_j\hat{\theta}_j\Big/\sum_{j=1}^{k}w_j$$

where $\hat{\theta}_j=\hat{p}_{1j}-\hat{p}_{2j}=X_{1j}\big/n_{1j}-X_{2j}\big/n_{2j}$ and w_j defined by the reciprocal of the variance as

$$w_j=\frac{1}{V\left(\hat{\theta}_j\right)}=\left(\frac{p_{1j}\left(1-p_{1j}\right)}{n_{1j}}+\frac{p_{2j}\left(1-p_{2j}\right)}{n_{2j}}\right)^{-1}$$

The non-random and non-negative weights w_j yield the minimum variance of the summary estimator $\hat{\theta}_{INV}$ for estimating θ. The variance of $\hat{\theta}_{INV}$ is just given by

$$V\left(\hat{\theta}_{INV}\right)=\frac{\sum_{j=1}^{k}w_j^2V\left(\hat{\theta}_j\right)}{\left(\sum_{j=1}^{k}w_j\right)^2}=\frac{\sum_{j=1}^{k}w_j^2\left[1\big/w_j\right]}{\left(\sum_{j=1}^{k}w_j\right)^2}=\frac{1}{\sum_{j=1}^{k}w_j}$$

However, the weights w_j cannot be used in practice since p_{1j} and p_{2j} are unknown. Therefore, it has become common practice to replace them by their sample estimators. It yields

$$\hat{w}_j^{-1}=\frac{\hat{p}_{1j}\left(1-\hat{p}_{1j}\right)}{n_{1j}}+\frac{\hat{p}_{2j}\left(1-\hat{p}_{2j}\right)}{n_{2j}}$$

This weight was suggested in several textbooks of epidemiology such as Kleinbaum *et al.* [16] or in textbooks of meta-analysis such as Petitt [17]. We assume that a normal approximation is reliable; the inverse-variance weighted test statistic for testing $H_0:\theta=\theta_0$ is

$$Z_{INV}=\frac{\left(\sum_{j=1}^{k}\hat{w}_j\hat{\theta}_j\big/\sum_{j=1}^{k}\hat{w}_j\right)-\theta_0}{\sqrt{\left(1\big/\sum_{j=1}^{k}\hat{w}_j\right)}}$$

where $\hat{w}_j^{-1}=\hat{V}\left(\hat{\theta}_j\middle|H_0\right)$. Also, the rule of H_0 rejection follows the same as the above standard normal test.

4. Monte Carlo Simulation

We perform simulations for estimating a common risk difference θ and testing the null hypothesis $H_0:\theta=\theta_0$ in the similar plans as follows:

Parameters Setting: Let the common risk difference θ be some constants varying from 0 to 0.6, with incremental steps of 0.1. Baseline proportion risks p_{2j} $\{p_{21}, p_{22}, \cdots, p_{2k}\}$ in the control arm for the j^{th} center $(j=1,2,\cdots,k)$ are generated from a uniform distribution over $[0,\ 0.95-\theta]$. The correspondent proportion risks p_{1j} for the treatment arm in the j^{th} center are obtained as $p_{1j}=p_{2j}+\theta$. For example, if $\theta=0.2$, then $p_{2j}\sim U(0,0.75)$ and $p_{1j}=p_{2j}+\theta\sim U(0.2,0.95)$. The sample sizes n_{1j} and n_{2j} are varied as 4, 8, 16, 32, 100. The number of centers k takes values 1, 2, 4, 8, 16, 32.

Statistics: Binomial random variables X_{1j} and X_{2j} in treatment and control arms are generated with parameters (n_{1j}, p_{1j}) and (n_{2j}, p_{2j}) for each center j.

Estimation: All summary estimates of θ are computed in a variety of different weights. The procedure is replicated 5000 times. From these replicates, bias, variance, and MSE (mean square error) are computed in the conventional way.

Type I Error: From the above parameter setting, we assign $\theta=\theta_0$ under a null $H_0:\theta=\theta_0$, so all tests are computed. The replication is treated 5000 times. From these replicates, the number of the null hypothesis rejections is counted for the empirical type I error $\hat{\alpha}$.

$$\hat{\alpha}=\frac{\text{Number of rejections of } H_0 \text{ when } H_0 \text{ is true}}{\text{Number of replications (5000 times)}}$$

The evaluation for two-sided tests in terms of the type I probability is based on Cochran limits [18] as follow.

At $\alpha=0.01$, the $\hat{\alpha}$ value is between $[0.005, 0.015]$.

At $\alpha=0.05$, the $\hat{\alpha}$ value is between $[0.04, 0.06]$.

At $\alpha=0.10$, the $\hat{\alpha}$ value is between $[0.08, 0.12]$.

If the empirical type I error $\hat{\alpha}$ lies within those of Cochran limits, then the statistical test can control type I error.

Power of Tests: Before evaluating tests with their powers, all comparative tests should be calibrated to have the same type I error rate under H_0; then any test whose power hits the maximum under H_1 would be the best test. To achieve the alternative hypothesis, we assume the random effect model for θ_j as

$$\theta_j=0.1+U_m=0.1+m(2U-1)$$

where U_m as an effect of between centers is assigned to be uniform $[-m,m]$ for a given $m\in[0,0.1]$, or equivalently, U is an uniform variable over $[0,1]$. That is, $E(\theta_j)=0.1$ and $Var(\theta_j)=(2m)^2/12$. Also, we have $p_{1j}=\theta_j+p_{2j}$ where p_{2j} be uniform distribution over $[0.1, 0.8]$. Binomial random variables X_{1j} and X_{2j} are drawn with parameters (n_{1j}, p_{1j}), and (n_{2j}, p_{2j}), respectively. All proposed test statistics are then computed. The procedure is replicated 5,000 times. From these replicates, the empirical power $1-\hat{\beta}$ of test is counted.

$$1-\hat{\beta}=\frac{\text{Number of rejections of } H_0 \text{ when } H_1 \text{ is true}}{\text{Number of replications (5000 times)}}$$

5. Results

Since it is difficult to present all enormous results from the simulation study, we just have illustrated some instances. Nevertheless, the main results are concluded perfectly.

5.1. Results for Estimating Risk Differences

Table 1 presents some results according to point estimation of a common risk difference θ. However, we can draw conclusions in the following.

- The number of centers, k, can not change the order of the MSE of all weighted estimators, even though an increase in k can decrease the variance and the MSE of all estimators, leading to the increasing efficiency. Also, increasing n_{1j} and n_{2j} can decrease the variance of all estimators while fixing k. The unbalanced cases of n_{1j} and n_{2j} for center j have a rare effect on the order of the MSE of all estimates.
- For most popular situations used under $\theta=0$, $\theta=0.1$, $\theta=0.2$, and $\theta=0.3$, the proposed summary estimator $\hat{\theta}_{cw}$ adjusted by $c=c_1=c_2=1$ including adjusted by $c=2$ is the best choice with the smallest MSE. The estimator $\hat{\theta}_{cw}$ adjusted by $c=0.5$ and the inverse-variance (INV) weighted estimator $(c=0)$ are close together and are the second choice with smaller MSE. The Cochran-Mantel-Haenszel (CMH) weight performs the worst in this simulation setting. This finding is very useful in general situations of most clinical trials and most causal relations between a disease and a suspected risk factor since the risk difference is often less than 0.25 [19].
- For $\theta=0.4$, the proposed estimator $\hat{\theta}_{cw}$ adjusted by $c=1$ performs best; for $\theta=0.5$, the proposed estimator $\hat{\theta}_{cw}$ adjusted by $c=0.5$ performs best; for $\theta=0.6$, the INV weighted estimator ($c=0$) performs best.

5.2. Results for Studying Type I Error

Table 2 presents some results for controlling the empirical type I error. We can conclude the performance of several tests according to the empirical alpha under H_0 as follows.

Table 1. Mean, variance, MSE for estimating θ .

θ	k	n_{1j}	n_{2j}	Measure	CMH	INV ($c = 0$)	$c = 0.5$	$c = 1$	$c = 2$
0.0	1	2	2	Mean:	–0.001700	–0.000850	–0.001130	–0.000850	–0.000570
				Var:	0.171245	0.042811	0.076109	0.042811	0.019027
				MSE:	0.171250	**0.042813**	0.076112	**0.042813**	**0.019028**
0.0	1	4	4	Mean:	–0.000800	0.000400	–0.000640	–0.000530	–0.000400
				Var:	0.088874	0.053058	0.056879	0.039499	0.022219
				MSE:	0.088875	0.053058	0.056880	**0.039500**	**0.022219**
0.0	1	8	8	Mean:	0.002625	0.001965	0.002333	0.002100	0.001750
				Var:	0.042575	0.035480	0.033641	0.027249	0.018923
				MSE:	0.042584	0.035483	0.033647	**0.027254**	**0.018926**
0.0	1	16	16	Mean:	–0.000050	0.000328	–0.000047	–0.000044	–0.000040
				Var:	0.021759	0.020761	0.019275	0.017193	0.013926
				MSE:	0.021759	0.020761	0.019275	**0.017193**	**0.013926**
0.0	1	32	32	Mean:	–0.001900	–0.001950	–0.001840	–0.001790	-0.001690
				Var:	0.010805	0.010674	0.010160	0.009572	0.008538
				MSE:	0.010809	0.010678	0.010164	**0.009575**	**0.008540**
0.0	1	100	100	Mean:	0.000566	0.000572	0.000560	0.000555	0.000544
				Var:	0.003482	0.003478	0.003413	0.003346	0.003219
				MSE:	0.003482	0.003478	0.003413	**0.003347**	**0.003219**
0.1	16	2	2	Mean:	0.102200	0.051100	0.068133	0.051100	0.034067
				Var:	0.178755	0.044689	0.079446	0.044689	0.019861
				MSE:	0.178759	**0.047080**	0.080462	**0.047080**	**0.024210**
0.1	16	4	4	Mean:	0.101900	0.071067	0.081520	0.067933	0.050950
				Var:	0.093292	0.056358	0.059708	0.041462	0.023323
				MSE:	0.093295	0.057194	0.060047	**0.042490**	**0.025729**
0.1	16	4	8	Mean:	0.091175	0.073915	0.078964	0.069820	0.056883
				Var:	0.068527	0.048536	0.047903	0.036184	0.023445
				MSE:	0.068605	0.049217	0.048345	**0.037095**	**0.025305**
0.1	16	4	16	Mean:	0.096425	0.086770	0.087330	0.080322	0.069865
				Var:	0.057752	0.041273	0.040889	0.032469	0.024164
				MSE:	0.057764	0.041448	0.041048	**0.032856**	**0.025072**
0.1	16	4	32	Mean:	0.103087	0.094537	0.095306	0.089488	0.080958
				Var:	0.052651	0.037007	0.037127	0.030458	0.025400
				MSE:	0.052662	0.037037	0.037149	**0.030568**	**0.025763**
0.1	16	8	8	Mean:	0.105625	0.091604	0.093890	0.084500	0.070417
				Var:	0.047621	0.041375	0.037626	0.030478	0.021165
				MSE:	0.047653	0.041446	0.037664	**0.030718**	**0.022040**
0.1	16	8	16	Mean:	0.100700	0.094838	0.093524	0.087382	0.077367
				Var:	0.035620	0.031899	0.029404	0.024987	0.019128
				MSE:	0.035620	0.031926	0.029445	**0.025147**	**0.019641**
0.1	16	8	32	Mean:	0.097381	0.093334	0.092488	0.088258	0.081217
				Var:	0.028539	0.025407	0.023764	0.020808	0.017542
				MSE:	0.028546	0.025452	0.023820	**0.020945**	**0.017895**
0.1	16	16	16	Mean:	0.099100	0.094834	0.093271	0.088089	0.079280
				Var:	0.023792	0.023050	0.021075	0.018798	0.015227
				MSE:	0.023793	0.023077	0.021120	**0.018941**	**0.015656**
0.1	16	32	32	Mean:	0.100794	0.099611	0.097741	0.094866	0.089594
				Var:	0.011022	0.010951	0.010364	0.009764	0.008709
				MSE:	0.011023	0.010951	0.010369	**0.009790**	**0.008817**
0.1	16	100	100	Mean:	0.100052	0.099934	0.099061	0.098092	0.096204
				Var:	0.003728	0.003725	0.003654	0.003583	0.003446
				MSE:	0.003728	0.003725	0.003655	**0.003587**	**0.003461**

Table 2. Empirical type I error for testing $H_0 : \theta = \theta_0$ at 5% significance level.

θ_0	k	n_{1j}	n_{2j}	CMH	INV ($c = 0$)	$c = 0.5$	$c = 1$	$c = 2$
0.0	1	4	4	3.42	3.42	3.42	3.42	3.42
		4	8	2.08	2.08	6.84	**4.76**	**4.76**
		4	16	3.00	3.00	6.52	**5.80**	8.18
		4	32	2.76	2.76	6.66	6.18	10.50
		4	100	2.54	2.54	7.30	6.46	14.40
		8	8	3.28	3.28	6.76	**4.16**	**4.16**
		8	16	**4.26**	**4.26**	6.54	**4.74**	**4.30**
		8	32	**4.34**	**4.34**	**5.58**	**4.22**	**5.10**
		8	100	**5.02**	**5.02**	6.58	**6.00**	8.90
		16	16	**4.74**	**4.74**	**4.48**	**4.48**	3.38
		16	32	**4.50**	**4.50**	**4.94**	**4.44**	3.90
		16	100	**5.02**	**5.02**	**5.30**	**4.58**	**5.10**
		32	32	**5.04**	**5.04**	**4.66**	**4.34**	3.88
		32	100	**5.22**	**5.22**	**5.16**	**4.46**	**4.34**
		100	100	**4.74**	**4.74**	**4.60**	**4.40**	**4.14**
0.0	4	4	4	3.68	3.68	3.68	3.68	3.68
		8	8	3.40	3.40	7.14	**4.56**	**4.56**
		16	16	**4.84**	**4.84**	**4.66**	**4.66**	3.54
		16	32	**4.52**	**4.52**	**5.00**	**4.52**	**4.10**
		16	100	**5.46**	**5.46**	**5.66**	**4.72**	**5.26**
		32	32	**4.74**	**4.74**	**4.42**	**4.18**	3.92
		32	100	**5.34**	**5.34**	**5.48**	**4.74**	**4.46**
		100	100	**5.04**	**5.04**	**4.98**	**4.86**	**4.64**
0.1	4	4	4	1.26	1.26	8.28	8.28	6.22
		8	8	**4.24**	**4.24**	7.6	**4.66**	**4.66**
		16	16	**5.18**	**5.18**	**5.76**	**5.04**	**4.06**
		16	32	**5.66**	**5.66**	**5.82**	**5.40**	**5.30**
		16	100	**58.6**	**5.86**	**62.0**	**4.84**	**4.88**
		32	32	**5.72**	**5.72**	**5.64**	**4.96**	**4.44**
		32	100	**5.88**	**5.88**	**5.44**	**5.20**	**4.82**
		100	100	**5.22**	**5.22**	**5.16**	**5.10**	**4.82**
0.2	4	4	4	1.74	1.74	**4.36**	**4.36**	8.00
		8	8	**4.66**	**4.66**	8.58	5.38	5.38
		16	16	7.54	7.54	6.32	6.28	6.58
		16	32	7.26	7.26	6.22	**5.56**	**5.60**
		16	100	6.24	6.24	6.18	**5.40**	**5.88**
		32	32	**5.46**	**5.46**	**5.40**	**5.46**	**5.08**
		32	100	**5.56**	**5.56**	**5.26**	**5.22**	**48.8**
		100	100	**5.34**	**5.34**	**5.16**	**5.10**	**5.22**
0.4	4	4	4	3.00	3.00	12.06	7.44	18.04
		8	8	8.00	8.00	6.82	9.18	12.04
		16	16	**5.78**	**5.78**	**5.92**	**5.16**	7.04
		16	32	6.82	6.82	6.56	6.16	7.56
		16	100	6.38	6.38	6.18	**5.80**	7.06
		32	32	**5.96**	**5.96**	**5.78**	**5.94**	6.28
		32	100	**5.92**	**5.92**	**5.80**	6.04	6.72
		100	100	**5.68**	**5.68**	**5.34**	**5.14**	**5.48**

Bold values denote that the statistical tests can control the type I error.

- The increasing k cannot change the order of the empirical type I error rates of all tests. Also, the unbalanced cases of n_{1j} and n_{2j} for center j have a slight effect on the order of the empirical type I error rates of all tests.
- None of tests can control type I error rates when sample size of treatment or control arm is very small ($n_{1j} \le 4$ or $n_{2j} \le 4$). There exists few tests that can control type I error when sample size is small ($n_{1j} = 8$ or $n_{2j} = 8$).
- For $\theta = 0$, almost all tests can control type I error rates when the sample size is moderate to large ($n_{1j} \ge 16$ or $n_{2j} \ge 16$). This finding frequently occurs in practical use of $H_0 : \theta = 0$.
- For $\theta = 0.2$, $\theta = 0.4$, and $\theta = 0.6$, almost all tests can control type I error rates when the sample size is large to very large ($n_{1j} \ge 32$ or $n_{2j} \ge 32$).

5.3. Results for Studying Power of Tests

Table 3 shows some more details of the powers. Fortunately, almost all tests under $H_0 : \theta = 0$ can control type I error rates when the sample size is moderate to large ($n_{1j} \ge 16$ or $n_{2j} \ge 16$). We ignore to consider the comparative tests when sample size is very small ($n_{1j} \le 4$ or $n_{2j} \le 4$) since all of tests can not control type I error rates. The performance of several weighted tests according to the powers under $H_1 : \theta_j = 0.1 + U_m$ can be concluded in the following:

- The empirical powers yield a similar pattern of results like the MSE. An increase in the number of centers, k, can increase the power but it can not change the order of power.
- Overall, the proposed weights adjusted by $c = 1$ including $c = 2$ perform best with the highest power in a multi-center study of size $k \ge 2$ when $n_{1j} \ge 16$ or $n_{2j} \ge 16$.
- The INV weight and the CMH weight are achieved with the highest powers in one center study when $n_{1j} \ge 16$ or $n_{2j} \ge 16$.
- When the sample size is large to very large ($n_{1j} \ge 32$ or $n_{2j} \ge 32$), all weights perform well.

6. Numerical Examples

Two examples are presented to illustrate the implementation of the related methodology. Pocock [20] presented data from a randomized trial studying the effect of placebo and metoprolol on mortality after heart attack (AMI: Acute Myocardial Infarction) classified by three strata of age groups, namely, 40 - 64, 65 - 69, 70 - 74 years. **Table 4** shows the data and weights corresponding to the CMH, the INV, and the proposed strategies. The estimated summary differences based on the CMH, the INV, and the proposed weights are 0.031, 0.024, 0.030, respectively. Also, the estimated standard errors of those of overall differences are 0.014, 0.013, 0.014, respectively. Since both of $Z_{CMH} = 2.237$ and $Z_{cw} = 2.197$ are greater than $Z_{\alpha/2} = 1.96$, the CMH and the proposed tests at $c = 1$ reject the null hypothesis at 5% level for two-sided test and lead to the conclusion of a significant difference between the placebo and metoprolol mortality rates whereas the INV test with $Z_{INV} = 1.823$ fails to reject the null hypothesis at 5% level.

Turner *et al.* [21] presented data from clinical trials to study the effect of selective decontamination of the digestive tract on the risk of respiratory tract infection of patients in intensive care units. See data and weights in **Table 5**. The estimated overall differences and their estimated standard errors are 0.152 (0.012), 0.140 (0.011), 0.162 (0.012) for the CMH, the INV, and the proposed weights at $c = 1$, respectively. All tests reject the null hypothesis with $Z_{CMH} = 12.584$, $Z_{INV} = 12.215$, $Z_{cw} = 13.719$ and lead to the conclusion of a significant difference between treatment effect of selective decontamination of the digestive tract on the risk of respiratory tract infection.

7. Conclusions and Discussion

In most general situations used by the risk difference lying on [0, 0.25], the results have confirmed that the minimum MSE weight of the proposed summary estimator $\hat{\theta}_{cw}$ adjusted by $c = c_1 = c_2 = 1$ (including $c = c_1 = c_2 = 2$) is the best choice with the smallest MSE under a constant of common risk difference θ over all k centers. The number of centers, k, cannot change the order of the MSE of all weighted estimators, even though an increase in k can decrease the variance and the MSE of all weighted estimators. Also, increasing n_{1j} and n_{2j} can decrease the variance of all estimators while fixing k. The unbalanced cases of n_{1j} and n_{2j} for center j have a slight effect on the order of the MSE of all estimates. The minimum MSE weight is designed to yield more precise estimate relative to the CMH and INV weights. Another benefit of the proposed weight is easy to compute because of its closed-form formula. With the basis of smallest MSE and the easy-to-compute formula, we have been solidly suggested to use the proposed weight. In addition, the various choices for c have been considered again. The use of $c = 0.5$ as a conventional correction term [22] should be revised. The better value of c in adding on the number of successes and the number of failures is suggested with at least for $c = 1$ (including $c = 2$). This result is supported by the ideas of Böhning and Viwatwongkasem [6], Agresti and Coull [1], and Agresti and Caffol [2] that recommended to use the appropriate values of c greater than or equal to 1.

Table 3. Empirical power (percent) at m = 0.04 after controlling the estimated type I error at the nominal 5% level.

			X = Controllable Type I error rates					Empirical power rates				
k	n_{1j}	n_{2j}	CMH	INV	c = 0.5	c = 1	c = 2	CMH	INV	c = 0.5	c = 1	c = 2
1	8	8				X	X				**6.8**	**6.8**
	8	16	X	X		X	X	7.3	7.3		**8.3**	**7.4**
	8	32	X	X	X	X	X	9.5	9.5	**11.5**	9.6	9.9
	8	100	X	X		X		11.3	11.3		**11.8**	
	16	16	X	X	X	X		**11.2**	**11.2**	10.6	10.6	
	16	32	X	X	X	X		12.2	12.2	**12.7**	11.8	
	16	100	X	X	X	X	X	**16.4**	**16.4**	15.4	14.8	14.6
	32	32	X	X	X	X		**17.6**	**17.6**	16.5	16.4	
	32	100	X	X	X	X	X	**21.4**	**21.4**	21.2	20.8	20.3
	100	100	X	X	X	X	X	**36.8**	**36.8**	**36.8**	36.5	36.1
4	8	8				X	X				**26.9**	**29.7**
	8	16				X	X				**29.5**	**32.8**
	8	32	X	X	X	X	X	20.8	23.8	31.5	**33.1**	**35.2**
	8	100	X	X		X		23.6	28.9		**36.8**	
	16	16	X	X	X	X		25.3	27.0	31.0	**33.4**	
	16	32	X	X	X	X	X	32.4	35.9	38.0	**40.6**	**43.6**
	16	100	X	X	X	X	X	39.1	44.6	45.2	**46.8**	**48.9**
	32	32	X	X	X	X		44.2	46.1	47.8	**49.5**	
	32	100	X	X	X	X	X	58.1	60.6	61.4	**62.8**	**64.6**
	100	100	X	X	X	X	X	85.6	87.0	86.7	**87.2**	**87.8**
8	8	8										
	8	16			X	X				44.0	**48.8**	
	8	32	X	X	X	X	X	35.3	39.5	50.2	**53.9**	**59.0**
	8	100	X	X				39.9	46.0			
	16	16	X	X	X	X		43.1	45.9	52.7	**56.9**	
	16	32	X	X	X	X	X	53.4	57.0	61.3	**64.3**	**68.5**
	16	100	X	X	X	X	X	65.7	69.3	72.0	**74.5**	**77.1**
	32	32	X	X	X	X	X	71.1	72.9	74.8	**76.9**	**80.4**
	32	100	X	X	X	X	X	86.1	87.7	88.3	**89.1**	**90.4**
	100	100	X	X	X	X	X	98.8	98.9	99.0	**99.1**	**99.1**
16	8	8				X	X				**68.3**	**77.5**
	8	16			X					68.5		
	8	32	X	X	X	X	X	60.9	64.6	74.2	**77.1**	**82.1**
	8	100	X	X		X		67.4	72.1		**82.0**	
	16	16	X	X	X	X		71.0	73.8	79.0	**82.2**	
	16	32	X	X	X	X	X	82.5	84.4	87.3	**89.2**	**92.0**
	16	100	X	X	X	X	X	90.3	90.8	93.1	**93.8**	**94.8**
	32	32	X	X	X	X		93.9	94.9	95.3	**96.0**	
	32	100	X	X	X	X	X	99.0	99.1	99.1	**99.2**	**99.3**
	100	100	X	X	X	X	X	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
32	8	8										
	8	16	X	X		X	X	81.8	83.2		**92.7**	**95.6**
	8	32	X	X	X	X	X	88.7	90.0	94.1	**95.1**	**96.7**
	8	100	X	X		X		92.2	93.4		**96.4**	
	16	16	X	X				94.5	95.0		**97.5**	
	16	32	X	X	X	X		98.1	98.5	99.0	**99.1**	
	16	100	X	X	X	X	X	99.7	99.5	99.8	**99.9**	**99.9**
	32	32	X	X	X	X		99.8	**99.9**	**99.9**	**99.9**	
	32	100	X	X	X	X	X	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
	100	100	X	X	X	X	X	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Table 4. Mortality data over three strata of age groups following Pocock.

Age	Placebo		Metoprolol			Weights		
j	x_{1j}	n_{1j}	x_{2j}	n_{2j}	$\hat{\theta}_j$	CMH	INV	$c=1$
40 - 64	26	453	21	464	0.012	0.66	0.79	0.69
65 - 69	25	174	11	165	0.077	0.24	0.16	0.25
70 - 74	11	70	8	69	0.041	0.10	0.05	0.06

Table 5. Respiratory tract infections following Turner *et al.*

Trial	Treatment I		Treatment II			Weights		
j	x_{1j}	n_{1j}	x_{2j}	n_{2j}	$\hat{\theta}_j$	CMH	INV	$c=1$
1	25	54	7	47	0.314	0.03	0.02	0.02
2	24	41	4	38	0.480	0.02	0.02	0.03
3	37	95	20	96	0.181	0.05	0.03	0.04
4	11	17	1	14	0.576	0.01	0.01	0.01
5	26	49	10	48	0.322	0.03	0.02	0.02
6	13	84	2	101	0.135	0.05	0.07	0.07
7	38	170	12	161	0.149	0.09	0.09	0.09
8	29	60	1	28	0.448	0.02	0.02	0.03
9	9	20	1	19	0.397	0.01	0.01	0.01
10	44	47	22	49	0.487	0.03	0.02	0.03
11	30	160	25	162	0.033	0.08	0.07	0.06
12	40	185	31	200	0.061	0.10	0.08	0.07
13	10	41	9	39	0.013	0.02	0.01	0.01
14	40	185	22	193	0.102	0.10	0.09	0.09
15	4	46	0	45	0.087	0.02	0.06	0.04
16	60	140	31	131	0.192	0.07	0.04	0.05
17	12	75	4	75	0.107	0.04	0.05	0.05
18	42	225	31	220	0.046	0.12	0.11	0.09
19	26	57	7	55	0.329	0.03	0.02	0.03
20	17	92	3	91	0.152	0.05	0.07	0.07
21	23	23	14	25	0.440	0.01	0.01	0.02
22	6	68	3	65	0.042	0.03	0.07	0.05

In terms of type I error estimates, when sample size is very small ($n_{1j} \le 4$ or $n_{2j} \le 4$), none of tests can control type I error rates. In addition, there exists few tests that can control type I error rates when sample size is small ($n_{1j} = 8$ or $n_{2j} = 8$). This result is consonant with the comments of Lui [23] that none of conventional tests/weights under sparse data is appropriate. This inappropriateness under sparse data can cope with the minimum MSE weights from this finding. The further work to seek some appropriate tests/weights in sparse data challenges for investigators to develop an innovation or to improve much more reasonable tests/weights. In general results, almost all tests can control type I error rates when sample size is moderate to large ($n_{1j} \ge 16$ or $n_{2j} \ge 16$).

In terms of power, we ignore to evaluate the power when sample size is very small ($n_{1j} \le 4$ or $n_{2j} \le 4$) because all tests can not control type I error rates. The results illustrate the same pattern like the MSE. The proposed weights adjusted by $c=1$ including $c=2$ perform best with the highest power in a multi-center study of size $k \ge 2$ when $n_{1j} \ge 16$ or $n_{2j} \ge 16$. The INV weight and the CMH weight are achieved with the highest powers in one center study when $n_{1j} \ge 16$ or $n_{2j} \ge 16$. When sample size is large to very large ($n_{1j} \ge 32$ or $n_{2j} \ge 32$), all tests perform well. We

strongly recommend to use the minimum MSE weight as an appropriate choice because of its highest power.

8. Acknowledgements

We would like to thank the editors and the referees for comments which greatly improved this paper. This study was partially supported for publication by the China Medical Board (CMB), Faculty of Public Health, Mahidol University, Bangkok, Thailand.

REFERENCES

[1] A. Agresti and B. A. Coull, "Approximate Is Better than Exact for Interval Estimation of Binomial Proportions," *American Statistical Association*, Vol. 52, 1998, pp. 119-126.

[2] A. Agresti and B. Caffo, "Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures," *The American Statistician*, Vol. 54, No. 4, 2000, pp. 280-288. doi:10.2307/2685779

[3] B. K. Ghosh, "A Comparison of Some Approximate Confidence Interval for the Binomial Parameter," *Journal of the American Statistical Association*, Vol. 74, No. 368, 1979, pp. 894-900. doi:10.2307/2286420

[4] R. G. Newcombe, "Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods," *Statistics in Medicine*, Vol. 17, No. 8, 1998, pp. 857-872. doi:10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E

[5] R. G. Newcombe, "Interval Estimation for the Difference between Independent Proportions: Comparison of Eleven Methods," *Statistics in Medicine*, Vol. 17, No. 8, 1998, pp. 873-890. doi:10.1002/(SICI)1097-0258(19980430)17:8<873::AID-SIM779>3.0.CO;2-I

[6] D. Böhning and C. Viwatwongkasem, "Revisiting Proportion Estimators," *Statistical Methods in Medical Research*, Vol. 14, No. 2, 2005, pp. 147-169. doi:10.1191/0962280205sm393oa

[7] G. Casella and R. L. Berger, "Statistical Inference," Duxbury Press, Belmont, 1990.

[8] A. E. Taylor and W. R. Mann, "Advanced Calculus," John Wiley & Sons, New York, 1972.

[9] K. J. Lui and K. C. Chang, "Testing Homogeneity of Risk Difference in Stratified Randomized Trials with Noncompliance," *Computational Statistics and Data Analysis*, Vol. 53, No. 1, 2008, pp. 209-221. doi:10.1016/j.csda.2008.07.016

[10] W. G. Cochran, "The Combination of Estimates from Different Experiments," *Biometrics*, Vol. 10, No. 1, 1954, pp. 101-129. doi:10.2307/3001666

[11] W. G. Cochran, "Some Methods for Strengthening the Common Chi-Square Test," *Biometrics*, Vol. 10, No. 4, 1954, pp. 417-451. doi:10.2307/3001616

[12] N. Mantel and W. Haenszel, "Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease," *Journal of the National Cancer Institute*, Vol. 22, 1959, pp. 719-748.

[13] J. Sanchez-Meca and F. Marin-Martinez, "Testing the Significance of a Common Risk Difference in Meta-Analysis," *Computational Statistics & Data Analysis*, Vol. 33, No. 3, 2000, pp. 299-313. doi:10.1016/S0167-9473(99)00055-9

[14] J. L. Fleiss, "Statistical Methods for Rates and Proportions," John Wiley & Sons Inc., New York, 1981.

[15] S. R. Lipsitz, K. B. G. Dear, N. M. Laird and G. Molenberghs, "Tests for Homogeneity of the Risk Difference When Data Are Sparse," *Biometrics*, Vol. 54, No. 1, 1998, pp. 148-160. doi:10.2307/2534003

[16] D. G. Kleinbaum, L. L. Kupper and H. Morgenstern, "Epidemiologic Research: Principles and Quantitative Methods," Lifetime Learning Publications, Belmont, 1982.

[17] D. B. Petitti, "Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine," Oxford University Press, Oxford, 1994.

[18] W. G. Cochran, "The Chi-Square Test of Goodness of Fit," *Annals of Mathematical Statistics*, Vol. 23, No. 3, 1952, pp. 315-345. doi:10.1214/aoms/1177729380

[19] K. J. Lui and C. Kelly, "Tests for Homogeneity of the Risk Ratio in a Series of 2 × 2 Tables," *Statistics in Medicine*, Vol. 19, No. 21, 2000, pp. 2919-2932. doi:10.1002/1097-0258(20001115)19:21<2919::AID-SIM561>3.0.CO;2-D

[20] S. J. Pocock, "Clinical Trials: A Practical Approach," Wiley Publication, New York, 1997.

[21] R. Tuner, R. Omar, M. Yang, H. Goldstein and S. Thompson, "A Multilevel Model Framework for Meta-Analysis of Clinical Trials with Binary Outcome," *Statistics in Medicin*e, Vol. 19, No. 24, 2000, pp. 3417-3432. doi:10.1002/1097-0258(20001230)19:24<3417::AID-SIM614>3.0.CO;2-L

[22] F. Yates, "Contingency Tables Involving Small Numbers and the Chi-Squared Test," *Journal of the Royal Statistical Society* (*Supplement*), Vol. 1, 1934, pp. 217-235.

[23] K. J. Lui, "A Simple Test of the Homogeneity of Risk Difference in Sparse Data: An Application to a Multicenter Study," *Biometrical Journal*, Vol. 47, No. 5, 2008, pp. 654-661. doi:10.1002/bimj.200410150

[24] C. A. Rencher, "Linear Models in Statistics," Wiley Series in Probability and Mathematical Statistics, New York, 2000.

[25] A. Sen and M. Srivastava, "Regression Analysis: Theory, Methods, and Applications," Springer-Verlag, New York, 1990.

Appendix

Under a true common risk difference θ over all k centers ($j=1,2,\cdots,k$), the mean square error of $\hat{\theta}_{cw}=\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}$ is given by

$$MSE\left(\hat{\theta}_{cw}\right)=E\left(\hat{\theta}_{cw}-\theta\right)^2=E\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}-\theta\right)^2$$

To obtain the optimal weights $\left\{f_{cj}\right\}$ subject to a constraint that $\sum_{j=1}^{k} f_{cj}-1=0$, we form the auxiliary function ϕ by following Lagrange's method to seek $\left\{f_{cj}\right\}$ that minimize

$$\phi=E\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}-\theta\right)^2+\lambda\left(\sum_{j=1}^{k} f_{cj}-1\right)$$

$$\phi=E\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}\right)^2-2\theta E\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}\right)+\theta^2+\lambda\left(\sum_{j=1}^{k} f_{cj}-1\right)$$

$$\phi=V\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}\right)+\left(E\left(\sum_{j=1}^{k} f_{cj}\hat{\theta}_{cj}\right)\right)^2$$

$$-2\theta\sum_{j=1}^{k} f_{cj}E\left(\hat{\theta}_{cj}\right)+\theta^2+\lambda\left(\sum_{j=1}^{k} f_{cj}-1\right)$$

$$\phi=\sum_{j=1}^{k} f_{cj}^2V\left(\hat{\theta}_{cj}\right)+\left(\sum_{j=1}^{k} f_{cj}E\left(\hat{\theta}_{cj}\right)-\theta\right)^2+\lambda\left(\sum_{j=1}^{k} f_{cj}-1\right)$$

Let $V_j=V\left(\hat{\theta}_{cj}\right)$ and $E_j=E\left(\hat{\theta}_{cj}\right)$. The partial derivatives with respect to λ and f_{cj} yield

$$\frac{\partial\phi}{\partial\lambda}=\sum_{j=1}^{k} f_{cj}-1,\quad \frac{\partial\phi}{\partial f_{cj}}=2f_{cj}V_j+2\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)E_j+\lambda$$

Setting $\frac{\partial\phi}{\partial\lambda}=0$ and $\frac{\partial\phi}{\partial f_{cj}}=0$, it yields

$$\sum_{j=1}^{k} f_{cj}=1$$

$$f_{cj}=-\frac{E_j}{V_j}\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)-\frac{\lambda}{2V_j}$$

Solving for λ by taking summation on f_{cj}, it yields

$$\sum_{j=1}^{k} f_{cj}=1=-\left(\sum_{j=1}^{k}\frac{E_j}{V_j}\right)\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)-\frac{\lambda}{2}\left(\sum_{j=1}^{k}\frac{1}{V_j}\right)$$

Let $a=\sum_{j=1}^{k}\frac{1}{V_j}=\sum_{j=1}^{k}V_j^{-1}$, $b=\sum_{j=1}^{k}\frac{E_j}{V_j}$, then $\lambda/2$ can be written as

$$\frac{\lambda}{2}=-\frac{1}{a}-\frac{b\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)}{a}$$

Hence,

$$f_{cj}=-\frac{E_j}{V_j}\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)+\frac{1}{V_j\,a}+\frac{b\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)}{V_j\,a}$$

$$(aV_j)\ f_{cj}=-a\,E_j\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)+1+b\left(\sum_{j=1}^{k} f_{cj}E_j-\theta\right)$$

$$(aV_j)\ f_{cj}=-a\,E_j\left(\sum_{j=1}^{k} f_{cj}E_j\right)+aE_j\theta+1$$

$$+b\left(\sum_{j=1}^{k} f_{cj}E_j\right)-b\theta$$

$$aV_j\,f_{cj}+\left(\sum_{j=1}^{k} f_{cj}E_j\right)\left(a\,E_j-b\right)=1+\left(a\,E_j-b\right)\theta$$

Let $\tau_j=aE_j-b$

$$aV_j\,f_{cj}+\left(\sum_{j=1}^{k} f_{cj}E_j\right)\tau_j=1+\tau_j\theta$$

$$aV_j\,f_{cj}+\left(f_{c1}E_1+f_{c2}E_2+\cdots+f_{ck}E_k\right)\tau_j=1+\tau_j\theta$$

Substitute each of the subscript j and rearrange terms.

$$\begin{aligned}
&j=1;\ \ (aV_1+E_1\tau_1)f_{c1}+f_{c2}E_2\tau_1+f_{c3}E_3\tau_1+\cdots+f_{ck}E_k\tau_1=1+\tau_1\theta\\
&j=2;\ \ f_{c1}E_1\tau_2+(aV_2+E_2\tau_2)f_{c2}+f_{c3}E_3\tau_2+\cdots+f_{ck}E_k\tau_2=1+\tau_2\theta\\
&j=3;\ \ f_{c1}E_1\tau_3+f_{c2}E_2\tau_3+(aV_3+E_3\tau_3)f_{c3}+\ldots+f_{ck}E_k\tau_3=1+\tau_3\theta\\
&\qquad\vdots\qquad\qquad\vdots\qquad\qquad\vdots\qquad\qquad\vdots\\
&j=k;\ \ f_{c1}E_1\tau_k+f_{c2}E_2\tau_k+f_{c3}E_3\tau_k+\ldots+(aV_k+E_k\tau_k)f_{ck}=1+\tau_k\theta
\end{aligned}$$

It can be written in the matrix form as $\mathbf{Hf}=\mathbf{y}$ where

$$\mathbf{H}=\begin{bmatrix}
aV_1+E_1\tau_1 & E_2\tau_1 & E_3\tau_1 & \cdots & E_k\tau_1\\
E_1\tau_2 & aV_2+E_2\tau_2 & E_3\tau_2 & \cdots & E_k\tau_2\\
E_1\tau_3 & E_2\tau_3 & aV_3+E_3\tau_3 & \cdots & E_k\tau_3\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
E_1\tau_k & E_2\tau_k & E_3\tau_k & \cdots & aV_k+E_k\tau_k
\end{bmatrix}$$

$$\mathbf{f}' = \begin{bmatrix} f_{c1} & f_{c2} & f_{c3} & \cdots & f_{ck} \end{bmatrix},$$
$$\mathbf{y}' = \begin{bmatrix} 1+\tau_1\theta & 1+\tau_2\theta & 1+\tau_3\theta & \cdots & 1+\tau_k\theta \end{bmatrix}$$

The matrix $\mathbf{H}$ can be illustrated as

$$\mathbf{H} = \mathbf{D} + \mathbf{t}\mathbf{e}'$$

where

$$\mathbf{D} = \begin{bmatrix} aV_1 & 0 & \cdots & 0 \\ 0 & aV_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & aV_k \end{bmatrix},$$

$$\mathbf{t}' = \begin{bmatrix} \tau_1 & \tau_2 & \cdots & \tau_k \end{bmatrix},$$
$$\mathbf{e}' = \begin{bmatrix} E_1 & E_2 & \cdots & E_k \end{bmatrix}$$

The inverse of $\mathbf{H}$ is suggested in several textbooks of linear model such as Rencher [24] and Sen and Srivastava [25]. It yields

$$\mathbf{H}^{-1} = \left(\mathbf{D} + \mathbf{t}\mathbf{e}'\right)^{-1} = \mathbf{D}^{-1} - \frac{\left(\mathbf{D}^{-1}\mathbf{t}\right)\left(\mathbf{e}'\mathbf{D}^{-1}\right)}{1+\mathbf{e}'\mathbf{D}^{-1}\mathbf{t}}$$

Therefore, $\mathbf{f} = \mathbf{D}^{-1}\mathbf{y} - \dfrac{\left(\mathbf{D}^{-1}\mathbf{t}\right)\left(\mathbf{e}'\mathbf{D}^{-1}\right)}{1+\mathbf{e}'\mathbf{D}^{-1}\mathbf{t}}\mathbf{y}$

$$\mathbf{f} = \begin{bmatrix} \frac{1}{a}\left[V_1^{-1}(1+\tau_1\theta)\right] \\ \frac{1}{a}\left[V_2^{-1}(1+\tau_2\theta)\right] \\ \vdots \\ \frac{1}{a}\left[V_k^{-1}(1+\tau_k\theta)\right] \end{bmatrix} - \begin{bmatrix} \left(\frac{V_1^{-1}\tau_1}{\left(a+\sum_{m=1}^{k}\tau_m E_m V_m^{-1}\right)}\right)\left(\frac{\sum_{m=1}^{k}V_m^{-1}E_m(1+\tau_m\theta)}{a}\right) \\ \left(\frac{V_2^{-1}\tau_2}{\left(a+\sum_{m=1}^{k}\tau_m E_m V_m^{-1}\right)}\right)\left(\frac{\sum_{m=1}^{k}V_m^{-1}E_m(1+\tau_m\theta)}{a}\right) \\ \vdots \\ \left(\frac{V_k^{-1}\tau_k}{\left(a+\sum_{m=1}^{k}\tau_m E_m V_m^{-1}\right)}\right)\left(\frac{\sum_{m=1}^{k}V_m^{-1}E_m(1+\tau_m\theta)}{a}\right) \end{bmatrix}$$

Therefore, for the j^{th} center, it yields

$$f_{cj} = \left(\frac{V_j^{-1}(1+\tau_j\theta)}{a}\right) - \left(\frac{V_j^{-1}\tau_j}{\left(a+\sum_{m=1}^{k}\tau_m E_m V_m^{-1}\right)}\right)\left(\frac{\sum_{m=1}^{k}V_m^{-1}E_m(1+\tau_m\theta)}{a}\right)$$

where $a = \sum_{j=1}^{k}\frac{1}{V_j} = \sum_{j=1}^{k}V_j^{-1}$,

$b = \sum_{j=1}^{k}\frac{E_j}{V_j}$, $\tau_j = aE_j - b$

$$E_j = E\left(\hat{\theta}_{cj}\right) = E\left(\hat{p}_{c1j} - \hat{p}_{c2j}\right) = \frac{n_{1j}p_{1j}+c_1}{n_{1j}+2c_1} - \frac{n_{2j}p_{2j}+c_2}{n_{2j}+2c_2}$$

$$V_j = V\left(\hat{\theta}_{cj}\right) = V\left(\hat{p}_{c1j} - \hat{p}_{c2j}\right) = \frac{n_{1j}p_{1j}\left(1-p_{1j}\right)}{\left(n_{1j}+2c_1\right)^2} + \frac{n_{2j}p_{2j}\left(1-p_{2j}\right)}{\left(n_{2j}+2c_2\right)^2}$$

In practice, we have to estimate the adjusted summary estimator by replacing the sample estimates for the unknown quantities: E_j, V_j, p_{1j}, p_{2j}, θ.

