Some tests based on the profile likelihood estimator for testing homogeneity of diagnostic odds ratios in meta-analysis

Khanokporn Donjdee\(^a\), Pratana Satitvipawee\(^a\), Prasong Kitidamrongusk\(^a\), Pichitpong Soontornpipit\(^a\), Jutatip Sillabutra\(^a\), Chukiat Viwatwongkasem\(^a\)*

\(^a\)Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok 10400, Thailand

Abstract

This research aims to propose some modified tests “\( \chi^2_{\text{PLE1}} \)” and “\( \chi^2_{\text{PLE2}} \)” based on the profile likelihood estimator for testing homogeneity of diagnostic odds ratios in meta-analysis and compare their performances with the conventional tests of \( Q_{\text{WLS}} \), \( Q_{\text{MH}} \), and \( \chi^2_{\text{Con}} \). According to the performance in terms of type I error rates under \( H_0 \) and power of tests under \( H_1 \), Monte Carlo simulation with R language was applied. The results found that all of tests cannot control type I error rates when sample sizes are small (\( n_i^D, n_i^H \leq 5 \)), regardless of study size (\( k \)). However, for \( k \geq 16 \) in combination with \( n_i^D, n_i^H \geq 50 \), three tests (\( \chi^2_{\text{PLE1}}, Q_{\text{WLS}}, \chi^2_{\text{Con}} \)) can control type I error rates in almost all situations. In addition, the profile test (\( \chi^2_{\text{PLE1}} \)) performs best with highest power when \( n_i^D, n_i^H = 50,100 \) for \( k \geq 16 \), while conventional tests of \( Q_{\text{WLS}} \) and \( \chi^2_{\text{Con}} \) perform well with the same power as the profile test (\( \chi^2_{\text{PLE1}} \)) when \( n_i^D, n_i^H = 500 \) for \( k \geq 16 \). Therefore, the \( \chi^2_{\text{PLE1}} \) is recommended to be used when \( k \geq 16 \) in combination with \( n_i^D, n_i^H \geq 50 \).

1. Introduction

Let diagnostic odds ratio estimate as an effect size be defined as \( \hat{\theta} = \frac{\hat{q}^D}{\hat{q}^H} \) where \( \hat{q}^D = \hat{p}^D / (1 - \hat{p}^D) \) and \( \hat{q}^H = \hat{p}^H / (1 - \hat{p}^H) \) are the estimated odds of positive risks \( \hat{p}^D = x^D / n^D \) and \( \hat{p}^H = x^H / n^H \) in the disease and healthy...
groups, where \( x^o \) and \( x^h \) are the number of positive data out of the total number of subjects \( n^o \) and \( n^h \), respectively. Mostly the point estimator \( \hat{\theta} \) is derived by a maximum likelihood method. But in this study, we are interested to seek an alternative estimate \( \hat{\theta} \) from the profile likelihood approach which widely used to eliminate a nuisance parameter and it also has an invariant property \(^1\). Usually, meta-analysis is a statistical technique for combining results of different studies into a summarizing result. However, before combing the diagnostic odds ratios of different studies to obtain an overall effect, the hypothesis testing is requested to evaluate whether there is true heterogeneity occurrence, or not. Cochran’s \( Q \) test is conventionally popular for testing the null hypothesis: \( H_0: \theta_1 = \ldots = \theta_k = \theta \) where \( \theta \) is a true diagnostic odds ratio over across study \( i \) \((i = 1, \ldots, k)\) \(^2\). However, Cochran’s \( Q \) test might have been low in the power of test when the number of studies \((k)\) included in the meta-analysis is small. The low power of test implies that a statistically non-significant test can occur even though the genuine heterogeneity of population effects is present. Many scientists try to increase the power of \( Q \) test with several methods. We also have an attempt to modify some tests by replacing the profile likelihood estimator into the variance of logarithm of diagnostic odds ratio. A comparison of the performance of tests in terms of type I error and the power is applied via a simulation study.

2. Methodology

The methods are divided into two parts: (1) providing the idea of creating modified tests for homogeneity of diagnostic odds ratios in meta-analysis and (2) comparing the efficiency between two new proposed tests \((\chi^2_{\text{PLE1}}, \chi^2_{\text{PLE2}})\) based on the profile likelihood estimator (PLE) and the conventional tests by simulation with R language in different situations.

Part 1: We followed the work of Böhning et al. \(^3\) who had already provided the estimate of the diagnostic odds ratios based on the profile likelihood method and the pooled diagnostic odds ratio estimator under homogeneity is obtained as:

\[
\hat{\theta}_{\text{PMLE}} = \frac{\sum_{i=1}^{k} (n_i^h - x_i^h) \chi_i^D}{\sum_{i=1}^{k} I_i(\hat{\theta}_{\text{PMLE}})}
\]

where \( q_i^h[\hat{\theta}] = \frac{-(x_i^o + x_i^h - n_i^h + (x_i^o + x_i^h - n_i^o)\hat{\theta})}{2(x_i^o + x_i^h - n_i^h - n_i^o)\hat{\theta}} \)

\[
r_i(\hat{\theta}) = -4(\hat{\theta} - \hat{\theta}_{\text{PMLE}}) + [(x_i^o + x_i^h - n_i^o)\hat{\theta} + (x_i^o + x_i^h - n_i^h)\hat{\theta} - (x_i^o + x_i^h - n_i^o + n_i^h)\hat{\theta}]^2
\]

\[
t_i(\hat{\theta}) = \frac{n_i^o(\hat{\theta} \times (x_i^h + q_i^h[\hat{\theta}] + q_i^h[\hat{\theta}]) - 1)}{1 + \hat{\theta} \times q_i^h[\hat{\theta}]}
\]

To obtain the modified tests for testing homogeneity of diagnostic odds ratios over \( k \) studies in meta-analysis, we replace the profile likelihood estimator \( \hat{\theta}_{\text{PMLE}} \) into the variance formula of logarithmic diagnostic odds ratio \( \text{vår}(\hat{\theta}) \) on the form of the below \( \chi^2 \):

\[
\chi^2 = \sum_{i=1}^{k} \frac{(\hat{\theta} - \hat{\theta}_{\text{PMLE}})^2}{\text{vår}(\hat{\theta})}
\]

where \( \hat{\theta} = \log \hat{\theta} = \log\left(\hat{p}_i^o/1 - \hat{p}_i^h\right) - \log\left(\hat{p}_i^h/1 - \hat{p}_i^o\right) \) and \( \hat{\theta}_{\text{PMLE}} = \log \hat{\theta}_{\text{PMLE}} \)

1. Conventional variance estimate from the delta’s method for \( \chi^2_{\text{Con}} \)

\[
\text{vår}(\hat{\theta}) = \frac{1}{(n_i^o \hat{p}_i^o(1 - \hat{p}_i^o) + n_i^h \hat{p}_i^h(1 - \hat{p}_i^h))} = \frac{1}{x_i^h} + \frac{1}{n_i^o - x_i^o} + \frac{1}{n_i^o - x_i^h} + \frac{1}{n_i^h - x_i^h}
\]

2. Modification of variance estimate for \( \chi^2_{\text{PLE1}} \) when \( \hat{p}_i^o = x_i^o / n_i^o \) and \( \hat{p}_i^h = \frac{\hat{\theta}_{\text{PMLE}} \hat{p}_i^h}{1 - \hat{p}_i^h + \hat{\theta}_{\text{PMLE}}} \)

\[
\text{vår}(\hat{\theta}) = \frac{1}{n_i^o\left(\frac{\hat{\theta}_{\text{PMLE}} \hat{p}_i^h}{1 - \hat{p}_i^h + \hat{\theta}_{\text{PMLE}}}\right)\left(1 - \frac{\hat{\theta}_{\text{PMLE}} \hat{p}_i^h}{1 - \hat{p}_i^h + \hat{\theta}_{\text{PMLE}}}\right)} + \frac{1}{n_i^h\hat{p}_i^h(1 - \hat{p}_i^h)}
\]
3. Modification of variance estimate for $\chi^2_{\text{PHL}}$ when $p_i^\theta = \frac{q_i^\theta [\hat{\theta}]}{1 + q_i^\theta [\hat{\theta}]}$ and $\hat{p}_i^\theta = \frac{\hat{\theta}_{\text{MLE}} \hat{p}_i^H}{(1 - \hat{p}_i^H + \hat{\theta}_{\text{MLE}} \hat{p}_i^H)}$

$$\text{var}(\hat{\theta}) = \left( \frac{(1 - \hat{p}_i^H + \hat{\theta}_{\text{MLE}} \hat{p}_i^H)^2}{n_i^H(\hat{\theta}_{\text{MLE}} \hat{p}_i^H)} + \frac{1}{n_i^H \hat{p}_i^H} \right) \left( \frac{1}{1 - \hat{p}_i^H} \right)$$

In addition, the conventional $Q$ tests can be found in the following:

4. The Mantel-Haenszel test ($Q_{\text{MH}}$) is

$$Q_{\text{MH}} = \sum_{i=1}^{k} \hat{w}_i^H (\hat{\delta} - \overline{\delta})^2$$

where $\hat{\delta}_i = \log \left( \frac{\hat{p}_i^0}{1 - \hat{p}_i^0} \right) - \log \left( \frac{\hat{p}_i^H}{1 - \hat{p}_i^H} \right)$, $\hat{w}_i^H = x_i^H \times \left( n_i^0 - x_i^0 \right)$, $\overline{\delta} = \frac{\sum_{i=1}^{k} w_i \hat{\delta}_i}{\sum_{i=1}^{k} w_i}$

5. Weighted Least squares test ($Q_{\text{WLS}}$) is

$$Q_{\text{WLS}} = \sum_{i=1}^{k} \hat{w}_i (\hat{\delta} - \overline{\delta})^2$$

where $\hat{\delta}_i = \log \left( \frac{\hat{p}_i^0}{1 - \hat{p}_i^0} \right) - \log \left( \frac{\hat{p}_i^H}{1 - \hat{p}_i^H} \right)$, $\overline{\delta} = \frac{\sum_{i=1}^{k} w_i \hat{\delta}_i}{\sum_{i=1}^{k} w_i}$, $w_i = \frac{1}{\text{var}(\hat{\delta}_i)} = \left( \frac{1}{\hat{x}_i^0} + \frac{1}{n_i^0 - x_i^0} + \frac{1}{\hat{x}_i^H} + \frac{1}{n_i^H - x_i^H} \right)^{-1}$

Part 2: The performance of tests will be compared by R language with 2,000 replications in different situations. The additional conditions of the simulation plan are:

Type I error under homogeneity of a null hypothesis: $H_0: \hat{\theta}_i = \dots = \hat{\theta}_k = \theta$ over all $k$ studies, the values of $p_i^\theta$, $\theta$, $p_i^0$ are related with equation $p_i^\theta = p_i^0 \theta / (1 - p_i^0 + \theta p_i^H)$. For instances, if $\theta = 5$ and let $p_i^H$ be false positive risk in healthy group for the $i^{th}$ study which is generated from a uniform $(0,0.35)$ distribution, then the correspondent true positive risk $p_i^0$ in disease group has the possible range as $(0,0.73)$. Each statistical test can control type I error when the actual type I error lies within Bradley limit $[0.025-0.075]$ at 0.05 significance level.

Power of test under the alternative hypothesis, $\hat{\theta}_i$ has been assumed a specific distribution, e.g. $H_1: \hat{\theta}_i = \hat{\theta}_o + \mu \in [-mm, mm]$ where $\mu$ is a uniform random variable for a given range $[-mm, mm]$ where $mm = 20\%$, $40\%$, $60\%$ for $\hat{\theta}_o$ at the values of 1, 5, 10; or $mm = 10\%$, $20\%$, $30\%$ for $\hat{\theta}_o$ at the values of 50,100.

3. Results

There are a lot of results for comparing the type I error and power of test among five statistic tests based on Chi-square test in the different situations. However, there are some situations of interest that all statistic tests can control type I error under the null hypothesis. We ignore to consider the $Q_{\text{MH}}$ and $\chi^2_{\text{PHL}}$ since these two tests cannot control type I error in almost all situations. This section will show only the results of the performance in terms of powers according to the same type I error rates for both equal and unequal cases when the value of $\hat{\theta}_o$ equals 5 because at this value all test statistics should be calibrated most with the same type I error rate before power comparisons.

**Equal case** ($n_i^0 = n_i^H$)

Figure 1 showed that at $mm = 0.2$ when the number of studies is small ($k = 4$) in combination with moderate to very large sample sizes $(n_i^H, n_i^H \geq 50)$, the $\chi^2_{\text{PLH}}$ test has the highest power, followed by $\chi^2_{\text{CM}}$ and $Q_{\text{WLS}}$, respectively. Meanwhile, when the number of studies is moderate ($k = 16$) in combination with moderate to very large sample sizes (in graph sample size code 3 means $n_i^H, n_i^H \geq 50$), the $\chi^2_{\text{PLH}}$ test has the highest power, followed by $Q_{\text{WLS}}$. When the number of studies is large ($k = 32$) in combination with very large sample size (in graph sample size code 5 means $n_i^H, n_i^H \geq 500$), the $\chi^2_{\text{CM}}, Q_{\text{WLS}}, \chi^2_{\text{PLH}}$ tests perform well with close high power.
Unequal case ($n_i^0 \neq n_i^H$)

From the Figure 2, the best performance of test statistics for small size of study ($k = 4$) and moderate sample size (code 8 in graph means $n_i^0, n_i^H = 50,100$) is the $\chi^2_{PLE1}$ test, followed by $Q_{WLS}$ and $\chi^2_{Con}$, respectively. When the number of studies is moderate to large ($k \geq 16$) and sample size is large (code 10 in graph means $n_i^0, n_i^H = 100,500$), $\chi^2_{PLE1}$ test performs best with the highest power, followed by $\chi^2_{Con}$ and $Q_{WLS}$.

4. Conclusion

The above results found that five tests cannot control type I error rates when sample sizes are very small ($n_i^0, n_i^H = 5$); that is, a need to find tests to meet this gap is still further investigated. However, best performance of $\chi^2_{PLE1}$ when $n_i^0, n_i^H = 50,100$ and $k \geq 16$ is very interesting, $\chi^2_{PLE1}$ yields the better test than the conventional $Q_{WLS}$ test. Therefore, the idea of replacement of profile likelihood estimates into the variance formulas of logarithm of diagnostic odds ratios works well. Furthermore, in cases of large sample sizes $n_i^0, n_i^H = 500$ for $k \geq 16$, the profile $\chi^2_{PLE1}$ test performs well with the close same power as $Q_{WLS}$ and $\chi^2_{Con}$.

Acknowledgements

This study was partially supported for publication by the China Medical Board (CMB), Faculty of Public Health, Mahidol University, Bangkok, Thailand.

References