Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n1p75

Post-hoc comparison tests for odds ratios
By Yilmaz, Altunay

Published: 20 May 2022

This work is copyrighted by Università del Salento, and is licensed under a Creative Commons Attribuzione - Non commerciale - Non opere derivate 3.0 Italia License.
For more information see:
http://creativecommons.org/licenses/by-nc-nd/3.0/it/

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 01, May 2022, 75-94
DOI: 10.1285/i20705948v15n1p75

Post-hoc comparison tests for odds ratios

Ayfer Ezgi Yilmaz*and Serpil Aktas Altunay

Hacettepe University
Department of Statistics
Ankara, Turkey

Published: 20 May 2022

The null hypothesis of homogeneity of odds ratio across the strata for a 2x2 tables is tested by several tests. When the null hypothesis is rejected, it means that at least one of the odds ratios significantly differs from others. Post-hoc tests are used after the null hypothesis of equality of groups is rejected Those tests aim to reveal the true differences between groups. In this paper, we propose three post-hoc tests that control family-wise type-I error. The first and second approaches are based on the difference between the two odds ratios. The third approach uses the modification of the Breslow-Day test. These tests provide the homogeneous groups by their odds ratios and enables to calculate the common odds ratio. The suggested tests are applied to several COVID-19 data sets and the results are discussed. The proposed methods can also be used to compare different risk factors on a certain event or outcome.

keywords: 2x2 tables, common odds ratio, post-hoc tests, homogeneity of odds ratios, COVID-19 data.

1 Introduction

Categorical scales measure the attitudes and opinions in social or behavioral sciences and categorical variables have measurement scales consisting of a set of categories (Agresti, 2002). Associations between two categorical variables are examined across two or more variables structured in contingency tables (Lawal, 2003). These tables can sometimes

*Corresponding author: ezgiyilmaz@hacettepe.edu.tr

©Università del Salento
ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

come from a single study that is stratified by some factor. The goal is usually to be able to combine the tables in order to have unified information across the tables. In such cases, meta-analysis is used to investigate the k independent but similar studies conducted at different centers (Borenstein et al., 2009). Xiao et al. (2021) suggested to use odds ratios when reporting the clinical study results. Aggregation of odds ratios calculated from a fourfold table over the k stratum might be required for many purposes. When the test of homogeneity of common odds test is rejected that means at least one of the odds ratios differs from others. In order to find out from which group the difference arises, similar methods to multiple comparisons methods (post-hoc) tests are needed. Chu et al. (2020) discussed that availability of the multiple comparison tests to estimate risk ratios from odds ratios and applicability to published studies included in meta-analyses when the original study data are inaccessible. The omnibus test increases the probability of type-I error. For instance, the pattern of difference between means is analyzed by some pairwise comparisons, when the ANOVA hypothesis is rejected. In practical applications of categorical data analysis, odds ratios are calculated and combined via meta-analysis, but interpreting the odds ratios for the multiple comparisons contexts is not studied well in practice.

In this article, we focus on the analysis of COVID-19 data when the odds ratios are heterogeneous in meta-analysis studies. In order to detect the differences between the heterogeneous odds ratios, we discuss the necessity of a multiple comparisons tests. The contributions of this study are that we i) propose Breslow-Day-based and Chi-squared-based LSD tests and the adjusted Breslow-Day test in multiple comparison tests of heterogeneous odds ratios, ii) demonstrate the importance of heterogeneous odds ratios in COVID-19 studies and discuss the use of multiple comparison tests to get reliable results.

The procedures for testing the homogeneity of odds ratios over k stratum in 2-by-2 contingency tables will be given in Section 2. Proposed tests are given in detail in Section 3 and, the proposed tests are applied to several COVID-19 data sets in Section 4. The results and conclusions are discussed in Sections 5 and 6.

2 Methods

Consider a meta-analysis of k similar and independent studies. When the studies have a dichotomous (binary) outcome, the results of each study can be presented in a 2-by-2 table (Table 1).

The odds ratio (OR) is a measure that measures whether the odds of a certain event or outcome is the same for two groups. The OR evaluates whether the odds of a certain event or outcome is the same for the exposed and unexposed groups. The above 2-by-2 table may be partitioned according to one or more variables into several 2-by-2 tables. Each single table is referred to as a "stratum". The calculation of the OR is straightforward as,

$$OR = \frac{p_1(1-p_1)}{p_2(1-p_2)}, \tag{1}$$

Table 1: A 2x2 cross-classification table.

	Outcome		
Group	Event	No event	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	a+c	b+d	n

where $p_1 = a/b$ represents the odds that the event of interest in the exposed group; $p_2 = c/d$ represents the odds that the event of interest in the unexposed group. Therefore the odds ratio for ith table over the k strata is calculated as,

$$OR_i = \frac{a_i d_i}{b_i c_i}. \tag{2}$$

The researchers wish to compare the odds ratios of several 2-by-2 tables by the following null hypothesis,

$H_0 : OR_1 = OR_2 = ... = OR_k$
H_A: At least one differs from the others.

This hypothesis yields that odds ratios are all equal to one against the alternative hypothesis that at least one odds ratio is different from unity. Some tests such as, The Mantel-Haenszel, Breslow-Day, Peto and Woolf chi-square tests this null hypothesis.

Mantel and Haenzsel (MH) estimate for an OR is defined as

$$\hat{OR}_{MH} = \frac{\sum_{i=1}^{k} \frac{a_i d_i}{n_i}}{\sum_{i=1}^{k} \frac{b_i c_i}{n_i}}, \tag{3}$$

where a_i, b_i, c_i, and d_i are the numbers of participants in the cells of the two-by-two table in the ith stratum of the confounding variable, and n_i represents the number of participants in the ith stratum. The MH test statistics is (Mantel and Haenszel, 1959)

$$\chi^2_{MH} = \frac{[A - E(A)]^2}{V(A)}, \tag{4}$$

where $A = \sum_{i=1}^{k} a_i$, the expected frequency is

$$E(A) = \sum_{i=1}^{k} \frac{(a_i + b_i)(a_i + c_i)}{n_i},$$

the variance of the first cell frequency is

$$V(A) = \sum_{i=1}^{k} \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2(n_i - 1)}.$$

Yusuf et al. (1985) proposed an alternative method to the usual Mantel-Haenszel method for pooling odds ratios across the strata of fourfold tables. This method is called Peto method. The pooled odds ratio is

$$\hat{OR}_{Peto} = exp\left[\frac{\sum_{i=1}^{k}(Oi - E_i)]}{\sum_{i=1}^{k} V_i}\right], \tag{5}$$

where the observed (O_i) and expected (E_i) number of events and its variance (V_i) at the ith stratum are

$$E_i = \frac{(a_i + b_i)(a_i + c_i)}{n_i},$$

$$V_i = \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2(n_i - 1)}.$$

Peto's method can be used to combine studies with dichotomous outcome data with studies using time-to-event analysis. The Peto test statistics is

$$\chi^2_{Peto} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{V(A)} - \frac{\left[\sum_{i=1}^{k}(Oi - E_i)\right]^2}{\sum_{i=1}^{k} V_i}, \tag{6}$$

with the degrees of freedom $df = k - 1$.

Breslow-Day Test for homogeneity of the OR is calculated to test of homogeneity of the odds ratios across strata to investigate if all k strata have the same OR (Breslow and Day, 1980; Breslow, 1996). The Breslow-Day test statistics is

$$\chi^2_{BD} = \sum_{i=1}^{k} \frac{\left[a_i - \hat{\mu}_i(\hat{OR}_{MH})\right]^2}{\hat{\sigma}^2(\hat{OR}_{MH})}, \tag{7}$$

$$\hat{\sigma}^2(\hat{OR}_{MH}) = \left[\frac{1}{\hat{\mu}_i} + \frac{1}{a_i + b_i - \hat{\mu}_i} + \frac{1}{a_i + c_i - \hat{\mu}_i} + \frac{1}{d_i - a_i - \hat{\mu}_i}\right]^{-1}, \tag{8}$$

where $\hat{\mu}_i(\hat{OR})$ and $\hat{\sigma}^2(\hat{OR})$ are the expected value and variance of a_i, under the null hypothesis of homogeneity of odds. Breslow-Day formula uses the Mantel-Haenszel OR to generate the expected values by conditional maximum likelihood method.

Tarone's adjustment is a special case of Tarone (1985) corrected of the MH odds ratio for the test of homogeneity. Tarone's test under the null hypothesis of homogeneity of OR has distributed chi-square with $df = k - 1$.

$$\chi^2_{Tarone} = \chi^2_{BD} - \frac{\left[\sum_{i=1}^{k} a_i - \sum_{i=1}^{k} \hat{\mu}_i(\hat{OR}_{MH})\right]^2}{\sum_{i=1}^{k} \hat{\sigma}^2(\hat{OR}_{MH})}. \tag{9}$$

Woolf (1955) statistic under the null hypothesis of homogeneity of OR has distributed chi-square with $df = k - 1$.

$$\chi^2_{Woolf} = \sum_{i=1}^{k} w_i [ln(OR_i)]^2 - \frac{[w_i ln(OR_i)]^2}{\sum_{i=1}^{k} w_i}, \tag{10}$$

where the weights are $w_i = \left[\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}\right]^{-1}$.

The methods discussed above are used to determine whether there are any statistically significant differences between independent or unrelated odds ratios calculated from 2-by-2 studies.

3 Post-Hoc Methods

Homogeneity of odds ratio tests are the omnibus test statistics that the alternative hypothesis indicates the heterogeneity of odds ratios. The alternative hypothesis does not say which specific groups are statistically significantly different from each other and, we only know that at least two groups are different. In meta-analysis, determining which of these groups differ from each other is important in many aspects. This comparison can be done via post-hoc tests. Type-I error correction can be used for post-hoc comparison. In this study, we propose three post-hoc tests where the odds ratios are heterogeneous. These tests control the family-wise type error that explains the probability of making at least one type-I error.

3.1 The adjustment methods

The Bonferroni and Dunn-Sidak adjustments consists of calculating a new significance level to keep the family-wise type-I error at (a) (Bonferroni, 1936; Dunn, 1961; Šidák, 1967; Félix and Menezes, 2018). Suppose k is the number of groups (strata) which are multiple compared and the number of hypotheses tested is $l = \frac{k(k-1)}{2}$. The Bonferroni adjusted p-value is

$$a' = a/l.$$

The Dunn-Sidak adjusted p-value is

$$a' = 1 - (1 - a)^{1/l}.$$

3.2 Proposed methods

We propose three post-hoc comparison approaches for the heterogeneous odds ratios. The first and second approaches are based on the difference between the two odds ratios. The third approach uses the modification of the Breslow-Day test.

3.2.1 BD-based LSD test method

Assume that OR_i be the odds ratios at strata i and OR_j be the odds ratios at strata j. Let δ be the difference between these two natural logarithms of odds ratios as

$$\delta = |ln(OR_i) - ln(OR_j)|. \tag{11}$$

The common variance is

$$Var(\hat{\delta}) = \frac{\sum_{i=1}^{k} \hat{\sigma}^2(\hat{OR}_{MH})}{k}, \tag{12}$$

where $\hat{\sigma}^2(\hat{OR}_{MH})$ is defined in Equation 8. We wish to test whether δ is statistically significant or not is tested with the null hypothesis $H_0 : \delta = 0$ against $H_A : \delta \neq 0$. Hence, the Z statistic testing the null hypothesis is

$$Z = \frac{\hat{\delta}}{\sqrt{Var(\hat{\delta})}} \tag{13}$$

and it follows the standard normal distribution with Z(0,1). More generally, the Maximum Likelihood Estimate divided by its standard error can be used as a test statistic for the null hypothesis that the difference value of the odds ratios equals zero. From this point of view, a critical region can be constructed based on differences of OR's ($ORDIF$) as below:

$$ORDIF = Z_{a/2}\sqrt{Var(\hat{\delta})}. \tag{14}$$

The null hypothesis is rejected if the difference is greater and equal to the critical value, as

$$\delta \geq ORDIF.$$

The idea behind the BD-based LSD test method comes from the observation that, when the null hypothesis is true, the value of $ORDIF$ evaluating the difference between stratum i and stratum j.

3.2.2 Chi-square-based LSD test method

The second approach is based on the first method in which the expected frequencies are based on the chi-square approach. The excepted values ($\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$, $\hat{d}_i$) are calculated from ordinary chi-square. The standard error for stratum i is

$$SE_i = \left[\frac{1}{\hat{a}_i} + \frac{1}{\hat{b}_i} + \frac{1}{\hat{c}_i} + \frac{1}{\hat{c}_i}\right]^{-0.5}. \tag{15}$$

Then, all steps used in the first method are followed to calculate the $ORDIF$.

3.2.3 Adjusted Breslow-Day test

The third approach is based on the Breslow-Day test.

$$\chi^2_{BD} = \frac{\left[a_i - \hat{\mu}_i(\hat{OR}_{MH})\right]^2}{\hat{\sigma}^2_i(\hat{OR}_{MH})} + \frac{\left[a_j - \hat{\mu}_j(\hat{OR}_{MH})\right]^2}{\hat{\sigma}^2_j(\hat{OR}_{MH})}, \tag{16}$$

where the expected frequencies are calculated based on overall MH. The calculated chi-square value is compared with the chi-square with $df = 1$.

4 COVID-19 Data

Three approaches for post-hoc comparison tests given in Section 3 are applied to COVID-19 data sets. These data sets are directly taken from Wu et al. (2020), Yuan et al. (2020), Zhou et al. (2020), Ruan et al. (2020), Wang et al. (2020), and Yang et al. (2020).

Table 2: Demographics and clinical characteristics of COVID-19 positive patients.

	Wu *et al.* (2020)*		Yuan *et al.* (2020)		Zhou *et al.* (2020)		Ruan *et al.* (2020)		Wang *et al.* (2020)		Yang *et al.* (2020)**	
Variables	Died	Survived	Died	Survived	Died	Survived	Died	Survived	Died	Survived	Died	Survived
	($n = 44$)	($n = 40$)	($n = 10$)	($n = 17$)	($n = 54$)	($n = 137$)	($n = 68$)	($n = 82$)	($n = 19$)	($n = 88$)	($n = 32$)	($n = 20$)
Gender(Male)	29	31	4	8	38	81	49	53	16	41	21	14
	(65.9%)	(77.5%)	(40.0%)	(47.1%)	(70.4%)	(59.1%)	(72.1%)	(64.6%)	(84.2%)	(46.6%)	(65.6%)	(70.0%)
Hypertension	16	7	5	0	26	32	29	23	10	16	-	-
	(36.4%)	(17.5%)	(50.0%)	(0%)	(48.1%)	(23.4%)	(42.6%)	(28.0%)	(52.6%)	(18.2%)	-	-
Cardiac disease	4	1	3	0	13	2	13	0	7	6	3	2
	(9.1%)	(2.5%)	(30.0%)	(0%)	(24.1%)	(1.5%)	(19.1%)	(0%)	(36.8%)	(6.8%)	(9.4%)	(10%)
Diabetes	11	5	6	0	17	19	12	13	5	6	7	2
	(25.0%)	(12.5%)	(60.0%)	0%	(31.5%)	(13.9%)	(17.6%)	(15.9%)	(26.3%)	(6.8%)	(21.9%)	(10%)
Cerebrovascular	-	-	1	0	-	-	7	5	3	3	7	0
	-	-	(10.0%)	(0%)	-	-	(10.3%)	(6.1%)	(15.8%)	(3.4%)	(21.9%)	(0%)
COPD$^+$	-	-	-	-	4	2	2	1	1	2	2	2
	-	-	-	-	(7.4%)	(1.5%)	(2.9%)	(1.2%)	(5.3%)	(2.3%)	(6.3%)	(10%)
Chronic kidney injury	-	-	-	-	2	0	2	0	1	2	-	-
	-	-	-	-	(3.7%)	(0%)	(2.9%)	(0%)	(5.3%)	(2.3%)	-	-
Malignancy	-	-	0	1	0	2	2	1	-	-	1	1
	-	-	(0%)	(5.9%)	(0%)	(1.5%)	(2.9%)	(1.2%)	-	-	(3.1%)	(5%)
Chronic liver injury	-	-	-	-	-	-	1	3	1	5	-	-
	-	-	-	-	-	-	(1.5%)	(3.7%)	(5.3%)	(5.7%)	-	-

* Patients with ARDS; ** Patients in ICU; + Chronic obstructive pulmonary disease

The COVID-19 positive patients are classified as either died or survived according to some demographics, clinical characteristics, signs, and symptoms. The number of patients and its percentage (%) are summarized in Table 2 and Table 3.
Table 4 presents the odds ratios [95 %CI] of death between the presence and absence of clinical characteristics, signs, and symptoms of COVID-19. Those with having some clinical characteristics are said to be at high-risk death from COVID-19 positive.

We calculated the common ORs with their confidence intervals for ORs across six strata for some demographics, clinical characteristics, signs, and symptoms (Table 5).

Table 3: Signs and symptoms of COVID-19 positive patients.

Variables	Wu *et al.* (2020)*		Yuan *et al.* (2020)		Zhou *et al.* (2020)		Ruan *et al.* (2020)		Wang *et al.* (2020)		Yang *et al.* (2020)**	
	Died ($n = 44$)	Survived ($n = 40$)	Died ($n = 10$)	Survived ($n = 17$)	Died ($n = 54$)	Survived ($n = 137$)	Died ($n = 68$)	Survived ($n = 82$)	Died ($n = 19$)	Survived ($n = 88$)	Died ($n = 32$)	Survived ($n = 20$)
Fever	39 (88.6%)	39 (97.5%)	6 (60.0%)	15 (88.2%)	51 (94.4%)	129 (94.2%)	59 (86.8%)	68 (82.9%)	19 (100%)	85 (96.6%)	31 (96.9%)	20 (100%)
Cough	33 (75.0%)	35 (87.5%)	5 (50.0%)	11 (64.7%)	39 (72.2%)	112 (81.8%)	51 (75.0%)	59 (72.0%)	11 (57.9%)	56 (63.6%)	25 (78.1%)	15 (75.0%)
Sputum	- -	- -	- -	- -	14 (25.9%)	30 (21.9%)	29 (42.6%)	28 (34.1%)	- -	- -	- -	- -
Fatigue	- -	- -	- -	- -	15 (27.8%)	29 (21.2%)	15 (22.1%)	22 (26.8%)	14 (73.7%)	55 (62.5%)	14 (43.8%)	4 (20.0%)
Myalgia	15 (34.1%)	12 (30.0%)	1 (10.0%)	2 (11.8%)	8 (14.8%)	21 (15.3%)	9 (13.2%)	10 (12.2%)	5 (26.3%)	28 (31.8%)	4 (12.5%)	2 (10.0%)
Dyspnea	29 (65.9%)	21 (52.5%)	10 (100%)	1 (5.9%)	- -	- -	59 (86.8%)	51 (62.2%)	15 (78.9%)	20 (22.7%)	21 (65.6%)	12 (60.0%)
Respiratory failure	- -	- -	- -	- -	53 (98.1%)	50 (36.5%)	58 (85.3%)	13 (15.9%)	- -	- -	- -	- -
ARDS^{+}	- –	- -	10 (100%)	1 (5.9%)	50 (92.6%)	7 (5.1%)	55 (80.9%)	7 (5.5%)	- -	- -	26 (81.3%)	9 (45.0%)
Acute kidney injury	- -	- -	- -	- -	27 (50.0%)	1 (0.7%)	21 (30.9%)	2 (2.4%)	- -	- -	12 (37.5%)	3 (15.0%)
Headache	- -	- -	- -	- -	- -	- -	- -	- -	0 (0%)	7 (8.0%)	2 (6.3%)	1 (5.0%)
Diarrhea	- -	- -	- -	- -	2 (3.7%)	7 (5.1%)	- -	- -	4 (21.1%)	3 (3.4%)	- -	- -
Vomiting	- -	- -	- -	- -	3 (5.6%)	4 (2.9%)	- -	- -	1 (5.3%)	2 (2.3%)	1 (3.1%)	1 (5.0%)

* Patients with ARDS; ** Patients in ICU; + Acute respiratory distress syndrome

These ORs indicate the relationship between a dichotomous risk factor and a dichotomous outcome. Testing the homogeneity of the six ORs is performed with Mantel-Haenszel, Peto, Breslow-Day, Tarone, and Woolf tests.

5 Results

The homogeneity tests indicate that the homogeneity of the odds ratios for all the variables except for cardiac disease, diabetes, dyspnea, ARDS, acute kidney injury, and diarrhea ($p>0.05$; Table 5). There is no statistically significant risk of death between males and females; between the patients with and without COPD, malignancy, chronic liver injury, fever, cough, sputum, fatigue, myalgia, headache, and vomiting ($p>0.05$; Table 5). It means there is no association between these variables and the mortality.

Similarly, the homogeneity tests indicate that the odds ratios were homogeneous ($p>0.05$) and the Mantel-Haenszel odds ratio is found statistically significant for hypertension, cerebrovascular disease, chronic kidney injury, and respiratory failure ($p<0.05$; Table 5). Patients with hypertension are more likely to die 2.92 times more than the other patients. Patients with cerebrovascular disease are more likely to die 3.75 times more than the other patients. Patients with chronic kidney injury is likely to die 9.90 times more than the other patients. Patients with respiratory failure is likely to die 45.04 times more than the other patients.

The homogeneity tests reveal that the odds ratios were heterogeneous for cardiac disease, diabetes, dyspnea, ARDS, acute kidney injury, and diarrhea ($p<0.05$; Table 5). Thus, the methods for multiple comparisons are summarized in Tables 6-9. The classical Peto, Breslow-Day, Tarone, and Wolf tests are applied to compare two odds ratios. The results of these tests are compared with the Bonferroni and Dunn-Sidak adjusted p-values (a'). The proposed methods are also applied to the data.

The differences and significance of all pairwise comparisons by the cardiac disease are shown in Table 6. These values indicate that which group is the most different among the stratum. Both BD-based and Chi-square-based LSD test methods imply that statistically significant differences between the log odds ratios of Yuan et al. (2020) and Yang et al. (2020), Zhou et al. (2020) and Yang et al. (2020), Ruan et al. (2020) and Yang et al. (2020) studies for cardiac disease ($p<0.05$; Table 6). There is no statistically significant difference between the odds ratios according to Bonferroni and Dunn-Sidak adjusted significance level. The adjusted Breslow-Day test gives statistically significant differences between the seven pairs of odds ratios for cardiac disease ($p<0.05$; Table 6). In general, the odds ratio for the cardiac disease of Yang et al. (2020) is statistically different and lower than others and the odds ratio of Ruan et al. (2020) is statistically higher than others. When the MH odds ratio is re-calculated with the other four studies, the homogeneity tests exhibit that the odds ratios are homogeneous ($p>0.05$) and Mantel-Haenszel odds ratio is found to be statistically significant ($p<0.05$). Consequently, patients with cardiac disease are expected to die 11.59 times more than the other patients.

Table 7 gives the post-hoc comparison results only for those patients who have diabetes and no diabetes. Both the BD-based and Chi-square-based LSD test methods points out that statistically significant differences between the log odds ratios of Yuan et al. (2020) and Ruan et al. (2020) studies for diabetes ($p<0.05$; Table 7. There is a statistically significant difference between the odds ratios of Yuan et al. (2020) and Ruan et al. (2020) for diabetes according to Bonferroni and Dunn-Sidak adjusted significance level. The adjusted Breslow-Day test gives that statistically significant differences between the six pairs of odds ratios for diabetes ($p<0.05$; Table 7). The results show that the odds ratio for the diabetes of Yuan et al. (2020) is higher than the other studies. If the MH odds ratio is calculated again with the other five studies, the homogeneity tests show that the odds ratios were homogeneous ($p>0.05$) and the Mantel-Haenszel odds ratio is found to be statistically significant ($p<0.05$). In conclusion, the patients with diabetes are more likely to die 2.19 times more than the other patients.

Table 8 gives the odds ratio post-hoc test results only for patients with or without dyspne. Both the BD-based and Chi-square-based LSD test methods implies that statistically significant differences between the log odds ratios of Wu et al. (2020) and Yuan et al. (2020), Yuan et al. (2020) and Ruan et al. (2020), Yuan et al. (2020) and Yang et al. (2020) studies for dyspnea ($p<0.05$; Table 8). The adjusted Breslow-Day test reveals that statistically significant differences between the eight pairs of odds ratios for dyspnea ($p<0.05$; Table 8). The results show that the odds ratio for dyspnea of Yuan et al. (2020) is statistically different and higher than the others, followed byWang et al. (2020). When the MH odds ratio is re-calculated with the other three studies,

the homogeneity tests gives that the odds ratios are homogeneous (p>0.05) and the Mantel-Haenszel odds ratio was found statistically significant (p<0.05). Consequently, it can be said that the patients who have dyspnea are expected to die 2.35 times more than the other patients.

Table 9 gives the odds ratio post-hoc test results only for patients with or without ARDS. BD-based LSD test method indicate statistically significant differences between the log odds ratios of Yuan et al. (2020) and Yang et al. (2020), Zhou et al. (2020) and Yang et al. (2020) studies for ARDS (p<0.05; Table 9). There are statistically significant differences between the odds ratios of Zhou et al. (2020) and Yang et al. (2020), Ruan et al. (2020) and Yang et al. (2020) studies for ARDS according to Bonferroni and Dunn-Sidak adjusted significance level. The adjusted Breslow-Day test marks statistically significant differences between all pairs of odds ratios for ARDS (p<0.05; Table 9). The results show that the odds ratio of Zhou et al. (2020), then Yuan et al. (2020) is higher; the odds ratio of Yang et al. (2020) is lower than the others.

Table 10 gives the odds ratio post-hoc test results only for patients with or without acute kidney injury. There is a statistically significant difference between the odds ratios of Zhou et al. (2020) and Yang et al. (2020) studies for acute kidney injury according to Bonferroni and Dunn-Sidak adjusted significance level. The results show that the odds ratio for acute kidney injury of Zhou et al. (2020) is statistically higher than the others, followed by Ruan et al. (2020). The lowest odds ratio is observed in Yang et al. (2020) study.

We can illustrate the suggested methods by using acute kidney injury data in Table 3. The odds ratios of acute kidney injury in Zhou et al. (2020), Ruan et al. (2020), and Yang et al. (2020) studies can be calculated as follows (see Table 4).

$$\hat{\theta}_{Zhou} = \frac{27 \times 136}{1 \times 27} = 136.00, \quad \hat{\theta}_{Ruan} = \frac{21 \times 80}{2 \times 47} = 17.87, \quad \hat{\theta}_{Yang} = \frac{12 \times 17}{3 \times 20} = 3.40,$$

and the natural logarithm transformed odds ratios are calculated as

$$ln(\hat{\theta}_{Zhou}) = 4.913, \quad ln(\hat{\theta}_{Ruan}) = 2.883, \quad ln(\hat{\theta}_{Yang}) = 1.224.$$

The homogeneity test results in Table 5 indicate that the non-homogeneity of the odds ratios for acute kidney injury ($p < 0.05$). Thus, the post-hoc tests are calculated (Table 10). The difference between the log-odds ratios are calculated as

$$\begin{aligned} \delta_{12} &= ln(\hat{\theta}_{Zhou}) - ln(\hat{\theta}_{Ruan}) = 2.029, \\ \delta_{13} &= ln(\hat{\theta}_{Zhou}) - ln(\hat{\theta}_{Yang}) = 3.689, \\ \delta_{22} &= ln(\hat{\theta}_{Ruan}) - ln(\hat{\theta}_{Yang}) = 1.659. \end{aligned}$$

The classical calculations of Peto, Breslow-Day, Tarone, and Wolf tests given in Equations 6–10 are used to compare two odds ratios. The ORDIF value of the BD-based LSD test method is calculated as follows.

$\hat{\sigma}^2(\hat{OR}_{MH})$, the variance of a_i under the assumption of homogeneous odds ratios, is calculated as

Study	$\hat{\mu}_i$	$(a_i + b_i - \hat{\mu}_i)$	$(a_i + c_i - \hat{\mu}_i)$	$(d_i - a_i - \hat{\mu}_i)$	$\hat{\sigma}^2(\hat{OR}_{MH})$
Zhou et al. (2020)	22.669	5.331	31.331	131.669	3.687
Ruan et al. (2020)	21.000	2.000	47.000	80.000	1.720
Yang et al. (2020)	14.150	0.850	17.850	19.150	0.738

The common variance is calculated as $Var(\hat{\delta}) = (3.686 + 1.719 + 0.737)/3 = 2.048$. Then, the ORDIF value of the BD-based LSD test method is calculated as $ORDIF = 1.96\sqrt{2.048} = 2.805$. The difference between the log-odds ratios are compared with the ORDIF value. BD-based LSD test method reveals statistically significant differences between the log odds ratios of Zhou et al. (2020) and Yang et al. (2020) studies for acute kidney injury (Table 10).

$$\delta_{12} = ln(\hat{\theta}_{Zhou}) - ln(\hat{\theta}_{Ruan}) = 2.029 < ORDIF = 2.805 \quad (H_0 \textit{ is not rejected}),$$
$$\delta_{13} = ln(\hat{\theta}_{Zhou}) - ln(\hat{\theta}_{Yang}) = 3.689 > ORDIF = 2.805 \quad (H_0 \textit{ is rejected}),$$
$$\delta_{22} = ln(\hat{\theta}_{Ruan}) - ln(\hat{\theta}_{Yang}) = 1.659 < ORDIF = 2.805 \quad (H_0 \textit{ is not rejected}).$$

The ORDIF value of the Chi-square-based LSD test method is calculated as follows. The variance for stratum i is calculated as following steps. Here, $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$, and $\hat{d}_i$ are the expected values calculated from the ordinary chi-square.

Study	$\hat{a}_i$	$\hat{b}_i$	$\hat{c}_i$	$\hat{d}_i$	Var_i
Zhou et al. (2020)	7.916	20.084	46.084	116.916	4.846
Ruan et al. (2020)	10.427	12.573	57.573	69.427	4.826
Yang et al. (2020)	9.231	5.769	22.769	14.231	2.526

The common variance is calculated as $Var(\hat{\delta}) = (4.846 + 4.826 + 2.526)/3 = 4.066$. Then, the ORDIF value of the chi-squared-based LSD test is calculated as $ORDIF = 1.96\sqrt{4.066} = 3.952$. The difference between the log-odds ratios are compared with the ORDIF value. The chi-squared-based LSD test method reveals no statistically significant differences between the log odds ratios of the three studies (Table 10).

$$\delta_{12} = ln(\hat{\theta}_{Zhou}) - ln(\hat{\theta}_{Ruan}) = 2.029 < ORDIF = 3.952 \quad (H_0 \textit{ is not rejected}),$$
$$\delta_{13} = ln(\hat{\theta}_{Zhou}) - ln(\hat{\theta}_{Yang}) = 3.689 < ORDIF = 3.952 \quad (H_0 \textit{ is not rejected}),$$
$$\delta_{22} = ln(\hat{\theta}_{Ruan}) - ln(\hat{\theta}_{Yang}) = 1.659 < ORDIF = 3.952 \quad (H_0 \textit{ is not rejected}).$$

The adjusted Breslow-Day test is calculated as follows.

Study	a_i	$\hat{\mu}_i$	$(a_i - \hat{\mu}_i)^2$	$\hat{\sigma}^2(\hat{OR}_{MH})$
Zhou et al. (2020)	27	22.669	18.758	3.687
Ruan et al. (2020)	21	21.000	0.000	1.720
Yang et al. (2020)	12	14.150	4.624	0.738

The test statistics for each pair of the studies are calculated as following. χ^2_{BD} values are compared with $\chi^2(1, 0.05) = 3.841$. The adjusted Breslow-Day test indicates statistically significant differences between all pairs of odds ratios for acute kidney injury (Table 10).

- Zhou et al. (2020) - Ruan et al. (2020):
  $\chi^2_{BD} = \frac{18.758}{3.687} + \frac{0.000}{1.720} = 5.087 > \chi^2(1, 0.05) = 3.841$ (H_0 is rejected).
- Zhou et al. (2020) - Ruan et al. (2020):
  $\chi^2_{BD} = \frac{18.758}{3.687} + \frac{4.624}{0.738} = 11.358 > \chi^2(1, 0.05) = 3.841$ (H_0 is rejected).
- Ruan et al. (2020) - Yang et al. (2020):
  $\chi^2_{BD} = \frac{0.000}{1.720} + \frac{4.624}{0.738} = 6.270 > \chi^2(1, 0.05) = 3.841$ (H_0 is rejected).

The results show that the odds ratio for diarrhea of Wang et al. (2020) was statistically different and higher than the Zhou et al. (2020)'s.

6 Discussions

When the test of homogeneity of odds ratios across the independent studies is rejected and gets a significant result, multiple comparisons methods are needed to reveal that which group (groups) make the true difference. In this study, three different post-hoc comparison approaches are suggested for homogeneity of the OR's. For this purpose, we used six COVID-19 China data sets available on the web. First, we apply the homogeneity of odds ratios tests. When the results of these tests are significant ($p<0.05$), we conclude that the odds ratios are heterogeneous and this means that at least one OR differ from others. Then, we apply the proposed methods and the classical tests for each pair. We use Bonferroni and Dunn-Sidak adjustments for the classical tests.

The disadvantages of adjustment methods was discussed (Perneger, 1998). The method is a very conservative so it is very difficult to find significant difference and the interpretation of a finding depends on the number of other tests performed. The results in Tables 6–10 show that Bonferroni and Dunn-Sidak adjustments were also very conservative when comparing the odds ratios. The Adjusted Breslow-Day test is the less conservative than the other methods.

When calculating the odds ratios, sampling zero problem occurred for some of the tables. Therefore, 0.5 was added to each of the cells and then the odds ratio was calculated over these adjusted cell counts (Agresti, 2002). The odds ratio in the presence of sparse data is questionable and a common problem is the presence of high odds ratios (ORs). For example, the odds ratio associated with dyspnea in Yuan et al. (2020) study was calculated as 231. To overcome this problem, a log transformation can be used to compare the odds ratios. For the sparse tables, BD-based and Chi-square-based LSD methods are suggested to be used instead of the Adjusted Breslow-Day test.

The proposed methods can also be used to compare different risk factors on a certain event or outcome.

References

Agresti, A. (2002). *Categorical data analysis*. John Wiley & Sons, New York.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, West Sussex.

Breslow, N. E. (1996). Statistics in epidemiology: the case-control study. *Journal of the American Statistical Association*, 91(433):14–28.

Breslow, N. S. and Day, N. E. (1980). The analysis of case-control studies. *Statistical methods in cancer research*, 1.

Chu, T.-C., Ojha, R. P., and VanderWeele, T. J. (2020). Empirical comparison of approaches for odds ratios to risk ratio transformations in meta-analyses of randomized controlled trials with common outcomes. *Annals of epidemiology*, 44:57–59.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American statistical association*, 56(293):52–64.

Félix, V. B. and Menezes, A. F. B. (2018). Comparisons of ten corrections methods for t-test in multiple comparisons via monte carlo study. *Electronic Journal of Applied Statistical Analysis*, 11(1):74–91.

Lawal, B. (2003). *Categorical data analysis with SAS and SPSS applications*. Lawrence Erlbaum Associates, Publishers, New Jersey.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute*, 22(4):719–748.

Perneger, T. V. (1998). What's wrong with bonferroni adjustments. *British Medical Journal*, 316(7139):1236–1238.

Ruan, Q., Yang, K., Wang, W., Jiang, L., and Song, J. (2020). Clinical predictors of mortality due to covid-19 based on an analysis of data of 150 patients from wuhan, china. *Intensive care medicine*, 46(5):846–848.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633.

Tarone, R. E. (1985). On heterogeneity tests based on efficient scores. *Biometrika*, 72(1):91–95.

Wang, D., Yin, Y., Hu, C., Liu, X., Zhang, X., Zhou, S., Jian, M., Xu, H., Prowly, J., Hu, B., et al. (2020). Clinical course and outcome of novel coronavirus covid-19 infection in 107 patients discharged from the wuhan hospital. *Research Square*.

Woolf, B. (1955). On estimating the relation between blood group and disease. *Ann hum genet*, 19(4):251–253.

Wu, C., Chen, X., Cai, Y., Zhou, X., Xu, S., Huang, H., Zhang, L., Zhou, X., Du, C., Zhang, Y., et al. (2020). Risk factors associated with acute respiratory distress

syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, china. *JAMA internal medicine*, 180(7):934–943.

Xiao, M., Chu, H., Cole, S., Chen, Y., MacLehose, R., Richardson, D., and Greenland, S. (2021). Odds ratios are far from" portable": A call to use realistic models for effect variation in meta-analysis. *arXiv preprint arXiv:2106.02673*.

Yang, X., Yu, Y., Xu, J., Shu, H., Liu, H., Wu, Y., Zhang, L., Yu, Z., Fang, M., Yu, T., et al. (2020). Clinical course and outcomes of critically ill patients with sars-cov-2 pneumonia in wuhan, china: a single-centered, retrospective, observational study. *The Lancet Respiratory Medicine*, 8(5):475–481.

Yuan, M., Yin, W., Tao, Z., Tan, W., and Hu, Y. (2020). Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in wuhan, china. *PloS one*, 15(3):e0230548.

Yusuf, S., Peto, R., Lewis, J., Collins, R., and Sleight, P. (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in cardiovascular diseases*, 27(5):335–371.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The lancet*, 395(10229):1054–1062.

Table 4: The odds ratios [95 %CI] of death between the presence and absence of clinical characteristics, signs, and symptoms of COVID-19.

Studies	Wu *et al.* (2020)	Yuan *et al.* (2020)	Zhou *et al.* (2020)	Ruan *et al.* (2020)	Wang *et al.* (2020)	Yang *et al.* (2020)
Demographics and Clinical Characteristics						
Gender(Male)	0.56	0.75	1.64	1.41	6.11*	0.82
	[0.21;1.48]	[0.15;1.48]	[0.84;3.23]	[0.70;2.83]	[1.66;22.49]	[0.25;2.72]
Hypertension	2.69	35.00*	3.05*	1.91	5.00*	-
	[0.97;7.48]	[1.66;738.69][a]	[1.57;5.92]	[0.97;3.77]	[1.75;14.30]	-
Cardiac disease	3.9	16.33 [0.75;356.90][a]	21.40*	40.14*	7.97*	0.93
	[0.42;36.46]	-	[4.64;98.76]	[2.34;689.12][a]	[2.29;27.75]	[0.14;6.12]
Diabetes	2.33	50.56*	2.85*	1.14	4.88*	2.52
	[0.73;7.44]	[2.38;1075.389][a]	[1.35;6.05]	[0.48;2.69]	[1.31;18.18]	[0.47;13.58]
Cerebrovascular	-	5.53	-	1.77	5.31	12.06
	-	[0.20;149.34][a]	-	[0.54;5.84]	[0.98;28.71]	[0.65;223.86][a]
COPD	-	-	5.4	2.46	2.39	0.6
	-	-	[0.96;30.40]	[0.22;27.67]	[0.21;27.78]	[0.08;4.64]
Chronic kidney injury	-	-	13.1	6.2	2.39	-
	-	-	[0.62;277.35][a]	[0.29;131.44][a]	[0.21;27.78]	-
Malignancy	-	0.52	0.5	2.45	-	0.61
	-	[0.02;14.10][a]	[0.02;10.53][a]	[0.22;27.67]	-	[0.04;10.39]
Chronic liver injury	-	-	-	0.39	0.92	-
	-	-	-	[0.04;3.87]	[0.10;8.838]	-
Signs and Symptoms						
Fever	0.20	0.20	1.05	1.35	1.60	0.51
	[0.22;1.79]	[0.029;1.40]	[0.27;4.13]	[0.54;3.34]	[0.08;32.19][a]	[0.02;13.19][a]
Cough	0.43	0.55	0.58	1.17	0.79	1.19
	[0.13;1.37]	[0.11;2.67]	[0.28;1.21]	[0.56;2.43]	[0.29;2.16]	[0.32;4.43]
Sputum	-	-	1.25	1.43	-	-
	-	-	[0.60;2.59]	[0.74;2.78]	-	-
Fatigue	-	-	1.43	0.77	1.68	3.11
	-	-	[0.70;2.95]	[0.36;2.64]	[0.54;5.09]	[0.85;11.41]
Myalgia	1.21	0.83	0.96	1.1	0.77	1.29
	[0.48;3.03]	[0.07;10.55]	[0.40;2.32]	[0.42;2.88]	[0.25;2.34]	[0.21;7.76]
Dyspnea	1.75	231.00*	-	3.99*	12.75*	1.27
	[0.73;4.22]	[8.58;6218.49]a		[1.74;9.15]	[3.80;42.77]	[0.40;4.04]
Respiratory failure	-	-	92.22*	30.79*	-	-
	-	-	[12.37;87.38]	[12.58;75.36]	-	-
ARDS	-	231.00*	232.14*	45.33*	-	5.30*
	-	[8.58;6218.49][a]	[65.13;27.48]	[16.97;121.08]	-	[1.52;18.50]
Acute kidney injury	-	-	136.00*	17.87*	-	3.40
	-	-	[17.72;1044.04]	[4.01;79.66]	-	[0.82;14.08]
Headache	-	-	-	-	0.28	1.27
	-	-	-	-	[0.02;5.09][a]	[0.11;14.95]
Diarrhea	-	-	0.71	-	7.56*	-
	-	-	[0.14;3.55]	-	[1.53;37.21]	-
Vomiting	-	-	1.96	-	2.39	0.61
	-	-	[0.42;9.05]	-	[0.21;27.78]	[0.04;10.39]

a: Calculated with 0.5 added to each cell; *: odds ratio is statistically significant ($p<0.05$).

Table 5: The homogeneity of odds ratios test results for COVID-19 positive patients.

				Test of Homogeneity			
		Mantel-Haenszel		Peto	Breslow-Day	Tarone	Woolf
	Number of Studies	OR [95% CI]	Chi-Squared	Q Statistic	Chi-Squared	Chi-Squared	Q Statistic
Demographics and Clinical Characteristics							
Gender(Male)	6	1.39 [0.96;1.99]	3.101 (0.097)	10.332 (0.066)	10.66 (0.059)	10.66 (0.059)	9.844 (0.080)
Hypertension	5	2.92** [1.97;4.32]	29.406** (p<0.001)	7.906 (0.095)	7.079 (0.132)	7.063 (0.133)	5.085 (0.279)
Cardiac disease	6	9.82** [4.55;21.19]	52.044** (p<0.001)	10.183 (0.070)	13.429* (0.020)	13.265* (0.021)	8.500 (0.131)
Diabetes	6	2.50** [1.61;3.89]	17.028** (p¡0.001)	12.203* (0.032)	11.186* (0.048)	11.146* (0.049)	8.040 (0.154)
Cerebrovascular	4	3.75** [1.49;9.42]	7.550** (0.006)	3.232 (0.357)	3.703 (0.295)	3.701 (0.296)	2.189 (0.534)
COPD	4	2.20 [0.79;6.13]	1.852 (0.174)	3.099 (0.377)	2.849 (0.415)	2.823 (0.420)	2.598 (0.458)
Chronic kidney injury	3	9.90** [1.45;67.55]	4.697** (0.030)	1.269 (0.530)	2.271 (0.321)	2.172 (0.338)	0.749 (0.688)
Malignancy	4	0.73 [0.16, 3.37]	0.001 (0.984)	1.903 (0.593)	2.212 (0.547)	2.212 (0.547)	0.978 (0.807)
Chronic liver injury	2	0.60 [0.12;2.94]	0.079 (0.778)	0.259 (0.611)	0.282 (0.596)	0.281 (0.596)	0.277 (0.599)
Signs and Symptoms							
Fever	6	0.82 [0.45;1.52]	0.207 (0.649)	6.585 (0.253)	6.654 (0.248)	6.654 (0.248)	5.158 (0.397)
Cough	6	0.77 [0.52;1.13]	1.524 (0.217)	3.431 (0.634)	3.439 (0.633)	3.439 (0.633)	3.409 (0.637)
Sputum	2	1.35 [0.83;2.20]	1.113 (0.287)	0.068 (0.794)	0.076 (0.783)	0.076 (0.783)	0.076 (0.783)
Fatigue	4	1.31 [0.85;2.02]	1.21 (0.271)	3.789 (0.285)	3.917 (0.271)	3.917 (0.271)	3.848 (0.278)
Myalgia	6	1.02 [0.65;1.61]	0.001 (0.978)	0.508 (0.992)	0.513 (0.992)	0.513 (0.992)	0.512 (0.992)
Dyspnea	5	3.71** [2.37;5.81]	35.351** (p<0.001)	20.904* (p<0.001)	21.558* (p<0.001)	21.398* (p<0.001)	15.998* (0.003)
Respiratory failure	2	45.04** [19.28;105.20]	127.344** (p<0.001)	0.423 (0.515)	1.206 (0.272)	1.111 (0.292)	0.956 (0.328)
ARDS	6	40.72* [21.87;75.80]	231.544** (p<0.001)	15.167* (0.002)	21.718* (p<0.001)	20.238* (p<0.001)	18.553* (p<0.001)
Acute kidney injury	3	17.87** [7.75;41.23]	82.834** (p<0.001)	15.303* (p<0.001)	11.358* (0.003)	10.585* (0.005)	8.688* (0.013)
Headache	2	0.4 [0.06;2.79]	0.211 (0.646)	0.916 (0.338)	1.569 (0.210)	1.533 (0.216)	0.606 (0.436)
Diarrhea	2	1.95 [0.70;5.42]	1.081 (0.298)	6.247* (0.012)	4.711* (0.030)	4.702* (0.030)	4.178* (0.041)
Vomiting	3	1.63 [0.49;5.38]	0.253 (0.615)	0.684 (0.710)	0.641 (0.726)	0.639 (0.727)	0.605 (0.739)

a: Calculated with 0.5 added to each cell; *: results are statistically significant at p<0.05; **: results are statistically significant at p<0.01.

Table 6: The multiple comparisons test results for odds ratios by cardiac disease.

Studies	Difference of Log-Odds Ratios*	Peto Q Statistic** (p-value)	Breslow-Day Chi-Squared** (p-value)	Tarone Chi-Squared** (p-value)	Woolf Chi-Squared** (p-value)	Breslow-Day 2 Chi-Squared*** (p-value)
Wu-Yuan	1.432	1.298 (0.255)	1.289 (0.256)	1.245 (0.265)	0.543 (0.461)	3.361 (0.067)
Wu-Zhou	1.703	3.160 (0.075)	1.640 (0.200)	1.551 (0.213)	1.518 (0.218)	1.743 (0.187)
Wu-Ruan	2.331	1.294 (0.255)	3.346 (0.067)	3.314 (0.069)	1.596 (0.206)	4.036 (0.045)
Wu-Wang	0.715	1.835 (0.176)	0.318 (0.573)	0.294 (0.588)	0.3 (0.584)	0.809 (0.368)
Wu-Yang	1.432	0.865 (0.352)	0.967 (0.325)	0.967 (0.326)	0.923 (0.337)	9.601 $(0.002)^{e}$
Yuan-Zhou	0.270	0.018 (0.894)	0.338 (0.561)	0.337 (0.561)	0.024 (0.878)	3.702 (0.054)
Yuan-Ruan	0.899	0.147 (0.702)	2.500 (0.114)	2.449 (0.118)	0.176 (0.674)	5.995 $(0.014)^{e}$
Yuan-Wang	0.717	0.009 (0.925)	0.887 (0.346)	0.869 (0.351)	0.179 (0.673)	2.767 (0.096)
Yuan-Yang	2.865^{ab}	3.64 (0.056)	3.721 (0.054)	3.662 (0.056)	2.414 (0.120)	11.559 $(0.001)^{e}$
Zhou-Ruan	0.629	0.727 (0.394)	1.064 (0.302)	1.009 (0.315)	0.146 (0.703)	4.377 $(0.036)^{e}$
Zhou-Wang	0.988	0.109 (0.741)	1.024 (0.311)	1.007 (0.316)	0.962 (0.327)	1.149 (0.284)
Zhou-Yang	3.135^{ab}	7.947 (0.005)	7.944 (0.005)	7.522 (0.006)	6.415 (0.011)	12.234 $(<0.001)^{e}$
Ruan-Wang	1.616	0.161 (0.688)	2.848 (0.091)	2.501 (0.144)	1.041 (0.308)	3.442 (0.064)
Ruan-Yang	3.764^{ab}	4.862 (0.027)	8.809 $(0.003)^{ab}$	8.670 $(0.003)^{ab}$	4.679 (0.031)	12.234 $(<0.001)^{e}$
Wang-Yang	2.147	5.403 (0.020)	3.848 (0.050)	3.639 (0.056)	3.471 (0.062)	9.007 $(0.003)^{e}$

*There is a statistically significant difference between the two log odds ratios according to a: BD-based LSD test method ($ORDIF$=2.184; p<0.05); b: Chi-square-based LSD test method ($ORDIF$=2.599; p<0.05).

** There is a statistically significant difference between the two odds ratios according to c: Bonferroni adjusted significance level (a'=0.0033); d: Dunn-Sidak adjusted significance level (a'=0.0034).

*** There is a statistically significant difference between the two odds ratios according to e: Adjusted Breslow-Day test.

Table 7: The multiple comparisons test results for odds ratios by diabetes.

Studies	Difference of Log-Odds Ratios*	Peto Q Statistic** (p-value)	Breslow-Day Chi-Squared** (p-value)	Tarone Chi-Squared** (p-value)	Woolf Chi-Squared** (p-value)	Breslow-Day 2 Chi-Squared*** (p-value)
Wu-Yuan	3.047	5.426 (0.020)	5.523 (0.019)	5.416 (0.020)	3.399 (0.065)	9.399 $(0.002)^{e}$
Wu-Zhou	0.201	0.250 (0.617)	0.082 (0.775)	0.081 (0.775)	0.082 (0.775)	0.133 (0.715)
Wu-Ruan	0.719	0.908 (0.341)	0.960 (0.327)	0.960 (0.327)	0.952 (0.329)	3.320 (0.068)
Wu-Wang	0.738	1.684 (0.194)	0.700 (0.403)	0.685 (0.408)	0.681 (0.409)	1.031 (0.310)
Wu-Yang	0.077	0.001 (0.989)	0.005 (0.941)	0.005 (0.941)	0.005 (0.941)	0.014 (0.907)
Yuan-Zhou	2.846	4.586 (0.032)	5.449 (0.020)	5.445 (0.020)	3.202 (0.074)	9.504 $(0.002)^{e}$
Yuan-Ruan	3.765^{ab}	0.147 $(0.002)^{cd}$	2.500 $(0.002)^{cd}$	2.449 $(0.002)^{cd}$	0.176 (0.019)	5.995 $(<0.001)^{e}$
Yuan-Wang	2.309	0.992 (0.319)	3.341 (0.068)	3.336 (0.068)	1.895 (0.169)	10.402 $(0.001)^{e}$
Yuan-Yang	2.970	4.437 (0.035)	4.471 $(0.034)^{d}$	4.166 (0.041)	2.835 (0.092)	9.385 $(0.002)^{e}$
Zhou-Ruan	0.920	2.872 (0.090)	2.521 (0.112)	2.516 (0.113)	2.492 (0.114)	3.425 (0.064)
Zhou-Wang	0.537	1.052 (0.305)	0.488 (0.485)	0.486 (0.486)	0.482 (0.487)	1.136 (0.286)
Zhou-Yang	0.124	0.152 (0.697)	0.018 (0.895)	0.017 (0.895)	0.017 (0.895)	3.306 (0.069)
Ruan-Wang	1.457	4.393 (0.036)	3.498 (0.061)	3.487 (0.062)	3.3 (0.069)	4.323 $(0.038)^{e}$
Ruan-Yang	0.796	0.627 (0.428)	0.693 (0.405)	0.693 (0.405)	0.68 (0.410)	3.306 (0.069)
Wang-Yang	0.661	1.318 (0.251)	0.383 (0.536)	0.365 (0.546)	0.368 (0.544)	1.017 (0.313)

*There is a statistically significant difference between the two log odds ratios according to a: BD-based LSD test method ($ORDIF$=3.483; p<0.05); b: Chi-square-based LSD test method ($ORDIF$=3.362; p<0.05).

** There is a statistically significant difference between the two odds ratios according to c: Bonferroni adjusted significance level (a'=0.0033); d: Dunn-Sidak adjusted significance level (a'=0.0034).

*** There is a statistically significant difference between the two odds ratios according to e: Adjusted Breslow-Day test.

Table 8: The multiple comparisons test results for odds ratios by dyspnea.

Studies	Difference of Log-Odds Ratios*	Peto Q Statistic** (p-value)	Breslow-Day Chi-Squared** (p-value)	Tarone Chi-Squared** (p-value)	Woolf Chi-Squared** (p-value)	Breslow-Day 2 Chi-Squared*** (p-value)
Wu-Yuan	4.883^{ab}	5.426 (<0.001)^{cd}	5.523 (<0.001)^{cd}	5.416 (<0.001)^{cd}	3.399 (<0.001)^{cd}	9.399 (<0.001)^{e}
Wu-Ruan	0.823	1.466 (0.226)	1.791 (0.181)	1.791 (0.181)	1.778 (0.182)	2.872 (0.090)
Wu-Wang	1.986	8.107 (0.004)^{cd}	7.13 (0.008)	7.077 (0.008)	6.771 (0.009)	7.128 (0.008)^{e}
Wu-Yang	0.318	0.180 (0.671)	0.185 (0.667)	0.185 (0.667)	0.184 (0.668)	5.538 (0.019)^{e}
Yuan-Ruan	4.060^{ab}	8.154 (0.004)^{cd}	9.257 (0.002)^{cd}	8.989 (0.003)^{cd}	5.490 (0.019)	13.993 (<0.001)^{e}
Yuan-Wang	2.897	1.625 (0.202)	4.474 (0.034)	4.281 (0.039)	2.619 (0.106)	18.249 (0.001)^{e}
Yuan-Yang	5.201^{ab}	12.652 (<0.001)^{cd}	14.401 (<0.001)^{cd}	14.139 (<0.001)^{cd}	8.536 (<0.001)^{cd}	16.659 (<0.001)^{e}
Ruan-Wang	1.163	3.872 (0.049)	2.470 (0.116)	2.430 (0.119)	2.410 (0.121)	4.312 (0.039)^{e}
Ruan-Yang	1.141	2.124 (0.145)	2.507 (0.113)	2.507 (0.113)	2.473 (0.116)	2.723 (0.099)
Wang-Yang	2.304	8.319 (0.004)^{cd}	7.653 (0.006)^{cd}	7.589 (0.006)^{cd}	7.292 (0.007)^{cd}	6.978 (0.008)^{e}

*There is a statistically significant difference between the two log odds ratios according to a: BD-based LSD test method ($ORDIF$=3.726; p<0.05); b: Chi-square-based LSD test method ($ORDIF$=3.933; p<0.05).

** There is a statistically significant difference between the two odds ratios according to c: Bonferroni adjusted significance level (a'=0.0050); d: Dunn-Sidak adjusted significance level (a'=0.0051).

*** There is a statistically significant difference between the two odds ratios according to e: Adjusted Breslow-Day test.

Table 9: The multiple comparisons test results for odds ratios by ARDS.

Studies	Difference of Log-Odds Ratios*	Peto Q Statistic** (p-value)	Breslow-Day Chi-Squared** (p-value)	Tarone Chi-Squared** (p-value)	Woolf Chi-Squared** (p-value)	Breslow-Day 2 Chi-Squared*** (p-value)
Yuan-Zhou	0.005	0.214 (0.643)	0.593 (0.441)	0.593 (0.441)	0.001 (0.998)	12.433 (<0.001)^{e}
Yuan-Ruan	1.628	0.840 (0.359)	2.103 (0.147)	2.070 (0.150)	0.863 (0.353)	4.651 (0.031)^{e}
Yuan-Yang	3.775^{a}	4.593 (0.032)	7.090 (0.008)	6.841 (0.009)	4.413 (0.036)	15.986 (0.001)^{e}
Zhou-Ruan	1.633	6.111 (0.013)	4.064 (0.044)	3.891 (0.049)	3.971 (0.046)	7.886 (0.005)^{e}
Zhou-Yang	3.780^{a}	13.316 (<0.001)^{cd}	18.939 (<0.001)^{cd}	17.038 (<0.001)^{cd}	17.266 (<0.001)^{cd}	19.222 (<0.001)^{e}
Ruan-Yang	2.147	3.846 (0.050)	7.257 (0.007)	7.114 (0.008)	7.000 (0.008)	11.440 (<0.001)^{e}

*There is a statistically significant difference between the two log odds ratios according to a: BD-based LSD test method ($ORDIF$=3.264; p<0.05); b: Chi-square-based LSD test method ($ORDIF$=4.528; p<0.05).

** There is a statistically significant difference between the two odds ratios according to c: Bonferroni adjusted significance level (a'=0.0083); d: Dunn-Sidak adjusted significance level (a'=0.0085).

*** There is a statistically significant difference between the two odds ratios according to e: Adjusted Breslow-Day test.

Table 10: The multiple comparisons test results for odds ratios by acute kidney injury.

Studies	Difference of Log-Odds Ratios*	Peto Q Statistic** (p-value)	Breslow-Day Chi-Squared** (p-value)	Tarone Chi-Squared** (p-value)	Woolf Chi-Squared** (p-value)	Breslow-Day 2 Chi-Squared*** (p-value)
Zhou-Ruan	2.029	7.375 (0.007)^{cd}	2.725 (0.099)	2.471 (0.116)	2.477 (0.116)	5.287 (0.024)^{e}
Zhou-Yang	3.689^{a}	13.613 (<0.001)^{cd}	11.358 (0.001)^{cd}	10.284 (0.001)^{cd}	8.468 (0.004)^{cd}	11.358 (0.001)^{e}
Ruan-Yang	1.659	2.041 (0.153)	2.669 (0.102)	2.656 (0.103)	2.488 (0.115)	6.270 (0.012)^{e}

*There is a statistically significant difference between the two log odds ratios according to a: BD-based LSD test method ($ORDIF$=2.805; p<0.05); b: Chi-square-based LSD test method ($ORDIF$=3.952; p<0.05).

** There is a statistically significant difference between the two odds ratios according to c: Bonferroni adjusted significance level (a'=0.0167); d: Dunn-Sidak adjusted significance level (a'=0.0170).

*** There is a statistically significant difference between the two odds ratios according to e: Adjusted Breslow-Day test.