Optimal Concordant Tests

Zhongxue Chen

Abstract: In meta-analyses, unlike model-based methods, such as fixed- or random-effect models, the p-value combining methods are distribution-free and robust. How to appropriately and powerfully combine p-values obtained from various sources remains an important but challenging topic in statistical inference. For cases where all or a majority of the individual alternative hypotheses have the same but unknown direction, concordant tests based on one-sided p-values can substantially improve the detecting power. However, there exists no test that is uniformly most powerful; therefore, figuring out how to choose a robust and powerful test to combine one-sided p-values for a given data set is desirable. In this paper, we propose and study a class of gamma distribution-based concordant tests. Those concordant tests are optimal under specific conditions. An asymptotically optimal concordant test is also studied. The excellent performances of the proposed tests were demonstrated through a numeric simulation study and real data example.

Keywords: chi-square test; constrained likelihood ratio test; gamma distribution; meta-analysis; uniformly most powerful test

1. Introduction

Due to recent technical developments, larger volume data, including genome-wide genomic data, can be generated rapidly and at a low cost. As a consequence, powerful statistical approaches, including p-value combination tests, are highly desirable for analyzing these data sets. For instance, meta-analyses, when applied to genome-wide association studies (GWASs), have discovered many associated genetic variants that could not be identified from a single GWAS. With the development and application of advanced powerful p-value combination tests, we expect that many more casual genetic variants will be identified through existing data.

In meta-analyses, p-value combination methods are important alternatives to model-based techniques [1,2], especially when the fixed- or random-effect models have a lack of fit [3]. In the literature, many p-value combination methods, although first proposed a long time ago, are still widely used nowadays. These methods include popular ones: Fisher test [4], Pearson test [5], minimal p-test [6], z-test [7], and chi-square test [8,9]. Recently, we studied a class of p-value combination tests based on gamma distribution, which include those popular tests as special cases [10]. However, as noticed by Birnbaum, there exists no uniformly most powerful (UMP) p-value combination test for all conditions [11]. Therefore, robust methods that have good power under many conditions are desirable [12].

Under some conditions, we may know or reasonably assume that all or most of the true effects under the individual alternative hypotheses have the same but unknown direction. In this case, the concordant tests which use one-sided p-values have been shown to be more powerful than those methods based on two-sided p-values, such as the chi-square test [1]. For instance, in GWASs, studies with the same phenotype are usually conducted among several different subpopulations, from which the effect of a causal genetic
variant may have the same direction but various sizes. Then, if the commonly used fixed- and random-effect meta-analyses are applied to combine the results from individual studies, they may lose the power to detect the associated genetic variants. To improve the detecting power, in this paper, we study a class of concordant tests based on gamma distribution, and then an optimal concordant test based on constrained likelihood ratio test (CLRT) is developed.

The rest of the manuscript is organized as follows. In Section 2, we first introduce the Pearson’s concordant test, which was studied by Owen [1] (Section 2.1). In Section 2.2, we describe our proposed generalized concordant tests based on gamma distribution and study their properties as UMP tests under some conditions. The asymptotically UMP concordant test based on CLRT is proposed and studied in Section 2.3. In Section 3, we compare the performances of the proposed concordant tests with some existing methods through a simulation study. In addition, an example of a real data application is illustrated to demonstrate the desired performance of the proposed tests. This paper concludes in Section 4 with a discussion and conclusions.

2. Methods

Suppose we have \( n \) independent studies. For each study, \( i \) (\( i = 1, \cdots, n \)), we can perform a two-sample test for testing the null hypothesis, \( H_{0i}: \mu_i = 0 \), against the alternative hypothesis, \( H_{1i}: \mu_i \neq 0 \), where \( \mu_i \) is the mean difference between two populations. The corresponding \( p \)-value is denoted as \( P_i \) (\( i = 1, \cdots, n \)). Throughout this paper, we assume \( P_i \sim U(0,1) \) under \( H_{0i} \), where \( U(0,1) \) stands for the uniform distribution between 0 and 1.

In a meta-analysis, we consider testing the global null hypothesis, \( H_0 = \bigcap H_{0i} \), vs. the global alternative hypothesis, \( H_1 = \bigcup H_{1i} \), using the \( n \) observed \( p \)-values. For a concordant test, we assume, under the global alternative, that the majority of the non-zero mean differences have the same but unknown direction (positive or negative): \( \mu_i \geq 0 \) and at least one strictly greater than holds, or \( \mu_i \leq 0 \) and at least one strictly less than holds (\( i = 1, \cdots, n \)). Although, under this setting, a two-sided \( p \)-value can still be calculated from each data set and then used in \( p \)-value combination, in this paper, we focus on the concordant tests which use one-sided \( p \)-values, \( P_i \), obtained from the directional alternatives \( \mu_i > 0 \) or \( \mu_i < 0 \) for all \( i \) values (\( i = 1, \cdots, n \)). In general, either the right- or left-sided \( p \)-values can be used, as they will produce the same result. Without loss of generality, we use the right-sided individual \( p \)-values hereafter, unless otherwise specified.

2.1. Pearson’s Concordant Test

Owen [1] revisited and studied the following Pearson’s concordant test, whose test statistic is defined as follows:

\[
Q^C = \max (-2 \sum_{i=1}^{n} \ln(P_i), -2 \sum_{i=1}^{n} \ln(1-P_i)),
\]

where \( P_i \) (\( i = 1, \cdots, n \)) represents the one-sided \( p \)-values obtained from the same directional alternatives. The two components, \(-2 \sum_{i=1}^{n} \ln(P_i)\) and \(-2 \sum_{i=1}^{n} \ln(1-P_i)\), in \( Q^C \) are obtained through the popular Fisher test for combining independent \( p \)-values [4]. Therefore, under the global null hypothesis, each component is the random variable, \( \chi^2_{2n} \), which has a chi-square distribution with degrees of freedom equal to \( 2n \). Furthermore, although \(-2 \sum_{i=1}^{n} \ln(P_i)\) and \(-2 \sum_{i=1}^{n} \ln(1-P_i)\) are not independent, the \( p \)-value of the Pearson’s concordant test defined in (1) can be easily approximated using the upper bound, \( 2P[\chi^2_{2n} > q^C] \), where \( q^C \) is the observed test statistic. In addition, the approximation is very accurate when the true \( p \)-value is small [1].

2.2. Concordant Tests Based on Gamma Distribution
The above Pearson’s concordant test can be extended to concordant tests based on gamma distribution. For this purpose, we first study a class of combination tests based on gamma distribution [10].

For given shape parameter $\alpha$ and scale parameter $\beta$, a $p$-value combination test statistic based on the gamma distribution can be constructed as follows:

$$T_{G(\alpha,\beta)}(P_1, P_2, \ldots, P_n) = \sum_{i=1}^{n} F_{G(\alpha,\beta)}^{-1}(1 - P_i),$$

where $F_{G(\alpha,\beta)}^{-1}(y)$ is the inverse function of the cumulative distribution function (CDF), $F_{G(\alpha,\beta)}(x)$, of the random variable $\text{Gamma}(\alpha, \beta)$, whose probability density function (PDF) is $f_{G(\alpha,\beta)}(x) = \beta^\alpha x^{\alpha-1} \exp(-\beta x)/\Gamma(\alpha)$ for $x > 0$, where the gamma function $\Gamma(z) = \int_{0}^{\infty} x^{z-1} e^{-x} dx$. The right-sided $p$-value of the test defined in (2) can be defined as follows:

$$P = P(\text{Gamma}(\alpha, \beta) > t) = 1 - F_{G(\alpha,\beta)}(t) = S_{G(\alpha,\beta)}(t),$$

where $\text{Gamma}(\alpha, \beta)$ is the random variable having gamma distribution with the shape parameter, $\alpha$, and the scale parameter, $\beta$; and $F_{G(\alpha,\beta)}$ and $S_{G(\alpha,\beta)}$ are the CDF and the survival function of $\text{Gamma}(\alpha, \beta)$.

It can be shown that, due to the property of gamma distribution, parameter $\beta$ has no effect on the test defined in (2) and (3) [10]. Therefore, for simplicity, we use $\beta = 1$ in the above test and denote $T_{G(\alpha,1)}$ hereafter. Furthermore, we can prove the following properties for the above class of gamma distribution-based tests [10]: (i) $T_{G(0,1)} = \lim_{\alpha \to 0^+} T_{G(\alpha,1)}$ is equivalent to the Tippett’s minimal-p test; (ii) $T_{G(0.5,1)}$ is the same as the chi-square test with degrees of freedom equal to $n$; (iii) $T_{G(1,1)}$ is equivalent to Fisher test; and (iv) $T_{G(\infty,1)} = \lim_{\alpha \to \infty} T_{G(\alpha,1)}$ is the same as the z-test. Therefore, the class of gamma distribution-based tests, $T_{G(\alpha,1)}$, includes many popular tests as special cases. This observation motivates us to generalize the Pearson’s concordant test.

We propose the gamma distribution-based concordant test with the following test statistic:

$$T_{G(\alpha)} = \max(T_{G(\alpha)}^L, T_{G(\alpha)}^R),$$

where $T_{G(\alpha)}^L = T_{G(\alpha)}(P_1, P_2, \ldots, P_n) = \sum_{i=1}^{n} F_{G(\alpha)}^{-1}(1 - P_i)$ and $T_{G(\alpha)}^R = T_{G(\alpha)}(1 - P_1, 1 - P_2, \ldots, 1 - P_n) = \sum_{i=1}^{n} F_{G(\alpha)}^{-1}(P_i)$ are the gamma distribution-based tests using the left- and right-sided $p$-values, respectively.

It is easy to see that when $\alpha = 1$, $T_{G(\alpha)}^C$ is equivalent to the Pearson’s concordant test $Q^C$. Therefore, $T_{G(\alpha)}^C$ is a generalization of the Pearson’s concordant test. We show that the $p$-value for the proposed concordant test, $T_{G(\alpha)}^C$, can also be estimated in a similar way as $Q^C$: $P[T_{G(\alpha)}^C > A] \approx 2P[T_{G(\alpha)}^L > A]$. Now, we study the properties of $T_{G(\alpha)}^C$. First we need the following definitions and lemmas [13].

**Definition 1.** A function $f$ on $\mathbb{R}^n$ is nondecreasing if it is nondecreasing in each of its $n$ arguments when the other $n-1$ values are held fixed.

**Definition 2.** Let $X_1, \ldots, X_n$ be random variables with a joint distribution. These random variables are associated if the covariance $\text{Cov}(f(X), g(X)) \geq 0$ holds for all nondecreasing functions, $f$ and $g$, for which the covariance is defined.

**Lemma 1.** (Theorem 2.1 of Esary, Proschan, and Walkup 1967 [13]) Independent random variables are associated.
Lemma 2. (Theorem 5.1 of Esary, Proschan, and Walkup 1967 [13]) Let \( X_1, \ldots, X_n \) be associated random variables, \( X = (X_1, \ldots, X_n), S_i = f_i(X), i = 1, \ldots, k \), then \( P[S_1 \leq s_1, \ldots, S_k \leq s_k] \geq \prod_{i=1}^{k} [S_i \leq s_i] \) and \( P[S_1 > s_1, \ldots, S_k > s_k] \geq \prod_{i=1}^{k} [S_i > s_i] \) for all \( s_1, s_2, \ldots, s_k \).

Lemma 3. For integer \( n \geq 1 \), let \( X = (X_1, \ldots, X_n) \in (0, 1)^n \) have independent components. Set \( T_{\alpha}^L = T_{G(a)}(X_1, \ldots, X_n) \) and \( T_{\alpha}^R = T_{G(a)}(1 - X_1, \ldots, 1 - X_n) \). Then, for any \( A_L > 0 \), and \( A_R > 0 \), \( P[T_{\alpha}^L > A_L, T_{\alpha}^R > A_R] \leq P[T_{\alpha}^L > A_L] P[T_{\alpha}^R > A_R] \).

Proof: Denote \( f(X) = T_{\alpha}^L = - \sum_i F_{\alpha}^{-1}(1 - P_{ij}) \) and \( g(X) = T_{\alpha}^R = \sum_i F_{\alpha}^{-1}(P_{ij}) \). It is easy to see that both \( f \) and \( g \) are nondecreasing functions of \( X \). And the components of \( X \) are independent and hence associated. Therefore, we have
\[
P[T_{\alpha}^L > A_L, T_{\alpha}^R > A_R] = P[-f(X) > A_L, g(X) > A_R] = P[f(X) < -A_L, g(X) > A_R] \leq P[g(X) > A_R] - P[f(X) > -A_L, g(X) > A_R] \leq -A_L] = P[T_{\alpha}^L > A_L] P[T_{\alpha}^R > A_R].
\]

□

The concordant test, \( T_{G(a)}^C \) defined in (4) has the following property.

Theorem 1. Let \( X = (X_1, \ldots, X_n) \in (0, 1)^n \) be a random vector with independent components, and denote \( T_{G(a)}^L = T_{G(a)}(X_1, \ldots, X_n) \), \( T_{G(a)}^R = T_{G(a)}(1 - X_1, \ldots, 1 - X_n) \), and \( T_{G(a)}^C = \max(T_{G(a)}^L, T_{G(a)}^R) \). For any \( A \in \mathbb{R} \), let \( P_L = P[T_{G(a)}^L > A] \), \( P_R = P[T_{G(a)}^R > A] \), and \( P_C = P[T_{G(a)}^C > A] \). Then \( P_L + P_R - P_L P_R \leq P_C \leq P_L + P_R \).

Proof: From Lemmas 2 and 3, we have \( P_C = P[T_{G(a)}^C > A] = P[T_{G(a)}^L > A \lor T_{G(a)}^R > A] = P[T_{G(a)}^L > A] + P[T_{G(a)}^R > A] - P[T_{G(a)}^L > A, T_{G(a)}^R > A] \geq P[T_{G(a)}^L > A] + P[T_{G(a)}^R > A] - P[T_{G(a)}^L > A] P[T_{G(a)}^R > A] = P_L + P_R - P_L P_R \). On the other hand, from Bonferroni inequality, we have \( P_C \leq P[T_{G(a)}^L > A] + P[T_{G(a)}^R > A] = P_L + P_R \). □

Under the condition specified in Theorem 1 (i.e., the global null hypothesis), it is easy to see that \( T_{G(a)}^L \) and \( T_{G(a)}^R \) have the same distribution; hence, \( P[T_{G(a)}^L > A] = P[T_{G(a)}^R > A] \).

From Theorem 1, we have the following result.

Corollary 1. Suppose \( X \sim U(0,1)^n \), for \( A > 0 \), let \( p_A = P[T_{G(a)}^L > A] \), then \( 2p_A - p_A^2 \leq P[T_{G(a)}^C > A] \leq 2p_A \).

From Corollary 1, the p-value of the concordant test, \( T_{G(a)}^C \), defined in (4), \( P[T_{G(a)}^C > A] \), can be easily approximated by twice of the p-value from \( T_{G(a)}^L \), i.e., \( P[T_{G(a)}^C > A] \approx 2P[T_{G(a)}^L > A] \). The approximation is very accurate if the true p-value (e.g., the p-value from the sampling distribution under the alternative) is small, as \( p_A^2 \) is even much smaller under this situation.

In the following subsection, we study an optimal concordant test when the optimal value for \( \alpha \) is unknown and needs to be estimated from the data.

2.3. Optimal Concordant Test

For a given set of one-sided p-values, the class of concordant tests, \( T_{G(a)}^C \), proposed in the previous section, Section 2.2, with different \( \alpha \) values can be applied. However, in order to control type I error, the \( \alpha \) value must be chosen before we see the data. But usually, the optimal value is unknown, and therefore how to choose an optimal value for \( \alpha \) is critical in practice. We may choose a set of \( \alpha \) values to apply the tests \( T_{G(a)}^C \) to the data and then use the smallest p-value among them. There are two limitations associated with
this procedure. First, we need to adjust for multiple comparisons, which may result in power loss. Second, the final result will depend on the chosen \( \alpha \) values; it might be difficult to determine those values. To circumvent these difficulties, we propose to estimate the optimal \( \alpha \) value from the data, and therefore an asymptotically optimal concordant test can be constructed accordingly.

In this subsection, we make the following assumption: \( P_1, P_2, \ldots, P_n \) are independently and identically distributed (i.i.d.) with the following density function for parameters \( \alpha > 0 \) and \( c < 1 \):

\[
f_{\alpha,c}(p) = (1-c)\alpha \exp \left[ c F_{G(\alpha)}^{-1}(1-p) \right] \quad \text{for } p \in (0,1).
\]

(5)

For the above density function, \( f_{\alpha,c}(p) \), it is not difficult to show that when \( c = 0 \), \( f_{\alpha,0}(p) = 1 \) (0 < \( p < 1 \)). Therefore, \( f_{\alpha,0}(p) \) corresponds to the global null hypothesis. In addition, we have the following property [10].

**Proposition 1.** Suppose that \( P_1, P_2, \ldots, P_n \) are i.i.d. with the common density function defined in (5) with parameters \( \alpha > 0 \) and \( c < 1 \). If both \( \alpha \) and \( c \) are known, then the gamma distribution-based test, \( T_{\alpha}(\chi) \), is uniformly most powerful (UMP).

When both parameters \( \alpha \) and \( c \) in (5) are unknown, they can be estimated via the constrained maximum likelihood estimation (CMLE), from which the following constrained likelihood ratio test (CLRT) can be defined:

\[
T_{\text{CLRT}}(P_1, \ldots, P_n) = 2 \hat{\alpha}_{\text{CLRT}} n \ln(1 - \hat{c}_{\text{CLRT}}) + 2(\hat{c}_{\text{CLRT}}) \sum_{i=1}^{n} F_{G(\alpha)}^{-1}(1 - P_i),
\]

(6)

where \( \hat{\alpha}_{\text{CLRT}} \) and \( \hat{c}_{\text{CLRT}} \) are the CMLEs for parameters \( \alpha \) and \( c \), respectively, through maximizing the log-likelihood function \( l(\alpha, c) = n \ln(1-c) + c \sum_{i=1}^{n} F_{G(\alpha)}^{-1}(1 - P_i) \) with the constrains \( c < 1 \) and \( \alpha > 0 \).

For the above CLRT-based test, \( T_{\text{CLRT}} \), we have the following result [10,14].

**Proposition 2.** The asymptotic distribution of the test \( T_{\text{CLRT}} \) is a mixture of chi-square distributions, \( \sum_{i=0}^{\infty} w_i \chi_i^2 \), where \( \chi_i^2 \) is the chi-square distribution with \( df = i \); \( \chi_0^2 \) is the random variable with probability 1 of being 0; and the weights \( w_0, w_1, \) and \( w_2 \) are determined by the null and the alternative hypothesis.

The above asymptotic result may not be directly applicable to estimate the \( p \)-value for this test. First, the number of \( p \)-values, \( n \), is usually small, and the asymptotic result may not provide approximate solutions. Second and more seriously, the above weights, \( w_i \)'s, are difficult to obtain. Hence, here we use a simple resampling method to approximate the null distribution and to estimate the \( p \)-value of \( T_{\text{CLRT}} \). More specifically, for a given sample size \( n \), we randomly sample \( n \) independent null \( p \)-values from the uniform distribution \( U(0,1) \) and then calculate the test statistic using (6). Repeat this process for \( B \) times (e.g., \( B = 10^5 \)) to get the empirical null distribution of the test statistic, which will be used to approximate the \( p \)-value of \( T_{\text{CLRT}}^p = \# \{ T_i \geq T_{\text{CLRT}}(p_1, \ldots, p_n) \} / B \), where \( T_i \) (\( i = 1, \ldots, B \)) is the test statistic from the \( i \)-th resampling, and \( T_{\text{CLRT}}(p_1, \ldots, p_n) \) is the observed test statistic from the data.

For \( T_{\text{CLRT}} \), we have the following result [10]:

**Proposition 3.** Under the conditions specified in (5), the CLRT-based test, \( T_{\text{CLRT}} \), is asymptotically UMP.

We now propose an optimal concordant test. Denote \( T_{\text{CLRT}}^L = T_{\text{CLRT}}(P_1, P_2, \ldots, P_n) \), and \( T_{\text{CLRT}}^R = T_{\text{CLRT}}(1 - P_1, 1 - P_2, \ldots, 1 - P_n) \), and then define the following concordant test statistic:

\[
T_{\text{CLRT}}^c = \max(T_{\text{CLRT}}^L, T_{\text{CLRT}}^R).
\]

(7)
Note that the null distributions of $T_{\text{LRT}}^L$ and $T_{\text{LRT}}^R$ are identical. For the concordant test, $T_{\text{LRT}}^C$, we have the following property.

**Theorem 2.** Let $X = (X_1, \ldots, X_n) \in (0,1)^n$ be a random vector with independent components, and let $T_{\text{LRT}}^L = \max_{a>b<c<d} \left[ n \ln(1-c) + c \sum_{i=1}^{n} F_{G(a)}(1-X_i) \right]$, $T_{\text{LRT}}^R = \max_{a>b<c<d} \left[ n \ln(1-c) + c \sum_{i=1}^{n} F_{G(a)}(X_i) \right]$, and $T_{\text{LRT}}^C = \max \left( T_{\text{LRT}}^L, T_{\text{LRT}}^R \right)$. For any $A \in \mathbb{R}$, let $P_L = P[T_{\text{LRT}}^L > A]$, $P_R = P[T_{\text{LRT}}^R > A]$, and $P_C = P[T_{\text{LRT}}^C > A]$. Then, $P_L + P_R - P_C \leq P_C \leq P_L + P_R$.

**Proof:** First notice that $f(X) = -\max_{a>b<c<d} \left[ n \ln(1-c) + c \sum_{i=1}^{n} F_{G(a)}(1-X_i) \right]$ and $g(X) = \max_{a>b<c<d} \left[ n \ln(1-c) + c \sum_{i=1}^{n} F_{G(a)}(X_i) \right]$ are both nondecreasing functions of associated random variables, $X$. The rest of the proof is similar to that for Theorem 1. □

In addition, we have the following result for $T_{\text{LRT}}^C$.

**Corollary 2.** Suppose $X \sim U(0,1)^n$; for $A > 0$, let $p_A = P[T_{\text{LRT}}^L > A]$, and then $2p_A - p_A^2 \leq P[T_{\text{LRT}}^C > A] \leq 2p_A$.

From Corollary 2, the $p$-value of the concordant test, $T_{\text{LRT}}^C$, defined in (7) can be estimated by the upper bound, $2p_A$, where $p_A$ is the $p$-value from the CLRT-based test, $T_{\text{LRT}}^L$, and can be estimated using the resampling method described previously. Again, when the true $p$-value is small, this approximation is very accurate.

3. Results

3.1. Numeric Results from Simulation Study

In this section, we assess the performance of the proposed tests through a simulation study. In the simulation, we compare the proposed optimal concordant test, $T_{\text{LRT}}^C$, with gamma distribution-based concordant tests, $T_{\text{G}(0)}^C$, $T_{\text{G}(1)}^C$ (i.e., the Owen’s Pearson concordant test), and $T_{\text{G}(\omega)}^C$, using one-sided $p$-values.

In the simulation study, we want to combine fifty ($n = 50$) randomly simulated independent $p$-values using different methods. We assume that a number, $m$ ($m = 10, 20, 40, 50$), out of these $n$ $p$-values are from the true individual alternative hypotheses, and the rest ($n-m$) are from the true individual null hypotheses. The $n-m$ $p$-values from the true null hypotheses are randomly sampled from the uniform distribution with range 0 and 1. For a true individual alternative hypothesis, $H_{1i}$ ($i = 1, \ldots, m$), the $p$-value $p_i$ is obtained via a normal variable, $z_i \sim N(\mu_i, 1)$. To consider the situations where the true effects from different studies may have opposite directions (positive and negative), we randomly set $k$ of the $m$ $\mu_i$’s to have the same direction (positive or negative), and the rest of $m-k$ has the other direction. A one-sided (right-sided) $p$-value for each true individual alternative hypothesis is obtained via the standard z-test, i.e., $p = P[Z > z]$, where $Z$ is the standard normal distribution, $N(0,1)$, and $z$ is the simulated number as described above.

For the hypotheses under true alternative, we consider three different scenarios for the effects of $\mu_i$’s. Scenario 1: We use $|\mu_i| = \mu v_i / \sum_{i=1}^{m} v_i$, where $v_i = 10^r_i$, $r_i \sim N(0,3,1)$, and $\mu = 0.8, 0.6, 0.4, 0.3$ when there are 10, 20, 40, and 50 true individual alternatives, respectively. Scenario 2: We set $|\mu_i| = \mu v_i / \sum_{i=1}^{m} v_i$, where $v_i \sim U(1,100)$ and $\mu = 1.2, 1.0, 0.6, 0.5$ when the number of true individual alternatives is $m = 10, 20, 40, 50$, respectively. Scenario 3: We consider $|\mu_i| = \mu / m$, $\mu = 1.5, 1.2, 0.8, 0.6$ when $m = 10, 20, 40, 50$, respectively. Note that, in the simulation, we consider the following: (i) The constants (e.g., $\mu$, the parameters in the normal distribution for $r_i$ and the uniform distribution for $v_i$) are chosen in a way that the empirical powers are appreciable for comparison. (ii) The sum of the absolute effect sizes is equal to $\mu$ for all the three scenarios.
(iii) The degree of heterogeneity of the effect sizes among the true individual alternatives decreases from Scenario 1 to Scenario 3 for each given \( m \).

The empirical power values of the tests are estimated as the rejection proportions based on 1000 replicates, with 0.05 as the significance level. For the CLRT-based concordant test, \( T_{CLRT}^C \), we use \( B = 10^5 \) samples to estimate the \( p \)-values from the resampling method described in Section 2.3.

When all of the 50 \( p \)-values are from null hypotheses, the empirical power is the type 1 error rate. All methods considered here were able to control the type I error rate: their empirical powers under this situation were close to 0.05 when the significance level 0.05 was used.

Figures 1–3 plot the empirical power values for the concordant tests based on minimal \( p \), \( T_{G(0)}^C \) (denoted as Min_CC), Fisher \( T_{G(1)}^C \) (denoted as Fisher_CC, i.e., Pearson’s concordant test), z-test \( T_{G(\infty)}^C \) (denoted as Z_CC), and the proposed concordant test, \( T_{CLRT}^C \) (denoted as LRT_CC), when one-sided \( p \)-values under Scenarios 1 to 3, respectively, are used.

**Figure 1.** Empirical power values of the tests based on one-sided \( p \)-values under Scenario 1.
From the simulation study, we have the following observations. First, under Scenario 1 (Figure 1), where the effect sizes under alternative hypotheses are extremely heterogeneous, usually the minimal p-based concordant test (i.e., $T^c_{G(1)}$) performs better than the Pearson’s concordant test ($T^c_{G(0)}$), which in turn outperforms the z-test-based concordant test ($T^c_{G(0)}$). Second, under Scenarios 2 and 3, when the degrees of heterogeneity of the
effect sizes among individual alternative hypotheses are less extreme than in Scenario 1, $T_c^{(1)}$ and the $T_c^{(0)}$ usually perform better or much better than $T_c^{(0)}$ (Figures 2 and 3). Third, for the three concordant tests with fixed $\alpha$ values, namely $T_c^{(0)}$, $T_c^{(1)}$, and $T_c^{(w)}$, one may perform well under some conditions but do poorly under others. Fourth, under all conditions considered, the CLRT-based concordant test, $T_{CLRT}$, always has the best or the second-best power. This demonstrates that, as expected, $T_{CLRT}$ is a robust test, meaning that, under many conditions, it has reasonable detection power compared with other tests.

3.2. Real Data Application

In this section, we demonstrate the usefulness of the proposed tests by applying them to a real-world problem. A data set from a meta-analysis is used. In that meta-analysis, results from 12 randomized trials—each examined the effect of patient rehabilitation designed for geriatric patients versus usual care on improving functional outcome at 3–12-month follow-ups—were collected [15,16]. The 12 estimated odds ratios (ORs) from the individual studies are listed in Table 1.

Table 1. Estimated odds ratio and its 95% CI from each study in a meta-analysis with 12 trials. Data were taken from Bachmann et al. [15] and Riley et al. [16].

<table>
<thead>
<tr>
<th>Study</th>
<th>OR</th>
<th>95% CI</th>
<th>Study</th>
<th>OR</th>
<th>95% CI</th>
<th>Study</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.11</td>
<td>0.51, 2.39</td>
<td>5</td>
<td>0.88</td>
<td>0.39, 1.95</td>
<td>9</td>
<td>1.06</td>
<td>0.63, 1.79</td>
</tr>
<tr>
<td>2</td>
<td>0.97</td>
<td>0.78, 1.21</td>
<td>6</td>
<td>1.28</td>
<td>0.71, 2.30</td>
<td>10</td>
<td>2.95</td>
<td>1.54, 5.63</td>
</tr>
<tr>
<td>3</td>
<td>1.13</td>
<td>0.73, 1.72</td>
<td>7</td>
<td>1.19</td>
<td>0.69, 2.08</td>
<td>11</td>
<td>2.36</td>
<td>1.18, 4.72</td>
</tr>
<tr>
<td>4</td>
<td>1.08</td>
<td>0.42, 2.75</td>
<td>8</td>
<td>3.82</td>
<td>1.37, 10.60</td>
<td>12</td>
<td>1.68</td>
<td>1.05, 2.70</td>
</tr>
</tbody>
</table>

The commonly used fixed-effect model for meta-analyses was inadequate for this data set, as the $p$-value from the Cochran’s test for homogeneity was 0.021. Hence, the random-effect model of the meta-analysis was applied, which estimated the overall OR as 1.36 with a 95% CI (1.08, 1.71) [16]. However, when we applied the goodness-of-fit test for the random-effect model, we obtained a $p$-value of 0.025 [3], which indicates a lack of fit of the random-effect model for this data set. We then applied the $p$-value combination methods to test whether there is an overall effect among these 12 independent trials.

To use the $p$-value combination methods, we need the individual $p$-values. For each trial, the corresponding $p$-value is calculated based on the reported 95% CI of the OR. If we use U and L to denote the upper and lower limits of the 95% CI, then the test statistic can be approximated as $z = \ln(U \times L) / \sqrt{4 \ln(U/L) / 3.92}$, which is asymptotically distributed as $N(0,1)$ under the null hypothesis that there is no difference between the new treatment and the control. Since the sample sizes of these 12 trials were relatively large, ranging from 108 to 1388, their $p$-values can be reasonably estimated using the asymptotic null distribution by comparing the test statistic with $N(0,1)$. For example, the right-sided $p$-value can be calculated as $P[Z > z]$, where $Z$ has a standard normal distribution, $N(0,1)$, and $z$ is the test statistic calculated as described above.

We first calculate the one-sided $p$-value for each trial and then use the values for the concordant tests. The $p$-values from the concordant tests $T_c^{(0)}$, $T_c^{(1)}$, $T_c^{(w)}$, and $T_{CLRT}$ are 0.013, 0.00017, 0.00031, and 0.00030, respectively. For comparison, we also applied the gamma distribution-based tests to this data set, using two-sided $p$-values calculated from each individual trial. The resulting $p$-values are 0.013, 0.0068, 0.075, and 0.0047 from $T_c^{(0)}$, $T_c^{(1)}$, $T_c^{(w)}$, and $T_{CLRT}$, respectively. As expected, under the situation that a majority of the trials have the same direction (10 out of the 12 estimated ORs were greater than 1), concordant tests are usually more powerful than the corresponding gamma distribution-based tests using two-sided $p$-values. Noticeably, the $z$-test-based concordant test, $T_{CLRT}$, obtained a much smaller $p$-value (0.00031) compared with the $z$-test using two-sided $p$-
values (0.075). Among the concordant tests, the proposed test has the second smallest \( p \)-value (0.00030).

4. Discussion and Conclusions

In this paper, we studied a class of gamma distribution-based concordant tests, which include the Pearson’s concordant test as a special case. We also proposed and studied a CLRT-based concordant test, which is asymptotically optimal, robust, and powerful under many conditions. The advantages of the proposed tests were demonstrated based on numeric simulation study and real data application.

In addition to their use for meta-analyses, the proposed concordant tests have many other applications. In a two-way contingency table, if at least one of the two variables is ordinal, a trend test may be more powerful than the popular unidirectional Pearson’s chi-square test [17]. One challenge in trend tests is how to appropriately assign scores for each level of the ordinal variable; results can be quite different with different score assignments. Our proposed concordant test, \( T_{CLRT}^C \), can circumvent this difficulty and provide another solution. Specifically, by using the asymptotically independent statistics with the same direction obtained by the Lancaster’s chi-square partition [18,19], we can obtain asymptotically independent one-sided \( p \)-values, to which \( T_{CLRT}^C \) can be applied.

The proposed tests will also be very useful for detecting casual genetic variants associated with diseases when applied to GWASs. It offers an alternative but powerful tool to combine information from different (subpopulations) but related studies (similar phenotype). We would expect that many genetic variants will be identified in the near future due to the new approach. In the future, we will also study how to combine dependent or correlated \( p \)-values.

**Funding:** This work was partially supported by the National Institutes of Health grants 1R03DE030259 and U11TR002529. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** The author declares no conflict of interest.

**References**


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.