Fisher’s method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations

QIZHAI LI*

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
liqz@amss.ac.cn

JIYUAN HU

School of Life Science, Institute of Biostatistics, Fudan University, Shanghai 200433, China

JUAN DING

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

GANG ZHENG

Office of Biostatistics Research, National Heart, Lung and Blood Institute, Bethesda, MD 20892, USA

SUMMARY

A classical approach to combine independent test statistics is Fisher’s combination of \( p \)-values, which follows the \( \chi^2 \) distribution. When the test statistics are dependent, the gamma distribution (GD) is commonly used for the Fisher’s combination test (FCT). We propose to use two generalizations of the GD: the generalized and the exponentiated GDs. We study some properties of mis-using the GD for the FCT to combine dependent statistics when one of the two proposed distributions are true. Our results show that both generalizations have better control of type I error rates than the GD, which tends to have inflated type I error rates at more extreme tails. In practice, common model selection criteria (e.g. Akaike information criterion/Bayesian information criterion) can be used to help select a better distribution to use for the FCT. A simple strategy of the two generalizations of the GD in genome-wide association studies is discussed. Applications of the results to genetic pleiotrophic associations are described, where multiple traits are tested for association with a single marker.

Keywords: Dependent tests; Fisher’s combination; Gamma distributions; Genetic pleiotropic associations; Genome-wide association studies; Type I error.

1. INTRODUCTION

Combining independent test statistics is common in biomedical research. One approach is to combine the \( p \)-values of one-sided tests using Fisher’s method (Fisher, 1932), referred to here as the Fisher’s
Fisher’s method of combining dependent statistics

Fisher’s method of combining dependent statistics

Fisher’s combination test (FCT). It has optimal Bahadur efficiency (Little and Folks, 1971). However, in general, it has a disadvantage in the interpretation of the results for combining two-sided tests. If one combines \( m \) independent \( p \)-values, the null distribution of the FCT is the \( \chi^2 \) distribution with \( 2m \) degrees of freedom \( \chi^2_{2m} \). If \( m \)-test statistics are dependent, the null distribution of the FCT is no longer \( \chi^2_{2m} \).

Most research on the FCT has focused on combining independent \( p \)-values and comparisons between the FCT and alternative approaches. In many genetic studies, test statistics are often dependent. In the analysis of Affymetrix expression array data, the FCT was employed to combine probe level two-sample \( t \)-tests, where the probes of the same probe set are correlated (Hess and Iyer, 2007). In genetic pleiotropic association studies, the association is tested between multiple traits collected from the same subjects and a marker. The FCT was used to combine the dependent single-trait analyses and was approximated using a gamma distribution (GD) (Zheng and others, 2012). Combining case–control and family-based designs or several genetic association studies can improve power to detect disease-associated markers (Infante-Rivard and others, 2009; Pfeiffer and others, 2009). When the two study designs share the cases or multiple association studies share controls, the dependent test statistics can be combined using the FCT. In genetic linkage and association studies under model uncertainty, multiple correlated tests are obtained under various genetic models (Joo and others, 2010).

Several approaches to handle the dependence in the FCT have been discussed. The GD is most commonly used. Assuming that \( m \)-tests follow a multivariate normal distribution, Brown (1975) and Kost and McDermott (2002) considered fitting the FCT with a GD, \( \Gamma(\xi, \eta) \), where \( \xi > 0 \) and \( \eta > 0 \) are the scale and shape parameters, respectively. Yang (2010) compared using the GD, normal approximations, and permutation methods for the FCT with the dependent \( p \)-values and showed that the GD approximation and permutation methods overall perform well, the latter of which also approximate the exact distribution of the FCT. Zheng and others (2012) derived an approximate GD for the FCT to test pleiotropic associations.

Since \( \chi^2_{2m} = \Gamma(2, m) \), the two-parameter \( \Gamma(\xi, \eta) \) is more flexible for the FCT when the tests are dependent. In this paper, we consider two generalizations of the GD with a third parameter. One is the generalized GD (GGD) with an extra shape parameter, and the other is the exponentiated GD (EGD) with an extra power parameter. Some properties of using the GD when the GGD or EGD is the true distribution for the FCT are studied, which provide insight into the performance of the FCT. Our results show that both the GGD and EGD have better control of type I error than the GD, especially at extremely small significance levels. Applications of the FCT to genetic pleiotropic associations are described. A simple strategy to apply the GGD and EGD to genome-wide association studies (GWASs) with several mixed types of traits is discussed. Simulations and applications using real data illustrate the performance of the proposed methods.

### 2. Fisher’s combination of \( p \)-values

Let \( m \geq 2 \) test statistics \( T_i, i = 1, \ldots, m \), be combined. Denote the \( p \)-value of \( T_i \) as \( p_i (i = 1, \ldots, m) \). The FCT test statistic, \( T_{\text{FCT}} \), is given by

\[
T_{\text{FCT}} = -2 \sum_{i=1}^{m} \log p_i .
\]  

(2.1)

If all the \( p \)-values are independent, \( T_{\text{FCT}} \sim \chi^2_{2m} \) under the null hypothesis \( H_0 \). Let \( X \sim \Gamma(\xi, \eta) \) with the density function

\[
\gamma(x; \xi, \eta) = x^{\eta-1} \exp(-x/\xi)/\{\Gamma(\eta)\}, \quad \Gamma(\eta) \text{ is the gamma function, and the distribution function is } \Gamma(x; \xi, \eta).
\]

When the \( p \)-values are dependent, model \( T_{\text{FCT}} \) with \( \Gamma(\xi, \eta) \) under \( H_0 \), where \((\xi, \eta)\) are estimated using the maximum likelihood estimates (MLEs) or the moments estimates (Kost and McDermott, 2002; Zheng and others, 2012). In our applications, all \( p \)-values are allowed to be
two-sided because all \( m \) \( p \)-values are obtained from testing \( m \) different null hypotheses. The following results, however, are valid when \( m \) one-sided \( p \)-values are combined using the FCT.

3. Generalizations of the GD

3.1 Applying the GGD

3.1.1 The GGD \( \Gamma(\xi, \eta, k) \). Its density function is \( \gamma(x; \xi, \eta, k) = k(x/\xi)^{k\eta-1} \exp\{-x/(\xi k)\}/(\xi \Gamma(\eta)) \) for \( x > 0 \), where \( \xi > 0 \) is the scale parameter and both \( \eta > 0 \) and \( k > 0 \) are shape parameters (Stacy, 1962). Its inference has been discussed in Stacy and Mihram (1965) and Huang and Hwang (2006). We use the MLEs of \( (\xi, \eta, k) \) and model \( T_{\text{FCT}} \) with \( \Gamma(\xi, \eta, k) \).

3.1.2 Fitting the GGD with the GD. Let \( (X_1, \ldots, X_n) \) be a simple random sample with the GGD \( \Gamma(\xi_2, \eta_2, k) \) for some \( (\xi_2, \eta_2, k) \). Denote the sample mean and variance as \( \bar{X} \) and \( S^2 \), respectively. Then, from Huang and Hwang (2006), as \( n \to \infty \),

\[
\bar{X} \to \xi_2 \frac{\Gamma(\eta_2 + 1/k)}{\Gamma(\eta_2)} \quad \text{and} \quad S^2 \to \xi_2^2 \frac{\Gamma(\eta_2 + 2/k)}{\Gamma(\eta_2)} - \xi_2^2 \left\{ \frac{\Gamma(\eta_2 + 1/k)}{\Gamma(\eta_2)} \right\}^2.
\]

If we fit the sample with the GD \( \Gamma(\xi_{11}, \eta_{11}) \), by the moments estimates,

\[
\hat{\xi}_{11} = S^2/\bar{X} \quad \text{and} \quad \hat{\eta}_{11} = \bar{X}^2/S^2. \tag{3.1}
\]

Denote \( \xi_{11} = \lim_{n \to \infty} \hat{\xi}_{11} \) and \( \eta_{11} = \lim_{n \to \infty} \hat{\eta}_{11} \). Then,

\[
\xi_{11} = \xi_2 \frac{\Gamma(\eta_2 + 2/k)\Gamma(\eta_2) - \Gamma^2(\eta_2 + 1/k)}{\Gamma(\eta_2 + 1/k)\Gamma(\eta_2)} \quad \text{and} \quad \eta_{11} = \frac{\Gamma^2(\eta_2 + 1/k)}{\Gamma(\eta_2 + 2/k)\Gamma(\eta_2) - \Gamma^2(\eta_2 + 1/k)}. \tag{3.2}
\]

Note that \( \xi_{11} \neq \xi_2 \) and \( \eta_{11} \neq \eta_2 \) unless \( k = 1 \). The following result provides more insight into fitting the GGD with the GD; the proof is given in supplementary material available at Biostatistics online. The monotonicity of \( \xi_{11} \) in \( k \) is shown with a condition \( \eta_2 + 1/k > 3/2 \), which is satisfied in our applications. In the proof, however, we show graphically that the result holds even if \( \eta_2 + 1/k \leq 3/2 \).

**Theorem 3.1** Let \( (\xi_2, \eta_2, k) \) and \( (\xi_{11}, \eta_{11}) \) be defined above and \( (\hat{\xi}_{11}, \hat{\eta}_{11}) \) be the moments estimates for the GD. Then \( \hat{\xi}_{11} \) and \( \hat{\eta}_{11} \) are asymptotically uncorrelated. Given \( (\xi_2, \eta_2) \), \( \xi_{11} \) is a decreasing function in \( k \) when \( \eta_2 + 1/k > 3/2 \) and \( \eta_{11} \) is an increasing function in \( k \).

Next, we compute the \( p \)-value of a test statistic \( T \) using the GD when the GGD is the true distribution. Let \( T \sim \Gamma(\xi_2, \eta_2, k) \) under \( H_0 \) for some \( (\xi_2, \eta_2, k) \). If we fit \( T \) with \( \Gamma(\hat{\xi}_{11}, \hat{\eta}_{11}) \) given in (3.2), then the \( p \)-values for \( T \) using the GGD and GD are given, respectively, by

\[
p_{\text{GGD}} = 1 - \Gamma((T/\xi_2)^k; 1, \eta_2) \quad \text{and} \quad p_{\text{GD}} = 1 - \Gamma(T/\hat{\xi}_{11}; 1, \hat{\eta}_{11}). \tag{3.3}
\]

We assume that all the parameters in the distributions are known for studying analytical behavior of the \( p \)-values. In practice, these parameters are estimated using the data. Theorem 3.2 shows that using the GD for \( T_{\text{FCT}} \) with dependent \( p \)-values may not be appropriate for extremely small significance levels when the GGD is the true distribution. In particular, when \( k < 1 \), type I error rate using the GD tends to be much larger than that using the GGD. A proof is given in supplementary material available at Biostatistics online.
Theorem 3.2 Let \( T \sim \Gamma(\xi_2, \eta_2, k) \) for some known \((\xi_2, \eta_2, k)\) with \( p \)-value \( p_{GGD} \). If we calculate the \( p \)-value \( p_{GD} \) with \( T \sim \Gamma(\xi_{11}, \eta_{11}) \), where \((\xi_{11}, \eta_{11})\) are given in (3.2), then \( \lim_{k \to \infty} p_{GD}/p_{GGD} = 0 \) if \( k < 1 \); 1 if \( k = 1 \); and \(+\infty\) if \( k > 1 \).

3.2 Applying the EGD

An alternative approach is the EGD, denoted as \( \Gamma(k; \xi, \eta) \) with distribution function is given by \( \Gamma^k(x; \xi, \eta) \) and a power parameter \( k > 0 \) (de Pascoa and others, 2011). We can find the MLEs for \((\xi, \eta, k)\) under \( H_0 \) and model \( T_{FCT} \) with \( \Gamma(k; \xi, \eta) \). In hypothesis testing, \( F^k(x) \) is also referred to as Lehmann alternatives (Lehmann, 1953). Like the GGD, it is more flexible to approximate \( T_{FCT} \) with the EGD than with the GD. For \( X \sim \Gamma(k; \xi_2, \eta_2) \) and some \( r \geq 0 \),

\[
E(X^r) = \frac{k\xi_2^r \Gamma(\eta_2 + r)}{\Gamma(\eta_2)} \int_0^{\infty} \gamma(x; 1, \eta_2 + r) \Gamma^{k-1}(x; 1, \eta_2) \, dx = \frac{k\xi_2^r \Gamma(\eta_2 + r)}{\Gamma(\eta_2)} I_r(\eta_2, k),
\]

where \( I_r(\eta_2, k) = \int_0^{\infty} \gamma(x; 1, \eta_2 + r) \Gamma^{k-1}(x; 1, \eta_2) \, dx \).

Let \((X_1, \ldots, X_n) \sim \Gamma(k; \xi_2, \eta_2)\) for some \((\xi_2, \eta_2, k)\). If we fit the sample with \( \Gamma(\xi_{12}, \eta_{12}) \) by the moments estimates and the same notation as in (3.1) and (3.2), we have

\[
\xi_{12} = \xi_2 \frac{(\eta_2 + 1)I_2(\eta_2, k) - k\eta_2 I_1^2(\eta_2, k)}{I_1(\eta_2, k)} \quad \text{and} \quad \eta_{12} = \frac{k\eta_2 I_1^2(\eta_2, k)}{(\eta_2 + 1)I_2(\eta_2, k) - k\eta_2 I_1^2(\eta_2, k)}. \tag{3.4}
\]

Note that \( \xi_{12} \neq \xi_2 \) and \( \eta_{12} \neq \eta_2 \) unless \( k = 1 \). Like (3.3), the \( p \)-values of observing the FCT \( T \) using the EGD and the GD can be written, respectively, as

\[
p_{EGD} = k \int_{T/\xi_2}^{\infty} \gamma(u; 1, \eta_2) \Gamma^{k-1}(u; 1, \eta_2) \, du \quad \text{and} \quad p_{GD} = 1 - \Gamma(T/\xi_{12}; 1, \eta_{12}). \tag{3.5}
\]

Although (3.4), similar to (3.2), is obtained, results similar to Theorem 3.1 are not available for the EGD due to the double integrations in the moments. Numerical results presented later and also online show that \( \xi_{12} > \xi_2 \) and \( \eta_{12} < \eta_2 \) when \( k < 1 \) and \( \xi_{12} < \xi_2 \) and \( \eta_{12} > \eta_2 \) when \( k > 1 \). A result similar to Theorem 3.2 is given below with proof given in supplementary material available at Biostatistics online.

Theorem 3.3 Let \( T \sim \Gamma(k; \xi_2, \eta_2) \) for some known \((\xi_2, \eta_2, k)\) with \( p \)-value \( p_{EGD} \). If we calculate \( p \)-value \( p_{GD} \) with \( T \sim \Gamma(\xi_{12}, \eta_{12}) \), where \((\xi_{12}, \eta_{12})\) are given in (3.4), then \( \lim_{k \to \infty} p_{GD}/p_{EGD} = 0 \) if \( \xi_{12} < \xi_2 \); 1 if \( \xi_{21} = \xi_2 \); and \(+\infty\) if \( \xi_{12} < \xi_2 \).

Since we observe \( \xi_{12} < \xi_2 \) when \( k > 1 \), \( p_{GD} \) tends to inflate type I error rate at extreme tail values when the EGD is the true distribution and \( k > 1 \).

3.3 Numerical results

We present the numerical values of \((\xi_{11}, \eta_{11}), (\xi_{12}, \eta_{12}), R_1 = p_{GD}/p_{GGD}, \) and \( R_2 = p_{GD}/p_{EGD} \) given a significance level \( \alpha \) and \((\xi_2, \eta_2, k)\) of either the GGD or EGD. When calculating \( R_1 \) and \( R_2 \), the value of the test statistic \( T \) is chosen so that \( p_{GGD} = \alpha \) or \( p_{EGD} = \alpha \) given \((\xi_2, \eta_2, k)\). The results are reported in Table 1, which confirm the analytic results, and show that there is more substantial impact on type I error using the GD when the data are drawn from the GGD or EGD with \( \alpha = 0.0001 \) than with \( \alpha = 0.05 \). Table 1 also shows that the patterns of \( R_1 \) and \( R_2 \) have opposite directions depending on whether or not
the same individuals are available, testing pleiotropic associations is more powerful than testing association in which multiple traits are tested for associations with a single marker. When multiple traits measured from k individuals are available, testing pleiotropic associations is more powerful than testing association in which multiple traits are measured from k individuals. Several approaches have been discussed for genetic pleiotropic associations; combining dependent traits here, although other types of traits, e.g. ordinal traits, can also be considered (Zhang, 2008). We consider binary and continuous traits here, although other types of traits, e.g. ordinal traits, can also be considered (Zhang and others, 2010). Several approaches have been discussed for genetic pleiotropic associations; combining dependent univariate statistics is a useful approach when the statistics do not follow a multivariate normal distribution (Yang and others, 2010).

### 4. Application to genetic pleiotropic association

We describe how to use the proposed methods with an application to genetic pleiotropic association studies, in which multiple traits are tested for associations with a single marker. When multiple traits measured from the same individuals are available, testing pleiotropic associations is more powerful than testing association with an individual trait (Klei and others, 2008; Zhu and Zhang, 2009). We consider binary and continuous traits here, although other types of traits, e.g. ordinal traits, can also be considered (Zhang and others, 2010). Several approaches have been discussed for genetic pleiotropic associations; combining dependent univariate statistics is a useful approach when the statistics do not follow a multivariate normal distribution (Yang and others, 2010).

### 4.1 Test statistics for pleiotropic association

Let the alleles of a genetic marker be A and B. Denote the traits as $t_1, \ldots, t_m$ ($m \geq 2$). Assume that the first U traits are binary and the other V = m − U traits are continuous. For the $i$th individual ($i = 1, \ldots, n$), denote the data as $(d_{iu}, y_{iv}, g_i)$, where $d_{iu}$ is the disease indicator for the $u$th binary trait ($u = 1, \ldots, U$), $y_{iv}$ is the $v$th continuous trait value ($v = 1, \ldots, V$), and $g_i$ is the 0, 1, or 2 genotype score corresponding to the numbers of allele B in the genotype.

For testing association with the $u$th binary trait with $r_u$ cases and $s_u$ controls ($n = r_u + s_u$), the trend test (Sasieni, 1997; Freidlin and others, 2002) can be written as

$$T_{1u} = \frac{n^{1/2}}{} \left[ \frac{(s_u r_{1u} - r_u s_{1u}) + 2(s_u r_{2u} - r_u s_{2u})}{r_u s_u \{n(n_{1u} + 4n_{2u}) - (n_{1u} + 2n_{2u})^2\}^{1/2}}, \quad u = 1, \ldots, U, \right]$$

where $(r_{0u}, r_{1u}, r_{2u})$ and $(s_{0u}, s_{1u}, s_{2u})$ are the counts of $(AA, AB, BB)$ in cases and controls, respectively, and $n_i = r_{iu} + s_{iu}$ ($i = 0, 1, 2$). Under $H_0$, $T_{1u}^2 \sim \chi^2_1$ asymptotically. For testing association with the $v$th continuous trait, the $F$-test can be written as

$$T_{2v} = \frac{\sum_{j=1}^n (y_{iv} - \bar{y}_v)^2}{Y_v' [I_n - X(X'X)^{-1}X'] Y_v}, \quad v = 1, \ldots, V,$$

where $(r_{0u}, r_{1u}, r_{2u})$ and $(s_{0u}, s_{1u}, s_{2u})$ are the counts of $(AA, AB, BB)$ in cases and controls, respectively, and $n_i = r_{iu} + s_{iu}$ ($i = 0, 1, 2$). Under $H_0$, $T_{1u}^2 \sim \chi^2_1$ asymptotically. For testing association with the $v$th continuous trait, the $F$-test can be written as

$$T_{2v} = \frac{\sum_{j=1}^n (y_{iv} - \bar{y}_v)^2}{Y_v' [I_n - X(X'X)^{-1}X'] Y_v}, \quad v = 1, \ldots, V,$$
where $\bar{y}_v = (1/n) \sum_{i=1}^n y_{iv}$, $Y_v = (y_{1v}, \ldots, y_{nv})^T$, $X = (I_n, G)$, $I_n$ is an $n$-dimensional vector with the elements being 1, and $G = (g_1, \ldots, g_n)^T$. Under $H_0$, $T_{2v}$ follows an asymptotic $F$-distribution with degrees of freedom $1$ and $n - 2$.

Denote the $p$-values of $T_{1v}^2$ and $T_{2v}$ as $p_{1v}$ and $p_{2v}$. All $p$-values in this application can be either two-sided or one-sided because all $p$-values are obtained from testing different associations, where each $p$-value corresponds to one trait. Then the FCT is $T_{\text{FCT}} = -2 \sum_{u=1}^U \log(p_{1u}) - 2 \sum_{v=1}^V \log(p_{2v})$. Under $H_0$, we assume $T_{\text{FCT}} \sim \Gamma(\xi, \eta, k)$ using the GGD or $T_{\text{FCT}} \sim \Gamma(K)(\xi, \eta)$ using the EGD, where $(\xi, \eta, k)$ are estimated separately for the two distributions.

4.2 Estimation procedures

4.2.1 For a single marker. The simulation procedure given in Zheng and others (2012) is modified below to estimate $(\xi, \eta, k)$ for both the GGD and EGD for a single marker. First, we simulate data under $H_0$ from the observed data. Then, we assume that either the GGD or EGD is the true distribution, so we can find the MLEs of the parameters for one of the two distributions. In the simulation, the correlation structure among the traits is retained but the associations between the traits and the marker are all removed. Denote the density function for either the GGD or EGD as $f(x; \xi, \eta, k)$.

In the $s$th replicate ($s = 1, \ldots, S$ for some large $S$), one keeps the observed traits $\{(d_{iu}, y_{iv}): u = 1, \ldots, U; v = 1, \ldots, V\}$ for the $i$th individual ($i = 1, \ldots, n$), but simulates his/her genotype $g_{is}^*$ from a multinomial distribution $\text{Mul}(1; \hat{q}_0, \hat{q}_1, \hat{q}_2)$, where $\hat{q}_i = n_i/n$ for $i = 0, 1, 2$. Then $T_{\text{FCT}}$ is computed using $\{(d_{iu}, y_{iv}, g_{is}^*): u = 1, \ldots, U; v = 1, \ldots, V; i = 1, \ldots, n\}$, denoted as $T_{\text{FCT}}^{(s)} (s = 1, \ldots, S)$. The empirical likelihood function with $S$ replicates is $L(\xi, \eta, k) = \prod_{s=1}^S f(T_{\text{FCT}}^{(s)}; \xi, \eta, k)$. Then find the MLEs $(\hat{\xi}, \hat{\eta}, \hat{k}) = \arg\max_{\xi, \eta, k} L(\xi, \eta, k)$ numerically. The MLEs depend on both the minor allele frequency (MAF) of the marker and the correlations among the traits. However, in our simulation, the traits are fixed. Thus, conditional on the traits (and their correlations), the MLEs for any two markers with different MAFs only depend on the MAFs. The above simulation does not require Hardy–Weinberg equilibrium (HWE). With HWE, one can generate $g^* \sim \text{Mul}(1; (1 - p)^2, 2p(1 - p), p^2)$ where $p$ is the estimated MAF of the marker.

Like a permutation test, the above simulation procedure can be used to find an approximate exact distribution for the FCT because $\{T_{\text{FCT}}^{(s)}, s = 1, \ldots, S\}$ forms an empirical distribution of the FCT when $S$ is large.

4.2.2 For a GWAS. In a GWAS, hundreds to millions of markers with different MAFs are tested. Although an exact FCT method described in Section 4.2.1 can achieve similar performance to the GGD or EGD, it is not computationally efficient for a GWAS as a set of simulation with large number of replicates has to be done for each marker with a typical genome-wide significance level $5 \times 10^{-8}$. We propose a computationally efficient method based on our estimation procedure in Section 4.2.1 for applying the GGD and EGD to GWASs. The method is given below with technical details and justifications given in supplementary material available at Biostatistics online (Section 5.2 and Tables S2–S3).

For the GGD, we first find the MLEs $(\hat{\xi}, \hat{\eta}, \hat{k})$ for each MAF from 0.1 to 0.5 with an increment of 0.05. Then we take the average of the MLEs over the nine MAFs as the final estimates, denoted as $(\bar{\xi}, \bar{\eta}, \bar{k})$. For the EGD, we first obtain $(\hat{\xi}, \hat{\eta}, \hat{k})$ and $(\bar{\xi}, \bar{\eta}, \bar{k})$ similar to the GGD, but we take $\bar{k}_{\text{max}}$ as the maximum of all $\bar{k}$ across the nine MAFs. Then we apply $(\hat{\xi}, \hat{\eta}, \bar{k}_{\text{max}})$ as the final estimates.

4.3 Real applications

4.3.1 Candidate genes. We consider the genetic data of the Genetic Analysis Workshop 16 (GAW16), which consist of 868 rheumatoid arthritis (RA) positive cases and 1194 controls, and 531689
Table 2. Estimates for the parameters of GD, GGD, and EGD and their p-values for pleiotropic associations with anti-CCP and IgM \((m = 2)\) with imputations of the missing quantitative traits in the controls using the FCT

<table>
<thead>
<tr>
<th>SNP</th>
<th>((\xi, \eta))</th>
<th>AIC/BIC</th>
<th>p-value</th>
<th>((\xi, \eta, k))</th>
<th>AIC/BIC</th>
<th>p-value</th>
<th>((\xi, \eta, k))</th>
<th>AIC/BIC</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs653667</td>
<td>(2.801, 1.433)</td>
<td>32.89/55.95</td>
<td>1.53E-06</td>
<td>(0.331, 4.142, 0.599)</td>
<td>30.45/65.04</td>
<td>2.24E-05</td>
<td>(3.967, 0.128, 15.161)</td>
<td>32.14/66.74</td>
<td>7.72E-06</td>
</tr>
<tr>
<td>rs2454170</td>
<td>(2.783, 1.434)</td>
<td>37.57/60.64</td>
<td>1.45E-07</td>
<td>(0.353, 4.047, 0.606)</td>
<td>33.56/68.16</td>
<td>4.84E-06</td>
<td>(3.895, 0.141, 13.766)</td>
<td>36.02/70.62</td>
<td>1.10E-06</td>
</tr>
<tr>
<td>rs4375229</td>
<td>(2.771, 1.447)</td>
<td>54.82/77.88</td>
<td>2.59E-11</td>
<td>(0.366, 4.014, 0.611)</td>
<td>43.67/78.26</td>
<td>3.48E-08</td>
<td>(4.002, 0.078, 25.269)</td>
<td>48.59/83.19</td>
<td>2.14E-09</td>
</tr>
<tr>
<td>rs12743229</td>
<td>(2.767, 1.443)</td>
<td>33.67/56.74</td>
<td>1.02E-06</td>
<td>(0.317, 4.218, 0.597)</td>
<td>31.01/65.61</td>
<td>1.68E-05</td>
<td>(3.855, 0.144, 13.716)</td>
<td>33.02/67.62</td>
<td>4.86E-06</td>
</tr>
<tr>
<td>rs4567320</td>
<td>(2.799, 1.429)</td>
<td>16.40/39.46</td>
<td>6.03E-03</td>
<td>(0.336, 4.116, 0.601)</td>
<td>18.28/52.88</td>
<td>7.79E-03</td>
<td>(3.933, 0.135, 14.363)</td>
<td>18.46/53.06</td>
<td>6.61E-03</td>
</tr>
<tr>
<td>rs11264329</td>
<td>(2.774, 1.441)</td>
<td>9.10/32.17</td>
<td>2.56E-01</td>
<td>(0.382, 3.932, 0.614)</td>
<td>11.14/45.74</td>
<td>2.41E-01</td>
<td>(4.021, 0.094, 20.658)</td>
<td>11.11/45.70</td>
<td>2.41E-01</td>
</tr>
<tr>
<td>rs429201</td>
<td>(2.786, 1.440)</td>
<td>28.62/51.69</td>
<td>1.29E-05</td>
<td>(0.387, 3.942, 0.617)</td>
<td>27.81/62.41</td>
<td>7.74E-05</td>
<td>(3.937, 0.118, 16.647)</td>
<td>28.82/63.42</td>
<td>3.98E-05</td>
</tr>
<tr>
<td>rs10917678</td>
<td>(2.758, 1.446)</td>
<td>24.81/47.88</td>
<td>8.65E-05</td>
<td>(0.322, 4.194, 0.598)</td>
<td>24.92/59.51</td>
<td>3.18E-04</td>
<td>(3.896, 0.130, 15.124)</td>
<td>25.70/60.30</td>
<td>1.85E-04</td>
</tr>
<tr>
<td>rs1323120</td>
<td>(2.790, 1.437)</td>
<td>52.22/75.29</td>
<td>9.56E-11</td>
<td>(0.346, 4.081, 0.604)</td>
<td>42.06/76.65</td>
<td>7.80E-08</td>
<td>(4.013, 0.094, 20.659)</td>
<td>46.70/81.29</td>
<td>5.54E-09</td>
</tr>
<tr>
<td>rs17505650</td>
<td>(2.802, 1.431)</td>
<td>31.64/54.71</td>
<td>2.86E-06</td>
<td>(0.341, 4.100, 0.602)</td>
<td>29.69/64.29</td>
<td>3.20E-05</td>
<td>(4.004, 0.100, 19.516)</td>
<td>31.12/65.72</td>
<td>1.29E-05</td>
</tr>
<tr>
<td>rs2985441</td>
<td>(2.774, 1.442)</td>
<td>58.02/81.09</td>
<td>5.22E-12</td>
<td>(0.379, 3.960, 0.614)</td>
<td>45.55/80.15</td>
<td>1.37E-08</td>
<td>(3.943, 0.119, 16.343)</td>
<td>51.31/85.91</td>
<td>5.45E-10</td>
</tr>
<tr>
<td>rs1923946</td>
<td>(2.749, 1.451)</td>
<td>15.31/38.37</td>
<td>1.03E-02</td>
<td>(0.354, 4.063, 0.609)</td>
<td>17.32/51.91</td>
<td>1.21E-02</td>
<td>(3.917, 0.112, 17.603)</td>
<td>17.44/52.04</td>
<td>1.07E-02</td>
</tr>
<tr>
<td>rs2802822</td>
<td>(2.780, 1.437)</td>
<td>11.97/35.04</td>
<td>5.69E-02</td>
<td>(0.381, 3.944, 0.614)</td>
<td>14.18/48.78</td>
<td>5.50E-02</td>
<td>(3.942, 0.127, 15.284)</td>
<td>14.20/48.80</td>
<td>5.27E-02</td>
</tr>
<tr>
<td>rs1567602</td>
<td>(2.793, 1.433)</td>
<td>40.91/63.98</td>
<td>2.74E-08</td>
<td>(0.295, 4.307, 0.588)</td>
<td>35.33/69.93</td>
<td>2.11E-06</td>
<td>(3.998, 0.105, 18.547)</td>
<td>38.23/72.83</td>
<td>3.76E-07</td>
</tr>
</tbody>
</table>
single-nucleotide polymorphisms (SNPs) (Amos and others, 2009). In addition to the RA status, two quantitative traits (anti-cyclic citrullinated peptide (anti-CCP) antibody and rheumatoid factor IgM) are also available only for the individuals with RA. It is known that anti-CCP and IgM levels are higher among individuals with RA than those without (Huizinga and others, 2005). For this application, we first consider testing association with anti-CCP and IgM. Because the controls have no measures of anti-CCP and IgM, as Zheng and others (2012) did for anti-CCP, we impute the unmeasured values of both traits of the controls by the respective minimum observed trait values of the cases and apply the F-test to each trait as if the imputed values were observed. Then the FCT is applied to testing the pleiotropic association. The null hypothesis is that an SNP is not associated with either trait. Zheng and others (2012) showed that this imputation has no impact under the null hypothesis. Here, we focus on the testing problem; if we consider this an estimation problem, then replacing a large amount of the data (controls) with a single value will likely produce biased estimations. All values (observed or imputed) are log-transformed. We focus on the SNPs listed in Zheng and others (2012, Table VII), but the overall significance level for each SNP test is $5 \times 10^{-8}$. The results are reported here in Table 2.

We present the moments estimates for the GD and the MLEs for the GGD and EGD (based on Section 4.2.1) along with common model selection criteria: Akaike information criterion (AIC) and Bayesian information criterion (BIC). The results show that the estimates for the GD and GGD are consistent across the SNPs. For the EGD, only the estimates for $(\xi, \eta)$ are consistent but not those for $k$. Using AIC/BIC, the GGD and/or EGD are often a better fit for the FCT than the GD. When AIC/BIC values are similar among the three distributions, they also have similar $p$-values. However, when the GD is a worse fit, its $p$-values are more significant. This indicates that we need to be cautious when using the GD for combining dependent statistics in the FCT.

For a comparison, Table S4 in supplementary material available at Biostatistics online shows results analogous to Table 2 except that there is no imputation, so controls are not used in the F-test. In this case, the two tests seem to be independent and the GD is the better model to use, although the $p$-values using all three distributions are similar. However, the $p$-values without imputations tend to be less significant than those with imputations because the former ignores the potential associations of these SNPs with RA. For another comparison, in supplementary material available at Biostatistics online, we report Table S5 for pleiotropic association with three traits (RA, anti-CCP, and IgM) without imputations. The three tests seem to be independent and result in similar $p$-values although the GD is a better fit for the FCT. When comparing the $p$-values using the GGD and EGD in Table 2 with those in Table S5, we note that $p$-values in Table 2 are often more significant, which indicates the benefit of the imputation as a way of testing pleiotropic association compared with testing all three traits.

Table 3. The moments estimates $(\hat{\xi}, \hat{\eta})$ for GD and the MLEs $(\hat{\xi}, \hat{\eta}, \hat{k})$ for EGD and GGD given $\rho$ and $m$

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\rho$</th>
<th>GD</th>
<th>GGD</th>
<th>EGD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>(2.011, 1.987)</td>
<td>(1.951, 2.029, 0.990)</td>
<td>(2.050, 1.889, 1.062)</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>(2.163, 1.848)</td>
<td>(1.529, 2.315, 0.897)</td>
<td>(2.501, 1.178, 1.712)</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>(2.453, 1.628)</td>
<td>(0.791, 3.101, 0.732)</td>
<td>(3.359, 0.392, 5.190)</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>(2.955, 1.354)</td>
<td>(0.222, 4.583, 0.554)</td>
<td>(4.127, 0.082, 22.58)</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>(3.603, 1.109)</td>
<td>(0.127, 4.772, 0.482)</td>
<td>(4.744, 0.065, 22.59)</td>
</tr>
<tr>
<td>3</td>
<td>(0, 0, 0)</td>
<td>(2.009, 2.981)</td>
<td>(1.876, 3.106, 0.979)</td>
<td>(2.057, 2.832, 1.064)</td>
</tr>
<tr>
<td></td>
<td>(0, 0.4, 0)</td>
<td>(2.224, 2.691)</td>
<td>(1.076, 4.006, 0.823)</td>
<td>(2.760, 1.512, 2.036)</td>
</tr>
<tr>
<td></td>
<td>(0, 0, 0.8)</td>
<td>(2.828, 2.110)</td>
<td>(0.337, 5.623, 0.619)</td>
<td>(4.340, 0.336, 8.492)</td>
</tr>
<tr>
<td></td>
<td>(0.4, 0.8, 0.32)</td>
<td>(3.171, 1.885)</td>
<td>(0.097, 7.664, 0.509)</td>
<td>(4.765, 0.170, 16.37)</td>
</tr>
<tr>
<td></td>
<td>(0.32, 0.4, 0.8)</td>
<td>(3.122, 1.919)</td>
<td>(0.106, 7.556, 0.515)</td>
<td>(4.649, 0.231, 11.94)</td>
</tr>
<tr>
<td></td>
<td>(0.8, 0.32, 0.4)</td>
<td>(3.176, 1.880)</td>
<td>(0.176, 6.565, 0.550)</td>
<td>(4.830, 0.153, 17.97)</td>
</tr>
</tbody>
</table>
4.3.2 **GWAS application.** We used the same dataset as in the previous application except that we analyzed all the SNPs from chromosomes 5, 6, and 10 after standard GWAS quality control procedures, where chromosome 6 also contains the well-known HLA region associated with RA. We applied our genome-wide methods in Section 4.2.2 and plotted the \( p \)-values using the GD, EGD, and GGD in Figure S2 in supplementary material available at *Biostatistics* online. Except for the HLA region, for which the GD, EGD, and GGD all have peaks, there are no significant pleiotropic associations at the GWAS level \( 5 \times 10^{-8} \) using the EGD and GGD. But there is a significant association on chromosome 5 and some marginally significant associations on chromosome 10 using the GD. For example, SNP rs6596147 on chromosome 5 has \( p \)-values \( 4.76 \times 10^{-10} \) (GD), \( 1.69 \times 10^{-7} \) (EGD), and \( 3.45 \times 10^{-7} \) (GGD). Based on the analytical and simulation results (presented next), one should be cautious when reporting this association based on the GD alone.

## 5. Simulations

We present part of the simulation results here and the rest online with a summary here. Based on our application, we simulate \( m \)-test statistics \((t_1, \ldots, t_m)\), for \( m = 2, 3, 4 \), from a multivariate normal distribution with mean zeros, unit variances, and a set of pair-wise positive correlations \( \rho \) among the tests. We also added results for \( m = 10 \) test statistics for some simulations. The GD, GGD, and EGD are, respectively, used for the FCT given in (2.1). In Table 3, we report the moments estimates \((\hat{\xi}, \hat{\eta}, \hat{k})\) for both the EGD and GGD with 100 000 samples of \((t_1, \ldots, t_m)\) given \( m = 2, 3, \) and \( \rho \). Results with \( m = 4, 10 \) are reported online (Tables S6–S7 of supplementary material available at *Biostatistics* online). The results show that \( \hat{\xi} \approx \hat{\eta} \approx 2 \) (\( \hat{\xi} \approx 2, \hat{\eta} \approx 3 \)) using either distribution when \( \rho = 0 \ (\rho = (0, 0, 0)) \), in which case \( \hat{k} \approx 1 \) so that all three distributions can be used. The results are consistent with the analytical results and show that the distributions are influenced by \( \rho \).

We also compare their empirical type I error rates with 1 million replicates and two nominal levels \((0.05 \text{ and } 0.0001)\) using the estimates in Table 3. The results are reported in Table 4 for \( m = 3 \), which show that all type I error rates are close to the nominal level when \( \alpha = 0.05 \). When \( \alpha = 0.0001 \), using the GD inflates type I error rates while using the GGD tends to be conservative, but the type I error rates of the EGD are still close to the nominal level. Q–Q plots of the \( p \)-values in \(-\log_{10} \) scale using the GD, GGD, and EGD are given in Figure 1 for \( m = 3 \). Q–Q plots with \( m = 2, m = 4, \) and \( m = 10 \) are reported online (Figures S3–S5 of supplementary material available at *Biostatistics* online). To summarize, the size of the FCT using the GGD is more conservative with \( m = 2 \) than with \( m = 4 \) and \( m = 10 \). The EGD has overall better size than the GGD. The GD has inflated type I error rates with a small significance level when dependent statistics are combined. When \( m = 10 \), the performance of the EGD and GGD are less reliable than when \( m = 2, 3, 4 \). More discussion on \( m = 10 \) can be found online.

### Table 4. Empirical type I error rates with \( \alpha = 0.05 \times 10^{-2} \) and with \( \alpha = 0.0001 \times 10^{-4} \) using GD, EGD, and GGD given \( \rho \) when \( m = 3 \)

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \alpha = 0.05 )</th>
<th>( \alpha = 0.0001 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (0, 0, 0) )</td>
<td>4.99 4.98 4.98</td>
<td>0.90 0.81 0.83</td>
</tr>
<tr>
<td>( (0, 0.4, 0) )</td>
<td>5.03 5.01 5.01</td>
<td>1.88 1.11 1.05</td>
</tr>
<tr>
<td>( (0, 0.8) )</td>
<td>5.19 5.20 5.20</td>
<td>2.42 0.60 0.63</td>
</tr>
<tr>
<td>( (0.4, 0.8, 0.32) )</td>
<td>5.10 5.15 5.11</td>
<td>2.96 0.58 1.17</td>
</tr>
<tr>
<td>( (0.32, 0.4, 0.8) )</td>
<td>5.17 5.11 5.51</td>
<td>2.92 0.54 1.17</td>
</tr>
<tr>
<td>( (0.8, 0.32, 0.4) )</td>
<td>5.11 5.18 5.47</td>
<td>2.77 0.57 0.99</td>
</tr>
</tbody>
</table>
Fisher’s method of combining dependent statistics

Theoretical values

Empirical values

The simulations for our GWAS strategies in Section 4.2.2 and the results are reported online. The results show that, when \( m = 2, 3, 4 \), using the EGD and GGD and the GWAS strategies has better control of type I error rates than using the GD, except for \( \rho = (0, 0, 0) \), under which type I error rates are similar using all three distributions. Simulation results with negative correlations can be found online.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

We thank two reviewers for their careful reading and thoughtful comments, which greatly improved our presentation. We would like to thank Neal Jeffries for his careful reading and editing our manuscript. Conflict of Interest: None declared.

FUNDING

The GAW16 data were gathered with the support of grants from the National Institutes of Health (NO1-AR-2-2263 and RO1-AR-44422, Peter K. Gregersen, PI), and the National Arthritis Foundation.
The use of the data was approved by GAW16. Research of Qizhai Li is partially supported by National Nature Science Foundation of China, Nos 61134013, 11371353.

REFERENCES


YANG, J. J. (2010). Distribution of Fisher’s combination statistic when the tests are dependent. *Journal of Statistical Computation and Simulation* 80, 1–12.
Fisher’s method of combining dependent statistics


[Received November 13, 2012; revised September 27, 2013; accepted for publication September 28, 2013]