Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects

C. H. Lee1, E. Eskin2,3 and B. Han1,*

1Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, Songpa-gu, Seoul 138-736, Korea, 2Department of Computer Science and 3Department of Human Genetics, University of California, Los Angeles, CA 90095, USA

*To whom correspondence should be addressed.

Abstract

Motivation: Meta-analysis is essential to combine the results of genome-wide association studies (GWASs). Recent large-scale meta-analyses have combined studies of different ethnicities, environments and even studies of different related phenotypes. These differences between studies can manifest as effect size heterogeneity. We previously developed a modified random effects model (RE2) that can achieve higher power to detect heterogeneous effects than the commonly used fixed effects model (FE). However, RE2 cannot perform meta-analysis of correlated statistics, which are found in recent research designs, and the identified variants often overlap with those found by FE.

Results: Here, we propose RE2C, which increases the power of RE2 in two ways. First, we generalized the likelihood model to account for correlations of statistics to achieve optimal power, using an optimization technique based on spectral decomposition for efficient parameter estimation. Second, we designed a novel statistic to focus on the heterogeneous effects that FE cannot detect, thereby, increasing the power to identify new associations. We developed an efficient and accurate p-value approximation procedure using analytical decomposition of the statistic. In simulations, RE2C achieved a dramatic increase in power compared with the decoupling approach (71% vs. 21%) when the statistics were correlated. Even when the statistics are uncorrelated, RE2C achieves a modest increase in power. Applications to real genetic data supported the utility of RE2C. RE2C is highly efficient and can meta-analyze one hundred GWASs in one day.

Availability and implementation: The software is freely available at http://software.buhmhan.com/RE2C.

Contact: buhm.han@amc.seoul.kr

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Genome-wide association studies (GWASs) have identified numerous single-nucleotide polymorphisms (SNPs) that are associated with human traits (Manolio, 2010; Welter et al., 2014). For many diseases, however, the identified variants explain only part of the known heritability, which indicates the existence of undetected variants with small effects (Evangelou and Ioannidis, 2013; Manolio, 2013). To scale up genetic discovery, meta-analysis of GWASs has become a popular tool to augment the sample size (Evangelou and Ioannidis, 2013; Manolio, 2013; Fleiss, 1993; Zeggini and Ioannidis, 2009). Recently, the use of meta-analysis in GWASs has expanded to new research designs, such as combining different related diseases (Kiryluk et al., 2012; Lee et al., 2014; Perry et al., 2012), populations (Liu et al., 2015), environments (Kang et al., 2014), tissues (Sul et al., 2013) and cancer types (Bhattacharjee et al., 2012; Petersen et al., 2010). These differences between studies can manifest as heterogeneity, which refers to effect-size differences. When heterogeneity exists, the commonly used fixed effects model (FE) is not optimal. The traditional random effects model (RE) (DerSimonian and Laird, 1986) is also conservative and is not powerful (Han and Eskin, 2011). To overcome this challenge, we recently developed a modified RE (RE2) that has higher power under condition of heterogeneity (Han and Eskin, 2011). RE2 has been used widely in cross-population human disease analyses (Chimusa

©The Author 2017. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
et al., 2014; Keller et al., 2014; Sapkota et al., 2014), cross-environment mouse trait analyses (Kang et al., 2014), cross-condition expression quantitative trait loci (eQTL) analyses (Sul et al., 2013; Ye et al., 2014), and cross-feature neuroimaging analyses (Hibar et al., 2012; Stein et al., 2012).

However, RE2 has some limitations. First, RE2 cannot perform meta-analysis of correlated statistics. Although the traditional assumption of independence of statistics has been valid in conventional study designs, it can be invalidated in new research designs. For example, in cross-disease meta-analyses, it is common that some controls are used in more than one study, which can cause correlations of statistics (Dichgans et al., 2014; Kar et al., 2016; Moskvina et al., 2013). Thus, in cross-disease analyses, both heterogeneity and correlations can occur. In a cross-tissue eQTL analysis (Sul et al., 2013), the intra-individual similarity of gene expression levels between different tissues can cause the correlations of statistics. To account for these correlations, Lin and Sullivan extended FE (Lin and Sullivan, 2009). However, for RE methods, no solutions have been suggested. Recently, Han et al. developed a decoupling approach that makes the statistics independent (Han et al., 2016). The transformed data can be used for RE2. However, the optimality of this approach has not been evaluated yet. The second limitation of RE2 is that the identified variants by RE2 and FE overlap substantially. This is because RE2 is designed as a stand-alone method that captures variants with and without heterogeneity. However, in most of the meta-analyses of GWASs, it is essential to apply FE before applying RE2, because detecting variants with homogeneous effects is of primary interest. To the best of our knowledge, all investigators who employed RE2 for meta-analyses of GWASs used RE2 coupled with FE. Considering this practical situation, the current implementation of RE2 could be suboptimal.

In the present study, we propose a new method, called RE2C, which increases the power of RE2 in two ways. First, we generalized the likelihood model of RE2 to account for correlations of statistics and to achieve optimal power. To estimate the maximum likelihood estimators of parameters efficiently, we developed an optimization procedure based on spectral decomposition of the variance-covariance matrix. Second, we modified the statistic to focus on the heterogeneous effects that cannot be detected by FE. This modification increased the power to identify new associations after the application of FE. The statistic does not follow a known asymptotic distribution; therefore, we developed an efficient and accurate P-value approximation procedure using analytical decomposition of the statistic. In our simulations, RE2C achieved a dramatic increase in power compared with competing approaches, such as the decoupling approach (71% vs. 21%) when the statistics were correlated. Even when the statistics were uncorrelated, RE2C achieved a modest increase in power. Applications to real genetic data demonstrated that RE2C improved the significances of the associated variants. RE2C is efficient and can meta-analyze one hundred GWASs within one day.

The software is available at http://software.buhmhan.com/RE2C.

2 Materials and methods

2.1 Existing meta-analysis methods for independent statistics

2.1.1 Fixed effects model

The FE method assumes that the magnitude of the true effect is common or fixed in every study in the meta-analysis. The inverse-variance-weighted effect-size method (Cochran, 1954; de Bakker et al., 2008; Fleiss, 1993; Mantel and Haenszel, 1959) and the weighted sum-of-z-scores method (de Bakker et al., 2008; Han and Eskin, 2011; Zaykin, 2011) are used widely. We only describe the former, because the two methods are approximately equivalent (Lee et al., 2016). Let \( X_1, \ldots, X_N \) be the effect-size estimates, such as log odds ratios or regression coefficients, in \( N \) independent studies.

Under the FE model, the observed effect \( X_i \) of study \( i \) is the sum of the true common effect \( \mu \) and the within-study error \( \epsilon_i \):

\[
X_i = \mu + \epsilon_i.
\]

If the sample sizes of the studies are sufficiently large, \( X_i \) is normally distributed. Let \( SE(X_i) \) be the standard error of \( X_i \) and let \( V_i = SE(X_i)^2 \). It is common practice to use the estimated sample variance for \( V_i \). Let \( W_i = V_i^{-1} \) be the inverse variance. The inverse-variance-weighted effect-size estimator is the sum of \( X_i \) weighted with weights \( W_i \):

\[
X_{FE} = \frac{\sum W_i X_i}{\sum W_i}.
\]

The variance of \( X_{FE} \) is

\[
V_{FE} = \frac{1}{\sum W_i}.
\]

It follows that the standard error of \( X_{FE} \) is \( SE(X_{FE}) = (\sum W_i)^{-1/2} \). Note that \( SE(X_{FE}) \) is minimized only if the weights are inverse variances, which explains the method’s name (Cochran, 1954; Greene, 2012; Lee et al., 2016). We can then build a summary z-score,

\[
Z_{FE} = \frac{X_{FE}}{SE(X_{FE})} = \frac{\sum W_i X_i}{\sqrt{\sum W_i}},
\]

which follows \( N(0,1) \) under the null hypothesis of no association \((H_0 : \mu = 0)\). The P-value can be calculated as

\[
p_{FE} = 2 \Phi(-|Z_{FE}|),
\]

where \( \Phi \) is the cumulative density function of the standard normal distribution.

2.1.2 Random effects model (traditional)

In contrast to FE, the RE method models heterogeneity explicitly and assumes that the true value of the effect size \( \mu_i \) of each study is sampled from an underlying distribution. Suppose that the distribution has mean \( \mu \) and variance \( \tau^2 \). The observed effect \( X_i \) is then the sum of the common effect \( \mu \) and the deviation of the \( i \)th study’s observed effect from \( \mu \), say \( \delta_i = (\mu_i - \mu) + \epsilon_i \) (Cochran, 1954) such that

\[
X_i = \mu + \delta_i,
\]

where the within-study error \( \epsilon_i \) is uncorrelated with the true effect sizes \( \mu_i \). The variance in \( X_i \) is the sum of the between-study variance and the within-study variance (Western and Blome, 2009),

\[
V(\delta_i) = W_i^{-1} + \tau^2.
\]

The most popular approach to estimate \( \tau^2 \) is the method of moments proposed by DerSimonian and Laird (DerSimonian and Laird, 1986, 2015). Given the estimated between-study variance \( \hat{\tau}^2 \), the RE effect size is calculated similarly to Equation (1):

\[
X_{RE} = \frac{\sum w_i^* X_i}{\sum w_i^*},
\]

where the weights are now \( w_i^* = \left(W_i^{-1} + \hat{\tau}^2\right)^{-1} \) instead of \( W_i \). Note that \( SE(X_{RE}) = \left(\sum w_i^*\right)^{-1/2} \).
Similarly to FE, we can construct a z-score statistic
\[
Z_{RE} = \frac{X_{RE}}{SE(X_{RE})},
\]
and the P-value is
\[
P_{RE} = 2\Phi(-|Z_{RE}|).
\]
The traditional RE approach is equivalent to a likelihood ratio test that assumes the same heterogeneity under both the null and the alternative hypotheses (Han and Eskin, 2011). This assumption can be conservative in GWASs, therefore, RE has limited power in GWASs (Han and Eskin, 2011).

2.1.3 RE2 (Han and Eskin)
Han and Eskin proposed a modified RE method (RE2) that has better power than RE or FE under conditions of effect size heterogeneity (Han and Eskin, 2011). The key difference between RE and RE2 is that the latter assumes no heterogeneity under the null hypothesis. This assumption is appropriate in many situations of GWASs where we expect that the effect sizes are all zero under the null hypothesis. The method is a likelihood ratio test that has the fixed parameters \(\mu = 0\) and \(\tau^2 = 0\) under the null hypothesis, as follows:
\[
L_0 = \prod_i \frac{1}{\sqrt{2\pi V_i}} \exp \left( \frac{-X_i^2}{2V_i} \right),
\]
\[
L_1 = \prod_i \frac{1}{\sqrt{2\pi (V_i + \tau^2)}} \exp \left( \frac{(X_i - \mu)^2}{2(V_i + \tau^2)} \right).
\]
The roots of the partial derivatives of the equation (3) are not in a closed form; therefore, the maximum likelihood (ML) estimates \(\hat{\mu}\) and \(\hat{\tau}^2\) must be determined by using an iterative procedure. Hardy and Thompson suggested a simple and efficient procedure based on the Newton–Raphson method (Han and Eskin, 2011; Hardy and Thompson, 1996). Given \(\hat{\mu}\) and \(\hat{\tau}^2\), the likelihood ratio statistic can be constructed as follows:
\[
S_{RE2} = \sum_i \log \left( \frac{V_i}{V_i + \hat{\tau}^2} \right) + \sum_i \frac{X_i^2}{V_i} - \sum_i \frac{(X_i - \hat{\mu})^2}{V_i + \hat{\tau}^2}.
\]
The value of \(\hat{\tau}^2\) is restricted to be non-negative; therefore, as shown by Self and Liang (Self and Liang, 1987), the statistic follows a 50:50 mixture of \(\chi^2_1\) and \(\chi^2_2\) asymptotically. Thus, the asymptotic P-value is
\[
P_{RE2} = 0.5 \cdot \Phi(\chi^2_1 \geq S_{RE2}) + 0.5 \cdot \Phi(\chi^2_1 \geq S_{RE2})
\]
In practice, because of the small number of studies \((N)\), a tabulated correction is necessary for an accurate P-value. We pre-calculated the P-value and the update P-value is
\[
P_{RE2} = \lambda(N, S_{RE2}) \cdot P_{RE2}
\]
where \(\lambda(N, S_{RE2})\) is the small sample correction factor.

2.2 Existing meta-analysis methods for correlated statistics
2.2.1 The Lin-Sullivan method
Historically, meta-analysis methods focused mainly on summarizing independent estimates. However, in recent research design, the statistics are often correlated, for example, because of overlapping subjects, which is common in cross-disease meta-analysis. Lin and Sullivan (Lin and Sullivan, 2009) developed a meta-analysis solution to account for these correlations. First, they showed that the correlations of statistics could be calculated analytically. For example, in a case/control design, the correlation between statistics of studies \(i\) and \(j\) is approximated as
\[
\gamma_{ij} \approx \left( \frac{n_{ij} \mu_{i} \mu_{j}}{n_{i} \mu_{i}^2 + n_{j} \mu_{j}^2} \right)^{1/2} \left( \frac{n_{ij} \mu_{i} \mu_{j}}{n_{i} \mu_{i}^2 + n_{j} \mu_{j}^2} \right)^{1/2}.
\]
where \(n_{ij}, n_{i}\) and \(n_{j}\) are the total number of \(r\)th and \(j\)th studies and the number of overlapping subjects between the two \((r\)th and \(j\)th), respectively. Subscripts 1 and 0 denote the case and control status. Let \(C = [\theta]_{N \times N}\) be the correlation matrix of \(X = \{X_1, \ldots, X_N\}\). Given \(C\), one can easily calculate the variance-covariance matrix, \(\Omega\). Lin and Sullivan suggested a statistic:
\[
Z_{LS} = \frac{\hat{\tau} \Omega^{-1} X}{\hat{\tau} \Omega^{-1} \hat{\tau}}
\]
where \(\hat{\tau}\) is an \(N \times 1\) vector with ones. The variance is \(Var(X_{LS}) = (\hat{\tau} \Omega^{-1} \hat{\tau})^{-1}\). Therefore, one can obtain a z-score as well as a P-value (Lin and Sullivan, 2009). This method does not assume heterogeneity; therefore, it can be considered as an extension of FE to account for correlations.

2.2.2 The decoupling method
Recently, Han et al. (Han et al., 2016) proposed a method called “decoupling” that can transform correlated data into independent data. As Lin and Sullivan showed, in many situations, the correlation matrix \(C\) can be approximated analytically before the meta-analysis. Han et al. calculate a transformed covariance structure:
\[
\Omega_{\text{decoupled}} = \text{diag} \left( \frac{\hat{\tau}}{\hat{\tau}} \cdot \text{diag}(s) \cdot C \cdot \text{diag}(s)^{-1} \right)^{-1}
\]
where \(s\) is the vector of standard errors, and \(\text{diag}(s)\) is a diagonal matrix whose diagonals are \(s\). The updated standard errors then become
\[
SE_{\text{decoupled}} = \sqrt{\Omega_{\text{decoupled}}[i, i]}
\]
where \(\Omega_{\text{decoupled}}[i, i]\) denotes the \(r\)th diagonal element of \(\Omega_{\text{decoupled}}\). The data become independent, and thus can be used for RE2 as well as FE. Han et al. showed that when the decoupled data are used for FE, the method is analytically equivalent to the Lin-Sullivan method. Han et al. also showed that under conditions of heterogeneity, RE2 with decoupling (Decoupling-RE2) shows a higher power than FE with decoupling. However, the optimality of Decoupling-RE2 has not been evaluated.

2.3 RE2C
In the present study, we propose RE2C, a powerful random effects method for meta-analysis of GWASs. RE2C is built upon RE2, but with two modifications that improve its power: (1) accounting optimally for correlations, and (2) focusing on heterogeneous effects conditioned on the application of FE. C in RE2C refers to both correlations and conditioning.

2.3.1 Optimizing for the meta-analysis of correlated datasets
We extended the RE2 model to include correlations between statistics. Let \(x\) be the length \(n\) vector denoting the observed effect sizes. Then, we could build a model
\[
x = \mu \hat{\tau} + u + \epsilon
\]
where \(u \sim N(0, \Sigma^2 I)\) are random effects reflecting between-study heterogeneity and \(\epsilon \sim N(0, \Sigma)\) are random errors. Given
the correlation of statistics C, which can be approximated analytically using the Lin-Sullivan approach (Lin and Sullivan, 2009), we have $\Sigma = \text{diag}(\alpha) \cdot C \cdot \text{diag}(\alpha)$ where $\alpha$ is the vector of the standard errors. Then, the variance-covariance matrix of $x$ is

$$H = \text{Var}(x) = \Sigma + \tau^2I.$$

The likelihood functions under the null and alternative hypotheses become

$$L_0 = (2\pi)^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} x^T \Sigma^{-1} x \right)$$

and

$$L_1 = (2\pi)^{-\frac{1}{2}} |H|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} x^T (x - \mu\hat{e})^T H^{-1} (x - \mu\hat{e}) \right).$$

To build a likelihood ratio test, we must find the maximum likelihood estimation (MLE) of the parameters $\mu$ and $\tau$. Previously, for independent statistics, RE2 utilized an iterative procedure suggested by Hardy and Thompson (Hardy and Thompson, 1996). However, their method only considers independent statistics. Therefore, we developed an optimization procedure that can be applied efficiently for both independent and correlated statistics. We chose to use the technique developed for the restricted maximum likelihood (REML) framework. The key idea of our optimization procedure is to transform the two-dimensional search into a one-dimension search using the technique that was developed by Patterson and Thompson (Patterson and Thompson, 1971). A similar technique has been used previously to correct for population stratifications (Kang et al., 2008).

We decomposed the observations using a direct sum, where one of the decomposed observations is the observation for the REML function after integrating out the fixed effects. That is, we decomposed $x$ into two matrix-vector multiplications of $S$ and $Q$ such as:

$$x = Sx + Qs.$$

where $S$ is a transformation matrix of rank $n - 1$ and $Q$ is a transformation matrix of rank 1. The specific forms of $S$ and $Q$ for our purpose are described below. The properties of the direct sum mean that the log-likelihood function of the mixed model ($l_1$) can be decomposed into two log-likelihood functions of independent observations as follows:

$$l_1 = l_0 + l_1^*.$$  

The projection matrix $S$ is an idempotent and symmetrical matrix that integrates out the fixed effects (mean) of the observation $x$. In our problem, matrix $S$ is:

$$S = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix} \frac{1}{n}.$$  

Here, $\hat{e}$ is a vector of ones of size $n$. The matrix $S$ satisfies $E(Se) = 0$, i.e. $S \cdot \hat{e} = 0$. Then, matrix $Q$ becomes:

$$Q = \hat{e}^T H^{-1}.$$  

Matrix $Q$ satisfies the conditions $\text{cov}(Sx, Qs) = 0$ and $\text{SHQ}' = 0$. Next, we considered the full log-likelihood $l_1$ with the parameters of interest $\tau^2$ and $\mu$ as follows

$$l_1(\mu, \tau^2 | x) = -\frac{1}{2} \left[ \text{log}(2\pi) + \text{log}|H| + (x - \mu\hat{e})^T H^{-1} (x - \mu\hat{e}) \right]$$

$S$ is an orthogonal projection matrix; therefore, $S$ is in the form of $AA^T$, where $A$ is an $n \times (n - 1)$ matrix with orthonormal columns, such that $AA^T = I$. To reduce the complexity of the restricted likelihood function $l_0^*$, for $Sx$, Harville (1974) suggested the use of the restricted likelihood function for $A'x$, where the MLE for the two likelihood functions are the same. As Harville showed, the restricted likelihood can be shown as:

$$l_1^*(\tau^2 | A'x) = -\frac{1}{2} \left[ (n - 1) \text{log}(2\pi) + \text{log}|A'HA| + (A'x)'(A'HA)^{-1}(A'x) \right]$$

$$= -\frac{1}{2} \left[ \text{log}(2\pi) + \text{log}|H| + (x - \mu\hat{e})^T H^{-1} (x - \mu\hat{e}) \right]$$

$$+ \frac{1}{2} \left[ \text{log}(2\pi) + \text{log}|\hat{e}^T \hat{e}| - \text{log}|\hat{e}^T H^{-1} \hat{e}| \right],$$

where $\hat{\mu} = x^T H^{-1} (x - \mu\hat{e})$. Let the orthogonal matrix, $B$, be the eigenvectors of the matrix $AHA$ such that $B' AHA$ is diagonal. Then $P = AB$. The matrix $P$ then has the following properties: (i) $PP' = I$, (ii) $PP' = S$, (iii) $SP = P$ and (iv) $PS = P'$. Using the spectral decomposition framework, the symmetric matrix $SHS$ can be shown as:

$$SHS = P \text{diag}(\xi_{S,S} + \tau^2 \bar{I})P'$$

Here, we considered the full (not restricted) likelihood function whose $\mu$ is substituted with $\hat{\mu}$.

$$l_1(\tau^2 | x) = -\frac{1}{2} \left[ \text{log}(2\pi) + \text{log}|H| + (x - \mu\hat{e})^T H^{-1} (x - \mu\hat{e}) \right].$$

For our problem of finding the MLE, this modified function is sufficient, because it satisfies that $\mu = \hat{\mu}$ at the MLE. Note that although we focused on the full likelihood function to build a likelihood ratio test, the same optimization procedure below can be applied to the restricted likelihood function. Following Equation (4), we could define the generalized inverse of the matrix $SHS$, $(SHS)^+$, which is

$$(SHS)^+ = P \text{diag}(\xi_{S,S} + \tau^2 \bar{I})^{-1} P'.$$

Next, we could transform $(x - \mu\hat{e})^T H^{-1} (x - \mu\hat{e})$ into a simpler expression as follows:

$$(x - \mu\hat{e})^T H^{-1} (x - \mu\hat{e}) = x^T S (SHS)^+ S x$$

$$= x^T P \text{diag}(\xi_{S,S} + \tau^2 \bar{I})^{-1} P x.$$

Thus, the likelihood becomes

$$l_1(\tau^2 | x) = -\frac{1}{2} \left[ \text{log}(2\pi) + \sum_{i=1}^{n} \text{log}(\xi_{S,i} + \tau^2) \right] + \sum_{i=1}^{n} \left( \frac{n}{\xi_{S,i} + \tau^2} \right).$$
Increasing the power of meta-analysis for heterogeneous effects

where the scalar value $\zeta_{X,l}$ is the $r$th eigenvalue of the matrix $\Sigma$, and $\eta_i$ is the $i$th component of the vector $P'x$. Now, the transformation has reduced the number of parameters to one ($\tau^2$). Thus, we can use a simple Newton-Raphson procedure to estimate the unknown parameter, $\tau^2$. The first and the second derivatives of the transformed log-likelihood functions are:

$$
\frac{d}{d\tau^2} \log L_{\tau^2} = -\frac{1}{2} \sum_{i=1}^{n} \left( \frac{1}{\tau^2} - \frac{1}{(\zeta_{X,l} + \tau^2)} \right) \left( \frac{\eta_i^2}{(S_{\tau^2} + \tau^2)^2} \right)
$$

$$
\frac{d^2}{d\tau^2} \frac{d}{d\tau^2} \log L_{\tau^2} = -\frac{1}{2} \sum_{i=1}^{n} \left( \frac{1}{\tau^2} - \frac{1}{(\zeta_{X,l} + \tau^2)} \right) \left( \frac{\eta_i^2}{(S_{\tau^2} + \tau^2)^2} \right)
$$

In summary, using this optimization procedure, the parameter estimation needs only the application of the Newton-Raphson method to a single parameter, which is very efficient. Thus, we have a high chance of obtaining the global optimum using a grid search as the starting point for the Newton-Raphson procedure. After we find the MLE, we can build a likelihood ratio test statistic,

$$
\log \left( \frac{\Sigma}{\Sigma + \tau^2} \right) + x'\Sigma^{-1}x - (x - \bar{\mu})'\left( \Sigma + \tau^2 I \right)^{-1}(x - \bar{\mu})
$$

which follows a 50:50 mixture of $\chi^2_1$ and $\chi^2_2$ asymptotically.

2.3.2 Focusing on heterogeneous effects

We then modified the test procedure of RE2 to focus on heterogeneous effects. In most meta-analyses of GWASs, detecting variants with homogeneous effects is of primary interest. For this reason, it is often essential to apply FE before applying RE2, while accounting for the increased multiple testing burden. We surveyed the literature that cited and used RE2, at least in all the > 50 papers that we examined, the studies used RE2 coupled with FE. Thus, considering this unique situation of meta-analysis of GWASs, where the prior application of FE is mandatory, we can improve the power of RE2 by focusing on the heterogeneous effects that would not be identified by FE.

Specifically, we designed a statistic as follows,

$$
S_{RE2C} = \begin{cases} 
S_{RE2} & \text{if } p_{RE2} \leq p_{FE} \\
0 & \text{if } p_{RE2} > p_{FE}
\end{cases}
$$

In short, this statistic can become significant only if the RE2 $P$-value is more significant than the FE $P$-value. Although the statistic looks simple, calculating the $P$-value of this statistic is non-trivial. Obviously, unlike RE2, this statistic does not follow a known asymptotic distribution. One possible way is to use a resampling approach that samples null z-scores repeatedly. However, $P$-values typically observed in GWASs are extremely small ($< 5 \times 10^{-8}$). To estimate such a small $P$-value using resampling, a large number of samplings are required. Thus, in GWASs where millions of markers are analyzed, resampling can be very slow.

To approximate the $P$-value of the new method efficiently, we used the following strategy. Recall that the RE2 statistic is a likelihood ratio statistic that measures the difference between the two likelihoods: $L_0$ in Equation (2) and $L_1$ in Equation (3). We introduced an intermediate likelihood function,

$$
L_{int} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi V_i}} \exp \left( -\frac{(X_i - \mu)^2}{2V_i} \right)
$$

which is similar to $L_1$, but with a restriction of $\epsilon^2 = 0$. Then, the RE2 statistic can be decomposed into the sum of the difference between $L_0$ and $L_{int}$ and the difference between $L_{int}$ and $L_1$, as follows (Han and Eskin, 2011):

$$
S_{RE2} = \log \left( \frac{\sup \{L_1(x^2, \mu(X_i, V_i)\} \}}{\sup \{L_0(\mu(X_i, V_i)\} \} \right)^2
$$

$$
= \log \left( \frac{\sup \{L_{int}(\mu(X_i, V_i)\} \}}{\sup \{L_0(\mu(X_i, V_i)\} \} \right)^2 + \log \left( \frac{\sup \{L_{int}(\mu(X_i, V_i)\} \}}{\sup \{L_{int}(\mu(X_i, V_i)\} \} \right)^2
$$

$$
= S_{FE} + S_{Het}
$$

where $\Theta$ indicates an empty set. The first statistic, $S_{FE}$, is equal to the square of the FE statistic ($\chi^2_{FE}$). The second statistic, $S_{Het}$, tests for nonzero between-study variance, similar to the Cochran’s Q test. The two statistics are independent under the null hypothesis (Self and Liang, 1987). Asymptotically, $S_{FE}$ follows $\chi^2_1$, and $S_{Het}$ follows a 50:50 mixture of 0 and $\chi^2_2$. However, the conditions for them to follow their asymptotic distributions are different. Under the assumption that the effect size ($X_i$) follows a normal distribution due to a large sample in each study, which is the case in GWASs, $S_{FE}$ follows $\chi^2_1$ regardless of the number of studies (N). However, even under the normality assumption, $S_{Het}$ follows a 50:50 mixture of 0 and $\chi^2_2$ only if $N$ is large. $N$ is small in typical meta-analysis of GWASs; therefore, the true distribution of $S_{Het}$ deviates greatly from the asymptotic distribution. For our method, we approximated and tabulated the distribution of $S_{Het}$ empirically for every possible $N$.

In the previous section, we extended the RE2 model to account for correlations between statistics. Equation (5) can also be decomposed into two parts,

$$
S_{FE} = [x'\Sigma^{-1}x - (x - \bar{\mu}X_{LS})'\Sigma^{-1}(x - \bar{\mu}X_{LS})],
$$

$$
S_{Het} = \log \left( \frac{\Sigma}{\Sigma + \tau^2} \right) + (x - \bar{\mu}X_{LS})'\Sigma^{-1}(x - \bar{\mu}X_{LS})
$$

$$
- (x - \bar{\mu})'\left( \Sigma + \tau^2 I \right)^{-1}(x - \bar{\mu})
$$

where $X_{LS}$ is the Lin-Sullivan estimator of $\mu$, which is equal to $\sup \{L_{int}(\mu(X_i, \Sigma)\}$. $S_{FE}$ is equivalent to the square of the z-score of the Lin-Sullivan method in this situation.

Now that the RE2 statistic can be decomposed into $S_{FE}$ and $S_{Het}$ whose null distributions are known, given an observed RE2 statistic, its $P$-value can be interpreted as an integral over a region in the two-dimensional space. Specifically, in Figure 1, the RE2 $P$-value is the volume of the region excluding the bottom left triangle (i.e. region $A + B$). However, in RE2C, we only consider the region where $p_{RE2} \leq p_{FE}$. Thus, for each $S_{Het}$, we can search for $S_{Het}$ that would satisfy $p_{RE2} \leq p_{FE}$, or

$$
\lambda(N, S_{FE} + S_{Het}) \cdot [0.5 \cdot P(\chi^2_1 > S_{FE} + S_{Het}) + 0.5 \cdot P(\chi^2_2 > S_{FE} + S_{Het})] \leq P(\chi^2_1 > S_{FE}).
$$

Let this lower boundary of $S_{Het}$ that satisfies $p_{RE2} \leq p_{FE}$ be $S_{Het, low}(S_{FE}, N)$. This boundary is plotted as a dashed line in Figure 1. Then, given an observed RE2 statistic $S_{RE2C}$, we calculated the $P$-value as follows. We divided the range of $S_{FE}$ into $K$
small bins (e.g., 1000 bins in [0, 50]), denoted as \(x_i (i = 1, \ldots, K)\). The approximated \(P\)-value is

\[
p_{\text{RE2C}} \approx \sum_{i=1}^{K} P \left( S_{\text{het}} > \max(S_{\text{RE2C}} - S_{\text{het}}) \right) \Delta x,
\]

where \(\Delta x\) is the width of the bins. That is, we calculated the probability that \(S_{\text{het}}\) would be large enough to satisfy \(S_{\text{RE2C}} \geq S_{\text{RE2C}}\) for every bin of \(S_{\text{RE2C}}\), and integrated them together. We took the maximum function because if \(S_{\text{het}}\) is smaller than \(S_{\text{het}}(x_i, N)\), then \(S_{\text{RE2C}} = 0\) by definition. Thus, we calculated the volume of region A in Figure 1. As a result, it always satisfies the equation:

\[
p_{\text{RE2C}} < p_{\text{RE2}}
\]

as long as \(p_{\text{RE2}} \leq p_{\text{FE}}\), because we have removed region B in Figure 1. This shows that the \(\text{RE2C}\) \(P\)-value can never be less significant than the \(\text{RE2}\) \(P\)-value when those methods are used coupled with \(\text{FE}\), for the variants with \(p_{\text{RE2}} \leq p_{\text{FE}}\). Note that the calculation of the \(P\)-value is efficient because we have pre-calculated \(S_{\text{het}}(x, N)\) for every \(x\) and \(N\) and the cumulative density function of \(S_{\text{het}}\) for every \(N\). Thus, the computational complexity is only \(O(K)\).

Moreover, the complexity is not dependent on how small the \(P\)-value is, unlike in the resampling approaches.

3 Results

3.1 Simulations

We evaluated the performance of \(\text{RE2C}\) using simulations. We assumed seven studies, each of which comprised 2000 individuals, half of which were controls and half were cases. We assumed a SNP with a minor allele frequency (MAF) of 0.1, following the Hardy-Weinberg equilibrium.

3.1.1 False positive rate

We assumed the null hypothesis of no association and evaluated the false positive rate of \(\text{RE2C}\). We repeated the null simulations \(10^9\) times and estimated the false positive rate as the proportion of the repeats whose \(P\)-value was \(\leq \alpha\), where \(\alpha \in \{5 \times 10^{-2}, 5 \times 10^{-4}, 5 \times 10^{-6}\}\). Table 1 shows that the false positive rates of \(\text{RE2C}\) were well calibrated. We then assumed that the statistics were correlated, with a correlation coefficient \(\rho = 0.4\). The false positive rates for the correlated statistics were also controlled (Table 1). There was a slight conservative tendency, which was possibly caused by the errors in our approximation of \(P\)-values using bins. However, the discrepancies were very small.

<table>
<thead>
<tr>
<th>(x)</th>
<th>(5.0 \times 10^{-2})</th>
<th>(5.0 \times 10^{-4})</th>
<th>(5.0 \times 10^{-6})</th>
<th>(5.0 \times 10^{-8})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent</td>
<td>4.8 (10^{-2})</td>
<td>4.8 (10^{-4})</td>
<td>4.7 (10^{-6})</td>
<td>5.5 (10^{-8})</td>
</tr>
<tr>
<td>Correlated ((\rho = 0.4))</td>
<td>4.7 (10^{-2})</td>
<td>4.6 (10^{-4})</td>
<td>4.5 (10^{-6})</td>
<td>4.0 (10^{-8})</td>
</tr>
</tbody>
</table>

3.1.2 Power for independent statistics

We compared the powers of \(\text{FE}, \text{RE2}\) and \(\text{RE2C}\). We generated 10,000 sets for meta-analysis, where we again assumed seven studies with sample size equal to 2000 and a MAF of 0.1. In our simulations, we considered the practical situations that \(\text{FE}\) was already applied before the application of \(\text{RE2}\) or \(\text{RE2C}\). Thus, we considered the combined use of \(\text{RE2}\) (or \(\text{RE2C}\)) with \(\text{FE}\) where multiple tests were accounted. Specifically, the power of \(\text{FE}\) was the proportion of the sets whose \(P\)-value exceeded the genome-wide threshold \(p_{\text{GWAS}} = 5 \times 10^{-8}\). The power of \(\text{RE2}\) (or \(\text{RE2C}\)) was the proportion of the sets whose \(\text{FE}\) or \(\text{RE2}\) (or \(\text{RE2C}\)) \(P\)-value exceeded \(p_{\text{GWAS}}/2 = 2.5 \times 10^{-8}\). To model the effect size heterogeneity in our simulations, we assumed four different effect size distributions. Let \(\mu\) be a specific, assumed target log odds ratio. The four distributions were as follows, in order of increasing amount of heterogeneity.

First, we assumed a unimodal distribution that was a normal distribution with mean \(\mu\) and standard error \(\mu\), truncated to \([0, 2\mu]\). Second, we assumed a uniform distribution spanning \([0, 2\mu]\). Third, we assumed a binomial distribution that followed \(N(0, \mu^2)\) truncated to \([0, \mu]\) with one half probability, and \(N(2\mu, \mu^2)\) truncated to \([\mu, 2\mu]\) with another half probability. These three distributions all had mean \(\mu\). Finally, we assumed a distribution representing opposite effects, which followed \(N[1(1, 2\mu, \mu^2)]\) with one-half probability and \(N[1(2\mu, \mu^2)]\) with another half probability. Although opposite effects between studies can be rare in genetic studies of the same disease, they can occur in cross-disease meta-analyses or cross-tissue eQTL analyses. Once we assumed one of the distributions above, we randomly sampled \(b_k\), the log odds ratio in study \(k \in \{1, \ldots, K\}\), from the distribution. We then sampled the minor allele counts in control and case samples assuming the control and case MAF, respectively. The control MAF was assumed to be the same as the population MAF (0.1), assuming a very small prevalence, and the case MAF was \(\text{MAF}_{\text{cases}} = \text{MAF}_{\text{control}}/(e^{b_k} - 1) \times \text{MAF}_{\text{control}} + 1\). For effective comparisons of power, we adjusted \(\mu\) for each distribution such that the power of the most powerful method was approximately 70%.

Figure 2 shows the power comparison results. The powers of \(\text{RE2}\) and \(\text{RE2C}\) are shown as stacked bars. We assumed a prior application of \(\text{FE}\) to random effect methods; therefore, we applied a different color scheme to the proportion of datasets determined as significant by \(\text{FE}\) (light grey) and the proportion of datasets where the random effect methods newly identified as significant (dark grey). Note that the height of light grey bar is slightly shrunk in \(\text{RE2}/\text{RE2C}\) compared to \(\text{FE}\), because the significance level was adjusted to one-half. As the heterogeneity increases, the combined use of the random effect methods with \(\text{FE}\) gave increasingly higher powers than compared with using \(\text{FE}\) alone, as expected. Under all tested scenarios of effect size distributions, \(\text{RE2C}\) was the most powerful. \(\text{RE2C}\) increased power of \(\text{RE2}\) by 1.55, 1.85, 2.07 and 2.98% for unimodal, uniform, bimodal and opposite effects, respectively. Although the increase in the absolute amount of power was modest, the increase in relative power gain compared with \(\text{FE}\) was non-negligible. For example, in the unimodal distribution, the power gain of \(\text{RE2C}\) from \(\text{FE}\) was 1.71%, which was more than 10 times greater than that of \(\text{RE2}\) (0.16%).

\[2\text{D} \text{representation of } S_{\text{RE}}, p_{\text{RE}} \text{ and } S_{\text{het}}\]
Increasing the power of meta-analysis for heterogeneous effects

**Fig. 2.** Power of FE, RE2 and our new RE2C method for the meta-analysis of independent statistics. Assuming the statistics are independent, we simulated various effect size distributions with differing amounts of heterogeneity. We considered the scenario that RE2 or RE2C is additionally applied to FE while accounting for multiple testing. The power of RE2 and RE2C are shown as two-color stacked bars, where we colored the proportion identified by FE as significant in light grey and the proportion that RE2/RE2C additionally identified as significant in dark grey.

**Fig. 3.** Power of Lin-Sullivan (LS), Decoupling-RE2 (DR2) and our new RE2C method for meta-analyzing correlated statistics. Assuming statistics are correlated with correlation coefficient \( \rho \), we simulated various effect size distributions with differing amount of heterogeneity. We considered the scenario that DR2 or RE2C is additionally applied to LS while accounting for multiple testing. DR2 and RE2C power is shown as two-color stacked bars, where we colored the proportion that LS was significant in light grey and the proportion that DR2/RE2C additionally identified as significant in dark grey.
correlated statistics but without the technique for focusing on heterogeneous effects. The most significant end, we used the cross-disease analysis data of Moskvina et al. 3.2 Applications to real data

$q$ sampled the effect sizes of the studies, we generated the observed effect under the situation that the statistics were correlated. After we 3.1.3 Power for correlated statistics

Using a similar simulation scheme, we evaluated the power of RE2C under the situation that the statistics were correlated. After we sampled the effect sizes of the studies, we generated the observed effect sizes assuming that they were correlated with correlation coefficient $\rho$. We assumed $\rho = 0.1$ and $\rho = 0.4$, and calculated the power for each setting. The value $\rho = 0.4$ was derived from assuming a cross-disease analysis with 2000 cases and 3000 shared controls (Wellcome Trust Case Control Consortium, 2007). The competing approaches in this simulation were the Lin-Sullivan (LS) and the Decoupling-RE2 (DR2) methods. As described in the Methods, the Lin-Sullivan method is an extended FE method to account for correlations. Decoupling-RE2 refers to the application of the transformed data by decoupling approaches where $\rho = 0.4$, RE2C achieved 71% power while the power of Lin-Sullivan method and Decoupling-RE2 were only 23.8% and 21.4% respectively. Surprisingly, Decoupling-RE2 performed poorly even for large heterogeneity when the correlations were high ($\rho = 0.4$). This demonstrates that although the application of the decoupled data to RE2 is possible, it may not provide optimal power.

3.2 Applications to real data

We wanted to evaluate the utility of RE2C for real data. To this end, we used the cross-disease analysis data of Moskvina et al. (Moskvina et al., 2013) who performed a meta-analysis of association results for the Alzheimer’s disease (AD) and the Parkinson’s disease (PD). Moskvina et al. examined the meta-analyses $P$-values of 10 loci known to be associated with AD and 18 loci known to be associated with PD. The two diseases shared some controls ($N_{\text{shared}} = 5571$); therefore, there were correlations between the statistics of the two diseases. To account for these correlations, Moskvina et al. used the Lin-Sullivan method. However, the same variant may have differing effects on the two diseases. Therefore, random effect methods might help in association tests. We obtained the reported effect sizes (OR) and $P$-values for these 28 loci from the table shown in their manuscript. We then calculated the standard errors from the OR and $P$-values, and used them for meta-analysis. We removed three loci whose OR was 1.00 (because the paper reported only two digits below zero), and applied RE2C to the remaining 25 loci.

Table 2 gives the details of the collected data and the meta-analysis results. Out of 25 loci, LS was the most significant in 13 loci. In all the remaining 12 loci, RE2C was the most significant. Note that for the 13 loci where LS was the most significant, RE2C $P$-values were completely non-significant ($p_{\text{RE2C}} = 1$). This is because RE2C was designed to be used with FE (LS), but focusing only on loci with heterogeneity. We also show the results of an RE2C implementation with optimization for correlated statistics but without the technique for focusing on heterogeneous effects (denoted as RE2C*), which shows that focusing on heterogeneous effects improved $P$-values at these 12 loci. Overall, these results showed that

<table>
<thead>
<tr>
<th>Chr</th>
<th>Base Pairs</th>
<th>SNP</th>
<th>Parkinson Disease</th>
<th>Alzheimer Disease</th>
<th>Methods</th>
<th>LS</th>
<th>DR2</th>
<th>RE2C*</th>
<th>RE2C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>207 819 492</td>
<td>1-207819492</td>
<td>0.61</td>
<td>0.062</td>
<td>0.50</td>
<td>0.058</td>
<td>0.016</td>
<td>0.019</td>
<td>0.02389</td>
</tr>
<tr>
<td>2</td>
<td>127 892 810</td>
<td>rs6733839</td>
<td>1.07</td>
<td>0.0098</td>
<td>1.23</td>
<td>5.2E-5</td>
<td>0.00029</td>
<td>0.00033</td>
<td>0.00017</td>
</tr>
<tr>
<td>3</td>
<td>47 327 032</td>
<td>rs9367271</td>
<td>1.11</td>
<td>0.0014</td>
<td>1.06</td>
<td>0.339</td>
<td>0.0017</td>
<td>0.00020</td>
<td>0.00029</td>
</tr>
<tr>
<td>4</td>
<td>143 106 884</td>
<td>rs7806047</td>
<td>0.87</td>
<td>0.001</td>
<td>0.89</td>
<td>0.151</td>
<td>0.0007</td>
<td>0.00008</td>
<td>0.00118</td>
</tr>
<tr>
<td>5</td>
<td>27 466 181</td>
<td>rs1532277</td>
<td>0.99</td>
<td>0.709</td>
<td>0.81</td>
<td>1.8E-6</td>
<td>0.024</td>
<td>0.0011</td>
<td>8.1E-05</td>
</tr>
<tr>
<td>6</td>
<td>60 045 900</td>
<td>rs7949816</td>
<td>0.95</td>
<td>0.073</td>
<td>0.82</td>
<td>0.00075</td>
<td>0.0084</td>
<td>0.010</td>
<td>0.00389</td>
</tr>
<tr>
<td>7</td>
<td>85 677 094</td>
<td>1-85677094</td>
<td>1.20</td>
<td>0.0055</td>
<td>1.26</td>
<td>0.057</td>
<td>0.0019</td>
<td>0.0002</td>
<td>0.00014</td>
</tr>
<tr>
<td>8</td>
<td>31 032 228</td>
<td>rs56059558</td>
<td>0.86</td>
<td>0.0023</td>
<td>0.84</td>
<td>0.05</td>
<td>0.0008</td>
<td>0.00009</td>
<td>0.0001302</td>
</tr>
<tr>
<td>9</td>
<td>45 592 254</td>
<td>rs6857</td>
<td>0.95</td>
<td>0.154</td>
<td>5.55</td>
<td>4.4E-92</td>
<td>0.0002</td>
<td>2.9E-53</td>
<td>3.24E-94</td>
</tr>
<tr>
<td>10</td>
<td>51 724 326</td>
<td>rs200656</td>
<td>1.06</td>
<td>0.089</td>
<td>1.06</td>
<td>0.23713</td>
<td>0.055</td>
<td>0.06</td>
<td>0.07461</td>
</tr>
</tbody>
</table>

We compared the results of the Lin-Sullivan method (LS), Decoupling-RE2 (DR2) and RE2C. RE2C* refers to an RE2C implementation with optimization for correlated statistics but without the technique for focusing on heterogeneous effects. The most significant $P$-value among all methods is in bold-face.
if RE2C is used in combination with LS, a high association test power to detect both loci with and without heterogeneity is obtained. Interestingly, RE2C found two loci (rs4698413 and rs2263418) as genome-wide significant ($p_{GWAS} < 2.5 \times 10^{-8}$) that were not identified by LS alone.

We also performed additional real data analyses where statistics were uncorrelated, to demonstrate the performance of RE2C for combining independent datasets. The results are shown in Supplementary Materials (Supplementary Table S1).

### 3.3 Efficiency

We evaluated the efficiency of the methods (Table 3). To this end, we measured the running time of methods for the meta-analysis of differing numbers of studies (from 2 to 100). We timed how long it took to analyze 1,000,000 SNPs. We used the software R to run FE, RE2, and Java to run RE2. RE2C was highly efficient. The estimated time to analyze a million SNPs in a meta-analysis combining 100 studies was 0.07 hours for RE2 and 0.44 hours for RE2C.

Our results imply that RE2C is suitable for future large-scale meta-analyses, where the number of datasets to be combined is expected to grow.

### 4 Conclusion

We proposed a new random effects model meta-analysis method RE2C, which has an improved power for the detection of heterogeneous effects between studies. We optimized the statistic for meta-analyzing correlated statistics, and modified the statistics to only focus on heterogeneous effects. We expect that our method will be applied to a wide range of study designs in the future, such as cross-disease or cross-population studies, to help identify new associations.

### Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) [grant number 2016R1C1B2013126].

Conflict of Interest: none declared.

### References


Patterson,N.H. and Thompson,R. (1971) recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545–554.


