



Published in final edited form as:

*Genet Epidemiol.* 2011 November ; 35(7): 581–591. doi:10.1002/gepi.20603.

## Optimal Methods for Meta-analysis of Genome-wide Association Studies

Baiyu Zhou<sup>1</sup>, Jianxin Shi<sup>2</sup>, and Alice S. Whittemore<sup>1,\*</sup>

<sup>1</sup>Department of Health Research and Policy, Stanford University, Stanford, California

<sup>2</sup>Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

### Abstract

Meta-analysis of genome-wide association studies involves testing single nucleotide polymorphisms (SNPs) using summary statistics that are weighted sums of site-specific score or Wald statistics. This approach avoids having to pool individual-level data. We describe the weights that maximize the power of the summary statistics. For small effect-sizes, any choice of weights yields summary Wald and score statistics with the same power, and the optimal weights are proportional to the square roots of the sites' Fisher information for the SNP's regression coefficient. When SNP effect size is constant across sites, the optimal summary Wald statistic is the well-known inverse-variance-weighted combination of estimated regression coefficients, divided by its standard deviation. We give simple approximations to the optimal weights for various phenotypes, and show that weights proportional to the square roots of study sizes are suboptimal for data from case-control studies with varying case-control ratios, for quantitative trait data when the trait variance differs across sites, for count data when the site-specific mean counts differ, and for survival data with different proportions of failing subjects. Simulations suggest that weights that accommodate inter-site variation in imputation error give little power gain compared to those obtained ignoring imputation uncertainties. We note advantages to combining site-specific score statistics, and we show how they can be used to assess effect-size heterogeneity across sites. The utility of the summary score statistic is illustrated by application to a meta-analysis of schizophrenia data in which only site-specific p-values and directions of association are available.

### Keywords

combining GWAS; effect-size heterogeneity; meta-analysis; noncentrality parameter; score statistics; Wald statistics

### Introduction

Combining data from multiple genome-wide association studies (GWAS) of a common outcome has emerged as a major tool for identifying susceptibility loci for human disease and other conditions [Scott et al., 2007; Zeggini et al., 2007; Shi et al., 2009]. The goals are to discover new variants missed by the individual studies, to identify variants for outcomes not considered when the data were collected (e.g. cancer survival), and, for associated variants, to assess effect-size and its possible variation across sites. A challenge to achieving

---

\*Correspondence to: Alice S. Whittemore, Department of Health Research and Policy, Stanford University School of Medicine, 259 Campus Drive, Stanford, CA 94305-5405, alicesw@stanford.edu, Tel: 650-723-5460.

these goals is the need to assess and accommodate inter-site differences in study characteristics, such as demographic attributes of the populations studied, phenotypic aspects such as disease severity and extent of censoring of survival data, the SNPs included in the genotyping platforms used, and the choice and coding of covariates in need of adjustment.

The pooled analysis of individual-level data on phenotypes, genotypes and covariates has several advantages. Quality control can be implemented uniformly for all sites, and SNP genotypes can be imputed using data from all sites. Covariates common to multiple studies can be coded uniformly and their regression coefficients can be estimated with all available data. Optimal likelihood-based methods can be used to test whether SNP regression coefficients are nonzero and whether they vary across study sites. Offsetting these advantages is the labor involved in assembling the raw data from each site and coding common covariates, and privacy issues which could limit the sharing of genotypes.

An alternative to pooled analysis is the meta-analysis of site-specific test statistics, each having approximately a standard normal distribution under the null hypothesis of no association. Here each site types or imputes genotypes for a common set of SNPs, and then calculates a covariate-adjusted statistic for each SNP. The site-specific statistics are either Wald statistics (ratios of regression coefficient estimates to their estimated standard deviations (SDs), or score statistics (ratios of efficient scores to estimates of their null SDs) [Soranzo et al., 2009; Tanaka et al., 2009]. The coordinating center then computes a weighted sum of these statistics as a summary test statistic for the SNP [Cantor et al., 2010]. For reviews of issues in GWAS meta-analysis, see [Ioannidis 2007; Ioannidis et al., 2007; de Bakker et al., 2008; Guan and Stephen, 2008; Zeggini and Ioannidis, 2009; Cantor et al., 2010].

When SNP effect size is constant across sites, the optimal summary Wald statistic is the well-known inverse-variance-weighted combination of estimated regression coefficients, divided by its estimated standard deviation [Cochran 1954]. However there are practical advantages to combining score statistics rather than Wald statistics for GWAS meta-analyses. First, the score statistics do not require iterative parameter estimates for each SNP to be tested. Estimates for the site-specific covariate effects need be calculated just once under the global null hypothesis of no effect for any SNP; thus score statistics provide fast assessment of the statistical significance of large numbers of SNPs. Second, a SNP score statistic can be computed even when the site has provided only the SNP's p-value and direction of association. Third, the summary score statistic may be easier to interpret when the sites differ with respect to the covariates included in their regression models or with respect to their phenotype definition, measurement or coding. For example, a GWAS of nicotine addiction among current and former smokers might include sites whose outcome is a count of smoking cessation attempts, and other sites that simply classified subjects according to presence or absence of a successful cessation attempt. For a meta-analysis such as this, a combination of site-specific score statistics may be easier to interpret than a combination of site-specific Wald statistics, since the SNP regression coefficients in the latter are not comparable. Finally, site-specific score statistics have straightforward extensions to accommodate uncertainty due to imputation of unobserved genotypes, and we shall show that they can be used to assess heterogeneity of effect-size across sites.

Here we focus on optimal weights for combining site-specific statistics to identify new variants and, for associated variants, to assess possible effect-size variation across sites. Asymptotically optimal weights for combining Wald statistics have been known for more than half a century [Cochran 1954; DerSimonian and Laird, 1986]. Surprisingly however, despite their practical advantages, optimal weights for combining score statistics are not well

known. In fact, some investigators suggest or use weights proportional to the square roots of the study-specific sample sizes [Soranzo et al., 2009; Willer et al., 2010; Hu et al., 2011]. We show that this strategy can be suboptimal when combining data from sites whose phenotypes have different null distributions. Instead, for the small effect-sizes expected of GWAS meta-analysis, the optimal weights for combining score statistics are essentially the same as those used to combine Wald statistics. We provide explicit forms for the optimal weights for a general class of phenotypes that includes binary outcomes, quantitative traits, counts of events and censored survival outcomes.

In the following Methods section we begin by reviewing the score and Wald statistics for testing the null hypothesis of no SNP-phenotype association at a single site, with application to case-control studies, quantitative traits, count data and censored survival data. We also show how the site-specific score statistics can be extended to handle imputation uncertainties, as noted by Marchini et al., [2007]. Then, for each of these phenotypes, we describe the optimal weights for combining site-specific score or Wald statistics. We also extend the weights to handle SNP effect-sizes that vary across sites, and we show how the site-specific score statistics can be used to assess inter-site heterogeneity. The Methods section is followed by an evaluation of power for various summary score and Wald statistics, based on simulated case-control and censored survival data. An application to schizophrenia data shows that for binary outcomes, the score statistics can be used to assess inter-site effect-size heterogeneity even when only site-specific SNP p-values and direction of association are available. The final section concludes with a brief discussion.

## Methods

### Score and Wald Statistics for One Site

For simplicity we model the effects of each SNP as additive on an appropriate scale (the so-called trend or gene dosage model), though other genotype models can be used. Let  $(y, g, x) = \{(y_j, x_j, g_j), j = 1, \dots, n\}$  represent the data for the  $n$  study subjects. For subject  $j$ ,  $y_j$  is the phenotype,  $x_j$  is a column vector of covariates whose first component may be one to accommodate an intercept, and  $g_j = (g_{j1}, g_{j2}, \dots)$  is a vector of minor allele counts for each of the SNPs under study. We focus on testing association for just one of these SNPs, and omit its subscript. We assume that the loglikelihood  $\ell(y|x, g; \theta)$  for the phenotypes  $y$  conditional on covariates  $x$  and genotypes  $g$  depends on a parameter  $\theta = (\alpha, \beta)$ . Here  $\alpha$  is a  $s$ -dimensional vector of parameters corresponding to an intercept and/or any covariates used by the site, and  $\beta$  specifies the trend relating phenotype to SNP minor allele count.

We wish to test the null hypothesis  $H_0: \theta = \theta_0 = (\alpha, 0)$ . To do so, we introduce the Fisher information for  $\theta$ , given by

$$i=i(\theta)=\begin{pmatrix} i_{\alpha\alpha} & i_{\alpha\beta} \\ i_{\beta\alpha} & i_{\beta\beta} \end{pmatrix}. \quad (1)$$

Here  $i_{\alpha\alpha}$  is the  $s \times s$  matrix whose entries are the negative expectations of the second derivatives of the loglikelihood function for  $\theta$  with respect to  $\alpha$ , and the other submatrices are defined analogously. Also let  $U=(U_{\alpha}^T, U_{\beta}^T)^T$  denote the  $(s+1)$ -dimensional score vector whose entries are the derivatives of the loglikelihood function with respect to  $\theta = (\alpha, \beta)$ . The covariate-adjusted efficient score for testing  $\beta = 0$  is  $U_{\beta} = U_{\beta}(\hat{\alpha}(0), 0)$ , where  $\hat{\alpha}(0)$  satisfies  $U_{\alpha}(\hat{\alpha}(0), 0) = 0$ . Under  $H_0$  and mild regularity conditions [Cox and Hinkley, 1979],  $U_{\beta}$  has an asymptotic Gaussian distribution with mean zero and variance

$$i_{\beta} = i_{\beta}(\theta) = i_{\beta\beta} - i_{\beta\alpha} i_{\alpha\alpha}^{-1} i_{\alpha\beta}, \quad (2)$$

evaluated at  $(\hat{\alpha}(0), 0)$ .

The score and Wald statistics are

$$Z_{score} = \frac{U_{\beta}}{\sqrt{\hat{i}_{\beta}}} \text{ and } Z_{Wald} = \frac{\hat{\beta}}{SD(\hat{\beta})} = \hat{\beta} \sqrt{\hat{i}_{\beta}}, \quad (3)$$

where  $\hat{i}_{\beta}$  is a consistent estimate of  $i_{\beta}$  evaluated at  $\beta = 0$ ,  $\hat{\beta}$  is the maximum likelihood estimate (MLE) of  $\beta$ , and  $SD(\hat{\beta}) = \hat{i}_{\beta}^{-1/2}$  evaluated at  $\hat{\beta}$ . The two statistics are locally asymptotically equivalent, i.e., they are asymptotically equivalent when  $\beta = 0$  and are approximately so when  $\beta$  is small in absolute value [Cox and Hinkley, 1979]. Their noncentrality parameters (NCPs) are  $\xi = \beta \sqrt{\hat{i}_{\beta}}$ , with  $i_{\beta}$  evaluated at  $\theta_0 = (\alpha, 0)$  for  $Z_{score}$  and at  $\theta = (\alpha, \beta)$  for  $Z_{Wald}$ . In the following we shall approximate  $i_{\beta}$  for the Wald statistic by evaluating it at its null value, and refer to a common NCP  $\xi$  for both statistics.

We now specialize the score and Wald statistics to specific phenotypes of relevance to GWAS. We first consider phenotypes whose relation to genotypes and covariates can be described by generalized linear models (GLMs) [McCullagh and Nelder, 1989]. Special cases of particular relevance are the binary phenotypes of case-control studies, quantitative traits such as height or weight, and counts of events such as attempts to quit smoking, and censored survival data.

### Example: generalized linear models

GLMs assume that the mean but not the dispersion parameter of the phenotype distribution depends on genotypes and covariates, so that their effects can be modeled as a function of a linear predictor  $\eta_j = \alpha x_j + \beta g_j$ . The first component of the covariate vector  $x_j$  may equal one, in order to accommodate an intercept. The loglikelihood is

$$\ell(y|x, g; \theta) = \sum_{j=1}^n \left[ \frac{y \eta_j - b(\eta_j)}{a(\phi)} + c(y; \phi) \right], \quad (4)$$

where  $a(\cdot)$ ,  $b(\cdot)$  and  $c(\cdot)$  are known functions, and  $\phi$  is the dispersion parameter of the model. The score for  $\beta$  corresponding to (4) is

$$U_{\beta} = \sum_{j=1}^n \frac{y - \hat{\mu}_j}{a(\hat{\phi})} g_j, \quad (5)$$

where  $\hat{\mu}_j$  is the fitted mean of  $y$  under  $H_0$ , and  $\hat{\phi}$  is a consistent estimate of  $\phi$ . The null Fisher information (1), with  $\theta = \theta_0$ , and  $\eta_j = \alpha x_j$  is

$$i = \sum_{j=1}^n \frac{b''(\eta_j)}{a(\phi)} E \begin{pmatrix} x_j x_j^T & g_j x_j \\ g_j x_j^T & g_j^2 \end{pmatrix}. \quad (6)$$

In the absence of covariates  $x_j = 1$ ,  $\eta_j = \alpha_1$  where  $\alpha_1$  is the intercept, and (6) reduces to

$$i = \frac{b''(\alpha_1)n}{a(\phi)} \begin{pmatrix} 1 & 2p \\ 2p & 2p(1+p) \end{pmatrix}, \quad (7)$$

where  $p = 1 - q$  is the SNP minor allele frequency (MAF). Using (7) in (2), we find that the null Fisher information for  $\beta$  is

$$i_{\beta} = \frac{2nb''(\alpha_1)pq}{a(\phi)}. \quad (8)$$

For binary phenotypes  $y_j$  is coded as one if subject  $j$  has the trait and zero otherwise, and the logistic model corresponds to  $\phi = a(\phi) = c(y; \phi) = 1$ , and  $b(\eta_j) = \ln[1 + e^{\eta_j}]$ . In a case-control study with  $n_1$  cases and  $n_0 = n - n_1$  controls,  $b''(\alpha_1)$  can be estimated consistently by  $n_0 n_1 / n^2$ .

Substituting these expressions in (8) gives

$$\widehat{i}_{\beta} = \frac{2n_0 n_1}{n} pq. \quad (9)$$

For quantitative traits whose distribution is (possibly after transformation) Gaussian with

mean  $\mu_j = \eta_j$  and variance  $\phi$ , we have  $a(\phi) = \phi$ ,  $b(\eta_j) = \eta_j^2 / 2$  and  $c(y; \phi) = -\frac{1}{2} [y/\phi + \ln(2\pi\phi)]$ . Thus  $b''(\alpha_1) = 1$ . Substitution of these values for  $a(\phi)$  and  $b''(\alpha_1)$  into (8) gives

$$\widehat{i}_{\beta} = \frac{2n}{\phi} pq, \quad (10)$$

where  $\widehat{\phi}$  is a consistent estimate of the phenotype variance.

For Poisson count data,  $a(\phi) = 1$ ,  $b(\eta_j) = e^{\eta_j}$  and  $c(y; \phi) = 0$ . Thus  $b''(\alpha_1) = e^{\alpha_1} = \mu$ , where  $\mu$  is the null mean of the counts. Substituting these values into (8) gives

$$\widehat{i}_{\beta} = 2n\mu pq. \quad (11)$$

### Example: censored survival data

Here the data are  $(y, x^T, g) = \{(t_j, \varepsilon_j, x_j^T, g_j); j=1, \dots, n\}$  where for subject  $j$ ,  $t_j$  is survival time and  $\varepsilon_j$  is a failure indicator. We assume independent times to failure and censoring, each

with hazard that may vary across studies. We also assume that the failure distribution follows a proportional hazards model of the form  $h(t; x_j, g_j) = h^*(t) e^{\alpha x_j + \beta g_j}$ , where  $h^*(t)$  is the baseline hazard, and that the  $m$  failures occur at distinct times. Inferences are based on the log of a partial likelihood function for  $\beta$ , given by

$\ell(\beta) = \sum_{i=1}^m \left( \alpha x_{i1} + \beta g_{i1} - \ln \sum_{j \in R_i} e^{\alpha x_j + \beta g_j} \right)$ , where  $m$  is the number of failures,  $g_{i1}$  and  $x_{i1}$  are, respectively, the genotype and covariates of the  $i^{\text{th}}$  failure, and  $R_i$  is the set of subjects at risk just before this failure,  $i = 1, \dots, m$ . The corresponding score for  $\beta$  is

$$U_\beta = \sum_{i=1}^m (g_i - \bar{g}_i), \text{ where } \bar{g}_i = \sum_{j \in R_i} u_{ij} g_{ij} \text{ and } u_{ij} = e^{-\hat{\alpha}^T x_{ij}} / \sum_{j' \in R_i} e^{-\hat{\alpha}^T x_{ij'}}. \quad (12)$$

In the absence of covariates, the null Fisher information for  $\beta$ , as given by (2) and (1), is

$$\hat{i}_\beta = \sum_{i=1}^m E \left[ \frac{1}{R_i} \sum_{j \in R_i} g_{ij}^2 - \left( \frac{1}{R_i} \sum_{j \in R_i} g_{ij} \right)^2 \right] = 2mpq. \quad (13)$$

### Accommodating imputed genotypes

Meta-analysis of site-specific statistics requires local imputation at the study sites. This involves calculating posterior probabilities  $P_{jg}$  for each subject's unobserved SNP minor allele count  $g$ , with  $P_{j0} + P_{j1} + P_{j2} = 1$ . Here  $P_{jg}$  is the conditional probability that subject  $j$  has allele count  $g$ , given the observed genotype data  $Obs$ , and the haplotypes in a reference panel such as HapMap2 [International HapMap Consortium et al., 2007]. Louis [1982] provided a likelihood-based extension to the score statistic to accommodate this imputation. Inferences are based on the probability of phenotypes conditional on observed genotype data  $Obs$ , and covariates  $x$ :

$$\Pr(y|Obs, x; \theta) = \sum_{g \in G} \Pr(g|Obs) \Pr(y|g, x; \theta). \quad (14)$$

Here  $g$  denotes a vector of genotypes for the  $n_k$  subjects, and the posterior probabilities  $\Pr(g|Obs) = \Pr(g) / \sum_{g' \in G} \Pr(g')$ , where  $\Pr(g)$  is based on the Hardy-Weinberg MAF, and the sum is taken over genotypes in the set  $G$  consistent with the observed data  $Obs$  and the haplotypes in the reference panel. The posterior probabilities can be computed using available imputation software [Li and Abecasis 2006; Scheet and Stephens, 2006; Browning BL and Browning SR, 2009; Howie et al., 2009]. As shown by Louis [1982], the score for  $\beta$  is

$$U_\beta = E \left[ U_\beta^c | Obs \right], \quad (15)$$

where  $U_\beta^c$  is the complete data score and the expectation is taken with respect to the posterior distribution  $\Pr(g|Obs)$  for the SNP of interest. The null variance of  $U_\beta$  is given by (2) and (1), but with the Fisher information for the observed data reduced by a penalty due to the uncertainty in genotypes:

$$i = E[(U_\beta)^2] = E[J^c | Obs] - \{E[(U_\beta^c)^2 | Obs] - (U_\beta)^2\}, \quad (16)$$

where  $J^c$  is the negative of the second derivative matrix for the complete data. The score statistic is  $Z_{score} = U_\beta / \hat{\beta}$ , where  $\hat{\beta}$  is obtained from (2) and (16).

The score  $U_\beta$  of (5) and (12) for GLMs and survival data depends linearly on the genotypes  $g_j$ . Thus in the presence of unobserved genotypes, the posterior expectation (15) can be obtained simply by replacing the unobserved allele counts with their posterior means, as noted by Schaid et al., [2002] and Marchini et al., [2007]. In addition, the entries of the information matrix (16) are quadratic forms in the genotypes. Thus their posterior expectations can be computed as functions of the first and second moments of the posterior genotype distributions. In practice however, imputation uncertainties are generally ignored for both score and Wald statistics; instead, subjects' unknown minor allele counts are replaced by their posterior means  $P_{j1} + 2P_{j2}$ .

## Meta-Analysis of Data from K Sites

Suppose now that each of  $K$  sites has computed a score or Wald statistic using a regression model that may include an intercept and/or covariates. We wish to combine these site-specific statistics into a single summary statistic to assess the overall evidence for or against association between a SNP and a given phenotype. We assume that subjects participate in only one study (see [Lin and Sullivan, 2009] for methods applicable to studies with overlapping study subjects). We first describe the weights optimal under the assumption that the SNP effect-size  $\beta$  is common across sites (homogeneous effect-size), and then consider the weights optimal when the effect-sizes  $\beta_1, \dots, \beta_K$  vary across sites (heterogeneous effect-size).

### Homogeneous effect-size

Here the null hypothesis to be tested is  $H_0: \theta = \theta_0 = (\alpha_1, \dots, \alpha_K, 0)$ , where  $\alpha_k$  is the  $s_k$ -dimensional nuisance parameter for study  $k$ ,  $k = 1, \dots, K$ . The summary statistics used for this purpose in GWAS meta-analysis are weighted sums  $Z = \sum_{k=1}^K w_k Z_k$ , where  $Z_k$  is either a site-specific score statistic or a site-specific Wald statistic, and where the weights  $w_k$  are nonnegative constants with  $\sum_k w_k^2 = 1$ . Under  $H_0$ , both summary statistics,  $Z_{score}$  and  $Z_{Wald}$ , have standard Gaussian distributions asymptotically [Cox and Hinkley, 1979]. The optimal weights  $w_k$  are defined as those that maximize the summary statistic's NCP, given by  $\xi = \sum_k w_k \zeta_k$ , where  $\zeta_k$  is the NCP for the score and Wald statistics at site  $k$ ,  $k = 1, \dots, K$ . We have seen that the NCP for both site-specific score and site-specific Wald statistics are  $\xi_k = \beta \sqrt{\hat{i}_{k\beta}}$ ; thus  $\xi = \beta \sum_k w_k \sqrt{\hat{i}_{k\beta}}$ . A straightforward application of the Cauchy-Schwartz inequality shows that the weights  $w_k$  that maximize  $\xi$  under local alternatives are proportional to  $\sqrt{\hat{i}_{k\beta}}$ . Thus the optimal summary score and Wald statistics are given by

$$Z_{score} = \sum_k w_k Z_{k:score} \text{ and } Z_{Wald} = \sum_k w_k Z_{k:Wald}, \text{ with } w_k = \sqrt{\frac{\hat{i}_{k\beta}}{\sum_{k'} \hat{i}_{k'\beta}}}. \quad (17)$$

They are asymptotically equivalent for testing  $\theta = \theta_0$  under the null hypothesis and under local alternatives. Replacing the terms  $Z_{k:score}$  with the score statistics of (3) in the first equality of (17) shows that the summary  $Z_{score}$  is just the well-known stratified score statistic, i.e., the one obtained from the sum of the  $K$  loglikelihood functions:

$$Z_{score} = \frac{\sum_k U_{k\beta}}{\sqrt{\sum_k \hat{i}_{k\beta}}}.$$

For case-control data,  $Z_{score}$  is the Mantel-Haenzel test for trend in a  $2 \times K$  table [Breslow and Day, 1980]. Moreover, replacing the terms  $Z_{k:Wald}$  with the Wald statistics of (3) in the second equality of (17) shows that the summary  $Z_{Wald}$  is the commonly deemed inverse-variance-weighted summary statistic proposed by Cochran [1954] and given by

$$Z_{Wald} = \frac{\hat{\beta}}{SD(\hat{\beta})}, \text{ where } \hat{\beta} = \sum_k \frac{\hat{i}_{k\beta}}{\sum_{k'} \hat{i}_{k'\beta}} \hat{\beta}_k \text{ and } SD(\hat{\beta}) = \left( \sum_k \hat{i}_{k\beta} \right)^{-1/2}. \quad (18)$$

The term “inverse-variance-weighted” refers to the fact  $\hat{\beta}$  is a weighted sum of the site-specific regression coefficients  $\hat{\beta}_k$ , weighted in inverse proportion to their estimated variances:

Table 1 gives the optimal weights for case-control data, quantitative trait data, Poisson count data and censored survival data, when the participating sites do not include covariates in the model, and do not adjust for imputation. The weight  $w_k$  for a given site and phenotype is proportional to the square root of the corresponding estimated null information  $\hat{i}_{k\beta}$  for  $\beta$ , as given by the appropriate choice of expressions (9-13). In contrast, the weights based on study size are proportional to  $\sqrt{n_k p_k q_k}$ . Comparison of these size-based weights to those in Table 1 shows that, even though SNP allele frequencies may be similar across sites, the size-based weights are suboptimal for case-control data from studies with varying case-control ratios, for quantitative trait data with varying trait variances, for count data with varying null expected counts, and for censored survival data with varying numbers of failures. Since the variance of a quantitative trait depends on the covariates included in the regression model which may vary across sites, using a common trait variance may in some circumstances be substantially suboptimal.

In the presence of site-specific adjustment for covariates and/or imputation uncertainty, the optimal weights are more complex than those shown in Table 1. According to (17), they are proportional to the square roots of the null SNP information  $\hat{i}_{k\beta}$ , which depend on the covariate distributions, covariate effects, and imputation uncertainty specific to each site. In principal, the values  $\hat{i}_{k\beta}$  could be provided by the sites. Instead, Zaitlen and Eskin [2010] have suggested accommodating imputation uncertainty by incorporating in each site's weight an estimate of the correlation coefficient  $r_k$  between true and imputed SNP allele counts. In the following section we use simulations to evaluate the power of this strategy.

### Heterogeneous effect-size

Site-specific effect-size variation is plausible for several reasons. First, the studies may have used different genotyping platforms with different numbers and locations of SNPs. This results in differences across sites in levels of genotype measurement error, creating different levels of regression coefficient attenuation across sites. Second, when covariates or their



coding differ across sites, different levels of control for confounding can affect SNP effect-sizes. For example, genetic ancestry may be associated with both phenotype and genotypes, and variable control for ancestry may induce differential levels of residual confounding. Third, for survival data, site-specific variation may occur if effect-size changes over time and the studies differ in level and type of censoring. Finally, the SNP of interest may interact with other risk factors whose prevalences vary across sites, which would cause variation in effect-size.

The site-specific Wald statistics have the form  $Z_{k:Wald} = \hat{\beta}_k / SD(\hat{\beta}_k)$ , where  $\hat{\beta}_k$  is the MLE for  $\beta$  and  $SD(\hat{\beta}_k) = \hat{i}_{k\beta}^{-1/2}$ . These quantities lend themselves naturally to evaluating SNP effect-size heterogeneity across sites, using the methods of Cochran [1954] and DerSimonian and Laird [1986]. Here we show that the components of the site-specific score statistics

$Z_{k:score} = U_{k\beta} / \hat{i}_{k\beta}^{1/2}$  also can be used for this purpose. Specifically, when  $\beta_k$  is small in absolute value, the asymptotic mean  $\varphi_k$  of  $U_{k\beta}$  can be approximated as  $\varphi_k \sim \beta_k \hat{i}_{k\beta}$  (see Appendix). This approximation suggests estimating  $\beta_k$  by  $\hat{\beta}_k = U_{k\beta} / \hat{i}_{k\beta}^{-1}$ , with  $SD(\hat{\beta}_k)$  estimated as  $\hat{i}_{k\beta}^{-1/2}$ ,  $k = 1, \dots, K$ . Then Cochran's homogeneity test can be applied with  $\hat{\beta}_k$  equal to either the MLE or the quantity  $U_{k\beta} / \hat{i}_{k\beta}^{-1}$ .

For either the score-based or Wald-based estimates  $\hat{\beta}_1, \dots, \hat{\beta}_K$ , Cochran's test is based on the statistic  $Q = \sum_k \hat{i}_k (\hat{\beta}_k - \hat{\beta})^2$ , where  $\hat{\beta}$  is given by (18). Under the homogeneity hypothesis  $\beta_1 = \dots = \beta_K = \beta$ ,  $Q$  has a chi-squared distribution on  $K - 1$  degrees of freedom (DF). When homogeneity is rejected, DerSimonian and Laird [1986] propose estimating the inter-site variance  $\tau^2$  of the  $\beta_k$  as

$$\hat{\tau}^2 = \max \left[ 0, \frac{Q - K - 1}{V_1 - (V_2/V_1)} \right],$$

where  $V_r = \sum_k \hat{i}_{k\beta}^r$ ,  $r = 1, 2$ . The overall effect-size is estimated as

$$\hat{\beta}_{HT} = \frac{\sum_k (\hat{\tau}^2 + \hat{i}_{k\beta}^{-1})^{-1} \hat{\beta}_k}{\sum_k (\hat{\tau}^2 + \hat{i}_{k\beta}^{-1})^{-1}}, \text{ with } SD(\hat{\beta}_{HT}) = \left[ \sum_k (\hat{\tau}^2 + \hat{i}_{k\beta}^{-1})^{-1} \right]^{-1/2}.$$

Corresponding summary statistics for testing  $\beta_1 = \dots = \beta_K = 0$  are

$$Z_{HT} = \frac{\hat{\beta}_{HT}}{SD(\hat{\beta}_{HT})} = \sum_k w_{k:HT} Z_k, \text{ where } w_{k:HT} = \sqrt{\frac{(\hat{\tau}^2 + \hat{i}_{k\beta}^{-1})^{-1}}{\sum_{k'} (\hat{\tau}^2 + \hat{i}_{k\beta}^{-1})^{-1}}}, \tag{19}$$

and  $Z_k$  denotes either  $Z_{k:score}$  or  $Z_{k:Wald}$ . When Cochran's test does not reject homogeneity and  $\hat{\tau}^2$  is set to zero, the two summary statistics (19) reduce to their usual forms (17) optimal for homogeneous effects.

In conclusion, one can choose between four types of summary statistics, each formed as a weighted sum of site-specific statistics. The four summary statistics differ according to the

site-specific statistics (score vs Wald) and whether the weights are chosen to be optimal under effect-size homogeneity ( $Z_{HM}$ ) or heterogeneity ( $Z_{HT}$ ).

## Simulations

We simulated case-control and survival data to evaluate the performances of the optimal summary score and Wald statistics, and to compare them to that of the size-weighted summary  $Z_{SZ}$ , obtained using weights proportional to the square roots of study sizes. For each replication, we generated data assuming both homogeneous and heterogeneous effect-sizes, with and without imputation error, and then computed the various summary statistics. We flagged those exceeding certain thresholds, and then computed for each statistic the proportion of all replications in which it was flagged.

### Type-1 Error

To evaluate type 1 error, we compared nominal to empirical type-1 error rates, defined as the proportion of replications in which the summary test statistic exceeded the standard Gaussian threshold for the nominal two-sided p-value. We considered two null hypotheses. The first *homogeneity* null hypothesis specifies that the true values of the  $\beta_k$  satisfy  $\beta_1 = \dots = \beta_K = 0$ . The second *heterogeneity* null hypothesis specifies that the  $\beta_k$  are a random sample from a Gaussian distribution with mean zero and variance  $\tau^2 > 0$ . Under the homogeneity null hypothesis, we found good agreement between nominal and empirical p-values for all summary test statistics considered. In contrast, under the heterogeneity null hypothesis all summary statistics showed inflated empirical error rates relative to their nominal values, with the extent of inflation somewhat attenuated when the summary statistics were heterogeneity-weighted compared to homogeneity-weighted (data not shown). We shall return to this finding after a comparison of test power.

### Power

**Case-control data**—We simulated the meta-analysis of  $K = 3$  case-control studies having the sample sizes shown in Table 2. We generated genotypes for a single SNP assuming Hardy-Weinberg allele frequencies, and then used a logistic regression model with study-specific intercepts  $(\alpha_1, \alpha_2, \alpha_3) = (-3.0, -2.5, -2.5)$  and one of three effect-size triples  $(\beta_1, \beta_2, \beta_3)$  to generate case and control phenotypes until we had achieved the targeted numbers of cases and controls. Two triples correspond to effect-size homogeneity, as given by  $\beta_1 = \beta_2 = \beta_3 = \ln(1.3) = 0.26$ , and  $\beta_1 = \beta_2 = \beta_3 = \ln(1.4) = 0.34$ . The third triple  $(\beta_1, \beta_2, \beta_3) = (0, 0.34, 0.26)$  corresponds to effect-size heterogeneity across the three sites. Power was taken as the proportion of the 1000 replications in which the test statistic exceeded the threshold for a p-value of  $10^{-8}$ .

Table 2 shows power for summary score and Wald statistics with weights determined by assumptions of homogeneous ( $Z_{HM}$ ) and heterogeneous ( $Z_{HT}$ ) effect-sizes, and with size-based weights ( $Z_{SZ}$ ). As expected given their asymptotic equivalence, the score-based and Wald-based summary statistics performed similarly for any given choice of weights. For all types of data, the homogeneity-weighted summary statistics were most powerful. When the data were generated under the homogeneous effect-size assumption, both homogeneity- and heterogeneity-weighted statistics  $Z_{HM}$  and  $Z_{HT}$  consistently outperformed the size-weighted statistic  $Z_{SZ}$ . However when the data were generated according to heterogeneous effect-sizes, the heterogeneity-weighted statistic  $Z_{HT}$  performed poorly compared to the other two. This finding, which is consistent with that reported in other applications [Clayton and Hills, 1993], reflects power loss due to sampling error in estimating the inter-site effect-size variance  $\tau^2$ .

**Survival data**—We also generated censored survival data for a total of 2,483 subjects of European ancestry from Heidelberg Germany, Ontario Canada, and Seattle US who have been genotyped for 600,000 SNPs using the Illumina 610Quad chip as part of an ongoing GWAS of early-onset breast cancer. To mimic a meta-analysis based on  $K = 6$  study sites, we split the subjects from each of the three regions into two subsets, which yielded six sites with the sample sizes shown in Table 3. We focused on 1000 SNPs in a 6-MB region of chromosome 2 (positions 1,000,713 to 6,488,209). To generate subjects' survival times, we assumed independent exponentially distributed times to failure and censoring with site-specific baseline and censoring hazards  $\lambda_k, \mu_k, k = 1, \dots, 6$ , independent of nongenetic covariates. These values are shown in Table 3. Subjects who had not failed by time  $t = 1$  were censored at that time. We assumed that the site-specific failure hazards also depend, via a proportional hazards model, on the minor allele counts of SNP rs17360123 at position 5,531,665 (MAF = 0.2) and SNP rs4072375 at position 3,565,156 (MAF = 0.4). Specifically, we assumed that the failure hazard at site  $k$  was  $\lambda_k \exp(\beta_k g_{kj})$ , where  $g_{kj}$  is the SNP minor allele count for the  $j^{\text{th}}$  subject at site  $k, k = 1, \dots, 6$ . For the homogeneity model we took the common  $\beta$  to be 0.25, 0.30 and 0.35, and for the heterogeneity model we took  $\beta = 0$  for sites 1 and 2, and  $\beta = 0.35$  for the remaining sites.

Table 4 shows the power of homogeneity-weighted, heterogeneity-weighted and size-weighted summary score and Wald statistics under these generating models. The same patterns found for the case-control data in Table 2 are evident here: the homogeneity weighted summary statistics are consistently most powerful, and when the data were generated with homogeneous SNP effects, the size-weighted statistic is least powerful. In the presence of heterogeneous SNP effects, however, the heterogeneity-weighted test statistic is least powerful, most likely because of power loss associated with error in estimating the variance parameter  $\tau^2$ .

The poor power of the heterogeneity-weighted test statistics mitigates against their use in GWAS meta-analysis. In general, these statistics are used to guard against false positives when the null hypothesis specifies varying site-specific effect-sizes whose overall mean is zero. However the primary null hypothesis in the GWAS setting specifies zero SNP effect-size at all  $K$  sites. This suggests the strategy of testing all SNPs using a homogeneity-weighted summary statistic, and investigating heterogeneity for those SNPs whose p-values suggest departures from this primary null hypothesis.

Finally, we evaluated how imputed genotypes affect the power of the homogeneity-weighted score-based test by masking and imputing some sites' genotypes in regions containing SNPs rs17360123 and rs4072375. These two SNPs were assumed associated with survival with a common effect-size  $\beta = 0.35$  across the six sites. We masked and imputed either: i) the associated SNPs and their nearest flanking markers (a total of three SNPs imputed); or ii) the associated SNPs and their nearest two flanking markers (a total of five SNPs imputed). SNPs were masked and imputed for: a) one study (study 1); b) three studies (studies 1,3,5); and c) five studies (studies 1-5). We conducted the imputations using the HapMap2 reference panel in an eight megabase region containing 9262 SNPs. The five studies with masked SNPs had estimated correlation coefficients between imputed and actual rs17360123 minor allele counts of  $(\hat{r}_1, \dots, \hat{r}_5) = (0.62, 0.66, 0.57, 0.56, 0.60)$  when three SNPs were imputed, and  $(0.59, 0.58, 0.51, 0.52, 0.49)$  when five SNPs were imputed. The corresponding values for rs4072375 were  $(0.82, 0.77, 0.57, 0.62, 0.68)$  and  $(0.54, 0.58, 0.49, 0.46, 0.45)$ . These estimates were obtained as described by Li and Abecasis [2006]. In total, we considered the 12 data sets described in Table 5. We imputed all masked SNPs using the Caucasian data from the HapMap2 reference panel and the imputation software BEAGLE [Browning BL and Browning SR, 2009] (<http://faculty.washington.edu/browning/beagle/beagle.html>). We evaluated the power of

three statistics adapted from the homogeneity-weighted summary score statistic  $Z_{HM}$ . The first is the likelihood-based score statistic given by equations (15, 16) that accounts for imputation uncertainty. The second statistic replaces subjects' unobserved SNP minor allele counts by their posterior means, ignoring imputation uncertainties. The third statistic also is based on subjects posterior mean minor allele counts, but with weights proportional to

$\hat{r}_k \sqrt{\hat{i}_{k\beta}}$ , a strategy similar to that proposed by Zaitlen and Eskin [2010].

Table 5 shows power for each of these three summary statistics as applied to each of the 12 data sets. For comparison, the power for the optimal tests with completely observed data was 0.827 for SNP rs17360123 and 0.954 for SNP rs4012375 (Table 4). Several results are noteworthy. First, as expected, the power of all tests dropped increasingly with the masking of increasingly many SNPs at increasingly many sites. Moreover the rs4012375 SNP suffers more power loss from masking five rather than three SNPs than does rs17360123. This difference mirrors the correspondingly larger loss of imputation accuracy for rs4012375 than rs17360123, as measured by their estimated correlation coefficients, described in the preceding paragraph. Second, substituting posterior means performed as well as the likelihood-based score statistics. In fact, substituting the posterior means for the unobserved genotypes gives exactly the observed data score  $U_{k\beta}^*$  (15), since the complete data score is a linear function of the unobserved genotypes. Moreover the Fisher information computed by substituting posterior means for unobserved genotypes is an approximation to the Fisher information  $i_k^*(\theta)$  for the observed data, given by (16), and it accounts for genotype uncertainty by its reduced inter-subject variation for poorly imputed SNPs. So it is not surprising that these two statistics show similar power. Third, the statistic formed by

weighting by  $\hat{r}_k \sqrt{\hat{i}_{k\beta}}$  did not perform as well, most likely due to inaccuracies in the estimates  $\hat{r}_k$  and sensitivity of the weights to such error. These results contrast with those of Zaitlen and Eskin [2010] who found slightly increased power using this strategy; further investigation of this matter would be useful.

## Application to GWAS Meta-Analysis

We illustrate the summary score-statistics by application to the meta-analysis of three case-control GWAS studies of schizophrenia in European-Americans described by Shi et al., [2009]. The three studies are the SGENE Consortium [Stefansson et al., 2009], The International Schizophrenia Consortium [ISC; International Schizophrenia Consortium, 2009] and The Molecular Genetics of Schizophrenia study, [MGS; Shi et al., 2009]. These studies enlisted a total of 8,008 cases and 19,077 controls, as shown in Table 6. For the meta-analysis, only site-specific p-values and the directions of association for the logistic trend test were available for each of a common set of SNPs in six regions containing SNPs satisfying a predetermined threshold for statistical significance in at least one of the three studies. For the meta-analysis the authors converted each p-value to a score statistic determined as the quantile corresponding to the appropriate tail of the standard Gaussian distribution. They then evaluated SNPs using the summary score statistic of (17), with weights proportional to the square roots of the estimates  $\hat{i}_{k\beta}$  of (9). GWAS tag SNPs are selected to have common frequencies in different Caucasian populations; thus the authors assumed that the SNP MAFs are approximately the same across the three sites, which gives weights  $w_k$  proportional to  $[(n_{k0}n_{k1}) / (n_{k0} + n_{k1})]^{1/2}$ .

We computed three test statistics for each of the 28 SNPs in the major histocompatibility complex class I and II identified by the authors as significantly associated with schizophrenia risk. These test statistics are: 1) the homogeneity-weighted statistic (17) used by Shi et al., [2009]; 2) the heterogeneity-weighted statistic (19); and 3) the size-weighted

statistic with weights proportional to  $(n_{k0} + n_{k1})^{1/2}$ . For the heterogeneity-weighted statistic, we found that the estimated inter-study variances  $\tau^2$  were zero for all SNPs, and Cochran's Q test did not reject homogeneity at a significance level of 0.1. (An alternate measure of heterogeneity, defined as  $I^2 = 1 - [(K - 1) / Q]$ , estimates the percentage of total inter-site variation beyond chance, with heterogeneity suggested by values  $I^2 > 0.5$  [Ioannidis et al., 2007]. For these data,  $I^2$  was negative for all 28 SNPs, showing no evidence of heterogeneity.) Thus the heterogeneity- and homogeneity-weighted score statistics were equal for all 28 SNPs. Figure 1 shows that this summary score statistic yielded smaller p-values than the size-weighted statistic.

This example illustrates three points. First, site-specific data can be combined optimally simply by sharing p-values, the directions of association, and the null information estimates (as obtained from case and control sample sizes). Second, heterogeneity across studies can be evaluated even with such minimal information from the sites. Finally, as expected from the theory and simulations, the optimally-weighted summary score-statistic is more powerful than the size-weighted score-statistic. For case-control data such as these, optimal weighting achieves the greatest power gain relative to size-based weighting when case and control numbers are roughly the same for some studies, but substantially different for others (as in the present example). Power gains also can be substantial for meta-analysis of censored survival data when censoring levels differ among the studies.

## Discussion

We have reviewed SNP association tests based on weighted sums of site-specific score and Wald statistics. We have shown that when SNP effect-size  $\beta$  is constant across sites, the optimal weights for both statistics are approximately proportional to the square roots of the sites' null Fisher information for the SNP of interest, and we have described these weights for the phenotypes commonly studied with GWAS. These optimal weights have simple forms in the absence of site-specific differences in covariates and their effects, and in the absence of accommodation for site-specific differences in imputation uncertainty. Their optimality in these circumstances is supported by simulation results. In the presence of covariate adjustment, the simple weights described here should be nearly optimal unless the sites differ strongly in choice of covariates or their effect-sizes. The simulations also suggest little power loss in using weights that ignore imputation uncertainties. This finding is consistent with the similarities in the two statistics noted in the Methods section.

We also have shown how to use the site-specific score statistics to obtain SNP effect-size estimates for assessing inter-site effect-size heterogeneity. An application to schizophrenia data shows that for case-control GWAS meta-analysis, such heterogeneity can be assessed even when only site-specific SNP p-values and directions of association are available. Both simulations and the data application support the conclusion of others that Cochran's test has poor power to detect effect-size heterogeneity, particularly when the number of sites is small [Clayton and Hills, 1993]. Thus we suggest performing GWAS meta-analysis using the homogeneity-weighted summary statistics and examining heterogeneity for the hit SNPs, with the caveat that lack of evidence for heterogeneity can be interpreted only as ruling out very large differences across sites.

A limitation of all meta-analysis is its need for site-specific estimates of regression coefficients for covariates that are common to all or several sites (e.g. age, sex, ancestry). The simulations of Lin and Zeng [2010a; 2010b] suggest that site-specific regression coefficient estimates for such covariates do not produce substantial bias or power loss. Although this conclusion seems plausible unless the covariates are strongly associated with both genotype and phenotype, more evidence would be useful.

## Acknowledgments

This research was supported by NIH grants R01 CA094069 and U01 CA122171. The authors are grateful to Habibur Ahsan, Irene Andrusis, Jenny Claude-Chang, Mohammed Kibriya, and Kathi Malone for the use of these genotype data, and to the reviewers of an earlier version of the manuscript for their helpful suggestions.

## Appendix: Approximating the Asymptotic Mean of the Score $u\beta$

Let  $U(\theta_0) = (U_\alpha(\theta_0), U_\beta(\theta_0))^T$  denote the vector of efficient scores for  $\theta$ , evaluated at  $\theta_0 = (\alpha, 0)$ . Thus  $E[U(\theta_0); \theta_0] = 0$ . To approximate this expectation over local alternatives for  $\theta$ , we evaluate it over a distribution governed by a value near to but not equal to  $\theta_0$ . Thus we write  $\theta = \theta_0 + \delta$ , where  $\delta = (0, \beta)^T$ , and  $\beta$  is small in absolute value. Using a multivariate version of the argument given by Cox and Hinkley [1979, p 109]; we can write

$$E[U(\theta_0); \theta_0 + \delta] = i(\theta_0)\delta + o(\|\delta\|) = i(\theta_0)\delta + o(|\beta|) = \beta \begin{pmatrix} i_{\alpha\beta} \\ i_{\beta\beta} \end{pmatrix} + o(|\beta|),$$

where  $i(\theta_0)$  is the Fisher information for  $\theta$  evaluated at  $\theta_0$ . Similarly, the covariance matrix of  $U(\theta_0)$  is

$$\text{var}\{U(\theta_0); \theta_0 + \delta\} = i(\theta_0) + O(|\beta|).$$

These results, when used with standard normal theory, show that, conditional on  $\theta_0 = (\hat{\alpha}(0), 0)$ , the asymptotic distribution of  $U_\beta(\hat{\alpha}(0), 0) = U_\beta$  is normal with mean  $\beta i_\beta + o(|\beta|)$  and variance  $i_\beta + O(|\beta|)$ . Here  $i_\beta$  is given by (2) with  $i(\theta) = i(\theta_0)$ .

## References

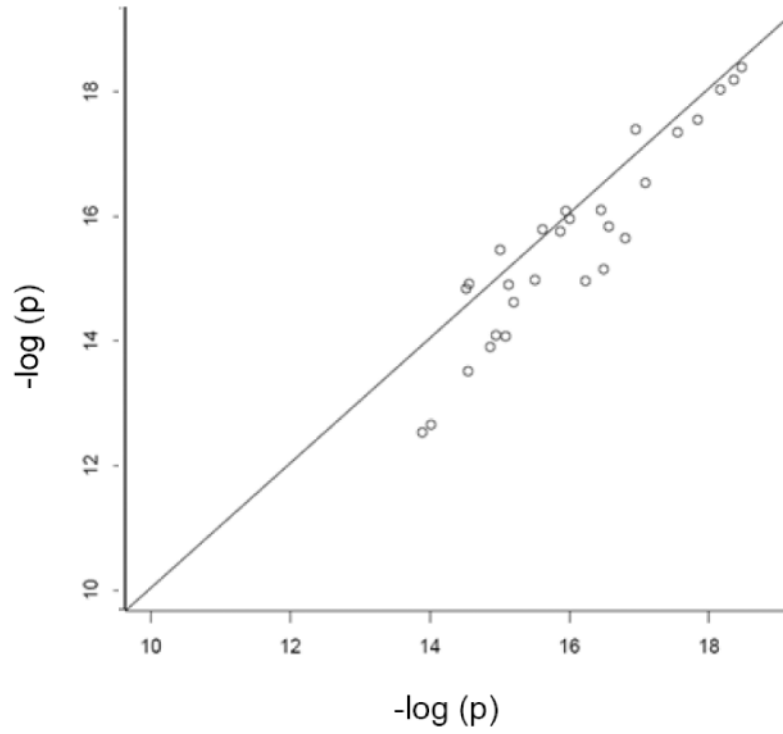
- Breslow, NE.; Day, NE. Statistical methods in cancer research, vol 1, the Analysis of Case-control Studies. Lyon: International Agency for Research on Cancer; 1980.
- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 2009; 84:210–23. [PubMed: 19200528]
- Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet.* 2010; 86:6–22. [PubMed: 20074509]
- Clayton, D.; Hills, M. Statistical models in epidemiology. Oxford: Oxford University Press; 1993.
- Cochran BG. The combination of estimates from different experiments. *Biometrics.* 1954; 10:101–29.
- Cox, DR.; Hinkley, CV. Theoretical statistics. London: Chapman and Hall; 1979.
- de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet.* 2008; 17:R122–8. [PubMed: 18852200]
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986; 7:177–88. [PubMed: 3802833]
- Guan Y, Stephens M. Practical issues in imputation-based association mapping. *PLoS Genet.* 2008; 4:e1000279. [PubMed: 19057666]
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009; 5:e1000529. [PubMed: 19543373]
- Hu X, Pickering E, Liu YC, Hall S, Fournier H, Katz E, Dechairo B, John S, Van Eerdewegh P, Soares H. Alzheimer's Disease Neuroimaging Initiative. Meta-analysis for genome-wide association study

identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. *PLoS ONE*. 2011; 6:e16616. [PubMed: 21390209]

- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; 449:851–61. [PubMed: 17943122]
- Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–52. [PubMed: 19571811]
- Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered*. 2007; 64:203–13. [PubMed: 17551261]
- Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One*. 2007; 2:e841. [PubMed: 17786212]
- Li Y, Abecasis GR. Rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet*. 2006; S79:2290.
- Lin DY, Sullivan PF. Meta-analysis of genome-wide association studies with overlapping subjects. *Am J Hum Genet*. 2009; 85:862–72. [PubMed: 20004761]
- Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol*. 2010a; 34:60–6. [PubMed: 19847795]
- Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika*. 2010b; 97:321–32.
- Louis TA. Finding the observed information matrix when using the EM algorithm. *Jr Stat Soc*. 1982; B44:226–33.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007; 39:906–13. [PubMed: 17572673]
- McCullagh, PJ.; Nelder, JA. Generalized linear models. 2nd. London: Chapman and Hall; 1989.

- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet.* 2002; 70:425–34. [PubMed: 11791212]
- Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 2006; 78:629–44. [PubMed: 16532393]
- Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science.* 2007; 316:1341–5. [PubMed: 17463248]
- Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whittemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman PV. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature.* 2009; 460:753–7. [PubMed: 19571809]
- Soranzo N, Rivadeneira F, Chinappan-Horsley U, Malkina I, Richards JB, Hammond N, Stolk L, Nica A, Inouye M, Hofman A, Stephens J, Wheeler E, Arp P, Gwilliam R, Jhamai PM, Potter S, Chaney A, Ghorri MJ, Ravindrarajah R, Ermakov S, Estrada K, Pols HA, Williams FM, McArdle WL, van Meurs JB, Loos RJ, Dermitzakis ET, Ahmadi KR, Hart DJ, Ouweland WH, Wareham NJ, Barroso I, Sandhu MS, Strachan DP, Livshits G, Spector TD, Uitterlinden AG, Deloukas P. Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet.* 2009; 5:e1000445. [PubMed: 19343178]
- Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OP, Mors O, Mortensen PB, Sigurdsson E, Gustafsson O, Nyegaard M, Tuulio-Henriksson A, Ingason A, Hansen T, Suvisaari J, Lonnqvist J, Paunio T, Børghlum AD, Hartmann A, Fink-Jensen A, Nordentoft M, Hougaard D, Norgaard-Pedersen B, Böttcher Y, Olesen J, Breuer R, Möller HJ, Giegling I, Rasmussen HB, Timm S, Mattheisen M, Bitter I, Réthelyi JM, Magnusdottir BB, Sigmundsson T, Olason P, Masson G, Gulcher JR, Haraldsson M, Fossdal R, Thorgeirsson TE, Thorsteinsdottir U, Ruggeri M, Tosato S, Franke B, Strengman E, Kiemenev LA, Genetic Risk and Outcome in Psychosis (GROUP), Melle I, Djurovic S, Abramova L, Kaleda V, Sanjuan J, de Frutos R, Bramon E, Vassos E, Fraser G, Ettinger U, Picchioni M, Walker N, Touloupoulou T, Need AC, Ge D, Yoon JL, Shianna KV, Freimer NB, Cantor RM, Murray R, Kong A, Golimbet V, Carracedo A, Arango C, Costas J, Jönsson EG, Terenius L, Agartz I, Petursson H, Nöthen MM, Rietschel M, Matthews PM, Muglia P, Peltonen L, St Clair D, Goldstein DB, Stefansson K, Collier DA. Common variants conferring risk of schizophrenia. *Nature.* 2009; 460:744–7. [PubMed: 19571808]
- Tanaka T, Scheet P, Giusti B, Bandinelli S, Piras MG, Usala G, Lai S, Mulas A, Corsi AM, Vestriani A, Sofi F, Gori AM, Abbate R, Guralnik J, Singleton A, Abecasis GR, Schlessinger D, Uda M, Ferrucci L. Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. *Am J Hum Genet.* 2009; 84:477–82. [PubMed: 19303062]
- Willer C, Li Y, Abecasis G. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010; 26:2190–1. [PubMed: 20616382]
- Zaitlen N, Eskin E. Imputation aware meta-analysis of genome-wide association studies. *Genet Epidemiol.* 2010; 34:537–42.
- Zeggini E, Ioannidis JP. Meta-analysis in genome-wide association studies. *Pharmacogenomics.* 2009; 10:191–201. [PubMed: 19207020]
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Morris AD, Morris AD, Doney AS, Wellcome Trust Case Control Consortium (WTCCC), McCarthy MI, Hattersley AT. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science.* 2007; 316:1336–41. [PubMed: 17463249]





**Fig. 1.** p-values of size-weighted statistics (y-axis) versus summary score statistics with homogeneity weights(x-axis) for 28 top schizophrenia SNPs

**Table 1**  
**Optimal Weights in GWAS Meta-analysis without Covariates**

Phenotype	Optimal Weight $w_k$ for $k^{\text{th}}$ Study <sup>a</sup>
Case-control Status <sup>b</sup>	$\left(\frac{n_{k0}n_{k1}}{n_k} p_k q_k\right)^{1/2}$
Quantitative Trait <sup>c</sup>	$\left(\frac{n_k}{\phi_k} p_k q_k\right)^{1/2}$
Poisson Count Data <sup>d</sup>	$(n_k \mu_k p_k q_k)^{1/2}$
Censored Survival Data <sup>e</sup>	$(m_k p_k q_k)^{1/2}$

<sup>a</sup>)  $w_k$  are optimal under effect size homogeneity across sites, and are given up to a proportionality factor determined by the constraint  $\sum_k w_k^2 = 1$ .  $p_k = 1 - q_k$  is the SNP MAF and  $n_k$  is the number of subjects in study  $k$ .

<sup>b</sup>)  $n_{kj}$  is the number of cases ( $j=1$ ) and controls ( $j=0$ ).

<sup>c</sup>)  $\phi_k$  is the trait variance.

<sup>d</sup>)  $\mu_k$  is the null expected count

<sup>e</sup>)  $m_k$  is the number of failures.

**Table 2**  
**Test Power<sup>a</sup> in Meta-analysis of Three Case-control Studies ( $\alpha = 10^{-8}$ )**

Study-site	1			2			3			Summary Test Statistic			
	500	1000	500	1000	5000	5000	500	1000	5000	Score-based		Wald-based	
Controls	500	1000	500	1000	5000	5000	$Z_{HM}^b$	$Z_{HT}^c$	$Z_{SZ}^d$	$Z_{SZ}^d$	$Z_{HM}$	$Z_{HT}$	$Z_{SZ}$
Odds-ratio MAF	1.3	1.3	1.3	1.3	1.3	1.3							
	0.2	0.2	0.2	0.2	0.2	0.2	0.356	0.349	0.301	0.348	0.343	0.291	
	0.3	0.3	0.3	0.3	0.3	0.3	0.638	0.632	0.563	0.634	0.629	0.555	
	0.3	0.3	0.3	0.2	0.2	0.2	0.543	0.537	0.426	0.530	0.527	0.418	
Odds-ratio MAF	1.4	1.4	1.4	1.4	1.4	1.4							
	0.2	0.2	0.2	0.2	0.2	0.2	0.887	0.886	0.840	0.886	0.885	0.836	
	0.3	0.3	0.3	0.3	0.3	0.3	0.974	0.974	0.951	0.974	0.974	0.949	
	0.3	0.3	0.3	0.2	0.2	0.2	0.953	0.950	0.908	0.951	0.950	0.904	
Odds-ratio MAF	1.0	1.4	1.3										
	0.2	0.2	0.2	0.2	0.2	0.2	0.188	0.086	0.170	0.185	0.086	0.164	
	0.3	0.3	0.3	0.3	0.3	0.3	0.457	0.230	0.420	0.452	0.221	0.409	
	0.3	0.3	0.3	0.2	0.2	0.2	0.323	0.168	0.260	0.309	0.166	0.254	

<sup>a</sup>) fraction of 1000 replications rejecting  $H_0$  at  $\alpha = 10^{-8}$

<sup>b</sup>) homogeneity weights of equation (17)

<sup>c</sup>) heterogeneity weights of equation (19)

<sup>d</sup>) weights proportional to square root of study size

**Table 3**  
**Sample Sizes and Hazard Rates for Data from Six Survival Studies**

Study	1	2	3	4	5	6
Sample size	468	470	328	395	296	526
$\lambda_k$	0.105	0.127	0.223	0.274	0.510	0.693
$\mu_k$	2.000	1.500	1.000	0.090	0.200	0.075
Expected no failures	20.5	23.9	63.0	90.7	108.0	254.4

$\lambda_k$  and  $\mu_k$  are failure and censoring hazards, respectively, for study  $k$ ,  $k=1, \dots, 6$

Table 4

Test Power<sup>a</sup> in Meta-analysis of Six Survival Studies

Associated SNP	MAF	Log hazard ratio $\beta$	Summary Test Statistic					
			Score-based			Wald-based		
			$Z_{HM}^b$	$Z_{HT}^c$	$Z_{SZ}^d$	$Z_{HM}$	$Z_{HT}$	$Z_{SZ}$
rs17360123	0.2	0.25	0.264	0.202	0.172	0.265	0.208	0.162
		0.30	0.530	0.449	0.416	0.550	0.487	0.403
		0.35	0.827	0.709	0.716	0.841	0.792	0.709
		Heterogeneous $\beta^e$	0.294	0.117	0.201	0.302	0.129	0.194
rs4012375	0.4	0.25	0.463	0.422	0.333	0.465	0.429	0.318
		0.30	0.787	0.719	0.673	0.787	0.738	0.665
		0.35	0.954	0.884	0.910	0.956	0.904	0.907
		Heterogeneous $\beta^e$	0.566	0.204	0.416	0.579	0.232	0.401

<sup>a</sup>) fraction of 1000 replications in rejecting  $H_0$  at  $\alpha = 10^{-5}$

<sup>b</sup>) homogeneity weights of equation (17)

<sup>c</sup>) heterogeneity weights of equation (19)

<sup>d</sup>) weights proportional to square root of study size

<sup>e</sup>)  $\beta = 0$  for sites 1,2, and  $\beta = 0.35$  for studies 3-6

**Table 5**  
**Test Power<sup>a</sup> in Score-based Meta-analysis of Six Survival Studies with SNP Imputation**

Associated SNP	SNPs Imputed <sup>b</sup>	Sites with Imputed SNPs	Imputation Method	Likelihood-based <sup>b</sup>	Posterior Mean <sup>c</sup>	r <sub>based</sub> <sup>d</sup>
rs17360123 MAF = 0.2	Three <sup>e</sup>	Site 1		0.802	0.803	0.798
		Sites 1,3,5		0.777	0.768	0.722
	Five <sup>e</sup>	Sites 1-5		0.737	0.726	0.649
		Site 1		0.803	0.802	0.798
		Sites 1,3,5		0.767	0.757	0.692
rs4072375 MAF = 0.4	Three	Sites 1-5		0.715	0.697	0.591
		Site 1		0.947	0.945	0.942
	Five	Sites 1,3,5		0.918	0.918	0.896
		Sites 1-5		0.854	0.843	0.789
		Site 1		0.934	0.933	0.933
		Sites 1,3,5		0.836	0.839	0.818
		Sites 1-5		0.654	0.647	0.601

<sup>a</sup>) fraction of 1000 replications rejecting  $H_0$  at  $\alpha = 10^{-5}$ , with data generated under a common effect size  $\beta = 0.35$

<sup>b</sup>) obtained from equations (15,16)

<sup>c</sup>) masked genotypes replaced by mean of posterior distributions

<sup>d</sup>) using posterior means with weights proportional to  $r_{ik}^2/2$

<sup>e</sup>) masking 3 SNPs (causal SNP plus one flanking SNP each side) or 5 SNPs (causal SNP plus 3 flanking SNPs each side)

**Table 6**  
**Sample Sizes in Meta-Analysis of K=3 Schizophrenia Case-control GWAS<sup>a</sup>**

Study <sup>b</sup>	MGS		ISC		SGENE		Total	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Number of Subjects	2681	2653	3322	3587	2005	12,837	8008	19,077

<sup>a</sup>) Shi et al., [2009]

<sup>b</sup>) MGS: Molecular Genetics of Schizophrenia; ISC : International Schizophrenia Consortium; SGENE : SGENE Consortium