

Estimating Odds Ratios in Genome Scans: An Approximate Conditional Likelihood Approach

Arpita Ghosh,¹ Fei Zou,^{1,2,3} and Fred A. Wright^{1,2,3,*}

In modern whole-genome scans, the use of stringent thresholds to control the genome-wide testing error distorts the estimation process, producing estimated effect sizes that may be on average far greater in magnitude than the true effect sizes. We introduce a method, based on the estimate of genetic effect and its standard error as reported by standard statistical software, to correct for this bias in case-control association studies. Our approach is widely applicable, is far easier to implement than competing approaches, and may often be applied to published studies without access to the original data. We evaluate the performance of our approach via extensive simulations for a range of genetic models, minor allele frequencies, and genetic effect sizes. Compared to the naive estimation procedure, our approach reduces the bias and the mean squared error, especially for modest effect sizes. We also develop a principled method to construct confidence intervals for the genetic effect that acknowledges the conditioning on statistical significance. Our approach is described in the specific context of odds ratios and logistic modeling but is more widely applicable. Application to recently published data sets demonstrates the relevance of our approach to modern genome scans.

Introduction

In genetic studies, it is widely recognized that the control of genome-wide error requires the use of stringent thresholds for significance testing. For genome-wide linkage scans, standard LOD significance thresholds in the range 3.0 to 4.0 correspond to point-wise p values in the range 10^{-4} to 10^{-5} , depending on the model and study design.¹ For modern genome-wide association scans (GWASs), 100,000 to 1 million SNP markers may be genotyped, and control of family-wise error or false discovery rates typically requires point-wise significance thresholds in the range 10^{-7} to 10^{-8} .²⁻⁴ The use of such stringent thresholds is offset somewhat by the belief that GWAS offer greater power than linkage studies for detecting complex disease genes.⁵ Nonetheless, the application of stringent thresholds distorts the inferential process, producing estimates of disease risk effect sizes that may be, on average, far greater in magnitude than the true effect.^{1,2,6-16} This phenomenon has been described as a form of “winner’s curse” by Zöllner and Pritchard¹⁶ and others, or as a form of regression to the mean,¹⁵ and has profound importance for genome scans. Although the problem has been described as primarily an issue of bias, we demonstrate below that the variance of risk estimates can also be greatly inflated by the selection procedure. Moreover, standard confidence intervals for risk estimates will have very poor coverage properties, although this issue seems to have received less attention.

Consider a genome association scan for a complex disease in which ten genomic regions contain disease genes, and each region has a 20% chance of meeting genome-wide significance. Assuming independence of regions, the genome scan has respectable power $1 - (1 - 0.2)^{10} = 0.89$

to achieve significance in at least one region. However, a repeated genome scan of equal size will have power of only 0.2 for any one region and thus probably not result in “replication” of the first study. A follow-up study might focus on a single significant region, with fewer markers and paying a lower penalty for multiple comparisons. But if the results of the initial genome scan are used as a guide, the follow-up study is likely to be underpowered, by relying on an inflated estimate of locus disease risk.

As a statistical phenomenon, the winner’s curse should not be confused with additional sources of bias, including variations due to genotyping technologies, or heterogeneity of patient populations from which samples are drawn.^{12,17,18} The winner’s curse is investigated in detailed simulations elsewhere,^{8,9,13-16} including a recent paper by Garner,⁸ who clarified that the bias can be understood predominantly through the behavior of Wald statistics for log-odds ratios.

Although the bias is simple to understand and to document, reducing or eliminating it may be nontrivial. Zöllner and Pritchard¹⁶ have described a likelihood approach that requires maximization over numerous parameters, including genotype frequencies and penetrance parameters, along with conditioning on declared statistical significance. Their procedure reduces the bias in risk estimation but cannot be performed with standard statistical software. Yu et al.¹⁵ have recently applied bootstrapping to correct for significance bias. Both of these bias correction approaches are technically feasible for genome scans, but they would be highly computationally intensive in that setting.

We describe our alternative approach for estimating genetic effects in terms of odds ratios, which have numerous advantages that have made them standard for analysis of

¹Department of Biostatistics, ²Carolina Center for Genome Sciences, ³Center for Environmental Bioinformatics, The University of North Carolina at Chapel Hill, NC 27599, USA

*Correspondence: fwright@bios.unc.edu

DOI 10.1016/j.ajhg.2008.03.002. ©2008 by The American Society of Human Genetics. All rights reserved.

case-control designs.¹⁹ A crucial advantage for case-control studies is that the odds ratio (OR) may be estimated consistently, whether the study design is prospective or retrospective,²⁰ and the OR has an interpretation distinct from nuisance parameters such as genotype frequencies. Moreover, in logistic models, the OR retains interpretability in the presence of covariates, and such a retention is increasingly important for complex disease investigations.

In this paper, we introduce a method to correct for significance bias in disease association studies, by using an approximate conditional likelihood. The approach is directly based on the $\log(\text{OR})$ estimate and its standard error as reported by standard statistical software and applies to dominant, recessive, or additive genetic models. No modification is necessary when covariates such as population-stratification variables have also been fit in the model. The approach may even be applied to published results without access to the original data. In addition, we develop a method to construct accurate confidence intervals for the OR.

We illustrate the performance of our approach via extensive simulations of a disease SNP analyzed by logistic regression. The simulations cover a range of models, disease allele frequencies, and OR values. Compared to naive OR estimation, our approach provides greatly reduced bias and mean squared error, particularly for the modest effect sizes likely to be encountered in complex diseases. In addition, our confidence-interval procedure provides coverage that is accurate or slightly conservative. Performing simulations for OR values near the null presents a challenge because significant results are very rare when applying genome-wide thresholds. We thus employ a screening approach in which a deterministic trend statistic is used to identify data sets potentially significant in logistic regression.

Material and Methods

We assume a genetic model with one parameter for the effect of disease genotype, which includes recessive, dominant, and additive models. We use $\beta = \log(\text{OR})$ to denote the true \log_e odds ratio for disease risk conferred by a referent genotype, or for the contribution of each allele in an additive model. A single locus test statistic for disease association can be expressed as an estimate for β divided by an estimate for its standard error,

$$Z = \frac{\hat{\beta}}{SE(\hat{\beta})}, \quad (1)$$

which is compared to the asymptotic null distribution $N(0,1)$. We will refer to $\hat{\beta}$ and $SE(\hat{\beta})$ as naive estimators because they are obtained from standard statistical procedures without acknowledging selection based on significance. For our problem, we wish to estimate β only when the SNP is significant in two-sided testing, i.e., $|z| > c$ for a value c corresponding to genome-wide significance. By explicitly considering this selection, below we obtain three new estimators and a confidence interval procedure. Our approach offers marked improvements over $\hat{\beta}$ and standard confidence intervals. Our exposition includes mathematical and moti-

national details that we believe will considerably demystify the problem, which has until now appeared more obscure and complex than necessary. The performance of our new estimators is described in subsection *Simulations*.

Significance Bias

When logistic regression is used to test for genetic association, the Wald statistic for genetic effect assumes the specific form of (1), with numerator and denominator obtained from maximum likelihood and the information matrix.^{20–22} However, the essence of our approach applies to a wide variety of testing procedures, for which the key requirements typically hold: (1) asymptotically normality of $\hat{\beta}$ and (2) consistency of the standard error estimate, so that $SE(\hat{\beta})/SE(\beta) \rightarrow 1$. Expressing the test statistic in the form of Equation (1) provides a straightforward illustration of significance bias and points the way toward corrected estimation procedures. Related test statistics that are based on maximum likelihood ratios or efficient scores, or that are directly based on contingency tables, are all asymptotically equivalent to (1) for local departures from the null hypothesis $H_0 : \beta = 0$,²³ although this asymptotic equivalence is not necessary to apply our approach. The remainder of this subsection is similar to Garner,⁸ but our explicit and expanded treatment provides the grounds for later development.

For large samples, $SE(\hat{\beta})$ does not vary markedly in repeated data realizations. Thus, the estimate $\hat{\beta}$ and its statistical significance are highly correlated,⁸ and the problem can be restated as single-parameter estimation for a truncated normal distribution. To see this, we define $\mu = \beta/SE(\hat{\beta})$, with $Z \sim N(\mu, 1)$. Our use of this approximation follows from the standard result $Z - \mu = (\hat{\beta} - \beta)/SE(\hat{\beta}) \xrightarrow{D} N(0, 1)$ for increasing sample size.²⁴ The statistical procedures to follow are developed entirely within this “ μ version” of the problem, which has been greatly simplified by the variance standardization.

Our naive estimate of μ is $\hat{\mu} = z$, and the expectation can be shown analytically to be

$$E_{\mu}(Z \mid |Z| > c) = \mu + \frac{\phi(c - \mu) - \phi(c + \mu)}{\Phi(-c + \mu) + \Phi(-c - \mu)}, \quad (2)$$

where ϕ and Φ are the density and cumulative distribution function of a standard normal, respectively (see Appendix A). This is the two-sided rejection version of a result given by Garner.⁸ As we detail in the Results, the bias can be substantial in realistic settings. In the special case of the null hypothesis $\mu = 0$, it is clear from Equation (2) that the naive estimate z is unbiased because the two-sided testing procedure is equally likely to falsely declare positive or negative risk (i.e., a protective effect of the referent genotype). It is not clear that the lack of bias for naive estimation under the null has been fully appreciated (e.g., Figure 2 in Zöllner and Pritchard¹⁶ does not display the exact null value). However, this lack of bias requires averaging over rejections for both positive and negative z . In any significant data set, $\hat{\mu}$ must be less than $-c$ or greater than c and so will be far from the truth under the null. In other words, the lack of bias under the null is offset by very large variance.

An Approximate Conditional Likelihood

The approximating distribution of Z suggests a correspondingly approximate likelihood for μ ,

$$L(\mu) = p_{\mu}(z) = \phi(z - \mu). \quad (3)$$

The likelihood applies generally to a wide variety of testing procedures, eliminating any nuisance parameters that have been

included in the modeling, including stratification variables, clinical covariates, or the effects of other SNP genotypes. It is easy to show that the maximum-likelihood estimate (MLE) is $\hat{\mu} = z$. A standard approach to likelihood testing for $H_0: \mu = 0$ (Ref. ²⁴) involves comparing the maximum log-likelihood ratio $LLR = 2 \log(L(\hat{\mu})/L(0))$ to a χ^2_1 density. It is also simple to show that here $LLR = z^2$, so in terms of both estimation and testing, the likelihood simply recapitulates the initial Equation (1). The advantage to Equation (3), however, is that it provides a simple and transparent approach to handle the conditioning. Acknowledging the event that the SNP is declared statistically significant, we have the conditional likelihood

$$L_c(\mu) = p_\mu(z | |Z| > c) = \frac{p_\mu(z)}{P_\mu(|Z| > c)} = \frac{\phi(z - \mu)}{\Phi(-c + \mu) + \Phi(-c - \mu)}. \quad (4)$$

Under Equation (4), the relationship between numerator and denominator is such that, for a given z , it is quite possible that the most likely value for μ is in the interval $[-c, c]$, even though z itself is conditioned to be outside that range.

By using this conditional approximate likelihood, we now derive improved estimators of μ . For any proposed value of μ , we can convert back to the desired log-odds ratio by using $\beta = \mu \hat{S}\hat{E}(\hat{\beta})$, where $\hat{S}\hat{E}(\hat{\beta})$ is obtained from standard approaches (i.e., does not consider the significance selection). One remarkable feature of our approach is that we can apply it to published summary results. To do so, we require only the significance threshold c , $\hat{\beta}$, and $\hat{S}\hat{E}(\hat{\beta})$. The standard error, if not provided directly, can be inferred from c , $\hat{\beta}$, and any one of the following: z , the p value, or an unconditional OR confidence interval.

The Conditional MLE

With the conditional likelihood, the maximum likelihood principle suggests the MLE estimator,

$$\tilde{\mu}_1 = \arg \max_{\mu} L_c(\mu),$$

which can be obtained with numerical maximization for any z and c (hereafter “~” will signify estimates based on the conditional likelihood). Note that in this setting, the conditional maximum-likelihood estimate provides no guarantee of unbiasedness or efficiency, a fact that does not appear to have been considered by other investigators. We have already applied large-sample assumptions in constructing the conditional likelihood (4), but as we show below, other estimators can provide reduced bias or mean-squared error for certain ranges of μ , and therefore $\hat{\beta}$.

Motivated by bias reduction, one might attempt to directly correct the bias in $\hat{\mu}$ by solving for μ in the equation $E_\mu(Z | |Z| > c) = z$. Such an estimator has intuitive appeal, representing the value of μ for which, after conditioning on significance, we would have expected to observe z . Perhaps surprisingly, this “bias-correction” estimator in fact turns out to be $\tilde{\mu}_1$. To see this, we take the derivative of the conditional likelihood with respect to μ , for which the identity $L'_c(\tilde{\mu}_1) = 0$ implies

$$z = \tilde{\mu}_1 + \frac{\phi(c - \tilde{\mu}_1) - \phi(c + \tilde{\mu}_1)}{\Phi(-c + \tilde{\mu}_1) + \Phi(-c - \tilde{\mu}_1)}. \quad (5)$$

Comparing Equation (2) to (5) implies that the bias-correction estimator and $\tilde{\mu}_1$ are the same. Similar estimators have been examined in the context of sequential clinical trials, in which effect parameters are estimated only after a stopping boundary has

been reached.²⁵ Despite its secondary motivation as a bias-correction estimator, the conditional MLE $\tilde{\mu}_1$ is not in fact unbiased because of nonlinearity in the bias of the naive estimator $\hat{\mu}$. Moreover, in this setting the conditional MLE has no special optimality properties, and other estimators may be reasonable. Nonetheless, we will show that $\tilde{\mu}_1$ is markedly improved over the naive estimator, both in terms of bias and mean squared error.

The Mean of the Normalized Conditional Likelihood

The motivation to reduce mean squared error (MSE) suggests another, perhaps less obvious estimator,

$$\tilde{\mu}_2 = \frac{\int_{-\infty}^{\infty} \mu L_c(\mu) d\mu}{\int_{-\infty}^{\infty} L_c(\mu) d\mu}, \quad (6)$$

which is easily calculated numerically. $\tilde{\mu}_2$ is the mean of the random variable following the distribution $L_c(\mu)$, normalized to be a proper density. $\tilde{\mu}_2$ has favorable MSE properties when averaged across a wide range of μ . This fact follows from an interpretation of $\tilde{\mu}_2$ as a posterior mean in a Bayesian treatment of the problem with a flat prior on μ .²⁶ However, $\tilde{\mu}_2$ is considered here as an entirely frequentist estimate, with bias and error examined at each value of μ and judged accordingly. For $|z|$ near the boundary c , $\tilde{\mu}_2$ typically represents a less aggressive shrinkage toward 0 compared to $\tilde{\mu}_1$.

A Compromise Estimator

In the treatment below, we will see that the conditional likelihood is typically skewed, and so $\tilde{\mu}_1$ and $\tilde{\mu}_2$ can differ appreciably for certain values of z . $\tilde{\mu}_2$ can show greater MSE than $\tilde{\mu}_1$ for μ near zero but is more favorable for μ away from zero. Thus, as a practical compromise, we also examine the estimator

$$\tilde{\mu}_3 = (\tilde{\mu}_1 + \tilde{\mu}_2)/2,$$

which balances the strengths of $\tilde{\mu}_1$ versus $\tilde{\mu}_2$.

Illustrations of the Conditional Likelihood

Figure 1 illustrates the conditional and unconditional likelihoods assuming an illustrative constant threshold $c = 5.0$. Figures 1A–1C correspond to $z = 5.2, 5.33,$ and 6.0 , respectively. For each panel, the unconditional likelihood is centered and maximized at z (indicated by a dot on each plot). For Figure 1A, in which z is only slightly above the threshold, the conditional likelihood is in contrast shifted aggressively toward zero ($\tilde{\mu}_1 = 0.66, \tilde{\mu}_2 = 2.53,$ and $\tilde{\mu}_3 = 1.60$). When z is well above the threshold ($z = 6.0$; Figure 1C), this shift is much smaller ($\tilde{\mu}_1 = 5.48, \tilde{\mu}_2 = 4.94,$ and $\tilde{\mu}_3 = 5.21$). For an intermediate z (Figure 1B), the shift is intermediate. Note that our estimates are obtained here for the μ version of the problem, and the conversion $\beta = \mu \hat{S}\hat{E}(\hat{\beta})$ must be performed before the results are interpreted on the log-odds scale.

As desired, the conditional likelihood shows a clear shift toward zero. But why is the shift so extreme, e.g., when $z = 5.2$? Such a z value (which is equivalent to $\hat{\mu}$) has already met genome-wide multiple-testing correction for statistical significance, but a shrinkage from $\hat{\mu} = 5.2$ to $\tilde{\mu}_1 = 0.66$ (for example) will effect a corresponding proportional reduction in the log-odds ratio. Thus, it seems our proposed estimation procedures can often adjust the estimated effect size to be practically insignificant. To see why the result is reasonable, consider that the conditional likelihood, as a frequentist construction, makes no judgment about the prior plausibility of various values of μ . When presented with a value z

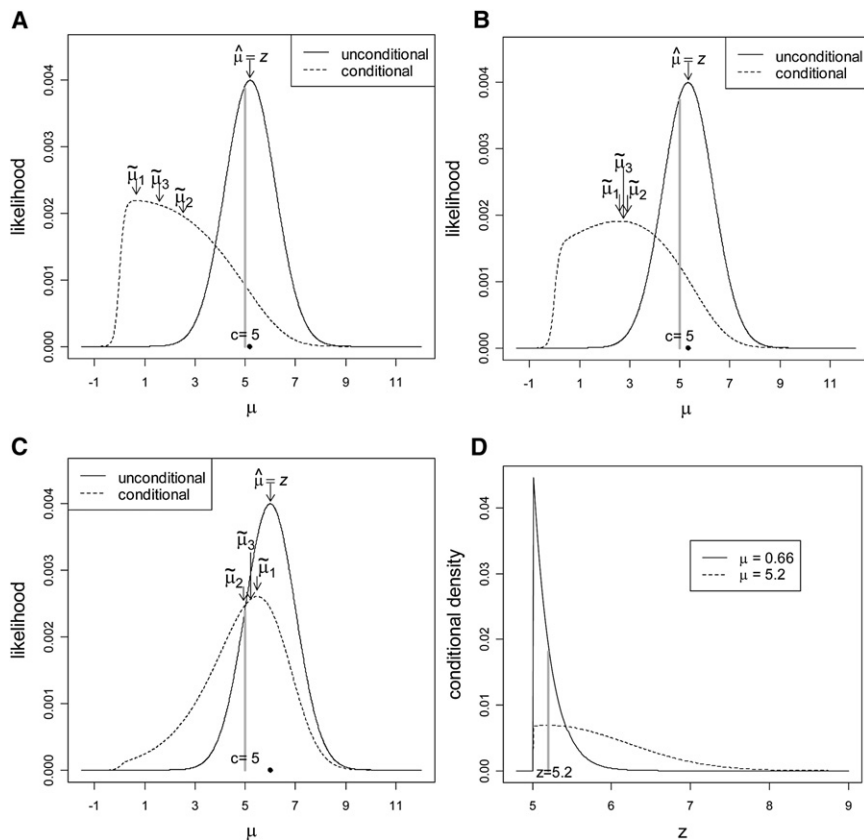


Figure 1. Behavior of the Unconditional and Conditional Likelihoods for μ Unconditional and conditional likelihoods of μ are presented for (A) $z = 5.2$, (B) $z = 5.33$, and (C) $z = 6$. The location of the observed z is indicated by a black dot on each plot. The conditional likelihood changes considerably for small changes in z near c . For larger z , the conditional likelihood approaches the unconditional likelihood. Likelihoods for $\mu < -c$ are negligible and not shown. (D) shows conditional densities of z for $\mu = 0.66$ and $\mu = 5.2$, with the relative likelihoods highlighted for a fixed value $z = 5.2$.

for each μ , it considers only the chance that z would have arisen, given that $|z| > c$. Figure 1D presents the (truncated normal) conditional densities for z under $\mu = 0.66$ and $\mu = 5.2$. These μ values were chosen because they represent the conditional and unconditional MLEs when $z = 5.2$. Note that these curves are conditional densities for z , not likelihoods. However, for a fixed value of z , the relative heights of the two curves reflect the conditional likelihoods for the two competing values of μ . From the curves, we can see the value $z = 5.2$ is $2.77\times$ more likely to arise when $\mu = 0.66$ than when $\mu = 5.2$. Expressed in another way, when μ values are truly of large magnitude, then z tends to overshoot the threshold c by a greater amount than was observed here for $z = 5.2$. Thus, in this instance, we would conclude that μ is not likely to be of large magnitude.

Our three proposed estimators can be easily computed numerically, and simple R and Excel programs to do so are available at our website. By using the threshold $c = 5.0$ for illustration, we have calculated the conditional expectations and MSEs for the three estimators, shown in Figures 2A and 2B. The three corrected estimators provide dramatically reduced bias compared to the naive estimator for much of the range of μ . For $\mu = 0$, by symmetry all estimators are unbiased. For $|\mu|$ considerably larger than c , all methods will give estimators near z and will be nearly unbiased. The corrected estimators tend to undercorrect for small μ and overcorrect for large μ . The conditional MLE $\hat{\mu}_1$ can be viewed as a first-order attempt to correct the bias, whereas z occupies the same range whether μ is small or large. In a sense, the corrected estimate splits the difference between the two extremes, leading to the observed pattern.

The MSE for $\hat{\mu} = z$ is extremely large for μ near zero, as predicted. MSEs for the corrected estimators are considerably smaller in the range of small to moderate μ . As described above, these estimators

are easily converted to the corresponding improved log(OR) estimators $\tilde{\beta}_1$, $\tilde{\beta}_2$, and $\tilde{\beta}_3$. Moreover, for large samples, the bias and MSE properties for μ will predominantly carry over to real data, essentially with a rescaling of the axes to convert μ to β .

Conditional Confidence Intervals

Proper interpretation of the corrected μ estimates requires an understanding of estimation error, conditioned on statistical

significance. Standard confidence interval (CI) procedures fail in this setting. For example, after conditioning on significance, a standard 95% CI for μ cannot contain 0, for otherwise it would not have been significant. Thus, when $\mu = 0$ the standard CI procedure has zero conditional coverage probability. Zöllner and Pritchard¹⁶ addressed this issue by using a standard maximum likelihood ratio approach applied to the conditional likelihood. In our setting, a $1 - \eta$ CI created in this manner would consist of all μ values such that $2\log(L_c(\tilde{\mu}_1)/L_c(\mu)) \leq q_{1-\eta}$, where $q_{1-\eta}$ is the $1 - \eta$ quantile of a χ^2_1 density. However, we have shown via numerical integration that in the μ version of the problem, the true coverage probability of this CI procedure can exhibit markedly conservative or anticonservative departures from $1 - \eta$, depending on the true μ . Approaches that use the second derivative at $\ln L_c(\tilde{\mu}_1)$ to estimate the error variance also fail. The difficulty arises because the conditional MLE is not normally distributed nor is the shape of $L_c(\mu)$ approximately normal for a realized data set.

To create confidence intervals with correct conditional coverage, we return to the original Neymanian concept of a confidence region,^{23,27} a concept that can always be applied when the distribution of a test statistic is known for each value of the unknown parameter. Let $A(\mu, 1 - \eta)$ be an acceptance region depending on μ such that $P_\mu(Z \in A(\mu, 1 - \eta) | |Z| > c) = 1 - \eta$. Given an observed z , the confidence region consists of all values μ such that $z \in A(\mu, 1 - \eta)$. It is straightforward to show that this approach gives exact coverage probability $1 - \eta$ for any μ . Among possible acceptance regions, we choose $A(\mu, 1 - \eta)$ as the interval between the $\eta/2$ and $1 - \eta/2$ quantiles of the conditional density $p_\mu(z | |Z| > c)$. Note that, although we have presented three competing point estimates for μ , our procedure yields only a single CI. Figure 2C shows the upper and lower confidence limits for our CI procedure for each z . Note that for $|z|$ near c , the confidence

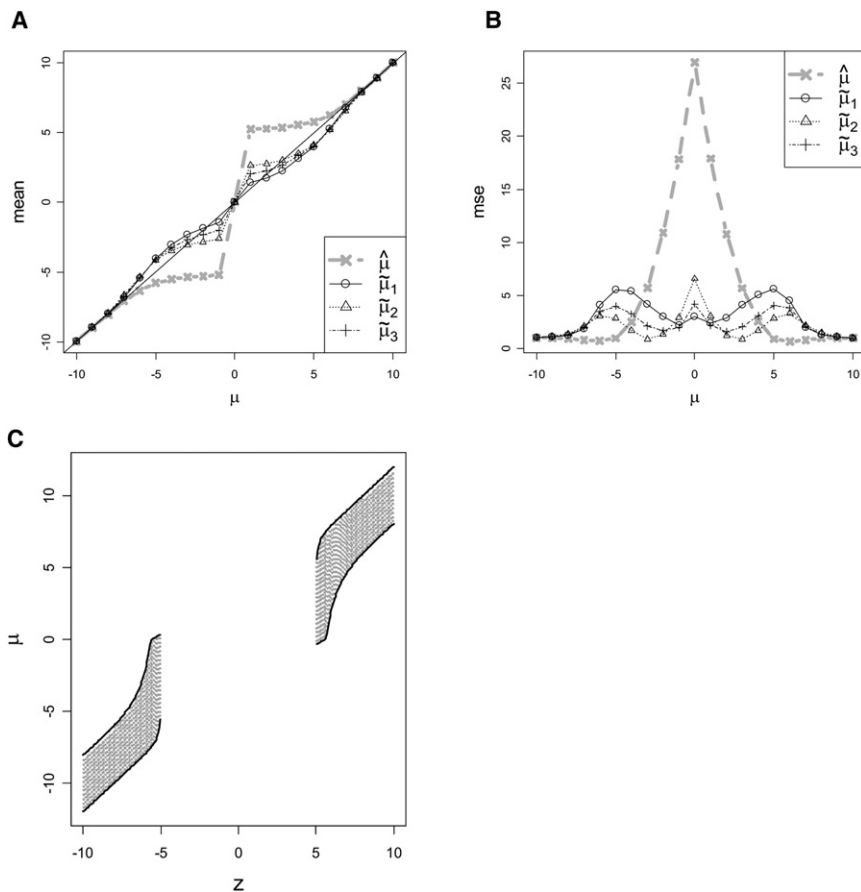


Figure 2. Estimators and Confidence Intervals for μ with Significance Threshold $c = 5$

(A) The expectation of naive estimator $\hat{\mu}$ shows substantial bias and (B) very large mean squared error for much of the range of μ , whereas the corrected estimators have lower bias and MSE (C) shows upper and lower confidence bounds for μ as a function of the observed statistic z .

$\pi = \sum_x \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} p(x)$ was used for calculating α . Finally, solving for the genotype probabilities conditioned on case-control status yields

$$P(X = x | Y = 1) = \frac{p(x)}{\pi} \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)},$$

$$P(X = x | Y = 0) = \frac{p(x)}{1 - \pi} \frac{1}{1 + \exp(\alpha + \beta x)}.$$

A standard result is that logistic modeling for β applies even when the data are sampled retrospectively.²⁰

Each data set was simulated and analyzed in *R* v.2.5.1. We will denote the total sample size $n = n_{cases} + n_{controls}$ and $n_{cases} = n_{controls}$ throughout. Most simulations consisted of $n = 1000$. This sample size is relatively small for a genome scan

interval can contain $\mu = 0$. This does not contradict the statistical significance—the intent of the procedure is to obtain correct coverage for any μ (including $\mu = 0$) after conditioning on significance. The conversion of the confidence limits to the β scale is $(\mu_{lower} SE(\hat{\beta}), \mu_{upper} SE(\hat{\beta}))$. Although our procedure is guaranteed correct conditional coverage in the idealized μ setting, our CI for β relies on large-sample normality assumptions for $\hat{\beta}$. Thus, we investigate empirical coverage of our procedure in the Results.

Simulations

To describe our simulations, we begin with basic notation for disease association studies. We let γ denote the disease status (0 = control, 1 = case) for an individual and x denote the SNP genotype predictor value. For a biallelic SNP with major allele *A* and minor allele *a*, x is defined as follows for genetic models with respect to *a*:

$$\begin{array}{ccc} \text{recessive} & \text{additive} & \text{dominant} \\ x = \begin{cases} 0, AA \\ 0, Aa \\ 1, aa \end{cases} & x = \begin{cases} 0, AA \\ 1, Aa \\ 2, aa \end{cases} & x = \begin{cases} 0, AA \\ 1, Aa \\ 1, aa \end{cases} \end{array}$$

We assume the logistic model for a randomly sampled individual in the population

$$\log(P(Y = 1 | x) / (1 - P(Y = 1 | x))) = \alpha + \beta x,$$

for some α , and β is the log-odds ratio for a unit increase in x . Rather than specify α directly, it is more interpretable to solve for α for a specified allele frequency and disease prevalence π . The marginal frequency of x is denoted $p(x)$ and is easily calculated from Hardy-Weinberg assumptions. With fixed disease prevalence, the identity

and was intentionally chosen to emphasize any departures from normality or difficulties in estimating $SE(\hat{\beta})$. We also examined larger sample sizes for several of the setups to examine the effect of sample size on bias, MSE, and confidence coverage. We assumed a disease prevalence of 0.01 throughout—the retrospective sampling is not very sensitive to this specification. We examined β ranging from -0.7 ($OR \approx 0.5$) to 0.7 ($OR \approx 2.0$). This range corresponds to biological plausibility for complex disease¹¹ and ensures that simulations span the range from low power to high power. For simplicity, we used $c = 5.0$, corresponding to a single p value of 5.7×10^{-7} , near the genome-wide threshold considered by others.²⁻⁴

For recessive models, we considered MAF values of 0.25 and 0.5—lower values created small expected cell counts that were problematic for sample sizes of 500 in each group. For the additive and dominant models, we considered minor allele frequency (MAF) values of 0.05, 0.1, 0.25, and 0.5. A single setup consisted of the genetic model, MAF, and β , and sufficient simulations were performed for each setup so that 1000 significant data sets were obtained. Setups with $\beta = 0$ required on the order of 10^9 to 10^{11} simulations for this rarified threshold. We sped up the analysis by first applying a chi-square test (Cochran-Armitage trend test for the additive model) to the data sets, which can be obtained without iterative maximization. The chi-square statistic was determined to have a close correspondence to z^2 obtained from the more computationally intensive logistic regression, and a chi-square statistic ≥ 24 was determined to capture essentially all data sets with $z^2 \geq c^2 = 25$. Data sets meeting the chi-square criterion were analyzed via logistic regression in *R* glm. For data sets achieving final significance as determined by logistic regression,

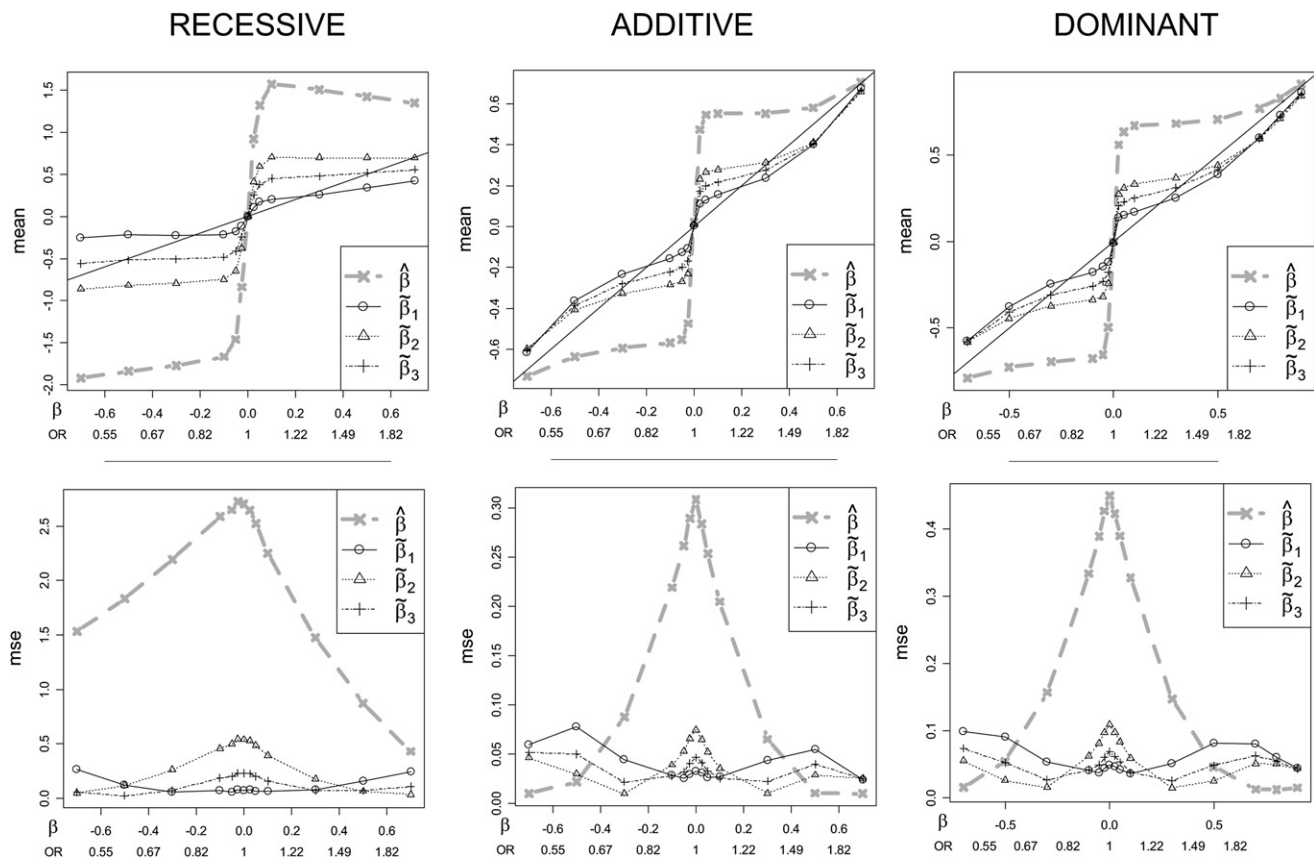


Figure 3. Expectations and Mean Squared Errors for the Three Genetic Models under MAF = 0.25

The corrected estimators show greatly improved performance for much of the range of β . The left panels correspond to the recessive model, the middle panels correspond to the additive model, and the right panels correspond to the dominant model. The top row shows expected values for the naive and conditional likelihood estimators versus β . The bottom row shows mean squared errors for the estimators. The y axes are rescaled to highlight details—the MSE is considerably larger for the recessive model because of scarcity of the risk homozygotes.

$\hat{\beta}$ and $S\hat{E}(\hat{\beta})$ were used for obtaining $\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3$, and conditional confidence intervals.

Results

In all scenarios described here, expectations and mean-squared errors are calculated conditional on significance, i.e., $|z| > c$.

Bias

The top row of Figure 3 plots the means for each of the naive and corrected estimators versus β (with corresponding OR values) for all models, with MAF = 0.25. The naive estimator shows very large bias, especially for moderate β . All of the corrected estimators show dramatically reduced bias across most of the range examined. For each model, the corrected estimates tend to undercorrect for small (magnitude) β while overcorrecting for large β . All of the methods become nearly unbiased for large β , as they must, because the conditional and unconditional likelihoods are nearly identical when $|z|$ is well beyond c . In terms of bias, $\tilde{\beta}_1$ performs best among the corrected estimates for small β . How-

ever, the overcorrection of the conditional MLE can be substantial for moderate to large β , especially for the recessive model. $\tilde{\beta}_2$ shrinks the estimates toward zero less dramatically, resulting in undercorrection for a larger part of the range of β . $\tilde{\beta}_3$ strikes a balance between the other two corrected estimates and has much improved bias for moderate β under the recessive model. All estimators are effectively unbiased for $\beta = 0$. A subtle asymmetry in the plots for positive and negative $\log(OR)$, most evident in the recessive model, occurs because MAF < 0.5 and, for a fixed prevalence, the logistic intercept α depends on β .

Mean Squared Error

The corresponding MSE values for the estimators are shown in the bottom row of Figure 3. The naive estimator $\hat{\beta}$ exhibits extremely large MSE for most β values examined. For $\beta = 0$, this is due to high variance, whereas for moderate β , the naive estimator has low variance but high bias. The corrected estimators show dramatically improved MSE for β in the interval $[-0.3, 0.3]$ (OR ranging from 0.74 to 1.35) that encompasses the bulk of significant associations thus far for complex diseases.^{3,4} The MSEs of $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are predominantly complementary. At $\beta = 0$, the

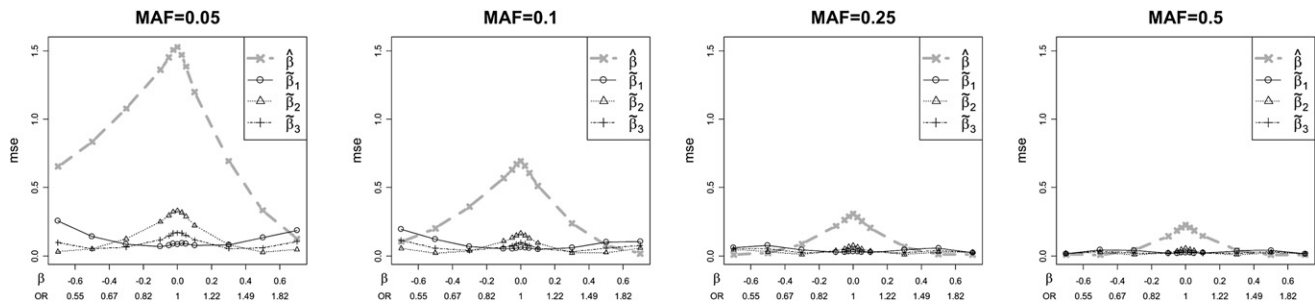


Figure 4. Mean Square Errors of the Estimators versus β for MAF Values Ranging from 0.05 to 0.5

The additive model is assumed, with $n = 1000$. The MSEs drop for larger MAF, but the relative performance of the estimators is maintained.

MSE ($\hat{\beta}_1$) is fairly low, whereas MSE ($\hat{\beta}_2$) peaks. For larger magnitude β , the roles reverse. As expected, $\hat{\beta}_3$ exhibits a more even MSE across the range, and represents a reasonable choice for stable error characteristics. For the additive and dominant models, $\hat{\beta}$ exhibits very low MSE for large β . This phenomenon is not as attractive as it appears, essentially resulting from a boundary effect in which $\hat{\beta}$ is nearly constant because z is just barely significant. In particular, for β outside of the plotted range, MSE ($\hat{\beta}$) rises again to the $\text{var}(\hat{\beta})$ value encountered in the unconditional setting.

The empirical bias and MSE observed in our simulations essentially follow the results from the μ version of the estimation problem, with a rescaling of the axes to convert μ to β . Our empirical results for the remaining MAF values

are plotted in Figure S1 available online and predominantly follow the results described for MAF = 0.25. Figure 4 shows a portion of these results for the additive model, in which the MSE is shown to drop for all estimators as the MAF increases. This occurs because for small MAF, the MSE is largely driven by the heterozygote genotype counts, which increase with the MAF. The key point of Figure 4 is that the relative advantages of the corrected estimators are preserved across a wide range of MAF values.

Confidence Coverage

Figure 5 presents the estimated coverage probabilities of 95% and 90% CIs with MAF = 0.25 for the three models. The top row shows the results for $n = 1000$. The coverage

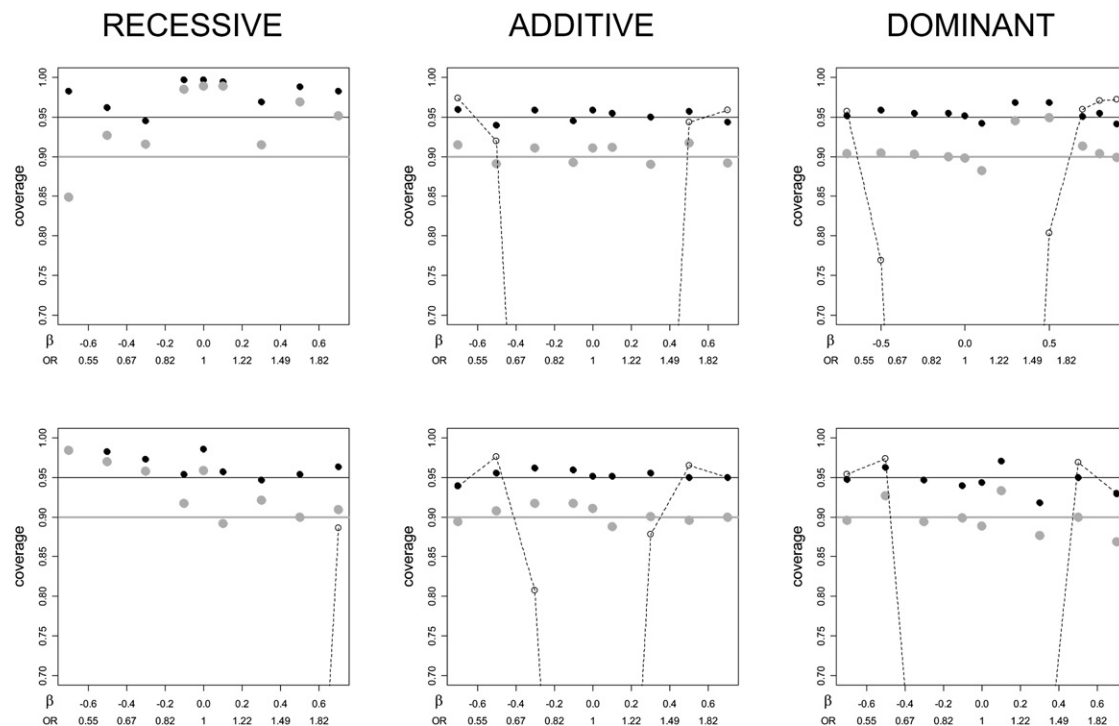


Figure 5. Estimates of the CI Coverage Probability Plotted against β for the Three Genetic Models, MAF = 0.25

Black dots correspond to 95% CIs; gray dots correspond to 90% CIs. The dashed curves represent coverage of standard 95% CIs that do not acknowledge the significance selection. The top row shows $n = 1000$ (500 cases and 500 controls). The bottom row shows $n = 2000$ (1000 cases and 1000 controls). Coverage is close to nominal, except for regions of overcoverage in the recessive model because of small cell counts (note that the y axis range begins at 0.7). For all models, the coverage will approach the nominal value as the sample size increases further.

is close to the nominal level for almost all the setups, except for conservativeness near $\beta = 0$ for the recessive model. The coverage of the naive confidence intervals is also depicted in the figure, dropping dramatically out of the axis range to zero coverage for β of small magnitude. For $n = 2000$, the coverage of the proposed procedure improves further, with a region of modest overcoverage for recessive models. Results for other MAF values are similar, and are presented in Figure S2.

Sample Sizes, Thresholds, and Covariates

Our setup conditions represent a wide range of realistic scenarios but cannot represent all situations and complicating factors. Fortunately, the large-sample behavior of the constructed approximate likelihood provides considerable robustness for our conclusions. Figure S3 shows the results of increasing sample size for several realistic β values for the additive model when MAF = 0.25. The bias and MSE for all the estimators are reduced as the sample size increases. For each sample size, the corrected estimators show superior bias and MSE compared to the naive estimator.

In maximum-likelihood settings, the distribution of the Wald test statistic is predominantly driven by $\beta/SE(\hat{\beta})$. This is also true for our conditional likelihood, because $\beta/SE(\hat{\beta})$ determines the noncentrality of the z statistic. For a fixed ratio $n_{cases} : n_{controls}$, the standard error is proportional to $1/\sqrt{n}$. Thus, for the setups in Figure 3 and Figure S1, a doubling of the sample size to $n = 2000$ (for example, and assuming cases and controls remain in the same ratio) would produce qualitatively similar results, with perhaps a slight improvement for the corrected estimates as the normality approximation improves. Moreover, we can make the results quantitatively comparable by appropriate rescaling. For example, for any value β for $n = 1000$, the comparable results for $n = 2000$ should correspond to $\beta' = \beta\sqrt{2}$. Figure S4A demonstrates an empirical example of this effective rescaling equivalence for the additive model, MAF = 0.25. Thus, the conclusions from our simulations extend to larger sample sizes.

Similarly, variations on the threshold c do not have much impact. A value of $c = 5.5$ would be considered quite conservative for genome scans, corresponding to Bonferroni control of family-wise error at 0.05 for 1.3 million SNPs. Empirical investigation requires many more simulations to achieve significance, but we find that the qualitative behavior of the estimators is unchanged (Figure S4B).

Finally, we simulated an example in which the additive model is fit (MAF = 0.25), and the logistic regression includes an additional continuous covariate [distributed $N(0,1)$, one fitted regression coefficient] and a discrete covariate [distributed $Binomial(2,0.05)$, two fitted coefficients]. The covariates were independent of case-control status and the test-locus genotype. The Wald statistic is relatively insensitive to inclusion of these extra parameters, and the relative change in degrees of freedom quite minimal. Accordingly, the results for our corrected estimators are virtually unchanged compared to the model without covariates

(Figure S4C—only $\tilde{\beta}_1$ is shown). Covariate considerations are increasingly important in genome scans, for example to control for confounding population stratification.

Analyses of Published Data Sets

Table 1 illustrates our reanalysis of an association study with a modest number of SNPs, as well as two GWASs, all of which had been analyzed with additive models. We begin with a brief description of the three studies, followed by our reanalysis. Yu et al.¹⁵ examined the lymphoma results described in Wang et al.,¹⁸ with 48 SNPs and a p value threshold $0.1/48 \approx 0.002$. The standard OR results were compared to bootstrap bias-corrected¹⁵ results, as well as the results from a larger pooled analysis involving seven studies.²⁸ The SNPs rs1800629 and rs909253 were found to be significant, with ORs 1.54 and 1.40, respectively. Todd et al.³ list four significant SNPs resulting from two type 1 diabetes (T1D [MIM 222100]) GWAS studies, highlighting for especial consideration those SNPs with p value less than 5×10^{-7} . In addition, the authors conducted a larger case-control follow-up study to confirm these results. Scott et al.⁴ performed numerous analyses of several type 2 diabetes (T2D [MIM 125853]) data sets (FUSION, DGI, and WTCCC/UKT2D). We consider here only the SNPs reported by the T2D authors using the declared genome-wide significance threshold ($p < 5 \times 10^{-8}$) for the combined analysis of all studies.

With only the published odds ratios, p values, and stated significance thresholds, we produced bias-corrected odds ratios for all of these studies. Our corrected β estimates are exponentiated so that odds ratios were obtained: For example, $OR_{\tilde{\beta}_1} = \exp(\tilde{\beta}_1)$. For the two lymphoma SNPs (Table 1, top section), the p values are slightly above the threshold, and our bias-corrected estimates shrink the naive OR estimates markedly. Our estimated values match well with the bootstrap-corrected values obtained by Yu et al.,¹⁵ as well as the pooled analysis results from Rothman et al.²⁸

For the four T1D SNPs (Table 1, middle section), our analysis results in noticeably less extreme OR estimates (Table 1, middle section) than those reported by Todd et al.³ The corrected ORs and CIs for the most extreme SNP, rs17696736, are only slightly changed from the published estimated of 1.37 because the result is so extreme ($p = 7.27 \times 10^{-14}$). However, the follow-up study obtained a considerably lower value (OR = 1.16), with the 95% CI not overlapping the earlier estimates, thereby suggesting possible heterogeneity in population sampling. For the two least significant T1D SNPs among those considered, the corrected ORs show a more substantial change. It is worth noting that the OR estimate corresponding to the SNP rs12708716 was shrunk from 0.77 to ~ 0.82 by our methods, whereas the estimated OR from the follow-up study was 0.83. We also note that for the four significant T1D SNPs, as well as an additional three SNPs approaching significance (Table 1 of Todd et al.³), the follow-up study always gave a less extreme OR estimate than the initial studies. This result is

Table 1. Original versus Corrected Odds-Ratio Estimates for Three Published Genetic Association Studies

Study	SNP	Minor Allele Frequency	p Value	Reported OR, ^a (95% CI)	Bootstrap ^b Estimates	Bias-Corrected Estimates ^c			Bias ^c - Corrected (95% CI)	Follow-Up ^d OR, (95% CI)
						\tilde{OR}_1	\tilde{OR}_2	\tilde{OR}_3		
Association Study of Lymphoma, Wang et al. ¹⁸ (318 Cases and 766 Controls)										
	rs1800629	0.217	5.7×10^{-4}	1.54	1.29	1.08	1.25	1.16	(0.94,1.85)	1.29
	rs909253	0.358	7.4×10^{-4}	1.4	1.18	1.06	1.18	1.12	(0.95,1.56)	1.16
GWAS of T1D, Todd et al. ³ (2000 Cases and 3000 Controls)										
	rs17696736	0.423	7.27×10^{-14}	1.37 (1.27,1.49)	-	1.37	1.36	1.37	(1.25,1.49)	1.16 (1.09,1.23)
	rs2292239	0.34	1.49×10^{-9}	1.3 (1.20,1.42)	-	1.26	1.23	1.25	(1.08,1.42)	1.28 (1.20,1.36)
	rs12708716	0.322	1.28×10^{-8}	0.77 (0.70,0.84)	-	0.82	0.84	0.83	(0.71,1.00)	0.83 (0.78,0.89)
	rs2542151	0.163	8.4×10^{-8}	1.33 (1.20,1.49)	-	1.04	1.15	1.09	(0.99,1.44)	1.29 (1.19,1.40)
GWAS of T2D, Scott et al. ⁴ (9521 Cases and 12183 Controls)										
	rs7903146	0.18	1.0×10^{-48}	1.37 (1.31,1.43)	-	1.37	1.37	1.37	(1.31,1.43)	
	rs4402960	0.30	8.9×10^{-16}	1.14 (1.11,1.18)	-	1.14	1.14	1.14	(1.10,1.18)	
	rs10811661	0.85	7.8×10^{-15}	1.2 (1.14,1.25)	-	1.2	1.2	1.2	(1.14,1.26)	
	rs8050136	0.38	1.3×10^{-12}	1.17 (1.12,1.22)	-	1.17	1.16	1.16	(1.10,1.22)	
	rs5219	0.46	6.7×10^{-11}	1.14 (1.10,1.19)	-	1.13	1.11	1.12	(1.05,1.19)	
	rs7754840	0.36	4.1×10^{-11}	1.12 (1.08,1.16)	-	1.11	1.1	1.11	(1.05,1.16)	
	rs1111875	0.52	5.7×10^{-10}	1.13 (1.09,1.17)	-	1.1	1.09	1.1	(1.00,1.17)	

^a Standard OR values as reported.

^b Bootstrap correction reported in Ref. ¹⁵.

^c Correction method proposed in this paper.

^d Replication or other follow-up result for the SNP.

strong empirical evidence for significance bias and showed that corrected OR approaches are needed.

The bottom portion of Table 1 gives the results for the combined T2D studies. All of the p values are considerably beyond the significance threshold, and so the corrected estimates are nearly unchanged from the original estimates. This phenomenon is hopeful, in the sense that with very large studies, OR estimates can be attained that will not be shrunk to irrelevance by corrected OR estimates.

Discussion

We have presented an approach that greatly reduces significance bias for odds ratios in genome association scans and that is much simpler than competing approaches. We favor the use of $\tilde{\beta}_3$ as a general-purpose estimator with fairly uniform MSE as a function of β . However, all of the three corrected estimators have greatly superior performance compared to the naive estimator. Although developed for case-control applications, our methodology is an effective blueprint to perform inference whenever a Wald-like statistic has been used to declare significance. Thus, the general approach can be used in numerous other settings, including regression-based quantitative-trait association analyses. Our results are qualitatively similar to those of other investigators^{15,16} (e.g., see bias curves similar to ours in Figure 2 of Zöllner and Pritchard¹⁶). Additional comparisons to these approaches should be performed in future work, although comparison is complicated by differing genetic models. To our knowledge, our approach is the only

method that can perform bias correction based only on published summary tables.

The widespread application of conditional likelihood estimators in genome scans will no doubt be discouraging to genetic investigators, who may expend considerable time and expense only to find that a significant SNP is estimated to have a very weak effect. Nonetheless, we view this process as healthy and necessary for the genetics community, in particular to tamp down expectations that significant findings will be easily replicated. The use of our estimators may also have an additional benefit of discouraging excessive massaging of data and trying various test procedures to achieve genome-wide significance. If a SNP suddenly becomes significant after numerous data manipulation procedures have been applied, its z statistic is likely to be only slightly above the threshold c . Thus, as we observed in the μ version of the problem, the conditional-likelihood estimator will be dramatically shrunk toward the null. Thus, the estimated SNP effect size will be very modest, as is appropriate here for a likely spurious finding.

Our current approach does not explicitly consider multi-stage or other sequential designs, in which SNPs meeting a loose standard of significance are used for further testing in a follow-up sample. However, for multistage designs in which almost all SNPs that will eventually be declared significant are carried forward to later stages, the approach may be used directly. Also, our results technically hold for a SNP randomly selected from those achieving the significance threshold, and thus an additional bias may be anticipated for the most highly significant SNPs among a collection of significant SNPs. Although we believe this second

source of bias is much less than that produced by significance selection, it is the subject of continuing investigation.

Our rejection-sampling scheme was feasible, but it required a massive number of simulations to provide accurate results. Future work in this area may benefit from the practical development of importance sampling or related computational techniques to provide flexible and accurate simulations conditioned on significance.

Appendix A

Derivation of the Conditional Mean for the Naive Estimator

$$\begin{aligned}
 E(Z \mid |Z| > c) &= K^{-1} \left[\int_{-c}^{-\mu} z\phi(z - \mu) dz + \int_c^{\infty} z\phi(z - \mu) dz \right], \\
 \text{where } K &= \Phi(-c + \mu) + \Phi(-c - \mu) \\
 &= \mu + K^{-1} \left[\int_{-\infty}^{-c-\mu} x\phi(x) dx + \int_{c-\mu}^{\infty} x\phi(x) dx \right], \\
 x &= z - \mu \\
 &= \mu + K^{-1} \left[(2\pi)^{-\frac{1}{2}} \int_{\frac{1}{2}(c+\mu)^2}^{\frac{1}{2}(c+\mu)^2} e^{-y} dy \right. \\
 &\quad \left. + (2\pi)^{-\frac{1}{2}} \int_{\frac{1}{2}(c-\mu)^2}^{\infty} e^{-y} dy \right], y = \frac{1}{2}x^2 \\
 &= \mu + K^{-1} \left[(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(c-\mu)^2} - (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(c+\mu)^2} \right] \\
 &= \mu + \frac{\phi(c-\mu) - \phi(c+\mu)}{K}
 \end{aligned}$$

Supplemental Data

Four figures are available at <http://www.ajhg.org/>.

Acknowledgments

This paper was supported in part by the Carolina Environmental Bioinformatics Center (EPA RD-83272001) and NIH grants R01 GM074175 and R01 HL068890. We thank the anonymous reviewers, who provided numerous suggestions that led to improvements in the manuscript.

Received: December 24, 2007

Revised: February 28, 2008

Accepted: March 5, 2008

Published online: April 17, 2008

Web Resources

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/>

R code and a simple Excel calculator, www.bios.unc.edu/~fwright/genomebias

References

- Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11, 241–247.
- Zondervan, K.T., and Cardon, L.R. (2007). Designing candidate gene and genome-wide case-control association studies. *Nat. Protocols* 2, 2492–2501.
- Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F., et al. (2007). Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* 39, 857–864.
- Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U., et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316, 1341–1345.
- Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516–1517.
- Allison, D.B., Fernandez, J.R., Heo, M., Zhu, S., Etzel, C., Beasley, T.M., and Amos, C.I. (2002). Bias in estimates of quantitative-trait-locus effect in genome scans: Demonstration of the phenomenon and a method-of-moments procedure for reducing bias. *Am. J. Hum. Genet.* 70, 575–585.
- NCI-NHGRI Working Group on Replication in Association Studies, Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., et al. (2007). Replicating genotype-phenotype associations. *Nature* 447, 655–660.
- Garner, C. (2007). Upward bias in odds ratio estimates from genome-wide association studies. *Genet. Epidemiol.* 31, 288–295.
- Goring, H.H., Terwilliger, J.D., and Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *Am. J. Hum. Genet.* 69, 1357–1369.
- Hirschhorn, J.N., Lohmueller, K., Byrne, E., and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet. Med.* 4, 45–61.
- Ioannidis, J.P., Ntzani, E.E., Trikalinos, T.A., and Contopoulos-Ioannidis, D.G. (2001). Replication validity of genetic association studies. *Nat. Genet.* 29, 306–309.
- Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S., and Hirschhorn, J.N. (2003). Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* 33, 177–182.
- Siegmund, D. (2002). Upward bias in estimation of genetic effects. *Am. J. Hum. Genet.* 71, 1183–1188.
- Sun, L., and Bull, S.B. (2005). Reduction of selection bias in genomewide studies by resampling. *Genet. Epidemiol.* 28, 352–367.
- Yu, K., Chatterjee, N., Wheeler, W., Li, Q., Wang, S., Rothman, N., and Wacholder, S. (2007). Flexible design for following up positive findings. *Am. J. Hum. Genet.* 81, 540–551.
- Zöllner, S., and Pritchard, J.K. (2007). Overcoming the winner's curse: Estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.* 80, 605–615.
- Balding, D.J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7, 781–791.
- Wang, W.Y., Barratt, B.J., Clayton, D.G., and Todd, J.A. (2005). Genome-wide association studies: Theoretical and practical concerns. *Nat. Rev. Genet.* 6, 109–118.
- Aschengrau, A., and Seage, G.R. (2003). *Essentials of Epidemiology in Public Health* (Sudbury, MA: Jones & Bartlett Publishers).
- McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models* (Boca Raton, FL: Chapman & Hall/CRC).

21. Agresti, A. (2002). *Categorical Data Analysis* (Hoboken, NJ: Wiley-Interscience).
22. Cox, D.R., and Snell, E.J. (1989). *Analysis of Binary Data* (Boca Raton, FL: Chapman & Hall/CRC).
23. Rao, C.R. (2002). *Linear Statistical Inference and Its Application* (Hoboken, NJ: Wiley-Interscience).
24. Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.* 54, 426–482.
25. Liu, A.Y., Troendle, J.F., Yu, K.F., and Yuan, V.W. (2004). Conditional maximum likelihood estimation following a group sequential test. *Biometrical Journal* 46, 760–768.
26. Leonard, T., and Hsu, J.S.J. (2001). *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers* (Cambridge: Cambridge University Press).
27. Lehmann, E.L., and Casella, G. (1998). *Theory of Point Estimation* (New York: Springer).
28. Rothman, N., Skibola, C.F., Wang, S.S., Morgan, G., Lan, Q., Smith, M.T., Spinelli, J.J., Willett, E., De Sanjose, S., Cocco, P., et al. (2006). Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: A report from the InterLymph Consortium. *Lancet Oncol.* 7, 27–38.

Note Added in Proof

While this paper was in press, a paper by Zhong, H. and Prentice, R.L. proposed a similar methodology, with some differences in approaches to point and interval estimation: Zhong, H. and Prentice, R.L. (2008). Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics*, in press.