Quantifying and Correcting for the Winner's Curse in Genetic Association Studies

Rui Xiao and Michael Boehnke[*]

Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Michigan

Genetic association studies are a powerful tool to detect genetic variants that predispose to human disease. Once an associated variant is identified, investigators are also interested in estimating the effect of the identified variant on disease risk. Estimates of the genetic effect based on new association findings tend to be upwardly biased due to a phenomenon known as the ''winner's curse.'' Overestimation of genetic effect size in initial studies may cause follow-up studies to be underpowered and so to fail. In this paper, we quantify the impact of the winner's curse on the allele frequency difference and odds ratio estimators for one- and two-stage case-control association studies. We then propose an ascertainment-corrected maximum likelihood method to reduce the bias of these estimators. We show that overestimation of the genetic effect by the uncorrected estimator decreases as the power of the association study increases and that the ascertainment-corrected method reduces absolute bias and mean square error unless power to detect association is high. *Genet. Epidemiol.* 33:453–462, 2009. © 2009 Wiley-Liss, Inc.

Key words: winner's curse; ascertainment bias; genome-wide association study; maximum likelihood

INTRODUCTION

Large-scale genetic association studies are now commonly used to localize genetic variants that predispose to a wide range of human diseases. In genetic association studies, once the disease-predisposing variants are identified, it is of interest to estimate the genetic effect of those variants on disease risk. The simplest method of estimating the effect size of the variant is to calculate the difference of the observed risk allele frequency between cases and controls or the corresponding odds ratio. However, these naïve estimators are likely to overestimate the true genetic effect size as a consequence of the ''winner's curse'' [Lohmueller et al., 2003], a phenomenon first described in the auction theory literature [Bazerman and Samuelson, 1983]. In auctions, participants place bids on an item. Even if the bids are unbiased, the winning bid is likely to overestimate the true item value since it is the highest among all the bids. In genetic association studies, an initial positive finding plays the role of the winning bid, since we generally focus on genetic effect size estimates only for the variants that yield significant evidence for association, resulting in effect size estimates that are upwardly biased. We refer to this bias as ''ascertainment bias'' since it is caused by ascertaining only those samples that result in significant association evidence. If the sample size calculation for a subsequent study is based on an overestimated effect size, replication studies are likely to be underpowered and so more likely to fail. A review of association studies [Ioannidis et al., 2001] has described

the overestimation in first positive reports, consistent with the winner's curse.

This problem has drawn attention from several investigators in the context of genetic linkage and association studies [Göring et al., 2001; Siegmund, 2002; Allison et al., 2002; Sun and Bull, 2005; Wu et al., 2006; Garner, 2007; Yu et al., 2007; Zöllner and Pritchard, 2007; Ghosh et al., 2008; Zhong and Prentice, 2008]. Göring et al. [2001] recommended the use of two independent data sets: one for locus mapping, the other for parameter estimation. An obvious disadvantage of this strategy is the power loss due to splitting the sample in two. Sun and Bull [2005] proposed resampling estimators that employ repeated random sample splitting of the data via cross-validation or the bootstrap. Wu et al. [2006] compared their bootstrap estimators for locus-specific quantitative trait linkage analysis, and, in the context of two-stage design, Yu et al. [2007] applied a bootstrap estimator to correct for stage 1 bias and improve sample size estimates for stage 2. Zöllner and Pritchard [2007] used computer simulation to evaluate the magnitude of the winner's curse effect in case-control studies and proposed a maximum likelihood method to correct for it. Their method estimates the frequencies of all genotypes and corresponding penetrance parameters based on a known population prevalence of the disease under different inheritance models. Garner [2007] studied the source of the upward bias in the odds ratio estimate in genome-wide association studies, but did not propose a method to correct for it. Zhong and Prentice [2008] and

Ghosh et al. [2008] recently proposed conditional-likelihood-based methods for point and interval estimation of the (logarithm of the) odds ratio in the context of logistic regression analysis of case-control status using genotype categories as a covariate.

In this paper, we take a direct approach to evaluate and correct for the effect of winner's curse in the context of case-control genetic association studies. In contrast to previous simulation-based evaluations, we calculate analytically the impact of the winner's curse on estimates of the allele frequency difference between cases and controls and the corresponding odds ratios as a function of sample size, allele frequencies, and statistical significance level. We then describe a simple ascertainment-corrected maximum likelihood method to estimate the risk allele frequency difference and odds ratio. Our method is most similar to that of Zöllner and Pritchard [2007], but in contrast to their method, ours estimates directly the allele frequency difference or odds ratio, instead of estimating the penetrance parameters. We compare the performance (bias, standard error, and mean square error (MSE)) of our ascertainment-corrected maximum likelihood estimators (MLEs) to that of the naïve, uncorrected estimators. We extend these calculations to two-stage association studies, in which all markers are genotyped on a set of individuals in stage 1, and the most promising markers are followed up by genotyping a second set of individuals in stage 2.

Consistent with Zöllner and Pritchard [2007], we find that (1) the factors that result in overestimation of the allele frequency difference can be summarized by study power, independent of sample size and allele frequency, and that overestimation decreases as power increases; and (2) compared to the uncorrected estimator of the allele frequency difference, the ascertainment-corrected estimator results in reduced absolute bias when study power is low or moderate, and has comparable absolute bias when power is high. Further, we find that (3) for the logarithm of the odds ratio (ln OR), overestimation can again be summarized by study power, independent of sample size and allele frequency, and that overestimation decreases as power increases; (4) compared to the uncorrected estimator, the ascertainment-corrected MLE of the OR generally results in reduced bias and MSE, and (5) for reasonable two-stage designs [Skol et al., 2007], results mirror those for the corresponding one-stage designs. We recommend use of this ascertainment-corrected maximum likelihood method for estimation of genetic effect size in large-scale genetic association studies.

METHODS

ONE-STAGE DESIGN

Model and assumptions. We assume independent samples of N cases and N controls genotyped at an autosomal disease locus with alleles D and d. Let p and $p + \delta$ ($\delta \neq 0$) denote the frequency of the risk allele D in controls and cases, respectively. For a complex disease, we expect the genetic effect size to be small, so that Hardy-Weinberg equilibrium predictions provide a good approximation to the genotype frequencies in both controls and cases. Under this assumption, the counts m_0 and m_1 of the

risk allele D in controls and cases follow independent binomial distributions on $2N$ trials with probabilities of success p and $p + \delta$, respectively.

Let X be the standard Pearson χ^2 test statistic for association in a 2×2 table of allele counts in cases and controls. Under the assumption of Hardy-Weinberg equilibrium, X follows a χ^2 distribution with one degree of freedom under the null hypothesis of no association ($\delta = 0$). We claim an association significant if X exceeds the critical value x_α at significance level α.

Uncorrected (naïve) MLEs. In practice, investigators generally estimate the allele frequency difference between cases and controls by its MLE $\hat{\delta}_{un} = m_1/2N - m_0/2N$, or the corresponding odds ratio by

$$\widehat{OR}_{un} = \frac{m_1(2N - m_0)}{m_0(2N - m_1)}$$

We call these uncorrected MLEs "naïve" because they ignore the bias associated with focusing on genetic markers with statistically significant association results.

To model the impact of the winner' curse, we calculate the expected value of the uncorrected MLE $\hat{\delta}_{un}$ of the allele frequency difference δ conditional on obtaining significant evidence for association:

$$E(\hat{\delta}_{un}|X > x_\alpha) = \frac{\sum_{(m_0,m_1) \in I} \hat{\delta}_{un} P(m_0, m_1)}{\sum_{(m_0,m_1) \in I} P(m_0, m_1)} \quad (1)$$

and from it the bias of the estimator as $E(\hat{\delta}_{un}|X > x_\alpha) - \delta$, and the proportional bias as $(E(\hat{\delta}_{un}|X > x_\alpha) - \delta)/\delta$. Here, $I = \{(m_0, m_1) : X(m_0, m_1) > x_\alpha\}$ is the set of allele count pairs that result in statistically significant evidence for association and

$$P(m_0, m_1) = \binom{2N}{m_0} p^{m_0}(1-p)^{2N-m_0}$$
$$\times \binom{2N}{m_1}(p+\delta)^{m_1}(1-p-\delta)^{2N-m_1}. \quad (2)$$

Note that the denominator in (1) is the power to detect association if we genotype the disease SNP.

The standard error of the uncorrected MLE $\hat{\delta}_{un}$ can be calculated as

$$SE(\hat{\delta}_{un}|X > x_\alpha) = \sqrt{E(\hat{\delta}_{un}^2|X > x_\alpha) - (E(\hat{\delta}_{un}|X > x_\alpha))^2} \quad (3)$$

where $E(\hat{\delta}_{un}^2|X > x_\alpha)$ may be calculated by replacing $\hat{\delta}_{un}$ by $\hat{\delta}_{un}^2$ in (1).

We also calculate the absolute bias of $\hat{\delta}_{un}$ as:

$$E(|\hat{\delta}_{un} - \delta||X > x_\alpha) = \frac{\sum_{(m_0,m_1) \in I} |\hat{\delta}_{un} - \delta| \times P(m_0, m_1)}{\sum_{(m_0,m_1) \in I} P(m_0, m_1)} \quad (4)$$

Analogous formulae allow us to calculate the conditional bias, standard error, and absolute bias of the uncorrected MLE of the odds ratio OR, and its logarithm ln OR.

Ascertainment-corrected MLEs. The naive estimators ignore the fact that we typically are interested in estimates of the allele frequency difference δ and the odds ratio OR only if we have strong evidence for association.

To address this, we propose an ascertainment-corrected maximum likelihood method that conditions on obtaining evidence for association. To this end, we calculate the conditional likelihood function

$$L(p, \delta | X > x_\alpha) = P(m_0, m_1 | X > x_\alpha)$$

$$= 1\{X > x_\alpha | m_0, m_1, N\} \frac{P(m_0, m_1)}{\sum_{(m_0, m_1) \in I} P(m_0, m_1)} \quad (5)$$

where the indicator function $1\{X > x_\alpha | m_0, m_1, N\}$ equals 1 or 0 depending on whether or not $X > x_\alpha$.

We maximize $L(p, \delta | X > x_\alpha)$ as a function of p and δ to obtain the ascertainment-corrected MLEs $\hat{p}_{as}$ and $\hat{\delta}_{as}$ by using the Nelder and Mead [1965] simplex method. We calculate the empirical standard errors of these estimators based on 1,000 simulation replicates, and the asymptotic-theory standard errors by calculating the observed information matrix (see Appendix) evaluated at the parameter estimates:

$$I(\hat{p}_{as}, \hat{\delta}_{as}) = -\partial_{p,\delta}^2 \log L(p, \delta | X > x_\alpha)|_{\hat{p}_{as}, \hat{\delta}_{as}} \quad (6)$$

The covariance matrix for $\hat{p}_{as}$ and $\hat{\delta}_{as}$ can be approximated by $I^{-1}(\hat{p}_{as}, \hat{\delta}_{as})$. We take advantage of the invariance property of the MLE to calculate the ascertainment-corrected MLEs for the odds ratio and its logarithm, and apply the delta method [Rao, 1965] to obtain their standard errors. We calculate the MSE for the estimators by taking the sum of the variance and the squared bias of the estimator.

TWO-STAGE DESIGN

Model and assumptions. We next consider two-stage association studies, in which N_1 cases and N_1 controls are genotyped for all markers, and only the most promising markers are genotyped in the second stage in an additional N_2 cases and N_2 controls. Let p_i and δ_i be the risk allele frequencies in controls and the allele frequency difference between cases and controls in stage i. Given genetic homogeneity between stages 1 and 2, $p_1 = p_2 = p$ and $\delta_1 = \delta_2 = \delta$. At each stage, we calculate the association test statistic using the data only from that stage

$$Z_i = \frac{\hat{p}_{i1} - \hat{p}_{i0}}{\sqrt{[\hat{p}_{i0}(1 - \hat{p}_{i0}) + \hat{p}_{i1}(1 - \hat{p}_{i1})]/(2N_i)}} \quad (7)$$

where $\hat{p}_{i0}$ and $\hat{p}_{i1}$ are the naïve MLEs of the risk allele frequencies in controls and cases respectively, at stage i, $\hat{p}_{ij} = (m_{ij}/2N_i)$ $(i = 1, 2; j = 0, 1)$. Under null hypothesis of no disease–marker association ($\delta = 0$), the association test statistic Z_i follows a standard normal distribution with mean 0 and variance 1.

We employ a joint analysis strategy for this two-stage study [Satagopan et al., 2002; Skol et al., 2006] by calculating

$$Z_{12} = \sqrt{\pi_{sample}}\, Z_1 + \sqrt{1 - \pi_{sample}}\, Z_2 \quad (8)$$

where $\pi_{sample} = N_1/(N_1 + N_2)$ is the proportion of individuals genotyped in stage 1. We claim significant association when both $|Z_1|$ and $|Z_{12}|$ exceed the relevant critical values C_1 and C_{12} in joint analysis. C_1 is calculated so that $P(|Z_1| > C_1) = \pi_{marker}$, where π_{marker} is the propor-

tion of markers to be genotyped in stage 2, and C_{12} by finding the threshold so that $P(|Z_1| > C_1, |Z_{12}| > C_{12}) = P(|Z_{12}| > C_{12} | |Z_1| > C_1) \times P(|Z_1| > C_1)$ results in the desired significance level [Skol et al., 2006].

Uncorrected (naïve) MLEs. The uncorrected MLE of the risk allele frequency difference for the two-stage design $\hat{\delta}_{12} = \pi_{sample}\hat{\delta}_1 + (1 - \pi_{sample})\hat{\delta}_2$, where $\hat{\delta}_i = m_{i1}/2N_i - m_{i0}/2N_i$, $i = 1, 2$. The bias of the uncorrected MLE $\hat{\delta}_{12}$ can be calculated exactly as for one-stage design by formula (1) and similarly the proportional bias. However, exact calculation becomes computationally difficult when N_1 or N_2 is large, so we simulated $n = 1,000$ data sets satisfying $|Z_1| > C_1$ and $|Z_{12}| > C_{12}$ and approximated the expectation and empirical standard error of $\hat{\delta}_{12}$ by calculating the mean and the standard error of the uncorrected MLE of the n simulated data sets:

$$E(\hat{\delta}_{12} | |Z_1| > C_1, |Z_{12}| > C_{12}) \approx \bar{\hat{\delta}}_{12}$$

$$= \frac{1}{n} \sum_{j=1}^{n} (\pi_{sample}\hat{\delta}_{1j} + (1 - \pi_{sample})\hat{\delta}_{2j}) \quad (9)$$

$$\mathrm{SE}(\hat{\delta}_{12} | |Z_1| > C_1, |Z_{12}| > C_{12}) = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (\hat{\delta}_{12,j} - \bar{\hat{\delta}}_{12})^2} \quad (10)$$

Ascertainment-corrected MLEs. In analogy to the one-stage design, the two-stage ascertainment-corrected likelihood

$$L(p, \delta | |Z_1| > C_1, |Z_{12}| > C_{12}) = P(m | |Z_1| > C_1, |Z_{12}| > C_{12})$$

$$= \frac{1\{|Z_1| > C_1, |Z_{12}| > C_{12} | m, N_1, N_2\}P(m)}{P(|Z_1| > C_1, |Z_{12}| > C_{12})} \quad (11)$$

Here, $m = (m_{10}, m_{11}, m_{20}, m_{21})$, $1(|Z_1| > C_1, |Z_{12}| > C_{12} | m, N_1, N_2)$ is an indicator function taking values of 1 or 0 depending on whether or not $|Z_1| > C_1$ and $|Z_{12}| > C_{12}$, and $P(m)$ is the product of four binomial probabilities. The denominator of (11) is again the power of the study, and can be evaluated as described by Skol et al. [2006]. We maximize the likelihood (11) to get MLEs of p and δ by using the Nelder-Mead simplex approach and obtain empirical standard errors based on 1,000 simulation replicates.

RESULTS

ONE-STAGE DESIGN

Bias of the uncorrected MLEs. For a locus showing association ($\delta \neq 0$), our analytical calculation demonstrates upward bias in the genetic effect size by the naïve estimator $\hat{\delta}_{un}$ of the allele frequency difference δ (Fig. 1). This bias is particularly severe when power is low, owing to the small sample size N and/or small allele frequency difference δ (Table I, Fig. 2A). As power approaches one, the bias disappears. Under the null hypothesis ($\delta = 0$), $\hat{\delta}_{un}$ is unbiased, since δ is equally likely to be over- or under-estimated. However, the absolute bias of this uncorrected estimator is extremely high when $\delta = 0$ or when δ is small (Fig. 1). Due to approximate symmetry, for the rest of the tables or figures, we only provide results for $\delta > 0$ (ln OR > 0).

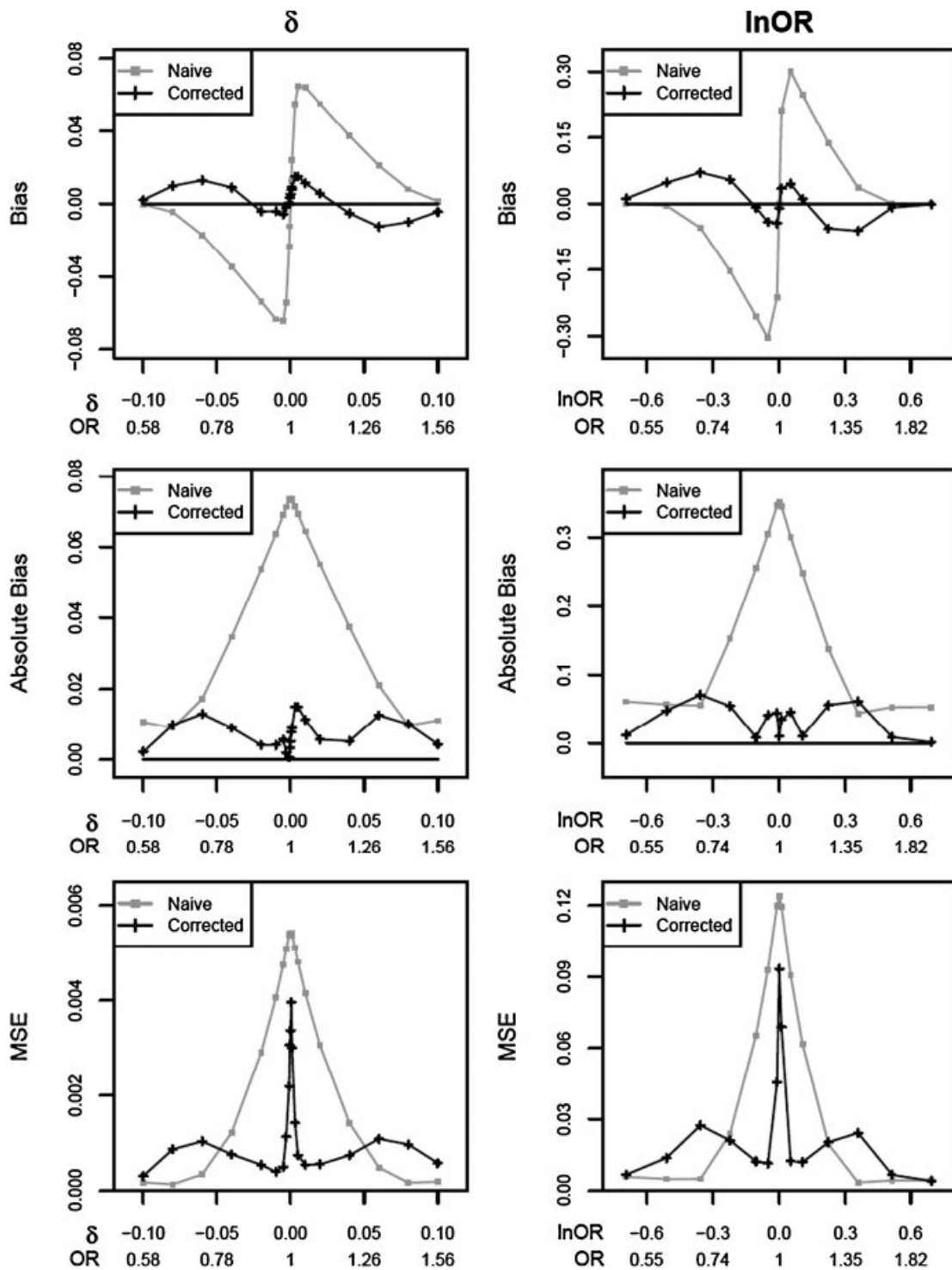**Fig. 1. Bias, absolute bias, and mean square error (MSE) for allele frequency difference δ and logarithm of odds ratio ln OR with sample size $N = 1,000$ and control allele frequency $p = 0.3$. Significance level $\alpha = 10^{-6}$.**

Given $N = 1,000$ cases and $N = 1,000$ controls, allele frequencies $p = 0.1$ and $p + \delta = 0.1258$ (OR = 1.295), and testing at significance level of $\alpha = 10^{-6}$ (resulting in power = 0.01), the expected value of the uncorrected estimator of δ is 0.0524 compared to the true value of 0.0258, a bias of 0.0266 and a proportional bias of 103%; similarly, the expected value of the uncorrected OR estimator is 1.699 compared to its true value of 1.295. In this case, a follow-up study designed to have 80% power at significance level $\alpha = 0.05$ would include 310

TABLE I. Proportional bias (%) of the uncorrected (naïve) and ascertainment-corrected MLEs of the allele frequency difference δ and odds ratio OR. Results are presented only for δ>0.

| p	N	Power	δ	$\frac{\hat{\delta}_{un}-\delta}{\delta}$	$\frac{\hat{\delta}_{as}-\delta}{\delta}$	OR	$\frac{\widehat{OR}_{un}-OR}{OR}$	$\frac{\widehat{OR}_{as}-OR}{OR}$
0.1	500	0.01	0.0376	101.9	−4.3	1.436	53.3	−1.8
		0.10	0.0541	47.9	−12.4	1.640	27.9	−5.7
		0.30	0.0665	26.2	−16.1	1.798	17.6	−7.2
		0.50	0.0752	16.2	−14.2	1.913	11.9	−6.5
		0.80	0.0898	6.0	−8.2	2.108	5.55	−1.7
	1,000	0.01	0.0258	103.1	−7.0	1.295	31.2	−1.9
		0.10	0.0370	48.1	−15.7	1.429	19.2	−4.2
		0.30	0.0453	26.5	−18.3	1.530	12.3	−7.1
		0.50	0.0512	16.0	−16.2	1.603	8.36	−6.3
		0.80	0.0609	6.1	−9.3	1.726	3.77	−4.1
0.5	500	0.01	0.0576	103.0	−9.0	1.260	27.2	−2.1
		0.10	0.0806	48.3	−16.8	1.384	17.3	−5.3
		0.30	0.0972	26.3	−18.3	1.483	11.2	−6.3
		0.50	0.1086	16.2	−14.3	1.555	7.72	−5.8
		0.80	0.1270	6.1	−9.2	1.681	3.51	−3.6
	1,000	0.01	0.0405	104.0	−13.1	1.176	18.5	−2.0
		0.10	0.0571	48.3	−18.6	1.258	11.8	−4.2
		0.30	0.0690	26.4	−18.3	1.320	7.7	−4.9
		0.50	0.0772	16.2	−15.0	1.365	5.4	−4.5
		0.80	0.0903	6.1	−10.4	1.441	2.4	−3.1

un: uncorrected; as: ascertainment-corrected; p: disease allele frequency in controls; N: sample size (number of cases and of controls). Assume testing at significance level $\alpha = 10^{-6}$.

cases and 310 controls, but would have actual power of only 30%.

We found that, for a fixed significance level α, the proportional bias in the uncorrected estimate of δ is solely a function of power, and is otherwise independent of sample size, allele frequency, or genetic model [Zöllner and Pritchard, 2007]. Consistent with intuition, proportional bias decreases as power increases (Fig. 2A), since the conditioning event becomes increasingly likely. At significance level $\alpha = 10^{-6}$, the uncorrected estimator of δ gives a proportional bias of ∼60% when power is 0.05 but is nearly unbiased when power is 95%. Interestingly, given fixed power, the proportional bias of the naïve estimator is consistently less when $\alpha = 10^{-6}$ than when $\alpha = 10^{-4}$ (data not shown).

We extended our analytical calculation to the uncorrected MLE of the odds ratio (Table I, Fig. 1), and observed the same general trend: substantial overestimation of the genetic effect given low to modest power to detect association and no bias given no association or sufficiently strong association. However, the proportional bias of the OR estimator, $\widehat{OR}_{un}$, cannot be explained by power alone, but depends on sample size, allele frequency, and genetic model (Fig. 2B). Interestingly, the proportional bias of the logarithm of the OR estimator, $\ln \widehat{OR}_{un}$, is a function of power, and follows a very similar pattern as the uncorrected MLE of the allele frequency difference δ.

Bias of the ascertainment-corrected MLEs. When we correct for ascertainment, the absolute bias of the MLE is substantially reduced (Fig. 1, Table I), and correction actually results in underestimation unless the genetic
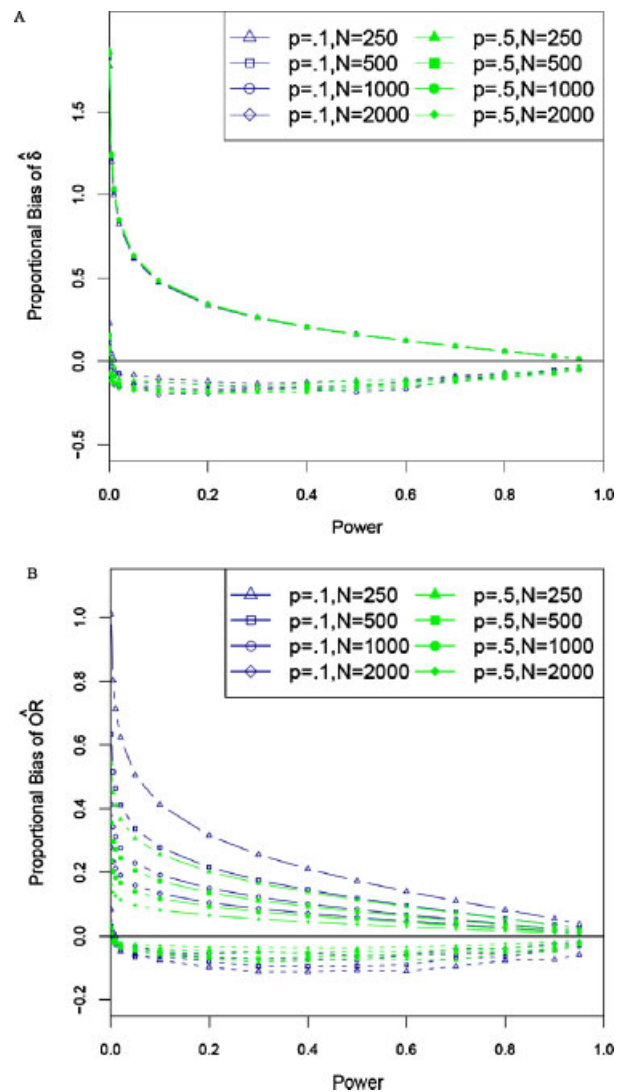
Fig. 2. Proportional bias versus power for the uncorrected (naïve) (solid lines) and corrected (dashed lines) estimators of the (A) allele frequency difference δ and (B) odds ratio OR. Significance level $\alpha = 10^{-6}$. Results are presented only for δ>0.

effect size is small or power is very low. For example, given $N = 1,000$ cases and $N = 1,000$ controls, allele frequencies $p = 0.1$ and $p+\delta = 0.1258$ (power = 0.01), and testing at significance level $\alpha = 10^{-6}$, the proportional bias of the corrected MLE of δ is −7%, compared to +103% before correction. In this case, a follow-up study designed to have 80% power at significance level $\alpha = 0.05$ would include 1,350 cases and 1,350 controls and have actual power 85%, whereas 1,150 cases and 1,150 controls actually would be sufficient to achieve 80% power. In the absence of association ($\delta = 0$), the corrected MLE is again nearly unbiased.

Reduction of the absolute bias is most pronounced when overestimation is most severe, and for fixed significance level α, bias reduction depends solely on study power. The relationship between power and proportional bias of the ascertainment-corrected MLE of δ is summarized in

Figure 2A. Although the corrected MLE $\hat{\delta}_{as}$ typically underestimates δ by 10–20% over the power range of 0.001–0.95 given testing at significance level of $\alpha = 10^{-6}$, the corrected MLE is considerably less biased than the uncorrected estimator unless power is high (typically > 60%). Even given high power, the magnitude of the bias of the ascertainment-corrected MLE $\hat{\delta}_{as}$ is not much greater than that of the uncorrected MLE $\hat{\delta}_{un}$, and it is of opposite sign. Interestingly, when power is greater than 0.1, the bias in the corrected MLE $\hat{\delta}_{as}$ decreases almost linearly as power increases (Fig. 2A).

The situation for the odds ratio is similar. With correction, the OR is typically underestimated by 5–10%, and this bias is in general smaller (although of opposite sign) than that for the uncorrected estimator for study powers ranging from 0.001 to 0.95 (Table I, Figs. 1 and 2B). Compared to the corrected MLE of δ whose proportional bias can be approximately summarized by power alone, the proportional bias for the corrected OR estimator does depends on sample size and allele frequency (Fig. 2B), while the proportional bias of the corrected estimator of the log odds ratio depends essentially on power alone and displays a very similar pattern as that of the corrected estimator for δ (Fig. 2A). Again, if we focus on the situations in which power < 60%, correction generally results in reduced absolute bias, and in many cases, absolute bias reduction is impressive. For example, given $N = 1,000$ cases and $N = 1,000$ controls, allele frequencies $p = 0.1$ and $p + \delta = 0.1258$ (OR = 1.295), and testing at significance level $\alpha = 10^{-6}$ (resulting in power = 0.01), the proportional bias of the corrected MLE of OR is −2%, compared to +31% before correction.

Standard errors and MSEs of the estimators. Table II summarizes the standard errors (SEs) for the MLEs of δ. We observed that the empirical SEs agree well with the asymptotic SEs for the corrected MLE, and both are two to six times greater than the SE of the uncorrected MLE which incorrectly ignores the fact of ascertainment. We also calculated the SE based on a random sample of the same sample size without ascertainment. All calculated SEs demonstrate that the genetic effect size estimates are quite variable in the settings described. The SEs of the corrected MLE are typically 1.5–2 times as large as those for an unascertained independent sample of the same size. This implies that while the ascertained sample is not as informative as a new random sample would be to estimate genetic effect size, the ascertained sample does provide 50–60% of the information in a new random sample, without the extra cost of collecting a new sample. We observed a very similar trend for SEs for the MLE of the odds ratio and its logarithm.

The MSE provides a measure of estimator quality that takes into account both bias and variance. Figure 1 displays the MSE for the naïve and corrected MLEs of δ and lnOR. In general, the naïve estimator has larger MSE than the ascertainment-corrected estimator unless the genetic effect size is sufficiently large to result in high power to detect association. In that case, biases for the two estimators are similar but the variance of the corrected estimator is larger than that of the naïve estimator (Table II).

TABLE II. Standard errors (SEs) for the uncorrected (naïve) and ascertainment-corrected MLEs of the allele frequency difference δ and for MLE obtained from an unascertained random sample. Results are presented only for δ > 0.

					SE			
p	N	Power	OR	δ	$\hat{\delta}_{un}$	$\hat{\delta}^{*\,a}_{as}$	$\hat{\delta}^{\dagger\,b}_{as}$	$\hat{\delta}_{rand}$
0.1	500	0.01	1.436	0.0376	0.0049	0.0263	0.0307	0.0142
		0.10	1.640	0.0541	0.0064	0.0291	0.0315	0.0150
		0.30	1.798	0.0665	0.0080	0.0304	0.0307	0.0148
		0.50	1.913	0.0752	0.0094	0.0309	0.0291	0.0153
		0.80	2.108	0.0898	0.0120	0.0291	0.0244	0.0154
	1,000	0.01	1.295	0.0258	0.0032	0.0179	0.0216	0.0099
		0.10	1.429	0.0370	0.0043	0.0200	0.0212	0.0103
		0.30	1.530	0.0453	0.0054	0.0216	0.0204	0.0099
		0.50	1.603	0.0512	0.0063	0.0218	0.0195	0.0107
		0.80	1.726	0.0609	0.0081	0.0195	0.0170	0.0102
0.5	500	0.01	1.260	0.0576	0.0069	0.0392	0.0433	0.0215
		0.10	1.384	0.0806	0.0091	0.0442	0.0460	0.0214
		0.30	1.483	0.0972	0.0114	0.0447	0.0439	0.0229
		0.50	1.555	0.1086	0.0133	0.0445	0.0409	0.0221
		0.80	1.681	0.1270	0.0168	0.0411	0.0348	0.0222
	1,000	0.01	1.176	0.0405	0.0049	0.0280	0.0319	0.0159
		0.10	1.258	0.0571	0.0065	0.0310	0.0320	0.0154
		0.30	1.320	0.0690	0.0081	0.0325	0.0305	0.0154
		0.50	1.365	0.0772	0.0095	0.0325	0.0286	0.0160
		0.80	1.441	0.0903	0.0120	0.0281	0.0247	0.0159

un: uncorrected; as: ascertainment-corrected; rand: random sample without ascertainment; p: disease allele frequency in controls; N: sample size (number of cases and of controls). Assume testing at significance level $\alpha = 10^{-6}$.
[a]Empirical.
[b]Asymptotic.

TWO-STAGE DESIGN

For both the allele frequency difference δ and the odds ratio OR, the naïve and ascertainment-corrected MLEs for optimal two-stage designs yield very similar results to those for the one-stage association designs described above (Fig. 3A). This is hardly surprising, since for optimal two-stage designs, statistical power is very close to that of the corresponding one-stage design in which all markers are genotyped on all samples, and power (approximately) determines proportional bias for δ and lnOR. Even for non-optimal two-stage designs, this continues to be true, except that the proportional bias of both the uncorrected and corrected estimators tends to increase modestly as π_{sample}, the fraction of the sample genotyped in stage 1, increases (Fig. 3B).

DISCUSSION

In genetic association studies, the genetic effect size for associated markers tends to be overestimated as a consequence of the winner's curse. This bias is due to the strong positive correlation between the association test statistic and the estimator of the genetic effect and the focus of investigators on markers that show statistically significant
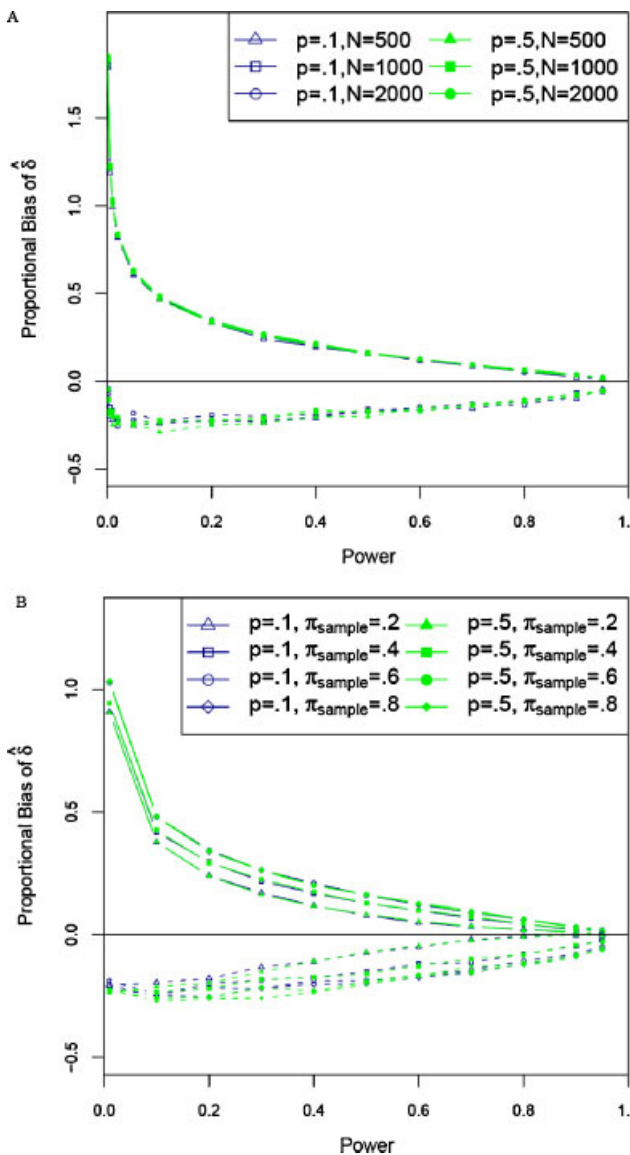
Fig. 3. Proportional bias versus power for the uncorrected (naïve) (solid lines) and corrected (dashed lines) estimators of the allele frequency difference δ for (A) optimal and (B) non-optimal two-stage designs. Significance level $\alpha = 10^{-6}$. Designs optimal for multiplicative disease model with disease prevalence 0.10, stage 2 to stage 1 genotype cost ratio 30. For non-optimal designs, $\pi_{marker} = 1\%$, and samples of $N = 1,000$ cases and $N = 1,000$ controls. Results are presented only for δ > 0.

evidence of association. In this paper, we studied the bias of the naïve MLEs for the allele frequency difference and the odds ratio that ignore this ascertainment; these measures are routinely used to estimate the strength of the effect in genetic association studies. We demonstrated that the proportional bias in the estimators decreases as power increases. Interestingly, at fixed significance level, the proportional biases of the allele frequency difference and the logarithm of odds ratio are functions of power, and otherwise are essentially independent of allele frequency or sample size (see also Zöllner and Pritchard, 2007).

We proposed a maximum likelihood method to correct for this ascertainment bias. The ascertainment-corrected MLEs for both the allele frequency difference and the (log) odds ratio are generally less biased than the uncorrected estimators unless study power is moderate to high (>60%). Since large-scale genetic association studies of complex traits typically are underpowered owing to small genetic effect sizes, our method should generally provide a more accurate estimate of genetic effect size in the context of genome-wide association studies and large-scale candidate gene studies. In high power situations, bias for both the naïve and corrected methods is small, so that ascertainment correction again is reasonable. Proportional bias of the corrected and uncorrected estimators for both the allele frequency difference and the odds ratio does show modest dependence on significance level α. For example, when the significance level $\alpha = 10^{-4}$, biases for all estimators are somewhat increased compared to the case of $\alpha = 10^{-6}$, and the advantage of ascertainment correction is increased slightly.

Zöllner and Pritchard [2007] used simulations to evaluate the impact of the winner's curse effect in genetic association studies and also proposed a maximum likelihood method to correct for it. Their method estimates the frequencies of all genotypes and corresponding penetrance parameters based on a known population prevalence of the disease under different inheritance models. In contrast, our method is simpler and focuses solely on the parameters of greatest interest: the allele frequency difference and odds ratio. This advantage of our method does require the assumption of Hardy-Weinberg equilibrium for our case and control samples. Such an assumption is entirely reasonable given the modest locus effect sizes for complex traits, but would not be reasonable in the context of a Mendelian major locus.

Our corrected MLEs for the allele frequency difference and odds ratio generally underestimate the true genetic effects [Zöllner and Pritchard, 2007]. Using computer simulation, we note that the empirical distribution of our corrected MLEs can reasonably be described as a two-component mixture, with one component near zero and the other appearing more nearly normal. Figure 4 illustrates this for the ascertainment-corrected estimator of the allele frequency difference. As power increases, the distribution becomes more nearly normal, and the asymptotic unbiasedness of the MLE comes into play.

We investigated the coverage of the asymptotic theory 95% confidence interval for the naïve and ascertainment-corrected MLEs for the allele frequency difference δ. The coverage of the ascertainment-corrected interval ranged from 82 to 100% for the cases we considered, reflecting the distribution and the bias of the ascertainment-corrected MLE, but still generally better than the coverage for the naïve estimator, which ranged from 0 to 92%.

Given the usual downward bias of our ascertainment-corrected estimators, one could consider an ad hoc bias correction. For the estimators of the allele frequency difference δ and the log odds ratio ln OR, the downward bias is 5–20% across the situations we considered (control allele frequency p=0.1–0.5, allele frequency difference δ = 0.018–0.159 (OR 1.11–2.30), case and control sample sizes 250–2,000, and statistical significance 10^{-4}–10^{-8}), so that multiplying the resulting estimate by 1.05–1.10 would generally reduce absolute bias. However, such an
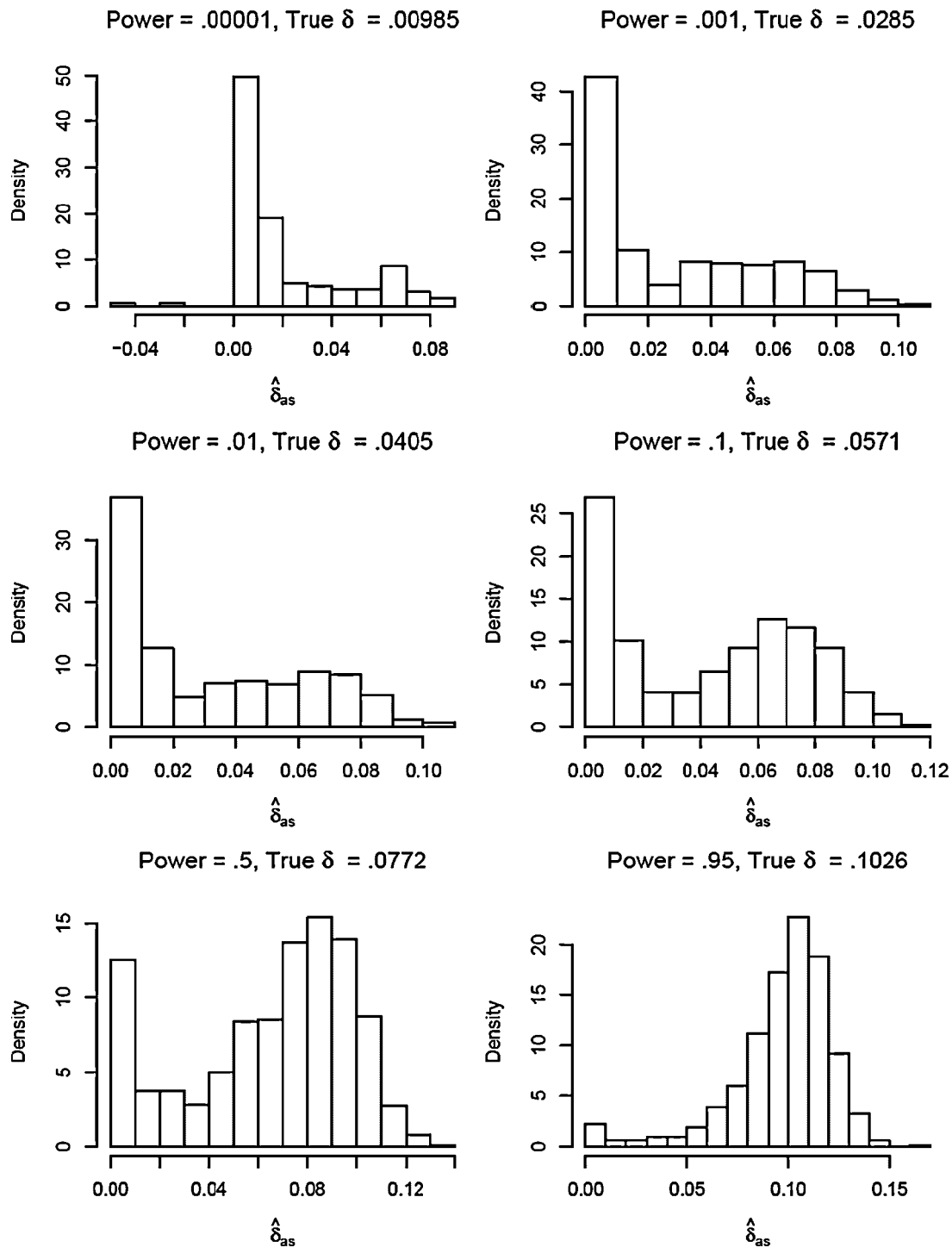
Fig. 4. Distribution of the ascertainment-corrected MLE of the allele frequency difference δ for different power levels. Results are presented only for δ > 0.

approach is counterproductive when power is very low (<0.005). The same criticism holds for taking a (weighted) average of the corrected and uncorrected estimators. More appealing might be to use an alternative estimation approach, and we currently are considering an empirical

Bayes method [Carlin and Louis, 2000] that uses information from genome-wide association studies to help define a prior distribution for the genetic effect size.

Realistically, precise and unbiased estimation of genetic effect size will best be obtained by collecting a large

sample specifically for this purpose, should resources be available to do so. However, given a sample in which an association is discovered, our ascertainment-corrected approach provides more accurate estimation of allele frequency difference and odds ratio than the naïve approach, and permits better design of subsequent replication studies or studies focused on estimating the population effect of the identified variant(s). Standard errors for the ascertainment-corrected MLEs were substantially larger than those for the naïve estimator based on an independent random sample of the same size, correctly reflecting the information loss for estimation based on a sample used for association detection.

In summary, we have presented analytic calculations that quantify the impact of the winner's curse in large-scale genetic association studies, and confirm that in realistic situations, it can result in substantial overestimation of the true genetic effect as measured by the case-control allele frequency difference or the corresponding odds ratio. We propose a MLE that corrects for the typical focus on statistically significant results, and demonstrate that this estimator results in reduced absolute bias compared to the naïve uncorrected estimator when study power is low or moderate ($<60\%$), a range that is typical for most large-scale genetic association studies, and similar absolute bias when power is high. Our method does not require specification of a genetic model and is easy to implement. We extended these calculations to two-stage association studies, and found similar results to those for one-stage studies. We recommend the use of this ascertainment-corrected method for estimation of genetic effect size in large-scale genetic association studies.

Software that carries out this analysis for case-controls data is available at http://csg.sph.umich.edu/boehnke/winner.

ACKNOWLEDGMENTS

REFERENCES

Allison DB, Fernandez JR, Heo M, Zhu S, Etzel C, Beasley TM, Amos CI. 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.

Bazerman MH, Samuelson WF. 1983. I won the auction but don't want the prize. J Conflict Resolut 27:618–634.

Carlin BP, Louis TA. 2000. Bayes and Empirical Bayes Methods for Data Analysis, 2nd edition. Boca Raton, FL: CRC Press.

Garner C. 2007. Upward bias in odds ratio estimates from genome-wide association studies. Genet Epidemiol 31: 288–295.

Ghosh A, Zou F, Wright FA. 2008. Estimating odds ratios in genome scans: an approximate conditional likelihood approach. Am J Hum Genet 82:1064–1074.

Göring HHH, Terwilliger JD, Blangero J. 2001. Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 69:1357–1369.

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. 2001. Replication validity of genetic association studies. Nat Genet 29:306–309.

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182.

Nelder JA, Mead R. 1965. A simplex method for function minimization. Computer J 7:308–313.

Rao CR. 1965. Linear Statistical Inference and its Applications. New York: Wiley.

Satagopan JM, Verbel DA, Venkatraman ES, Offit KE, Begg CB. 2002. Two-stage design for gene-disease association studies. Biometrics 58:163–170.

Siegmund D. 2002. Upward bias in estimation of genetic effects. Am J Hum Genet 71:1183–1188.

Skol AD, Scott LJ, Abecasis GR, Boehnke M. 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38: 209–213.

Skol AD, Scott LJ, Abecasis GR, Boehnke M. 2007. Optimal designs for two-stage genome-wide association studies. Genet Epidemiol 31:776–788.

Sun L, Bull SB. 2005. Reduction of selection bias in genome-wide studies by resampling. Genet Epidemiol 28:352–367.

Wu LY, Sun L, Bull SB. 2006. Locus-specific heritability estimation via bootstrap in linkage scans for quantitative trait loci. Hum Hered 62:84–96.

Yu K, Chatterjee N, Wheeler W, Li Q, Wang S, Rothman N, Wacholder S. 2007. Flexible design for following up positive findings. Am J Hum Genet 81:540–551.

Zhong H, Prentice RL. 2008. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. Biostatistics, in press.

Zöllner S, Pritchard JK. 2007. Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am J Hum Genet 80:605–615.

APPENDIX

Calculate the observed information matrix I for one-stage study:

$$I(p, \delta) = \begin{pmatrix} a & b \\ b & d \end{pmatrix},$$

where

$$a = -\frac{m_0}{p^2} - \frac{2N - m_0}{(1-p)^2} - \frac{m_1}{(p+\delta)^2}$$
$$\quad - \frac{2N - m_1}{(1-p-\delta)^2} + \frac{A^2}{F^2} + \frac{B}{F},$$
$$b = -\frac{m_1}{(p+\delta)^2} - \frac{2N - m_1}{(1-p-\delta)^2} + \frac{AD}{F^2} + \frac{C}{F},$$
$$d = -\frac{m_1}{(p+\delta)^2} - \frac{2N - m_1}{(1-p-\delta)^2} + \frac{D^2}{F^2} + \frac{E}{F},$$

where A, B, C, D, E, and F are calculated as follows:

$$A = \sum_{(x_0,x_1)\in I}\left\{\left(\frac{x_0}{p} - \frac{2N-x_0}{1-p} + \frac{x_1}{p+\delta} - \frac{2N-x_1}{1-p-\delta}\right)P(x_0,x_1)\right\},$$

$$B = \sum_{(x_0,x_1)\in I}\left\{\left(\frac{x_0}{p^2} + \frac{2N-x_0}{(1-p)^2} + \frac{x_1}{(p+\delta)^2} + \frac{2N-x_1}{(1-p-\delta)^2}\right.\right.$$
$$\left.\left. - \left(\frac{x_0}{p} - \frac{2N-x_0}{1-p} + \frac{x_1}{p+\delta} - \frac{2N-x_1}{1-p-\delta}\right)^2\right)P(x_0,x_1)\right\},$$

$$C = \sum_{(x_0,x_1)\in I}\left\{\left(\frac{x_1}{(p+\delta)^2} + \frac{2N-x_1}{(1-p-\delta)^2} - \left(\frac{x_0}{p} - \frac{2N-x_0}{1-p}\right)\right.\right.$$
$$\left.\left. \times \left(\frac{x_0}{p} - \frac{2N-x_0}{1-p} + \frac{x_1}{p+\delta} - \frac{2N-x_1}{1-p-\delta}\right)\right)P(x_0,x_1)\right\},$$

$$D = \sum_{(x_0,x_1)\in I}\left\{\left(\frac{x_1}{p+\delta} - \frac{2N-x_1}{1-p-\delta}\right)P(x_0,x_1)\right\},$$

$$E = \sum_{(x_0,x_1)\in I}\left\{\left(\frac{x_1}{(p+\delta)^2} + \frac{2N-x_1}{(1-p-\delta)^2}\right.\right.$$
$$\left.\left. - \left(\frac{x_1}{p+\delta} - \frac{2N-x_1}{1-p-\delta}\right)^2\right)P(x_0,x_1)\right\},$$

$$F = \sum_{(x_0,x_1)\in I} P(x_0,x_1),$$

where $I = \{(x_0,x_1) : X(x_0,x_1) > x_\alpha\}$ and $P(x_0,x_1)$ is calculated by formula (2) in the paper. Our calculation for the asymptotic SE for p and δ was based on the observed information matrix evaluated at $\hat{p}_{as}, \hat{\delta}_{as}$.