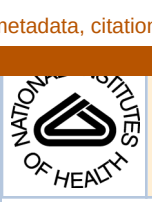

Published in final edited form as:

Hum Genet. 2010 December ; 128(6): 597–608. doi:10.1007/s00439-010-0880-x.

Using Public Control Genotype Data to Increase Power and Decrease Cost of Case-Control Genetic Association Studies

Lindsey A. Ho[1] and **Ethan M. Lange**[1,2]

[1] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

[2] Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Abstract

Genome-wide association (GWA) studies are a powerful approach for identifying novel genetic risk factors associated with human disease. A GWA study typically requires the inclusion of thousands of samples to have sufficient statistical power to detect single nucleotide polymorphisms (SNPs) that are associated with only modest increases in risk of disease given the heavy burden of a multiple test correction that is necessary to maintain valid statistical tests. Low statistical power and the high financial cost of performing a GWA study remains prohibitive for many scientific investigators anxious to perform such a study using their own samples. A number of remedies have been suggested to increase statistical power and decrease cost, including the utilization of free publicly available genotype data and multi-stage genotyping designs. Herein, we compare the statistical power and relative costs of alternative association study designs that use cases and screened controls to study designs that are based only on, or additionally include, free public control genotype data. We describe a novel replication-based two-stage study design, which uses free public control genotype data in the first stage and follow-up genotype data on case-matched controls in the second stage, that preserves many of the advantages inherent when using only an epidemiologically matched set of controls. Specifically, we show that our proposed two-stage design can substantially increase statistical power and decrease cost of performing a GWA study while controlling the type I error rate that can be inflated when using public controls due to differences in ancestry and batch genotype effects.

Keywords

Case-Control; Association Study; Genome-wide; Two-stage; Power

Introduction

Large-scale commercial genotyping platforms have facilitated the identification of numerous common single nucleotide polymorphisms (SNPs) that are associated with complex genetic diseases. The high cost of genome-wide association (GWA) studies has lead to the utilization of multi-stage study designs. Two-stage genotyping designs typically involve genotyping a fraction of the entire sample on a commercial genotyping platform containing all SNPs of interest in stage 1, performing systematic tests of association using stage 1 samples, and genotyping stage 2 samples on only the SNPs of greatest interest as determined

Address correspondence to: Ethan Lange, Ph.D. Department of Genetics, 5111 Genetic Medicine Building, University of North Carolina, Chapel Hill, North Carolina, 27599-7264, Phone (919) 966-3356, Fax (919) 843-4682, elange@med.unc.edu.

The original publication is available at www.springerlink.com.

in stage 1 (Satagopan et al., 2002). Two-stage genotyping designs have been shown to maintain power comparable to a single-stage study employing all samples while substantially decreasing overall genotyping costs (Kraft, 2006;Satagopan et al., 2002;Satagopan et al., 2004;Skol et al., 2006;Skol et al., 2007;Thomas et al., 2004;Wang et al., 2006). The data collected from the second stage of a two-stage GWA study is either analyzed separately as a replication-based sample or the data is combined with data from the first stage and the combined data is analyzed jointly. A recent alternative approach for reducing the cost of a large-scale case-control genetic association study and to increase the statistical power to detect an association when present is to use freely available genotype data on a large number of subjects from previous genome-wide association scans as control data in the current study. The effective use of a large public control dataset for comparison with multiple case datasets for different phenotypes was illustrated by the Wellcome Trust's Case Control Collaboration (WTCCC) GWA study on 14,000 cases of seven common diseases and 3,000 shared controls (Wellcome Trust Case Control Consortium, 2007). In this study, based on British subjects of European descent, the WTCCC identified 24 independent associations ($p < 5 \times 10^{-7}$) for bipolar disorder, coronary artery disease, Crohn's disease, rheumatoid arthritis, type 1 diabetes and type 2 diabetes using 2,000 independent cases for each disorder.

For investigators that have collected a well-matched group of cases and controls who wish to preserve many of the benefits of their sample collection design, we describe a replication-based two-stage case-control genetic association study design that uses free genotype data from public controls in stage 1, well-matched study controls in stage 2, and study cases distributed over stages 1 and 2. We compare the power and relative cost of our two-stage approach to single-stage approaches that strictly use either free public control genotype data or genotype data from study controls and to the single-stage approach that combines public and study controls. We discuss the advantages and limitations of each of the four sampling designs while considering the impact of ancestrally poorly-matched public controls and batch genotype effects. We show that the proposed replication-based two-stage design controls the overall type I error rate and has increased power over studies that exclude public controls.

Material and Methods

We assumed an investigator had a sample of N_A study cases, N_U study controls and access to free genotype data on N_{PU} public controls. We further assumed that study controls were screened for disease and that public controls were not screened for disease. We performed a series of calculations over a range of alternative models comparing the power achieved in an association study using four different sampling approaches: 1) a single-staged association study that used all N_A study cases and N_U study controls; 2) a single-staged association study that utilized all N_A study cases and N_{PU} public controls; 3) a single-staged association study that used all N_A cases and combined all N_U study and N_{PU} public controls; 4) a two-staged replication-based study that used all N_{PU} public controls in stage 1, all N_U study controls in stage 2 and all N_A cases apportioned between stages 1 and 2. We assumed an underlying multiplicative genetic mode-of-inheritance risk model for a bi-allelic locus with alleles D and d and corresponding allele frequencies of f_{D} and f_{d}, respectively. For each alternative model, we set the population frequency of the susceptibility allele D in the general population, the prevalence (K) of the disease in the population, and the locus specific genetic relative risk (GRR) = Pen(DD)/Pen(Dd) = Pen(Dd)/Pen(dd), where Pen(dd), Pen(Dd), and Pen(DD) were the penetrances for the dd, Dd, and DD genotypes, respectively. Consistent with many genetic power calculators, our power calculations are for the main effects of a directly genotyped locus and, as such, do not rely on additional assumptions regarding the extent of linkage disequilibrium between this locus and an untyped causal locus. All power

analyses were programmed into the freely available statistical software R version 2.4.1 (R Development Core Team, 2006).

Single-stage Power Calculations

Assuming Hardy-Weinberg equilibrium in the general population from which the cases and controls were selected, we used our model assumptions (allele frequencies, disease prevalence and GRR) to calculate the penetrance functions and we used Bayes' theorem to ascertain the conditional probability of each genotype given affection status, P_{ji}, where j indexes affection status and i = 0 (dd), 1 (Dd), 2 (DD) indexes genotype. Namely, for the cases these probabilities were P_{A0} = Pr(dd | case), P_{A1} = Pr(Dd | case), and P_{A2} = Pr(DD | case) and for the unaffected (screened) controls the probabilities were P_{U0} = Pr(dd | unaffected control), P_{U1} = Pr(Dd | unaffected control), P_{U2} = Pr(DD | unaffected control). We assumed no disease misclassification among study cases or screened study controls. Derivations of the conditional genotype probabilities are provided for the multiplicative model in the Supplementary materials. For public controls, the genotype probabilities were set to the genotype probabilities in the general population, namely $P_{PU0} = f_d^2$, $P_{PU1} = 2f_d f_D$, $P_{PU2} = f_D^2$, since affection status was not assumed to be known.

We calculated asymptotic power for the Cochran-Armitage trend test (Armitage, 1955;Cochran, 1954) by specifying the non-centrality parameter based on work by Chapman and Nam (Chapman and Nam, 1968) and we set the vector of scores to x = (0, 1, 2) for genotypes (dd, Dd, DD), respectively (Slager and Schaid, 2001). In particular, the non-centrality parameter, explicitly stated by Ahn et al. (Ahn et al., 2007), was

$$\lambda = N_A N_U \frac{\left[\sum_{i=0}^{2} x_i (P_{Ai} - P_{Ui})\right]^2}{\sum_{i=0}^{2} x_i^2 (N_A P_{Ai} + N_U P_{Ui}) - \left[\sum_{i=0}^{2} x_i (N_A P_{Ai} + N_U P_{Ui})\right]^2 / (N_A + N_U)}$$

where N_A and N_U (or optionally N_{PU}) were the sample sizes of the cases and screened (or public) controls, respectively, x_i was the score for the i-th genotype (i = 0, 1, 2 for genotypes dd, Dd, DD), and P_{Ai} and P_{Ui} were the probabilities of the i-th genotype for the cases and controls, respectively. Power was then taken to be $1 - \beta$, where β was the type II error of the non-central χ^2 distribution with 1 degree of freedom and non-centrality parameter λ, evaluated at the $100(1 - \alpha_{Bonferroni})$ percentile of the central χ^2 distribution with 1 degree of freedom. For single-stage designs, the overall family-wise error rate was set to $\alpha = 0.05$ by using a Bonferroni corrected significance threshold $\alpha_{Bonferroni} = 0.05/M$, where M was the number of markers evaluated.

Replication-Based Two-stage Power Calculations

Using the formulas described above for one-stage power, we calculated power for a replication-based two-stage design. For a replication-based two-stage design, the overall power for a SNP was simply calculated as the product of the power for the first stage times the power of the second stage. Following the notation in Skol et al. (Skol et al., 2006), the power for the first-stage was calculated using a significance threshold defined as the proportion of markers followed in stage 2, $\pi_{markers}$. Power for the second-stage was calculated using a significance threshold equal to $2\alpha/(M\,\pi_{markers})$, i.e. the Bonferroni corrected cutoff for a one-sided test that requires the direction of the SNP main effect to be the same in stage 1 and 2 samples.

We restricted the number of SNPs for follow-up analysis in stage 2 to be values that approximate numbers that would typically be considered given today's currently available commercial genotyping platforms. Namely, we considered follow-up platforms of size 100, 1,500, 7,500, and 16,500 SNPs. For each follow-up genotyping platform, we found the optimal proportion of cases, π_{cases}, to be genotyped in stage 1 that optimized the power of the two-stage design. Specifically, we used the "optimize" function in R to search for the maximum power in the continuous space of π_{cases}. This method combines the golden section search and successive parabolic interpolation algorithms.

Examples of Power Approximations for 1- and 2-Stage Designs

We calculated power for a GWA scan on M = 500,000 SNPs based on a study sample of N_A = 2,000 study cases and N_U = 2,000 study controls to demonstrate the difference in power between the competing approaches. We assumed a multiplicative model with a GRR = 1.3, and a susceptibility allele frequency f_D = 0.3 in the general population. We considered a wide range of disease prevalence values of $K = 1\times10^{-4}$, 0.01, 0.05, 0.1, 0.25, and 0.5 and we assumed available genotype data on samples of N_{PU} = 2,000, 5,000 and 10,000 public controls. We calculated power for the single-stage designs (using only study controls, only public controls, or both control samples combined) and for the optimal replication-based two-stage design. For each optimal two-stage model we provided the power estimate from the follow-up platform that provided the greatest power. Finally, in order to test how power of the 2-stage designs for the four proposed follow-up platforms were impacted by different combinations of susceptibility allele frequency and GRR, we calculated power with K =0.10 (assuming N_{PU} = 5,000) using susceptibility allele frequencies of f_D = 0.1, 0.3 and 0.5 and GRRs ranging from 1.2 to 1.5. Additional power calculations for other study designs using the Cochran-Armitage trend test, the general two-degree of-freedom test of association, and both dominant and recessive models are provided in the Supplementary material.

Impact on Power of Ancestrally Poorly-Matched Public Controls and Batch Genotype Effects

Ancestrally poorly-matched public controls and batch genotype effects that can occur when genotyping study samples and public controls from different populations at different times can have detrimental effects on power and type I error. We evaluated the impact of these factors for a study design that included 2,000 study cases, 2,000 study controls, and 5,000 public controls for a multiplicative disease model with susceptibility allele frequency f_D = 0.3, K = 0.10 and GRR = 1.3 (Model 1).

We assessed the impact of ancestrally poorly-matched public controls on power by simulating two genetically-admixed populations (our study population and the public control population). We assumed our study population consisted of individuals derived from two ancestral populations (*POP1* and *POP2*), with subjects having a mean proportion of *POP1* equal to 0.25 and standard deviation equal to 0.15. Furthermore, we assumed that study cases had a mean proportion of *POP1* ancestry equal to 0.221 and study controls had a mean proportion of *POP1* ancestry equal to 0.253. These values are consistent with the estimated proportion of European ancestry in African American prostate cancer cases and controls (Haiman et al., 2007). For public controls, we allowed the mean proportion of *POP1* ancestry to vary between 0.10 and 0.40 and assumed a fixed standard deviation of 0.15. We maintained the overall frequency of the susceptibility allele to be f_D = 0.30 in the study population, but varied f_D between 0.00 and 0.60 in *POP1* (in both the study and public control populations). Finally, we assumed that any extreme outliers, such as individuals with misreported ethnicity, would be identified and removed prior to testing for association. Simulated data sets (n = 10,000) were analyzed using logistic regression models, with covariate adjustment for the proportion of *POP1* descent of each subject to control for

population stratification as would be routinely done in a GWA study using analytic methods such as principal components, for the one- and two-stage designs under the null model (GRR = 1.0) and the alternative model described above.

Modern commercial genotyping platforms have increasingly high accuracy in genotype calling, but small systematic biases in genotyping calling when genotyping cases and controls at different times or on different platforms can create artificial differences in genotype frequencies between them. Careful study design can alleviate these concerns when genotyping study cases and controls at the same time, but using public controls will always be a concern. We assessed the impact of batch genotype effects on power for the one- and two-stage designs. We assumed that batch genotype effects would result in systematic and over or under calling of the susceptibility allele on an allele-by-allele basis (i.e. the probability an allele in a genotype being miscalled was assumed to be independent of the calling for the other allele). We additionally assumed that all subjects genotyped at the same time were subject to the same batch effect, effecting any cases and controls genotyped together equally. Over or under calling of the susceptibility allele was allowed to occur with different probabilities for each genotyping platform (e.g. for our proposed two-stage design there were three different genotyping platforms where the susceptibility allele could either be over or under called – the platform used for public controls, the genome-wide panel for a subset of the cases, and the follow-up platform for the remaining cases and all study controls). Genotype batch effects were modeled by modifying the genotype probabilities in our power calculator. For example, under Model 1 the probabilities of the three genotypes for public controls ($f_D = 0.30$) when there were no genotype batch effects were set to 0.09, 0.42 and 0.49 for the DD, Dd and dd genotypes, respectively. If the batch effects resulted in the susceptibility allele D systematically being over-called with error probability + 0.01 (i.e. the non-susceptibility allele d is erroneously reported as the susceptibility allele D with probability 0.01), then the probabilities for the three genotypes DD, Dd and dd were set to 0.094249 $[0.09 + 0.01(0.42) + (0.01)^2(0.49)]$, 0.425502 $[(0.99)(0.42) + 2(0.99)(0.01)(0.49)]$ and 0.480249 $[(0.99)^2(0.49)]$, respectively. This calculation takes into consideration that subjects with true genotype dd could be mistakenly scored Dd (with probability = $2\times0.99\times0.01$) or DD (with probability $(0.01)^2$) and subjects with true genotype Dd could be mistakenly scored DD (with probability 0.01). We did not assume any additional random error in our calculations (in the above example, subjects with true genotype DD were assumed to be scored DD with probability = 1). We considered systematic allele calling error probabilities for a given genotype platform of + 0.01, 0.00, and − 0.01, where + (−) corresponds to erroneously over (under) calling of the susceptibility allele.

The impact of batch genotype effects on power were evaluated under the null hypothesis of no association and under the alternative model described above for the one and two-stage designs. These calculations were performed under the assumption that either the specific SNP under consideration was the only SNP subject to batch effects or that there existed a systematic bias due to batch effects that impacted all SNPs. Under the former scenario, we used the same significance thresholds described previously for the one- and two-stage studies. For the latter scenario, we assumed a baseline 10% mean systematic inflation of the chi-square test statistics across the genome (mean value test statistic, $\mu=1.10$) for all SNPs evaluated in the one-stage study design based only on public controls. We assumed no inflation ($\mu=1.00$) for the test statistics of the one-stage study design based only on study controls and for the test statistics from stage 2 analyses of the replication-based two-stage study designs (because cases and controls would be genotyped at the same time in stage 2). The mean systematic inflation of the test statistics, $\mu - 1$, is equal to $\lambda \sim Np(1-p)\Delta^2$ (i.e., the non-centrality parameter of a chi-square test with 1 df), where N is the total sample size, p is the proportion of cases in the sample and Δ is the metric reflecting the difference in genotype frequencies between cases and controls due to batch effects. Consequently, the

magnitude of the systematic inflation of the test statistics does not impact all study designs equally. Hence, we recalculated the mean inflation of the test statistics for the one-stage design with both study and public controls and for stage 1 of each replication-based two-stage study design (based on the number of cases included in stage 1). For the one-stage studies that include public controls, we calculated power after correcting for the systematic batch effects across all SNPs by multiplying the critical value of the 1 df chi-square distribution corresponding to $p = 1.0\times10^{-7}$ (i.e. 28.373) by the mean test statistic value, μ (Reich and Goldstein, 2001). For each two-stage study, the stage 1 test-statistic critical value was multiplied by the corresponding mean test statistic value, μ, determined by the stage 1 sample composition.

Example of Genotyping Costs for Different Genotype Sampling Strategies

To understand the financial impact of the different genotyping sampling strategies, we estimated the relative total experimental cost of each genotype sampling design for a GWA study based on M = 500,000 SNPs using N_A = 2,000 study cases, N_U = 2,000 screened study controls and N_{PU} = 5,000 public controls. We assumed a multiplicative trait with a prevalence K = 0.1, GRR = 1.3 and f_D = 0.3 (Model 1). We calculated the relative total costs of performing the three single-stage studies that used either study or public controls or both. For these single-stage sampling designs, all study samples were assumed to be genotyped on all 500,000 SNPs; genotype data for public controls were assumed to be available at no expense. In addition, we calculated the relative total cost of the optimal (highest power) replication-based two-stage study design for each follow-up platform. For the purpose of our calculations, we assumed the Illumina Human660W-Quad platform would be used for genotyping 500,000 viable SNPs in stage 1 and Illumina's GoldenGate 96, 384 and 1,536 SNP panels and Illumina's Custom iSelect Infinium 7,600 and 16,720 SNP panels would be used as the follow-up platforms for stage 2. Given that genotyping costs are constantly changing, rather than using dollar amounts, we report the relative total cost of genotyping all study subjects based on current prices. Using the cost of genotyping 500,000 SNPs for an individual sample on a genome-wide panel as a baseline, the relative total cost of genotyping 16,000, 7,500, 1,500, and 100 SNPs for that sample were assumed to be 1/2, 1/3, 1/5, 1/10 and 1/12 of the cost, respectively, based on the most recent genotype prices at the CIDR genotyping facility (www.cidr.jhmi.edu/pricing.pdf).

Skol et al. (Skol et al., 2006) demonstrated that a joint-analysis two-stage study design could effectively achieve equivalent power to a single-stage study for a fraction of the cost. Consequently, for the three single-stage sampling designs, we also estimated the relative cost of performing a joint-analysis two-stage association study for each follow-up platform. For each combination of sampling design and follow-up platform, we performed a series of simulations to identify the least expensive joint-analysis two-stage sampling design that obtained an estimated power within 0.01 of the power obtained from the corresponding single-stage study. For the sampling design that used only public controls, cases were divided between stages 1 and 2 while all public controls were assumed to be available in stage 1. For the sampling design that included both study and public controls, all study controls were assumed to be genotyped in stage 2, and all public controls were assumed to be available in stage 1. Cases were divided between stages 1 and 2.

Results

We performed power calculations for a range of study designs and disease models. Power is described for the frequency of the risk allele in the general population (the frequency of the risk allele in cases and study controls for different values of K are provided in the table footnotes). Not surprisingly, our results showed that including free genotype data from public controls increases statistical power over studies that do not include these data (Table

1). The single-stage study with both public and study controls noticeably outperformed the replication-based two-stage study using the same samples. Power for the proposed replication-based two-stage design was typically greater than the power of the one-stage design based only on study controls. Overall, the same general patterns of results were observed when varying GRR and frequency of the disease susceptibility allele (Supplementary Figure 1), when analyzing the genotype data using a general (co-dominant) 2-df inheritance model (Supplementary Table 2), and when considering dominant or recessive genetic inheritance models (Supplementary Tables 3 and 4, respectively).

For the replication-based two-stage design, we observed that the optimal choice of the proportion of cases, π_{cases}, to be genotyped in stage 1 varied considerably between the different choices of stage 2 genotyping platforms (as expected, a larger proportion of cases were necessary to be genotyped in stage 1 for the smaller follow-up platforms) but, importantly, varied little within a given platform across the considered range of GRRs and disease allele frequencies (Table 2). We note that for a given follow-up platform, the optimal choice of π_{cases} was also insensitive to analytic strategy (i.e. similar optimal values of π_{cases} were observed for the general 2-df test as for the trend test) (Supplementary Table 5) and genetic inheritance model (i.e., similar optimal values of π_{cases} were also observed for the dominant and recessive models) (Supplementary Tables 6 and 7). These results suggest that it is reasonable to use the proportion of cases, π_{cases}, to be genotyped in stage 1 that optimizes power for a replication-based two-stage study design based on a single specific alternative model and expect that power should be near optimized by this choice of π_{cases} across a range of alternative genetic models when using the same follow-up platform.

Ancestrally Poorly-Matched Public Controls and Batch Genotype Effects

All study designs maintained power near levels obtained under negligible population stratification when the proportion of *POP1* ancestry in public controls was within 0.05 of the proportion of *POP1* ancestry in the study population regardless of the frequencies of the disease-susceptibility allele in the *POP1* and *POP2* ancestral populations (Table 3). Power for the one-stage design that only included public controls dropped noticeably when the proportion of *POP1* ancestry in public controls was either 0.10 or 0.40. Under these two scenarios, the greatest decline in power was observed when the susceptibility allele was more (for proportion *POP1* = 0.10) or less (for proportion *POP1* = 0.40) frequent in the *POP1* ancestral population. Power for the two-stage design with follow-up on 100 of the best SNPs dropped when the proportion of *POP1* ancestry in public controls was equal to 0.40. Power for the one-stage design with both public and study controls and the remaining two-stage designs stayed relatively stable across the range of considered proportions of *POP1* ancestry in public controls.

Systematic genotyping errors (over or under calling of the susceptibility allele) decreased power modestly for the one-stage design with only study controls (Table 4). For the one-stage designs that included public controls, genotyping errors in opposite directions on the two genotyping platforms (e.g. over calling susceptibility allele in public controls and under calling susceptibility allele in study samples) had a major impact on power. For the single-stage study with only public controls, in the absence of batch effects (for both the individual SNP and for all other SNPs across the genome) the power was 0.90. Over calling the susceptibility allele in the public controls and under calling the susceptibility allele in cases (each with probability of 0.01 per allele) resulted in power decreasing to 0.53. Conversely, if the susceptibility allele was under called in the public controls and over called in the cases then power increased to 0.99. A similar pattern, but less dramatic differences, were observed for the single-stage study design with both public and study controls. Both single-stage studies that include public controls experienced some loss in power when accounting for the mean systematic inflation in the test statistics for all other SNPs across the genome due to

batch genotype effects. Regardless of batch effects, the single-stage study that includes both public and study controls always had greater power than the single-stage study with only study controls.

Batch effects also impacted the power of the two-stage studies (Table 5). The two-stage studies that were based on the larger follow-up platforms were less impacted by batch effects than the studies that were based on the smaller follow-up platforms (the range in power for the 16500 platform was 0.74–0.87 while the range for the 100 platform was 0.70–0.95, before factoring in the impact of the mean systematic inflation of test statistics across the genome). Accounting for systematic batch effects on other SNPs across the genome (which increased the significance threshold in stage 1 required for inclusion of the SNP into stage 2) had near negligible impact on power for the larger follow-up platforms but resulted in small declines in power for the smaller follow-up platforms. Even after factoring in the impact of batch genotype effects, the two-stage study designs typically outperformed the one-stage study design based only on study controls.

Cost Savings Including Public Controls

In addition to increased power, in Table 6 we illustrate that substantial cost savings can be achieved for a GWA study when including public controls. We compared the relative cost of one- and two-stage study designs that include study controls, public controls or both. We required the power of the joint-analysis two-stage study designs to be within 0.01 of the corresponding one-stage designs. As expected, the most expensive study designs were the one-stage study designs that genotyped all samples (excluding public controls – which provide genotype data at no expense) on all SNPs. Significant cost savings were observed when using the joint-analysis two-stage design described by Skol et al. (Skol et al., 2006). For example, when utilizing the joint-analysis two-stage design following-up the top 1,500 SNPs (corresponding to the 1,536 SNP Illumina GoldenGate custom panel) in stage 2, a 36%, 44% and 60% cost savings was achieved relative to the corresponding one-stage designs that included only study controls, only public controls and both public and study controls, respectively. The total cost of our proposed replication-based two-stage design was consistently less than the joint-analysis two-stage design that included both public and study controls, though the latter design maintained greater power. The joint-analysis two-stage design with only public controls was the least expensive design and had modestly greater power than the replication-based two-stage design for most follow-up platforms. In addition to having the lowest power among two-stage designs, the joint-analysis two-stage design that included only study controls was substantially more expensive than any other two-stage sampling design.

Discussion

We have performed a series of calculations to evaluate the statistical power of alternative study designs that either includes public controls, study controls or both. We also describe a novel replication-based two-stage design that uses freely available public control data in stage 1, study controls in stage 2 and study cases genotyped in stages 1 and 2. For each study design, we assessed the impact of both systematic ancestry differences between public controls and study samples and batch genotype effects that could occur due to genotyping public controls and study samples at different times on different genotyping platforms. Not surprisingly, the single-stage study design with both public and study controls had the greatest power under all circumstances considered. These results are entirely consistent with previous reports that have shown the negative effects of disease misclassification on power can be overcome by a using large number of unscreened controls (Edwards et al., 2005;Moskvina et al., 2005;Wellcome Trust Case Control Consortium, 2007;Zheng and Tian, 2005). While the single-stage study using only public controls generally had good

power when the number of available public controls was large, we noted that inclusion of the study controls protected power when there were strong differences in ancestry between public controls and study cases or when there were relatively strong batch genotype effects present. Under most circumstances, the proposed replication-based two-stage study had greater power than the single-stage study with only study controls and, depending on the circumstance, greater or lesser power than the single-stage study based only on public controls.

Clearly the greatest cause for concern when using public control genotype data is that observed allele frequency differences between public controls and study cases may be the consequence of systematic bias due to population stratification or batch genotype effects from differential allele calling between the two samples (Moskvina et al., 2006;Neale and Purcell, 2008). Greater differences in background ancestry will likely occur between public controls and study cases than between study cases and a carefully selected set of study controls from the same community. The impact of population stratification can be largely remedied by employing appropriate analytic methods (Price et al., 2006;Roeder and Luca, 2009;Yu et al., 2008), though these methods may not adequately alleviate biased results for a relatively small number of genetic markers under strong selective pressure such as those witnessed by the WTCCC study, that found highly significant differences in allele frequencies for a small number of loci between individuals of Caucasian descent from different communities in Great Britain (Wellcome Trust Case Control Consortium, 2007). Results from several GWA studies that have included public control genotype data on Caucasian samples have not revealed strong systematic differences in allele frequencies between previously genotyped public controls and study samples (Hom et al., 2008;Luca et al., 2008;Silverberg et al., 2009;Wrensch et al., 2009;Yu et al., 2008). However, results from a recent study that used public controls have raised concerns about the impact of batch genotype effects when cases and controls are genotyped on different platforms (Sebastiani et al., 2010). In our examples, modest systematic differences in ancestry between public and study samples had little impact on power (Table 3) or type I error (data not shown) when the estimated proportion of ancestry for each subject was included as a covariate in the model. These results are consistent with a recent report advocating the use of public controls (Zhuang et al., 2010). Batch genotype effects, before and after accounting for systematic batch effects across all SNPs, had either a negative or positive impact on power (Tables 4 and 5) and resulted in increased type I error rates for the SNP under consideration (Supplementary Table 8). Still, the impact of batch genotype effects on the overall family-wise error rate was small after applying the Genomic Control method (Reich and Goldstein, 2001) to account for the systematic inflation.

It is important to note that our examples did not include a small number of SNPs under strong selective pressure or SNPs with extreme batch effects. The impact of including these types of SNPs on overall statistical power would be minimal because, *a priori*, the probability that a SNP under the alternative model is under such pressure is likely small. However, a small number of SNPs under selective pressure or with extreme batch effects would substantially inflate the family-wise error rate for the GWA study and this inflation likely cannot be controlled by analytic methods. While a single-stage or a joint-analysis two-stage study that includes both public and study controls provides the greatest power, there is an increased possibility that any given significant result could be due to population stratification or extreme batch effects. In contrast, the proposed replication-based two-stage study maintains similar control of the overall type I error rate compared to a study based only on study controls, making any single significant result more reliable. For example, suppose that a single SNP, under the null hypothesis, with minor allele frequency of 0.30 has the major allele mistakenly scored the minor allele with probability = 0.15 in public controls (increasing the frequency of the minor allele in this public control sample to 0.405).

In a single-stage study design with 2,000 study cases, 2,000 study controls and 5,000 public controls, this SNP would be significantly associated with the outcome with probability near 1 under the null of no true association between the SNP and outcome, hence the family-wise error rate of the experiment would also be 1. In contrast, in the proposed replication-based two-stage study the SNP would almost certainly be included in stage 2 but the overall experimental type I error would be well controlled because the allele frequencies of the SNP in study controls and remaining cases in stage 2 are unaffected by the batch effects in the public controls in stage 1.

When considering which study design to use when including public controls, investigators should consider the trade-offs between increased power and increased false positives when choosing how to include public control genotype data in their study. Investigators should also consider that reporting a large number of false positives could impact the power of future replication studies due to the increased multiple test burden from following up a larger number of variants. Study designs that include public controls in a single-stage or joint-analysis two-stage design have more power than the proposed replication-based two-stage design, however, these designs are also more susceptible to increased type I error rates. The recent report by Sebastiani et al. (Sebastiani et al., 2010), that found evidence for associations between 150 genetic variants and longevity based on a case-control sample where cases and controls were disproportionately genotyped on different genotyping platforms, highlights the potential severe impact of batch genotype effects. Our proposed replication-based two-stage design is designed to protect the overall type I error of the experiment while still increasing power and decreasing study costs compared to studies that exclude public controls.

Finally, we have performed power calculations assuming fixed sample sizes rather than fixed costs. As we have shown (Table 6), the inclusion of public control genotype data can dramatically decrease the cost of genotyping in addition to increasing statistical power. With this in mind, the increase in power between approaches that include public control data compared to those that do not is even greater than what we have presented in situations where the sample size of study subjects is limited by costs and not sample availability. Optimizing power with respect to cost for each study design would require an iterative application of the methods we have described. We have R software code that is available for investigators who would like to calculate power and make the comparisons for their own studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health grants CA120082 and CA136621. We would like to express our appreciation to Yunfei Wang for his assistance in programming and three anonymous reviewers for their helpful suggestions. Please site the manuscript: Ho LA, Lange EM. Using public control genotype data to increase power and decrease cost of case-control genetic association studies. Human Genetics 2010 Dec;128(6): 597–608.

Reference List

Ahn K, Haynes C, Kim W, Fleur RS, Gordon D, Finch SJ. The effects of SNP genotyping errors on the power of the Cochran-Armitage linear trend test for case/control association studies. Ann Hum Genet. 2007; 71:249–261. [PubMed: 17096677]

Armitage P. Tests for Linear Trends in Proportions and Frequencies. Biometrics. 1955; 11:375–386.

Chapman DG, Nam JM. Asymptotic power of chi square tests for linear trends in proportions. Biometrics. 1968; 24:315–327. [PubMed: 5683872]

Cochran WG. Some Methods for Strengthening the Common Chi-Squared Tests. Biometrics. 1954; 10:417–451.

Edwards BJ, Haynes C, Levenstien MA, Finch SJ, Gordon D. Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. BMC Genet. 2005; 6:18. [PubMed: 15819990]

Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le ML, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet. 2007; 39:638–644. [PubMed: 17401364]

Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapaa-Dahlqvist S, Petri M, Manzi S, Seldin MF, Ronnblom L, Syvanen AC, Criswell LA, Gregersen PK, Behrens TW. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med. 2008; 358:900–909. [PubMed: 18204098]

Kraft P. Efficient two-stage genome-wide association designs based on false positive report probabilities. Pac Symp Biocomput. 2006:523–534. [PubMed: 17094266]

Luca D, Ringquist S, Klei L, Lee AB, Gieger C, Wichmann HE, Schreiber S, Krawczak M, Lu Y, Styche A, Devlin B, Roeder K, Trucco M. On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. Am J Hum Genet. 2008; 82:453–463. [PubMed: 18252225]

Moskvina V, Craddock N, Holmans P, Owen MJ, O'Donovan MC. Effects of differential genotyping error rate on the type I error probability of case-control studies. Hum Hered. 2006; 61:55–64. [PubMed: 16612103]

Moskvina V, Holmans P, Schmidt KM, Craddock N. Design of case-controls studies with unscreened controls. Ann Hum Genet. 2005; 69:566–576. [PubMed: 16138915]

Neale BM, Purcell S. The positives, protocols, and perils of genome-wide association. Am J Med Genet B Neuropsychiatr Genet. 2008

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: 2006.

Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol. 2001; 20:4–16. [PubMed: 11119293]

Roeder K, Luca D. Searching for disease susceptibility variants in structured populations. Genomics. 2009; 93:1–4. [PubMed: 18519160]

Satagopan JM, Venkatraman ES, Begg CB. Two-stage designs for gene-disease association studies with sample size constraints. Biometrics. 2004; 60:589–597. [PubMed: 15339280]

Satagopan JM, Verbel DA, Venkatraman ES, Offit KE, Begg CB. Two-stage designs for gene-disease association studies. Biometrics. 2002; 58:163–170. [PubMed: 11890312]

Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, Achkar JP, Goyette P, Scott R, Xu W, Barmada MM, Klei L, Daly MJ, Abraham C, Bayless TM, Bossa F, Griffiths AM, Ippoliti AF, Lahaie RG, Latiano A, Pare P, Proctor DD, Regueiro MD, Steinhart AH, Targan SR, Schumm LP, Kistner EO, Lee AT, Gregersen PK, Rotter JI, Brant SR, Taylor KD, Roeder K, Duerr RH. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet. 2009; 41:216–220. [PubMed: 19122664]

Sebastiani P, Solovieff N, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB, Myers RH, Steinberg MH, Montano M, Baldwin CT, Perls TT. Genetic signatures of exceptional longevity in humans. Science. 2010 (in press).

Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006; 38:209–213. [PubMed: 16415888]

Skol AD, Scott LJ, Abecasis GR, Boehnke M. Optimal designs for two-stage genome-wide association studies. Genet Epidemiol. 2007; 31:776–788. [PubMed: 17549752]

Slager SL, Schaid DJ. Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. Hum Hered. 2001; 52:149–153. [PubMed: 11588398]

Thomas D, Xie R, Gebregziabher M. Two-Stage sampling designs for gene association studies. Genet Epidemiol. 2004; 27:401–414. [PubMed: 15543639]

Wang H, Thomas DC, Pe'er I, Stram DO. Optimal two-stage genotyping designs for genome-wide association scans. Genet Epidemiol. 2006; 30:356–368. [PubMed: 16607626]

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]

Wrensch M, Jenkins RB, Chang JS, Yeh RF, Xiao Y, Decker PA, Ballman KV, Berger M, Buckner JC, Chang S, Giannini C, Halder C, Kollmeyer TM, Kosel ML, LaChance DH, McCoy L, O'Neill BP, Patoka J, Pico AR, Prados M, Quesenberry C, Rice T, Rynearson AL, Smirnov I, Tihan T, Wiemels J, Yang P, Wiencke JK. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. Nat Genet. 2009; 41:905–908. [PubMed: 19578366]

Yu K, Wang Z, Li Q, Wacholder S, Hunter DJ, Hoover RN, Chanock S, Thomas G. Population substructure and control selection in genome-wide association studies. PLoS ONE. 2008; 3:e2551. [PubMed: 18596976]

Zheng G, Tian X. The impact of diagnostic error on testing genetic association in case-control studies. Stat Med. 2005; 24:869–882. [PubMed: 15568195]

Zhuang JJ, Zondervan K, Nyberg F, Harbron C, Jawaid A, Cardon LR, Barratt BJ, Morris AP. Optimizing the power of genome-wide association studies by using publicly available reference samples to expand the control group. Genet Epidemiol. 2010

Table 1

Power of the Cochran-Armitage trend test for 1- and 2-stage study designs across a range of sample sizes, SNPs in stage 1, and disease prevalences. Study controls are assumed to be screened and disease free.

		Number of Public Controls								
		2,000			5,000			10,000		
K[a]	Study Controls Only	Public Controls Only	Study + Public Controls	Two-Stage[b]	Public Controls Only	Study + Public Controls	Two-Stage[b]	Public Controls Only	Study + Public Controls	Two-Stage[b]
2,000 Cases and 2,000 Screened Study Controls/500,000 SNPs										
0.0001	0.57	0.56	0.85	0.71	0.90	0.94	0.82	0.97	0.97	0.84
0.01	0.59	0.56	0.86	0.72	0.90	0.94	0.82	0.97	0.97	0.85
0.05	0.68	0.56	0.89	0.76	0.90	0.96	0.86	0.97	0.98	0.88
0.1	0.79	0.56	0.92	0.82	0.90	0.96	0.89	0.97	0.98	0.91
0.25	0.98	0.56	0.99	0.95	0.90	0.99	0.97	0.97	0.99	0.97
0.5	1.00	0.56	1.00	1.00	0.90	1.00	1.00	0.97	1.00	1.00

[a] Population prevalence of disease

[b] Optimal replication-based two-stage design using all public controls in stage 1 and all screened controls in stage 2, 2-sided test in stage 1 and 1-sided test in stage 2

Risk allele frequency in general population (f_D) = 0.3, genetic relative risk (GRR) = 1.3 assuming a multiplicative model, overall type I error (α) = 0.05

f_D = 0.3 corresponds to f_D = 0.358 in cases for all K and f_D = 0.300, 0.299, 0.297, 0.293, 0.278, 0.228 in screened controls for K = 0.0001, 0.01, 0.05, 0.1, 0.25, and 0.5, respectively

Table 2

Power for the Cochran-Armitage trend test and the proportion of cases in stage 1 that optimizes power (in parenthesis) in a replication-based two-stage GWA study with 2,000 Cases/5,000 public controls (stage 1)/2,000 screened controls (stage 2).

	Genetic Relative Risk						
f_D[a]	1.20	1.25	1.30	1.35	1.40	1.45	1.50
Follow-up Platform: 16,500							
0.1	0.01(0.32)	0.05(0.30)	0.17 (0.29)	0.39 (0.28)	0.63 (0.27)	0.82 (0.27)	0.93 (0.27)
0.3	0.19 (0.29)	0.53 (0.27)	0.84 (0.27)	0.97 (0.28)	1.00 (0.29)	1.00 (0.30)	1.00 (0.31)
0.5	0.27 (0.28)	0.64 (0.27)	0.90 (0.28)	0.98 (0.29)	1.00 (0.30)	1.00 (0.31)	1.00 (0.32)
Follow-up Platform: 7,500							
0.1	0.01 (0.36)	0.06 (0.35)	0.19 (0.33)	0.41 (0.32)	0.65 (0.31)	0.84 (0.31)	0.94 (0.31)
0.3	0.20 (0.33)	0.55 (0.32)	0.85 (0.31)	0.97 (0.32)	1.00 (0.33)	1.00 (0.33)	1.00 (0.34)
0.5	0.28 (0.33)	0.66 (0.32)	0.91 (0.32)	0.98 (0.33)	1.00 (0.33)	1.00 (0.34)	1.00 (0.35)
Follow-up Platform: 1,500							
0.1	0.01 (0.45)	0.07 (0.44)	0.21 (0.43)	0.45 (0.41)	0.69 (0.41)	0.86 (0.40)	0.95 (0.4)
0.3	0.22 (0.42)	0.59 (0.41)	0.87 (0.40)	0.97 (0.40)	1.00 (0.41)	1.00 (0.41)	1.00 (0.41)
0.5	0.31 (0.42)	0.69 (0.41)	0.92 (0.41)	0.99 (0.41)	1.00 (0.41)	1.00 (0.42)	1.00 (0.42)
Follow-up Platform: 100							
0.1	0.02 (0.59)	0.08 (0.58)	0.24 (0.57)	0.49 (0.57)	0.73 (0.56)	0.89 (0.55)	0.96 (0.55)
0.3	0.25 (0.57)	0.62 (0.56)	0.89 (0.55)	0.98 (0.55)	1.00 (0.54)	1.00 (0.53)	1.00 (0.53)
0.5	0.34 (0.57)	0.72 (0.56)	0.93 (0.55)	0.99 (0.55)	1.00 (0.54)	1.00 (0.54)	1.00 (0.53)

[a] Risk allele frequency

Population Prevalence of Disease (K) = 0.10

Number of markers on genome-wide platform (M) = 500,000

Overall type I error (α) = 0.05

Table 3

Statistical power calculations accounting for poor matching of ancestry between study samples and public controls. Calculations are for one- and replication-based two-stage study designs including study controls (n = 2,000), public controls (n = 5,000) or both. Calculations assume 2,000 cases, M = 500,000 markers in stage 1, a multiplicative genetic model with susceptibility allele frequency = 0.3, K = 0.10 and GRR = 1.3. Study samples have on average a proportion = 0.25 *POP1* ancestry (cases = 0.221, study controls = 0.253, SD = 0.15). Power is calculated across a range of proportions of *POP1* ancestry (0.1 – 0.4) in public controls and for three pairs of values of allele frequencies of the risk allele (f_D) in the *POP1* and *POP2* ancestral populations.

Proportion of *POP1* in Public Controls	f_D in *POP1*	f_D in *POP2*	One-Stage Study Designs			Two-Stage Study Design: Follow-Up Platform			
			Study Controls Only	Public Controls Only	Study + Public Controls	100	1,500	7,500	16,500
0.10	0.00	0.40	0.76	0.83	0.96	0.84	0.83	0.82	0.80
	0.30	0.30	0.77	0.80	0.96	0.83	0.83	0.82	0.81
	0.60	0.20	0.75	0.75	0.95	0.79	0.81	0.80	0.79
0.20	0.00	0.40	0.75	0.90	0.96	0.87	0.85	0.82	0.82
	0.30	0.30	0.77	0.89	0.97	0.87	0.86	0.83	0.83
	0.60	0.20	0.76	0.87	0.95	0.85	0.84	0.82	0.80
0.25	0.00	0.40	0.76	0.89	0.96	0.86	0.86	0.83	0.81
	0.30	0.30	0.76	0.89	0.96	0.86	0.86	0.84	0.82
	0.60	0.20	0.75	0.88	0.95	0.85	0.84	0.82	0.80
0.30	0.00	0.40	0.76	0.84	0.95	0.84	0.85	0.82	0.81
	0.30	0.30	0.77	0.87	0.96	0.86	0.85	0.82	0.82
	0.60	0.20	0.76	0.86	0.95	0.84	0.83	0.81	0.80
0.40	0.00	0.40	0.76	0.61	0.92	0.73	0.78	0.78	0.78
	0.30	0.30	0.77	0.65	0.92	0.75	0.80	0.79	0.79
	0.60	0.20	0.75	0.68	0.92	0.74	0.78	0.78	0.78

Table 4

Statistical power calculations for one-stage study designs accounting for batch genotype effects between study samples and public controls. Calculations assume 2,000 study cases, 2,000 study controls and 5,000 public controls and M = 500,000 markers. Power calculated for a multiplicative genetic model with susceptibility minor allele frequency = 0.3, K = 0.10 and GRR = 1.3 across different combinations of error rates for the two genotyping platforms both before and after adjustment for mean systematic inflation in test statistics for SNPs across the genome due to batch genotype effects.

		Power	
Error Rate Public Controls	**Error Rate Study Samples**	**No Systematic Inflation Across Genome Due to Batch Effects (μ =1.0)**	**Systematic Inflation Across Genome Due to Batch Effects***
Single Stage Design with Study Controls Only			
n.a.	+ 0.01	0.76	0.76
n.a.	0.00	0.79	0.79
n.a.	− 0.01	0.77	0.77
Single Stage Design with Public Controls Only			
+ 0.01	+ 0.01	0.89	0.83
	0.00	0.68	0.59
	− 0.01	0.53	0.43
0.00	+ 0.01	0.98	0.96
	0.00	0.90	0.85
	− 0.01	0.82	0.74
− 0.01	+ 0.01	0.99	0.98
	0.00	0.95	0.92
	− 0.01	0.90	0.84
Single Stage Design with Study and Public Controls			
+ 0.01	+ 0.01	0.96	0.94
	0.00	0.89	0.86
	− 0.01	0.83	0.78
0.00	+ 0.01	0.99	0.99
	0.00	0.97	0.96
	− 0.01	0.94	0.92
− 0.01	+ 0.01	1.00	0.99
	0.00	0.98	0.98
	− 0.01	0.96	0.95

* μ = 1.000, 1.100 and 1.056 for Study Controls Only, Public Controls Only, and Public and Study Controls, respectively.

Table 5

Statistical power calculations for replication-based two-stage design accounting for batch genotype effects between study samples and public controls. Calculations assume 2,000 study cases (spread across stages 1 and 2), 5,000 public controls (stage 1), 2,000 public controls (stage 2) and M = 500,000 markers in stage 1. Power was calculated for a multiplicative genetic model with susceptibility minor allele frequency = 0.3, K = 0.10 and GRR = 1.3 across different combinations of error rates for the three genotyping platforms both before and after adjustment for mean systematic inflation in test statistics for SNPs across the genome due to batch effects. Power was calculated for each replication-based two-stage follow-up study design.

			Power							
Error Rate Public Controls	**Error Rate Stage 1**	**Error Rate Stage 2**	**No Systematic Inflation Across Genome Due to Batch Effects (μ = 1.0)**				**Systematic Inflation Across Genome Due to Batch Effects***			
			16,500	7,500	1,500	100	16,500	7,500	1,500	100
+ 0.01	+ 0.01	+ 0.01	0.81	0.83	0.85	0.87	0.81	0.82	0.84	0.86
		0.00	0.83	0.84	0.86	0.88	0.83	0.84	0.86	0.87
		− 0.01	0.82	0.83	0.86	0.88	0.82	0.83	0.85	0.86
	0.00	+ 0.01	0.77	0.78	0.79	0.78	0.77	0.77	0.78	0.76
		0.00	0.79	0.80	0.80	0.79	0.79	0.79	0.79	0.77
		− 0.01	0.78	0.79	0.79	0.79	0.78	0.78	0.78	0.76
	− 0.01	+ 0.01	0.74	0.74	0.74	0.70	0.74	0.74	0.72	0.68
		0.00	0.76	0.76	0.75	0.72	0.76	0.76	0.74	0.69
		− 0.01	0.75	0.75	0.75	0.71	0.75	0.75	0.73	0.68
0.00	+ 0.01	+ 0.01	0.84	0.86	0.89	0.93	0.84	0.86	0.89	0.92
		0.00	0.86	0.88	0.90	0.93	0.86	0.87	0.90	0.93
		− 0.01	0.85	0.87	0.90	0.93	0.85	0.87	0.89	0.92
	0.00	+ 0.01	0.82	0.83	0.85	0.88	0.81	0.83	0.85	0.87
		0.00	0.84	0.85	0.87	0.89	0.83	0.84	0.86	0.88
		− 0.01	0.83	0.84	0.86	0.89	0.82	0.84	0.86	0.87
	− 0.01	+ 0.01	0.80	0.81	0.83	0.84	0.80	0.80	0.82	0.82
		0.00	0.82	0.83	0.84	0.85	0.81	0.82	0.83	0.83
		− 0.01	0.81	0.82	0.83	0.84	0.81	0.81	0.83	0.83
− 0.01	+ 0.01	+ 0.01	0.85	0.87	0.90	0.94	0.85	0.87	0.90	0.94
		0.00	0.87	0.89	0.92	0.95	0.87	0.88	0.91	0.94

			Power							
Error Rate Public Controls	**Error Rate Stage 1**	**Error Rate Stage 2**	**No Systematic Inflation Across Genome Due to Batch Effects (μ = 1.0)**				**Systematic Inflation Across Genome Due to Batch Effects***			
		− 0.01	0.86	0.88	0.91	0.95	0.86	0.88	0.91	0.94
	0.00	+ 0.01	0.83	0.85	0.87	0.91	0.83	0.84	0.87	0.90
		0.00	0.85	0.86	0.89	0.92	0.85	0.86	0.88	0.91
		− 0.01	0.84	0.86	0.88	0.91	0.84	0.85	0.88	0.90
	− 0.01	+ 0.01	0.82	0.83	0.85	0.88	0.81	0.83	0.85	0.86
		0.00	0.83	0.85	0.87	0.89	0.83	0.84	0.86	0.87
		− 0.01	0.83	0.84	0.86	0.88	0.82	0.83	0.85	0.87

*μ = 1.063, 1.048, 1.039, and 1.034 for the first stage of replication-based two-stage studies with follow-up on 100, 1500, 7500, and 16500 SNPs, respectively. For the second stage, μ = 1.000 for all follow-up studies.

Table 6

Estimated relative total cost* (power/proportion of total study samples genotyped in stage 1) of GWA study (M = 500,000 SNPs) for one- and two-stage study designs that include only study controls (n = 2,000), only public controls (n = 5,000) or both. Relative cost estimates assume 2,000 cases, a multiplicative genetic model with susceptibility minor allele frequency = 0.3, K = 0.10 and GRR = 1.3. The relative costs of genotyping 16,000, 7,500, 1,500, and 100 SNPs was assumed to be 1/2, 1/3, 1/5, and 1/12 of the cost of genotyping all 500,000 SNPs on GWA panel, respectively.

One–Stage Genotype Design(All Samples Genotyped on GWA Panel)	**Study Controls Only** **$1.00 (0.78/1.00)**		**Public Controls Only** **$0.50 (0.90/1.00)**	**Study + Public Controls** **$1.00 (0.97/1.00)**
Two-Stage Genotype Design	Joint-Analysis Study Controls Only[a]	Joint-Analysis Public Controls Only[b]	Joint-Analysis Public + Study Controls[c]	Replication- Based Public + Study Controls[c]
Follow-up Platform				
100 SNPs	$0.725 (0.78/0.70)	$0.317 (0.89/0.60)	$0.395 (0.95/0.68)	$0.335 (0.89/0.55)
1,500 SNPs	$0.640 (0.78/0.55)	$0.280 (0.89/0.45)	$0.400 (0.95/0.50)	$0.360 (0.87/0.40)
7,500 SNPs	$0.640 (0.78/0.46)	$0.287 (0.89/0.36)	$0.448 (0.95/0.39)	$0.437 (0.85/0.31)
16,500 SNPs	$0.700 (0.78/0.40)	$0.328 (0.89/0.31)	$0.585 (0.95/0.34)	$0.568 (0.84/0.27)

*Costs are calculated relative to the cost of genotyping 2,000 study cases and 2,000 study controls on GWA marker panel

[a]For joint-analysis two-stage design, proportion of samples genotyped in stage 1 represents proportion of cases and study controls (both included in stage 1 genotyping)

[b]For joint-analysis two-stage design, proportion of samples genotyped in stage 1 represents proportion of cases genotyped in stage 1. No study controls are genotyped in stage 1 or stage 2.

[c]For joint-analysis and replication-based two-stage designs, proportion of samples genotyped in stage 1 represents proportion of cases genotyped in stage 1. All study controls are genotyped in stage 2.