HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

Neuroimage. 2015 November 1; 121: 136–145. doi:10.1016/j.neuroimage.2015.07.058.

A Cautionary Note on Using Secondary Phenotypes in Neuroimaging Genetic Studies

Junghi Kim[1] and **Wei Pan**[1] **for the Alzheimer's Disease Neuroimaging Initiative**[2]

[1] Division of Biostatistics, University of Minnesota

[2] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/howtoapply/ADNIAcknowledgementList.pdf.

Abstract

Almost all genome-wide association studies (GWASs), including Alzheimer's Disease Neuroimaging Initiative (ADNI), are based on the case-control study design, implying that the resulting case-control data are likely a biased, not random, sample of the target population. Although association analysis of the disease (e.g. Alzheimer's disease in the ADNI) can be conducted using a standard logistic regression by ignoring the biased case-control sampling, a standard linear regression analysis on a secondary phenotype (e.g. any neuroimaging phenotype in the ADNI) may in general lead to biased inference, including biased parameter estimates, inflated Type I errors and reduced power for association testing. Despite of this well known result in genetic epidemiology, to our surprise, all the published studies on secondary phenotypes with the ADNI data have ignored this potential problem. Here we aim to answer whether such a standard analysis of a secondary phenotype is valid or problematic with the ADNI data. Through both real data analyses and simulation studies, we found that, strikingly, such an analysis was generally valid (with only small biases or slightly inflated Type I errors) for the ADNI data, though cautions must be taken when analyzing other data. We also illustrate applications and possible problems of two methods specifically developed for valid analysis of secondary phenotypes.

Keywords

ADNI; biased sampling; case-control design; GWAS; inverse probability weighting; linear regression; logistic regression; SPREG

Correspondence author: Wei Pan, Telephone: (612) 626-2705, Fax: (612) 626-0660, weip@biostat.umn.edu, Address: Division of Biostatistics, MMC 303, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455–0392, U.S.A..

1 Introduction

Genome-wide association studies (GWASs) have become popular for identifying genetic variants associated with complex diseases and other secondary phenotypes. Most existing GWASs adopt the case-control design, in which a certain number of disease-affected and disease-free individuals are sampled from the corresponding subpopulations respectively (Hunter et al. 2007; Scott et al. 2007; Thomas et al. 2008). Due to its separate samplings on the subjects conditional on their disease status, a key feature of a case-control sample is that it is not a random sample from the population; though both the case sample and the control sample are a random sample from the corresponding subpopulation, the combined case-control sample is biased for the population because, for example, a fraction of the cases (often close to 50%) much larger than that of the population are included in the case-control sample. Interestingly, when a standard logistic regression model is applied to a case-control sample to assess the disease and a (genetic or other) risk factor association, the case-control sample can be treated as a random sample from the population, though the estimated disease prevalence (i.e. the intercept) is biased (Prentice and Pyke 1979). However, when a linear regression model is applied to other secondary phenotypes to assess their associations with a risk factor, if no adjustment is made for the biased case-control sample, estimation and inference result except under some special situations (Lin and Zeng 2009). These conclusions apply to neuroimaging genetic studies. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI) collected its samples based on participants' disease status: specifically, in ADNI (or more precisely, ADNI-1 as used throughout), 200 healthy controls (HCs),biased 400 subjects with mild cognitive impairment (MCI) and 200 patients with Alzheimer's Disease (AD) were recruited; the set of the ADNI participants is not expected to be a random sample of the age-matched general population. For instance, in the general population, ten to twenty percent of people age 65 or older is known to have mild cognitive impairment (MCI) (Lopez et al 2003; Roberts et al 2008; Hanninen et al 2002), but nearly a half of the ADNI samples consists of MCI individuals. Hence, although a standard logistic regression model can be applied to draw unbiased inference for genetic associations with the risk of AD, a standard regression model (without any suitable adjustment) may lead to biased inference of genetic associations with secondary phenotypes, such as many neuroimaging phenotypes. On the other hand, surprisingly, to our knowledge, all the publications on analyses of secondary phenotypes for the ADNI data have relied on standard linear regression without any adjustment to or even any discussion on possible problems with the biased ADNI sample (e.g. Shen et al. 2010; Stein et al. 2010a; Meda et al. 2012; Hibar et al. 2015b). Biased inference may lead to not only biased parameter estimates, but also inflated Type I error rates and reduced power. It is the primary goal of this paper to address whether such standard linear regression really leads to biased inference for secondary phenotypes using the ADNI data as an example; if so, to what extent.

As a result, findings from previous studies may be questioned. A number of strategies have been proposed for correct inference for secondary phenotypes, including inverse probability weighted regression (Schifano et al. 2013; Monsees et al. 2009), use of retrospective likelihoods (Lin and Zeng 2009; Wei et al 2013; Ghosh et al 2014) and conditional and other methods (Chen et al 2013; Tchetgen 2014). Since the retrospective likelihood method of Lin

and Zeng (2009) is statistically efficient (but technically more challenging to extend to other more complex situations), while inverse probability weighted regression is easier to implement (but less efficient statistically), we use them as the references against the standard linear regression. In addition, since some imaging genetics studies (Potkin et al. 2010; Hibar et al. 2015a) have considered a variation of the standard linear regression by adjusting for the (primary phenotype) disease status, we also consider this method.

We first briefly review the above four methods for association analysis of a genetic variant and a secondary phenotype. We then apply the methods to the ADNI data in section 3. In section 4, realistic simulation studies mimicking the ADNI data are conducted to further investigate possible problems when analyzing a secondary phenotype. Section 5 provides a simple toy example to demonstrate the problem and offers some intuitive explanations. A summary of our conclusions is given in section 6.

2 Methods

Let $\{x_i, Y_i, Z_i, D_i\}$ be the observed data for subject $i = 1, \ldots, n$, where x_i is an additive genotype score of an SNP of interest, Y_i is a univariate and quantitative secondary phenotype, $D_i = 1$ or 0 is an indicator of the disease (i.e. primary phenotype), and $Z_i = (Z_{i1}, \ldots, Z_{il})'$ is a vector of covariates. Define the number of controls (with $D_i = 0$) as n_0, and that of cases (with $D_i = 1$) as n_1. A major characteristic of a case-control study is that the disease-status is identified at the beginning of the study, and the sampling of the subjects is conditional on their disease status. One implication is that the combined case-control data may not be a random sample from the population. A proper analysis should take account of the sampling scheme; otherwise biases may result. This paper considers following four approaches: the first two are standard approaches currently widely used in imaging genetics, while the last two were specifically developed for valid analysis of secondary phenotypes.

2.1 Unadjusted linear model

A standard linear model regressing the secondary phenotype (Y_i) on the genotype score (x_i) has been used for testing association between the two:

$$E(Y_i|x_i, Z_i) = \beta_0 + \beta_1 x_i + \beta_z' Z_i, \quad (1)$$

and it is assumed that the conditional distribution $f(Y_i|x_i, Z_i)$ is Normal, $N(\beta_0 + \beta_1 \chi_i + \beta_z' Z_i, \sigma^2$. Accordingly, based on the likelihood $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i)$, maximum likelihood is used to draw inference on β_1. For example, as used in the following, the Wald test is applied to test the null hypothesis $H_0 : \beta_1 = 0$ based on the maximum likelihood estimate (MLE) $\hat{\beta}_1$.

Note that with a case-control sample, in general the above likelihood function $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i)$ is not appropriate, failing to account for the conditional sampling. Hence, in general the above inference is expected to be biased.

2.2 Disease status adjusted linear model

A simple way to adjust for the case-control sampling is to adjust for the disease status (D_i) in a standard linear regression model (Potkin et al. 2010):

$$E(Y_i|x_i, Z_i, D_i) = \beta_0 + \beta_1 x_i + \beta_z' + \beta_d D_i,$$

and it is assumed that the distribution density $f(Y_i|x_i, Z_i, D_i)$ is $N\left(\beta_0 + \beta_1 x_i + \beta_z' Z_i + \beta_d D_i, \sigma^2\right)$. The likelihood function $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i, D_i)$ is used for inference of β_1 in the framework of maximum likelihood. Again note that the likelihood $L = \prod_{i=1}^{n} f(Y_i|x_i, Z_i, D_i)$ is in general invalid for the case-control data.

2.3 Inverse probability weighted regression

To properly account for biased case-control sampling, a weighted likelihood (or weighted estimating equations) can be used (Richardson et al. 2007; Monsees et al. 2009; Schifano et al. 2013). The weight for each subject is defined to be proportional to the inverse probability of the subject's being sampled into the case-control data. Intuitively, for instance, if the disease is rare in the population, but an equal number of cases and controls are sampled, the weight is used to up-weight the controls and down-weight the affected individuals so that the weighted case-control sample is like a random sample from the population. Monsees et al. (2009) discussed such an inverse probability weighted (IPW) regression approach, offering unbiased inference of genotype-secondary phenotype associations, though its statistical efficiency may be low. Following Schifano et al. (2013), in this study, the weight (w_i) for subject i was specified as $w_i = p/\pi$ if $D_i = 1$, and $w_i = (1 - p)/(1 - \pi)$ if $D_i = 0$, where $p = P(D = 1)$ is the disease prevalence in the population, and $\pi = p(D = 1|\text{sampled})$ is the proportion of affected individuals in the case-control sample, which is always substituted with $n_1/(n_0 + n_1)$ throughout. The regression model is the same as equation (1), but the likelihood is weighted with w_i, i.e. $L_w = \prod_{i=1}^{n} f(Y_i|x_i, Z_i)^{w_i}$, where $f(Y_i|x_i, Z_i)$ is the density function for a normal distribution, $N\left(\beta_0 + \beta_1 x_i + \beta_z' Z_i, \sigma^2\right)$. Maximum likelihood is used for inference. We implemented the above IPW regression approach using geeglm() function in R.

2.4 A retrospective likelihood approach

Lin and Zeng (2009) and Ghosh et al. (2014) proposed retrospective likelihoods to properly account for the fact that the case-control data should be conditioned on the disease status. Specifically, a regression model for secondary phenotype data (SPREG) proposed by Lin and Zeng (2009) is based on a retrospective likelihood $f(Y_i, x_i, Z_i|D_i) =$

$$\left\{\frac{P(D_i=1|x_i, Y_i, Z_i)\, f(Y_i|x_i, Z_i)\, f(x_i, Z_i)}{P(D_i=1)}\right\}^{D_i} \times \left\{\frac{P(D_i=0|x_i, Y_i, , Z_i)\, f(Y_i|x_i, Z_i)\, f(x_i, Z_i)}{P(D_i=0)}\right\}^{1-D_i} \quad (2)$$

where $P(D_i = 1) = \int_Y \int_{x,Z} P(D_i = 1|Y_i, x_i, Z_i) f(Y_i|x_i, Z_i) f(x_i, Z_i) d_Y d_{x,Z}$, $P(D_i = 0) = 1 - P(D_i = 1)$, $P(D_i = 1|Y_i, x_i, Z_i)$ determined by

$$logitP\left(D_i=1|x_i, Y_i, Z_i\right)=\alpha_0+\alpha_1 x_i+\alpha_2 Y_i+\alpha_z^{'} Z_i,$$

$f(Y_i|x_i, Z_i)$ is the density function of $N(\beta_0 + \beta_1 x_i + \beta_z Z_i, \sigma^2)$, and $f(x_i, Z_i)$ is treated as nuisance parameters. A profile likelihood is used to eliminate nuisance parameters, which is then maximized by the Newton-Raphson algorithm; maximum likelihood is used to draw inference on β_1. Since this method is likelihood-based, it is efficient. However, due to the presence of high-dimensional nuisance parameters, the (profile) likelihood may be difficult to maximize, leading to some numerical problems as pointed out by Lutz et al (2014) and to be confirmed later, especially if the disease prevalence is unknown or estimated inaccurately (Chen et al 2013). Software for SPREG was downloaded from http://dlin.web.unc.edu/software/spreg-2/.

3 ADNI data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (**adni.loni.usc.edu**). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (**adni.loni.usc.edu**). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 60 million, 5-year public private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

3.1 ADNI data analysis

We considered a univariate and quantitative secondary phenotype, volume of right hippocampus, for its possible association with each of several SNPs, rs429358, rs2075650, rs7526034, rs10932886, rs7647307, rs7610017, rs4692256 and rs6463843, which were chosen because they were shown to be highly associated with multiple imaging phenotypes

when using the standard (unadjusted) linear regression method (Shen et al. 2010). From the ADNI baseline data, we extracted the secondary phenotype, the SNPs and five covariates: gender, education, handedness, age, and intracranial volume (ICV) for association testing.

We regressed hippocampus volume on each genotype score and five covariates using the four methods: standard linear regression without adjusting for disease status (unadj-lm), with adjustment for disease status (D-adj-lm), IPW regression (lm-w), and SPREG (Lin and Zeng 2009). For the latter two methods, an estimate of the AD prevalence in the population is needed, which was obtained based on the following data. In 2014, it was reported that one in nine people age 65 and older (11 percent) had AD, and one third of people age 85 and older (32 percent) had AD; in 2012, 13 percent people age 65 and older were believed to have AD, and nearly half of people age 85 and older had AD (Alzheimer's Association, 2014, 2012; Hebert et al. 2013). It was not straightforward to determine the disease prevalence, since the AD prevalence for an aging population varies over time and it is not always clear what is the age-matched population based on the given case-control sample. The subjects in our collected data had mean age 75.68 with minimum 56, the first and third quantiles 71.75 and 75.68 respectively. Thus we estimated that the AD prevalence (p) in the population ranged from 0.10 to 0.30. Accordingly we considered disease prevalence $p \in \{0.10, 0.13, 0.16, 0.20, 0.23, 0.27, 0.30\}$, investigating how the results depended on the chosen p. For IPW regression, a subject's weight (w_i) was calculated based on a given p as discussed before; for SPREG, a given p was input to the software program.

We applied the methods to the ADNI data including all $n_0 = 180$ healthy controls (HCs) and $n_1 = 144$ AD patients available from the ADNI baseline data. The results are summarized in Table 1. Unadj-lm, lm-w and SPREG suggested significant associations between rs429358/rs2075650 and right hippocampus volume. This is consistent with the results from previous studies (Shen et al. 2010; Kim et al. 2002; Lu et al. 2011; Mori et al. 2002). Unadj-lm and SPREG showed more significant p-values. When the disease prevalence p=0.10 or 0.13 was assumed, none of the p-values given by lm-w could reach the genome- wide significance level (5×10^{-8}); however, if p=0.27 or 0.30 was used, rs429358 became highly significant, demonstrating that the results of lm-w were sensitive to the estimate of the disease prevalence p. The dependence of SPREG on p was to a lesser degree. It is noted that disease adjusted linear model (D-adj-lm) gave no significant p-value for any SNP. Interestingly, when the disease prevalence p=0.23 was assumed, which was reasonable, unadj-lm and SPREG showed p-values close to each other.

To confirm the results in Table 1 with a larger sample size, we included additional 311 MCI subjects in the ADNI data. We treated the MCI subjects as controls, and applied the methods with 491 controls and 144 AD patients. In Table 2, all p-values became smaller but only rs429358 and rs2075650 showed strong associations with the right hippocampus volume, in agreement to that in Table 1. Again, when assuming disease prevalence p = 0.16 or 0.20, which was reasonable (because MCIs were treated as controls), unadj-lm, lm-w and SPREG, but not D-adj-lm, all gave similar results.

3.2 An association scan on chromosome 19

Rather than drawing our conclusions based on only a few SNPs, we conducted a genome-wide scan on chromosome 19. The secondary phenotype was still the right hippocampus volume. We included all the SNPs with minor allele frequency (maf) ≥ 0.05, genotyping rate more than 90%, and surviving the Hardy-Weinberg Equilibrium test with p-value > 0.001, resulting in 9184 SNPs to be tested. Subjects with more than 10% missing genotypes were excluded; only non-Hispanic Caucasians whose right hippocampus volume was measured at baseline were included. As in the previous section, two sample sizes were considered: (1) n_0 = 180 controls (HCs) and n_1 = 144 AD patients; (2) n_0 = 491 controls (including both HCs and MCIs) and n_1 = 144 AD patients.

The quantile-quantile (Q-Q) plots in Figures 1 and 2 show the distributions of the observed p-values against those of the expected (null) p-values. For each method, the pattern shown on the two plots is similar. Surprisingly, both unadj-lm and D-adj-lm had their estimated inflation factors (λ) (almost) 1 in each case, and the observed p-values were in close agreement with the expected ones, suggesting no obvious inflation of their Type I errors. Although the estimated inflation factors for lm-w were also close to 1, there were a few more points falling outside of the confidence regions. Depending on the population disease prevalence *p* used, the estimated inflation factors of SPREG ranged from 1.06 to 1.16, which were not too bad; however, most strikingly, in every Q-Q plot, there were many observed p-values far more significant than expected, implying a large portion of likely false positives, presumably due to some numerical problems for those SNPs in SPREG. In our experience, especially for secondary phenotypes with large variances such as brain volumetric measures, SPREG might not converge, and scaling a phenotype by its standard deviation improved its convergence; even with scaling, in this example, SPREG failed to converge for about 1000 SNPs (10%) when the disease prevalence (*p*) was set to be less than 0.23.

In summary, in an association scan on chromosome 19 with two sample sizes, all methods seemed to give reasonable estimates of inflation factors. In addition, the two unadjusted methods and IPW regression did not show any obvious problem in Type I error inflations; in contrast, SPREG had some numerical problems, giving many SNPs more significant p-values than expected.

For more generalizable conclusions, we also conducted a genome-wide scan on chromosome 19 with each of 27 other FreeSurfer phenotypes defined as volumetric or cortical thickness measures extracted from the ADNI data (Table S1 in Supplementary Materials). In Supplementary Materials, Figure S1 summarizes the results of the methods when applied to the 27 secondary phenotypes. For each phenotype, unadj-lm had an inflation factor close to 1 and showed a similar pattern to that of lm-w in the Q-Q plots. In addition, Tables S2 and S3 show the p-values of several candidate SNPs when the methods were applied to two selected secondary phenotypes; again unadj-lm gave the results similar to those of lm-w and SPREG (with a suitable *p*), while D-adj-lm gave less significant p-values for rs429358.

4 Simulations

4.1 Simulation set-ups

We conducted simulation studies with realistic set-ups to mimic the ADNI data. First we selected two SNPs, rs429358 and rs6463843 in Table 1, to represent two association patterns. SNP rs429358 (in gene APOE) is well known for its strong associations with both hippocampus volume and AD (Kim et al. 2002; Lu et al. 2011; Mori et al. 2002); by choosing rs429358, we had a representative case where both the SNP and secondary phenotype are highly associated with the disease risk. On the other hand, rs6463843 (in gene NXPH1) was chosen to reflect an opposite scenario where both the SNP and the secondary phenotype are only moderately associated with the disease. Next, we used the ADNI data to estimated various association parameters for each SNP. Specifically, we fitted a linear regression model with the right hippocampus volume as the secondary phenotype and an SNP (x) and covariates (Z, including gender, education, handedness, age, ICV) as predictors, obtaining the estimated regression coefficients, β_{xy} and β_{zy}. Then a logistic regression model was fitted to determine the effects of SNP (x) and the phenotype (Y) on the disease (D), obtaining the estimated regression coefficients $\beta_{Dy}$and β_{Dx}. The parameter values for the two SNPs/set-ups are given in Table 3, which were used as the true parameter values for generating simulated data.

To maintain the true correlation structures among the five covariates, we sampled $Z_i = (Z_{i1}, \ldots, Z_{i5})$ from the ADNI data in each simulation. An additive genotype score (x_i) was randomly generated from a binomial distribution Bin(2, maf) with maf=0.27 and 0.45 for the two SNPs respectively. The secondary phenotype was generated from a Normal distribution based on the simulated covariates and genotype score $\{Z_i, x_i\}$:

$$Y_i \tilde{} N\left(\phi \cdot \beta_{xy} x_i + \beta'_{zy} Z_i, \hat{\sigma}_y^2\right) \quad (3)$$

where β_{xy} and β_{zy} are presented in Table 3, $\hat{\sigma}_y^2$ was obtained from the sample variance of hippocampus volume, and φ is a scaling parameter controlling the association strength between x and Y. When $\varphi = 0$, we created a null case with no association; when $\varphi = 1$, the effect size was equal to the estimate from the ADNI data.

For each subject i, the disease status D_i was generated from a Bernoulli distribution with probability $P(D_i = 1|x_i, Y_i)$ determined by

$$LogitP\left(D_i = 1 | x_i, Y_i\right) = \beta_{D0} + \beta_{Dy} Y_i + \beta_{Dx} x_i,$$

where the values of β_{Dy} and β_{Dx} are shown in Table 3, and $\beta_{D0} = \text{logit}^{-1} p$. The disease prevalence was set at $p = 0.23$ or 0.10 to mimic that for the ADNI data. Note however that $p = 0.23$ was more reasonable for AD.

To generate a simulated data set, we repeated simulating observations $\{Z_i, x_i, Y_i, D_i\}$ until reaching the predefined sample size of n_1 cases and n_0 controls; any simulated observations not used in the case-control sample were added back to the case-control sample to form a

cohort sample. Since a cohort sample was a random sample from the population, while a case-control sample was not, we used the results from cohort samples as benchmarks.

We also used each cohort sample to obtain an estimate $\hat{p}$ of the disease prevalence for the corresponding case-control sample. To investigate the effects of the specified disease prevalence on analysis, three different disease prevalence rates, $\hat{p}-0.05$, $\hat{p}$, and $\hat{p}+0.05$, were input to lm-w and SPREG.

For each simulation set-up, the results were based on 10^4 independent simulation replicates.

4.2 Results

In Table 4, we report the empirical Type I errors for the methods for the null case (with $\varphi = 0$). SPREG and lm-w had valid Type I errors in all cases, while the results of unadj-lm and D-adj-lm largely depended on the simulation set-ups and the true disease prevalence. In set-up 1, where both the SNP and the secondary phenotype were highly associated with the disease risk, unadj-lm, lm-w and SPREG showed proper type I error rates, with the true prevalence $p = 0.23$; however, D-adj-lm gave highly inflated ones. Yet when the true disease prevalence was set at $p = 0.10$, only SPREG had type I errors close to the nominal level (0.05), and lm-w (with $\hat{p}$ applied) gave slightly inflated ones. However, the numerical results suggested that both SPREG and lm-w were sensitive to the pre-specified disease prevalence.

In set-up 2 where the SNP or the secondary phenotype was not highly associated with the disease risk, the Type I error rates of all the methods except D-adj-lm were controlled, though the inflations by D-adj-lm were small to moderate.

In order to ensure the above results were not due to a small sample size, we increased the sample size to $n_0 = n_1 = 500$ and $n_0 = n_1 = 1000$. As shown in Supplementary Materials (Table S4), the empirical Type I error rates of SPREG and lm-w were reliable as compared to unadj-lm and D-adj-lm. In Supplementary Materials, a more extreme disease prevalence $p = 0.01$ was also considered, in which the Type I errors of unadj-lm were more inflated, while D-adj-lm performed well as pointed out in Monsees et al. (2009). The corresponding Q-Q plots for Table 4 are presented in Supplementary Figures S2 and S3.

The empirical power of each method is presented in Table 5. In set-up 1, unadj-lm had the highest power (but recall that it had slightly inflated Type I errors), followed by SPREG, then by lm-w. Note the dramatic power loss of D-adj-lm in spite of its severely inflated Type I errors. In set-up 2, D-adj-lm was most powerful but, due to its inflated Type I errors, it should not count; the other three methods were similarly powered.

Figures 3 and 4 illustrate the distributions of the parameter estimates $\hat{\beta}_1$ by each method. In set-up 1 (Figure 3) lm-w and SPREG provided almost unbiased estimates, while D-adj-lm always yielded largely biased estimates; unadj-lm gave almost unbiased estimates for $p = 0.23$, but slightly biased ones for $p = 0.10$. For set-up 2 (Figure 4), only D-adj-lm gave obviously biased estimates.

More detailed numerical results are presented in Supplementary Tables S5 and S6. In all cases, lm-w and SPREG yielded unbiased estimates, while the performance of unadj-lm and D-adj-lm largely depended on the disease prevalence (and simulation set-ups).

In summary, under practical situations mimicking the ADNI data, the standard linear regression method unadj-lm, but not D-adj-lm, performed satisfactorily, giving results similar to the other two valid methods.

5 An Illustrative Example

Finally we used a simple toy example to illustrate the problems with Unadj-lm and D-adj-lm. For better visualization, we took a continuous x and no covariate Z; it is easy to see that the main points carry over to the case with a genotype score x and with Z. We assumed a finite population (or equivalently, a random sample from a super-population) containing 9000 controls (with $D = 0$) and 1000 cases (with $D = 1$). For controls, we had a predictor $x \sim N(0, 1)$, while $x \sim N(2, 1)$ for cases. A secondary phenotype Y was distributed as $Y \sim N(2D, 1)$.

Based on the assumed model, we can see that conditional on D, Y was not associated with x, which is confirmed in the left panel of Figure 5: for either the control or case group, regressing Y on x yielded a horizontal line; the OLS estimates for the slope parameter of the two groups were 0.005 (SE=0.01) and 0.004 (SE=0.03), respectively. On the other hand, marginally Y was associated with x: the OLS estimate of the slope parameter was 0.260 (SE=0.009).

Now we consider a case-control sample. To minimize the influence of the sampling errors, for simplicity, we took a random sample of 1000 controls and all 1000 cases. As shown in the right panel of Figure 5, applying Unadj-lm and D-adj-lm led to the OLS estimates of the slope parameter for x as 0.511 (SE=0.019) and 0.006 (SE=0.022) respectively; that is, Unadj-lm over-estimated the population marginal association (i.e. 0.511 versus 0.260), while D-adj-lm was on the target for the conditional association (0.006 versus 0) but again off from the marginal association (0.006 versus 0.260). We also applied weighted regression with lm-w: based on the sampling proportions, a weight 9 was assigned to each control and weight 1 to each case in the case-control sample; then we regressed Y on x; the WLS estimate of the slope parameter was 0.279 (SE=0.021), very close to the population marginal association (i.e. 0.279 versus 0.260).

It is simple why Unadj-lm may not work for a case-control sample: a case-control sample may not represent the population. More importantly, this example also clearly demonstrates that, even with the data from the whole population (or a large random sample), marginal association based on Unadj-lm and conditional association based on D-adj-lm may be quite different. More formally, we are interested in inference for β_1 in a marginal model

$$E(Y|x) = \beta_0 + \beta_1 x.$$

However, D-adj-lm is based on a conditional model

$$E(Y|x, D) = b_0 + b_1 x + b_2 D,$$

from which we can derive

$$E(Y|x) = E[E(Y|x, D)] = b_0 + b_1 x + b_2 E(D|x).$$

If D is associated with x, say $E(D|x) = a_0 + a_1 x$, then we have

$$E(Y|x) = E[E(Y|x, D)] = b_0 + a_0 + (b_1 + b_2 a_1) x,$$

based on which we may have $\beta_1 \neq b_1 + b_2 a_1$ unless $b_2 = 0$ or $a_1 = 0$.

6 Conclusions and Discussion

We set out to address whether standard linear regression of secondary phenotypes in a practical neuroimaging genetic study would lead to biased inference, i.e. biased estimates, inflated Type I errors and reduced power. This is an important question given that in general it will lead to biased inference while the current practice in neuroimaging genetics has largely ignored this potential problem. Using the ADNI data as an example, we conducted both real data analyses and simulation studies. Our main conclusion was the following: under practical situations similar to the ADNI data, using standard linear regression without any adjustment (unadj-lm), but not the one adjusting for the disease status (D-adj-lm), to assess SNP-secondary phenotype associations did not appear to cause any severe problem, though cautions still must be taken.

Of course, our main conclusion is only specific to the ADNI data, and is not applicable in general; some geenral principles were discussed, which might offer some guidelines to practitioners for other applications. The main theoretical reason for our conclusion to hold for the ADNI data (and possibly other data) is due to the high prevalence of the AD (or other disease) in the target population, leading to its small difference from the sampling proportion of the cases in the case-control sample, which is usually close to 50%. In other words, the key issue is how much biased is the case-control sample for the target population. For example, if the disease is less common, say at 10% in the general population, while as usual about a half of the case-control sample are cases, then a suitable adjustment in analysis is more likely to be necessary. There is also another factor influencing the validity of the standard unadjusted methods: the association strength between the disease and a secondary phenotype. For example, if the disease and the secondary phenotype are not associated, then a standard unadjusted analysis for the secondary phenotype is fine (Lin and Zeng 2009). However, in neuroimaging genetic studies, often a secondary phenotype is of interest simply because it is treated an intermediate phenotype for the disease, suggesting its likely association with the disease. Nevertheless, if the secondary phenotype-disease association is weak, a standard unadjusted analysis of the secondary phenotype may be only slightly biased. We have also discussed why simply adjusting for disease status (D-adj-lm) might not work: in addition to the possible poor representation of a case-control sample for the

population, D-adj-lm targets the conditional association between the secondary phenotype and an SNP (after adjusting for possible covariates), not the marginal association of interest, which may be quite different from the conditional association as shown in our toy example in section 5.

It is fair to ask why we do not always use one of the valid methods that properly correct for the sampling bias of case-control studies. In this paper we have considered two representative methods, IPW regression and SPREG; the former is general, more robust (Tapsoba et al 2014) and easier to implement but less efficient, while the latter is the opposite. For SPREG, it is challenging to extend it (or other retrospective likelihood methods) to more complex study designs beyond the simple case-control design, such as with longitudinal phenotypes or familial relatedness. Although IPW regression is general and easy to implement, its loss of power may hinder its wide use, especially for small neuroimaging GWASs. In addition, there may be numerical problems with the use of SPREG (see Figures 1 and 2, and Supplementary Materials Table S7). Furthermore, both of the methods require an estimate of the disease prevalence in the target population, and their results may be sensitive to the estimate; however, it may not be easy to obtain an accurate estimate, as in the ADNI data, since the target population is not well defined, e.g. with respect to the study participants' age while the AD (or MCI) prevalence largely depends on the age.

Although we have only considered single quantitative secondary phenotype–single SNP associations, we anticipate that our conclusions will be likely to hold for other cases, such as for binary secondary phenotypes (Wang and Shete 2010; Chen et al 2013), multiple secondary phenotypes (Lin et al 2012; Zhang et al 2014; Zhu et al 2014), longitudinal secondary phenotypes (Skup et al 2012; Xu et al 2014), or for gene-gene or gene-environment interactions (Ge et al 2015; Hibar et al 2015b), though further studies are needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

The authors are grateful to the reviewers for constructive comments. This research was supported by NIH grants R01GM113250, R01HL105397, R01HL116720 and R01GM081535, and by the Minnesota Supercomputing Institute.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neu- roimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Com- pany; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & De- velopment, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Med- pace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Rev December 5, 2013 Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern

California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Alzheimer's Association. Alzheimer's Disease Facts and Figures, Alzheimer's and Dementia. 2012; 8(2)

Alzheimer's Association. Alzheimer's Disease Facts and Figures, Alzheimer's and Dementia. 2014; 10(2)

Chen HY, Kittles R, Zhang W. Bias correction to secondary trait analysis with casecontrol design. Stat Med. 2013; 32(9):1494–1508. [PubMed: 22987618]

Ge T, Nichols TE, Ghosh D, Mormino EC, Smoller JW, Sabuncu MR, Alzheimer's Disease Neuroimaging Initiative. A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. NeuroImage. 2015; 109(1):505–514. [PubMed: 25600633]

Ghosh A, Wright F, Zou F. Unified analysis of secondary traits in case-control association studies. Journal of the American Statistical Association. 2014; 108:566–576.

Hanninen T, Hallikainen M, Tuomainen S, Vanhanen M, Soininen H. Prevalence of mild cognitive impairment: A population-based study in elderly subjects. Acta Neurol Scand. 2002; 106:148–154. [PubMed: 12174174]

Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. Alzheimer disease in the U.S. population: Prevalence estimates using the 2000 Census. Archives of Neurology. 2013; 60(8):1119–22. [PubMed: 12925369]

Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. Nature. 2015a; 520:224–229. [PubMed: 25607358]

Hibar DP, Stein JL, Jahanshad N, Kohannim O, Hua X, Toga AW, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Alzheimer's Disease Neuroimaging Initiative, Weiner, M.W. Thompson PM. Genome-wide interaction analysis reveals replicated epistatic effects on brain structure. Neurobiol Aging. 2015b; 36(Suppl 1):S151–S158. [PubMed: 25264344]

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmeno-pausal breast cancer. Nat. Genet. 2007; 39:870–874. [PubMed: 17529973]

Kim DH, Payne ME, Levy RM, MacFall JR, Steffens DC. APOE genotype and hippocampal volume change in geriatric depression. Biol Psychiatry. 2002; 51(5):426–429. [PubMed: 11904138]

Lin DY, Zeng D. Proper analysis of secondary phenotype data in case-control association studies. Genet. Epidemiol. 2009; 33:256–265. [PubMed: 19051285]

Lin J, Zhu H, Knickmeyer R, Styner M, Gilmore J, Ibrahim JG. Projection Regression Models for Multivariate Imaging Phenotype. Genet Epidemiol. 2012; 36:631–641. [PubMed: 22807230]

Lin JA, Zhu H, Mihye A, Sun W, Ibrahim JG, the Alzheimer's Neuroimaging Initiative. Functional-Mixed Effects Models for Candidate Genetic Mapping in Imaging Genetic Studies. Genet Epidemiol. 2014; 38:680–691. [PubMed: 25270690]

Lopez OL, Jagust WJ, DeKosky ST, Becker JT, Fitzpatrick A, Dulberg C, et al. Prevalence and classification of mild cognitive impairment in the cardiovascular health study cognition study. Arch Neurol. 2003; 60:1385–1389. [PubMed: 14568808]

Lu PH, Thompson PM, Leow A, Lee GJ, Lee A, Yanovsky I, Parikshak N, Khoo T, Wu S, Geschwind D, Bartzokis G. Apolipoprotein E genotype is associated with temporal and hippocampal atrophy rates in healthy elderly adults: a tensor-based morphometry study. J Alzheimers Dis. 2011; 23(3): 433–42. [PubMed: 21098974]

Lutz SM, Hokanson JE, Lange C. An alternative hypothesis testing strategy for secondary phenotype data in case-control genetic association studies. Frontiers in Genetics. 2014; 5:188. [PubMed: 25071819]

Meda SA, Narayanan B, Liu J, Perrone-Bizzozero NI, Stevens MC, Calhoun VD, Glahn DC, Shen L, Risacher SL, Saykin AJ, Pearlson GD. A large scale multivariate parallel ICA method reveals novel imaging-genetic relationships for Alzheimer's disease in the ADNI cohort. Neuroimage. 2012; 60(3):1608–1621. [PubMed: 22245343]

Mori E, Lee K, Yasuda M, Hashimoto M, Kazui H, Hirono N, Matsui M. Accelerated hippocampal atrophy in Alzheimer's disease with apolipoprotein E epsilon4 allele. Ann Neurol. 2002; 51(2): 209–14. [PubMed: 11835377]

Monsees GM, Tamimi RM, Kraft P. Genome-wide association scans for secondary traits using case-control samples. Genet Epidemiol. 2009; 33:717–728. [PubMed: 19365863]

Potkin SG, Macciardi F, Guffanti G, Fallon JH, Wang Q, Turner JA, Lakatos A, Miles MF, Lander A, Vawter MP, Xie X. Identifying gene regulatory networks in schizophrenia. Neuroimage. 2010; 53(3):839–847. [PubMed: 20600988]

Prentice RL, Pyke R. Logistic Disease Incidence Models and Case-Control Studies. Biometrika. 1979; 66(3):403–411.

Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, et al. The Mayo clinic study of aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology. 2008; 30:58–69. [PubMed: 18259084]

Richardson DB, Rzehak P, Klenk J, Weiland SK. Analyses of case-control data for additional outcomes. Epidemiology. 2007; 18:441–445. [PubMed: 17473707]

Schifano ED, Li L, Christiani DC, Lin X. Genome-wide association analysis for multiple continuous secondary phenotypes. Am J Hum Genet. 2013; 92:744–759. [PubMed: 23643383]

Shen L, Kim S, Risachera SL, Nho K, Swaminathan S, Westa JD, Foroudd T, Pankratz N, Mooree JH, Sloane CD, Huentelmanf MJ, Craig DW, DeChairog BM, Potkinh SG, Jack CR Jr, Weiner MW, Saykin AJ, the Alzheimer's Disease Neuroimaging Initiative. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. Neuroimage. 2010; 53(3):1051–1063. [PubMed: 20100581]

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007; 316:1341–1345. [PubMed: 17463248]

Skup M, Zhu H, Zhang H. Multiscale Adaptive Marginal Analysis of Longitudinal Neuroimaging Data with Time-Varying Covariates. Biometrics. 2012; 68:1083–1092. [PubMed: 22551084]

Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Dechairo BM, Potkin SG, Weiner MW, Thompson P, Alzheimer's Disease Neuroimaging Initiative. Voxelwise genome-wide association study (vGWAS). Neuroimage. 2010a; 53(3):1160–1174. [PubMed: 20171287]

Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, et al. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. NeuroImage. 2010b; 51(2):542–554. [PubMed: 20197096]

Tapsoba JD, Kooperberg C, Reiner A, Wang CY, Dai JY. Robust estimation for secondary trait association in case-control genetic studies. Am J Epidemiol. 2014; 179(10):1264–1272. [PubMed: 24723002]

Tchetgen EJT. A general regression framework for a secondary outcome in case-control studies. Biostatistics. 2014; 5(1):117–128.

Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet. 2008; 40:310–315. [PubMed: 18264096]

Wang J, Shete S. Estimation of odds ratios of genetic variants for the secondary phenotypes associated with primary disease. Genet Epidemiol. 2011; 35:190–200. [PubMed: 21308766]

Wei J, Carroll RJ, Muller UU, Keilegom IV. Robust estimation for homoscedastic regression in the secondary analysis of case-control data. J Roy Stat Soc B. 2013; 75:185–206.

Xu Z, Shen X, Pan W, the Alzheimer's Disease Neuroimaging Initiative. Longitudinal Analysis Is More Powerful than Cross-Sectional Analysis in Detecting Genetic Association with Neuroimaging Phenotypes. PLoS ONE. 2014; 9(8):e102312. [PubMed: 25098835]

Zhang Y, Xu Z, Shen X, Pan W, the ADNI. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. NeuroImage. 2014; 96:309–325. [PubMed: 24704269]

Zhu H, Khondker Z, Lu Z, Ibrahim JG. Bayesian Generalized Low Rank Regression Models for Neuroimaging Phenotypes and Genetic Markers. Journal of the American Statistical Association. 2014; 109:977–990.

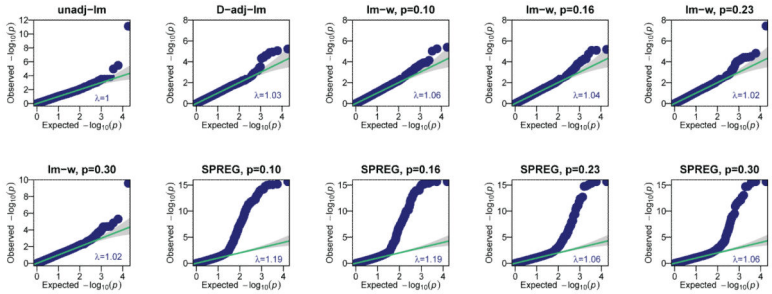

Figure 1.

Q-Q plots of the p-values for each methods when applied to SNPs on chromosome 19 for the ADNI data.

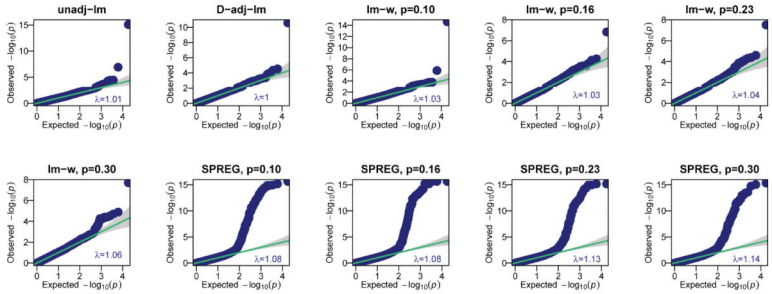

Figure 2.

Q-Q plots of the p-values for each methods when applied to SNPs on chromosome 19 for the ADNI data. All subjects with MCI were included as controls.

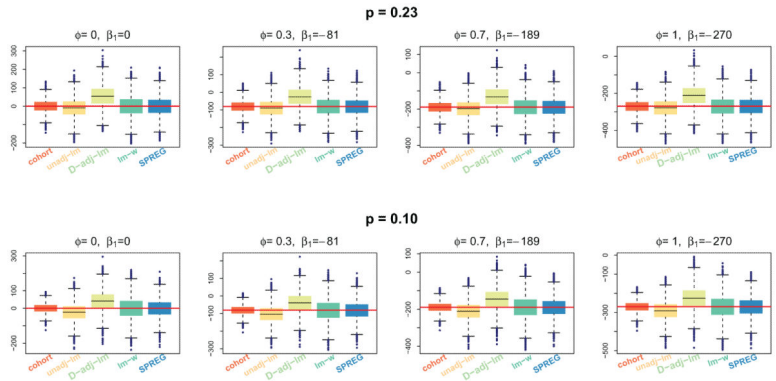

Figure 3.

Simulation set-up 1: Distributions of the estimates $\hat{\beta}_1$ from each method with two different values of the disease prevalence *p*.

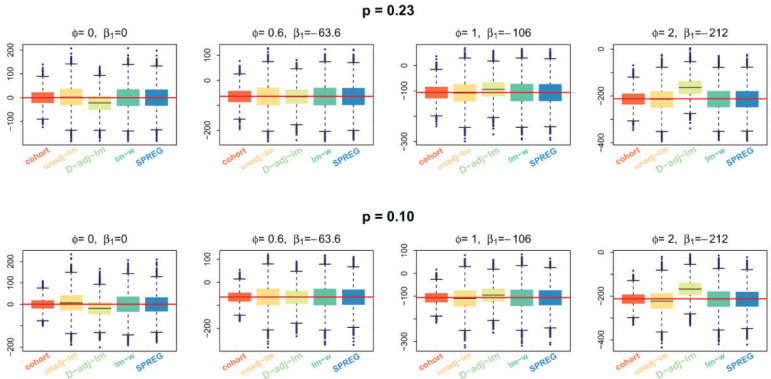

Figure 4.

Simulation set-up 2: Distributions of the estimates $\hat{\beta}_1$ from each method with two different values of the disease prevalence *p*.

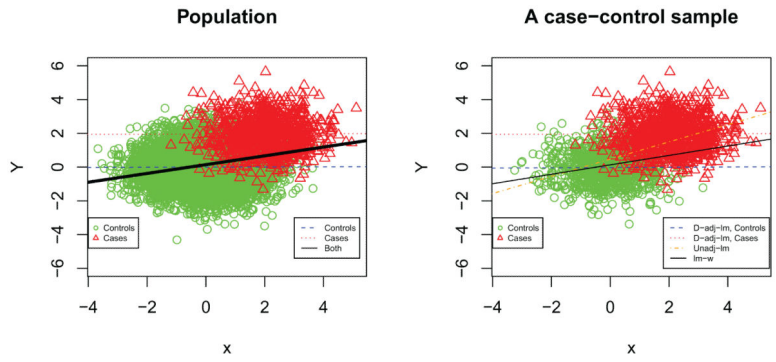

Figure 5.

An illustrative example. The left panel is for a population with 9000 controls and 1000 cases, while the right panel is for a case-control sample with 1000 controls and 1000 cases.

Table 1

P-values for association testing between right hippocampus volume and each candidate SNP with 324 subjects in the ADNI data.

					lm-w						
SNP	chr	maf	unadj-lm	D-adj-lm	*p*=0.10	*p*=0.13	*p*=0.16	*p*=0.20	*p*=0.23	*p*=0.27	*p*=0.30
rs429358	19	0.27	7.78e-12	2.03e-02	8.06e-04	7.99e-05	3.63e-06	3.61e-07	3.61e-07	2.02e-09	2.60e-10
rs2075650	19	0.24	5.18e-08	6.11e-02	9.63e-03	2.25e-03	3.42e-04	8.81e-05	8.81e-05	4.77e-06	1.58e-06
rs7526034	1	0.12	6.46e-02	5.68e-01	3.56e-01	2.83e-01	2.17e-01	1.83e-01	1.83e-01	1.32e-01	1.18e-01
rs10932886	2	0.33	3.75e-02	3.31e-01	1.28e-01	9.39e-02	6.62e-02	5.38e-02	5.38e-02	3.93e-02	3.66e-02
rs7647307	3	0.44	1.05e-03	1.33e-01	2.63e-03	1.48e-03	8.19e-04	5.94e-04	5.94e-04	4.02e-04	3.87e-04
rs7610017	3	0.03	4.80e-01	2.11e-02	4.28e-01	4.58e-01	4.93e-01	5.14e-01	5.14e-01	5.43e-01	5.47e-01
rs4692256	4	0.46	2.98e-03	2.87e-02	9.09e-02	5.69e-02	3.21e-02	2.19e-02	2.19e-02	1.02e-02	7.66e-03
rs6463843	7	0.45	6.27e-03	4.65e-02	1.49e-01	9.16e-02	5.11e-02	3.49e-02	3.49e-02	1.72e-02	1.36e-02

			SPREG						
SNP	chr	maf	*p*=0.10	*p*=0.13	*p*=0.16	*p*=0.20	*p*=0.23	*p*=0.27	*p*=0.30
rs429358	19	0.27	1.22e-09	2.21e-10	5.42e-11	1.19e-11	4.75e-12	1.83e-12	1.08e-12
rs2075650	19	0.24	1.44e-06	5.23e-07	2.30e-07	9.59e-08	5.67e-08	3.27e-08	2.40e-08
rs7526034	1	0.12	1.28e-01	1.12e-01	9.92e-02	8.66e-02	7.95e-02	7.24e-02	6.84e-02
rs10932886	2	0.33	6.24e-02	5.29e-02	4.67e-02	4.16e-02	3.92e-02	3.71e-02	3.61e-02
rs7647307	3	0.44	6.11e-03	4.02e-03	2.85e-03	1.97e-03	1.59e-03	1.27e-03	1.12e-03
rs7610017	3	0.03	2.15e-01	2.58e-01	2.97e-01	3.42e-01	3.70e-01	4.02e-01	4.22e-01
rs4692256	4	0.46	2.65e-03	2.55e-03	2.50e-03	2.47e-03	2.45e-03	2.43e-03	2.42e-03
rs6463843	7	0.45	6.28e-03	5.78e-03	5.51e-03	5.35e-03	5.32e-03	5.33e-03	5.36e-03

notes: maf based on the 324 subjects

Table 2

P-values for association testing between right hippocampus volume and each candidate SNP when subjects with MCI were included as controls in the ADNI data.

					lm-w						
SNP	chr	maf	unadj-lm	D-adj-lm	*p*=0.10	*p*=0.13	*p*=0.16	*p*=0.20	*p*=0.23	*p*=0.27	*p*=0.30
rs429358	19	0.31	8.12e-16	2.52e-11	2.22e-15	2.22e-16	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
rs2075650	19	0.26	1.21e-07	4.61e-05	1.31e-06	3.99e-07	1.46e-07	5.23e-08	5.23e-08	2.12e-08	2.00e-08
rs7526034	1	0.13	1.19e-03	4.71e-03	2.68e-03	2.29e-03	2.12e-03	2.16e-03	2.16e-03	2.97e-03	3.69e-03
rs10932886	2	0.33	1.26e-01	3.42e-01	2.31e-01	1.88e-01	1.58e-01	1.31e-01	1.31e-01	1.11e-01	1.09e-01
rs7647307	3	0.44	1.43e-03	1.89e-02	1.26e-03	1.05e-03	9.60e-04	9.87e-04	9.87e-04	1.44e-03	1.85e-03
rs7610017	3	0.04	9.36e-01	7.23e-01	7.49e-01	8.15e-01	8.87e-01	9.90e-01	9.90e-01	8.24e-01	7.46e-01
rs4692256	4	0.46	8.15e-04	4.52e-03	7.06e-03	3.94e-03	2.28e-03	1.18e-03	1.18e-03	4.74e-04	3.55e-04
rs6463843	7	0.47	7.87e-04	1.10e-03	1.14e-03	8.94e-04	7.60e-04	6.87e-04	6.87e-04	7.33e-04	8.03e-04

			SPREG						
SNP	chr	maf	*p*=0.10	*p*=0.13	*p*=0.16	*p*=0.20	*p*=0.23	*p*=0.27	*p*=0.30
rs429358	19	0.31	5.33e-15	1.55e-15	4.44e-16	2.22e-16	0.00e+00	0.00e+00	0.00e+00
rs2075650	19	0.26	6.98e-07	3.41e-07	1.90e-07	1.02e-07	7.03e-08	4.72e-08	3.72e-08
rs7526034	1	0.13	1.50e-03	1.33e-03	1.21e-03	1.11e-03	1.06e-03	1.00e-03	9.72e-04
rs10932886	2	0.33	1.82e-01	1.61e-01	1.46e-01	1.31e-01	1.22e-01	1.14e-01	1.09e-01
rs7647307	3	0.44	3.87e-03	2.74e-03	2.05e-03	1.49e-03	1.22e-03	9.93e-04	8.77e-04
rs7610017	3	0.04	8.54e-01	8.79e-01	9.00e-01	9.23e-01	9.37e-01	9.53e-01	9.63e-01
rs4692256	4	0.46	1.15e-03	9.67e-04	8.53e-04	7.59e-04	7.14e-04	6.73e-04	6.51e-04
rs6463843	7	0.47	6.67e-04	6.55e-04	6.57e-04	6.74e-04	6.94e-04	7.27e-04	7.55e-04

notes: maf based on the 635 subjects

Table 3

Simulation set-ups: parameter values used.

Set-up	SNP	maf	p	φ	β_{xy}	β_{Dx}	p-value	β_{Dy}	p-value
1	rs429358	0.27	0.23, 0.10	$0 \le \varphi \le 1$	−270	1.76	4.61e-16	−4.6e-04	2.29e-15
2	rs6463843	0.45	0.23, 0.10	$0 \le \varphi \le 2$	−106	0.61	8.66e-05	−3.3e-04	4.58e-08

Table 4

Empirical Type I error rates based on 10^4 simulations with sample size 324.

						lm-w			SPREG		
Set-up	***p***	***φ***	**cohort**	**unadj-lm**	**D-adj-lm**	$\hat{p}$ **− 0.05**	$\hat{p}$	$\hat{p}$ **+ 0.05**	$\hat{p}$ **− 0.05**	$\hat{p}$	$\hat{p}$ **+ 0.05**
1	0.23	0.050	0.0494	0.0565	0.1457	0.0573	0.0546	0.0550	0.0529	0.0541	0.0545
		0.010	0.0102	0.0101	0.0484	0.0120	0.0124	0.0128	0.0125	0.0117	0.0111
		0.005	0.0057	0.0051	0.0287	0.0056	0.0059	0.0059	0.0060	0.0058	0.0051
		0.001	0.0016	0.0009	0.0099	0.0014	0.0010	0.0009	0.0011	0.0008	0.0009
	0.10	0.050	0.0498	0.0747	0.1103	0.0671	0.0614	0.0621	0.0626	0.0561	0.0601
		0.010	0.0114	0.0188	0.0328	0.0166	0.0133	0.0131	0.0149	0.0124	0.0139
		0.005	0.0059	0.0109	0.0198	0.0078	0.0065	0.0076	0.0083	0.0062	0.0072
		0.001	0.0010	0.0030	0.0055	0.0017	0.0012	0.0015	0.0017	0.0014	0.0016
2	0.23	0.050	0.0492	0.0482	0.0837	0.0534	0.0543	0.0544	0.0540	0.0538	0.0532
		0.010	0.0091	0.0104	0.0221	0.0116	0.0107	0.0113	0.0109	0.0109	0.0111
		0.005	0.0051	0.0044	0.0114	0.0064	0.0057	0.0057	0.0061	0.0058	0.0056
		0.001	0.0009	0.0006	0.0028	0.0011	0.0014	0.0015	0.0010	0.0009	0.0009
	0.10	0.050	0.0492	0.0537	0.0698	0.0549	0.0549	0.0560	0.0562	0.0548	0.0557
		0.010	0.0085	0.0115	0.0177	0.0129	0.0128	0.0126	0.0132	0.0130	0.0130
		0.005	0.0048	0.0063	0.0113	0.0067	0.0070	0.0072	0.0070	0.0071	0.0071
		0.001	0.0013	0.0012	0.0034	0.0014	0.0018	0.0016	0.0023	0.0020	0.0019

Table 5

Empirical power based on 10^4 simulations with sample size 324.

						lm-w			SPREG		
Set-up	***p***	***φ***	**cohort**	**unadj-lm**	**D-adj-lm**	$\hat{p}$ **− 0.05**	$\hat{p}$	$\hat{p}$ **+ 0.05**	$\hat{p}$ **− 0.05**	$\hat{p}$	$\hat{p}$ **+ 0.05**
1	0.23	0.10	0.1240	0.1029	0.0722	0.0712	0.0829	0.0937	0.0775	0.0861	0.0949
		0.30	0.6567	0.3969	0.0676	0.2571	0.3078	0.3520	0.3172	0.3528	0.3749
		0.50	0.9744	0.7829	0.2574	0.6069	0.6811	0.7328	0.7138	0.7444	0.7654
		0.70	0.9996	0.9622	0.5924	0.8747	0.9184	0.9437	0.9414	0.9514	0.9567
		1.00	1.0000	0.9993	0.9377	0.9931	0.9976	0.9985	0.9987	0.9990	0.9992
	0.10	0.10	0.1723	0.1690	0.0619	0.0694	0.0878	0.1062	0.0708	0.0902	0.1129
		0.30	0.8540	0.5470	0.1088	0.1969	0.2744	0.3533	0.2937	0.3803	0.4381
		0.50	0.9982	0.8794	0.3607	0.4688	0.6018	0.7119	0.6895	0.7729	0.8166
		0.70	1.0000	0.9876	0.7083	0.7628	0.8729	0.9270	0.9340	0.9638	0.9757
		1.00	1.0000	1.0000	0.9702	0.9659	0.9920	0.9980	0.9963	0.9996	0.9999
2	0.23	0.30	0.1577	0.0961	0.1874	0.1066	0.1025	0.0999	0.1101	0.1057	0.1026
		0.60	0.4817	0.2462	0.3570	0.2642	0.2639	0.2611	0.2757	0.2664	0.2606
		1.00	0.8815	0.5688	0.6262	0.5698	0.5784	0.5816	0.6000	0.5900	0.5844
		1.50	0.9960	0.8877	0.8827	0.8860	0.8917	0.8950	0.9020	0.8973	0.8942
		2.00	1.0000	0.9887	0.9770	0.9865	0.9882	0.9888	0.9897	0.9896	0.9895
	0.10	0.30	0.2020	0.0847	0.1707	0.1028	0.0975	0.0949	0.1144	0.1062	0.1001
		0.60	0.5938	0.2350	0.3429	0.2387	0.2430	0.2448	0.2899	0.2763	0.2662
		1.00	0.9433	0.5642	0.6155	0.5241	0.5417	0.5571	0.6216	0.6090	0.5990
		1.50	0.9992	0.8925	0.8833	0.8533	0.8719	0.8865	0.9169	0.9117	0.9073
		2.00	1.0000	0.9909	0.9787	0.9800	0.9857	0.9890	0.9929	0.9927	0.9925