On the adjustment for covariates in genetic association analysis: A novel, simple principle to infer direct causal effects

Stijn Vansteelandt, Ghent University
Sylvie Goetgeluk, Ghent University
Sharon Lutz, Harvard School of Public Health
Irwin Waldman, Emory University
Helen Lyon, Children’s Hospital Boston
Eric E. Schadt, Rosetta Inpharmatics LLC
Scott T. Weiss, Harvard School of Public Health
Christoph Lange, Harvard School of Public Health

Journal Title: Genetic Epidemiology
Volume: Volume 33, Number 5
Publisher: Wiley | 2009-06-12, Pages 394-405
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1002/gepi.20393
Permanent URL: https://pid.emory.edu/ark:/25593/s5b2f

Final published version: http://dx.doi.org/10.1002/gepi.20393

Copyright information:
© 2009 Wiley-Liss, Inc.

Accessed October 15, 2019 10:54 AM EDT
On the Adjustment for Covariates in Genetic Association Analysis: A Novel, Simple Principle to Infer Direct Causal Effects

Stijn Vansteelandt, Sylvie Goetgeluk, Sharon Lutz, Irwin Waldman, Helen Lyon, Eric E. Schadt, Scott T. Weiss, and Christoph Lange

1Department of Applied Mathematics and Computer Sciences, Ghent University, Ghent, Belgium
2Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts
3Channing Laboratory, Harvard Medical School, Boston, Massachusetts
4Department of Psychology, Emory University, Atlanta, Georgia
5Children’s Hospital Boston, Boston, Massachusetts
6Rosetta Inpharmatics, LLC, Seattle, Washington

Abstract

In genetic association studies, different complex phenotypes are often associated with the same marker. Such associations can be indicative of pleiotropy (i.e. common genetic causes), of indirect genetic effects via one of these phenotypes, or can be solely attributable to non-genetic/environmental links between the traits. To identify the phenotypes with the inducing genetic association, statistical methodology is needed that is able to distinguish between the different causes of the genetic associations. Here, we propose a simple, general adjustment principle that can be incorporated into many standard genetic association tests which are then able to infer whether an SNP has a direct biological influence on a given trait other than through the SNP’s influence on another correlated phenotype. Using simulation studies, we show that, in the presence of a non-marker related link between phenotypes, standard association tests without the proposed adjustment can be biased. In contrast to that, the proposed methodology remains unbiased. Its achieved power levels are identical to those of standard adjustment methods, making the adjustment principle universally applicable in genetic association studies. The principle is illustrated by an application to three genome-wide association analyses.

Keywords
causal diagram; direct effect; genetic pathways; mediation; pleiotropy

INTRODUCTION

It is a well-studied phenomenon that the findings of genetic association studies can be confounded and biased by genetic and/or phenotypic heterogeneity which is not accounted...
for in the statistical analysis. Consequently, much effort has been devoted to the
development of statistical analysis techniques that minimize the impact of such effects, in
particular to approaches to control for population admixture and stratification [Pritchard and
Rosenberg, 1999; Devlin and Roeder, 1999; Price et al., 2006; Epstein et al., 2007].
However, relatively little is known and understood about situations in which a genetic
association with the disease phenotype is confounded by an association between the same
marker and another phenotype [Smoller et al., 2000; Robins et al., 2001]. If the two
phenotypes are linked other than just through the tested marker locus, a true association with
one phenotype can cause an association between the same marker locus and the other
phenotype, despite the lack of an independent, direct genetic effect on that phenotype and
thus even in the absence of pleiotropy. Such situations can be encountered in genome-wide
association studies (GWAS), in pathway analysis of candidate genes, and in association
studies that incorporate genomic information into the association analysis.

Recent GWAS revealed strong associations between the same SNPs and different
phenotypes [Hung et al., 2008; Amos et al., Thorgerisson et al., 2008; Bierut et al., 2007;
Frayling et al., 2007]. Scans for lung cancer showed highly significant associations with a
set of SNPs [Hung et al., 2008; Amos et al.]. In an independent study, associations between
the same SNPs and smoking behavioral phenotypes reached also genome-wide significance
[Thorgerisson et al., 2008]. Given the strong non-genetic link between both phenotypes,
lung cancer and smoking behavior, there is currently no conclusive evidence whether the
identified SNPs describe a lung cancer gene or a smoking-behavior gene [Chanock and
Hunter, 2008]. A similar situation occurred in GWAS for obesity [Frayling et al., 2007;
Scuteri et al., 2007], in which the strongest association for body-mass index is observed for
an SNP in the FTO gene. The SNP was also identified as one of the “top hits” by a GWAS
for smoking-behavior [Bierut et al., 2007]. Again, given the non-genetic link between both
phenotypes [Li et al., 2003; Zimlichman et al., 2005], the identification of the causative
association is not straightforward. In view of these quandaries, our goal in this article is to
infer whether a given marker is (causally) associated with the disease phenotype other than
through its association with a secondary, intermediate phenotype.

A standard epidemiological approach to remove the effect of a secondary phenotype/
covariate on the target phenotype is to regress the target phenotype on the secondary
phenotype (and possibly also other influencing factors/covariates) and to use the
corresponding residuals as the phenotype of interest in the association analysis. An
alternative, common approach is to test whether the SNP is associated with the target
phenotype after adjusting for the other phenotype. Both approaches are also used to reduce
variability and can be fallible. For the first approach, this is partly because by removing the
association between both phenotypes (via residuals), one risks to remove also part of the
effect of the SNP on the target phenotype (see the section “The Limitation of Standard
Analysis Approaches for the Detection of Direct Genetic Effects”). For the second approach
this is because adjustment for phenotypes, or more generally covariates, is only valid under
the assumption that the influencing factors/covariates are not associated with the marker
locus [Rosenbaum, 1984; Cole and Hernan, 2002]. In particular in the context of GWAS
which interrogate the entire human genome, this assumption is generally untenable, e.g., the
adjustment for physiological factors with relatively high heritability such as height [Carmichael and McGue, 1995; Hirschhorn et al., 2001].

A conceptually different situation is encountered in genetic association studies that incorporate genomic data, e.g., expression profiles [Schadt et al., 2005]. Here, associations between the same marker and both traits, the expression profile and the disease phenotype of interest, can be especially informative. If one is able to conclude that the observed association between the marker and the disease phenotype is caused by the association with the expression profile, this provides stronger evidence for the validity of a finding than an association with just the disease phenotype.

Common to all scenarios that were discussed above, is that the non-marker related link between the different phenotypes can bias the results of the association analysis. Specifically, an association between a marker and a particular phenotype can cause an apparent association between the same marker and another phenotype if there is a causal link between both phenotypes. In order to identify the true genetic pathways of complex diseases, it will be crucial to have the ability to distinguish whether an observed genetic association with a particular phenotype is attributable to a link with another phenotype that is itself associated with the marker, or whether the observed association is independent of that, thus indicating a direct causal genetic relationship between the marker locus and the examined phenotype.

To address this problem, we propose a simple, general principle to adjust the phenotype of interest for a potential association between the examined marker and another, related phenotype. Using causal inference methodology, we develop an adjustment approach that is applicable to both binary and quantitative traits, and that is computationally simple. The adjusted phenotype can be incorporated into many standard genetic association tests which then test for a direct genetic effect between the phenotype of interest and the marker (other than through the given intermediate phenotype). The principle is outlined first by application to quantitative phenotypes both in population-based designs and in family-based studies. The integration of the approach for binary traits, e.g., association studies of unrelated cases and controls in cohort studies, is also discussed.

Using both theoretical considerations and simulation studies, we demonstrate that, in the presence of a non-genetic link between the phenotype of interest and the other phenotype, association tests with adjustment procedures that are based on intuitive regression approaches can be biased and provide incorrect results. Further, the simulation studies show that the proposed adjustment principle is sufficiently powered to detect direct genetic effects for realistic sample sizes. When the adjustment principle is applied in situations of no relationship between both phenotypes, the adjusted association test still achieves power levels that are virtually identical to those of standard regression approaches, making the proposed principle a generally applicable approach for covariate-adjustment in genetic association studies.

As a real data example, we present results from a 100 k-scan in the Framingham Heart Study that suggested an association between an SNP and two phenotypes, FEV$_1$ and BMI, which
was detected using a standard regression-based covariate adjustment. The application of the proposed adjustment method suggests that the “true” effect of this SNP may have fully originated from its association with BMI. This conclusion is then confirmed by replication analysis in two independent GWAs, the CAMP study [C. A. M. P. R. Group, 1999] and the British Birth Cohort.

THE LIMITATION OF STANDARD ANALYSIS APPROACHES FOR THE DETECTION OF DIRECT GENETIC EFFECTS

Suppose that in the study of interest, \( n \) subjects have been genotyped at a specified marker locus and their coded genotype is denoted by \( X_i, i = 1, \ldots, n \). If the selected sample is a family-based study, we further assume that additional genotype data on other family members are available so that the expected marker score, \( E(X_i|S_i) \), can be computed conditional on Mendelian transmissions. When parental data are available, the variable \( S_i \) denotes the parental genotypes; otherwise it represents the sufficient statistic by Rabinowitz and Laird [Rabinowitz and Laird, 2000].

Two phenotypes, \( K_i \) and \( \Upsilon_i \), have been recorded for the \( i \)th subject, where \( \Upsilon_i \) is the target phenotype for the primary association analysis of the study. The phenotype \( K_i \) is of secondary interest in the study, e.g., an endo/intermediate phenotype or covariate for the target phenotype \( \Upsilon_i \). Both phenotypes may be influenced by shared, non-genetic confounding variables and may thus be correlated. The phenotype \( K_i \) has been tested for association with the marker locus, and a significant association has been observed. Now, given the established association between the marker locus and the secondary phenotype \( K_i \), our goal is to test for an association between the target phenotype \( \Upsilon_i \) and the marker locus that cannot be explained by the existing genetic association with phenotype \( K_i \) and the correlation between both phenotypes.

To understand the problems of standard intuitive regression approaches when they are applied to adjust for confounding in the context of genetic association studies, we use causal diagrams [Pearl, 1995; Robins, 2001]. These postulate the causal relationships between all measurements of interest by means of directed arrows, as in Figure 1. If the study of interest has a family-based design, the variable \( S \) encodes the parental genotypes. For study designs with unrelated probands, \( S \) denotes measured factors inducing population admixture. Further, \( L \) is a collection of measured prognostic factors for the secondary phenotype \( K \). As we will see later, the presence of \( S \) allows for the association between the marker \( X \) and both phenotypes to be confounded by measured factors. Likewise, the presence of \( L \) allows for the association between both phenotypes to be confounded by measured factors.

For the diagram in Figure 1 to infer causal genetic effects, it must satisfy the following two basic assumptions:

1. The absence of an arrow between any two variables \( A \) and \( B \) encodes the assumption that \( A \) exercises no direct causal effect on \( B \). Figure 1, thus, postulates that the common confounders \( S \) and \( L \) cannot affect the target phenotype \( \Upsilon \) other than by affecting the marker \( X \) and phenotype \( K \), respectively.
We will relax this restriction in the Appendix. Figure 1 further postulates that the target phenotype $\Upsilon$ is affected by the intermediate phenotype $K$ and not the other way around. We will elaborate on this restriction in the “Discussion” section.

2. The diagram includes all variables that jointly affect any two variables in the diagram. Figure 1, thus, postulates that all risk factors for the secondary phenotype $K$, which are also associated with the target phenotype $\Upsilon$, have been measured and are contained in $L$. Note that, to add realism, we allow for the relationship between the primary phenotype and the confounding factors $L$ and $S$ to be partly non-causal by means of the unmeasured common cause $U$.

Two variables in a causal diagram may be statistically associated along all paths that have no converging arrows. That is, they may be associated along all unbroken sequences of edges between those two variables, disregarding the direction of the arrows, in which there are no so-called colliders (i.e., nodes in which two arrows converge, or, point to each other) [Pearl, 1995; Robins, 2001]. Specifically, the genotype $X$ and primary phenotype $T$ in Figure 1 may be associated for three reasons: (1) because of a direct genetic effect (i.e., along the path $X \rightarrow T$); (2) because of an indirect genetic effect (i.e., along the paths $X \rightarrow K \rightarrow T$ and $X \rightarrow L \rightarrow K \rightarrow T$) and (3) because of population admixture (i.e., along the paths $X \leftarrow S \leftarrow U \rightarrow T$ and $X \leftarrow S \leftarrow U \rightarrow L \rightarrow K \rightarrow T$). Importantly, note that when the target phenotype $T$ is not directly genetically affected, then (in the absence of population admixture) no association can be detected between the genotype and the target phenotype, unless there is a causal link between both phenotypes $T$ and $K$ (i.e., unless $K$ causally affects $T$). A non-causal relationship between both phenotypes will not induce an association between the genotype and the target phenotype because the converging arrows along the paths $X \rightarrow K \leftarrow L \leftarrow U \leftarrow T$ and $X \rightarrow L \leftarrow U \rightarrow T$ transmit no association.

We will now use the causal diagram in Figure 1 to gain insight into the lack of validity of common regression approaches to test the hypothesis whether the SNP directly affects the target phenotype $\Upsilon$ other than through the intermediate phenotype $K$. A first common approach is to eliminate the effects of the established association with the other phenotype $K_i$ by regressing the target phenotype $\Upsilon_i$ on $K_i$ and using the residuals of $\Upsilon_i$ as the new phenotype in the association test. This approach has the disadvantage that the residuals remove the overall association between both phenotypes, which mixes the effect of the intermediate phenotype on the target trait (i.e., along the path $K \rightarrow \Upsilon$), with spurious (i.e., non-causal) associations through the SNP $X$ (i.e., along the paths $K(\leftarrow L) \leftarrow X \rightarrow T$ and $K(\leftarrow L) \leftarrow X \leftarrow S \leftarrow U \rightarrow T$) and by spurious association other than through the SNP $X$ (i.e., along the path $K \leftarrow L \leftarrow U \rightarrow T$). By not solely removing the causal effect of the intermediate phenotype on the target trait, these residuals may be associated with the SNP, even when it has no influence on the target trait. In particular, suppose that the SNP directly influences $K$, but not $T$, and that $K$ has no effect on $T$. Then the SNP has neither a direct, nor an indirect effect on $T$. Nonetheless, the residual, say $\Upsilon - \gamma K$, will have $\gamma \neq 0$ because $T$ is spuriously associated with $K$ along the path $K \leftarrow L \leftarrow U \rightarrow T$, and will be associated with the SNP by the fact that the phenotype $K$ is influenced by it.

An alternative common approach to test the hypothesis that the SNP has a direct influence on the target phenotype other than through the phenotype $K$, is to measure the association...
between the genotype $X$ and the target phenotype $ϒ$ conditional on $K$ (e.g., by regressing $ϒ$ on $X$ and $K$). To appreciate the impact of such adjustment for the phenotype $K$, note that the adjustment for a variable $K$ on a path between two variables $X$ and $ϒ$ in the causal diagram blocks (i.e., removes) the association between those variables along that path [Pearl, 1995; Robins, 2001]. This is true except when $K$ is a collider or a descendant of a collider along that path, in which case an association is induced [Pearl, 1995; Robins, 2001]. Stratification of the analysis on the secondary phenotype $K$, thus, removes the indirect effect of the marker $X$ on the target phenotype $ϒ$ by blocking the path $X \rightarrow K \rightarrow ϒ$, but at the same time induces a spurious association along the path $X \rightarrow K \leftarrow L \leftarrow U \rightarrow ϒ$ because the phenotype $K$ is a collider along that path. Additional adjustment for the confounders $L$ blocks this association, but induces a new, non-causal association along the path $X \rightarrow L \leftarrow U \rightarrow ϒ$, unless the confounders $L$ are not affected by the marker $X$.

In summary, traditional approaches for estimating/testing direct genetic effects, either based on an analysis of residuals or based on ordinary regression adjustment, may yield biased inferences for direct genetic effects whenever the association between the intermediate and target phenotype is confounded by a non-genetic link between the two phenotypes. Traditional regression adjustment for measured confounders remains problematic when (some of) these confounders are themselves influenced by the target SNP. We will later verify these theoretical considerations by simulation experiments. In the next section, we propose an alternative adjustment principle which is valid even when confounders/covariates in a genetic association analysis are genetically affected by the same genetic locus.

A GENERAL PRINCIPLE TO TEST FOR CAUSAL DIRECT GENETIC EFFECTS

For simplicity, we illustrate the principle first for the association analysis of quantitative traits, either in population-based designs or in family-based designs. For both scenarios, i.e., population-based studies and family-based studies, we assume that the selected association test $T$ (e.g., a standard score test, Wald test or likelihood ratio test) has the general form

$$T = \sum_{i=1}^{n} T_i,$$

(1)

where $T_i$ denotes the contribution of the $i$th subject to the test statistic. To keep the notation simple and without loss of generality, we assume that, under the null hypothesis of no association between the selected phenotype and the marker locus, the expected value of the test statistic $T$ is 0, i.e., $E(T) = 0$.

In order to derive a valid association test for the direct association between phenotype $ϒ$ and the marker locus that is not influenced by the existing association between the phenotype $K$ and the same disease-specific locus (DSL), we have to assess/estimate the effect of $K$ on the target phenotype $ϒ$. Inferring this requires knowledge on common risk factors of both phenotypes, as this allows for disentangling a spurious association between both traits from a real effect. A test for a direct genetic effect between the DSL and the target-phenotype $ϒ_i$
will, therefore, require measurements on all shared, non-genetic factors of both phenotypes. As we will see in the simulation study, this assumption applies only to the major non-genetic factors of both phenotypes, which are often known.

For simplicity, we assume here that the target phenotypes $T_i$ and the phenotype $K_i$ share a common risk factor $L_i$. Then, in a population-based study, the following linear model can be used to assess how the phenotype $K_i$ influences the target phenotype $T_i$

$$E(Y_i) = \gamma_0 + \gamma_1 K_i + \gamma_2 X_i + \gamma_3 L_i. \quad (2)$$

In a family-based study, the expected marker-score, $E(X_i|S_i)$, would be added to the model to maintain robustness against population admixture (see the Appendix), i.e.,

$$E(Y_i) = \gamma_0 + \gamma_1 K_i + \gamma_2 X_i + \gamma_3 L_i + \gamma_4 E(X_i|S_i). \quad (3)$$

In both equations, $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ and $\gamma_4$ denote the mean parameters. Importantly, note that models (2) and (3) include the offspring genotype $X_i$ as well as the common risk factor $L_i$ of both phenotypes, in order to ensure that $g_1$ represents the true effect of $K_i$ on $T_i$ and no spurious association. This will guarantee later during the computation of the adjusted phenotypes/residuals that only the effect of $K_i$ is removed from the target phenotype, but a potential direct association between $T_i$ and the marker locus is maintained. Indeed, it follows upon applying the principles outlined in the previous section to the causal diagram in Figure 1, that adjustment for $X_i$, $L_i$ and $S_i$ blocks all non-causal paths from $K_i$ to $T_i$ and thus reveals the causal effect of $K_i$ on $T_i$.

Using ordinary least squares to estimate all parameters in model (2) or (3), the target-phenotype $T_i$ can be adjusted for just the effect that the phenotype $K_i$ has on the target phenotype $T_i$

$$\tilde{T}_i = T_i - \bar{y} - \hat{\gamma}_1 (K_i - \bar{k}), \quad (4)$$

where $\hat{\gamma}_1$ is the ordinary least squares estimate for $g_1$ in model (2) or (3). The quantities $\bar{y}$ and $\bar{k}$ are the observed phenotypic means of $T$ and $K$, respectively, in the sample. The phenotype adjustment (4) here deliberately only involves the other phenotype $K_i$, but not the shared risk factor $L_i$. Although this may seem counterintuitive and contrary to standard practice, including factors such as $L_i$ in the phenotypic adjustment would introduce bias if the common risk factor $L_i$ is itself associated with the DSL. As shown in the Appendix, if the residuals $\tilde{T}_i$ are computed based on the adjustment formula (4), this problem is avoided.

Using now the adjusted phenotype $\tilde{T}_i$ as the target phenotype, we can construct standard association tests for quantitative traits in either population-based or family-based designs. For example, in a population-based setting, the adjusted phenotype $\tilde{T}_i$ can be tested for
association with a standard regression approach, e.g. each subject’s contribution to the test statistic is given by

\[ T_i = X_i \tilde{T}_i. \quad (5) \]

For family-based studies, we can construct an association test based on a standard FBAT statistic [Laird et al., 2000] by defining the contribution of each offspring to be

\[ T_i = \{ X_i - E(X_i|S_i) \} \tilde{T}_i. \quad (6) \]

Both association tests, (5) and (6), will then test for a direct association between the DSL and the target-phenotype \( \Upsilon_i \) (where “direct” means “other than through its association with the phenotype \( K_i \)”). More generally, as we show in the Appendix, any association test statistic \( T \) which is linear in the phenotype and which selects the adjusted phenotype \( \tilde{\Upsilon}_i \) as the target phenotype will provide a valid and robust test for the null hypothesis that there is no direct effect between the target phenotype \( T_i \) and the DSL, i.e., that the apparent association between the target phenotype \( T_i \) and the DSL is solely attributable to the established association between the DSL and the other phenotype \( K_i \). Under this condition, the expected value of the test statistic \( T \) will be zero, \( E(T) = 0 \) when \( T \) is computed based on the adjusted phenotype (defined in equation (4)). Further, for family-based tests, we show that, under the null hypothesis of no direct effect, the modified FBAT-statistic (6) remains robust against confounding due to population admixture, provided that population admixture on both phenotypes is w.r.t. unrelated causes (since there would otherwise exist unmeasured common causes of both phenotypes).

The proposed phenotype adjustment (equation (4)) of the association test \( T \) for a direct genetic effect includes a parameter estimate for \( \gamma_1 \) that is obtained by fitting model (2) or (3). Since the imprecision of the estimate for \( g_1 \) must be acknowledged in the computation of the asymptotic variance of the test statistic, the standard variance of the selected association test is not applicable here anymore. In the Appendix, we show that the standardized association test statistic for the adjusted phenotype \( T_i \) follows a \( \chi^2 \) distribution with 1 degree of freedom under the null hypothesis of no direct effect, where the variance of the test statistic, \( \Sigma \), is given by

\[ \sum = \text{Var}(\tilde{T}_i) \]

with

\[ \tilde{T}_i = T_i - E[T_i^2, K_i] \left( K_i - \mu^{(i)} \right) \frac{\sigma^2}{\sigma^2_K} \epsilon_i, \]
where \( T_i \) denotes the contribution of the \( i \)th subject to the association test statistic and \( T_i' \) the first order derivative of \( T_i \) w.r.t. \( \hat{\gamma}_i \) (e.g., for population-based tests, we have \( T_i = X_i \) in (5) and, for family-based tests, \( T_i = X_i - E(X_i|S_i) \) in (6)). The variable \( e_i \) is the residual in model (2) or (3). In population-based designs, the parameters \( \mu_K \) and \( \sigma_K^2 \) are obtained by fitting a linear regression for \( K_i \) with the covariates \( L_i \) and \( X_i \). For family-based studies, the covariate \( E(X_i|S_i) \) has to be included well. The predicted value for \( K_i \) is then defined by \( \mu_K^{(i)} = E(K_i|L_i, X_i) \) or by \( \mu_K^{(i)} = E(K_i|L_i, E(X_i|S_i)) \). The residual variance in the model is denoted by \( \sigma_K^2 \).

When the phenotype of interest \( T_j \) is dichotomous, e.g., affection status, the proposed adjustment can be extended provided that a relative risk model and a log-link function is assumed. The technical details are discussed in the Appendix.

**DATA ANALYSIS: AN APPLICATION TO THE FRAMINGHAM HEART STUDY, THE BRITISH BIRTH COHORT AND THE CAMP STUDY**

We evaluated the practical relevance of the proposed adjustment principle by an application to three GWAS: a 100 K Affymetrix scan in the family-plates of the Framingham Heart Study (1,400 probands) [Herbert et al., 2006], a 550 K Illumina scan in the British Birth Cohort (genotype data on 1,430 probands, http://www.b58cgene.sgu.ac.uk/) and a 550 K Illumina scan in 440 trios of the CAMP study [C. A. M. P. R. Group, 1999]. As target phenotype, we selected the lung-function measurement FEV1, which was available in all three studies. Since the three studies were genotyped on different platforms (the Framingham Heart Study on 100 K Affymetrix, the British Birth Cohort and CAMP on Illumina 550 K), we selected the 32,121 SNPs for the analysis that are common among both platforms.

As the first step, we analyzed FEV1 at exam 1 in the family-plates of the Framingham Heart Study. Using a standard regression approach, we adjusted FEV1 for height [Bierut et al., 2007], gender, weight and age, and then used the residuals as the target phenotype in the analysis. All statistical analysis was conducted under an additive mode of inheritance. Since the Framingham Heart Study is a family-based study, we applied the weighted Bonferroni-testing strategy by Ionita-Laza et al. [2007] [Van Steen et al., 2005; Laird and Lange, 2006]. Based on the conditional mean model approach [Lange et al., 2003], the testing strategy evaluates the evidence for association at a population level and then estimates the conditional power of the FBAT-statistic for each marker in the first step. In the second step of the testing strategy, FBAT-statistics are computed for all markers. Their significance is assessed based on individually adjusted \( \alpha \)-levels that maintain the overall type-1 error and that are weighted based on the conditional power estimate for the corresponding marker.

When the weighted Bonferroni approach was applied to the 32,121 SNPs that are on both genotyping platforms, none of the SNPs reached genome-wide significance. However, the SNP (rs2415815) with the highest conditional power estimate had an unadjusted FBAT \( P \)-value of 0.023, warranting additional analysis. When the covariates were tested for association with rs2415815, an association with weight was observed (\( P \)-value of 0.0054 for...
Both associations between the SNP and the two phenotypes were then verified in the CAMP study and the British Birth Cohort (Table I). For weight, SNP rs2415815 has nominal significant $P$-values in CAMP and the British Birth Cohort (measured as BMI), while the association tests with FEV1 are not significant in either studies. Given the established non-genetic link between FEV1 and weight [Oliveti et al., 2006; Yuan et al., 2002; Gessner and Chimonas, 2007; Sin et al., 2004; Demissie et al., 1998; Taveras et al., 2006; Camargo Jr et al., 1999] and the inconsistent replication results, both phenotypes, FEV1 and weight, were re-analyzed with the proposed adjustment procedure in CAMP and FHS. For the British Birth Cohort, we did not have access to the raw data and the proposed adjustment could not be computed.

The lung-function measurement FEV1 (i.e., $T$) was adjusted for its covariates gender, age and height (i.e., $L$) and for a potential association with the SNP (i.e., $X$), the expected marker score given the parental genotypes (i.e., $B(X|S)$) and weight (i.e., $K$) (see Fig. 1). Using the estimated association with body weight of 0.118 (SE 0.050) in the FHS, an adjusted phenotype was calculated as $T - \bar{y} - 0.118(K - \bar{K})$. Based on this adjusted phenotype, the FBAT-statistics in FHS (and similarly in CAMP) were re-calculated (Table I). The association tests for FEV1 in FHS and in CAMP were no longer significant (1), suggesting a 1.34 (95% confidence interval: −2.94–0.27) unit reduction in expected FEV1 per unit increase in marker score when body weight is controlled at a fixed level for all.

Because we cannot rule out the possibility of causal effects of FEV1 on body weight, we repeated the analysis with the roles of FEV1 and body weight reversed. In these analyses, the associations with weight remained significant. In the Appendix we show that, if the causal diagram underlying this analysis is wrong because body weight affects FEV1 and not vice versa, then the resulting test can be interpreted as an overall test for genetic association with body weight provided that FEV1 is not directly affected by the studied SNP.

In summary, our results suggest that the originally observed association with FEV1 in the FHS was likely attributable to the association with weight. More precisely, we found no evidence of a direct genetic effect of rs2415815 on FEV1 other than through its association with weight.

**SIMULATION STUDY**

Using simulation studies, we assess the type-1 error, the power and the robustness of the new approach and compare it to the two standard approaches earlier described. The new principle is evaluated under various conditions, including scenarios in which there are unmeasured risk factors that are common for both phenotypes and that are not included in the adjustment formula. In all simulations, we focus on quantitative traits and assume that there is no ascertainment condition. A sample size of 1,000 probands is selected. All simulation results that are presented in this article are based on 5,000 replicates. With the data analysis example in the Framingham Heart Study in mind, the phenotype of interest $T$ is simulated so that it resembles the FEV1 phenotype in that application. The second phenotype $K$ is weight and the set of common confounding variables is given by height and age, which are denoted by $L$ and $S$, respectively. First the genotype data are generated by drawing from a Binomial distribution with the specified marker allele frequency. Using the genotype data, all
phenotypic variables are simulated from normal distributions under the causal diagrams of Figure 2, with phenotypic means and variances that were observed in the application to the Framingham Heart Study. In addition, unless otherwise specified, effect sizes were also chosen to match the data application.

Five tests are evaluated each time. The first two test the association between the target phenotype \( \Upsilon \) and the marker locus with standard Wald tests that are applied to the adjusted phenotypic residuals, where the residuals are obtained from linear regressions adjusting for either \((S, K)\) or \((S, K, L)\). The next two test this association with standard Wald tests that are applied to the phenotype itself, adjusting for either \((S, K)\) or \((S, K, L)\). Finally, we evaluate the proposed adjustment principle.

**EMPIRICAL SIGNIFICANCE LEVEL**

The first set of simulation studies is conducted under the null hypothesis of no direct genetic effect on the target phenotype \( \Upsilon \). In order to assess the robustness of the adjustment principle against spurious associations, we consider the following scenarios corresponding to the causal diagrams in Figure 2:

- In the first scenario, we assume that there is a direct genetic effect of the marker on the intermediate phenotype \( K \) and on the common covariate \( L \). Each genetic effect has a locus specific heritability of 1%. The intermediate phenotype \( K \) explains 1% of the phenotypic variation in \( \Upsilon \), creating an association between the SNP and \( \Upsilon \).

- In the second simulation experiment, the first scenario is modified so that there is no genetic effect on the confounder \( L \), but the genetic association with the intermediate phenotype \( K \) is still present. Under these conditions, a standard adjustment for the covariate \( L \) will provide correct results for the Wald test.

- The third simulation experiment varies from the first scenario with respect to the association between the phenotypes \( \Upsilon \) and \( K \). While the common confounder \( L \) is still associated with the marker locus in the third scenario, there is no link anymore between the intermediate phenotype \( K \) and the target phenotype \( \Upsilon \), making an adjustment for the intermediate phenotype \( K \) in the association analysis unnecessary.

- In the fourth simulation experiment, the second scenario is modified so that there is also no genetic effect on the intermediate phenotype \( K \).

For a variety of allele frequencies, the estimated nominal significance levels are shown in Table II for the five different tests (performed at the 5% significance level), along with the empirical bias and standard error of the proposed direct effect estimate.

The Wald test based on the new adjustment principle maintains the specified significance level well in all four scenarios and throughout the entire range of allele frequencies. This is true regardless of whether the intermediate phenotype affects the target phenotype \( \Upsilon \) or not, and of whether the confounders \( L \) for the association between both phenotypes are genetically affected. However, Wald tests that are based on the standard adjustments for \((S, K, L)\)
$K, L$) or $(S, K)$ generally fail to preserve the theoretical $\alpha$-level, with the exception of Scenarios II and IV. In these scenarios, the Wald test based on the standard adjustment for the confounding variables $(S, K, L)$ maintains the significance level because $L$ is not genetically affected in these scenarios. In Scenario IV, Wald tests that are based on residuals, adjusting for $(S, K, L)$ or $(S, K)$, additionally maintain the significance level because $\Upsilon$ and $K$, and thus the residuals, are jointly independent of the genotype in this scenario, after adjustment for $S$. This is also true in Scenario II where there are only genetic effects on $K$, which are so weak that they do not induce bias of a detectable magnitude. With stronger genetic effects on $K$, also the approaches ignoring $L$ would be biased. While, as in these two scenarios, there are instances in which standard adjustment provides correct $\alpha$-levels for the Wald tests, this can only be achieved if the underlying genetic architecture is known and all necessary confounding variables are included in the standard adjustment. The proposed adjustment principle requires less prior knowledge about potential links and genetic associations between the phenotypes and the covariates, and maintains the significance level in all considered scenarios.

ESTIMATED STATISTICAL POWER

To assess whether the Wald tests based on the new adjustment principle have sufficient power to detect genetic effects of realistic magnitudes, we repeat the simulation study under the assumption that, in all the four scenarios, there is a direct genetic effect of the marker locus on the target phenotype $\Upsilon$. The locus-specific heritability of the genetic effect is specified to be 0.33%. The estimated power levels for the Wald tests based on the proposed adjustment principle are displayed in Table III. In general, we find high power levels in all simulation experiments, except, as usual, at low allele frequencies (<0.10).

Comparison of the attained power levels of the new adjustment principle is especially relevant vis-a-vis the $(S, K, L)$-adjustment in Scenario II and vis-a-vis all other tests in Scenario IV. In these scenarios, these other approaches are also valid and are expected to yield higher power levels by the fact that they involve stronger assumptions. Interestingly, these standard approaches provide power levels that are essentially identical to those of the proposed adjustment principle. Since these scenarios are essentially the best case scenarios for the standard approaches, these results illustrate the potential of the proposed adjustment principle as a generally applicable tool in genetic association studies.

ROBUSTNESS OF THE ADJUSTMENT PRINCIPLE WHEN COMMON CONFOUNDING VARIABLES OF BOTH PHENOTYPES ARE NOT INCLUDED IN THE ADJUSTMENT PRINCIPLE

The final series of simulation experiments is aimed to evaluate the robustness of the proposed approach against the omission of common confounders for both phenotypes. We, therefore, introduce a (normally distributed) non-genetic risk factor $U$ to explain 1 and 10%, respectively, of the phenotypic variation in both phenotypes. The variable $U$ will be considered unmeasured in the analysis. In the presence of such a variable, the proposed adjustment approach, like the standard approaches, will be biased because it will estimate the effect of the intermediate phenotype on the target phenotype without incorporating the common, unmeasured risk factor.
For a variety of allele frequencies, Table IV shows the estimated nominal significance levels for Wald tests that are based on the proposed adjustment principle. Although, as predicted by our theoretical considerations, our approach no longer maintains the specified significance level, the impact appears negligible for applications. For an unknown confounding variable that explains 1% of the phenotypic variation ($r^2 = 0.01$), no observable departure from the theoretical 5%-level can be detected. When a confounding variable that explains 10% of the phenotypic variation in both phenotypes is omitted in the adjustment principle, the theoretical significance level is not maintained for small allele frequencies (<15%). Given the current epidemiologic knowledge and understanding of the phenotypes studied in genetic association study, the omission of common confounding variable with $r^2$ of 10% is rather unlikely. Even for a $r^2$ range of about 1%, most confounding variables for phenotypes of complex disease are often known.

In summary, our simulation studies suggest that the proposed adjustment principle performs well under realistic conditions. The approach maintains the significance level in the presence of spurious associations. For realistic genetic effect sizes, the approach achieves sufficient power, even in situations in which an adjustment is not required and standard approaches are optimal. The disadvantage of the approach, the required knowledge of all common confounding variables, turns out to be of lesser concern in applications where the degree of unmeasured confounding is anticipated to be weak.

**DISCUSSION**

In order to understand the genetic architecture of complex diseases, it is important to gain insight into the multifaceted relationships between complex phenotypes and their genetic associations. The origin of an observed genetic association between a SNP and the target phenotype can be attributable to a direct genetic effect or can be caused by a non-genetic link with another phenotype that is itself influenced by the marker locus of interest. In order to prioritize the follow-up of large-scale association studies in terms of replication strategies in other populations or, even more importantly, in terms of functional work, it is crucial to be able to distinguish between these different sources of genetic association.

In this article, we proposed an adjustment that can be incorporated into many genetic association tests and, thereby, enables the test to assess whether an observed association is caused by a genetic association with another phenotype, or whether it is attributable to a direct genetic effect. The approach is computationally simple and can be easily implemented in most software packages. If the principle is applied to correct for a false-positive association between the marker and the secondary phenotype, the adjusted test remains valid and still achieves power levels that are identical to those of standard adjustments. These properties make the proposed procedure a universally applicable adjustment principle in genetic association studies.

The proposed direct effects approach relies on the assumption that the primary phenotype $T$ does not affect the secondary phenotype $K$. In many situations, both phenotypes may mutually influence each other over time. To accommodate this, we will extend our work to repeated measures data on both phenotypes. In addition, we will extend it to allow for
ascertainment conditions, such as encountered in case-control and family studies (see also the Appendix).

Acknowledgments

The first two authors acknowledge support from IAP research network grant nr. P06/03 from the Belgian government (Belgian Science Policy).

APPENDIX A

The proposed adjustment principle forms a special case of the “unweighted estimator” in Goetgeluk et al. [2009], which allows for more general complexities, such as arbitrary non-linear models for the expected outcome and gene-environment interactions between the genotype and intermediate phenotype. Below, we demonstrate the validity of this principle for the setting that we have considered in this article.

By using the adjusted phenotype \( \tilde{\gamma}_j \) in the test statistic (5) or (6), we remove the arrow from \( K \) to \( \Upsilon \), and thus the indirect effect of \( X \) on \( \Upsilon \). This can be seen using the principles of causal diagrams (see the section “The Limitation of Standard Analysis Approaches for the Detection of Direct Genetic Effects”) and explains intuitively why a standard association test, using the adjusted phenotype \( \tilde{\gamma}_j \), is valid for testing direct genetic effects. More formally, suppose that the null hypothesis is true that \( X \) has no effect on \( \Upsilon \) other than through \( K \). Let

\[
E(\Upsilon | X, K, U) = \Phi \{ \omega(U) + \gamma_1 K \}, \quad (A1)
\]

where \( \Phi \) is the identity link (\( \Phi(x) = x \)) or the exponential link (\( \Phi(x) = \exp(x) \)) and where \( \omega(U) \) is an arbitrary function. This model does not involve \( X \) because we are working under the null hypothesis of no direct effect. Furthermore, the parameter \( \gamma_1 \) in this model is the same as in model

\[
E(\Upsilon | X, K, L, S) = \Phi \{ \omega^*(X, L, S) + \gamma_1 K \} \quad (A2)
\]

for some function \( \omega^*(X, L, S) \) of \( (X, L, S) \) (cfr. model (2)), which can be seen by inferring this model from model (A1) upon noting that \( \Upsilon \perp (L, S) | K, X, U \) and \( U \perp K | L, X, S \) under the diagram of Figure 1, where \( A \perp B | C \) for variables \( A, B \) and \( C \) means that \( A \) is conditionally independent of \( B \), given \( C \). It now follows that the test statistic (6) with \( \tilde{\gamma} = \gamma - \gamma_1 K \) when \( \Phi \) is the identity link and \( \tilde{\gamma} = \Upsilon \exp(-\gamma_1 K) \) when \( \Phi \) is the exponential link (and likewise the test statistic (5)) has mean zero at the null hypothesis. For the test statistic (6), for instance, this is because

\[
E[\{X - E(X|S)\} \tilde{\gamma}] = E[\{X - E(X|S)\} \Phi \{\omega(U)\}]
\]
and by the fact that $U|X|S$ (which follows because the genotype $X$ is only directly affected by $S$).

We now derive the distribution of the test statistic (6). The derivation is analogous for the test statistic (5) upon substituting $X_i - E(X_i|S_i)$ with $X_i - E(X_i)$. Using M-estimation arguments (van der Vaart, 1998, p 48–60), the test statistic (6) may be adjusted for estimating $\gamma_1$ by calculating the adjusted test statistic

$$\tilde{T}_i = \{X_i - E(X_i|S_i)\} \tilde{Y}_i - \lambda[K_i - E(K_i|L_i, X_i, S_i)]\epsilon_i,$$  \hspace{1cm} (A3)

which guarantees that the test statistic (A3) is uncorrelated with the scores needed for estimation of $\gamma_1$. Under model (3), $e_i$ is the residual from that model,

$$\lambda = \frac{E[\{X_i - E(X_i|S_i)\}] K_i}{\text{Var}(K_i|L_i, X_i, S_i)}$$

and $E[K|L, X, S]$ is the fitted value from a linear regression of $K_i$ on $(L, X, S)$. For binary traits or counts obeying the multiplicative model

$$E(Y_i) = \exp(\gamma_0 + \gamma_1 K_i + \gamma_2 X_i + \gamma_3 S_i + \gamma_4 L_i),$$  \hspace{1cm} (A4)

the adjusted phenotype can be computed as

$$\tilde{Y}_i = Y_i \exp(-\hat{\gamma}_1 K_i) - \mu_i,$$  \hspace{1cm} (A5)

where $\hat{\gamma}_1$ now denotes the maximum likelihood estimate obtained by fitting the Poisson regression model (A4). In that case, $e_i$ is the residual from that model,

$$\lambda = \frac{E[\{X_i - E(X_i|S_i)\}] Y_i \exp(-\hat{\gamma}_1 K_i)K_i}{\text{Var}(K_i|L_i, X_i, S_i)}.$$

$\mu_i$ is the fitted value under model (10), $E[K|L, X, S]$ is the fitted value from a weighted linear regression of $K_i$ on $(L, X, S)$, with weights $\mu_i$ and $\text{Var}(K_i|L, X, S) = E[\mu_i(K_i - E[K|L, X, S])^2]$. By the Central Limit Theorem, $\sum_{i=1}^{n} \tilde{T}_i$ has a normal distribution in large samples with mean zero at the null hypothesis and variance $\Sigma$ which can be estimated by the sample variance of $\tilde{T}_i$. Squaring and noting that

$$\sum_{i=1}^{n} \{K_i - E(K_i|L_i, X_i, S_i)\} \epsilon_i = 0$$

by construction, yields the distribution of the test statistic as reported in the section “The Limitation of Standard Analysis Approaches for the Detection of Direct Genetic Effects”.

---

*Genet Epidemiol. Author manuscript; available in PMC 2017 September 29.*
Remark: population admixture

When there is population admixture on both phenotypes w.r.t. unrelated causes (i.e., when the diagram of Fig. 1 includes an additional unmeasured variable $U$ which simultaneously affects $S$ and $K$), then the proposed adjustment principle remains valid. This is because the principles of causal diagrams show that adjustment for $S$, $X$ and $L$ is then still sufficient to estimate the causal effect of $K$ on $\Upsilon$, where the adjustment for $S$ in model (3) happens through the adjustment for $E(X|S)$ (this involves the implicit assumption that adjustment for $E(X|S)$ is sufficient for controlling for $S$).

Remark: confounder effects on target phenotype

Note that Figure 1 implicitly assumes that common risk factors $L$ of both phenotypes do not themselves affect the target phenotype. Our test remains valid without this assumption. In that case, it tests for an association between the target SNP and target phenotype $\Upsilon$, other than through the given intermediate phenotype $K$ (i.e., either direct or indirect via $L$). Our results also remain valid when $S$ affects $T$ because conditioning on $S$ blocks all paths from $S$ to $T$.

Remark: ascertainment

Note that when $\mu = 0$ in (A5) and the outcome $T$ is dichotomous (0 or 1), the proposed test statistic continues to have mean zero at the null hypothesis of no direct effect even when the data are sampled conditional on the outcome. This is because it then has mean zero conditional on $T=0$ (by the fact that $\gamma_1$ is then zero) and thus also conditional on $T=1$ (by the fact that also the overall mean is zero). The proposed test thus remains valid under ascertainment conditions provided that the expected genotype $E(X|S)$ is estimated from population samples and logistic regression is used to approximate the mediator effect $\gamma_1$. Details for various ascertainment schemes will be reported in a companion article.

Remark: direction of the edge between $K$ and $\Upsilon$

In some applications, it may not be clear whether $K$ affects $T$ or vice versa. Suppose, therefore, that the causal diagram in Figure 1 holds, but that the roles of $K$ and $T$ are incorrectly reversed in the analysis. Suppose furthermore that model (A2) holds with $\Phi$ the identity link and with $(K, T)$ bivariate normally distributed conditional on $(X, L, S)$. Then it follows that

$$E(K|X, T, L, S) = \omega^{**}(X, L, S) + \gamma_1^* T$$

for some function $\omega^{**}(X, L, S)$ of $(X, L, S)$, where $\gamma_1^* = \gamma_1 \sigma_K^2 / (\gamma_1^2 \sigma_K^2 + \sigma_T^2)$ with $\sigma_K^2 \equiv \text{Var}(K|X, L, S)$ and $\sigma_T^2 \equiv E\{\text{Var}(T|K, X, L, S|X, L, S)\}$. It follows that the proposed adjusted phenotype $\tilde{K} \equiv K - \gamma_1^* T$ equals
\[
\frac{\sigma_T^2 K - \gamma_1 \sigma_K^2 (T - \gamma_1 K)}{\gamma_1^2 \sigma_K^2 + \sigma_T^2 T}.
\]

Because \( T - \gamma_1 K \) is independent of \( X \) (conditional on \( S \)) under the null hypothesis, standard association tests based on this residual are effectively testing for an overall association with \( K \) (under that null hypothesis).

**References**


*Genet Epidemiol.* Author manuscript; available in PMC 2017 September 29.


Fig. 1.
Causal diagram illustrating the confounding of the genetic association between the target phenotype \( \Upsilon \) and the marker locus \( X \). The variable \( K \) denotes the intermediate phenotype, \( X \) the marker score, and \( S \) the parental genotypes (in the case of a family-based study, and measured factors inducing population admixture otherwise). \( U \) denotes a collection of unmeasured factors that allow for confounding due to population admixture and of common risk factors of both phenotypes.
Fig. 2.
Causal diagrams illustrating the data generating mechanism under simulation scenarios I–IV.
TABLE I

Association with rs2415815 in the Framingham Heart Study (FHS), the CAMP Study and the British Birth Cohort (BBC): overall association (overall), direct association based on standard residuals adjusting for $K$ (direct-standard) and based on proposed residuals (direct-proposal)

<table>
<thead>
<tr>
<th>Target phenotype $\mathcal{T}$</th>
<th>Intermediate phenotype $K$</th>
<th>Adjustment</th>
<th>FHS</th>
<th>CAMP</th>
<th>BBC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FEV1</td>
<td>Overall</td>
<td>-0.35</td>
<td>1.86</td>
<td>-0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct-standard</td>
<td>-0.35</td>
<td>2.38 to 1.68</td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct-proposal</td>
<td>-1.34</td>
<td>-2.94 to 0.27</td>
<td>-0.45</td>
</tr>
<tr>
<td>Weight</td>
<td>FEV1</td>
<td>Overall</td>
<td>-1.20</td>
<td>-2.22 to -0.18</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct-standard</td>
<td>-1.20</td>
<td>-2.22 to -0.18</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct-proposal</td>
<td>-1.34</td>
<td>-2.32 to -0.35</td>
<td>-0.16</td>
</tr>
</tbody>
</table>

* Measured as BMI.
TABLE II

Empirical type-1 errors at 5% significance level of Wald tests for genetic effects, (a) based on residuals adjusted for \((S, K, L)\), \((S, K)\), (b) directly adjusted for \((S, K, L)\), \((S, K)\) and (c) adjusted with the proposed adjustment principle \((D\), along with bias and standard error of the direct effect size estimate\)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Wald test on residuals adjusted for</th>
<th>Wald test on trait adjusted for</th>
<th>Proposed adjustment principle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((S, K, L)) ((S, K)) ((S, K, L)) ((S, K))</td>
<td>((D) Bias \times 10^3 SE)</td>
<td>(</td>
</tr>
<tr>
<td>I</td>
<td>0.05 0.080 0.150 0.044 0.134 0.051</td>
<td>−2.05 0.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1 0.085 0.079 0.065 0.075 0.047</td>
<td>3.1 0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.15 0.092 0.059 0.079 0.057 0.053</td>
<td>−2.3 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2 0.083 0.047 0.075 0.047 0.047</td>
<td>−4.0 0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25 0.089 0.057 0.081 0.056 0.056</td>
<td>5.0 0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3 0.093 0.058 0.086 0.058 0.057</td>
<td>6.1 0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.35 0.083 0.051 0.075 0.051 0.047</td>
<td>5.1 0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4 0.085 0.055 0.078 0.055 0.052</td>
<td>5.0 0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.45 0.095 0.056 0.086 0.056 0.053</td>
<td>1.4 0.17</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>0.05 0.054 0.052 0.054 0.052 0.053</td>
<td>−1.7 0.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1 0.045 0.045 0.045 0.045 0.048</td>
<td>−4.1 0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.15 0.047 0.045 0.047 0.045 0.047</td>
<td>−6.9 0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2 0.048 0.049 0.048 0.049 0.049</td>
<td>1.4 0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25 0.049 0.050 0.049 0.050 0.048</td>
<td>−0.023 0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3 0.049 0.048 0.049 0.048 0.049</td>
<td>4.0 0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.35 0.054 0.052 0.054 0.052 0.054</td>
<td>−0.97 0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4 0.052 0.050 0.052 0.050 0.052</td>
<td>−1.9 0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.45 0.052 0.051 0.052 0.051 0.054</td>
<td>2.9 0.17</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>0.05 0.085 0.144 0.050 0.131 0.049</td>
<td>1.9 0.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1 0.083 0.077 0.065 0.074 0.046</td>
<td>−4.0 0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.15 0.083 0.061 0.072 0.060 0.050</td>
<td>1.2 0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2 0.090 0.051 0.083 0.051 0.050</td>
<td>−6.8 0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.25 0.085 0.048 0.078 0.048 0.046</td>
<td>0.32 0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3 0.081 0.050 0.075 0.050 0.049</td>
<td>−4.9 0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.35 0.085 0.054 0.076 0.054 0.052</td>
<td>−1.6 0.18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4 0.083 0.051 0.075 0.050 0.048</td>
<td>−2.2 0.17</td>
<td></td>
</tr>
<tr>
<td>Scenario</td>
<td>Freq</td>
<td>Wald test on residuals adjusted for (S, K, L)</td>
<td>Wald test on trait adjusted for (S, K, L)</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>IV</td>
<td>0.45</td>
<td>0.095</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.049</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.051</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.049</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.052</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.049</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.052</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.049</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.047</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.052</td>
<td>0.052</td>
</tr>
</tbody>
</table>
TABLE III

Empirical power at 5% significance level of Wald tests for genetic effects, (a) based on residuals adjusted for \((S, K, L)\), \((S, K)\), (b) directly adjusted for \((S, K, L)\), \((S, K)\) and (c) adjusted with the proposed adjustment principle \((D, along with bias and standard error of the direct effect size estimate)\)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Freq</th>
<th>Wald test on residuals adjusted for</th>
<th>Wald test on trait adjusted for</th>
<th>Proposed adjustment principle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>((S, K, L))</td>
<td>((S, K))</td>
<td>((D))</td>
</tr>
<tr>
<td>I</td>
<td>0.05</td>
<td>0.520</td>
<td>0.119</td>
<td>0.406</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.787</td>
<td>0.437</td>
<td>0.743</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.904</td>
<td>0.709</td>
<td>0.891</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.955</td>
<td>0.857</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.971</td>
<td>0.906</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.980</td>
<td>0.930</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.988</td>
<td>0.943</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.992</td>
<td>0.957</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.993</td>
<td>0.962</td>
<td>0.992</td>
</tr>
<tr>
<td>II</td>
<td>0.05</td>
<td>0.405</td>
<td>0.401</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.655</td>
<td>0.655</td>
<td>0.655</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.805</td>
<td>0.802</td>
<td>0.804</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.883</td>
<td>0.878</td>
<td>0.883</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.929</td>
<td>0.927</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.947</td>
<td>0.945</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.964</td>
<td>0.962</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.973</td>
<td>0.970</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.976</td>
<td>0.974</td>
<td>0.977</td>
</tr>
<tr>
<td>III</td>
<td>0.05</td>
<td>0.519</td>
<td>0.114</td>
<td>0.396</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.787</td>
<td>0.442</td>
<td>0.742</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.897</td>
<td>0.691</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.959</td>
<td>0.859</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.975</td>
<td>0.910</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.984</td>
<td>0.927</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.986</td>
<td>0.945</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.991</td>
<td>0.959</td>
<td>0.991</td>
</tr>
<tr>
<td>Scenario</td>
<td>Freq</td>
<td>Wald test on residuals adjusted for (S, K, L)</td>
<td>Wald test on trait adjusted for (S, K, L)</td>
<td>Wald test on trait adjusted for (S, K)</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>IV</td>
<td>0.45</td>
<td>0.995</td>
<td>0.964</td>
<td>0.994</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.407</td>
<td>0.400</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.647</td>
<td>0.643</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.803</td>
<td>0.802</td>
<td>0.804</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.887</td>
<td>0.882</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.927</td>
<td>0.926</td>
<td>0.927</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.953</td>
<td>0.950</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.962</td>
<td>0.961</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.968</td>
<td>0.967</td>
<td>0.968</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.977</td>
<td>0.976</td>
<td>0.977</td>
</tr>
</tbody>
</table>
TABLE IV

Empirical type-1 errors at 5% significance level of Wald tests for genetic effects, (a) based on residuals adjusted for \((S, K, L)\), \((S, K)\), (b) directly adjusted for \((S, K, L)\), \((S, K)\) and (c) adjusted with the proposed adjustment principle \((D, \text{ along with bias and standard error of the direct effect size estimate})\), in the presence of unmeasured confounding:

<table>
<thead>
<tr>
<th>(t^*)</th>
<th>Wald test on residuals adjusted for</th>
<th>Wald test on trait adjusted for</th>
<th>Proposed adjustment principle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((S, K, L))</td>
<td>((S, K))</td>
<td>((S, K, L))</td>
</tr>
<tr>
<td>1%</td>
<td>0.05</td>
<td>0.061</td>
<td>0.172</td>
</tr>
<tr>
<td>0.1</td>
<td>0.071</td>
<td>0.065</td>
<td>0.093</td>
</tr>
<tr>
<td>0.15</td>
<td>0.076</td>
<td>0.049</td>
<td>0.067</td>
</tr>
<tr>
<td>0.2</td>
<td>0.075</td>
<td>0.053</td>
<td>0.064</td>
</tr>
<tr>
<td>0.25</td>
<td>0.070</td>
<td>0.051</td>
<td>0.064</td>
</tr>
<tr>
<td>0.3</td>
<td>0.078</td>
<td>0.056</td>
<td>0.069</td>
</tr>
<tr>
<td>0.4</td>
<td>0.067</td>
<td>0.048</td>
<td>0.063</td>
</tr>
<tr>
<td>0.45</td>
<td>0.054</td>
<td>0.117</td>
<td>0.228</td>
</tr>
<tr>
<td>0.5</td>
<td>0.054</td>
<td>0.117</td>
<td>0.072</td>
</tr>
<tr>
<td>0.2</td>
<td>0.057</td>
<td>0.054</td>
<td>0.050</td>
</tr>
<tr>
<td>0.25</td>
<td>0.058</td>
<td>0.050</td>
<td>0.053</td>
</tr>
<tr>
<td>0.3</td>
<td>0.056</td>
<td>0.054</td>
<td>0.051</td>
</tr>
<tr>
<td>0.35</td>
<td>0.052</td>
<td>0.050</td>
<td>0.046</td>
</tr>
<tr>
<td>0.4</td>
<td>0.059</td>
<td>0.055</td>
<td>0.052</td>
</tr>
<tr>
<td>0.45</td>
<td>0.056</td>
<td>0.059</td>
<td>0.050</td>
</tr>
</tbody>
</table>