Journal of Biometrics & Biostatistics

Lee, et al. J Biomet Biostat 2011, 2:4
DOI: 10.4172/2155-6180.1000119

Research Article Open Access

Heritability Estimation using Regression Models for Correlation

Hye-Seung Lee[1*], Myunghee Cho Paik[2], Tatjana Rundek[3], Ralph L Sacco[3], Chuanhui Dong[3] and Jeffrey P Krischer[1]

[1]*Pediatrics Epidemiology Center, Department of Pediatrics, University of South Florida, Tampa, FL33612, USA*
[2]*Department of Biostatistics, Columbia University, New York, NY10032, USA*
[3]*Department of Neurology, Epidemiology and Public Health, and Human Genetics, Miller School of Medicine, University of Miami, Miami, FL33136, USA*

Abstract

Heritability estimates a polygenic effect on a trait for a population. Reliable interpretation of heritability is critical in planning further genetic studies to locate a gene responsible for the trait. This study accommodates both single and multiple trait cases by employing regression models for correlation parameter to infer the heritability. Sharing the properties of regression approach, the proposed methods are exible to incorporate non-genetic and/or non-additive genetic information in the analysis. The performances of the proposed model are compared with those using the likelihood approach through simulations and carotid Intima Media Thickness analysis from Northern Manhattan family Study.

Keywords: Heritability; Variance component model; Regression models for Correlation; Likelihood estimation

Introduction

Heritability is a fundamental concept to quantify the genetics on a quantitative trait. Through the random effect model for a quantitative trait, locus specific heritability, which is for linkage analysis, is estimated after adjusting for polygenic heritability. Heritability has been originally developed in animal studies, which have more control on environmental effects through the study design. Recent attention to quantitative trait analysis has been made in human genetic studies for common diseases. Since common diseases are mostly multifactorial, there are likely to be multiple genetic an environmenttal factors involved in the disease process, and each factor by itself will have limited influence on the disease. Thus, one effort to enhance power to detect a genetic influence on a common disease is to examine subcomponent traits for the disease when the subcomponent traits are considered to be closer to the action of a gene. In family studies, analyzing quantitatively measured subcomponent traits can be of benefit because they can provide genetic information for both affected and unaffected family members. However, existing analytic methods have been limited to control for non-genetic differences between two family members in the analysis, which is more critical in common disease analyses.

In quantitative genetics, the covariance of a trait is described as the sum of the covariance of a random genetic effect and the covariance of a random environmental effect [1,2]. The genetic variance is the variance of a random genetic effect given the correlation of random genetic effects between family members. The environmental variance is the variance of a random environmental effect given the correlation of random environmental effects between family members. While the correlation of random genetic effects is the probability of sharing genes, which can be determined based on their familial relationship, the correlation of random environmental effects is the probability of sharing environments between them. In non-human genetics, a proper study design can make feasible the estimation of the probability of sharing environments on a trait. But in human genetics, especially for large family studies including multi-generations, it is not trivial to make any reasonable estimation on the probability of sharing environments between two family members or the structure for each family.

There has been discussion on the impact of pedigree structure on heritability estimation. Mallinckrodt et al. [3] found that there was no difference in the precision and accuracy of heritability estimate from simulated data for random or fixed pedigree structures. However, [4] compared heritability estimates for pulse pressure from three different family studies and showed that multi-generation studies produced significantly lower heritability estimates than sib-pair studies. Recently, [5] revisited this issue using simulated data and concluded that there is hardly any impact of pedigree structure on heritability estimates. These contradictory findings can be due to the fact that simulated data can comply any necessary assumptions such as zero correlation of environmental effects between two family members for all families but not real data.

To reduce the bias due to non-genetic effects, one strategy would be to conduct analysis adjusted for environmental covariates. Issues in environmental covariate adjustments have been discussed in the setting of quantitative trait linkage analysis. As compared in [6], individual specific covariates are included in the analysis to reduce the residual error of the mean for the individual levels, while pair specific covariates are to allow or adjust for pair specific differences between pairs on the covariance, which is critical in quantitative genetics. Although both variance component model and regression approach can handle pair specific covariates, regression approach provides more exibility to incorporate pair specific covariates (e.g.,age difference for a pair) in the analysis. In addition, [7] did simulation study for variance component models incorporating covariates in the means, covariance and residuals, under the normality assumption, and concluded that potentially important environmental covariates should be adjusted in the absence of a correlation between the covariates and the locus.

In this paper, we propose a new analytic method to estimate heritability for both single and multiple trait cases. When phenotypic similarity is translated into the covariance, heritability is commonly estimated as the ratio between genetic variance estimate and total phenotypic variance estimate. Unlike existing approaches, the

***Corresponding author:** Hye-Seung Lee, Pediatrics Epidemiology Center, Department of Pediatrics, University of South Florida,Tampa, FL33612, USA, E-mail: Hye-Seung.Lee@epi.usf.edu

Received July 18, 2011; **Accepted** October 31, 2011; **Published** November 15, 2011

Citation: Lee HS, Paik MC, Rundek T, Sacco RL, Dong C, et al. (2011) Heritability Estimation using Regression Models for Correlation. J Biomet Biostat 2:119. doi:10.4172/2155-6180.1000119

proposed method estimates the heritability itself as a regression parameter, without estimating both variances. For single trait analysis, we implement the efficient three generalized estimating equations (GEEs) proposed by [8]; for multiple trait analysis, the method for multivariate familial correlation estimation proposed by [9]. While this new approach shares advantages using regression approaches under the scheme of GEE, it also allows us to handle various types of relative pairs under one model and all parameters at the same stage, without requiring the normality assumption. The performances of the proposed model are compared with those using variance component model through simulations and carotid Intima Media Thickness (IMT) analysis from Northern Manhattan Family Study (NOMAS).

Materials and Methods

Random effect model for a quantitative trait

One's level of a quantitative trait can be viewed as a combination of genetic and environmental effects. With the conventional assumption that genetic and environmental effects are independent, a model for the trait of interest Y_{ij} from i^{th} family and jth member is written as follows:

$$Y_{ij} = \mu_{ij} + g_{ij} + \varepsilon_{ij} \quad (1)$$

where μ_{ij} is the mean of the trait, while gij and εij are the random genetic and environmental effects, respectively.

Considering the nature of the genetic effect gij, the model (1) becomes as follows:

$$Y_{ij} = \mu_{ij} + A_{ij} + d_{ij} + \varepsilon_{ij} \quad (2)$$

where A_{ij} is the additive genetic effect whose variance is σ_A^2, d_{ij} is the dominance effect whose variance is σ_d^2, and ε_{ij} is the environmental effect whose variance is σ_e^2. Note that epistasis was not considered in this model. From the independence assumption, total phenotypic variance $Var(Y_{ij}) = \sigma^2 =$

$\sigma_A^2 + \sigma_d^2 + \sigma_e^2$, and for j ≠ k,

$$Cov(Y_{ij}, Y_{ik}) = Cov(A_{ij}, A_{ik}) + Cov(d_{ij}, d_{ik}) + Cov(\varepsilon_{ij}, \varepsilon_{ik})$$
$$= 2\theta_{jk}\sigma_A^2 + \Delta_{ij}\sigma_d^2 + \gamma_{ijk}\sigma_e^2$$

where $2\theta_{jk}$ = Corr(A_{ij}, Aik), Δ_{jk}= Corr(d_{ij}, dik), and γ_{ijk}= Corr(ε_{ij}, ε_{ik}). Note that the θ_{jk} and Δ_{jk} are determined as a probability of sharing additiveIBD allele and a probability of sharing dominance IBD allele, which can be expected from the familial relationship of j and k, respectively, irrespective of i. Two alleles (one each from two individuals) are said to be IBD (identity by descent) when they are in the same form and descended from the same ancestral allele. Table 1 shows the θ_{jk} and Δ_{jk} in each familial relationship.

Proposed methods

The correlation between two levels from a pair is expressed as

$$Corr(Y_{ij}, Y_{ik}) = 2\theta_{jk}\frac{\sigma_A^2}{\sigma^2} + \Delta_{jk}\frac{\sigma_d^2}{\sigma^2} + \gamma_{ijk}\frac{\sigma_e^2}{\sigma^2} \quad (3)$$

where the $\frac{\sigma_A^2}{\sigma^2}$ is referred to as `(narrow sense) heritability' in the literature, which measures the additive genetic effect on a quantitative trait. The poly-genic heritability is conventionally defined as the $\frac{\sigma_A^2}{\sigma^2}$ after assuming $\gamma_{ijk} = 0$.

To estimate the heritability, existing methods estimate the total variance σ_2 and σ_A^2, given the $2\theta_{jk}$ and Δ_{jk} for a pair (j, k) in the ith family. Likelihood function assuming multivariate normal distribution of the trait vector of interest from members of a family is commonly used in the analysis of extended families. The inference for the heritability estimator under variance component approach has been well established based on restricted maximum likelihood under the general framework by [10,11]. In addition, Ekstrϕm [5] provided explicit expression of the asymptotic variance of each variance estimator.

Degree	Relationship	$2\theta_{jk}$	Δ_{jk}
1	Parent-offspring	1/2	0
1	Full sib	1/2	1/4
2	Grandparent-grandoffspring	1/4	0
2	Half sibs	1/4	0
2	Uncle-nephews	1/4	0
3	First cousins	1/8	0

Table 1: P(IBD allele share) expected from the familial relationship.

We propose an alternative approach to interpret the heritability from the phenotypic correlation. Single trait analyses may employ Pearson's correlation, and multiple trait analyses may employ maximum canonical correlation. By interpreting the correlation for a pair (j, k) in the ith family as a regression model $2\theta_{jk}\beta_1 + \Delta_{jk}\beta_2 + \beta_{ijk}\beta_3$, the inference on the regression parameter β_1 explains the heritability of the trait, $\frac{\sigma_A^2}{\sigma^2}$. To model Pearson's correlation, we employed the 3GEEs by [8] and a joint estimating equation by [9] for maximum canonical correlation.

Our proposed approach differs from the existing approach in three folds: first, we do not assume normality and use an estimating equation approach rather than likelihood approach, which makes flexible to incorporate non- genetic share between two individuals; second, we parameterize in such a way that the ratio of σ_A^2 to σ^2 can be directly estimated, rather than estimating each variance component separate; lastly, multiple trait analysis can be utilized to examine a pleiotropy, which occurs when a single gene influences multiple traits.

Results

Simulation study

Simulation studies compared the performance of the proposed approach with the standard analysis implementing restricted maximum likelihood estimation for single trait cases. Bias and efficiency were compared in two situations when the trait does not follow the normality and when non-zero correlation of environmental effects exists between two family members. In each situation, we generated 100 simulated data including 100 families with three sib-pairs. For siblings, the correlation of additive genetic effects, $2\theta_{jk}$, is 0.5, while the correlation of dominant genetic effects, Δ_{jk}, is 0.25. For the multivariate normal, t or gamma distribution of a trait, when true polygenic heritabilities were 0.4, 0.6 and 0.8, the bias and variances of each heritability estimate were compared by varying the correlation of environmental effects for each pair from 0, 0.3, 0.6 to 0.9. That is, the data generation with zero correlation of environmental effects complies the defnition of the conventional heritability. Table 2 summarizes total variances and correlations implemented in this simulation.

First, we examined the impact of a positive environmental correlation on current restricted maximum likelihood approach for the conventional heritability assuming $\gamma_{ijk} = 0$. Table 3 presents the simulation results. As expected, the likelihood approach performed the best when the true correlation of environmental effect was zero and the trait followed the normal distribution. The bias got greater as the

true correlation of environmental effect was greater, regardless of the size of true heritability or distribution. Incorporating the correlation of environmental effects into the likelihood ap-proach and proposed approach, we compared their performances when the true non-zero correlation of environmental effects were known for families. Table 4 presents the results for the heritability estimates including non-zero correlation of environmental effects in the model. We found that the likelihood method produces better results with the estimates of the correlation of environmental effects than with the specification of the true correlation for the simulated data (results are not shown). Thus, we present the simulation results for the likelihood approach when the correlation of environmental effects was estimated from each simulated data with the correct specification of the structure for each family, while the proposed method implemented the correct specification of the correlation of environmental effects. Including non-zero correlation parameter of environmental effect in the model reduced the bias in both approaches, especially when the true correlation of environmental effects was high, as expected (Table 3 vs. 4). The likelihood approach performed better when the data were generated from the normal distribution, but the proposed method showed greater improvement as the data deviated more from the normality. However, the likelihood approach requires the correlation of environmental effects to be estimable or the number of family members to be the same. The variance estimates of the proposed model estimates were more efficient in terms of being close to the simulation variances. The performance of the proposed method was stable in most of the cases considered for simulations.

Data analysis

Sacco et al. [12] presented heritability and linkage analysis for eight carotid IMT measures from 1390 subjects and 100 Caribbean Hispanic families. The families were selected when probands from the NOMAS were high-risk Caribbean Hispanic members de_ned by the following criteria: (1) reporting a sibling with a history of myocardial infarction or stroke; or (2) having 2 of 3 quantitative risk phenotypes (maximal carotid plaque thickness, left ventricular mass, or homocysteine level above the 75th percentile in the NOMAS cohort). Families were enrolled if the proband was able to provide a family history, obtain the family members permission for the research staff to contact them, and had at least 3 first-degree relatives able to participate. Although the proband was identified in Northern Manhattan, they enrolled family members in New York at Columbia University and in the Dominican Republic at the Clinicas Corazones Unidos in Santo Domingo. All subjects provided informed consent, and the study was approved by the Institutional Review Boards of Columbia University, University of Miami, the National Bioethics Committee, and the Independent Ethics Committee of Instituto Oncologico Regional del Cibao in the Dominican Republic. See [13] for the details of study design.

Heritability	Covariance	Genetic		Environmental	Total*
		Additive	Dominant		
0.4	Variance	40	30	30	100
	Correlation	0.5	0.25	0	0.275
		0.5	0.25	0.3	0.365
		0.5	0.25	0.6	0.455
		0.5	0.25	0.9	0.545
0.6	Variance	60	10	30	100
	Correlation	0.5	0.25	0	0.325
		0.5	0.25	0.3	0.415
		0.5	0.25	0.6	0.505
0.8	Variance	80	5	15	100
	Correlation	0.5	0.25	0	0.4125
		0.5	0.25	0.3	0.4575
		0.5	0.25	0.6	0.5025
		0.5	0.25	0.9	0.5475

Table 2: Specified correlation and variance of each eect for simulation study.

Carotid IMT were measured by high-resolution B-mode ultrasound and expressed as the mean and mean of the maximum. Ultrasound measures of carotid IMT have been demonstrated to be valid measures of pathologically defined atherosclerosis, highly reproducible, associated with vascular risk factors, and predictive of stroke and myocardial infarction. The carotid IMT protocols yield measurements of the distance between lumen-intima and media-adventitia ultrasound echoes, from which the IMT and arterial diameter are derived for the 3 carotid segments. The carotid segments were defined as follows: (1) near and far wall of the segment extending from 10 to 20 mm proximal to the tip of the flow divider into the common carotid artery (CCA); (2) near and far wall of the carotid bifurcation beginning at the tip of the flow divider and extending 10 mm proximal to the ow divider tip (BIF); and (3) near and far wall of the proximal 10 mm of the internal carotid artery (ICA). Total IMT was calculated as a mean composite measure of the means of the near and the far wall IMT of all carotid sites (IMTx), and the maximum of the near and the far wall IMT of all carotid sites (IMTm). Carotid segment-specific IMT phenotypes were also examined (BIFx, BIFm, CCAx, CCAm, ICAx, ICAm).

We analyzed the data used for [12] to examine the performance of our proposed method. We first estimated the heritability assuming zero environmental correlations in the proposed model to compare the results from the SOLAR, which implements the restricted maximum likelihood approach. Our method is to interpret the relationship between phenotypic correlation and sharing genes as the heritability. Assuming $\gamma_{ijk} = 0$, the

$$Corr(Y_{ij}, Y_{ik}) = 2\theta_{jk}\frac{\sigma_A^2}{\sigma^2} + \Delta_{jk}\frac{\sigma_d^2}{\sigma^2}$$

Then, the estimate of $\frac{\sigma_A^2}{\sigma^2}$.

which is the heritability, would be roughly the sum of phenotypic correlation estimates for each degree of relatives over the sum of $2\theta_{jk}$ for each degree of relatives. For example, as for the trait IMTx, the ($2\theta_{jk}$, Correlation of the residual IMTx from the mean model) for each degree of relative pairs would be (0.5,0.406) for Parent-Offspring pairs, (0.5,0.379) for sib-pairs, (0.25, 0.294) for second degree, (0.125, 0.121) for third degree, (0.0625, 0.105) for fourth degree. Thus, the estimate of $\frac{\sigma_A^2}{\sigma^2}$ is 0.908 as the sum of phenotypic correlations is 1.305 and the sum of 2 θ_{jk} is 1.4375. As shown in Table 5, the heritability from the proposed model varied from 0.28 to 0.91, while the heritability reported in [12] was from 0.41 to 0.65, after adjusting for the same confounders in the mean model for each carotid IMT. Although two estimates were quite different, the trend (small to large) was consistent, as well as the statistical significance.

In this study, the mean family size was 14 (standard deviation (SD)=8); for family member pairs, age difference varied from 0 to 82 years (mean=20, SD=14, first quartile(Q1)=7, median(Q2)=18 and third quartile(Q3)=29), and BMI difference was from 0 to 32.7 (mean=6.0, SD=5.0, Q1=2.2, Q2=4.7 and Q3=8.4). A common practice is to analyze the residuals after adjusting for potential environmental

Citation: Lee HS, Paik MC, Rundek T, Sacco RL, Dong C, et al. (2011) Heritability Estimation using Regression Models for Correlation. J Biomet Biostat 2:119. doi:10.4172/2155-6180.1000119

		Normal			T			Gamma		
h^2	ρ_e	Bias	ASV	EMP	Bias	ASV	EMP	Bias	ASV	EMP
0.4	0	0.0023	0.0047	0.0025	0.0178	0.0258	0.0600	-0.0184	0.0230	0.0128
0.4	0.3	0.1458	0.0061	0.0024	0.0931	0.0292	0.0496	0.1353	0.0303	0.0134
0.4	0.6	0.2892	0.0077	0.0022	0.2287	0.0365	0.0441	0.2814	0.0384	0.0152
0.4	0.9	0.4326	0.0095	0.0019	0.3906	0.0463	0.0338	0.4162	0.0469	0.0104
0.6	00.0	0.0076	0.0094	0.0360	0.0007	0.0314	0.0498	-0.0227	0.0296	0.0127
0.6	0.6	0.2876	0.0094	0.0018	0.2695	0.0466	0.0360	0.2735	0.0466	0.0121
0.6	0.9	0.4311	0.0115	0.0014	0.4513	0.0595	0.0235	0.4277	0.0576	0.0072
0.8	0	-0.0005	0.0080	0.0020	0.0172	0.0412	0.0387	-0.0061	0.0397	0.0122
0.8	0.3	0.0712	0.0088	0.0019	0.0684	0.0441	0.0354	0.0639	0.0440	0.0129
0.8	0.6	0.1430	0.0098	0.0017	0.1475	0.0491	0.0341	0.1306	0.0483	0.0104
0.8	0.9	0.2147	0.0108	0.0015	0.2044	0.0531	0.0327	0.2043	0.0534	0.0081

NOTE: h^2: heritability; ρ_e: correlation of environmental eects; REML: Restricted maximum likelihood estimates; correlation regression: correlation regression estimates; Bias: estimate of h^2-h^2; ASV: Asymptotic variance of REML; EMP: Simulation variance.

Table 3: Simulation results for heritability assuming zero non-genetic correlations for relative pairs using REML: Multivariate normal, T and Gamma distribution.

			REML			correlation regression		
Distribution	h^2	ρ_e	Bias	ASV	EMP	Bias	SAN	EMP
Normal	0.4	0.3	-0.0003	0.0000	0.0001	0.0100	0.0061	0.0062
	0.4	0.6	-0.0003	0.0000	0.0001	0.0045	0.0020	0.0020
	0.4	0.9	-0.0004	0.0000	0.0001	0.0018	0.0006	0.0006
	0.6	0.3	0.0003	0.0000	0.0001	0.0087	0.0058	0.0059
	0.6	0.6	0.0004	0.0000	0.0001	0.0037	0.0018	0.0019
	0.6	0.9	0.0005	0.0001	0.0001	0.0691	0.0014	0.001
	0.8	0.3	0.0010	0.0004	0.0008	0.0075	0.0054	0.0056
	0.8	0.6	0.0011	0.0004	0.0009	0.0037	0.0018	0.0019
	0.8	0.9	0.0012	0.0004	0.0010	0.0018	0.0006	0.0006
	0.4	0.3	-0.0175	0.0001	0.0016	-0.0066	0.0161	0.0342
	0.4	0.6	-0.0166	0.0001	0.0013	-0.0273	0.0057	0.0098
	0.4	0.9	-0.0127	0.0001	0.0016	-0.0066	0.0018	0.0039
	0.6	0.3	0.0202	0.0001	0.0010	0.0106	0.0158	0.0295
	0.6	0.6	0.0185	0.0001	0.0012	-0.0123	0.0055	0.0112
	0.6	0.9	0.0225	0.0001	0.0014	0.0610	0.0042	0.0078
	0.8	0.3	0.0482	0.0005	0.0111	-0.0447	0.0159	
	0.8	0.6	0.0446	0.0005	0.0121	-0.0234	0.0049	0.0015
	0.8	0.9	0.0638	0.0006	0.0112	0.0016	0.0017	
Gamma	0.4	0.3	0.2012	0.0000	0.0000	-0.0066	0.0076	0.0077
	0.4	0.6	0.1977	0.0000	0.0000	-0.0030	0.0025	0.0031
	0.4	0.9	0.1921	0.0000	0.0000	-0.0043	0.0007	0.0008
	0.6	0.3	0.2026	0.0001	0.0003	-0.0067	0.0071	0.0080
	0.6	0.6	0.1993	0.0001	0.0003	-0.0070	0.0023	0.0028
	0.6	0.9	0.1948	0.0001	0.0003	0.0643	0.0018	0.0018
	0.8	0.3	0.1015	0.0002	0.0007	-0.0068	0.0068	0.0087
	0.8	0.6	0.1022	0.0002	0.0007	-0.0069	0.0022	0.0025
	0.8	0.9	0.1007	0.0003	0.0009	-0.003	7 0.0007	0.0008

SAN: Sandwich variance estimate; Others are same with the previous table.

Table 4: Simulation results for heritability estimation with non-zero environmental correlation parameter: Multivariate normal, T and Gamma distribution.

confounders in the mean model. However, it is still not clear if the mean model only adjustment is sufficient to correct for the bias when the covariance is modeled. Table 6 shows the raw correlation and the residual correlation from the mean model for CCAx, BIFx and ICAx stratified by the quartiles of age or BMI difference for pairs. The raw correlations became smaller when age difference or BMI difference was greater, regardless of sharing genes. Residual correlations stratified by the quartiles of age difference does not show the trend anymore, but those stratified by BMI difference seem to retain the trend.

We applied our proposed methods including the intercept in the model. For three different mean measures of carotid segments, CCAx, BIFx and ICAx, we studied the heritability of each trait and the heritability of three measures together. Table 7 presents the single trait analysis for each CCAx, BIFx and ICAx. The $\frac{\sigma_A^2}{\sigma^2}$ from Model 1 implies the heritability after allowing a common correlation for all relative pairs if they did not share any genes. For those traits, the common non-genetic correlation for relative pairs was not significantly different from zero, which may imply a strong genetic information on them (p=0.287, 0.522 and 0.522, respectively). The adjusted heritabilities from Model 1 were 0.581 for CCAx, 0.774 for BIFx and 0.375 for ICAx. The $\frac{\sigma_A^2}{\sigma^2}$ from Model 2 implies the heritability if there was no difference in age or BMI in relative pairs, after allowing a common correlation. In each carotid IMT measure, the age difference and BMI difference were not signigicantly different from 0, although p-values were close to 0.05 in some of situations. The adjusted heritabilities from Model 2 were 0.540 for CCAx, 0.732 for BIFx and 0.339 for ICAx. Table 8 presents

Citation: Lee HS, Paik MC, Rundek T, Sacco RL, Dong C, et al. (2011) Heritability Estimation using Regression Models for Correlation. J Biomet Biostat 2:119. doi:10.4172/2155-6180.1000119

	REML using SOLAR		correlation regression	
	$h^2 \pm SE^1$	p	$h^2 \pm SE^2$	p
IMTx	0.65±0.05	<0.0001	0.91±0.02	<0.0001
BIFx	0.58±0.05	<0.0001	0.86±0.02	<0.0001
ICAx	0.47±0.05	<0.000 1	0.45±0.01	<0.0001
CCAx	0.56±0.05	<0.0001	0.73±0.01	<0.0001
IMTm	0.62±0.05	<0.0001	0.75±0.01	<0.0001
BIFm	0.51±0.05	<0.0001	0.63±0.01	<0.0001
ICAm	0.41±0.06	<0.0001	0.28±0.01	0.002
CCAm	0.48±0.05	<0.0001	0.54±0.01	<0.0001

h^2: heritability; REML: Restricted maximum likelihood estimates; correlation regression: correlation estimates from Generalized estimating equation; SE^1: standard error from SOLAR: SE^2: standard error using correlation regression; p: p-value.

Table 5: Conventional heritability estimates for carotid Intima-Media Thickness (IMT) to be comparable with the results using SOLAR.

		CCAx		BIFx		ICAx	
		Raw	Residual	Raw	Residual	Raw	Residual
Overall		0.245	0.194	0.257	0.223	0.168	0.141
Age difference	<Q1	0.555	0.223	0.286	0.143	0.490	0.207
	Q1-Q2	0.406	0.132	0.193	0.118	0.372	0.175
	Q2-Q3	0.288	0.189	0.169	0.160	0.336	0.292
	≥Q3	0.224	0.245	0.152	0.150	0.257	0.227
BMI Difference	<Q1	0.298	0.240	0.241	0.202	0.312	0.267
	Q1-Q2	0.245	0.208	0.179	0.137	0.259	0.225
	Q2-Q3	0.214	0.171	0.171	0.127	0.243	0.217
	≥Q3	0.207	0.156	0.079	0.096	0.201	0.182

Q1: rst quartile; Q2: Median; Q3: third quartile; Raw: measured value; Residual: residual from the mean model.

Table 6: Correlation of carotid Intima-Media Thickness (IMT) stratied by age dierence and BMI dierence quartiles.

the multiple trait analysis for CCAx, BIFx and ICAx, simultaneously. We estimated the three trait heritability for CCAx, BIFx and ICAx to examine if there could be a common gene effect on those three measures. The multiple trait heritability from Model 1 was 0.738 and 0.784 from Model 2. Unlike the single trait analysis, the size became greater in Model 2 as the age difference was significantly different from zero. In conclusion, all of those three traits showed a strong genetic effect on them.

Discussion

In this paper, we proposed new heritability estimation methods implementing regression models for correlation. Unlike existing methods that estimate each variance component, the proposed approach estimates the heritability as a regression parameter through estimating equations without requiring normality assumption. In conjunction with linkage analysis, there have been a number of new quantitative trait analysis methods developed in extended family design to employ the GEE, regression based statistics or score statistics to be robust to the normality assumption [14-17]. However, to our knowledge, none of the methods estimate the heritability itself. By interpreting the heritability as a regression parameter under the GEE framework, the proposed approach has the following advantages: 1) exibility to model the genetic and non-genetic differences between two family members; 2) not requiring the normality assumption on the trait of interest; 3) simplicity to express the heritability adjusted for pair specific non-genetic factors; 4) robust estimation of the variance of the heritability estimate; 5) simplified inference as discussed in [18].We also extended our approach to accommodate multiple trait heritability analysis. Although [19] discussed multiple trait analysis using variance component models, there is no analytic method available for multiple trait analysis using correlations from multi-generation families.

Simulation results showed that there was improvement using the proposed method when true heritability was high on the trait of interest or when the trait does not follow the normality. The likelihood approach increases the bias and greatly underestimate the actual variance when the model is misspecified. However, when the trait follows the normality, the likelihood approach performed better than the proposed method. The model (2) has been extended for multipoint quantitative trait linkage analysis [20]. In that context, zero environmental correlation is often assumed. For the heritability estimation assuming zero environmental correlation, the biases from both approaches were not negligible when the environmental correlation was high as expected, which made both approaches inaccurate (Results were presented for the likelihood approach only). Since zero environmental correlation is very unlikely true for all relative pairs, it is desirable to interpret the heritability adjusted for non-zero correlation of environmental effects in the model to measure the genetics on a quantitative trait. The proposed methods provide a direct interpretation for the adjusted heritability from the model including pair specific genetic and non-genetic effects in the analysis. Data analysis implemented the heritability adjusted for a common non-genetic correlation for all pairs and pair specific differences in the model. Non-genetic correlations, which is an intercept in the correlation regression model, were weak for the three carotid IMT measures (CCAx, BIFx and ICAx) in single trait analysis. In single trait analysis, those differences in age or BMI were not significant, but age difference was significantly different from zero in multiple trait analysis. Regardless, both analyses showed strong genetic components in them.

Model			CCAx			BIFx			ICAx		
			PE	SE	p	PE	SE	p	PE	SE	p
Corr	Model1	Int	0.037	0.001	0.287	0.023	0.001	0.522	0.022	0.001	0.522
		σ_b^2	0.581	0.019	0.000	0.774	0.017	0.000	0.375	0.010	0.000
		σ_e^2	0.389	0.011	0.000	0.191	0.011	0.073	0.549	0.007	0.000
	Model2	Int	0.034	0.002	0.441	0.033	0.003	0.560	0.036	0.002	0.452
		σ_b^2	0.540	0.018	0.000	0.732	0.017	0.000	0.339	0.011	0.001
		σ_e^2	0.440	0.010	0.000	0.235	0.013	0.043	0.578	0.009	0.000
		Dage	0.025	0.000	0.054	0.027	0.000	0.065	0.021	0.000	0.157
		dbmi	-0.019	0.000	0.235	-0.031	0.000	0.048	-0.029	0.000	0.064

PE: parameter estimate; SE: standard error from the sandwich variance estimate; p: p-value; Model1: unadjusted for pair specic age and BMI dierences; Model2: adjusted for pair specic age and BMI dierences; dage: age dierence; dbmi: BMI dierence

Table 7: Heritability estimates for CCAx, BIFx and ICAx using the proposed method: Single trait analysis

Current paper is focused on polygenic heritability, but locus specific heritability, which uses genotypes, can be a direct extension from this model. This new regression approach may reduce the bias resulting from not accounting for environmental effects in quantitative genetics and may increase the power to locate genes on quantitative traits. It is also well known that the marginal approach can produce negative variance component estimates. In likelihood approach, a hierarchical model can accommodate to have a positive variance estimate. Our approach can also produce a negative heritability estimate, but a transformation for the regression parameter, such as logit, can range the heritability estimate from 0 to 1.

Acknowledgment

This work was supported by grants from the National Institutes of Health (NIH) U54RR019259-01 and from the National Institutes of Diabetes and Digestive and Kidney Diseases (NIDDK) 1-U01-DK63790-02. This work was also supported by grant from the National Institute of Neurological Disorders and Stroke (NINDS) R01 NS40807.

Reference

1. Falconer DS, Mackay TFC (1996) Introduction to quantitative ge-netics. (4thedn) Harlow UK: Longman.
2. Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer Associates.
3. Mallinckrodt CH, Golden BL, Reverter A (1997) Approximate con_- dence intervals for heritability from method R estimates. J Anim Sci 75: 2041-2046.
4. Hsu FC, Zaccaro DJ, Lange LA, Arnett DK, Langefeld CD, et al. (2005) The impact of pedigree structure on heritability estimates for pulse pressure in three studies. Hum Hered 60: 63-72.
5. Ekstrøm CT (2009) The impact of pedigree structure on heritability esti- mates. Hum Hered 68: 243-251.
6. Schaid DJ, Olson JM, Gauderman WJ, Elston RC (2003) Regression models for linkage: Issues of traits, covariates, heterogeneity, and interpreta- tion. Hum Hered 55: 86-96.
7. Zeegers M, Rijsdijk F, Sham P (2004) Adjusting for covariates in variance components QTL linkage analysis. Behavior Genetics 34: 127-133.
8. Lee HS, Paik MC, Lee JH (2008) Genotype adjusted familial corre-lation analysis using three generalized estimating equations. Statistics in Medicine 27: 5471-5483.
9. Lee HS, Paik MC, Lee JH (2009) Estimating a multivariate famil-ial correlation using Joint models for canonical correlations: Application to memory score analysis from familial Hispanic Alzheimer's disease study. Bio-metrics 65: 463-469.
10. Lange K, Westlake J, Spence M (1976) Extensions to pedigree analysis.III. Variance components by the scoring method. Ann Hum Genet 39: 485-491.
11. Hopper JL, Mathews JD (1982) Extensions to multivariate normal mod-els for pedigree analysis. Ann Hum Genet 46: 373-383.
12. Sacco RL, Blanton SH, Slifer S, Beecham A, Glover K (2009) Heritability and linkage analysis for carotid intima-media thickness: the family study of stroke risk and carotid atherosclerosis. Stroke 40: 2307-2312.
13. Sacco RL, Sabala EA, Rundek T, Juo SH, Huang S, et al. (2007) Design of a Family Study Among High Risk Caribbean Hispanics: The Northern Manhattan Family Study. Ethnicity and Disease 17: 351-357.
14. Amos CI (1994) Robust variance components approach for assessing ge-netic linkage in pedigrees. Am J Hum Genet 54: 535-543.
15. Sham PC, Purcell S, Cherny SS, Abecasis GR (2002) Powerful regression-based quantitative-trait linkage analysis of general pedigrees. Am J Hum Genet 71: 238-253.
16. Wang K (2005) A likelihood approach for quantitative-traitlocusmapping with selected pedigrees. Biometrics 61: 465-473.
17. Bhattacharjee S, Kuo C-L, Mukhopadhyay N, Brock GN, Weeks DE, et al.(2008) Robust Score Statistics for QTL Linkage analysis. Am J Hum Genet 82: 567-582.
18. Rotnitzky A, Jewell NP (1990) Hypothesis testing of regression pa-rameters in semiparametric generalized linear models for cluster correlated data. Biometrika 77: 485-497.
19. Sung YJ, Dawson G, Munson J, Estes A, Schellenberg GD, et al. (2005) Genetic investigation of quantitative traits related to Autism:Use of multivariate polygenic models with ascertainment adjustment. Am J Hum Genet 76: 68-81.
20. Almasy L, Blangero J (1998) Multipoint quantitative trait linkage analysis in general pedigrees. Am J Hum Genet 62: 1198-1211.