CrossMark

ORIGINAL RESEARCH

A New Look at Genetic and Environmental Architecture on Lipids Using Non-Normal Structural Equation Modeling in Male Twins: The NHLBI Twin Study

Sheng-Hui Wu[1] · Koken Ozaki[2] · Terry Reed[3] · Ruth E Krasnow[4] · Jun Dai[5,6]

Abstract This study examined genetic and environmental influences on the lipid concentrations of 1028 male twins using the novel univariate non-normal structural equation modeling (nnSEM) ADCE and ACE models. In the best fitting nnSEM ADCE model that was also better than the nnSEM ACE model, additive genetic factors (A) explained 4%, dominant genetic factors (D) explained 17%, and common (C) and unique (E) environmental factors explained 47% and 33% of the total variance of high-density lipoprotein cholesterol (HDL-C). The percentage of variation explained for other lipids was 0% (A), 30% (D), 34% (C) and 37% (E) for low-density lipoprotein cholesterol (LDL-C); 30, 0, 31 and 39% for total cholesterol; and 0, 31, 12 and 57% for triglycerides. It was concluded that additive and dominant genetic factors simultaneously affected HDL-C concentrations but not other lipids. Common and unique environmental factors influenced concentrations of all lipids.

Keywords Genetic effect · Environmental effect · Lipids · Twin · Non-normal structural equation modeling

Edited by Sarah Medland.

Sheng-Hui Wu and Koken Ozaki have contributed equally to this work.

Introduction

In a twin study, a trait or phenotype is conceptually determined by four latent effects: additive genetic (A), dominant genetic (D), common environmental (C), and unique environmental (E) effects. These latent effects can be statistically estimated from phenotype data collected from monozygotic and dizygotic twins using structural equation modeling. Traditional structural equation modeling (SEM) cannot estimate A, D, C, and E effects simultaneously because C and D are confounded in the classical twin study (Neale and Maes 1992; Rijsdijk and Sham 2002). As a result, in twin studies, C and D together cannot be included in the SEM. Instead, ACE and ADE, but not ADCE or

✉ Jun Dai
Jun.Dai@dmu.edu

1 Department of Epidemiology & Biostatistics-Laredo Campus, The University of Texas Health at San Antonio, Laredo, TX, USA

2 Graduate School of Business Sciences, University of Tsukuba, Tokyo, Japan

3 Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

4 Center for Health Sciences, SRI International, Menlo Park, CA, USA

5 Department of Public Health, Des Moines University, Des Moines, IA, USA

6 Division of Epidemiology, Department of Medicine, Institute of Medicine and Public Health, Vanderbilt Center for Translational and Clinical Cardiovascular Research (VTRACC), Vanderbilt University Medical Center, Nashville, TN, USA

CDE, are used for the SEM (Neale and Maes 1992; Rijsdijk and Sham 2002). However, it is biologically possible that D and C affect one trait simultaneously. Ignoring D and C co-existence would lead to biased estimates (Neale and Maes 1992; Rijsdijk and Sham 2002). To address this analytic limitation, Ozaki et al. recently published non-normal structural equation modeling (nnSEM) that could estimate additive and dominant genetic influences on and environmental contribution to univariate continuous traits simultaneously using ADCE model (Ozaki et al. 2011). In the ADCE model, each of A, D, C or E designates a single latent factor. A four-factor model is defined as all of the four factors (ADCE) affecting the trait. A "three-factor" model is a model in which only three of these four factors are presumed to contribute to the variance of a trait. The sum of additive (A) and dominant (D) genetic influences is a global genetic influence (G) and can be estimated if either three-factor (i.e., ADE) or four-factor (i.e., ADCE) nnSEM (Ozaki et al. 2011) is the best fitting model. Appendix A shows the brief statistical introduction to the ADCE model using the nnSEM.

Although genetic and environmental contributions to lipids have been assessed using the traditional SEM (Goode et al. 2007; Heller et al. 1993; Jermendy et al. 2011; O'Connell et al. 1988; Snieder et al. 1999), to date, no studies have applied the novel nnSEM to elucidate additive and dominant genetic influences on lipids along with environmental impact. Circulating levels of high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC) and triglycerides (TG), reflecting the lipid profile, are well known factors for cardiovascular diseases (Brownson et al. 1998). Using the SEM, prior twin studies have found that additive genetic factors contribute to plasma levels of HDL-C (Goode et al. 2007; Heller et al. 1993; Jermendy et al. 2011), LDL-C (Goode et al. 2007; Heller et al. 1993), TC (Goode et al. 2007; Heller et al. 1993) and TG (Goode et al. 2007; Jermendy et al. 2011). For example, estimated from the ACE SEM model, the longitudinal patterns of additive genetic effect on a lipid ranged from 0.46 to 0.57 for TC, 0.49 to 0.64 for LDL-C, 0.50 to 0.62 for HDL-C and 0.28 to 0.61 for TC among 456 twin pairs in the National Heart, Lung and Blood Institute (NHLBI) Twin Study (Goode et al. 2007; Jermendy et al. 2011). Since simultaneous inclusion of both C and D in the SEM is not possible, estimation of the genetic and environmental effect in those prior studies might be biased. It is worthy to note that the SEM and the nnSEM are not comparable as they analyze different sample statistics and are based on different statistical methods. However, it is intriguing to understand how different ACE and ADCE models would be. Therefore, the method to estimate three-factor models using the nnSEM was developed for our study (Appendix B), which allowed the comparison

between nnSEM ADCE and nnSEM ACE models using model fit indices. In this reported study, using the nnSEM as an alternative approach, we attempted to shed new light on the genetic and environmental influence on circulating fasting lipid concentrations.

Materials and methods

Study population

The National Heart, Lung, and Blood Institute (NHLBI) Twin Study has been widely described (Dai et al. 2013; Reed et al. 1993; The U.S.National Heart Lung and Blood Institute (NHLBI) 2005). The NHLBI Twin Study was designed to prospectively investigate the genetic and environmental role in cardiovascular disease risk through inclusion of 514 white male middle-aged veteran twin pairs born in 1917–1927 at baseline (1969–1973) (Dai et al. 2013; Reed et al. 1993). Based on zygosity ascertained by eight red blood cell antigen groups (serotyping 22 erythrocyte antigens) in the 1960s and variable number of tandem repeat DNA markers in the 1980s (The U.S.National Heart Lung and Blood Institute (NHLBI) 2005), the baseline sample included 253 MZ and 261 DZ pairs (The U.S.National Heart Lung and Blood Institute (NHLBI) 2005). Informed consent was obtained from all individual participants included in the study.

In this study, we included a total of 1028 twins who had baseline data on fasting plasma lipid profile. The reported study was approved by the Institutional Review Boards of Vanderbilt University.

Measures of the lipid profile

Blood was drawn from the forearm vein after an overnight fast into EDTA tubes and immediately placed on ice. After centrifugation the plasma was aliquoted and frozen at −70 °C (Feinleib et al. 1977; Reed et al. 1994). Plasma lipid fractions were measured in one of three laboratories following standards of the Centers for Disease Control, located in San Francisco, Indianapolis, and Framingham (Selby et al. 1991). LDL-C was estimated (Friedewald et al. 1972) for individuals with measured TG concentration less than 400 mg/dL (Sampson et al. 1975). Plasma samples from a twin pair were assayed in the same analytical run without knowing zygosity.

Statistical analysis

Plasma levels of HDL-C, LDL-C, TC and TG were the continuous variables used for the analyses. Univariate analysis was performed using the nnSEM. This

model assumes that variations in phenotypic traits can be decomposed into latent factors: additive genetic (A), common environmental (C), dominant genetic (D) and unique environmental (E) factors. In nnSEM, the "non-normal" means that some independent variables (like D, C, and/or E in our study) and dependent variables (like lipid phenotypes in our study) are non-normally distributed. We first analyzed three different nnSEM ADCE full models, in which A was normally distributed, but C, D and E could not be normally distributed simultaneously in one model. The distribution of C, D and E in the three ADCE models was non-normal for C and E in model 1; non-normal for D and E in model 2; and non-normal for C, D, and E in model 3 (Ozaki et al. 2011). The best model was selected based on the goodness of fit to data. A smaller Bayesian Information Criterion (BIC) or Root Mean Square Error of Approximation (RMSEA) indicates a better fit to the data (Ozaki et al. 2011). Unlike the SEM in which (first- and) second-order moments, namely, (means and) covariances are used as information, the nnSEM uses higher order moments as well as (first- and) second-order moments (Ozaki et al. 2011) (Appendix A). The genetic effect (heritability) was represented with the "A" and "D" components using the nnSEM. The sum (G) of additive (A) and dominant (D) genetic effect (i.e. $G = A + D$) is a good estimator of the global genetic effect for both three and four factor cases (Ozaki et al. 2011). In order to compare the nnSEM ADCE full model with its nested reduced models, we developed the nnSEM method for the ACE model in the reported study given that $r_{MZ} < 2r_{DZ}$ from the Supplemental Table 1, (code can be accessed at http://www010.upp.so-net.ne.jp/koken/bg.html) where r_{MZ} and r_{DZ} are correlation coefficients

between co-twins for MZ and DZ twins, respectively. This new nnSEM ACE method can estimate four ACE models, four CE models, one AE model and one E model using 2nd and 3rd order moments. For the comparisons among the nnSEM ADCE full and its nested reduced models, generally, a smaller BIC or RMSEA indicates a better fit to the data (Ozaki et al. 2011). However, if the BIC for one model is slightly larger than the other model but the RMSEA values at or greater than 0.05 for both models, the model that has a smaller RMSEA may be better (for example, the ACE model 4 vs CE model 8 for TC in Table 3). In another situation, if the BIC for the ADCE full model vs the other reduced model is small but the RMSEA is slightly large, the BIC indicates the ADCE full model as the best fit model while the RMSEA indicates the reduced model as the best fit model. Given that both BIC and RMSEA are the model fit evaluation indices that incorporate the model parsimony (Loelin 2004), it is arbitrary to use either the BIC or the RMSEA to select the best fit model. Theoretically or conceptually, the possibility of the importance of all of four factors (i.e. A, D, C, and E) to explain the phenotype cannot be exclusively ruled out; thus, in this study, the ADCE full model is preferred as the best fit model. The reduced model might also fit well. The detailed statistical method was shown in Appendix B.

The nnSEM was conducted using the R software (Ozaki et al. 2011). To provide a full view of genetic and environmental contributions to a phenotype, we additionally conducted the traditional SEM-using the Mx software (Neale et al. 1994; Posthuma and Boomsma 2005). The other analyses were performed using the SAS 9.1 (SAS Institute, Cary, NC) statistical package.

Table 1 Age and lipid profile in the fasting plasma at baseline in NHLBI twins study

	Monozygotic twins	Dizygotic twins	Total
n	506 (253 pairs)	522 (261 pairs)	1028
Age, mean (SD), years	47.8 (3.1)	47.9 (3.2)	47.8 (3.1)
HDL-C[a]			
Mean (SD), mg/dL	44.5 (12.8)	46.2 (14.7)	45.3 (13.8)
LDL-C			
Mean (SD), mg/dL	144.3 (33.6)	142.9 (38.4)	143.6 (36.1)
TC			
Mean (SD), mg/dL	220.6 (35.1)	220.0 (41.6)	220.3 (38.5)
TG			
Mean (SD), mg/dL	135.1 (88.6)	131.1 (101.8)	131.1 (95.5)

A total of 1028 twins (253 MZ and 261 DZ twin pairs) were included for the analyses

HDL-C high density lipoprotein cholesterol; *LDL-C* low-density lipoprotein cholesterol; *SD* standard deviation; *TC* total cholesterol; *TG* triglyceride

[a] $P < 0.05$ for comparing MZ and DZ twins. *P* values were derived from mixed models for continuous variables to account for clustering within twin pairs

Results

Characteristics of the study population

Among 1028 twins (506 MZ and 522 DZ twins), the average age was 47.8 years, and similar between MZ and DZ twins. The mean (SD) of the concentrations of lipid markers were also similar between MZ and DZ twins (Table 1). The concentrations of HDL-C were higher in DZ than MZ twins, while those of LDL-C, TC and TG tended to be lower in DZ than MZ twins (Table 1).

Univariate nnSEM ADCE models

The ADCE model 1 with non-normally distributed C and E was the best fitting model for each of the lipid markers using the nnSEM (Table 2), and was used to estimate genetic and environmental contribution. Supplemental Table 1 shows the descriptive statistics for MZ and DZ pairs used in the nnSEM estimation. The genetic and environmental influences on HDL-C was 0.04 [95% confidence interval (CI): 0.00, 0.29] for additive genetic effect (A), 0.17 (95% CI: 0.00, 0.35) for dominant genetic effect (D), 0.47 (95% CI: 0.37, 0.56) for common environmental effect (C) and 0.33 (95% CI: 0.30, 0.36) for unique environmental effect (E) (Model 1). The global genetic contribution (i.e., heritability) to HDL-C was 21%. The A, D, C, and E estimates from the model 1 for LDL-C was 0.00 (95% CI: 0.00, 0.00), 0.30 (95% CI: 0.31, 0.38), 0.34 (95% CI: 0.26, 0.41), and 0.37 (95% CI: 0.33, 0.40), respectively. The heritability for LDL-C (30%) was mainly from dominant genetic variance. Estimated from the model 1 for TC, A was 0.30 (95% CI: 0.20, 0.40); D was 0.00 (95% CI: 0.00, 0.00); C was 31% (95% CI: 23%, 39%) and E was 39% (95% CI: 35%, 42%). The heritability for TC (30%) was from A. The ADCE for TG was 0.00 (95% CI: 0.00, 0.00) for A, 0.12 (95% CI: 0.07, 0.17) for D, 0.31 (95% CI: 0.26, 0.36) for C and 0.57 (95% CI: 0.50, 0.64) for E. The heritability from TG (12%) was mainly due to D.

Table 2 Parameter estimates from univariate nnSEM ADCE model for plasma lipid concentrations in monozygotic and dizygotic twins

Model fitting[a]	Components of variance estimates[b]					Goodness-of-fit indices[c]	
	a^2 (95% CI)	c^2 (95% CI)	d^2 (95% CI)	e^2 (95% CI)	G (A + D)[a]	BIC	RMSEA
HDL-C[d]							
Model 1[e]	0.04 (0.00, 0.29)	0.47 (0.37, 0.56)	0.17 (0.00, 0.35)	0.33 (0.30, 0.36)	0.21	−43.1	0.029
Model 2	0.00 (0.00, 0.00)	0.30 (0.24, 0.36)	0.41 (0.35, 0.48)	0.29 (0.26, 0.31)	0.41	−36.0	0.038
Model 3	0.00 (0.00, 0.00)	0.40 (0.32, 0.47)	0.30 (0.21, 0.38)	0.31 (0.28, 0.34)	0.30	−31.4	0.051
LDL-C[d]							
Model 1[e]	0.00 (0.00, 0.00)	0.34 (0.26, 0.41)	0.30 (0.21, 0.38)	0.37 (0.33, 0.40)	0.30	−46.2	0.013
Model 2	0.00 (0.00, 0.00)	0.32 (0.25, 0.40)	0.32 (0.24, 0.40)	0.36 (0.32, 0.39)	0.32	−40.6	0.017
Model 3	0.10 (0.00, 0.34)	0.32 (0.20, 0.43)	0.22 (0.05, 0.39)	0.37 (0.33, 0.40)	0.32	−40.0	0.021
TC[d]							
Model 1[e]	0.30 (0.20, 0.40)	0.31 (0.23, 0.39)	0.00 (0.00, 0.00)	0.39 (0.35, 0.42)	0.30	−38.4	0.044
Model 2	0.04 (0.00, 0.32)	0.44 (0.31, 0.56)	0.12 (0.00, 0.31)	0.40 (0.36, 0.43)	0.16	−33.7	0.045
Model 3	0.00 (0.00, 0.50)	0.40 (0.35, 0.46)	0.21 (0.15, 0.27)	0.39 (0.35, 0.42)	0.21	−32.1	0.049
TG[d]							
Model 1[e]	0.00 (0.00, 0.00)	0.31 (0.26, 0.36)	0.12 (0.07, 0.17)	0.57 (0.50, 0.64)	0.12	−44.4	0.024
Model 2	0.00 (0.00, 0.00)	0.16 (0.12, 0.20)	0.37 (0.26, 0.49)	0.47 (0.41, 0.53)	0.37	−36.0	0.038
Model 3	0.00 (0.00, 0.00)	0.26 (0.19, 0.33)	0.21 (0.10, 0.33)	0.53 (0.45, 0.60)	0.21	−39.0	0.026

CI confidence interval; *HDL-C* high-density lipoprotein cholesterol; *LDL-C* low-density lipoprotein cholesterol; *nnSEM* non-normal structural equation modeling; *TC* total cholesterol; *TG* triglyceride

[a]A, C, D, E, and G refer to additive genetic, common environmental, dominance, unique environmental and global genetic influences, respectively. G is equivalent to g^2 ($g^2 = a^2 + d^2$)

[b]a^2, c^2, d^2, and e^2 are estimates of the proportion of additive genetic, common environmental, dominant genetic, and unique environmental components of variance, respectively, calculated for the different structural equation models

[c]Goodness-of-fit statistic. Model fits are summarized by Bayesian Information Criterion (BIC) or Root Mean Square Error of Approximation (RMSEA), with lower BIC or RMSEA indicating a better fit

[d]Model 1: C and E are non-normally distributed; Model 2: D and E are non-normally distributed; Model 3: C, D, and E are non-normally distributed

[e]Best fitting model if the BIC or RMSEA is the lowest among three models

Table 3 Parameter estimates for ACE, CE, AE and E Models from univariate nnSEM model for plasma lipid concentrations in monozygotic and dizygotic twins

Model fitting[a]	Components of variance estimates[b]				Goodness-of-fit indices[c]	
	a² (95% CI)	c² (95% CI)	e² (95% CI)	G (A)[a]	BIC	RMSEA
HDL[d]						
ACE						
Model 1	0.29 (0.19, 0.39)	0.38 (0.27, 0.49)	0.33 (0.29, 0.36)	0.29	−24.5	0.048
Model 2[e]	0.25 (0.17, 0.33)	0.42 (0.33, 0.52)	0.33 (0.29, 0.36)	0.25	−29.9	0.043
Model 3	0.25 (0.15, 0.34)	0.44 (0.31, 0.56)	0.32 (0.28, 0.35)	0.25	−12.7	0.082
Model 4	0.38 (0.26, 0.51)	0.21 (0.13, 0.29)	0.41 (0.36, 0.45)	0.38	−10.0	0.087
CE						
Model 5		0.62 (0.53, 0.71)	0.38 (0.35, 0.42)	0.00	−22.0	0.064
Model 6[f]		0.62 (0.53, 0.71)	0.38 (0.35, 0.42)	0.00	−28.2	0.058
Model 7		0.64 (0.55, 0.73)	0.36 (0.33, 0.39)	0.00	−14.0	0.083
Model 8		0.52 (0.46, 0.58)	0.48 (0.44, 0.52)	0.00	−6.2	0.094
AE						
Model 9	0.62 (0.53, 0.7)		0.39 (0.35, 0.42)	0.62	−11.9	0.086
E						
Model 10			1.00 (0.93, 1.07)	0.00	84.0	0.170
LDL[d]						
ACE						
Model 1	0.46 (0.28, 0.63)	0.18 (0.08, 0.28)	0.37 (0.32, 0.41)	0.46	−29.3	0.026
Model 2	0.42 (0.26, 0.58)	0.21 (0.11, 0.32)	0.36 (0.32, 0.40)	0.42	−33.8	0.028
Model 3	0.45 (0.27, 0.62)	0.19 (0.09, 0.29)	0.36 (0.32, 0.40)	0.45	−34.1	0.026
Model 4[e]	0.45 (0.27, 0.63)	0.18 (0.08, 0.28)	0.37 (0.33, 0.41)	0.45	−34.2	0.026
CE						
Model 5		0.56 (0.49, 0.64)	0.44 (0.4, 0.48)	0.00	−24.4	0.059
Model 6		0.56 (0.49, 0.64)	0.44 (0.4, 0.48)	0.00	−29.2	0.056
Model 7		0.57 (0.49, 0.65)	0.43 (0.39, 0.47)	0.00	−29.7	0.055
Model 8[f]		0.56 (0.49, 0.64)	0.44 (0.34, 0.54)	0.00	−29.8	0.055
AE						
Model 9	0.64 (0.55, 0.73)		0.36 (0.32, 0.40)	0.64	−38.2	0.029
E						
Model 10			1.00 (0.92, 1.08)		72.4	0.163
TC[d]						
ACE						
Model 1	0.25 (0.15, 0.35)	0.36 (0.25, 0.47)	0.40 (0.35, 0.44)	0.25	−21.0	0.058
Model 2	0.30 (0.19, 0.41)	0.31 (0.22, 0.40)	0.39 (0.35, 0.43)	0.30	−25.9	0.055
Model 3	0.27 (0.16, 0.37)	0.36 (0.25, 0.46)	0.38 (0.33, 0.42)	0.27	−17.6	0.073
Model 4[e]	0.25 (0.15, 0.35)	0.35 (0.25, 0.46)	0.40 (0.35, 0.44)	0.25	−26.7	0.053
CE						
Model 5		0.55 (0.48, 0.62)	0.45 (0.41, 0.49)	0.00	−21.8	0.064
Model 6		0.55 (0.48, 0.61)	0.45 (0.41, 0.49)	0.00	−24.3	0.066
Model 7		0.56 (0.49, 0.63)	0.44 (0.30, 0.58)	0.00	−18.2	0.076
Model 8[f]		0.55 (0.48, 0.61)	0.46 (0.42, 0.49)	0.00	−27.5	0.059
AE						
Model 9	0.64 (0.55, 0.72)		0.37 (0.32, 0.41)	0.64	−18.5	0.076
E						
Model 10			1.00 (0.93, 1.07)	0.00	94.4	0.175
TG[d]						
ACE						
Model 1	0.42 (0.27, 0.57)	0.18 (0.09, 0.26)	0.41 (0.36, 0.45)	0.42	−31.26	0.005
Model 2[e]	0.42 (0.27, 0.57)	0.18 (0.09, 0.26)	0.41 (0.36, 0.45)	0.42	−37.39	0.000

Model fitting[a]	Components of variance estimates[b]				Goodness-of-fit indices[c]	
	a² (95% CI)	c² (95% CI)	e² (95% CI)	G (A)[a]	BIC	RMSEA
Model 3	0.42 (0.27, 0.57)	0.18 (0.09, 0.26)	0.41 (0.36, 0.45)	0.42	−37.36	0.000
Model 4	0.42 (0.27, 0.57)	0.17 (0.09, 0.26)	0.41 (0.36, 0.45)	0.42	−37.36	0.000
CE						
Model 5		0.51 (0.43, 0.58)	0.50 (0.45, 0.54)	0.00	−25.37	0.055
Model 6		0.50 (0.43, 0.57)	0.50 (0.45, 0.54)	0.00	−31.92	0.049
Model 7[f]		0.50 (0.43, 0.57)	0.50 (0.45, 0.54)	0.00	−32.04	0.049
Model 8		0.50 (0.43, 0.57)	0.50 (0.45, 0.54)	0.00	−31.98	0.049
AE						
Model 9	0.61 (0.52, 0.7)		0.39 (0.35, 0.43)	0.61	−40.9	0.015
E						
Model 10			1.00 (0.92, 1.08)	0.00	53.6	0.149

Table 3 (continued)

CI confidence interval; *HDL-C* high-density lipoprotein cholesterol; *LDL-C* low-density lipoprotein cholesterol; *nnSEM* non-normal structural equation modeling; *TC* total cholesterol; *TG* triglyceride

[a] A, C, E, and G refer to additive genetic, common environmental, unique environmental and global genetic influences, respectively

[b] a^2, c^2, and e^2 are estimates of the proportion of additive genetic, common environmental, and unique environmental components of variance, respectively, calculated for the different structural equation models. The R script for the nnSEM ACE model can be accessed at http://www010.upp.sonet.ne.jp/koken/bg.html

[c] Goodness-of-fit statistic. Model fits are summarized by Bayesian Information Criterion (BIC) or Root Mean Square Error of Approximation (RMSEA), with lower BIC or RMSEA indicating a better fit

[d] Model 1: ACE model: C and E are non-normally distributed and have the different skewness; Model 2: ACE model: C and E are non-normally distributed and have the same skewness; Model 3: ACE model: only C is non-normally distributed; Model 4: ACE model: only E is non-normally distributed; Model 5: CE model: C and E are non-normally distributed and have the different skewness; Model 6: CE model: C and E are non-normally distributed and have the same skewness; Model 7: CE model: only C is non-normally distributed; Model 9: AE model: E is non-normally distributed; Model 10: E model: E is non-normally distributed

[e] Best fitting model if the BIC or RMSEA is the lowest among the four ACE models

[f] Best fitting model if the BIC or RMSEA is the lowest among the four CE models

Univariate nnSEM ACE model and its nested reduced models

Table 3 shows the estimates for 10 nnSEM ACE models including 4 ACE full models and its nested reduced models (i.e., 4 CE, 1 AE, and 1 E models). For HDL-C, model 2 was the best nnSEM ACE full model and model 6 was the best CE model. Among 10 nnSEM ACE models, the best fitting model was the ACE full model 2 for HDL-C. Similarly, among 10 nnSEM ACE models, the best fitting model was AE model for LDL-C, ACE model 4 for TC and AE for TG.

The best fitting nnSEM ADCE model 1 for all lipids was compared with the lipid-specific best fitting nnSEM ACE model using BIC or RMSEA as the model fit indices. The nnSEM ADCE model 1 provided better fitting for HDL-C relative to the nnSEM ACE model 2 (Tables 2, 3: −43.1 vs 29.9 for BIC and 0.029 vs 0.043 for RMSEA); for LDL-C relative to the nnSEM AE model (Tables 2, 3: −46.2 vs

−38.2 for BIC and 0.013 vs 0.029 for RMSEA); and for TC, the nnSEM ADCE model 1 relative to the nnSEM ACE model 4 (Tables 2, 3: −38.4 vs −26.7 for BIC and 0.044 vs 0.053 for RMSEA). For TG, the nnSEM ADCE model 1 generated smaller BIC but slightly larger RMSEA than the nnSEM AE model (Tables 2, 3: −44.4 vs −40.9 for BIC and 0.024 vs 0.015 for RMSEA). Given the smaller BIC for the nnSEM ADCE model 1 for TG, the nnSEM ADCE model was better than the nnSEM AE model for TG. The nnSEM ADCE model 1 was the best fitting model for each of the lipid markers with the nnSEM.

Discussion

Using the novel nnSEM that could analyze the influence of common environment and dominant genes simultaneously on a trait, we found that the additive and dominant genetic, and common and unique environmental influences

simultaneously affected the variance of HDL-C, but either additive or dominant genetic and environmental factors influenced total variance of LDL-C, TC and TG.

To our knowledge, this is the first study to estimate the genetic and environmental influences on lipid markers using the nnSEM ACE and ADE models. The ADCE model could not be identified by the traditional SEM because C and D are confounded in the classical twin study (Neale and Maes 1992; Rijsdijk and Sham 2002). In contrast, the nnSEM could specify an ADCE model and compare which model is the best model among three types of models, none of which are nested submodels of one another. We newly developed an nnSEM method to estimate its nested reduced models, which were compared with the nnSEM ADCE full model, and found that the ADCE is the best fitting model for all markers in this study. If the ADCE model is the true model, the ACE and ADE models yield biased estimates (Ozaki et al. 2011). Using the ADCE model, the nnSEM could estimate global genetic effect (G) [the sum of additive genetic (A) and dominant genetic (D) effect], which cannot be realized in the classic SEM. Furthermore, the global genetic effect is also a good estimator of the global genetic effect for the three factor case like the nnSEM ADE model (Ozaki et al. 2011).

Although there are no methods to compare the traditional SEM and our nnSEM, it might be interesting to explore the apparent differences in heritability estimated from the SEM. The heritability from our study estimated by the nnSEM was a little lower than those using the SEM from our study (Supplemental Table 2) and previous studies for HDL-C (0.36–0.76) (Goode et al. 2007; Heller et al. 1993; Jermendy et al. 2011; O'Connell et al. 1988; Snieder et al. 1999) and LDL-C (0.22-1.00) (Goode et al. 2007; Heller et al. 1993; Snieder et al. 1999). The results from the SEM may overestimate the additive genetic component but underestimate the dominant genetic component (Neale and Maes 1992; Ozaki et al. 2011; Rijsdijk and Sham 2002). However, estimated from the nnSEM, the heritability for LDL-C, TC and TG in our study was in the range of previously reported heritability, TC (0.00–0.80) (Goode et al. 2007; Heller et al. 1993; Jermendy et al. 2011; O'Connell et al. 1988; Snieder et al. 1999) and TG (0.19–0.81) (Goode et al. 2007; Jermendy et al. 2011; Snieder et al. 1999). The dominant genetic effect of LDL-C (35%) was also observed in 12,000 Swedish twins born between 1911 and 1958 estimated by the ADE model using the SEM (Rahman et al. 2009). Compared to previous studies, their findings of dominant genetic effects may be due to the enhanced power of the large and homogenous sample in their study, enabling them to detect weaker variance components underlying the phenotypic traits. Another contributing factor may be the older age of their study participants, possibly leading

to decreased influences from shared familial environment. In addition, the effect of common environment is under-estimated. In our study, the best nnSEM model for all four lipid markers was model 1 with little additive genetic effects: common environment accounted for 31–47% of the variance. By contrast, the best SEM model for all four markers was the AE model, from which the effect of common environment was estimated as zero. Even the SEM ACE full model estimated the smaller effect of common environment than the nnSEM best model (model 1) for each lipid marker. When all four factors (i.e. ADCE) affect a trait, given the mathematical comparisons between the nnSEM and the SEM, the SEM always underestimates variance of common environment (Neale and Maes 1992; Rijsdijk and Sham 2002) but the nnSEM does not (Ozaki et al. 2011).

Potential limitations of our study require acknowledgement. Given that the univariate nnSEM ADCE model is developed only for univariate continuous traits, we could not analyze the genetic and environmental influence on a categorical trait such as hypercholesterolemia, and we were unable to adjust for the effect of potential confounding factors such as age, obesity and other lifestyle factors on the estimations. Our twins were white men, thus our results may not be generalizable to women and other ethnic groups.

In conclusion, as shown by the ADCE model fit via the novel nnSEM (which cannot be performed with the traditional SEM), additive- and dominant-genetic, and common- and unique- environmental influences simultaneously affected the concentrations of HDL-C; additive or dominant genetic effect as well as common and unique environmental effect simultaneously affected concentrations of LDL-C, TC and TG. Genetic factors and environmental factors were important determinants in concentrations of lipid markers.

Compliance with ethical standards

Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Appendix A

In the classical twin study using the SEM, only (first- and) second-order moments, namely covariance matrix of both MZ pairs and DZ pairs, are used. However, in the ADCE model using the nnSEM, in addition to (first- and) second-order moments, third-order moments are used. When there are two phenotypes $p1$ and $p2$ (for twin 1 and twin 2, respectively), sample third-order moments are as follows:

$$S_{p_1^3} = \frac{1}{n}\sum_{i=1}^{n}\left(\text{pi1} - \overline{\text{p1}}\right)^3$$

$$S_{p_1^2 p2} = \frac{1}{n}\sum_{i=1}^{n}\left(\text{pi1} - \overline{\text{p1}}\right)^2\left(\text{pi2} - \overline{\text{p2}}\right)$$

$$S_{p1 p_2^2} \frac{1}{n}\sum_{i=1}^{n}\left(\text{pi1} - \overline{\text{p1}}\right)\left(\text{pi2} - \overline{\text{p2}}\right)^2$$

$$S_{p_2^3} = \frac{1}{n}\sum_{i=1}^{n}\left(\text{pi2} - \overline{\text{p2}}\right)^3$$

where i indicates an observation (in this case, an observation pair). These sample statistics become information in the estimation of a, c, d, and e effects.

When three latent variables C, D, and E are assumed to follow non-normal distributions, the expected third-order moments for the MZ twins are as follows:

$$E\left[p_1^3(mz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D^3} + e^3\sigma_{E^3}$$

$$E\left[p_1^2 p2(mz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D^3}$$

$$E\left[p_1 p_2^2(mz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D^3}$$

$$E\left[p_2^3(mz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D^3} + e^3\sigma_{E^3}$$

In addition, the expected third-order moments for the DZ twins are as follows:

$$E\left[p_1^3(dz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D^3} + e^3\sigma_{E^3}$$

$$E\left[p_1^2 p_2(dz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D1^2 D2}$$

$$E\left[p_1 p_2^2(dz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D1 D2^2}$$

$$E\left[p_2^3(dz)\right] = c^3\sigma_{C^3} + d^3\sigma_{D^3} + e^3\sigma_{E^3}$$

Note that A is assumed to follow a normal distribution when many loci affect the phenotype of interest. Here, σ_{c^3}, σ_{d^3} and σ_{E^3} are the skewness of C, D, and E, respectively. (In this case, the variances of these independent factors are fixed to 1 s. Therefore, these parameters are the

skewness of the factors.) $\sigma_{D1^2 D2^2}$ expresses the expected value of the second power of D of twin 1 times D of twin 2, and $\sigma_{D1 D2^2}$ expresses the expected value of D of twin 1 times the second power of D of twin 2, respectively. $\sigma_{D1^2 D2}$ and $\sigma_{D1 D2^2}$ are assumed to be equal, because the order of twins is arbitrary. Therefore, ADCE model using the nnSEM estimates A, C, D, and E influences by minimizing the discrepancy function, between the sample statistics (second- and third-order moments) and expected values (second- and third-order moments) using asymptotically distribution-free (ADF) method. For more details please see Ozaki et al. (2011) (Ozaki et al. 2011).

Appendix B

By dropping some parameters in the ADCE model using the nnSEM, nested reduced ACE, CE, and E models using 2nd and 3rd order moments can be identified. For example, four nnSEM ACE models can be identified by dropping D parameter in the ADCE model and by assuming: (1) C and E are non-normally distributed and different in the skewness, (2) C and E are non-normally distributed but have the same skewness, (3) only C is non-normally distributed, and (4) only E is non-normally distributed. In all of the four models A is normal. In this way, we can identify four ACE models, four CE models (1) C and E are non-normally distributed and different in the skewness, (2) C and E are non-normally distributed and have the same skewness, (3) only C is non-normally distributed, (4) only E is non-normally distributed, and one E model (E is non-normally distributed).

Because all of the ten models use the same sample statistics and are analyzed using the same estimation method, we can compare among the nnSEM ADCE and its nested reduced models using some fit indices.

References

Brownson RC, Remington PL, Davis JR (1998) Chronic disease epidemiology and control, 2nd edn. American Public Health Association, Washington.

Dai J, Krasnow RE, Liu L, Sawada SG, Reed T (2013) The Association between postload plasma glucose levels and 38-year mortality risk of coronary heart disease: the prospective NHLBI twin study. PLoS ONE 8(7):1–7

Feinleib M, Garrison RJ, Fabsitz R, Christian JC, Hrubec Z, Borhani NO, Kannel WB, Rosenman R, Schwartz JT, Wagner JO (1977) The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. Am.J.Epidemiol 106(4):284–285.

Friedewald WT, Levy RI, Fredrickson DS (1972) Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 18(6):499–502

Goode EL, Cherny SS, Christian JC, Jarvik GP, deAndrade M (2007) Heritability of longitudinal measures of body mass index and lipid and lipoprotein levels in aging twins. Twin Res Human Genet 10(5):703–711.

Heller DA, DeFaire U, Pedersen NL, Dahlen G, McClearn GE (1993) Genetic and environmental influences on serum lipid levels in twins. N Engl J Med 328(16):1150–1156.

Jermendy G, Horvath T, Littvay L, Steinbach R, Jermendy AL, Tarnoki AD, Tarnoki DL, Metneki J, Osztovits J (2011) Effect of genetic and environmental influences on cardiometabolic risk factors: a twin study. Cardiovasc Diabetol 10:96.

Loelin JC (2004) Latent variable models, 4th edn. Lawrence Erlbaum Associates, Mahwah

Neale MC, Maes HHM (1992) Methodology for genetic studies of twins and families. Kluwer Academic Publishers B.V., Dordrecht

Neale MC, Eaves LJ, Kendler KS (1994) The power of the classical twin study to resolve variation in threshold traits. Behav Genet 24(3):239–258

O'Connell DL, Heller RF, Roberts DC, Allen JR, Knapp JC, Steele PL, Silove D (1988) Twin study of genetic and environmental effects on lipid levels. Genet Epidemiol 5(5):323–341

Ozaki K, Toyoda H, Iwama N, Kubo S, Ando J (2011) Using non-normal SEM to resolve the ACDE model in the classical twin design. Behav Genet 41(2):329–339

Posthuma D, Boomsma DI (2005) Mx scripts library: structural equation modeling scripts for twin and family data. Behav Genet 35(4):499–505

Rahman I, Bennet AM, Pedersen NL, DeFaire U, Svensson P, Magnusson PK (2009) Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. Twin Res Human Genet 12(3):286–94.

Reed T, Carmelli D, Christian JC, Selby JV, Fabsitz RR (1993) The NHLBI male veteran twin study data. Genet Epidemiol 10(6):513–517

Reed T, Tracy RP, Fabsitz RR (1994) Minimal genetic influences on plasma fibrinogen level in adult males in the NHLBI twin study. Clin Genet 45(2):71–77

Rijsdijk FV, Sham PC (2002) Analytic approaches to twin data using structural equation models. Brief Bioinform 3(2):119–133

Sampson EJ, Demers LM, Krieg AF (1975) Faster enzymatic procedure for serum triglycerides. Clin Chem 21(13):1983–1985

Selby JV, Newman B, Quiroga J, Christian JC, Austin MA, Fabsitz RR (1991) Concordance for dyslipidemic hypertension in male twins. JAMA 265(16):2079–2084

Snieder H, van Doornen LJ, Boomsma DI (1999) Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: lessons from twin studies. Arterioscler Thromb Vasc Biol 19(12):2826–2834.

The U.S. National Heart Lung and Blood Institute (NHLBI) (2005) National heart, lung, and blood institute twin study. Clin Trials gov. Accessed 29 March 2013.