



Original article

Longitudinal analysis of sibling correlation on blood pressure using mixed modeling



Qihua Tan, MD, PhD^{a, b, *}, Hongmei Duan, MD^c, Ancong Wang, MD, PhD^d,
Dongyi Zhu, MD^d, Shuxia Li, MD, PhD^b

^a Department of Public Health, Epidemiology and Biostatistics, University of Southern Denmark, Odense, Denmark

^b Department of Clinical Research, Unit of Human Genetics, University of Southern Denmark, Odense, Denmark

^c Institute for Clinical Medicine, Copenhagen University, Denmark

^d Centre for Reproductive Medicine, Linyi People's Hospital, Linyi, China

ARTICLE INFO

Article history:

Received 5 November 2018

Accepted 20 February 2019

Available online 27 February 2019

Keywords:

Blood pressure

Sibling correlation

Longitudinal

Mixed-effects model

Fractional polynomials

ABSTRACT

Purpose: Although moderate to high genetic contribution to blood pressure variation have been estimated in numerous studies, the genetic control over the longitudinal change in blood pressure has been less frequently investigated because of the requirement of longitudinal design.

Methods: Based on blood pressure data from a large-scale family-based longitudinal survey, we introduced hierarchical modeling of longitudinal family data in combination with fractional polynomials for fitting nonlinear age patterns of blood pressure and the mixed-effect models for estimating sibling correlation on blood pressure to assess the genetic and shared environmental effects on blood pressure level as well as on the rate of change in blood pressure over ages.

Results: Significant sibling correlations were estimated on the levels of systolic blood pressure (0.2, 95% CI: 0.10–0.30) and diastolic blood pressure (0.28, 95% CI: 0.18–0.38), whereas for the longitudinal change or the rate of change, significant correlation was estimated only for diastolic blood pressure (0.13, 95% CI: 0.04–0.23). In the sex-specific analysis, similar pattern is observed, but statistical significance was only reached in female siblings with correlation estimates higher than the overall sample.

Conclusion: The rate of change in blood pressure is mainly influenced by individual's unique environment; and the genetic and common family environment may play a role in regulating the longitudinal change of diastolic but not systolic blood pressure.

© 2019 Elsevier Inc. All rights reserved.

Introduction

Hypertension is a comorbidity to both ischemic heart disease and stroke, the two topmost leading causes of death according to the WHO's recent update [1]. Hypertension also contributes significantly to the risk of metabolic syndrome [2], a condition that affects a large proportion of the world population. The levels of blood pressure trait have been shown to be affected by both genetic and environmental factors in twin and family-based studies [3, 4] and in recent genetic association studies [5, 6]. Although

moderate to high genetic contribution to blood pressure variation have been estimated in numerous studies, the genetic control over longitudinal changes in blood pressure has been less frequently investigated because of the requirement of longitudinal design. On the other hand, the level of blood pressure could change with age and some of the age patterns may not be linear. Proper adjustment for the confounding factors such as age is needed to correctly assess the genetic effect on the level and the rate of longitudinal change in blood pressure.

In the past decades, large-scale family-based longitudinal studies on complex diseases have been performed, such as the San Antonio Family Heart Study and San Antonio Family Diabetes/Gallbladder Study. The large sample sizes available allow for powerful analysis of traditional epidemiological data and genetic data resulting in rich research findings [7]. Although the results from intensive analysis on genetic data are interesting, analysis of the sizable and valuable epidemiological data can also provide

The authors declare no conflict of interest.

* Corresponding author. Department of Public Health, Epidemiology and Biostatistics, Faculty of Health Science, University of Southern Denmark, J. B. Winsløws Vej 9B, DK-5000 Odense C, Denmark. Tel.: +45 65503536; fax: +45 65411911.

E-mail address: qtan@health.sdu.dk (Q. Tan).

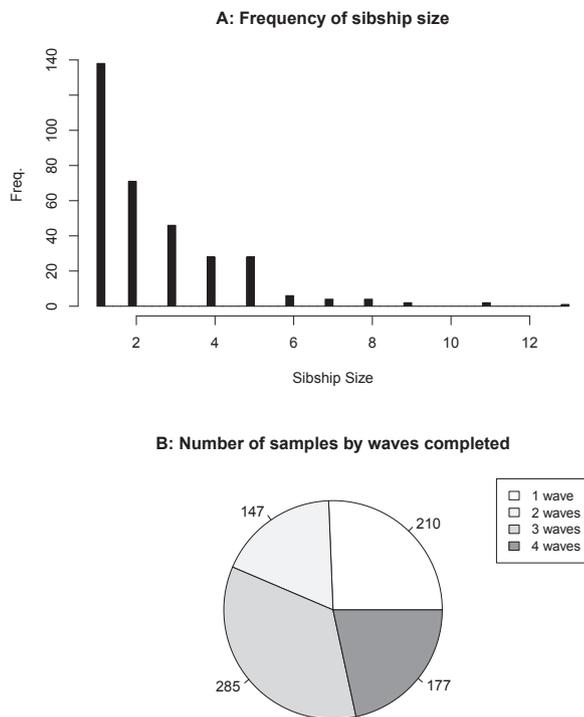


Fig. 1. The histogram for sibship size (A) and the pie plot for the number of siblings who underwent 1 wave, 2 waves, 3 waves, and 4 waves of examination (B). The largest portion of samples had 3 waves of follow-up.

useful information concerning the overall genetic and shared environmental background in the variation of clinical variables, for example, blood pressure.

Indeed, the analysis of longitudinal family data can be statistically challenging because of the complex data structure including repeated measurements over the follow-up waves and within family relatedness and nonlinear patterns of blood pressure over ages. This article introduced hierarchical modeling of longitudinal family data [8] in combination with the fractional polynomials (FPs) [9] for fitting nonlinear age trajectories and the mixed-effect models for estimating sibling correlation on blood pressure to assess the genetic and shared environmental effects on blood pressure. Sex-combined and sex-specific analyses have been done to allow comparison on sex-dependent patterns of sibling correlations.

Materials and methods

The samples

Large pedigree data on blood pressure were collected by the San Antonio Family Heart Study [10] and San Antonio Family Diabetes/

Gallbladder Study [11], which are pedigree-based studies conducted in San Antonio, Texas, and data on 20 Mexican American pedigrees were made available to the 18th Genetic Analysis Workshop [7]. The available data contain 27 to 107 individuals per pedigree with median pedigree size of 69 individuals and 1043 participants in total, from which blood pressure on 932 individuals was measured longitudinally over four times in a period of 30 years. For each examination, blood pressure was measured in sitting position, three times after a five-minute rest with a random zero sphygmomanometer [12]. The numbers provided here are the averages of the second and third readings.

Besides blood pressure, information concerning hypertension diagnosis (systolic blood pressure [SBP] > 140, diastolic blood pressure [DBP] > 90), antihypertension medicine intake, and tobacco smoking are also collected at each examination. From the 20 large pedigrees, a total of 819 siblings were found with age ranging from 16.8 to 76.3 years. Figure 1A displays the sibship size counted from each family. The frequency distribution and basic statistics on demography, blood pressure, hypertension medication, and smoking at each of the four waves are shown in Table 1. Figure 1B divides individuals by the number of consecutive waves undertaken during the survey with the biggest proportion (35%) of participants underwent three waves in a row. The effects of confounding were adjusted by regressing blood pressure on medication and smoking and keeping the residuals for subsequent analysis, resulting in 575 siblings in total (213 males and 362 females) (Table 2) because of missing observations on medication and smoking.

Statistical analysis

Hierarchical linear models

Hierarchical linear models (HLMs) represent a complex form of regression analysis [13] referred to as random coefficient models with two defining features. First, the data appropriate for HLMs are structured with different levels with lower level or level-1 units (here blood pressure measurements over time for each individual) nested within the higher level or level-2 units (here family). Second, the parameters of the level-1 model characterize linear relationship occurring between level-1 units (here the blood pressure trajectory over time). These parameters can be modeled as a function of level-2 units (family, age at entry, and sex of siblings within family). Our two-level HLM takes the form of regression models developed separately for level-1 and level-2 units. For clarity, the level-1 model for each individual can be shown as

$$Y_{ij} = \beta_{0,i} + \beta_{1,i}X_{ij} + e_{ij} \quad (1)$$

Here, Y_{ij} is the blood pressure measured for individual i at measurement j ; X_{ij} is the age at measurement j for individual i ; $\beta_{0,i}$ and $\beta_{1,i}$ are the intercept and slope parameters for individual i ; e_{ij} is a random error associated with individual i at measurement j , which is normally distributed with $E(e_{ij}) = 0$, $var(e_{ij}) = \sigma^2$.

Table 1
Basic statistics (95% CI) of sibling samples

	Examination 1	Examination 2	Examination 3	Examination 4
N*	762	530	573	202
Year	1981–1996	1997–2000	1998–2006	2009–2011
Age	37.5 (16.8–75)	40.7 (20–75.4)	44.6 (23.6–76.3)	49.2 (31.4–75.6)
Male (%)	366 (48)	251 (47.4)	276 (48.2)	121 (59.9)
SBP	120.5 (94–167.3)	123.5 (95.2–171.8)	124.5 (99–164.7)	127.1 (96.5–157.5)
DBP	71.3 (51.7–92)	71.9 (54–92)	71 (50.7–90)	77.9 (61–98)
Meds, %	8	16.5	25.4	40.3
Smoking, %	22.6	19.3	20.4	11.9

* Number with blood pressure measurements.

Table 2
Estimated ICCs and 95% CIs for sibling correlation on blood pressure

	Level		Change	
	SBP	DBP	SBP	DBP
All, N = 575				
Variance				
Sibship	33.24	21.36	0.06	0.38
Residual	149.26	63.75	3.67	1.17
ICC	0.20	0.28	0.02	0.13
95% CI	0.10–0.30	0.18–0.38	0.00–0.10	0.04–0.23
ICC*	0.18	0.25	0.01	0.10
F(age), P	2.00e–16	8.24e–07	0.27	1.22e–07
Males, n = 213				
Variance				
Sibship	15.13	15.16	0.48	0.00
Residual	213.77	73.67	4.00	1.72
ICC	0.09	0.19	0.11	0.00
95% CI	0.00–0.31	0.00–0.40	0.00–0.32	0.00–0.19
ICC*	0.07	0.17	0.11	0.00
F(age), P	8.98e–05	0.05	0.62	0.01
Females, n = 362				
Variance				
Sibship	38.85	30.47	0.23	0.26
Residual	110.51	51.98	3.08	1.13
ICC	0.30	0.40	0.07	0.22
95% CI	0.15–0.43	0.26–0.52	0.00–0.21	0.07–0.36
ICC*	0.26	0.37	0.07	0.19
F(age), P	2.00e–16	3.13e–07	0.29	2.53e–07

* ICC estimated by the linear model.

In the level-2 model, the regression coefficients from level-1 model are regressed on the level-2 group variables in a mixed-effect model.

$$E(\beta_{0,i}|r_{0,k}) = \gamma_{00} + \gamma_{01}age_i + \gamma_{02}sex_i + r_{0,k} \tag{2}$$

$$E(\beta_{1,i}|r_{1,k}) = \gamma_{10} + \gamma_{11}age_i + \gamma_{12}sex_i + r_{1,k} \tag{3}$$

In (2) and (3), γ_{00} and γ_{10} are the intercepts or overall means for $\beta_{0,i}$ and $\beta_{1,i}$ after adjusting for the fixed effects of age and sex; $r_{0,k}$ and $r_{1,k}$ are the random effects for $\beta_{0,i}$ and $\beta_{1,i}$ in family k after adjusting for age and sex and are normally distributed with $E(r_{0,k}) = E(r_{1,k}) = 0$, $SD(r_{0,k}) = v_0$, and $SD(r_{1,k}) = v_1$. Here, the most interesting parameters are $r_{0,k}$ and $r_{1,k}$, which can be used for estimating sibship correlation calculated as the intraclass correlation coefficient (ICC).

Following Andersen and Skovgaard [14], upon fitting the mixed-effect models, if the residual SDs for $\beta_{0,i}$ and $\beta_{1,i}$ for given $r_{0,k}$ and $r_{1,k}$ are denoted as s_0 and s_1 , then the marginal SDs for $\beta_{0,i}$ and $\beta_{1,i}$ are

$$SD(\beta_{0,i}) = \sqrt{v_0^2 + s_0^2}, \quad SD(\beta_{1,i}) = \sqrt{v_1^2 + s_1^2}.$$

Then, the correlation between observations $\beta_{0,1}$ and $\beta_{0,2}$ in family k is

$$ICC_0 = \frac{v_0^2}{v_0^2 + s_0^2}. \tag{4}$$

Likewise, the correlation between observations $\beta_{1,1}$ and $\beta_{1,2}$ in family k is

$$ICC_1 = \frac{v_1^2}{v_1^2 + s_1^2}. \tag{5}$$

We use ICC_0 and ICC_1 to measure sibling correlation on the level and rate of change of blood pressure, respectively.

The mixed model with FPs

As to be shown in the results section, some of the age patterns for the level and the rate of change of blood pressure are monotonously nonlinear. To model the nonlinear relationship, we introduce the first-order FPs as proposed by Royston and Altman [15] to the mixed-effect models, such that (2) and (3) become

$$E(\beta_{0,i}|r_{0,k}) = \gamma_{00} + \gamma_{01}age_i^p + \gamma_{02}sex_i + r_{0,k} \tag{6}$$

$$E(\beta_{1,i}|r_{1,k}) = \gamma_{10} + \gamma_{11}age_i^p + \gamma_{12}sex_i + r_{1,k} \tag{7}$$

where the power p is restricted to a special set $\phi = \{-2, -1, -1/2, 0, 1/2, 1, 2, 3\}$ with age^0 denotes the natural log transformation, $log(age)$. For each p in ϕ , a model was fitted and the best fitting model determined as the one with the lowest Akaike Information Criterion [16].

All statistical analyses were performed using the R Statistical software (<https://www.r-project.org/>). The mixed-effects model was fitted using the R package lme4.

Results

Following the steps described in the Method section, we first estimated coefficients (the intercept and slop) from the level-1 model, that is, Equation (1), for each individual for SBP and DBP. In fitting the model, the age at entry was subtracted from X_{ij} so that the intercept is actually the level of blood pressure at intake and the slope is the rate of change over the follow-up period. Figure 2 plots the intercepts (left panel) and slopes (right panel) of SBP against individual age at entry. The red lines are the smoothing average over ages (solid for males and dashed for females). The level of SBP increases nearly linearly with age in males. It starts higher in females at younger ages but converges to males after around the age of 40 years. Interestingly, the rate of change keeps constant around zero over the ages. The age pattern of DBP is plotted in Figure 3. Different from SBP, the level of DBP increases with age but starts to decline at around the age of 45 years both in males and females. Again, different from SBP, the rate of change for DBP decreases linearly with increasing age in both sexes.

The age trajectories of blood pressure levels in Figures 2 and 3 suggest that it is necessary to introduce nonlinear models to improve goodness of fit. As described in the Method section, we fitted mixed-effect models with first-order FPs to the sibling data. In Table 2, we show the estimated ICCs for the level and rate of change for SBP and DBP. In the sex-combined sample of 819 siblings, ICC estimates for the levels of SBP and DBP are 0.2 (95% CI: 0.10–0.30) and 0.28 (95% CI: 0.18–0.38), respectively. For comparison, we also fitted linear models to SBP and DBP levels, which estimated lower ICCs of 0.18 and 0.25. For the rate of change, the estimated ICC for SBP was not significant (ICC = 0.02, 95% CI: 0.00–0.10), whereas a significant but low ICC (0.13) was estimated for DBP (95% CI: 0.04–0.23). The P value for the FP function of age, F(age), shows that there is no age-dependent pattern in the rate of change in SBP ($P = .27$).

Considering the differential sex-dependent trajectories for the levels of blood pressure in the left panels of Figures 2 and 3, we next performed analysis on males (338 siblings) and females (481 siblings) separately (Table 2). The estimated ICCs for males are all very low with their 95% CI including 0. In females, however, all ICC estimates are significantly different from 0 except the rate of change for SBP (ICC = 0.07, 95% CI: 0.00–0.21), all higher than that for males and for sex-combined analysis. The estimated ICCs for females are 0.30 (95% CI: 0.15–0.43) and 0.40 (95% CI: 0.26–0.52) for the levels of SBP and DBP, respectively. Again, Table 2 shows that

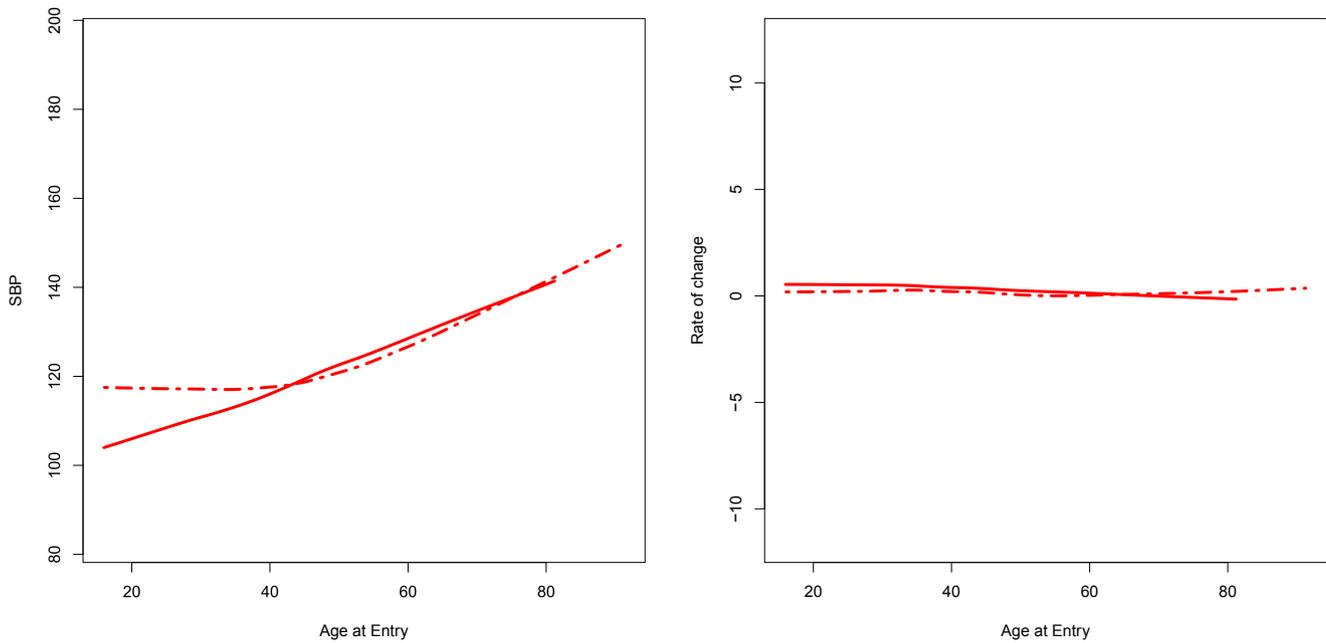


Fig. 2. Scatter plot for the level-1 estimates of levels (left panel) and rate of change (right panel) in SBP plotted against individual age at entry with the curves showing the smoothing averages for males (solid) and females (dashed). The solid circles are male samples and the empty circles are female samples.

the linear model underestimates ICCs as compared with the nonlinear model. For the rate of change, only DBP has a significant ICC estimate (ICC = 0.22, 95% CI: 0.07–0.36). The rate of change for SBP has a minor ICC estimate (0.07) which is not different from 0 (95% CI: 0.00–0.21).

Discussion

By fitting hierarchical linear models to longitudinal data and introducing the random effect to account for sibship, we were able to estimate ICCs for sibling correlation on both the level of blood

pressure and its rate of change over ages. Significant ICCs were estimated for the levels of SBP and DBP in the sex-combined sample (Table 2). Because two times sibling correlation gives the upper bound of additive genetic heritability, we have the upper limit of heritability estimate for the level of SBP as 0.40 (95% CI: 0.20–0.60). Likewise, the heritability estimate for DBP is 0.56 (95% CI: 0.36–0.76). These estimates are, in general, lower than the estimates using twins because the latter controls for multiple confounding factors. However, the estimate for DBP is in good agreement with the heritability estimates in Chinese twins (0.57, 95% CI: 0.47–0.65) [3]. Although the estimated heritability for SBP

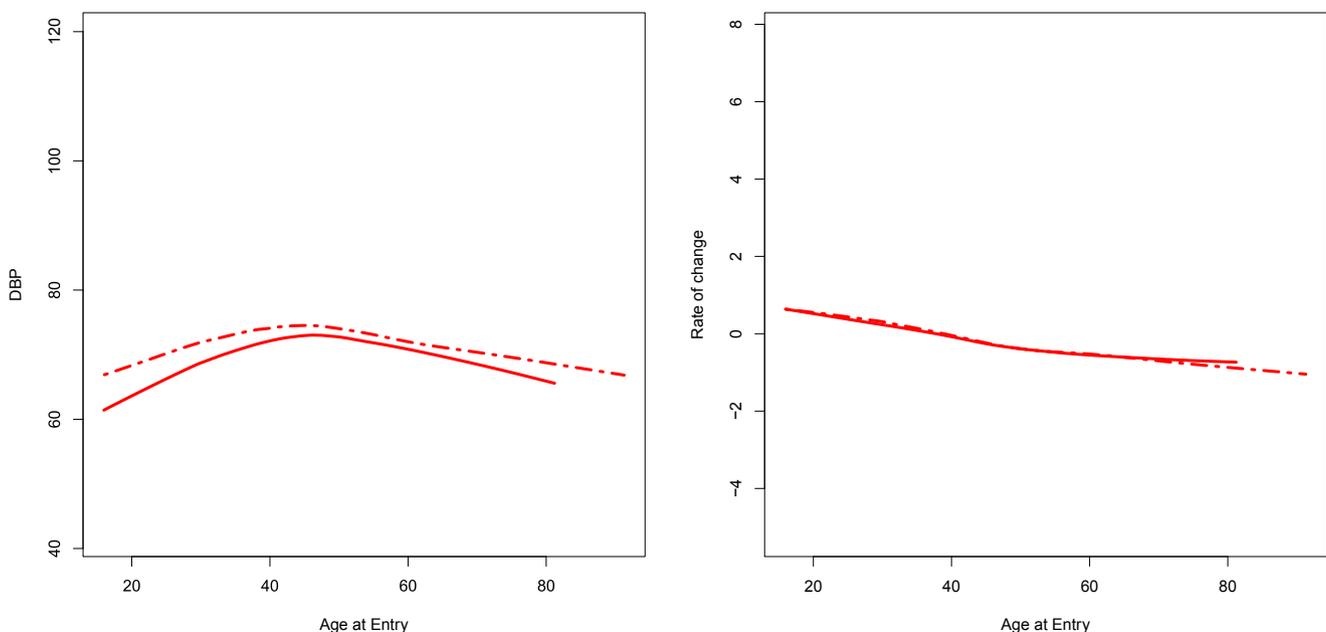


Fig. 3. Scatter plot for the level-1 estimates of levels (left panel) and rate of change (right panel) in DBP plotted against individual age at entry with the curves showing the smoothing averages for males (solid) and females (dashed). The solid circles are male samples and the empty circles are female samples.

is lower than the estimate in Chinese twins (0.54, 95% CI: 0.44–0.63), their confidence intervals overlap a lot. Different from the results on levels of blood pressure, the estimated ICC on the rate of change in SBP is not different from 0 (0.02, 95% CI: 0.00–0.10) and that for DBP is very low but significant (0.13, 95% CI: 0.04–0.23). Comparing our estimated ICCs for the level and for the rate of change in blood pressure, it is clear that genetic and shared family factors have only minor influence in determining the dynamic change in blood pressure. That is, the rate of change in blood pressure is mainly resulted from the individual's unique environment. This conclusion is consistent with that from a recent longitudinal study on Danish and Chinese twins [17], although our estimates are in general lower. The practical implications of these findings are two folds. First, genetic association studies should focus on genetic variants accounting for blood pressure variations but not the longitudinal changes; second, promoting individual healthy lifestyle can be the key for controlling the age-related progression of blood pressure.

In Table 2, the estimated ICCs for the levels of SBP and DBP in females correspond to heritability estimates of 0.60 (95% CI: 0.3–0.86) and 0.80 (95% CI: 0.52–1.00), respectively. These estimates are much higher than the sex-combined estimates, suggesting that the latter are in fact dominated by the correlation in female siblings. Although the ICC for the rate of change of SBP is not significant, the estimated upper bound of heritability for DBP is 0.44 (95% CI: 0.14–0.72), a clear indication that genetic and common family environment play a role in regulating the longitudinal change of DBP but not SBP in females. The observed sex difference in ICC estimates can be due to either biology or the different sample sizes used. Considering the sample size for males (213 individuals) is less than 60% of the sample size for females (362 individuals), the insignificant results in males can well be affected by the low power in analyzing the male siblings.

As described in the Method section, we fitted models with and without FPs, the latter assumes linear relationship between blood pressure and age. In Table 2, the ICC estimates from the linear models are presented as ICC'. It is clear that $ICC' \leq ICC$ in all analyses suggesting that the nonlinear models provide higher ICC estimates. This is understandable as the FPs helped to better account for the effect of age, which, otherwise, will serve to increase residual variance. In the literature of twin modeling, the nonlinear age effect is usually modeled by adding extra covariates defined as age, age², age³, etc. This arbitrary assignment of variables takes away power for statistical testing and does not capture the nonlinear age pattern efficiently. In this case, fitting the FPs should be encouraged.

Finally, although our hierarchical model with FPs estimated significant genetic contributions to both the level and the rate of change of blood pressure traits, the estimates are overall genetic effects together with influences from shared family environmental factors. Family and individual environmental factors (e.g., social economic status, diet habit, drinking behavior), if available, can be easily included in the level-2 model and potentially help to improve

the estimates on genetic contributions. Likewise, the inclusion of molecular markers (e.g., SNP genotypes, CpG methylation levels, gene expression) can help with identifying molecular targets of blood pressure variations together with their interaction effects with observed environmental exposures.

Acknowledgment

This study was jointly supported by the Independent Research Fund Denmark (grant number DFF-6110-00114) and the Novo Nordisk Foundation Medical and Natural Sciences Research grant [grant number NNF13OC0007493].

References

- [1] Kesarwani M, Perez A, Lopez VA, Wong ND, Franklin SS. Cardiovascular comorbidities and blood pressure control in stroke survivors. *J Hypertens* 2009;27(5):1056–63.
- [2] Djoussé L, Gaziano JM. Metabolic syndrome and hypertension. *Hypertension* 2008;52:799–800.
- [3] Li S, Pang Z, Zhang D, Duan H, von Bornemann Hjelmberg J, Tan Q, et al. Probing genetic overlap in the regulation of systolic and diastolic blood pressure in Danish and Chinese twins. *Hypertens Res* 2014;37(10):954–9.
- [4] Feng Y, Zhang T, Xu X, Xu X. Familial aggregation of metabolic syndrome and its components in a large Chinese population. *Obesity (Silver Spring)* 2008;16(1):125–9.
- [5] Palmer ND, Freedman BI. Genetics: genetic factors in the regulation of blood pressure. *Nat Rev Nephrol* 2016;12(12):716–7.
- [6] Wirtwein M, Melander O, Sjögren M, Hoffmann M, Narkiewicz K, Gruchala M, et al. Genetic risk factors influence nighttime blood pressure and related cardiovascular complications in patients with coronary heart disease. *Hypertens Res* 2018;41:53–9.
- [7] Bickeböller H, Bailey JN, Beyene J, Cantor RM, Cordell HJ, Culverhouse RC, et al. Genetic analysis workshop 18: methods and strategies for analyzing human sequence and phenotype data in members of extended pedigrees. *BMC Proc* 2014;8(Suppl 1):S1.
- [8] Tan Q, Hjelmberg JB, Thomassen M, Jensen AK, Christiansen L, Christensen K, et al. Hierarchical linear modeling of longitudinal pedigree data for genetic association analysis. *BMC Proc* 2014;8(Suppl 1):S82.
- [9] Tan Q, Thomassen M, Hjelmberg JB, Clemmensen A, Andersen KE, Petersen TK, et al. A growth curve model with fractional polynomials for analysing incomplete time-course data in microarray gene expression studies. *Adv Bioinformatics* 2011;2011:261514.
- [10] Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, et al. Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio family heart study. *Circulation* 1996;94(9):2159–70.
- [11] Hunt KJ, Lehman DM, Arya R, Fowler S, Leach RJ, Göring HH, et al. Genome-wide linkage analyses of type 2 diabetes in Mexican Americans: the San Antonio family diabetes/gallbladder study. *Diabetes* 2005;54(9):2655–62.
- [12] Kronmal RA, Rutan GH, Manolio TA, Borhani NO. Properties of the random zero sphygmomanometer. *Hypertension* 1993;21(5):632–7.
- [13] Woltman H, Feldstain A, MacKay JC, Rocchi M. An introduction to hierarchical linear modelling. *Tutor Quant Methods Psychol* 2012;8:52–69.
- [14] Andersen PK, Skovgaard LT. Regression with linear predictors. New York: Springer; 2010. p. 433.
- [15] Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling (with discussion). *Appl Stat* 1994;43:429–67.
- [16] Akaike H. A new look at the statistical model identification. *IEEE Trans Automatic Control* 1974;19:716–23.
- [17] Li S, Kyvik KO, Pang Z, Zhang D, Duan H, Tan Q, et al. Genetic and environmental regulation on longitudinal change of metabolic phenotypes in Danish and Chinese adult twins. *PLoS One* 2016;11(2):e0148396.