

## Analysis of gene $\times$ environment interactions in sibships using mixed models

Jill S Barnholtz-Sloan<sup>\*1</sup>, Laila M Poisson<sup>2</sup>, Steven W Coon<sup>2</sup>, Gary A Chase<sup>2</sup> and Benjamin A Rybicki<sup>2</sup>

Address: <sup>1</sup>Department of Internal Medicine (Oncology), Wayne State University School of Medicine and Karmanos Cancer Institute, 110 East Warren, Detroit, Michigan, USA and <sup>2</sup>Department of Biostatistics and Research Epidemiology, Henry Ford Health Sciences Center, One Ford Place, Detroit, Michigan, USA

Email: Jill S Barnholtz-Sloan<sup>\*</sup> - jbsloan@med.wayne.edu; Laila M Poisson - lpoisso1@hfhs.org; Steven W Coon - scoon1@hfhs.org; Gary A Chase - gchase@hes.hmc.psu.edu; Benjamin A Rybicki - brybick1@hfhs.org

<sup>\*</sup> Corresponding author

from Genetic Analysis Workshop 13: Analysis of Longitudinal Family Data for Complex Diseases and Related Risk Factors  
New Orleans Marriott Hotel, New Orleans, LA, USA, November 11–14, 2002

Published: 31 December 2003

BMC Genetics 2003, 4(Suppl 1):S18

This article is available from: <http://www.biomedcentral.com/1471-2156/4/s1/S18>

### Abstract

**Background:** Gene  $\times$  environment models are widely used to assess genetic and environmental risks and their association with a phenotype of interest for many complex diseases. Mixed generalized linear models were used to assess gene  $\times$  environment interactions with respect to systolic blood pressure on sibships adjusting for repeated measures and hierarchical nesting structures. A data set containing 410 sibships from the Framingham Heart Study offspring cohort (part of the Genetic Analysis Workshop 13 data) was used for all analyses. Three mixed gene  $\times$  environment models, all adjusting for repeated measurement and varying levels of nesting, were compared for precision of estimates: 1) all sibships with adjustment for two levels of nesting (sibs within sibships and sibs within pedigrees), 2) all sibships with adjustment for one level of nesting (sibs within sibships), and 3) 100 data sets containing random draws of one sibship per extended pedigree adjusting for one level of nesting.

**Results:** The main effects were: gender, baseline age, body mass index (BMI), hypertensive treatment, cigarettes per day, grams of alcohol per day, and marker GATA48G07A. The interaction fixed effects were: baseline age by gender, baseline age by cigarettes per day, baseline age by hypertensive treatment, baseline age by BMI, hypertensive treatment by BMI, and baseline age by marker GATA48G07A. The estimates for all three nesting techniques were not widely discrepant, but precision of estimates and determination of significant effects did change with the change in adjustment for nesting.

**Conclusion:** Our results show the importance of the adjustment for all levels of hierarchical nesting of sibs in the presence of repeated measures.

### Background

Gene  $\times$  environment models are widely used to assess genetic and environmental risks and their association with a phenotype of interest for many different complex

diseases [1,2]. The Framingham Heart Study began in 1948 with the aim of gathering the longitudinal family data needed for a comprehensive study of genetic and environmental risks for cardiovascular disease. In 1971, a

second-generation group was enrolled called the Offspring Cohort and has been followed every four years since. The Framingham data has led to the discovery of major cardiovascular risk factors (e.g., high blood pressure, high blood cholesterol, smoking, obesity, diabetes, physical inactivity), important related factors (e.g., age, gender, psychosocial factors, blood triglyceride, and lipid levels), as well as genetic risk factors [3].

Studies have examined environmental and genetic variables influencing blood pressure in the Framingham data as well as in other large epidemiologic data sets [3,4]. These studies were performed in extended pedigree, sibship, and case-control data. Genetic studies have found multiple regions on the genome that may contain a candidate gene for systolic blood pressure (SBP) and/or hypertension. Among these regions are areas on chromosomes 10 and 17 [5,6]. Chromosome 17 contains the angiotensin-I converting enzyme (ACE) gene and there is good supporting evidence that this gene is involved in hypertension [7-9].

Sibship studies are routinely used when parents and other members of the extended pedigree are not available for study. Full sibs share approximately half of their genome and usually share a common environment for a period of their lives, making them good candidates for gene  $\times$  environment risk studies. Longitudinal studies, which involve taking measurements of the same factors over time on an individual, can be used to enhance the accurate assessment of gene  $\times$  environment models in sib studies. With repeated measures data, stability over time of measurements can be evaluated, whereas in a cross-sectional study these measurements can only be evaluated at one point in time. However, repeated measures add complexity to variance estimation.

In addition to the correlation of data points collected in a serial manner, there is also the hierarchical nesting structure of sibs within sibships and of sibships within pedigrees. In cases when there is more than one sibship available per extended pedigree, most studies use either all sibships available, without accounting for the correlation between them, or one sibship only is randomly drawn from the extended pedigree and used for analysis. Mixed generalized linear models are frequently used to account not only for the situation in which there are repeated measures on an individual, but also when there is a complex hierarchical nesting structure present within the data. Hence, we used mixed generalized linear models to assess gene  $\times$  environment interactions with respect to SBP on sibships from the Framingham Heart Study offspring cohort data available for the Genetic Analysis Workshop 13 (GAW13). We compared precision of estimates of significant effects obtained from using all availa-

ble sibships, while controlling for the nesting of sibs within sibships and of sibships within extended pedigrees and repeated measures, to results from analyses in which the sibship nesting within extended pedigrees is either ignored or obviated by selection of one sibship per pedigree.

## Methods

### Study Population

Sibship data from the Framingham Heart Study offspring cohort (part of the GAW13 data) was used for all analyses, in which each sib could have participated in any or all of five exam periods. The data consisted of 410 sibships total (from 330 extended pedigrees) and for the data where one sibship was randomly drawn from each extended pedigree, 330 sibships were included. The phenotype of interest was SBP, in its continuous form. The covariates of interest were: baseline age, baseline gender and height, weight, hypertensive treatment, number of cigarettes per day, and number of grams of alcohol per day all measured over the five exam periods. Body mass index (BMI) was used rather than height and weight separately.

### Statistical Analysis

Mixed generalized linear models were implemented using PROC MIXED in SAS Version 8 software [10-12], to account for the repeated covariate measurements over the five exam periods and for the hierarchical nesting structure of sibs within sibships and of sibships within extended pedigrees. Markers of interest were chosen using FBAT [13,14], in which tests for association with high SBP for each multiallelic marker in the genome scan were performed using a dichotomous form of SBP ( $\geq 140$  and  $< 140$ ). Of the markers highly associated with high SBP (dominant and additive models,  $p = 0.05$ ), three markers were used: GATA48G07A and GGAA5D10 on chromosome 10 and GGAA7D11 on chromosome 17, because previous studies showed association with locations on chromosomes 10 and 17 [5-9]. The SBP-associated alleles at the three markers were then pooled to result in 3, 4, and 3 alleles for markers GATA48G07A, GGAA5D10, and GGAA7D11, respectively, with all alleles not associated with SBP pooled into the last allele (see Table 1).

Univariable and multivariable models were fitted and decisions for inclusion/exclusion were based on statistical significance from the models. Explanatory factors shown to be significant in previous studies of SBP were also kept in the model [3,4], where the phenotype of interest was SBP in its continuous form. The results were used to develop a gene  $\times$  environment model, using all sibship data available (410 sibships total from 330 extended pedigrees, where the average sibship has three individuals (median = 3)), adjusting for both nesting of sibs within sibships and nesting of sibships within extended

**Table 1: Reassignment of allele names for the three markers of interest: GATA48G07A, GGAA5D10 and GGAA7D11**

Marker	Allele No.	Alleles Pooled
GATA48G07A	1	350 bp
	2	362 bp
	3	All others
GGAA5D10	1	93 bp
	2	109 bp
	3	105, 121 bp
	4	All others
GGAA7D11	1	274 bp
	2	286 bp
	3	All others

pedigrees and adjusting for repeated measures. This gene  $\times$  environment model was then fitted again to the same full data set, while adjusting for only one level of nesting (sibs within sibships) and for repeated measures. Then the gene  $\times$  environment model was fitted to 100 data sets containing random draws of one sibship per extended pedigree (every pedigree is represented by one sibship; 60 out of 330 pedigrees has multiple sibships), adjusting for repeated measures and one level of nesting (sibs within sibships), in which estimate values, standard errors and  $p$ -values were averaged over all 100 data sets. The precision of parameter estimates for main and interaction effects were compared across all three nesting and repeated measures adjustment techniques by empirical evaluation of the resulting standard errors and  $p$ -values. For a specific effect, the estimate having the lowest standard error and  $p$ -value of the three models was determined to be the most precise estimate.

## Results

The main effects included in the model were: gender, baseline age, BMI, hypertensive treatment, cigarettes per day, grams of alcohol per day, and marker GATA48G07A (heterozygote genotype with alleles 1 (allele 350 bp) and 3 (all other alleles) (57% of total individuals) contrasted with all other possible genotypes). The interaction fixed effects were: baseline age by gender, baseline age by cigarettes per day, baseline age by hypertensive treatment, baseline age by BMI, hypertensive treatment by BMI, and baseline age by marker GATA48G07A (significant interactions from fitting all second-order interactions on the full data set). The estimates and standard errors for all main and interaction fixed effects included in the gene  $\times$  environment model performed using all three nesting adjustment techniques with repeated measures adjustment are given in Table 2.

The model including all available sibships adjusting for both levels of nesting and repeated measures was the most precise. The standard errors of the estimates for the main and interaction fixed effects were smaller (or equal to) standard errors of the model fitted on the same data adjusting for one level of nesting only and the model fitted on 100 random draws of one sibship per pedigree, adjusting for one level of nesting. Across the 100 data sets with random draws of one sibship per extended pedigree, the effects and standard error estimates were highly stable (data not shown).

For most of the main and interaction fixed effects, as compared to the full data set adjusted for both levels of nesting and repeated measures, the  $p$ -values either increase (but are still statistically significant), or they increase such that the effect would no longer be considered statistically significant, when the model is fitted on the full data set adjusting for only one level of nesting or fitted on the data when one sibship per extended pedigree is drawn at random (see Table 2 with interest in cigarettes per day, BMI, marker GATA48G07A, baseline age and cigarettes per day, baseline age and BMI, hypertensive treatment and BMI, and baseline age and marker GATA48G07A).

## Discussion

Mixed generalized linear models allow for adjustment for repeated measures as well as hierarchical nesting in data. When performing an analysis on sib data from extended pedigrees within a longitudinal study, a modeling scheme needs to consider not only the correlation in repeated measurements on each individual, but also the correlation resulting from the nesting of sibs within sibships and the nesting sibships within the larger family structure. The precision and significance of model estimates vary depending on the adjustments that are used.

Using all data available, i.e., all possible sibships, we fitted a gene  $\times$  environment model, adjusting for nesting of sibs within sibships and nesting of sibships within extended pedigrees and adjusting for repeated measures from each of five time periods. The model included all environmental main effects available, because they were either statistically or biologically essential to the prediction of SBP (see Table 2). The only univariable genetic effect that was significant was marker GATA48G07A (heterozygote genotype with alleles 1 and 3 as compared with all other possible genotypes) on chromosome 10, which is in a region that has previously been shown to be associated with SBP [5,6]. Multiple interaction effects were significant using the full data set with all adjustments (see Table 2). This model was then fitted to the full data set, adjusting for one level of nesting only and repeated measures and to 100 data sets consisting of random draws of one

**Table 2: Adjusted fixed effects estimates and standard errors from the gene × environment model**

Effect	FULL DATA-ADJ2 <sup>A</sup>			FULL DATA-ADJ1 <sup>B</sup>			ONE SIBSHIP DATA-ADJ1 (100 random draws) <sup>C</sup>		
	Estimate	SE	P-value (t-test)	Estimate	SE	P-value (t-test)	Avg Estimate	Avg SE	Avg P-value (t-test)
Intercept	95.72	6.91	<0.0001	95.99	6.90	<0.0001	87.67	7.81	<0.0001
Gender	-10.71	1.89	<0.0001	-9.98	2.07	<0.0001	-4.71	2.32	0.0001
Baseline age	0.15	0.14	0.29	0.19	0.14	0.18	0.16	0.16	0.36
BMI	0.49	0.16	0.003	0.51	0.16	0.002	0.63	0.18	0.001
Hypertensive treatment	-2.44	4.55	0.59	-3.32	4.60	0.47	-0.20	5.34	0.68
Cigarettes per day	0.11	0.05	0.04	0.13	0.05	0.02	0.10	0.06	0.11
Grams of alcohol per day	0.07	0.01	<0.0001	0.06	0.01	<0.0001	0.06	0.01	<0.0001
Marker GATA48G07A	5.49	2.24	0.01	4.04	2.07	0.05	2.84	2.33	0.02
Baseline age and gender	-0.19	0.06	0.0005	-0.17	0.06	0.006	-0.16	0.07	0.03
Baseline age and cigarettes per day	-0.005	0.00	0.003	-0.005	0.00	0.001	-0.004	0.002	0.01
Baseline age and hypertensive treatment	0.14	0.07	0.04	0.13	0.07	0.06	0.19	0.08	0.03
Baseline age and BMI	0.02	0.00	0.0005	0.02	0.00	0.001	0.02	0.006	0.008
Hypertensive treatment and BMI	-0.29	0.12	0.01	-0.30	0.12	0.01	-0.26	0.13	0.07
Baseline age and marker GATA48G07A	0.14	0.07	0.04	0.11	0.06	0.08	0.16	0.07	0.03

<sup>A</sup>FULL DATA-ADJ2: full data set using all available sibships adjusting for both levels of nesting (sibs within sibships and sibships within pedigrees); <sup>B</sup>FULL DATA-ADJ1: full data set using all available sibships adjusting for one level of nesting (sibs within sibships); <sup>C</sup>ONE SIBSHIP DATA-ADJ1 (100 random draws): data set using one randomly drawn sibship per each extended pedigree, adjusting for one level of nesting (sibs within sibships) performed 100 times.

sibship per extended pedigree, adjusting for one level of nesting only and repeated measures.

The estimates for all three nesting techniques were not widely discrepant, but precision of estimates of significant effects did change with the change in adjustment for nesting, i.e., when all levels of nesting were adjusted for as well as repeated measures, the precision was at the highest level.

**Conclusions**

Our results show the importance of the adjustments for all levels of hierarchical nesting of sibs in the presence of repeated measures in these analyses. We have shown that precision of estimates of statistically significant effects is negatively affected when all levels of a hierarchical nesting structure are not taken into account when using sibships from extended pedigree data.

**Acknowledgments**

This work was supported in part by NCI grant K07 CA91849-01 for JSB-S, by NEHS grant R01 ES11126 for BAR.

**References**

1. Liang KY, Beaty TH: **Statistical designs for familial aggregation.** *Stat Methods Med Res* 2000, **9**:543-562.

2. Clayton D, McKeigue PM: **Epidemiological methods for studying genes and environmental factors in complex diseases.** *Lancet* 2001, **358**:1356-1360.

3. **Framingham Heart Study** [<http://www.nhlbi.nih.gov/about/framingham/index.html>]

4. Timberlake DS, O'Conner DT, Parmer RJ: **Molecular genetics of essential hypertension: recent results and emerging strategies.** *Curr Opin Nephrol Hypertens* 2001, **10**:71-79.

5. Rice T, Rankinen T, Province MA, Chagnon YC, Perusse L, Borecki IB, Bouchard C, Rao DC: **Genome-wide linkage analysis of systolic and diastolic blood pressure.** *Circulation* 2000, **102**:1956-1963.

6. Williams SM, Addy JH, Phillips JA III, Dai M, Kpodonu J, Afful J, Jackson H, Joseph K, Eason F, Murray MM, Epperson P, Aduonum A, Wong LJ, Jose PA, Felder RA: **Combinations of variations in multiple genes are associated with hypertension.** *Hypertension* 2000, **36**:2-6.

7. O'Donnell CJ, Lindpaintner K, Larson MG, Rao VS, Ordovas JM, Schaefer EJ, Myers RH, Levy D: **Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not women in the Framingham Heart Study.** *Circulation* 1998, **97**:1766-1772.

8. Levy D, DeStefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavras H, Cupples LA, Myers RH: **Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham Heart Study.** *Hypertension* 2000, **36**:477-483.

9. Zhu X, Bouzekri N, Southam L, Cooper RS, Adeyemo A, McKenzie CA, Luke A, Chen G, Elston RC, Ward R: **Linkage and association analysis of angiotensin I-converting enzyme (ACE)-gene polymorphisms with ACE concentration and blood pressure.** *Am J Hum Genet* 2001, **68**:1139-1148.

10. Diggle PJ, Liang KY, Zeger SL: **Analysis of Longitudinal Data.** Oxford, Oxford University Press 1996.

11. Goldstein H: **Multilevel Statistical Models.** New York, John Wiley & Sons Inc 21995.

12. SAS Institute Inc: **Statistical Analysis Software, Version 8.1.** Cary, North Carolina, SAS Institute, Inc 2000.
13. Horvath S, Wei E, Xu X, Palmer LJ, Baur M: **Family-based association test method: age of onset traits and covariates.** *Genet Epidemiol* 2001, **21(suppl 1)**:S403-S408.
14. Horvath S, Xu X, Laird NM: **The family based association test method: strategies for studying general genotype-phenotype associations.** *Eur J Hum Genet* 2001, **9**:301-306.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

