

Multilevel modeling for the analysis of longitudinal blood pressure data in the Framingham Heart Study pedigrees

Laurent Briollais*^{1,2}, Anjela Tzontcheva^{1,2} and Shelley Bull^{1,2}

Address: ¹Division of Epidemiology and Biostatistics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada, M5G 1X5 and ²Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada, M5G 1X5

Email: Laurent Briollais* - laurent@mshri.on.ca; Anjela Tzontcheva - tsoncheva@utstat.toronto.edu; Shelley Bull - bull@mshri.on.ca

* 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):S19

This article is available from: <http://www.biomedcentral.com/1471-2156/4/s1/S19>

Abstract

Background: The data arising from a longitudinal familial study have a complex correlation structure that cannot be modeled using classical methods for the analysis of familial data at a single time point.

Methods: To fit the longitudinal systolic blood pressure (SBP) pedigree data arising from the Framingham Heart Study, we proposed to use multilevel modeling. That approach was used to distinguish multiple levels of information with individual repeated measurements (Level 1) being made within individuals (Level 2), and individuals clustered within pedigrees (Level 3). Residuals from the subject-specific and pedigree-specific regression models were summed both for the mean SBP and slope of SBP change over time, in order to define two new outcomes that were then used in a genome-wide linkage analysis.

Results: Evidence for linkage for the two outcomes (mean SBP and slope) was found in several chromosomal regions with a maximum LOD score of 3.6 on chromosome 8 and 3.5 on chromosome 17 for the mean SBP, and 2.5 on chromosome 1 for SBP slope. However, the linkage on chromosome 8 was only detected when the sample was restricted to subjects between age 25 and 75 and with at least four exams (Cohort 1) or 3 exams (Cohort 2).

Discussion: Multilevel modeling is a powerful approach to detect genes involved in complex traits when longitudinal data are available. It allows for complex hierarchical data structure to be taken into account and therefore, a better partitioning of random within-individual variation from other sources of variability (genetic or nongenetic).

Background

The Framingham Heart Study provides long-term repeated measurements of blood pressure and other phenotypes in two large cohorts of related individuals. Longitudinal studies are efficient designs for the investigation of individual changes over time. In the context of familial studies, such designs might be of particular interest to assess the proportion of the trait variability explained by

within-individual variation or other sources of variation. However, the data arising from a longitudinal familial study have a complex correlation structure that cannot be modeled using classical methods for the analysis of familial data at a single time point. In this study, we proposed to use multilevel modeling to fit the complex data structure arising from the Framingham Heart Study. Multilevel modeling, also known as hierarchical regression,

generalizes ordinary regression modeling to distinguish multiple levels of information in a model [1]. It might be appropriate to model the Framingham Heart Study data that form a natural hierarchy with individual repeated measurements (Level 1) being made within individuals (Level 2), and individuals clustered within pedigrees (Level 3). The use of appropriate random effects at each level allows one to adjust for the influence of a wide variety of correlation structures and to estimate variance, covariance, and correlation which are of particular interest in familial studies. In this paper, multilevel models are first used to fit the repeated systolic blood pressure (SBP) measurements. Residuals from the subject-specific and pedigree-specific regression models were summed both for the mean SBP and slope of SBP change over time, in order to define two new outcomes that were then used in a genome-wide linkage analysis. Both phenotypes are of interest because genes involved in the variation of SBP with time could differ from genes affecting long-term mean SBP.

Methods

Data

The Framingham Heart Study data includes 330 pedigrees originally selected for a genome-scan analysis. The pedigrees consisted of 4692 subjects, of whom 2885 have participated in the Framingham Heart Study. Longitudinal SBP data were analyzed for 25,263 examinations on 2662 individuals. Height, weight, gender, age, and hypertensive treatment information were required but if height was missing, the most recent measurement was imputed. Because there might be important variation in individual SBP measurement among younger and older subjects, we also restricted the sample to individuals aged between 25 and 75 years, as in Levy et al. [2]. The following selection criteria were also defined: 1) There had to be at least 10 years between a subject's initial and final examinations within the age range; 2) at least four examinations within the age range were required for the original cohort and at least three for offspring cohort participants [2]. Data from 24,840 examinations on 2530 individuals were available in the selected sample. For the genome-wide scan analysis, 1702 genotyped individuals were included (394 from the Cohort 1 and 1308 from the Cohort 2).

Multilevel analysis of the longitudinal SBP model

Let the random variable Y_{ijk} denote the SBP measurement at the i th examination for the j th individual in pedigree k . We then assume that Y_{ijk} satisfies the following general multilevel model:

Within-subject model – Level 1

$$Y_{ijk} = b_{0jk} + b_{1jk} (Age_{ijk} - \overline{Age}_{jk}) + b_{2jk} (Age_{ijk} - \overline{Age}_{jk})^2 + \beta_1 (BMI_{ijk} - \overline{BMI}_{jk}) + \beta_2 Treat_{ijk} + \beta_3 Treat_{ijk} \times (\overline{Age}_{jk} - Age_{ijk}) + \varepsilon_{ijk},$$

where $i = 1, \dots, 21$ for Cohort 1 subjects and $i = \{11, 15, 17, 19, 21\}$ for Cohort 2 subjects. Age_{ijk} , BMI_{ijk} , $Treat_{ijk}$ are the age, body mass index and hypertension treatment (1 for subjects treated and 0 for subjects untreated) at the i th exam for the j th individual in pedigree k , \overline{Age}_{jk} and \overline{BMI}_{jk} are the mean values across all exams for the j th individual, and ε_{ijk} are the error components that account for the within-individual variability. The ε_{ijk} are assumed to be normally distributed with mean vector zero and variance-covariance matrix Σ defined by a first-order autoregressive structure. The intercept b_{0jk} represents the average SBP for an untreated subject of average age and BMI across all of the subject's examinations. The regression coefficient b_{1jk} is used to model the linear variation of SBP with age. We found that every individual profile could be well approximated by a quadratic function of time, measured by the age at examination. We also tested a cubic effect, but it was not significant when we allowed for the individual's linear time trend to differ in each treatment group (interaction between age and treatment). Random effects were added to reflect the natural heterogeneity in the population. In this model, both the intercept and the linear effect for age were allowed to vary across individuals and the individual-specific regression coefficients (random effects) were defined at the second level:

Subject random-intercept model – Level 2

$$b_{0jk} = b_{00k} + \beta_{01} (\overline{Age}_{jk} - \overline{Age}) + \beta_{02} (\overline{BMI}_{jk} - \overline{BMI}) + \beta_{03} Sex_{jk} + \beta_{04} Cohort_{jk} + u_{0jk},$$

Subject random-slope model – Level 2

$$b_{1jk} = b_{10k} + \beta_{11} (\overline{BMI}_{jk} - \overline{BMI}) + \beta_{12} Sex_{jk} + \beta_{13} Cohort_{jk} + u_{1jk}, j = 1, \dots, n_k,$$

where $\begin{pmatrix} u_{0jk} \\ u_{1jk} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$ and n_k is the number of individuals in pedigree k .

\overline{Age} and \overline{BMI} are the sample means for age and body mass index, Sex and $Cohort$ are two indicator variables, coded 1 for males, 0 for females and 0 for Cohort 1 subjects, 1 for Cohort 2 subjects. The random components u_{0jk} and u_{1jk} measure the variation of each individual's mean SBP and slope from their average in pedigree k . The intercept b_{00k} represents the average SBP in pedigree k for males in Cohort 1 with average age and BMI and the intercept b_{10k} represents the average slope in pedigree k for males in Cohort 1 with average BMI. To account for the correlation of individuals within a pedigree, these two intercepts were allowed to vary between pedigrees. The

Table 1: Estimates of multilevel model fixed effects and random effects variances (± SE) in the selected and unselected samples with or without adjustment for BMI

	Unselected Sample		Selected Sample	
	Adjusted (1a)	Unadjusted (1b)	Adjusted (2a)	Unadjusted (2b)
Fixed effects estimates				
Intercept	132.46 (0.56)***	132.00 (0.57)***	131.16 (0.58)***	131.80 (0.59)***
Age effect ^A	0.60 (0.02)***	0.63 (0.02)***	0.61 (0.02)***	0.63 (0.02)***
Age ² effect ^A	0.013 (0.001)***	0.010 (0.001)***	0.014 (0.001)***	0.001 (0.001)***
BMI effect ^A	1.45 (0.05)***	-	1.4 (0.05)***	-
Treat effect	-2.13 (0.42)***	-1.45 (0.43)***	-2.2 (0.43)***	-1.52 (0.44)***
Age * Treatment effect	-0.35 (0.04)***	-0.34 (0.04)***	-0.34 (0.04)***	-0.34 (0.04)***
<i>Subject random intercept model</i>				
Mean age effect ^A	0.53 (0.03)***	0.65 (0.03)***	0.62 (0.04)***	0.72 (0.04)***
Mean BMI effect ^A	1.16 (0.06)***	-	1.21 (0.07)***	-
Cohort effect	-6.56 (0.67)***	-5.21 (0.70)***	-5.74 (0.71)***	-4.68 (0.74)***
Gender effect	2.27 (0.51)***	3.59 (0.53)***	1.71 (0.53)***	3.09 (0.56)***
<i>Subject random slope model</i>				
Mean BMI effect ^A	-0.014 (0.003)***	-	-0.015 (0.003)***	-
Cohort effect	-0.46 (0.03)***	-0.29 (0.03)***	-0.42 (0.03)***	-0.27 (0.03)***
Gender effect	-0.03 (0.03)	-0.07 (0.03)*	-0.04 (0.03)	-0.08 (0.03)**
Covariance Parameter Estimates				
Measurement error variance	140.85 (1.64)***	148.11 (1.73)***	141.80 (1.68)***	149.01 (1.77)***
AR(1) correlation	0.25 (0.01)***	0.26 (0.01)***	0.25 (0.01)***	0.26 (0.01)***
Subject level var(u _{0jk})	146.25 (5.01)***	161.79 (5.49)***	147.63 (5.30)***	164.31 (5.84)***
var(u _{1jk})	0.17 (0.01)***	0.20 (0.01)***	0.17 (0.01)***	0.20 (0.01)***
cov(u _{0jk} , u _{1jk})	1.41 (0.19)***	0.85 (0.21)***	1.42 (0.20)***	0.86 (0.22)***
Pedigree level var(v _{00k})	27.59 (4.17)***	24.41 (4.07)***	27.04 (4.31)***	23.22 (4.13)***
var(v _{01k})	0.008 (0.005)*	0.005 (0.005)	0.009 (0.005)*	0.006 (0.005)
cov(v _{00k} , v _{01k})	0.52 (0.12)***	0.37 (0.11)**	0.49 (0.12)***	0.37 (0.12)**
-2 REML	202,695.75	205,876.11	194,947.18	198,006.24

^AVariables centered * p < 0.05 ** p < 0.01 *** p < 0.001

random effects at different levels of the model are assumed independent.

Pedigree random-intercept model – Level 3

$$b_{00k} = \beta_{000} + v_{00k}, k = 1, \dots, N$$

Pedigree random-slope model – Level 3

$$b_{10k} = \beta_{010} + v_{01k}, k = 1, \dots, N,$$

where $\begin{pmatrix} v_{00k} \\ v_{10k} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \delta_{00} & \delta_{01} \\ \delta_{10} & \delta_{11} \end{pmatrix} \right]$ and N is the number of pedigrees.

The random components v_{00k} and v_{01k} measure the variation of each pedigree's mean SBP and mean slope from their average in the whole sample.

Statistical tests in the multilevel model

Analyses were conducted in both the unselected and selected samples and with and without adjustment for BMI. Multilevel models were fitted using SAS PROC MIXED [3]. Parameter estimates are obtained by restricted maximum likelihood estimation (REML). An F-statistic was used to test the significance of the fixed effects with number of degrees of freedom computed using the containment method [4]. The likelihood ratio statistic based on REML likelihoods was used to test the significance of the random effects. The null distribution of this statistic is a mixture of χ^2_{q+1} and χ^2_q with equal weights 0.5, where q and q + 1 are the number of random effects estimated under H₀ and H₁, respectively.

Genome-wide linkage analysis

We used the estimates of the random effects at the subject and pedigree levels to define two new outcomes that were

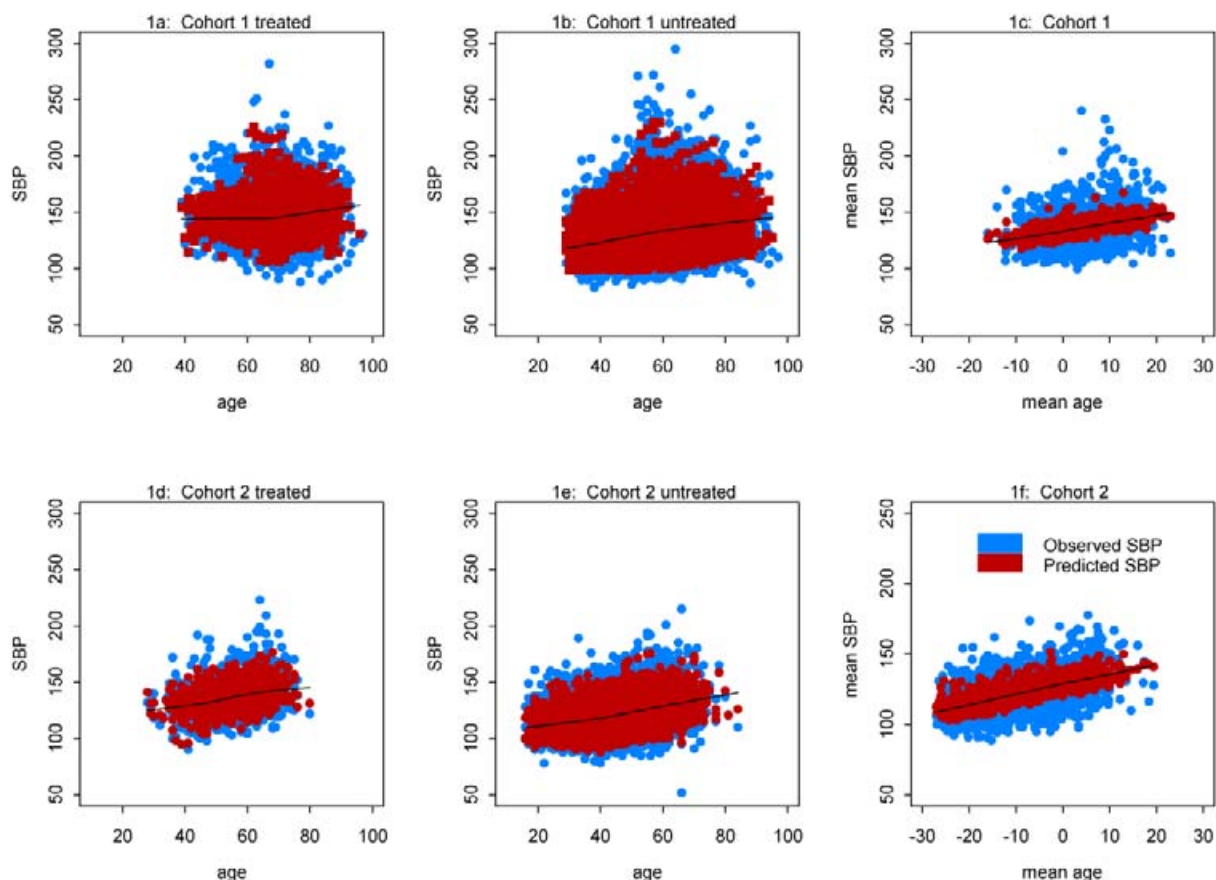


Figure 1
Comparisons of observed SBP and predicted values for the selected sample A, B, D, E are based on the multilevel model. C and F are based on the sample-wide regression for the selected sample [2].

used in the genome-wide linkage analysis. The two outcomes were defined as $\hat{u}_{0jk} + \hat{v}_{00k}$ and $\hat{u}_{1jk} + \hat{v}_{01k}$, which measure the random variation of each individual's SBP mean and slope, respectively, from the sample average after adjustment for the fixed effects. A third outcome was also defined using the residuals from a sample-wide regression in which each individual's mean SBP (across all exams) was regressed on his mean age (centered), mean BMI (centered), gender and cohort, as in Levy et al.'s paper [2]. Estimation of heritability and two-point linkage analyses were performed on the pedigree data using the variance component models implemented in the SOLAR package [5].

Results

Multivariate analysis of longitudinal SBP

All fixed effects included in the model were highly significant in the subject random slope model (Table 1) except for gender. Most of the SBP variability (316.8 in Model 1a, Table 1) was explained by within-subject (140.8, 44%) and between-subject (146.2, 46%) variability in the mean SBP and to a lesser extent by between-pedigree variability (27.6, <9%). Much less variability was explained by variability in the slope (0.17+0.008, <0.06%). Pedigree effects of mean SBP and SBP slope were more significant when the multilevel analyses were adjusted for body mass index. As shown in Figure 1, the multilevel model fit well

Table 2: Results of two-point linkage analysis (LOD scores ≥ 2 are in bold)

Markers	Location	Unselected Sample			Selected Sample		
		Mean SBP Multilevel Model	Mean SBP ^A Sample- wide Regression	SBP slope Multilevel Model	Mean SBP Multilevel Model	Mean SBP ^A Sample- wide Regression	SBP slope Multilevel Model
GATA48B01	1 (212 cM)	0.30	0.23	2.00	0.30	0.21	2.47
GATA11H10	2 (38 cM)	1.47	1.36	0	2.32	2.20	0.18
GATA6F06	3 (79cM)	2.16	1.78	0	1.17	0.81	0.14
GATA4A10	3 (153 cM)	0.31	0.21	2.27	0.33	0.31	1.29
GATA72C10	8 (37 cM)	3.57	2.54	1.00	3.57	2.82	1.71
ATA34E08	11 (33 cM)	0.16	0.11	2.13	0.03	0.01	1.54
GATA7G10	13 (64 cM)	2.02	1.92	0	1.79	1.63	0
GATA25A04	17 (62 cM)	2.13	1.78	0	3.54	3.42	0

^AResiduals from the sample-wide regression of the mean SBP on the mean age, mean BMI, gender, and cohort effects as in Levy et al. [2].

the data while the sample-wide regression does not capture all the SBP variability.

Heritability

Heritability estimates were 54.3% (SE = 3.1) and 55.6% (SE = 3.4) for the mean SBP, 31.9% (SE = 3.5) and 28.9% (SE = 3.5) for SBP slope over time in the unselected and selected samples, respectively. The heritability estimates for the subject-specific residuals from the sample-wide regression of the mean SBP were 47.7% (SE = 3.4) and 49.7% (SE = 3.8) in the unselected and selected samples, respectively.

Genome-wide linkage analysis

Evidence for linkage for the two outcomes (mean SBP and slope) was found in several chromosomal regions with a maximum LOD score of 3.6 on chromosome 8 and 3.5 on chromosome 17 for the mean SBP and 2.5 on chromosome 1 for SBP slope (Table 2). However, linkage on chromosome 8 for the mean SBP was only found in the selected sample. The decrease in LOD score in the unselected sample on chromosome 17 was important in several pedigrees that included individuals with a single SBP measurement, as illustrated in Figure 2. Adjusting the analyses for BMI showed stronger evidence for linkage, which could suggest that BMI is determined by other genetic factors (Table 3). Not adjusting the analysis for treatment effect did not change the results of the mean SBP, but yielded lower LOD scores for SBP slope (Table 3).

Discussion

Our study demonstrates the value of multilevel modeling in the search for genetic determinants of complex traits when longitudinal pedigree data are available. For the

mean SBP, we were able to replicate the linkage result on chromosome 17 previously reported by Levy et al. [2] and detect a new linkage on chromosome 8 that was not reported before. For SBP slope, we also found suggestive results for linkage for both mean SBP and SBP slope on several other chromosomal regions, including chromosomes 1, 2, 3, 11, and 13. Using residuals from the multilevel model in a genome-wide linkage analysis gave stronger evidence for linkage than using residuals from a sample-wide regression as in the Levy et al.'s paper [2]. This might be because this latter approach does not correctly account for within-individual and between-individual variability. Multilevel modeling, which can take into

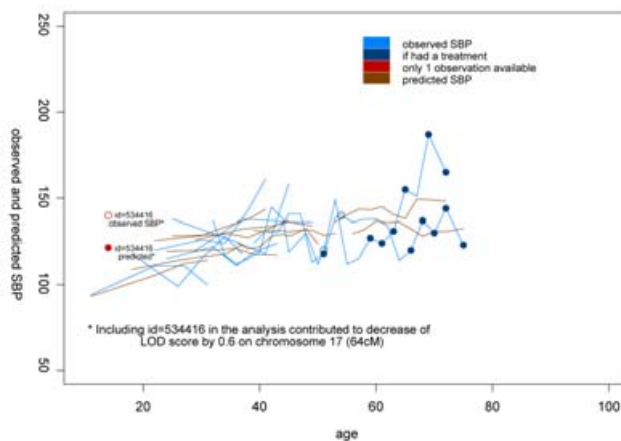


Figure 2
Observed and predicted SBP profiles for pedigree 32,445 members

Table 3: Results of two-point linkage analysis in the selected sample using residuals from the multilevel model (LOD scores ≥ 2 are in bold)

Markers	Location	Mean SBP Analysis adjusted for ^A			SBP Slope Analysis adjusted for ^A		
		BMI, treatment	BMI only	Treatment only	BMI, treatment	BMI only	Treatment only
GATA48B01	1 (212 cM)	0.30	0.30	0.17	2.47	1.83	1.53
GATA11H10	2 (38 cM)	2.32	2.53	0.93	0.18	0.25	0
GATA6F06	3 (79 cM)	1.17	1.21	1.49	0.14	0.02	0.06
GATA4A10	3 (153 cM)	0.33	0.32	0.02	1.29	0.74	0.70
GATA72C10	8 (37 cM)	3.57	3.52	1.32	1.71	1.13	0.68
ATA34E08	11 (33 cM)	0.03	0.02	0.01	1.54	1.19	0.93
GATA7G10	13 (64 cM)	1.79	1.68	0.71	0	0	0
GATA25A04	17 (62 cM)	3.54	3.36	1.48	0	0	0

^AAll analyses are also adjusted for age effect in Level 1 and age, sex, and Cohort effects in Level 2.

account the hierarchical structure of the data, may help disentangle the proportion of the trait variability explained by fundamental variation in the mean SBP and in the SBP slope from the proportion explained by random within-individual variability. A more general hierarchical structure could have included a nuclear family level nested within the pedigree level. However, such a multilevel model would be more difficult to fit. In our analysis we only included a fixed cohort effect that could account for differences between generations within a pedigree. Treating the pedigrees as random effects also allowed for between-pedigree heterogeneity in our model, which improved the accuracy of the random effect estimates at the individual level. Although there may be some concern about using a two-stage approach for detecting linkage, other studies based on similar strategies using linear mixed models in simulated data did not report an inflation of type I error for the test of linkage in the context genome-wide linkage analysis [6,7]. The linkage on chromosome 17 for mean SBP was only found in the selected sample. A important decrease in LOD score (>0.1) in the unselected sample was observed in several pedigrees comprising individuals with a single extreme SBP measurement, as illustrated in Figure 2. This suggests that a single SBP measurement may not provide a reliable characterization for an individual, especially when a familial study of SBP is designed. Adjusting the analyses for BMI showed stronger evidence for linkage, which could suggest that BMI is determined by other genetic factors. No correction was applied to the SBP value of subjects who received a hypertensive treatment. The analyses with the multilevel model were adjusted for treatment effect so that the residuals obtained from this model correspond to the untreated group. Taking into account an interaction between age and treatment in the multilevel model may also have reduced the bias due to treatment effect. How-

ever, our linkage results were insensitive to whether the analyses were adjusted for treatment effect. The multilevel modeling approach is also known to be robust to missing data, under the assumption that they are missing at random [4]. Future work could include the development of an integrated approach to perform linkage analysis within the multilevel framework.

Acknowledgments

This research was partially supported by a project grant from the Network of Centres of Excellence in Mathematics (Canada). SBB is a Senior Investigator of the Canadian Institutes for Health Research.

References

- Leyland AH, Goldstein H, Eds: **Multilevel Modelling of Health Statistics**. Chichester, John Wiley and Sons 2001.
- Levy D, DeStefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavvas 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.
- Littell R, Milliken G, Stroup W, Wolfinger R: **SAS System for mixed models**. Cary, NC, SAS Institute, Inc 1996.
- Verbeke G, Molenberghs G: **Linear Mixed Models in Practice. A SAS-Oriented Approach**. New York, Springer 1997.
- Almasy L, Blangero J: **Multipoint quantitative trait linkage analysis in general pedigrees**. *Am J Hum Genet* 1998, **62**:1198-1211.
- Palmer L, Jacobs K, Scurrah K, Xu X, Horvath S, Weiss S: **Genome-wide linkage analysis in a general population sample using sigma 2A random effects (SSARs) fitted by Gibbs sampling**. *Genet Epidemiol* 2001, **21**(suppl 1):S674-S679.
- Scurrah K, Tobin T, Burton P: **Longitudinal variance components models for systolic blood pressure, fitted using Gibbs sampling**. *BMC Genetics* 2003, **4**(suppl 1):S25.