Genome-Wide Association Studies for Bivariate Sparse Longitudinal Data

Kiranmoy Das^{a–c} Jiahua Li^{b, c} Guifang Fu^{b, c} Zhong Wang^{c, d} Rongling Wu^{b–d}

^aDepartment of Statistics, Temple University, Philadelphia, Pa., ^bDepartment of Statistics, The Pennsylvania State University, University Park, Pa., ^cCenter for Statistical Genetics, The Pennsylvania State University, and

^dDepartment of Public Health Sciences, The Pennsylvania State College of Medicine, Hershey, Pa., USA

Key Words

Bivariate response • Functional GWAS • Multiple testing • SNP

Abstract

Objective: Longitudinal measurements with bivariate response have been analyzed by several authors using two separate models for each response. However, for most of the biological or medical experiments, the two responses are highly correlated and hence a separate model for each response might not be a desirable way to analyze such data. A single model considering a bivariate response provides a more powerful inference as the correlation between the responses is modeled appropriately. In this article, we propose a dynamic statistical model to detect the genes controlling human blood pressure (systolic and diastolic). **Methods:** By modeling the mean function with orthogonal Legendre polynomials and the covariance matrix with a stationary parametric structure, we incorporate the statistical ideas in *functional genome-wide association studies* to detect SNPs which have significant control on human blood pressure. The traditional false discovery rate is used for multiple comparisons. **Results:** We analyze the data from the Framingham Heart Study to detect such SNPs by appropriately consider-

ing gender-gene interaction. We detect 8 SNPs for males and 7 for females which are most significant in controlling blood pressure. The genotype-specific mean curves and additive and dominant effects over time are shown for each significant SNP for both genders. Simulation studies are performed to examine the statistical properties of our model. The current model will be extremely useful in detecting genes controlling different traits and diseases for humans or non-human subjects.

Copyright © 2011 S. Karger AG, Basel

Introduction

The successful completion of the Human Genome Project (in 2005) made a revolution in detecting genes controlling various traits and diseases by genome-wide association studies (GWAS). In the past few years, GWAS have successfully identified a number of loci responsible for complex traits and diseases and by extensive genome-wide scans of genes, the genetic interactions and anomalous nature of complex genes have been better studied by geneticists and biomedical experts. This revolution is of paramount importance in developing advanced treat-

ment methods and powerful drugs for diseases with high complexity and enticed the researchers to focus on such studies.

Despite their potential effect on biomedical sciences, traditional GWAS suffer from several severe limitations. While in clinical trials each subject is measured at several time points not necessarily evenly spaced, GWAS consider a single time point measurement per subject and hence cannot reflect the underlying subject-specific trajectory for the longitudinal phenotype. Secondly, most GWAS have found only a small proportion of genetic variation and hence cannot explain the considerable genetic variances observed quite often in many traits. Also when correlated phenotypic traits are measured longitudinally for many subjects, traditional GWAS focus on single phenotype-genotype analyses without taking into account the possible dependence among the biomarkers. By integrating the basic idea for GWAS into powerful statistical approaches, Das et al. [1] proposed a new so-called *functional GWAS* (or *f GWAS*) approach which can well address the fundamental limitations of traditional GWAS.

For the analysis of univariate repeated measures, the linear mixed model-based approach [2] is quite popular and readily available in standard statistical packages. When more than one biomarker is measured repeatedly, longitudinal multivariate models should be used. A bivariate linear mixed model approach using the standard statistical package (SAS) was proposed by Thiebaut et al. [3]; Sithole and Jones [4] used such a model for detecting prescribing change in two drugs simultaneously with correlated errors. Since then it has been well recognized that for multiple correlated biomarkers, a multivariate linear model is more powerful than using a separate model for each biomarker.

In this article, we have used a similar bivariate longitudinal model to analyze data from the Framingham Heart Study (FHS) with the aim to unlock the door of the complex genetic architecture of human blood pressure which may lead to serious cardiovascular diseases. Two responses, systolic and diastolic blood pressure, are measured for 977 healthy subjects at different ages. Different subjects are measured at different time points and the numbers of measurements are quite different from subject to subject. Since the body mass index (BMI) is known to have a significant effect on blood pressure, we consider it as a covariate in our model and in order to incorporate the gender-gene interaction, we perform separate analyses for the male and female population. We detect single nucleotide polymorphisms (SNPs) which are most significant for controlling human blood pressure for the

male and female population. Simulation studies and power analyses have been performed to verify the effectiveness and importance of our approach.

Subjects and Methods

We use our model and method to analyze the bivariate longitudinal trait collected in the FHS. The objective of this study was to unlock the door of genetic architecture of cardiovascular diseases. The study began in the year 1948 with 5,209 healthy men and women aged 30–60 from Framingham, Mass., USA. Currently, the FHS project is in its third generation of participants. The study recorded the medical history of all its participants and laboratory tests were conducted to have actual data from the subjects. Different subjects were measured at different time points, but collectively the data provide records of phenotypes at many time points. Our data contain 977 subjects (all Caucasians), 500 males and 477 females, who had their systolic and diastolic blood pressure, BMI and many other variables measured at multiple time points. The number of serial measurements for a subject can be as low as 3, but the times and intervals of measurements are highly variable among the different subjects. Thus, despite high sparsity, our data set contains a large range of time points when all subjects are projected on a single time scale.

We have approximately 550,000 SNP data from the entire human genome. Following the general tradition of GWAS, we excluded SNPs with minor allele frequencies (MAF) of less than 0.10. The numbers and percentages of non-rare allele SNPs vary among different chromosomes and range from 4,417 to 28,771 and from 0.64 to 0.72, respectively. Our model, based on a direct relationship between phenotype and genotype, is used at the SNP level to detect significant SNPs associated with blood pressure by considering the gender-gene interaction.

Statistical Model

In the FHS, longitudinal traits are measured at irregular and possibly subject-specific time points. Let $y_{ik} = [y_{ik}(t_{i1}), y_{ik}(t_{i2}), \dots, y_{ik}(t_{iT_i})]$ denote the vector of trait of type k (for example, $k = 1$ for systolic blood pressure and $k = 2$ for diastolic blood pressure) for the i -th subject measured at time points $t_i = [t_{i1}, \dots, t_{iT_i}]$. Using this denotation, we can capture subject-specific differences in the number and interval of time points. In what follows, we consider n subjects measured at different time points. Consider a SNP with two alleles, A and a, generating three genotypes: AA (coded as 1) with n_1 observations, Aa (coded as 2) with n_2 observations and aa (coded as 3) with n_3 observations. The k -th type phenotypic value for subject i at time $t_{i\tau}$ ($\tau = 1, \dots, T_i$) is expressed as

$$y_{ik}(t_{i\tau}) = \sum_{j=1}^3 \xi_{ij} \mu_{jk}(t_{i\tau}) + \beta x_i(t_{i\tau}) + e_{ik}(t_{i\tau}), \quad (1)$$

where ξ_{ij} is an indicator variable taking value 1 if the i -th subject is of genotype j and 0 if otherwise, $\mu_{jk}(t_{i\tau})$ is the mean value for genotype j of the k -th response at time $t_{i\tau}$, x_i is the time-variant covariate, and β is the regression coefficient of the covariate. Since in the present situation the covariate is the BMI which is assumed to have the same effect on both systolic and diastolic blood pressure, we do not use subscript k for β . However, this might not be the case in general and depending on the context,

one has to decide whether to include the subscript or not. The residual error for subject i , $e_{ik}(t_{ir})$, is assumed to be distributed as multivariate normal with mean = 0 and covariance matrix = Σ_i .

The first important task in f GWAS is to model the genotype-specific mean curves and within-subject covariance structure. For modeling the mean curves, we used a nonparametric approach based on orthogonal Legendre polynomials (LP). These polynomials have already been proven a powerful tool to model longitudinal or functional data by several authors for nonparametric regression [5–7] and the robustness of nonparametric regression using LP has been investigated effectively. The LP are solutions to the Legendre differential equation

$$(1-x^2)\frac{d^2z}{dx^2} - 2x\frac{dz}{dx} + r(r+1)z = 0.$$

The general form of a LP of order r is given by the following sum

$$P_r(x) = \sum_{l=0}^r (-1)^l \frac{(2r-2l)!}{2^r l!(r-l)!(r-2l)!} x^{r-2l},$$

where

$$L = \frac{r}{2} \text{ or } \frac{r-1}{2}$$

whichever is an integer. These polynomials are defined over $[-1, 1]$ and are orthogonal to each other in this interval in the sense that $\int_{-1}^1 P_r(x)P_s(x)dx = 0$ when $r \neq s$. Let $P_r(t)$ denote the r -th order LP at time t . By appropriate transformation of the measurement times, we make proper adjustment to fit the orthogonal function range $[-1, 1]$. This can be achieved very easily by making the following transformation of the original time points $t_1, t_2, t_3, \dots$, to $t'_1, t'_2, t'_3, \dots$, where

$$t' = -1 + 2 \left(\frac{t - t_{\min}}{t_{\max} - t_{\min}} \right),$$

where $t_{\min}$ and $t_{\max}$ are the shortest and the longest time points, respectively. A family of orthogonal LP is denoted by $P(t) = [P_0(t), P_1(t), \dots, P_r(t)]^T$ and the genotype-specific mean value is expressed as a linear combination of the polynomials, such as

$$\mu_{jk}(t) = u_{jk}^T P(t), \quad (2)$$

where $u_{jk} = (u_{jk0}, u_{jk1}, \dots, u_{jkr})^T$ is called base vector. The order of the LP is chosen by information criterion like AIC/BIC, etc.

In the Statistics literature, there are a number of ways to model the time-dependent mean curves for longitudinal data. This includes Fourier series, B-spline, Wavelets, etc. If the trait under consideration is well studied and well understood, then a known parametric structure might be more effective and biologically meaningful to model the mean function. For example, in modeling the growth for plants and animals, a logistic equation might be used [8]. Triple logistic equations might be effective for human body growth [9]. Since these mathematical equations are derived from fundamental biological principles, one can incorporate these models in GWAS or f GWAS to capitalize their power in modeling the trait under consideration in an accurate and precise way. Semiparametric models [5] can also be implemented in some situations for a better understanding of the process under consideration.

In longitudinal data analysis, it is also fundamentally important to model the within-subject covariance structure in a robust and powerful way. If the trait variation within subjects is constant, one might use a stationary autoregressive model for modeling the covariance. The advantage of this approach is that one needs to estimate only two parameters, the variance and the correlation for this structure. While the stationary assumption is not worth, nonstationary approaches like the structured antedependence model can be used. In order to handle more complex covariance structures, the autoregressive moving average model (ARMA(p, q)) might be implemented too. More recently, Fan et al. [10] and Fan and Wu [11] have developed a semiparametric kernel-based covariance function which is robust and effective for modeling subject-specific covariance structure for longitudinal data analysis.

Here, we model the covariance matrix for the bivariate response longitudinal data as a Kronecker product of two covariance structures following Sithole and Jones [4]. For a fixed k ($= 1$ and 2), we assume that the longitudinal measurements from the same subject have an AR(1) structure, while the correlation between the systolic and diastolic blood pressure remains the same over time for a fixed subject. This kind of covariance structure can be expressed as $UN \otimes AR(1)$, where UN is a 2×2 symmetric matrix with σ_s^2 and σ_d^2 as the diagonal elements and σ_{sd} as the off-diagonal element. For instance, if we have three repeated measures from a subject, then we assume a covariance matrix with the structure $UN \otimes AR(1)$, where UN and $AR(1)$ are

$$\begin{bmatrix} \sigma_s^2 & \sigma_{sd} \\ \sigma_{sd} & \sigma_d^2 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix},$$

respectively. This specification of the covariance structure assumes that an intra-blood pressure correlation is the same for systolic and diastolic. Moreover, it assumes that an inter-blood pressure correlation is proportional to an intra-blood pressure correlation. However, this assumption is not too restrictive and such a covariance function can be modeled easily using the standard statistical packages. Such a covariance structure has been proposed by Thiebaut et al. [3] and Sithole and Jones [4] for regular longitudinal data when all subjects are measured at the same time points. We performed an exploratory data analysis to investigate the validity of such a structure for irregular longitudinal data using SAS, and it turned out that SAS is equally powerful in this case as well. Hence, when different subjects are measured at different time points (which is true for our FHS), one can model subject-specific covariance matrices in the way mentioned above, assuming the underlying covariance process is the same for all subjects.

The additive (a) and dominant (d) effects over time of the SNP under consideration can be expressed as follows

$$a_k(t) = \frac{1}{2} [\mu_{1k}(t) - \mu_{3k}(t)],$$

$$d_k(t) = \mu_{2k}(t) - \frac{1}{2} [\mu_{1k}(t) + \mu_{3k}(t)].$$

Fig. 1. Raw data for systolic and diastolic blood pressure for males (a) and females (b), respectively.

Table 1. BIC values for selecting the optimum order (r) of the LP

Order	BIC (male)	BIC (female)
1	4.62	3.85
2	1.62	1.95
3	3.84	2.77
4	1.99	2.01

Assuming the subjects under study are independent from each other, the joint likelihood function can be written as

$$L(y) = \prod_{i=1}^{n_1} f_1(y_i) \prod_{i=1}^{n_2} f_2(y_i) \prod_{i=1}^{n_3} f_3(y_i), \quad (3)$$

where $y_i = [y_{i1}, y_{i2}]^T$ is the response vector for the i -th subject, $f_j(y_i)$ is the multivariate normal density for the i -th subject carrying SNP genotype j ($j = 1, 2, 3$), with the following mean vectors

$$\begin{aligned} &[\mu_{11}(t_{i7}) + \beta x_i(t_{i7}), \mu_{12}(t_{i7}) + \beta x_i(t_{i7})]^T, \text{ for genotype AA,} \\ &[\mu_{21}(t_{i7}) + \beta x_i(t_{i7}), \mu_{22}(t_{i7}) + \beta x_i(t_{i7})]^T, \text{ for genotype Aa,} \\ &[\mu_{31}(t_{i7}) + \beta x_i(t_{i7}), \mu_{32}(t_{i7}) + \beta x_i(t_{i7})]^T, \text{ for genotype aa,} \end{aligned}$$

and subject-specific covariance matrix Σ_i , modeled as described earlier. In the traditional likelihood-based approach, we maximize the log likelihood to get the maximum likelihood estimate for the model parameters.

Hypothesis Tests

Once the model parameters are estimated, it is important to perform hypothesis tests in order to detect the significant genetic effect of a particular SNP. The hypotheses can be formulated in the following way: $H_0: \mu_{1k} = \mu_{2k} = \mu_{3k}$, for $k = 1$ and 2 , versus H_1 : at least one equality in H_0 does not hold.

Note that under H_0 , the mean functions for the three genotypes are the same, indicating no significant genetic effect for the SNP under consideration. By computing the likelihood function under H_0 and H_1 , one can compute the likelihood ratio (LR). The LR test statistic $-2\log(LR)$ asymptotically follows a χ^2 distribution with the degree of freedom equal to the difference in the numbers of unknown parameters under H_0 and H_1 . Because of the large

number of SNPs under consideration, we need to adjust for multiple comparison and here we use a standard false discovery rate (FDR) approach as proposed by Benjamini and Hochberg [12]. However, one can also apply more advanced FDR approaches but we avoid those for computational simplicity.

Results

Gene Detection

The raw data for systolic and diastolic blood pressure for males and females are shown in figure 1. In the exploratory data analysis, we first look for the optimal order (r) of the LP that fits our data best. To get the optimum order, we compute BIC values for $r = 1, 2, 3$ and 4 using the standard approach. Table 1 provides the corresponding BIC values for different orders and we see that a second order polynomial gives the best fit among the four different values of r which is therefore then taken as the optimum order.

Next, we consider one SNP at a time and use the model as calculated in equation 1 to test for a significant effect of the marker on the blood pressure trajectories. We have different numbers of SNPs for different chromosomes and we test for each SNP in each chromosome. We have the fixed effect of the genotype and also the effect of the BMI as the covariate. After estimating the model parameters, we perform the hypothesis testing as described in the previous section. The multiple testing problem is handled by controlling the FDR using Benjamini and Hochberg's algorithm [12].

As mentioned earlier, we have nearly 550,000 SNPs, we need to adopt a multiple hypothesis-testing approach to have control over the overall type 1 error probability. Although Bonferroni's method is a widely used approach, we did not apply this approach here simply because it is

Fig. 2. Manhattan plot for males: selection of the most significant SNPs controlling blood pressure.

Fig. 3. Manhattan plot for females: selection of the most significant SNPs controlling blood pressure.

too conservative in nature. We applied a FDR control for the multiple comparisons. With a significance level $\alpha = 10^{-6}$ (with an estimated FDR = 0.002), we select 8 significant SNPs for the male and 7 significant SNPs for the female population. The Manhattan plots for males and females are displayed in figures 2 and 3, respectively. For males, the detected SNPs are rs66475406 on chromosome

2, rs66149495 on chromosome 3, rs66299297 and rs66501706 on chromosome 5, rs66484226 on chromosome 8, rs66379521 on chromosome 12, rs66154967 on chromosome 17 and rs66092412 on chromosome 19. For females, we got rs66076226 on chromosome 1, rs66053327, rs66123984, rs66114880 and rs66347789 on chromosome 2, rs66447584 on chromosome 6 and rs66225752 on chro-

Fig. 4. Genotype-specific mean curves for the most significant SNPs for males.

mosome 18. Table 2 provides the names, positions, MAF and p values for all these sex-specific significant SNPs. Figures 4 and 5 show the sex-specific and genotype-specific mean trajectories for systolic and diastolic blood pressure. Figures 6 and 7 show the observed (black) versus expected (grey) p values in $-\log$ scale with base 10 for the male and female population, respectively.

Based on the fitted mean curves for different genotypes, we computed the additive and dominant effects over time. Figure 8 displays the age-specific changes of additive and dominant effects on blood pressure in the

different sexes. Those age-specific changes show a complicated pattern and in order to explain the molecular basis of these phenomena, one might want to refer to Kacser and Burns [13] and Keightley and Kacser [14].

Simulation, Power and False Positive Rate Analysis

Like for most GWAS results in the literature, we also performed a cross-validation analysis to investigate the reproducibility of bivariate *f* GWAS as proposed in this article. We split the data into the two sex groups which were of nearly equal size. We applied bivariate *f* GWAS

Fig. 5. Genotype-specific mean curves for the most significant SNPs for females.

to each group as well as on the complete sample and found that the same SNPs are detected as significant for males and females. To examine the power and the statistical properties of the newly proposed methodology, we performed a simulation study. Sparse trait values were simulated from the model in equation 1 and the covariance structure we used for original data analysis. We assumed a segregating SNP with different MAF associated with the trait under consideration, i.e. blood pressure. Table 3 shows the true model parameters, estimated model parameters and their standard errors. We note

that the estimated values are very close to the true parameter values with small standard errors, which indicates that the method we propose is accurate and precise enough and hence applicable for practical purposes. In figure 9, we show the simulation result by plotting actual (solid) versus fitted (dashed) curves for different genotypes.

The practical applicability and usefulness of the proposed joint analysis was assessed by further simulation studies. Under two different setups, when the bivariate data are affected by pleiotropic genes and when they are

Fig. 6. Observed versus expected P-plot for males.

Fig. 7. Observed versus expected P-plot for females.

Table 2. Associations between SNPs and age-specific changes of blood pressure for males and females

Chromosome	SNP	Position	Alleles	MAF	p value
Male-specific					
2	rs66475406	81,518,943	C/T	0.1652(C)	0.5353×10^{-6}
3	rs66149495	16,140,422	G/T	0.4038(G)	0.1511×10^{-6}
5	rs66299297	104,206,688	A/G	0.1137(A)	0.3244×10^{-6}
5	rs66501706	147,356,971	A/G	0.3481(A)	0.7316×10^{-6}
8	rs66484226	91,327,474	C/A	0.1125(C)	0.2436×10^{-6}
12	rs66379521	130,758,789	T/G	0.4158(T)	0.5921×10^{-6}
17	rs66154967	29,846,491	A/G	0.1505(A)	0.7514×10^{-6}
19	rs66092412	56,060,316	C/T	0.3490(C)	0.5090×10^{-6}
Female-specific					
1	rs66076226	180,385,227	A/G	0.4884(A)	0.5618×10^{-6}
2	rs66053327	10,213,937	C/T	0.1762(C)	0.3833×10^{-6}
2	rs66123984	10,211,140	C/T	0.3388(C)	0.3266×10^{-6}
2	rs66114880	10,208,865	G/C	0.3307(G)	0.031×10^{-6}
2	rs66347789	10,193,197	A/C	0.3702(A)	0.00014×10^{-6}
6	rs66447584	95,313,484	T/C	0.1846(T)	0.1409×10^{-6}
18	rs66225752	58,498,179	A/T	0.1226(A)	0.0566×10^{-6}

not affected by such genes, power and false positive rate (FPR) analyses are performed with different values for MAF. We considered two different sample sizes ($n = 1,000$ and $2,000$) for each situation. Table 4 shows the results when pleiotropic genes are there and it is clear that in such a situation joint analysis performs much better

in terms of power and FPR, i.e. the analysis results in higher power and lower FPR. However, table 5 shows the results when there is no pleiotropy: separate analysis should be recommended. In such a situation, joint analysis will have lower power and higher FPR than a separate analysis.

Fig. 8. Additive and dominant effects for the most significant SNPs for males (a) and females (b), respectively.

Fig. 9. Simulation results: actual (solid) versus fitted (dashed) curves for different genotypes.

Discussion

In last three years, there have been a number of successful applications of GWAS which is indeed a consequence of the successful completion of the Human Genome Project and its derivative hapmap project. By analyzing thousands of SNPs, this approach can effectively identify the function of multiple genetic changes that are responsible for polygenic traits and diseases. Despite of several limitations, GWAS play a major role in biomedical researches and in medical genetics too.

Most of the traditional GWAS consider one time point measurement per subject for the trait under consider-

ation. In all senses, a curve is more informative than a single point. This motivated us to develop GWAS with longitudinal traits. In most practical studies, the subjects are measured at irregular time points which give rise to irregular sparse longitudinal data. Advanced statistical approaches can effectively model the mean and covariance structure for these data, and hence they can produce more meaningful inferences. By incorporating the functional aspects of the observed trait in the traditional GWAS, a new approach (*f*GWAS) has been developed.

When we have a bivariate response variable and we collect data longitudinally per subject, a simple extension of the *f*GWAS can work better than the traditional ap-

Table 3. Simulation results mimicking the original data

Parameter	True value	Estimate	Standard error
u_{110}	27.09	26.62	0.0990
u_{111}	2.84	2.72	0.0899
u_{112}	1.78	1.64	0.0183
u_{120}	21.71	20.66	0.0617
u_{121}	1.67	1.65	0.0716
u_{122}	0.97	0.94	0.0091
u_{210}	29.19	28.33	0.0697
u_{211}	2.17	2.39	0.0978
u_{212}	-1.56	-1.58	0.0361
u_{220}	27.91	27.75	0.0721
u_{221}	1.77	1.81	0.0851
u_{222}	-1.41	-1.38	0.0098
u_{310}	34.19	33.17	0.0305
u_{311}	1.17	1.153	0.0503
u_{312}	1.79	1.73	0.105
u_{320}	30.15	31.01	0.0104
u_{321}	1.14	1.13	0.0310
u_{322}	1.84	1.81	0.1153
σ_s	1.34	1.38	0.0598
σ_d	1.86	1.88	0.0255
σ_{sd}	1.17	1.13	0.0658
ρ	0.65	0.62	0.0260

Table 4. Theoretical power and FPR of bivariate data affected by pleiotropic genes for separate and joint analysis

Sample size	MAF	Separate analysis		Joint analysis	
		power	FPR	power	FPR
1,000	0.1	0.67	0.115	0.72	0.103
	0.2	0.65	0.114	0.71	0.0914
	0.3	0.70	0.102	0.77	0.0681
	0.4	0.73	0.085	0.82	0.0502
	0.5	0.72	0.082	0.83	0.0601
2,000	0.1	0.71	0.110	0.81	0.0931
	0.2	0.71	0.098	0.80	0.0773
	0.3	0.74	0.092	0.82	0.0504
	0.4	0.73	0.088	0.83	0.0501
	0.5	0.72	0.081	0.83	0.0519

proaches. The only challenge here is to model the covariance structure appropriately. The present article deals with nonparametric modeling of the mean function and a parametric covariance structure assuming that the intra-response correlation is the same for each response and the inter-response correlation is directly proportion-

Table 5. Theoretical power and FPR of bivariate data not affected by pleiotropic genes for separate and joint analysis

Sample size	MAF	Separate analysis		Joint analysis	
		power	FPR	power	FPR
1,000	0.1	0.68	0.112	0.61	0.118
	0.2	0.67	0.110	0.60	0.117
	0.3	0.73	0.093	0.62	0.114
	0.4	0.74	0.094	0.63	0.112
	0.5	0.74	0.088	0.63	0.099
2,000	0.1	0.70	0.097	0.63	0.109
	0.2	0.69	0.089	0.62	0.098
	0.3	0.75	0.085	0.65	0.096
	0.4	0.75	0.085	0.66	0.096
	0.5	0.74	0.082	0.65	0.092

al to the intra-response correlation. However, authors are currently developing a new approach where this assumption will be relaxed and a more powerful f GWAS can be performed in the near future. Once genes having significant effects on blood pressure are identified, heart diseases which are the consequences of abnormal blood pressures can be treated and controlled more effectively. Also the same procedure could be applied to detect genes controlling other traits and diseases too.

In this article, we mainly focused on bivariate traits like systolic and diastolic blood pressure. The underlying assumption for this work is that when we observe one trait for a specific subject at a particular time point, we must also have measurement for the second trait at that point. For traits like blood pressure, this is true in almost all cases, but it might not be the case in a general setting. A more sophisticated method needs to be developed to handle those situations. For our method, both traits need to be measured concurrently. Even a simple extension of this method works for multivariate traits under the same underlying assumption.

One limitation of the approach we propose in this article is that we treat all subjects under study independent of each other. However, in reality, the subjects might be genetically related or even come from the same family. Once we have the information for genetic dependence of the subjects, we can incorporate that in our method and this will reflect biostatistically more informative results. We will also need to consider the impact of genetic imprinting [15]. Once we do that, we can actually address many meaningful and challenging issues in genetics and biomedical sciences.

References

- 1 Das K, Li J, Wang Z, Tong C, Fu G, Li Y, Xu M, Ahn K, Mauger D, Li R, Wu RL: A dynamic model for genome-wide association studies. *Human Genetics* 2011; DOI [10.1007/s00439-011-0960-6](https://doi.org/10.1007/s00439-011-0960-6).
- 2 Laird NM, Ware JH: Random effects model for longitudinal data. *Biometrics* 1982;38: 963–974.
- 3 Thiebaut R, Jacqmin-Gadda H, Chene G, Lepoint C, Commenges D: Bivariate linear mixed models using SAS proc MIXED. *Comput Methods Programs Biomed* 2002; 69:249–256.
- 4 Sithole JS, Jones PW: Bivariate longitudinal model for detecting prescribing change in two drugs simultaneously with correlated errors. *J Appl Stat* 2007;34:339–352.
- 5 Cui Y, Zhu J, Wu RL: Functional mapping for genetic control of programmed cell death. *Physiol Genomics* 2006;25:458–469.
- 6 Meyer K: Random regressions to model phenotypic variation in monthly weights of Australian beef cows. *Livest Prod Science* 2000; 65:19–38.
- 7 Huskova M, Sen PK: On sequentially adaptive asymptotically efficient rank statistics. *Seq Anal* 1985;4:125–151.
- 8 Guiot C, Degiorgis PG, Delsanto PP, Gabrielle P, Deisboeck TS: Does tumor growth follow a ‘universal law’? *J Theor Biol* 2003;225: 147–151.
- 9 Bock RD, Thissen D: Fitting multi-component models for growth in stature. *Proceedings of the Ninth International Biometrics Conference* 1976;1:431–442.
- 10 Fan J, Huang T, Li R: Analysis of longitudinal data with semi-parametric estimation of covariance function. *J Am Statist Assoc* 2007; 102:632–641.
- 11 Fan J, Wu Y: Semiparametric estimation of covariance matrixes for longitudinal data. *J Am Statist Assoc* 2008;103:1520–1533.
- 12 Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B* 1995;57:289–300.
- 13 Kacser H, Bums JA: The molecular basis of dominance. *Genetics* 1981;97:639–666.
- 14 Keightley PD, Kacser H: Dominance, pleiotropy and metabolic structure. *Genetics* 1987; 117:319–329.
- 15 Luedi PP, Hartemink AJ, Jirtle RL: Genome-wide prediction of imprinted murine genes. *Genome Res* 2005;15:875–884.

Erratum

In the article 'Genome-Wide Association Studies for Bivariate Sparse Longitudinal Data' by Das et al. [Hum Hered 2011;72:110–120], the following acknowledgement needs to be added:

Acknowledgements

The Framingham Heart Study project is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (N01 HC25195). The Genetic Analysis Workshop is supported by NIH grant R01 GM31575. The simulated data was supported by the Washington University Institute of Clinical and Translational Sciences, NIH grant 1U54RR023496. The GAW16 Framingham and simulated data used for the analyses described in this paper were obtained through dbGaP (accession numbers). The authors acknowledge the investigators that contributed the phenotype, genotype, and simulated data for this study. The manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or the NHLBI.