[47] Chadwick R. The Icelandic database—do modern times need modern sagas? *BMJ* 1999;**319:**441–44.

[48] Helgadottir A, Manolescu A, Thorleifsson G *et al*. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet* 2004;**36:**233–39.

[49] Magnus P, Arnesen E, Holmen J *et al*. CONOR (Cohort NORway): historie, formĺl og potensiale. *Norsk Epidemiologi* 2003;**13:**79–82.

[50] Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust NZ J Public Health* 1999;**23**:453–59.

[51] Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994;**54:**535–43.

[52] Soler J, Blangero J. Longitudinal familial analysis of blood pressure involving parametric (co)variance functions. *BMC Genet* 2003;**4(Suppl.):**86.

[53] Yang Q, Chazaro I, Cui J *et al*. Genetic analyses of longitudinal phenotype data: a comparison of univariate methods and a multivariate approach. *BMC Genetics* 2003;**4(Suppl. 1):**S29.

[54] MacGregor S, Knott SA, White I, Visscher PM. Longitudinal variance-components analysis of the Framingham Heart Study data. *BMC Genetics* 2003;**4(Suppl. 1):**S22.

[55] Burton PR, Palmer LJ, Jacobs K, Keen KJ, Olson JM, Elston RC. Ascertainment adjustment: where does it take us? *Am J Hum Genet* 2000;**67:**1505–14.

[56] Burton PR. Correcting for non-random ascertainment in generalized linear mixed models (GLMMs) fitted using Gibbs sampling. *Genet Epidemiol* 2003;**24:**24–35.

[57] Levy D, DeStefano AL, Larson MG *et al*. 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:**477–83.

[58] Chandler PJ, Bock RD. Age-changes in adult stature—trend estimation from mixed longitudinal data. *Ann Hum Biol* 1991;**18:**433–40.

[59] Knuiman MW, Divitini ML, Bartholomew HC, Welborn TA. Spouse correlations in cardiovascular risk factors and the effect of marriage duration. *Am J Epidemiol* 1996;**143:**48–53.

[60] Palmer LJ, Knuiman MW, Divitini ML *et al*. Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. *Eur Respir J* 2001;**17:**696–702.

[61] Boomsma DI, Vink JM, van Beijsterveldt T *et al*. Netherlands Twin Register: a focus on longitudinal research. *Twin Res* 2002;**5:**401–06.

[62] Hansen J, de Klerk NH, Croft M, Alessandri P, Burton P. The Western Australian Twin Child Health (WATCH) Study: work in progress. *Austr Epidemiol* 2000;**7:**16–20.

[63] Hansen J, Allesandri PT, Croft ML, Burton PR, de Klerk NH. The Western Australian Register of Childhood Multiples: effects of questionnaire design and follow-up protocol on response rates and representativeness. *Twin Res* 2004;**7:**149–61.

[64] Scurrah KJ, Sheehan NA, Burton PR. Association and linkage for age at onset of a common oligogenic disease using genetic variance component models. *Genet Epidemiol* 2001;**21(Suppl. 1):** S680–85.

[65] Mancia G, Sega R, Grassi G, Cesana G, Zanchetti A. Defining ambulatory and home blood pressure normality: further considerations based on data from the PAMELA study. *J Hypertens* 2001;**19:**995–99.

[66] Gilmour AR *et al*. ASREML Manual. Orange, NSW, Australia: New South Wales Department of Agriculture, 2002.

[67] Meyer K. DFREML Version 2.1—Programs to estimate variance components by restricted maxiumum likelihood using a derivative-free algorithm. Armidale, NSW, Australia: AGBU, University of New England, 1992.

[68] SAS. Inc., SAS Institute. 99–2001. Cary, NC, USA, SAS Institute Inc.

Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2005;**34:**1077–1079

doi:10.1093/ije/dyi156

Commentary: Models for longitudinal family data

W James Gauderman* and David V Conti

Cohort studies will become increasingly important in understanding the aetiology of complex human traits.1 While the longstanding approach of analysing cross-sectional data to identify genetic and/or environmental factors for disease or quantitative traits has resulted in some success, there have been many inconclusive results and far too few replications. There are recognized explanations that are often put forth for this, including low power and heterogeneity across study samples. However, a reason that is not often cited is that a single cross-sectional examination of data may not capture the essential aetiological mechanisms. For example, a specific variant genotype might cause an increase in a trait value that cumulates as a person ages. That is, a specific gene may affect the trajectory of the trait over time. Thus, two studies, one of young-aged subjects and the other of older-aged subjects, would likely come to different conclusions with respect to that locus owing to the different part of the gene–age trajectory that was examined. In a similar manner, different genes may also act at different time periods in the disease process, such as disease initiation or

Department of Preventive Medicine, University of Southern California School of Medicine, 1540 Alcazar Street, CHP-200, Los Angeles, CA 90089-9011, USA

* Corresponding author. E-mail: jimg@usc.edu

progression. In such situations, underlying genes may only be identified through longitudinal studies that accurately capture the dynamic nature of the phenotype.

The standard cohort study involves longitudinal follow-up of individuals (unrelated subjects). Although this is an effective design for studying measured factors (candidate genes and environmental exposures), it does not permit estimation or adjustment for unmeasured genetic or environmental factors that are shared within families. Understanding the distribution of traits with respect to shared genetic and environmental factors is an important first step in the progression from descriptive genetic epidemiology to targeted studies of specific loci or genomewide searches using linkage or association methods.

The paper by Burton et al.2 in this issue appropriately points out the need to integrate longitudinal and family studies, and discusses several advantages of combining these data types. By basing their modelling framework on generalized linear mixed models (GLMMs) and using a Bayesian estimation procedure, they develop a flexible approach for estimating the fraction of variance in both trait level and trait slope over time that can be attributed to additive genetic and shared (within family) environmental effects. Moreover, the model is applicable to repeated quantitative or binary traits. This, combined with their previous work for survival traits,3 encompasses a class of models and estimation methods that should handle almost any type of outcome one would collect in a longitudinal family study.

In their analysis of systolic blood pressure in the Framingham Heart Study data, Burton et al.[2] estimated the narrow-sense heritability for slope over time of only 9%, but a much larger heritability (44.3%) for the intercept. As the authors point out, the latter refers to the proportion of total variance attributable to additive genetic effects at baseline. Care must always be taken when interpreting what is meant by 'baseline'. To provide a context for understanding this issue, we show their linear model for SBP:

$$\mathrm{SBP}_{ijk} = \beta_0 + \beta_{0ij} + \beta_T\,(T_{ijk} - T^*) + \beta_{Tij}\,(T_{ijk} - T^*) + \boldsymbol{\gamma X}_{ijk} + e_{ijk},$$

where i, j, and k index family, individual, and measurement number, respectively, T denotes age, and T* is a fixed age value. Remaining terms include a vector of measured covariates (X) and an error (e) that is assumed to be normally distributed with mean zero. The authors set T* to 52.7 years, the mean age in the sample. The parameters bT and bTij measure the overall average slope of SBP on age and the subject-specific deviation in slope from that average, respectively. The latter is treated as a random effect, is modelled as a function of genetic and shared environmental components of variance, and is the source of the estimated 9% heritability in slopes. These slope parameters are invariant to the choice of T* and estimate the change over the time-period for which the longitudinal measures have been obtained.

In contrast, what do the intercept parameters measure? Both the overall average (b0) and subject-specific deviation (b0ij) parameterize the mean SBP at baseline, where baseline refers to the covariate profile at which all other terms in the model drop out. In the above model, this is when age 5 T* 5 52.7 years and each X has a zero value (e.g. female of average weight, height, etc.). Therefore, the reported heritability of 44.3% for baseline SBP is referable to the distribution of SBP at age 52.7 and reflects cumulative heritability up to that age. It is important to recognize that the intercept-based heritability is not invariant to the choice of T*. In other words, one will get a different estimated heritability for the intercept (mean SBP) with different choices of T*. This fact can be used to advantage to better understand the effect of underlying genes on longitudinal trajectory. For example, one could run repeated analyses setting T* to 0, 10, 20, etc. to estimate heritability at birth, age 10, age 20, etc. The estimated heritability at age 10, for example, represents the cumulative effect of heritability at birth plus genetic effects on the growth slope that occurred between birth and age 10.

We note that the linear model above could be generalized (e.g. using a linear spline model), to allow both the growth slope and heritability estimates on slopes to vary over time. This would be important for a trait such as lung function, which increases rapidly through childhood and then decreases slowly over time in adulthood. For this trait, one could imagine that there might be different genetic influences in the growth and decline processes. Studying the change in heritability across ages might suggest the possibility of environmental factors that act through time to enhance or suppress gene expression.

In the light of currently available technologies, one is unlikely to be satisfied with only estimates of heritability. Instead, investigators are likely to want to study candidate genes and perform genome screens by association-based methods. Certainly, having age-specific estimates of heritability in hand should be viewed as a key step in designing an optimal study to test specific genetic loci. However, one may be tempted to then discard the family-based design in place of easier-to-conduct case–control or cohort studies of unrelated individuals. In our view, the family-based cohort study can still play an important role in the context of gene-association studies. First, tests of gene associations for trait-average and trait-slope based on within-family comparisons will be free from biases owing to population stratification. Second, one will have the opportunity to conduct joint tests of linkage and association. Joint models can yield a more powerful test than either a linkage or association test alone, and can be useful for distinguishing a marker from a true underlying trait locus.4 Third, one can monitor estimates of heritability for both intercepts and slopes as measured genes are added (as covariates X) to the model. This would provide one way of determining how fruitful it might be to continue searching for additional genes and may suggest targeting additional searches to specific age groups. Although, such extensions to measured genotypes can quickly lead to complex models, the flexibility of the GLMM framework and the ability of Bayesian estimation procedures to handle large integrations make such extensions feasible, albeit computationally demanding (see ref. 5 for an application to measured genotypes using GLMMs and Bayesian estimation).

In summary, the work by Burton et al.2 and others (see ref. 6 for a summary) highlights the potential importance of longitudinal family-based studies. These studies will be costly and more difficult to perform than either a family-based cross-sectional study or a longitudinal study of unrelated individuals. However, the combination of these two designs is likely to pay large dividends in our attempts to understand genetic and

environmental determinants of complex human traits. The longitudinal family-based design should be given careful consideration as we plan new studies.

References

[1] Collins FS. The case for a US prospective cohort study of genes and environment. *Nature* 2004;**429:**475–77.

[2] Burton P, Scurrah K, Tobin M, Palmer LJ. Covariance components models for longitudinal family data. *Int J Epidemiol* 2005;**34:** 1063–77.

[3] Scurrah K, Palmer LJ, Burton P. Variance components analysis for pedigree-based censored survival data using generalized linear mixed models (GLMMs) and Gibbs sampling in BUGS. *Genet Epidemiol* 2000;**19:**127–48.

[4] Millstein J, Siegmund K, Conti D, Gauderman W. Testing association and linkage using affected sib-parent study designs. *Genet Epidemiol* (in press).

[5] Conti DV, Gauderman WJ. SNPs, haplotypes, and model selection in a candidate gene region: The SIMPle analysis for multilocus data. *Genet Epidemiol* 2004;**27:**429–41.

[6] Gauderman W, Macgregor S, Briollais L et al. Longitudinal data analysis in pedigrees. *Genet Epidemiol* 2003;25(Suppl.1):S18–28.