Letter to the Editor

PBAT: Tools for Family-Based Association Studies

To the Editor:

A large number of computer programs are available for family-based association tests (FBATs), including QTDT (Abecasis et al. 2000), FBAT (Horvath et al. 1998, 2001; Laird et al. 2000; Lake et al. 2000), and PDT (Monks and Kaplan 2000). These programs primarily focus on the computation of various test statistics. Here, we discuss a new integrated software package called "PBAT" that contains tools for the planning of family-based association studies, as well as tools for the data analysis of such studies.

For nuclear families, PBAT provides power calculations for virtually any study design. PBAT's data-analysis tools can handle nuclear families, as well as extended pedigrees. The data-analysis functions include univariate and multivariate tests for various trait types, procedures for effect-size estimation, and screening techniques to select the most "promising" combinations of markers and phenotypes. All *P* values can be computed on the basis of both asymptotic theory and permutation tests. PBAT implementations for Windows XP, Linux, and Unix Solaris are freely available at the authors' Web site.

For dichotomous and continuous traits, PBAT computes the power for study designs that consist of different family types with varying numbers of additional offspring, with missing parental genotypes, and under different ascertainment conditions. The data-analysis tools of PBAT offer a unique variety of FBAT statistics (Laird et al. 2000) for univariate and multivariate traits, for gene-covariate interactions, for time-to-onset data, and for repeated measurements. Most of these tests are not available in other programs. All FBATs can be adjusted for covariates, and their *P* values can be computed either by asymptotic theory or by permutation tests. For the observed data set, PBAT provides functions that assess the power of the observed data set, compute the most powerful test statistic, estimate the genetic-effect sizes in different ways, and provide screening techniques/testing strategies that select the optimal combinations of markers and phenotypes for testing (Lange et al. 2003*b*, 2003*c*). Especially in the context of large-scale associ-ation studies, in which vast numbers of SNPs are available, PBAT's screening techniques become crucial tools to handle the multiple-comparison problem and to establish overall significance of associations between SNPs and traits.

PBAT is based on two key components: the approach to conditional power calculations for a general class of family-based association tests by Lange and Laird (2002*a*, 2002*b*) and the conditional-mean–model approach for FBATs introduced elsewhere (Lange et al. 2003*b*, 2003*c*). The first core function of PBAT is an implementation of the approach to conditional power calculations. For nuclear families and extended pedigrees, this function computes the distribution of any FBAT statistic under the null and alternative hypotheses. This function is derived from the algorithm proposed by Rabinowitz and Laird (2000). The second core function estimates all parameters of the conditional mean model for FBATs (Lange et al. 2003*b*, 2003*c*), using generalized estimating equations (Liang and Zeger 1986) without biasing the significance level of any FBAT statistic that is computed subsequently.

The design section of PBAT contains functions that assist the user in planning family-based association studies. For virtually any given study design and set of ascertainment conditions, the user can assess the power of the FBAT statistic (Laird et al. 2000) and decide whether the planned study has sufficient power. PBAT's interactive design allows the user to change the design, the ascertainment conditions, and the underlying genetic model/mode of inheritance. The effects of such design changes can be examined without much effort by the user. The PBAT tools for power calculations are a software implementation of the approaches to analytical power calculations for FBATs described elsewhere (Lange and Laird 2002*a*, 2002*b*; Lange et al. 2002*a*). PBAT computes the power of FBATs for a variety of scenarios. PBAT handles dichotomous/binary traits, continuous traits, missing parental information, multiple offspring per family, and any combination of different family types. Furthermore, the user can specify different genetic models, and ascertainment conditions and the linkage disequilibrium between the marker and disease locus. All power calculations can be verified by Monte-Carlo simulations.

PBAT's data-analysis tools contain a unique set of

functions for the statistical analysis of family-based association studies. The theory for PBAT's data-analysis tools has been discussed in a series of articles (Lange and Laird 2002*a*, 2002*b*; Lange et al. 2002*b*, 2003*a*, 2003*b*, 2003*c*, in press; DeMeo et al. 2002; Lyon et al., in press). These tools allow for a variety of analysis possibilities: computation of the marker distribution under the null and alternative hypotheses for nuclear families and extended pedigrees, transformation tools for phenotypes/traits, multivariate FBATs (FBAT-GEE based on the generalized-estimating-equation approach), FBATs for repeated measurements (FBAT-PC based on principal components that maximize the heritability), and time-to-onset FBATs (log rank FBAT, Wilcoxon FBAT, and FBAT-EXP). The *P* values for all of these test statistics can be obtained by use of either asymptotic theory or permutation tests. Conditional upon the sufficient statistic for each family (Rabinowitz and Laird 2000), the permutation-test option of PBAT permutes all possible configurations of observed marker scores on the basis of their probabilities under the null hypothesis.

Furthermore, PBAT provides functions for conditional power calculations for all implemented FBATs and for the construction of the most powerful FBAT statistic. PBAT can accommodate gene-environment/drug interactions in the FBAT statistic and can estimate the genetic-effect size without biasing the significance level of subsequently computed FBAT statistics. PBAT can also screen all combinations of markers and phenotypes without biasing the significance level of the subsequently computed FBAT statistic.

PBAT is still being refined, and our goal is to accommodate developments of the field into PBAT. The power-calculation functions of PBAT will be extended so that the required sample size for the desired significance can be retrieved directly and optimal designs can be computed on the basis of user-defined cost functions for the screening process, for the genotyping, and for the phenotyping. Another development will be to extend all PBAT functions for power calculations and data analysis to haplotype analysis.

Acknowledgments

Christoph Lange,[1] Dawn DeMeo,[2]
Edwin K. Silverman,[2] Scott T. Weiss[2]
and Nan M. Laird[1]
[1]*Department of Biostatistics, Harvard School of Public Health, and* [2]*Department of Medicine, Brigham and Women's Hospital, Boston*

Electronic-Database Information

The URL for data presented herein is as follows:

PBAT, http://www.biostat.harvard.edu/~clange/default.htm (for PBAT implementations)

References

Abecasis GR, Cardon LR, Cookson WOC (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292

DeMeo DL, Lange C, Silverman EK, Senter JM, Drazen JM, Barth MJ, Laird NM, Weiss ST (2002) Univariate and multivariate family based analysis of the arg130gln polymorphism of the *IL13* gene in the childhood asthma management program. Genet Epidemiol 23:335–348

Horvath S, Laird NM (1998) A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet 63:1886–1897

Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype-phenotype associations. Eur J Hum Genet 9:301–306

Laird NM, Horvath S, Xu X (2000) Implementing a unified approach to family based tests of association. Genet Epidemiol Suppl 19:36–42

Lake SL, Blacker D, Laird NM (2000) Family-based tests of association in the presence of linkage. Am J Hum Genet 67:1515–1525

Lange C, Blacker D, Laird NM. Family-based association tests for survival and times-to-onset analysis. Stat Med (in press)

Lange C, DeMeo DL, Laird NM (2002*a*) Power and design considerations for a general class of family-based association tests: quantitative traits. Am J Hum Genet 71:1330–1341

Lange C, DeMeo DL, Silverman EK, Weiss ST, Laird NM (2003*a*) Using the noninformative families in family-based association tests: a powerful new testing strategy. Am J Hum Genet 79:801–811

Lange C, Laird NM (2002*a*) Analytical sample size and power calculations for a general class of family-based association tests: dichotomous traits. Am J Hum Genet 71:575–584

—— (2002*b*) On a general class of conditional tests for family-based association studies in genetics: the asymptotic distribution, the conditional power and optimality considerations. Genet Epidemiol 23:165–180

Lange C, Lyon H, DeMeo DL, Raby B, Silverman EK, Weiss ST (2003*b*) A new powerful non-parametric two-stage approach for testing multiple phenotypes in family-based association studies. Hum Hered 56:10–17

Lange C, Silverman EK, Xu X, Weiss ST, Laird NM (2003*c*) A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics 4:195–206

Lange C, Whittaker JC, Macgregor AJ (2002*b*) Generalized estimating equations: a hybrid approach for mean parameters in multivariate regression models. Stat Model 2:163–181

Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22

Lyon H, Lange C, Lake SL, Silverman EK, Randolph AG,

Kwiatkowski DJ, Raby B, Lazarus R, Weiland KM, Laird NM, Weiss ST. IL10 gene polymorphisms are associated with asthma phenotypes in children. Genet Epidemiol (in press)

Monks SA, Kaplan NL (2000) Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. Am J Hum Genet 66:576–592

Rabinowitz D, Laird NM (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum Hered 50:211–223

Address for correspondence and reprints: Dr. Christoph Lange, Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115. E-mail: clange@hsph.harvard.edu