Genetic Epidemiology 33: 549–558 (2009)

Statistical Validation of Endophenotypes Using a Surrogate Endpoint Analytic Analogue

Guan-Hua Huang,[1]* Chin-Chiang Hsieh,[1] Chen-Hsin Chen,[2,3] and Wei J. Chen[3]

[1]*Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan*
[2]*Institute of Statistical Science, Academia Sinica, Taipei, Taiwan*
[3]*Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan*

Endophenotypes, which involve the same biological pathways as diseases but presumably are closer to the relevant gene actions than diagnostic phenotypes, have emerged as an important concept in the genetic studies of complex diseases. In this report, we develop a formal statistical methodology for validating endophenotypes. The proposed method was motivated by the conditioning strategy used for surrogate endpoints commonly seen in clinical research. We define an endophenotype to be ''a trait for which a test of null hypothesis of no genetic heritability implies the corresponding null hypothesis based on the phenotype of interest''. An index, the proportion of heritability explained, is used as an operational criterion of validation. Statistical inferences on this index are also developed. Usefulness of the proposed method is demonstrated through computer simulations and a study of assessing the Continuous Performance Test as an endophenotype of the schizophrenia spectrum. *Genet. Epidemiol.* 33:549–558, 2009.

Key words: Continuous Performance Test; heritability; liability threshold model; schizophrenia-related personality disorder; variance component analysis

Contract grant sponsor: National Science Council, Taiwan; Contract grant numbers: NSC96-2118-M-009-001-MY2; NSC96-3112-B-001-017; NSC97-3112-B-001-020; Contract grant sponsor: the Academia Sinica Thematic Project; Contract grant number: AS91ISS1PP; Contract grant sponsor: National Health Research Institute, Taiwan; Contract grant numbers: DOH83-HR-306 to DOH87-HR-306; NHRI-EX91-9113PP.
*Correspondence to: Guan-Hua Huang, Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 30010, Taiwan. E-mail: ghuang@stat.nctu.edu.tw
Received 16 August 2008; Revised 12 November 2008; Accepted 22 December 2008
Published online 4 February 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20407

INTRODUCTION

Genetic studies based on phenotypes of diseases that do not follow typical Mendelian inheritance patterns may not be optimal [Gottesman and Gould, 2003]. These ''complex'' diseases are influenced by multiple genes, environmental factors, and their interactions on phenotypes. Additionally, diagnosis of phenotypes is complicated by the possibility of mild forms of disease, fluctuation of clinical features of disease over time, and comorbidity with other diseases [Almasy and Blangero 2001; Gottesman and Gould, 2003]. As a result, the direct relationship between the phenotype and the genotype is disrupted. To facilitate the identification of influential genes of complex diseases, the endophenotype approach has been advocated. In the literature, synonyms of endophenotypes such as intermediate phenotypes, biological markers, and sub-clinical traits have been used interchangeably with slightly different implications. Gottesman and Gould [2003] argued that putative endophenotypes should provide a means for identifying the ''downstream'' traits of clinical phenotypes, as well as the ''upstream'' consequences of genes. Also see Pan et al. [2006] that delineates ''upstream intermediate phenotypes'' as the equivalent of endophenotypes with examples related to several complex diseases. Hence, endophenotypes are closer to underlying genes than phenotypes in the course of disease's natural history. Endophenotype-based genetic analysis is more likely to succeed than phenotype-based one in terms of search for susceptibility genes; nevertheless, there are emerging needs of systematic statistical methods for endophenotype-based analysis.

On the other hand, surrogate endpoints have been frequently utilized in clinical research, when the primary endpoint is costly or time-consuming to obtain. For example, psychiatric disorders are often assessed by applying standardized criteria to patients' report of symptoms [Eaton et al., 1989]. Biomarkers are used very often as substitutes for observing new cases of cancer in testing treatments for cancer prevention, where event rates are low and a long time may be needed to obtain cancer cases [Piantadosi, 1997]. Ultimately, one may demonstrate the treatment effect on the primary endpoint if a treatment effect is shown on the surrogate. A good deal of statistical research in evaluation of surrogate endpoints has been undertaken for about two decades. Prentice [1989] presented a landmark definition of surrogate endpoints and operational criteria for validating them. It has entailed extensive discussion on stringency and verifiability of the criteria; see Begg and Leung [2000], De Gruttola et al.

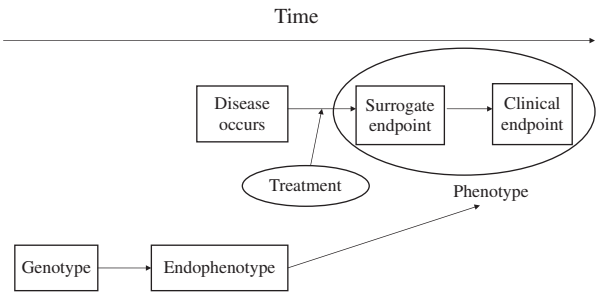

Fig. 1. An endophenotype versus a surrogate endpoint in the disease process.

[2001] and references therein. Freedman et al. [1992] introduced "the proportion of treatment effect on the primary endpoint explained" (PTE) by the surrogate to supplement Prentice's criteria.

Given that statistical methods for evaluation of surrogate endpoints in clinical studies have been relatively well developed, it would be interesting to (i) examine, conceptually, the similarities and distinctions between surrogate endpoints and endophenotypes and (ii) discuss how methods developed for the former may be utilized and/or modified for the latter. In general, an endophenotype is downstream from genes and upstream from a phenotype, and a surrogate is an upstream event to a clinical endpoint as illustrated in Figure 1. Noticeably, the causal pathway from an intervention to the primary endpoint via surrogate endpoints in surrogate analysis can be viewed as an analogy of the pathway from genotypes to the phenotype via endophenotypes in the endophenotype-based analysis. This article implements this connection and develops a formal statistical methodology for assessing the utility of endophenotypes. Our proposed methodology is especially useful for the situation where the underlying genotype is unknown and researchers use endophenotypes to increase opportunities of finding susceptible disease genes, not to verify whether specific genes are the cause of a disease.

MODELS

CURRENT CRITERIA FOR AN ENDOPHENOTYPE

Gottesman and Gould [2003] suggested the following five criteria for identification of endophenotypes:

1. The endophenotype is heritable.
2. The endophenotype found in affected family members is found in non-affected family members at a higher rate than in the general population.
3. The endophenotype is associated with illness in the population.
4. Within families, the endophenotype and illness cosegregate.
5. The endophenotype is primarily state-independent (manifests in an individual whether or not illness is active).

Criteria 1 and 2 imply that there are genetic effects underlying the endophenotype. Criteria 3 and 4 require the endophenotype to be associated with the phenotype on both population-and family-levels. These criteria suggest that the common genetic effects among relatives may cause this correlation. The reason for requiring criterion 5 is that relatives may be studied before the age of onset and one would expect the endophenotype to be manifested in the well relatives for correctly evaluating the relationship between the endophenotype and the phenotype.

These criteria are straightforward and several researchers have proposed designs and analyses to evaluate these criteria [Waldman, 2005]. However, in determining the feasibility of candidate endophenotypes, few studies have met all the criteria listed earlier [Gottesman and Gould, 2003]. Furthermore, they only "indirectly" imply that the genetic effects underlying the endophenotype can also affect the phenotype (criteria 3 and 4). It is important to demonstrate that the endophenotype and the phenotype share common causal genes because one does not expect susceptibility genes detected through the endophenotype-based analysis are only unique to the endophenotype. In the following section, we first propose our definition of endophenotypes, which directly addresses the concern of "common genes", and then provide an index for operationally evaluating the proposed definition.

STATISTICAL VALIDATION OF ENDOPHENOTYPES

Endophenotypes are useful for theorizing about clinical phenotypes and can mark the path between the genotype and the phenotype. Verification of existence of the pathway from genotypes to phenotypes via endophenotypes is the key of validating endophenotypes. Both the endophenotype and the surrogate endpoint lie in a biological pathway, but with two important differences: (i) the endophenotype is expected to be closer to the upstream genotype to increase the chance of identifying it, while the surrogate endpoint intends to substitute the downstream primary endpoint, and (ii) when the purpose of the study is to identify responsible genes for the phenotype, the genotype information is usually unknown, whereas treatment status in validating a surrogate endpoint is known.

Prentice [1989] defined a surrogate endpoint to be "a response variable for which a test of null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true (clinical) endpoint". Prentice's definition can be written as

$$f(S|X) = f(S) \Leftrightarrow f(T|X) = f(T),$$

where T denotes the status of a primary endpoint, S denotes the status of a surrogate endpoint, X is the treatment variable, $f(S)$ is the distribution of S, and $f(S \mid X)$ is the conditional distribution of S given X [Prentice, 1989; Buyse and Molenberghs, 1998].

By analogy, we define an endophenotype to be "a trait for which a test of null hypothesis of no genetic heritability implies the corresponding null hypothesis based on the phenotype of interest". More specifically, suppose P is the phenotype of interest, E is the selected endophenotype, and G represents an underlying genetic structure that fulfills specified assumptions in calculating heritability, then the proposed definition is

$$f(E|G) = f(E) \Rightarrow f(P|G) = f(P). \quad (1)$$

Furthermore, the endophenotype that fulfils (1) can share parts of genes that underlie the phenotype P.

The proposed definition has two important features. First, "imply" replaces the "if and only if" statement in Prentice's definition of surrogate endpoints. This change places the endophenotype in higher upstream of the pathway from the genotype to the phenotype, instead of retaining the relationship between the genotype and the endophenotype similar to that between the genotype and the phenotype. Second, genetic heritability is used as the measure of association with an underlying genetic structure. Heritability represents the proportion of variability attributable to genetic factors and can be obtained in a variance component approach [Hopper, 2002]. This is a perfect fit to our situation since it does not require knowledge of specific culprit genes and allows the likelihood of multiple gene influences.

We develop an operational criterion of the proposal definition as follows. Since

$$\begin{aligned} f(P|G) &= \int f(P,E|G)\,\mathrm{d}E \\ &= \int f(P|E,G)f(E|G)\,\mathrm{d}E, \end{aligned}$$

it follows from $f(E|G) = f(E)$ in (1) that

$$f(P|G) = \int f(P|E,G)f(E)\,\mathrm{d}E.$$

If the condition

$$f(P|E,G) = f(P|E) \tag{2}$$

holds, then

$$f(P|G) = \int f(P|E)f(E)\,\mathrm{d}E = f(P).$$

In pursuit of a feasible approach, we thus take (2) as an operational criterion for the proposed endophenotype definition together with the variance component model as a vehicle for the heritability. It then requires heritability of the phenotype becomes null, conditioning on the putative endophenotype; and it implies genetic heritability of the phenotype is captured by the endophenotype.

Given a phenotype of continuous measurements, significance of (2) can be judged through the following variance component model for quantitative traits [Burton and Tobin, 2003]:

$$\begin{aligned} P_{ij} &= \alpha + \gamma E_{ij} + \tau Z_{ij} + G_{ij} + \varepsilon_{ij}, \\ \varepsilon_{ij} &\sim \mathrm{Normal}(0, \sigma_{\mathrm{R}}^2), \\ G_{ij} &\sim \mathrm{Normal}(0, [\sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{D}}^2 + \sigma_{\mathrm{C}}^2]), \\ \mathrm{Cov}(G_{ij}, G_{ik}) &= 2\phi_{ij,ik}\sigma_{\mathrm{A}}^2 + \Delta_{ij,ik}\sigma_{\mathrm{D}}^2 + \lambda_{ij,ik}\sigma_{\mathrm{C}}^2, \quad j \neq k, \end{aligned} \tag{3}$$

where P_{ij} is the observed phenotype of the jth member in the ith family, E_{ij} is his/her corresponding specified endophenotype, Z_{ij} is his/her other covariates, G_{ij} is the random effect for the underlying genetic structure, and ε_{ij} is the residual error term representing the effect of non-family factors. The components σ_{A}^2, σ_{D}^2 and σ_{C}^2 represent the variance arising from polygenic additive effects, polygenic dominance effects, and shared environmental effects, respectively. For two members j and k in the ith family, $\phi_{ij,ik}$ is the probability of randomly drawing a single allele in j that is identical by descent to a single allele at the same locus randomly drawing from k, $\Delta_{ij,ik}$ is the probability that both alleles at a locus are shared identical by descent by both members, and $\lambda_{ij,ik}$ is the binary indicator showing both members live together ($=1$) or apart ($=0$). The (broad sense) heritability of P_{ij}, conditional on E_{ij} is

$$h_{P|E} = \frac{\sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{D}}^2}{\sigma_{\mathrm{A}}^2 + \sigma_{\mathrm{D}}^2 + \sigma_{\mathrm{C}}^2 + \sigma_{\mathrm{R}}^2}.$$

A statistically significant result of rejecting the null hypothesis $h_{P|E} = 0$ gives an evidence against (2).

For a discrete phenotype of ordinal scale, the liability threshold model can be used in the preceding variance component setting [Falconer, 1989; Duggirala et al., 1997]. The model postulates the existence of an unobserved continuous trait (i.e., liability L_{ij}), and a set of thresholds $t_1, t_2, \ldots, t_{K-1}$ that partition the liability distribution into intervals corresponding to distinct phenotypic states:

$$P_{ij} = \begin{cases} 1 & \text{if } L_{ij} < t_1, \\ 2 & \text{if } t_1 \le L_{ij} < t_2, \\ \vdots & \vdots \\ K & \text{if } t_{K-1} \le L_{ij}. \end{cases}$$

The liability L_{ij} is then assumed to follow the same distribution as the P_{ij} in model (3) and the heritability can be obtained based on the liability.

If the endophenotype mediates all effects of the genotype on the phenotype, $h_{P|E} = 0$ can imply (2). If the genotype has a direct effect on the phenotype that is not mediated through the endophenotype, $h_{P|E} = 0$ might be difficult to be satisfied in practice. Similar to validating surrogate endpoints, various indices can be used. Motivated by the PTE in Freedman et al. [1992], we define the proportion of heritability explained (PHE) by the endophenotype as

$$\mathrm{PHE} = 1 - \frac{h_{P|E}}{h_P},$$

where h_P is the heritability calculated from the variance component model (3) by excluding the endophenotype E as one covariate. A good endophenotype explains a large proportion of heritability; thus, the larger the PHE value, the more likely E is an endophenotype.

INFERENCES

ESTIMATION

Parameters in variance component model (3) can be estimated by applying generalized estimating equations approach [Amos, 1994; Almasy and Blangero, 1998]. The computer package SOLAR (Sequential Oligogenic Linkage Analysis Routines) [Blangero et al., 2004] can be used to perform the variance component analysis with (3). As a result, a nature estimate of PHE is given by $\widehat{\mathrm{PHE}} = 1 - (\hat{h}_{P|E} / \hat{h}_P)$, where $\hat{h}_{P|E}$ and $\hat{h}_P$ are estimates of $h_{P|E}$ and h_P, obtaining from the SOLAR.

To derive the variance of $\widehat{\text{PHE}}$, we first reparameterize variance components $(\sigma_A^2, \sigma_D^2, \sigma_C^2, \sigma_R^2)$ as

$$h_1 = \frac{\sigma_A^2}{\sigma_A^2+\sigma_D^2+\sigma_C^2+\sigma_R^2},$$
$$h_2 = \frac{\sigma_D^2}{\sigma_A^2+\sigma_D^2+\sigma_C^2+\sigma_R^2},$$
$$h_3 = \frac{\sigma_C^2}{\sigma_A^2+\sigma_D^2+\sigma_C^2+\sigma_R^2},$$
$$h_4 = \sigma_A^2+\sigma_D^2+\sigma_C^2+\sigma_R^2.$$

Let $h_r^{(1)}$s and $h_r^{(2)}$s be the corresponding items from the variance component model (3) with and without adjusting for E_{ij}, respectively. Notice that $h_{P|E} = h_1^{(1)} + h_2^{(1)}$ and $h_P = h_1^{(2)} + h_2^{(2)}$. The first-order Taylor approximation about $(E(\hat{h}_{P|E}), E(\hat{h}_P))$ [as described in Casella and Berger, 2001] gives

$$\begin{aligned}\widehat{\text{var}(\text{PHE})} \approx & \frac{1}{E^2(\hat{h}_P)}\{\text{var}(\hat{h}_1^{(1)}) + \text{var}(\hat{h}_2^{(1)}) \\ & + 2\text{cov}(\hat{h}_1^{(1)}, \hat{h}_2^{(1)})\} + \frac{E^2(\hat{h}_{P|E})}{E^4(\hat{h}_P)} \\ & \times \{\text{var}(\hat{h}_1^{(2)}) + \text{var}(\hat{h}_2^{(2)}) + 2\text{cov}(\hat{h}_1^{(2)}, \hat{h}_2^{(2)})\} \\ & - 2\frac{E(\hat{h}_{P|E})}{E^3(\hat{h}_P)}\{\text{cov}(\hat{h}_1^{(1)}, \hat{h}_1^{(2)}) + \text{cov}(\hat{h}_1^{(1)}, \hat{h}_2^{(2)}) \\ & + \text{cov}(\hat{h}_2^{(1)}, \hat{h}_1^{(2)}) + \text{cov}(\hat{h}_2^{(1)}, \hat{h}_2^{(2)})\}. \end{aligned} \quad (4)$$

To obtain approximate covariances of $\hat{h}_r^{(1)}$s and $\hat{h}_r^{(2)}$s, we apply generalized estimating equations for covariances [Prentice and Zhao, 1991] to the variance component analysis (3). Hence, the robust covariance is given by

$$\begin{aligned}&\text{cov}(\hat{h}_r^{(t)}, \hat{h}_{r^*}^{(t^*)}) \\ &\approx \left\{\sum_{i=1}^{I}\left[\left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right)^{\mathrm{T}}(\mathbf{W}_i^{(t)})^{-1}\left(\frac{\partial \mathbf{W}_i^{(t)}}{\partial h_r^{(t)}}\right)\right.\right. \\ &\quad \times (\mathbf{W}_i^{(t)})^{-1}(\mathbf{S}_i^{(t)} - \mathbf{V}_i^{(t)}) \\ &\quad \left.\left. + \left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right)^{\mathrm{T}}(\mathbf{W}_i^{(t)})^{-1}\left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right)\right]\right\}^{-1} \\ &\quad \times \left\{\sum_{i=1}^{I}\left[\left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right)^{\mathrm{T}}(\mathbf{W}_i^{(t)})^{-1}(\mathbf{S}_i^{(t)} - \mathbf{V}_i^{(t)})\right.\right. \\ &\quad \left.\left. \times (\mathbf{S}_i^{(t^*)} - \mathbf{V}_i^{(t^*)})^{\mathrm{T}}(\mathbf{W}_i^{(t^*)})^{-1}\left(\frac{\partial \mathbf{V}_i^{(t^*)}}{\partial h_{r^*}^{(t^*)}}\right)\right]\right\} \\ &\quad \times \left\{\sum_{i=1}^{I}\left[\left(\frac{\partial \mathbf{V}_i^{(t^*)}}{\partial h_{r^*}^{(t^*)}}\right)^{\mathrm{T}}(\mathbf{W}_i^{(t^*)})^{-1}\left(\frac{\partial \mathbf{W}_i^{(t^*)}}{\partial h_{r^*}^{(t^*)}}\right)\right.\right. \\ &\quad \times (\mathbf{W}_i^{(t^*)})^{-1}(\mathbf{S}_i^{(t^*)} - \mathbf{V}_i^{(t^*)}) \\ &\quad \left.\left. + \left(\frac{\partial \mathbf{V}_i^{(t^*)}}{\partial h_{r^*}^{(t^*)}}\right)^{\mathrm{T}}(\mathbf{W}_i^{(t^*)})^{-1}\left(\frac{\partial \mathbf{V}_i^{(t^*)}}{\partial h_{r^*}^{(t^*)}}\right)\right]\right\}^{-1}, \end{aligned} \quad (5)$$

TABLE I. The covariance components for relative pairs

Relationship	Covariance in σ_i^2s[a]	Covariance in h_is
Same person	$\sigma_A^2+\sigma_D^2+\lambda\sigma_C^2+\sigma_R^2$	$h_4-(1-\lambda)h_3h_4$
Parent-child	$\frac{1}{2}\sigma_A^2+\lambda\sigma_C^2$	$\frac{1}{2}h_1h_4+\lambda h_3h_4$
Full sibling	$\frac{1}{2}\sigma_A^2+\frac{1}{4}\sigma_D^2+\lambda\sigma_C^2$	$\frac{1}{2}h_1h_4+\frac{1}{4}h_2h_4+\lambda h_3h_4$
Half sibling	$\frac{1}{4}\sigma_A^2+\lambda\sigma_C^2$	$\frac{1}{4}h_1h_4+\lambda h_3h_4$
Monozygous twins	$\sigma_A^2+\sigma_D^2+\lambda\sigma_C^2$	$h_1h_4+h_2h_4+\lambda h_3h_4$
Grandparent-grandchild	$\frac{1}{4}\sigma_A^2+\lambda\sigma_C^2$	$\frac{1}{4}h_1h_4+\lambda h_3h_4$
Uncle/aunt-nephew/niece	$\frac{1}{4}\sigma_A^2+\lambda\sigma_C^2$	$\frac{1}{4}h_1h_4+\lambda h_3h_4$
First cousins	$\frac{1}{8}\sigma_A^2+\lambda\sigma_C^2$	$\frac{1}{8}h_1h_4+\lambda h_3h_4$
Double first cousins	$\frac{1}{4}\sigma_A^2+\frac{1}{16}\sigma_D^2+\lambda\sigma_C^2$	$\frac{1}{4}h_1h_4+\frac{1}{16}h_2h_4+\lambda h_3h_4$
Spouses	$\lambda\sigma_C^2$	λh_3h_4

[a]λ is a binary indicator denoting whether two individuals live together ($\lambda = 1$) or apart ($\lambda = 0$).

where $t, t^* = 1, 2, r, r^* = 1, 2, 3, 4$; by assuming there are I families and n_i members in the ith family, $\mathbf{S}_i^{(t)} = (r_{i1}^{(t)}r_{i1}^{(t)}, \ldots, r_{i1}^{(t)}r_{in_i}^{(t)}, r_{i2}^{(t)}r_{i2}^{(t)}, \ldots, r_{i2}^{(t)}r_{in_i}^{(t)}, \ldots, r_{in_i}^{(t)}r_{in_i}^{(t)})^{\mathrm{T}}$, $r_{ij}^{(1)} = P_{ij} - (\alpha^{(1)} + \gamma^{(1)}E_{ij} + \tau^{(1)}Z_{ij})$, $r_{ij}^{(2)} = P_{ij} - (\alpha^{(2)} + \tau^{(2)}Z_{ij})$, $\mathbf{V}_i^{(t)} = E(\mathbf{S}_i^{(t)})$ is the variance-covariance matrix for phenotypes among members of the ith family in vector form with its components for selected relative pairs given in Table I, and $\mathbf{W}_i^{(t)}$ is the "working" variance-covariance matrix for $\mathbf{S}_i^{(t)}$ with the component for the (j,k)th and (l,m)th pairs being

$$E(r_{ij}^{(t)}r_{il}^{(t)})E(r_{ik}^{(t)}r_{im}^{(t)}) + E(r_{ij}^{(t)}r_{im}^{(t)})E(r_{ik}^{(t)}r_{il}^{(t)})$$

under the Gaussian working model [Prentice and Zhao, 1991]. Details for obtaining (5) are outlined in Appendix. We may estimate $\text{var}(\widehat{\text{PHE}})$ by replacing $E(\hat{h}_{P|E})$ with $\hat{h}_{P|E}$, $E(\hat{h}_P)$ with $\hat{h}_P$, and $\text{cov}(\hat{h}_r^{(t)}, \hat{h}_{r^*}^{(t^*)})$ with its estimate.

HYPOTHESIS TESTING

To evaluate the importance of the endophenotype, we perform the following one-sided test:

$$\begin{cases} H_0 : \text{PHE} = a, \\ H_1 : \text{PHE} > a, \end{cases}$$

where a is some pre-specified critical value. Under the significance level α, we reject H_0 if the lower bound of the one-sided confidence interval of PHE,

$$\widehat{\text{PHE}} - z_{1-\alpha} \times \sqrt{\widehat{\text{var}}(\widehat{\text{PHE}})},$$

is greater than a, where $z_{1-\alpha}$ is the lower $100 \times (1-\alpha)$th percentile of the standard normal distribution.

In the following simulation study, we consider some different values of a and evaluate appropriateness of the normality assumption, hopefully, to construct useful criteria for validating endophenotypes.

SIMULATION STUDIES

STUDY DESIGN

The simulation studies evaluate the utility of the proposed index, PHE, under two different scenarios (Fig. 2). In scenario I, the endophenotype (*E*) and the phenotype (*P*) are influenced by a single disease gene (*G*). Scenario II postulates a more complex, but more practical, situation where *E* and *G* share effects from the disease gene *G*1, but genes *G*2 and *G*3 only have effects on *E* and *P*, respectively

Our simulations assumed both *E* and *P* to be continuous measurements. The quantitative trait y and its influential genes were assumed to have a linear relation as described in Almasy and Blangero [1998]:

$$y = \mu + \sum_{i=1}^{n} \eta_i + \varepsilon, \tag{6}$$

where μ is the grand mean, η_i is the random effect of the ith disease gene, and ε represents a random non-family deviation. Both η_i and ε are assumed to be normally distributed and uncorrelated. In these simulations, dominance effects and shared environmental effects were not included, and therefore $\text{var}(\eta_i) = \sigma^2_{A_i}$ (the variance from the additive effects of disease gene *i*). To construct scenario I, two quantitative traits *E* and *P* controlled by the same gene *G* were simulated. Each of *E* and *P* had the single-gene contribution from *G* through the linear model (6). The non-family deviation of *E* (ε_E) and the non-family deviation of *P* (ε_P) were assumed to have a correlation ρ_ε. The multiple-gene effects in scenario II included the action of gene *G*1 on *E* and *P*, the action of *G*2 on *E* and the action of *G*3 on *P*.

The simulated data contained 200 nuclear families, each with two siblings. The disease genes were assumed to be bi-allelic and the population frequency of the common allele was 0.9. In scenario I, the heritability of *P* due to *G* ($h_P(G)$) was assumed to be either 0.15 or 0.42 and the heritability of *E* due to *G* ($h_E(G)$) allowed being 0, 0.15, 0.42, or 0.74. The correlation between non-family deviations of *E* and *P* (ρ_ε) was 0 or 0.5. In scenario II, the parameter values can be found in the following result tables. The correlation between non-family deviations was assumed to be the same as scenario I. In the simulation, we fixed the heritability of *P* on 0.15 and 0.42, to imply that the phenotype had shown weak and strong genetic underpinning, respectively. Various heritabilities of *E* mimicked unknown relationship among the phenotype, endophenotype, and genotype. Notice that, in the situation where the endophenotype's heritability due to shared genes is higher than the phenotype's heritability due to shared genes (i.e., $h_E(G) > h_P(G)$ in scenario I and $h_E(G1) > h_P(G1)$ in scenario II), the linkage or association studies based on the endophenotype are more likely to detect these common genes than the studies based on the phenotype (as shown in Almasy and Blangero [1998] and Williams et al. [1999]). When evaluating the power of the proposed hypothesis testing, the significant level $\alpha = 0.05$ and the critical value was set to be 0, 0.25, 0.5, or 0.75. Two hundred replications were performed for each specified situation.

The command "simqtl" of the computer package SOLAR was used to simulate the data in two scenarios. The variance component analysis (3) was performed using the SOLAR command "polymod". Estimates of the standard error of $\widehat{\text{PHE}}$ were calculated by using R software.

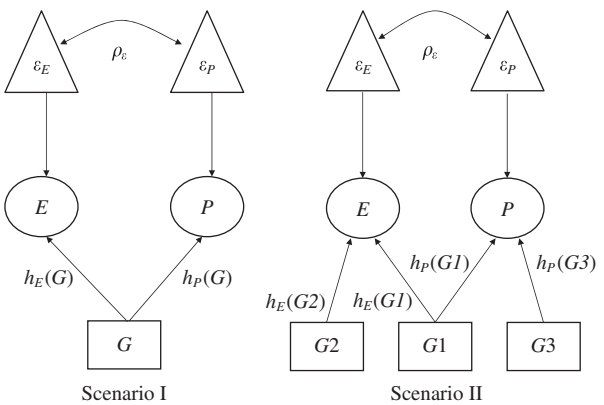

Fig. 2. Two scenarios considered in the simulation studies: endophenotype (*E*), phenotype (*P*), underlying disease genes (*G*, *G*1, *G*2 and *G*3), random non-family effects (ε_E and ε_P), $h_E(G')$ means the heritability of *E* due to G', $h_P(G')$ means the heritability of *P* due to G', and the correlation between non-family effects (ρ_ε), where G' may be *G*, *G*1, *G*2 or *G*3.

RESULTS

Table II contains results under scenario I. When the heritability of *P* due to *G* was fixed, the higher the $h_E(G)$, the lower the $h_{P|E}$ and the closer the PHE values to 1. No matter what the correlation between non-family deviations of *E* and *P* was (either 0 or 0.5), the trend was always retained. Also, the standard errors of $\widehat{\text{PHE}}$ for the phenotype showing weak genetic underpinning ($h_P(G) = 0.15$) were larger than the standard errors for

TABLE II. Simulation results for parameter estimation based on scenario I

$h_P(G)$	$h_E(G)$	ρ_ε	$h_{P\|E}$	PHE	ESE[a]	CSE[b]
Weak genetic underpinning for the phenotype						
0.15	0	0	0.146	−0.002	0.066	0.227
		0.5	0.188	−0.643	1.871	3.851
	0.15	0	0.110	0.307	0.219	0.478
		0.5	0.053	0.675	0.314	0.351
	0.42	0	0.059	0.672	0.239	0.415
		0.5	0.024	0.859	0.202	0.357
	0.74	0	0.029	0.854	0.170	0.377
		0.5	0.031	0.814	0.254	0.389
Strong genetic underpinning for the phenotype						
0.42	0	0	0.405	−0.002	0.009	0.025
		0.5	0.473	−0.201	0.138	0.215
	0.15	0	0.337	0.202	0.079	0.128
		0.5	0.269	0.322	0.158	0.151
	0.42	0	0.183	0.562	0.138	0.107
		0.5	0.075	0.816	0.149	0.087
	0.74	0	0.053	0.875	0.125	0.084
		0.5	0.028	0.937	0.093	0.075

[a]ESE = empirical standard error of PHE based on simulation replications.

[b]CSE = mean of calculated standard errors of PHE.

the phenotype showing strong genetic underpinning ($h_P(G) = 0.42$). To check the accuracy of standard error estimates of $\widehat{\text{PHE}}$, we compared the mean of estimated standard errors of $\widehat{\text{PHE}}$ based on (4) and (5) with the empirical standard error of $\widehat{\text{PHE}}$ based on simulation replications. When $h_P(G) = 0.15$, the proposed method can overestimate the standard errors. This overestimation may be partially due to the poor fit of the first-order Taylor approximation for the variance of $\widehat{\text{PHE}}$ in equation (4). It has been shown that the first-order Taylor approximation can overestimate the variance of a ratio of two random variables when the expectation of the variable in the denominator is small [Herson, 1975]. When $h_P(G) = 0.42$, we may overestimate or underestimate the standard errors; nevertheless, both absolute errors of the overestimates and the underestimates were small.

Table III shows the results when there existed multiple disease genes (scenario II). The higher the heritability of E due to $G1$ (the shared gene), the larger the PHE values, which was consistent with scenario I. However, we can find that the heritabilities due to non-shared genes ($h_P(G3)$ and $h_E(G2)$) also influenced the PHE values. The higher the $h_P(G3)$ or $h_E(G2)$, the smaller the PHE values. Besides, if there existed a correlation between non-family deviations of E and P ($\rho_\varepsilon = 0.5$), the PHE values were disrupted. For standard error estimates of $\widehat{\text{PHE}}$, a pattern similar to scenario I was observed.

We used the Shapiro-Wilk test to evaluate the appropriateness of normality assumption in establishing a confidence interval of PHE. One can find that the normality assumption of PHE estimates did not hold in most situations under scenario I (Table IV). Under scenario II, the normality assumption showed better fit in situations where $h_E(G1) \geq 0.42$ (Table V). We further examined histograms of PHE estimates under situations where the normality assumption did not fit. It seemed that these distributions of $\widehat{\text{PHE}}$ were usually left skewed (i.e., the left tail of the distribution was longer). Therefore, the use of standard normal percentile $z_{1-\alpha}$ in constructing the confidence interval of PHE can result in a lower bound smaller than what it should be. As a result, the proposed hypothesis testing can be conservative in rejecting $H_0 : \text{PHE} = a$ when the normality assumption did not fit.

Table IV shows the proportions of simulation replicates that rejected the hypothesis $H_0 : \text{PHE} = a$ and preferred $H_1 : \text{PHE} > a$, under various conditions of scenario I. When the single disease gene G did not have effects on the putative endophenotype (i.e., $h_E(G) = 0$), all replications accepted H_0 under various as. When $h_E(G) \neq 0$, the rejection proportion increased dramatically for the critical value $a = 0$. In situations where $h_P(G) = 0.42$ and $h_E(G) \geq h_P(G)$, $a = 0.25$ can still lead to rejection proportions > 80%. When there existed multiple disease genes, the proposed testing was especially useful in rejecting the critical value $a = 0$ if $h_P(G1) = 0.42$ and $h_E(G1) \geq h_P(G1)$

TABLE III. Simulation results for parameter estimation based on scenario II

$h_P(G1)$/ $h_E(G1)$	$h_P(G3)$/ $h_E(G2)$	ρ_ε	$h_{P\|E}$	PHE	ESE[a]	CSE[b]
Weak genetic underpinning for the phenotype						
0.15/0	0.25/0.3	0	0.397	−0.00005	0.009	0.026
		0.5	0.482	−0.226	0.112	0.198
	0.25/0.7	0	0.396	0.001	0.009	0.028
		0.5	0.458	−0.162	0.072	0.165
0.15/0.15	0.25/0.25	0	0.371	0.064	0.040	0.102
		0.5	0.429	−0.092	0.121	0.198
	0.25/0.59	0	0.375	0.054	0.041	0.109
		0.5	0.437	−0.110	0.094	0.191
0.15/0.42	0.25/0.12	0	0.330	0.171	0.067	0.132
		0.5	0.382	0.028	0.127	0.185
	0.25/0.23	0	0.334	0.159	0.067	0.134
		0.5	0.398	−0.012	0.120	0.189
0.15/0.74	0.25/0.08	0	0.298	0.251	0.085	0.137
		0.5	0.359	0.088	0.121	0.174
	0.25/0.13	0	0.303	0.237	0.084	0.138
		0.5	0.357	0.092	0.115	0.172
Strong genetic underpinning for the phenotype						
0.42/0	0.17/0.3	0	0.580	−0.001	0.005	0.012
		0.5	0.653	−0.138	0.065	0.101
	0.17/0.7	0	0.582	−0.0001	0.006	0.012
		0.5	0.639	−0.096	0.047	0.079
0.42/0.15	0.17/0.25	0	0.530	0.093	0.040	0.077
		0.5	0.581	−0.004	0.095	0.113
	0.17/0.59	0	0.536	0.073	0.041	0.082
		0.5	0.613	−0.049	0.074	0.109
0.42/0.42	0.17/0.12	0	0.424	0.273	0.087	0.089
		0.5	0.463	0.193	0.112	0.105
	0.17/0.23	0	0.434	0.243	0.074	0.093
		0.5	0.512	0.132	0.106	0.105
0.42/0.74	0.41/0.08	0	0.682	0.174	0.069	0.053
		0.5	0.769	0.057	0.072	0.057
	0.17/0.08	0	0.329	0.426	0.109	0.086
		0.5	0.408	0.294	0.136	0.097

[a]ESE = empirical standard error of PHE based on simulation replications.
[b]CSE = mean of calculated standard errors of PHE.

TABLE IV. Simulation results for hypothesis testing based on scenario I

$h_P(G)$	$h_E(G)$	ρ_ε	$P(0)$[a]	$P(0.25)$	$P(0.5)$	$P(0.75)$	S.W. P-value[b]
Weak genetic underpinning for the phenotype							
0.15	0	0	0	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
	0.15	0	0.08	0.005	0	0	<0.001
		0.5	0.735	0.53	0.285	0.025	<0.001
	0.42	0	0.775	0.385	0.075	0	<0.001
		0.5	0.865	0.78	0.5	0.075	<0.001
	0.74	0	0.925	0.765	0.38	0.005	<0.001
		0.5	0.835	0.735	0.435	0.04	<0.001
Strong genetic underpinning for the phenotype							
0.42	0	0	0	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
	0.15	0	0.55	0	0	0	<0.001
		0.5	0.715	0.195	0.01	0	0.039
	0.42	0	0.99	0.815	0.255	0	0.698
		0.5	0.995	0.98	0.825	0.365	<0.001
	0.74	0	1	1	0.945	0.52	<0.001
		0.5	1	1	0.99	0.78	<0.001

[a]$P(a)$ = proportion of simulation replicates whose lower bound of the one-sided 95% confidence interval of PHE is greater than the critical value a.
[b]S.W. P-value = mean of P-values for Shapiro-Wilk test of normality.

TABLE V. Simulation results for hypothesis testing based on scenario II

$h_P(G1)/h_E(G1)$	$h_P(G3)/h_E(G2)$	ρ_ε	$P(0)$[a]	$P(0.25)$	$P(0.5)$	$P(0.75)$	S.W. P-value[b]
Weak genetic underpinning for the phenotype							
0.15/0	0.25/0.3	0	0	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
	0.25/0.7	0	0	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
0.15/0.15	0.25/0.25	0	0.01	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
	0.25/0.59	0	0.005	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
0.15/0.42	0.25/0.12	0	0.26	0	0	0	0.015
		0.5	0.045	0	0	0	0.162
	0.25/0.23	0	0.215	0	0	0	0.018
		0.5	0.015	0	0	0	0.030
0.15/0.74	0.25/0.08	0	0.62	0.02	0	0	0.080
		0.5	0.12	0	0	0	0.087
	0.25/0.13	0	0.56	0.01	0	0	0.080
		0.5	0.11	0	0	0	0.108
Strong genetic underpinning for the phenotype							
0.42/0	0.17/0.3	0	0	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
	0.17/0.7	0	0	0	0	0	<0.001
		0.5	0	0	0	0	<0.001
0.42/0.15	0.17/0.25	0	0.275	0	0	0	0.3
		0.5	0.04	0	0	0	<0.001
	0.17/0.59	0	0.12	0	0	0	<0.001
		0.5	0	0	0	0	0.016
0.42/0.42	0.17/0.12	0	0.9	0.09	0	0	0.004
		0.5	0.585	0.025	0	0	0.047
	0.17/0.23	0	0.845	0.015	0	0	0.555
		0.5	0.415	0.02	0	0	<0.001
0.42/0.74	0.41/0.08	0	0.87	0.01	0	0	0.777
		0.5	0.35	0	0	0	0.02
	0.17/0.08	0	0.97	0.645	0.04	0	0.002
		0.5	0.805	0.225	0.015	0	0.07

[a]$P(a)$ = proportion of simulation replicates whose lower bound of the one-sided 95% confidence interval of PHE is greater than the critical value a.
[b]S.W. P-value = mean of P-values for Shapiro-Wilk test of normality.

TABLE VI. Variance component analysis for the schizophrenia spectrum and the CPT

Response variable	Schizophrenia spectrum			Undegraded d'		Degraded d'	
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Covariate[a]							
Female	0.01 (0.21)	0.01 (0.22)	0.03 (0.21)	−0.18 (0.17)	−0.08 (0.12)	−0.04 (0.20)	0.08 (0.12)
Age	−0.003 (0.008)	−0.003 (0.008)	−0.0003 (0.008)	−0.009 (0.006)	0.005 (0.004)	−0.03 (0.009)	−0.01 (0.005)
Education years	−0.05 (0.03)	−0.03 (0.03)	−0.02 (0.03)	0.12 (0.02)	0.05 (0.01)	0.10 (0.03)	−0.003 (0.02)
Undegraded d'	–	−0.11 (0.07)	–	–	–	–	0.91 (0.06)
Degraded d'	–	–	−0.12 (0.06)	–	0.66 (0.03)	–	–
Index							
Heritability	0.34 (0.32)	0.38 (0.36)	0.49 (0.35)	0.42 (0.09)	0.01 (0.09)	0.86 (0.29)	0.25 (0.36)
PHE	–	−0.14 (0.15)	−0.45 (0.40)	–	0.97 (0.11)	–	0.71 (0.36)

[a]The estimated value is for the regression coefficient in model (3).

(Table V). In both scenarios, the $P(a)$ values (i.e., the proportions of rejecting H_0) for $h_P(G)$ (or $h_P(G1)$) = 0.15 were smaller than the values for $h_P(G)$ (or $h_P(G1)$) = 0.42. This can be caused by the larger $\widehat{\text{PHE}}$ standard errors and the likelihood to overestimate the $\widehat{\text{PHE}}$ standard errors when $h_P(G)$ (or $h_P(G1)$) = 0.15, as seen in Tables II and III.

To evaluate the sample size effect on our estimation, we had performed the simulation with the number of families

equalling to 500 (data not shown). For both scenarios, the large sample size did not seem to affect the estimated PHE values, but can reduce the standard errors of the estimates. The larger the sample size, the better the normality assumption for the distribution of $\widehat{\text{PHE}}$. Also, the large sample size can increase the rejection proportion in Tables IV and V.

CONCLUSION

In summary, the PHE index reflects the heritability of the putative endophenotype due to the genes shared with the phenotype of interest. The genetic effects that are not shared between the phenotype and the endophenotype can adversely reduce the PHE values. Our proposed $\widehat{\text{PHE}}$ standard error estimate is reasonably accurate when the phenotype shows sensible genetic underpinning, but the proposed method can overestimate the standard error when the phenotype is less heritable.

Rejecting $H_0 : \text{PHE} = 0$ implies that parts of the genes underlying the putative endophenotype can also affect the phenotype and, thus, the putative endophenotype is useful in detecting the disease genes. Rejecting $H_0 : \text{PHE} = 0.25$ reveals strong evidences that the heritability of the endophenotype due to the genes shared with the phenotype is higher than the heritability of the phenotype due to these common genes and, thus, the endophenotype has greater likelihood of detecting disease genes than the phenotype. When the phenotype is less heritable and the sample size is not large enough, the proposed method for hypothesis testing tends to be conservative in rejecting H_0.

EXAMPLE

The Continuous Performance Test (CPT) [Rosvold et al., 1956] has been widely used to measure sustained attention deficits in psychotic disorders, and has become an increasingly important focus in the search for potential endophenotypes for genetic susceptibility to schizophrenia [Chen and Faraone, 2000; Cornblatt and Keilp, 1994]. In the following analyses, we will use the proposed methodology to assess the CPT as an endophenotype of schizophrenia based on the data from a family genetic study of schizophrenia in Northern Taiwan. Details of study design and eligibility criteria were described previously [Chen et al., 1998b, 2004; Chang et al., 2002]. Briefly, schizophrenic patients fulfilling some inclusion criteria were included as the probands of the present family genetic study. All probands' first-degree relatives aged 16 years or older were also included in this family study. In total, 91 schizophrenic probands and 231 first-degree relatives were included in this analysis.

The phenotype examined in the following analyses was the schizophrenia spectrum, including schizophrenia and schizophrenia-related personality disorders as well; the latter was defined as having a definite or possible diagnosis of schizotypal, schizoid or paranoid personality disorder [Chang et al., 2002]. Two different CPT scores were used as candidate endophenotypes: the index of sensitivity d' on the undegraded test and d' on the 25% degraded test [Chen et al., 1998a]. Variance component analysis (3) was then performed using the SOLAR computer package. More specifically, we fit a liability threshold model for discrete trait model with additive and dominance effects [Duggirala et al., 1997]. The shared environmental effect was not included because the current nuclear-family data could not appropriately distinguish this effect from genetic effects; thus the model with both environmental and genetic effects might result in over fitting [Hopper and Visscher, 2002]. Since our proband ascertainment was based on his/her schizophrenia status, the ascertainment correction provided in the SOLAR was employed. Covariates gender, age, and education years were also included in the models to control for possible confounding.

Results of variance component analyses for the schizophrenia spectrum on the CPT scores are shown in Table VI. Although the heritability for the schizophrenia spectrum was 0.34, it was not significantly different from zero. Heritabilities for both the undegraded and degraded d' were high and significantly different from zero. PHEs for evaluating various CPT scores as endophenotypes of the schizophrenia spectrum were all not significantly greater than zero. This indicated that the CPT and the schizophrenia spectrum, although both were highly heritable, might not share common causal genes. The PHE for evaluating the degraded d' as an endophenotype of the undegraded d' was greater than the PHE for evaluating the undegraded d' as an endophenotype of the degraded d', and both PHEs were very significant. These results showed evidence that the degraded d' was a more heritable indicator, which mediated almost all effects of the genotype on the undegraded d'. This is consistent with previous studies in supporting the proposition that the more difficult versions of the CPT are stable vulnerability indicators, whereas the simpler versions might be more state-dependent [Chen et al., 2004].

Our analyses in assessing various CPT scores as endophenotypes of the schizophrenia spectrum covering a variety of personality disorders did not reveal significant results. To some extent, the results are not surprising. A previous analysis indicated that schizophrenia-related personality disorders, though part of schizophrenia spectrum, do not necessarily lead to a higher statistical power for future genetic analysis: the value of recurrence risk ratio by incorporating these personality disorders in the spectrum became lower than the corresponding figure for schizophrenia itself [Chang et al., 2002]. We might also be handicapped by the data structure (nuclear families only) and/or the sample size (91 schizophrenia families) of our study. Nevertheless, these results do not rule out the possibility that CPT performance scores might be endophenotypes for other schizophrenia-related traits, such as certain symptom dimensions or neurocognitive functions.

DISCUSSION

In this report, we have attempted to provide a criterion for using to validate an endophenotype. This criterion is initially motivated by the operational criteria for validating surrogate endpoints, but these two sets of criteria have distinct implications. Endophenotypes need to be less removed from relevant genes than phenotypes and thus provide greater power for genetic studies, and genotype information is usually unknown when validating endophenotypes. Simulation results show that the proposed index is useful in validating endophenotypes. Comparing with the criteria proposed by Gottesman and Gould [2003],

our proposal can verify a postulated biological pathway from genotypes to phenotypes via endophenotypes directly and provide a formal statistical evaluation of the significance of the relationship.

To obtain the heritability in validating endophenotypes, we proposed to fit a liability threshold model for discrete phenotypes. One possible alternative is to use generalized linear mixed models (GLMMs) [Breslow and Clayton, 1993; Burton and Tobin, 2003]. When a phenotype is continuous, GLMMs are the typical variance component analysis of quantitative traits. For ordinal phenotypes (traits), GLMMs can be fit by selecting appropriate link functions. Although GLMMs are flexible, there are two major challenges in implementing them. First, difficulty arises when attempting to find an analogue to σ_R^2 in calculating the heritability. Consequently, it is inappropriate to adopt a heuristic approach, for example: to map the raw residuals ($P_{ij} - E(P_{ij})$) to the scale of link function and then to treat the mean square as an analogue to σ_R^2 [Breslow and Clayton, 1993; Burton et al., 1999]. Second, readily available software for fitting GLMMs (e.g., R package: lme4) is not intended for genetic analyses, thus the within-family covariance matrix can only be updated as a composite whole; various genetic variance components in model (3) cannot be obtained. Burton et al. [1999] had attempted to reparameterize GLMMs to obtain various genetic variance components for studies with nuclear families only. More efforts need to be put on modifying available software to allow specification of variance components σ_A^2, σ_D^2 and σ_C^2, or on reparameterizing GLMMs to accommodate general pedigrees of arbitrary size and complexity.

There are some possible extensions of the proposed method. The present study regards the genotype as an unmeasured random effect. If measured genotype or markers are available, one can modify equation (3) to incorporate the measured marker information as done in equation (1) of Almasy and Blangero [1998] to help detect candidate genes.

ACKNOWLEDGMENTS

This research was partially supported by grants from the National Science Council, Taiwan (NSC96-2118-M-009-001-MY2 to G.H.H., NSC96-3112-B-001-017 to C.H.C., and NSC97-3112-B-001-020 to G.H.H. and C.H.C.), the Academia Sinica Thematic Project (AS91ISS1PP), and the National Health Research Institute, Taiwan (DOH83-HR-306 to DOH87-HR-306, and NHRI-EX91-9113PP). The authors thank Dr. Kung-Yee Liang for his stimulating comments and helpful discussions.

REFERENCES

Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211.

Almasy L, Blangero J. 2001. Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. Am J Med Genet B Neuropsychiatr Genet 105:42–44.

Amos CI. 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54:535–543.

Begg C, Leung DHY. 2000. On the use of surrogate end points in randomized trials (with comments). J R Stat Soc A 163:15–28.

Blangero J, Lange K, Dyer T, Almasy L, Goring H, Williams J, Peterson C. 2004. SOLAR v.2.1.4. http://www.sfbr.org/solar/index.html. San Antonio: Southwest Foundation for Biomedical Research.

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. J Am Stat Assoc 88:9–25.

Burton PR, Tobin MD. 2003. Epidemiology and genetic epidemiology. In: Balding DJ, Bishop M, Cannings C, editors. Handbook of Statistical Genetics, 2nd edition. West Sussex: Wiley, p 855–879.

Burton PR, Tiller KJ, Gurrin LC, Cookson WOCM, Musk AW, Palmer LJ. 1999. Genetic variance components analysis for binary phenotypes using generalized linear mixed models GLMMs and Gibbs sampling. Genet Epidemiol 17:118–140.

Buyse M, Molenberghs G. 1998. Criteria for validation of surrogate endpoints in randomized experiments. Biometrics 54:1014–1029.

Casella G, Berger RL. 2001. Statistical Inference, 2nd edition. Belmont: Brooks/Cole, p 240–245.

Chang CJ, Chen WJ, Liu SK, Cheng JJ, Ou Yang WC, Chang HJ, Lane HY, Lin SK, Yang TW, Hwu HG. 2002. Morbidity risk of psychiatric disorders among the first degree relatives of schizophrenia patients in Taiwan. Schizophr Bull 28:379–392.

Chen WJ, Faraone SV. 2000. Sustained attention deficits as markers of genetic susceptibility to schizophrenia. Am J Med Genet 97:52–57.

Chen WJ, Hsiao CK, Hsiao LL, Hwu HG. 1998a. Performance of the continuous performance test among community samples. Schizophr Bull 24:163–174.

Chen WJ, Liu SK, Chang CJ, Lien YJ, Chang YH, Hwu HG. 1998b. Sustained attention deficit and schizotypal personality features in nonpsychotic relatives of schizophrenic patients. Am J Psychiatry 155:1214–1220.

Chen WJ, Chang CH, Liu SK, Hwang TJ, Hwu HG, collaborators from the Multidimensional Psychopathology Group Research Project. 2004. Sustained attention deficits in nonpsychotic relatives of schizophrenic patients: a recurrence risk ratio analysis. Biol Psychiatry 55:995–1000.

Cornblatt BA, Keilp JG. 1994. Impaired attention, genetics, and the pathophysiology of schizophrenia. Schizophr Bull 20:31–46.

De Gruttola VG, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedman L, Gail MH, Prentice R, Wittes J, Zeger SL. 2001. Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health Workshop. Control Clin Trials 22:485–502.

Duggirala R, Williams JT, Williams-Blangero S, Blangero J. 1997. A variance component approach to dichotomous trait linkage analysis using a threshold model. Genet Epidemiol 14:987–992.

Eaton WW, Dryman A, Sorenson A, McCutcheon A. 1989. DSM-III major depressive disorder in the community: a latent class analysis of data from the NIMH epidemiologic catchment-area program. Br J Psychiatry 155:48–54.

Falconer DS. 1989. Introduction to Quantitative Genetics, 3rd edition. New York: Wiley.

Freedman LS, Graubard BI, Schatzkin A. 1992. Statistical validation of intermediate endpoints for chronic diseases. Stat Med 11:167–178.

Gottesman II, Gould TD. 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 160:636–645.

Herson J. 1975. Fieller's theorem versus the delta method for significance intervals for ratios. J Stat Comput Sim 3:265–274.

Hopper JL. 2002. Heritability. In: Elston RC, Olson JM, Palmer L, editors. Biostatistical Genetics and Genetic Epidemiology. West Sussex: Wiley, p 371–372.

Hopper JL, Visscher PM. 2002. Variance component analysis. In: Elston RC, Olson JM, Palmer L, editors. Biostatistical Genetics and Genetic Epidemiology. West Sussex: Wiley, p 778–788.

Pan W-H, Lynn K-S, Chen C-H, Wu Y-L, Lin C-Y, Chang H-Y. 2006. Using endophenotypes for pathway clusters to map complex disease genes. Genet Epidemiol 30:143–154.

Piantadosi S. 1997. Clinical Trials: A Methodologic Perspective. New York: Wiley.
Prentice RL. 1989. Surrogate endpoints in clinical trials: Definitions and operational criteria. Stat Med 8:431–440.
Prentice RL, Zhao LP. 1991. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47:825–839.
Rosvold HE, Mirsky AF, Sarason I, Bransome ED, Beck LH. 1956. A continuous performance test of brain damage. J Consult Psychol 20:343–350.
Waldman ID. 2005. Statistical approaches to complex phenotypes: Evaluating neuropsychological endophenotypes for attention-deficit/hyperactivity disorder. Biol Psychiatry 57:1347–1356.
Williams JT, Eerdewegh PV, Almasy L, Blangero J. 1999. Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. Am J Hum Genet 65:1134–1147.

APPENDIX

DERIVATION OF $\mathrm{cov}(\hat{h}_r^{(t)}, \hat{h}_{r^*}^{(t^*)})$

As describing in Amos [1994], $h_r^{(t)}$ can be estimated by solving the generalized estimating equation for covariances [Prentice and Zhao, 1991]

$$U(h_r^{(t)}) = \sum_{i=1}^{I} \left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right)^{\mathrm{T}} (\mathbf{W}_i^{(t)})^{-1}(\mathbf{S}_i^{(t)} - \mathbf{V}_i^{(t)}) = 0,$$

where $\mathbf{S}_i^{(t)}$, $\mathbf{V}_i^{(t)}$, and $\mathbf{W}_i^{(t)} = \mathbf{W}_i^{(t)}(h_r^{(t)})$ are as defined in the text. Under regularity conditions, we can have

$$I^{1/2}(\hat{h}_r^{(t)} - h_r^{(t)}) = \left[-I^{-1}\left(\frac{\partial U(h_r^{(t)})}{\partial h_r^{(t)}}\right)\right]^{-1} \times [I^{-1/2}U(h_r^{(t)})],$$

where

$$\begin{aligned}\frac{\partial U(h_r^{(t)})}{\partial h_r^{(t)}} &= -\sum_{i=1}^{I}\left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right)^{\mathrm{T}}(\mathbf{W}_i^{(t)})^{-1} \\ &\quad \times \left(\frac{\partial \mathbf{W}_i^{(t)}}{\partial h_r^{(t)}}\right)(\mathbf{W}_i^{(t)})^{-1}(\mathbf{S}_i^{(t)} - \mathbf{V}_i^{(t)}) \\ &\quad - \sum_{i=1}^{I}\left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right)^{\mathrm{T}}(\mathbf{W}_i^{(t)})^{-1}\left(\frac{\partial \mathbf{V}_i^{(t)}}{\partial h_r^{(t)}}\right) \\ &= -\sum_{i=1}^{I} A_i - \sum_{i=1}^{I} B_i.\end{aligned}$$

Then, we essentially follow Appendix 1 of Prentice and Zhao [1991] to obtain (5), with an exception of assuming that $I^{-1}\sum_{i=1}^{I} A_i$ converges to a constant, not $o_p(1)$, as $I \to \infty$. This modification can improve the small-sample distributional approximation.