

The Power of the Classical Twin Study to Resolve Variation in Threshold Traits

Michael C. Neale,^{1,3} Lindon J. Eaves,² and Kenneth S. Kendler^{1,2}

Received 3 Dec. 1993—Final 15 Feb. 1994

We explore the power of the twin study to resolve sources of familial resemblance when the data are measured at the binary or ordinal level. Four components of variance were examined: additive genetic, nonadditive genetic, and common and specific environment. Curves are presented to compare the power of the continuous case with those of threshold models corresponding to different prevalences in the population: 1, 5, 10, 25, and 50%. Approximately three times the sample size is needed for equivalent power to the continuous case when the threshold is at the optimal 50%, and this ratio increases to about 10 times when 10% are above threshold. Some power may be recovered by subdividing those above threshold to form three or more ordered classes, but power is determined largely by the lowest threshold. Non-random ascertainment of twins (i) through affected twins and examining their cotwins or (ii) through ascertainment of all pairs in which at least one twin is affected increases power. In most cases, strategy i is more efficient than strategy ii. Though powerful for the rarer disorders, these methods suffer the disadvantage that they rely on prior knowledge of the population prevalence. Furthermore, sampling from hospital cases may introduce biases, reducing their value. A useful approach may be to assess the population with a screening instrument; the power calculations indicate that sampling all concordant and half of the discordant pairs would be efficient, as long as the cost of screening is not too high.

KEY WORDS: Twin study; threshold traits; variance power; research design; ascertainment; sampling; selection.

INTRODUCTION

The classical twin study, which uses monozygotic (MZ) and dizygotic (DZ) twins reared together from birth to adulthood, is perhaps the most widely used behavior genetic research design. Although twins share prenatal and postnatal environments to a much greater degree than other relatives, the physical and psychological consequences of this sharing *per se* appear to be relatively minor, with the exception of birth weight and some verbal IQ measures (Record et al., 1970). Thus a strong advantage of twin studies is that twins are representative of the general

population for a wide variety of phenotypes. Other assumptions of the method, e.g., that the degree of sharing of environmental factors is equal for MZ and DZ pairs, where tested, have almost always been found valid (Kendler *et al.*, 1993b; Loehlin and Nichols, 1976).

Previous exploration of the statistical power of twin study has usually been limited to the continuous, normally distributed variables. Eaves (1969, 1972) and Martin *et al.* (1978) studied the power of the twin study to resolve additive (A) and non-additive (D) genetic and common (C) and specific (E) environment effects. Likewise, Heath and Eaves (1985) studied the power of twin-family and adoption studies to resolve cultural transmission and assortative mating. We know of only two brief reports of twin study power for ordinal data. First, Neale

¹ Department of Psychiatry, Medical College of Virginia, Box 710, Richmond, Virginia 23298-0710.

² Department of Human Genetics, Medical College of Virginia, Box 3, Richmond, Virginia 23298-0003.

³ To whom correspondence should be addressed.

et al. (1989) showed that sample sizes required for ordinal data were generally much larger than for the continuous case. Second, Neale and Cardon (1992) illustrated the methods used for power calculations for both the continuous and the binary variable case. This treatment was for didactic purposes and used only one example of a true model with 30% A, 20% C, and 50% E. Thus there is no published account of a detailed study of power for binary or ordinal twin data. Such data are common in the study of medical or psychiatric illnesses, where the presence or absence of disease may be the only possible assessment.

Our aim here is to calculate the power for variables that are measured at the binary or ordinal level. A comprehensive study of these factors would prove prohibitively long, so we limit our treatment to the exploration of several key issues. First, we vary the *variance components* of the population under study (the "true world model"). Second, the parameter (*a*, *c*, or *d*) that will be fixed to zero under the false model is changed. Third, we alter the threshold of affection, corresponding to various prevalence rates in the population. Fourth, the ratio of MZ-to-DZ pairs is varied from 2:1 to 1:2 for some simple cases. Fifth, we explore the advantages of ordinal data (with more than two ordered classes) over binary data. Sixth, the value of two schemes for nonrandom ascertainment is examined: probandwise (single selection) and pairwise (or truncate). The relation between these schemes and common research strategies is discussed. Finally, because the individual researcher may wish to explore other research designs or to find the power to test other specific hypotheses, we present sample Mx scripts (Neale, 1991) for power calculation.

METHODS

Statistical Theory

The basic theory behind power calculations has been described in detail in several earlier papers (e.g., Martin *et al.*, 1978; Heath *et al.*, 1985; Neale and Cardon, 1992). Essentially we want to know what chance we have of rejecting a particular false model, when some other true world model generated the data. As mentioned by Neale and Cardon, this probability, or *power*, will depend on at least six factors: (i) the effect being considered, e.g., a^2 or c^2 ; (ii) the size of effect in the true world; (iii)

the probability level selected for rejection of the model; (iv) the sample size of the study; (v) the proportion of MZ-to-DZ twins in the sample; and (vi) the type of measurement—ordinal or continuous.

To compute the power to reject a particular hypothesis, e.g., that $a^2 = 0$, we first generate data from the true world model. The second step is to fit the false model to these generated data and obtain a χ^2 goodness-of-fit of this model, which is known as the *centrality parameter* (λ) of the non-central χ^2 distribution. Tables of λ were published by Pearson and Hartley (1972) and Haynam *et al.* (1970); they indicate critical values for probability β of rejecting the model at significance level α for a variety of degrees of freedom. β is known as the power of the test. Several computer programs will compute these values and will interpolate for values not found in the table (NAG, 1990; IMSL, 1987). Once the power of a given test is known, it is relatively straightforward to calculate the sample size required for a given level of power (λ_g), using the formula

$$N^* = \frac{\lambda_g N_0}{\lambda}$$

where N^* is the sample size required, and N_0 is the sample size of the generated data.

For complete data, we have two choices for calculating the centrality parameters. We could follow the procedure described by Neale and Cardon and use PRELIS (Jöreskog and Sörbom, 1993) to calculate tetrachoric correlations and their asymptotic weights and fit the (false) model by weighted least squares. Alternatively, we could generate contingency tables and fit the model directly to these data by maximum likelihood. For non-randomly ascertained samples we do not have this choice because PRELIS will not compute correlations for such data. However, we can still use the maximum-likelihood method as long as we correct for ascertainment, so for consistency we use this method throughout.

Threshold Model

For this study, we assume a threshold model based on an underlying normal distribution of liability (Falconer, 1960; Reich *et al.*, 1979; Neale *et al.*, 1986; Neale and Cardon, 1992, pp. 41–49). With ordinal data that have p categories, the model uses $p - 1$ thresholds to subdivide the liability distribution. The expected proportion of individuals