*Behavior Genetics, Vol. 18, No. 2, 1988*

# Resolving Causes of Developmental Continuity or "Tracking." I. Longitudinal Twin Studies During Growth

**J. K. Hewitt,[1] L. J. Eaves,[1] M. C. Neale[1], and J. M. Meyer[1]**

*Received 5 Dec. 1986—Final 23 June 1987*

*Models of developmental continuity and change in quantitative phenotypes may be tested using longitudinal data from twins. We illustrate a procedure for establishing the power and required sample sizes for detecting developmental transmission against an alternative common-factor hypothesis. We explore the general effects of different heritabilities, different fidelities of environmental and genetic developmental transmission, and varying numbers of occasions of measurement. In addition, a constraint of wide application is postulated for the action of the environment; either environmental effects are transmitted (learned) and occasion specific or they exert a constant influence which is not transmitted (learned). While the situations we examine are necessarily restricted here, our explorations of power show that, providing that we measure on at least four occasions, it is easy to detect developmental transmission with workable sample sizes.*

**KEY WORDS:** genes; environment; development; growth; twins.

## INTRODUCTION

Developmental continuity or "tracking" is so ubiquitous a phenomenon that its observation hardly draws comment. For physical characteristics such as height and weight or psychological characteristics such as intelligence, a positive correlation of measurements from one occasion to the

This work was supported in part by Grants AG04954, MH40828, GM30250, AA06781, and HL31010 from the National Institutes of Health.

[1] Department of Human Genetics, Medical College of Virginia, Richmond, Virginia 23298-0001.



0001-8244/88/0300-0133$06.00/0 

next is expected during both growth and adulthood. For a phenotype such as blood pressure the tracking phenomenon is also well established (e.g., Harlan *et al.*, 1973; Sneiderman *et al.*, 1976); individuals with higher blood pressures early in life tend to be those with higher blood pressures later on. Two mechanisms may give rise to such continuity. First, a common genetic or environmental disposition may influence an individual's observed score or phenotype whenever he/she is measured. For example, we may suppose that throughout early adulthood the constitutional properties of an individual's cardiovascular system give rise to consistently high or low blood pressure. The perturbations around this mean level, perhaps associated with an illness or an especially stressful period for the individual, are transient and whatever caused this perturbation has no effect on later occasions. We can call this a common-factor mechanism or pleiotropy.

The alternative possibility is that earlier influences are transmitted from occasion to occasion and so give rise to the measured continuity. For some cognitive tests a consequence of learning may be a transmission of environmental influences from occasion to occasion. For blood pressure, the effects of high or low levels at one age may persist and influence levels at later ages. In such a case, intervention to prevent the expression of high blood pressure at one age would have directly transmitted benefits later on.

We would like to be able to distinguish between common-factor and transmission mechanisms, or determine the relative influence of each, and to estimate environmental and genetic contributions to these. For a disease process such as hypertension, which is associated with extreme scores on an underlying continuum, failure to reject a common-factor model for longitudinal continuity during adulthood would suggest that changes in exercise habits or diet have had little or no implications for risk; at most, the effects are transient. Rejection of the common-factor model in favor of the transmission model suggests the etiological importance of environmentally determined changes in exposure to risks during adulthood only if the transmission is shown not to be solely through genetic influences.

Given its importance, this problem has received relatively little attention in behavior genetics. While there are a number of discussions of developmental data and their analysis (e.g., Nesselroade and Baltes, 1979; Rogosa and Willett, 1985), only recently have these been set specifically in the context of genetic analysis. Where they have (e.g., Province and Rao, 1985), developmental changes have been described by arbitrary statistical functions which provide few insights into the causal process. However, Eaves *et al.* (1986b) and McArdle (1986) have outlined some general

implications of a priori developmental genetic models for quantitative traits. Boomsma and Molenaar (1987) have confirmed in a specific case how it is possible to reject a common-factor model when the true mechanism is an occasion-to-occasion transmission. In this paper we cast the general approach of Eaves *et al.* (1986b) in the linear structural equation model of Joreskog and Sorbom (1984) and show how to determine the power of a longitudinal twin study to reject an inappropriate common-factor hypothesis and to detect environmental and genetic occasion-to-occasion transmission during growth. Throughout the current discussion we restrict ourselves to situations in which our observations begin at the initiation of a developmental transmission process, and we assume that the effects of this process are superimposed on otherwise stationary contemporary contributions from common and specific factors. Using restricted examples we illustrate general properties of the models and the method for calculating the power of any particular study to reject an inappropriate common factor model (Satorra and Saris, 1985). We then discuss and illustrate less restricted examples of the approach and demonstrate that quite small longitudinal studies will readily detect transmission.

## THE MODEL

A model for development, adapted from Eaves *et al.* (1986b), is shown in Fig. 1. The phenotype is measured on $m + 1$ occasions, giving the observations $P_0, P_1, \ldots, P_i, \ldots, P_m$. The observed phenotype is assumed to be a simple additive function of latent environmental and genetically determined phenotypes, $E'_i$ and $G'_i$, so that $P_i = eE'_i + hG'_i$. We may set $h^2 + e^2 = 1$ so that $h^2$ is the heritability at $i = 0$. On any particular occasion with $i > 0$, the latent environmental influence may be an additive function of three variables: $E_c$, the constant environmental factor; $E_s$, the environmental factor specific to an occasion; and $E'_{i-1}$, the environmental effect on the previous occasion. The paths from these influences are $e_{ci}$, $e_{si}$, and $z_i$, respectively. Corresponding genetic influences have paths $g_{ci}$, $g_{si}$, and $j_i$, respectively. The common-factor model is represented by the special case when $j_i = z_i = 0$, $i = 1, \ldots, m$. The transmission model is the special case when $e_{ci} = g_{ci} = 0$, $i = 0, \ldots, m$.[2]

[2] Our reviewers have pointed out that in certain circumstances, namely, when $j$ (or $z$) = 1 and $g_s$ (or $e_s$) = 0, the common-factor model is formally equivalent to the transmission model. Therefore the distinction between a common-factor model and a transmission model should be conceived of as a relative one.

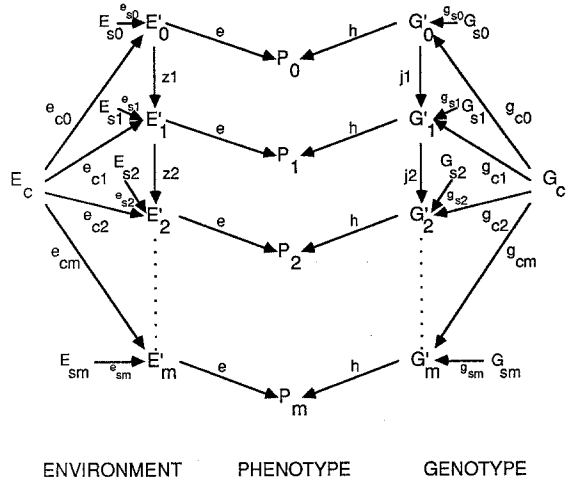


**Fig. 1.** The developmental path model for an individual.

Figure 2 shows the extension of the basic model to include members of a twin pair. The correlation $r$ is the genetic correlation between the individuals, which is 1 for monozygotic (MZ) twins and .5 for dizygotic (DZ) twins under an additive gene action and random mating. (For easier reading the $i$ subscripts are dropped for the common-factor and transmission paths in Fig. 2.)

Although it is not necessary to do so, this general model may be conveniently cast in terms of the widely used scheme for analyzing linear structural relationships available as LISREL VI (Joreskog and Sorbom,

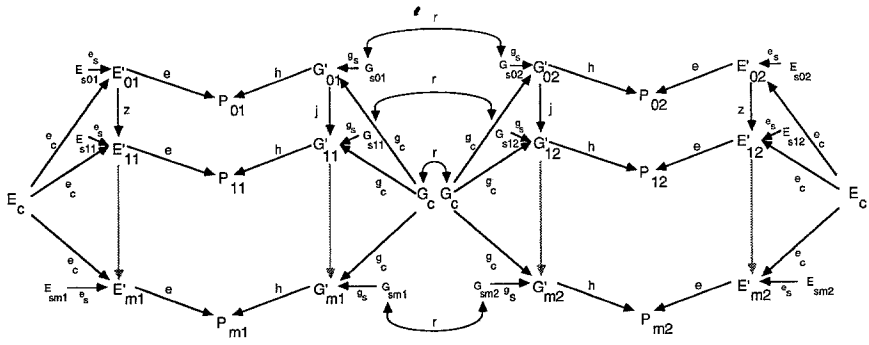

**Fig. 2.** The developmental path model for a pair of twins.

1984). We observe the $2(m + 1)$ phenotypes $y = [P_{01}, P_{02}, P_{11}, P_{12}, \ldots, P_{i1}, P_{i2}, \ldots, P_{m1}, P_{m2}]'$, where subscript $i$ is the occasion and 1,2 are the individuals in a twin pair, and we model the resulting $2(m + 1) \times 2(m + 1)$ variance–covariance matrix. The structural equations are

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi}$$

and then

$$\mathbf{Y} = \boldsymbol{\Lambda}\boldsymbol{\eta},$$

where

$$\boldsymbol{\eta}[4(m + 1) \times 1] = [G'_{01}, G'_{02}, E'_{01}, E'_{02}, \ldots, G'_{m1}, G'_{m2}, E'_{m1}, E'_{m2}]',$$

$$\boldsymbol{\xi}[4(m + 2) \times 1] = [G_{c1}, G_{c2}, E_{c1}, E_{c2}, G_{s01}, G_{s02}, E_{s01}, E_{s02}, \ldots, G_{sm1}, G_{sm2}, E_{sm1}, E_{sm2}]',$$

and

$\boldsymbol{\Gamma}[4(m + 1) \times 4(m + 2)]$

$$= \left[\begin{array}{cccc|cccc||cccc} g_{c0} & 0 & 0 & 0 & g_{s0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & g_{c0} & 0 & 0 & 0 & g_{s0} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e_{c0} & 0 & 0 & 0 & e_{s0} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e_{c0} & 0 & 0 & 0 & e_{s0} & 0 & 0 & 0 & 0 \\ \hline\hline g_{cm} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{sm} & 0 & 0 & 0 \\ 0 & g_{cm} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{sm} & 0 & 0 \\ 0 & 0 & e_{cm} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & e_{sm} & 0 \\ 0 & 0 & 0 & e_{cm} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & e_{sm} \end{array}\right],$$

$\mathbf{B}[4(m + 1) \times 4(m + 1)]$

$$= \left[\begin{array}{cccccccc||cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ j_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & j_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & z_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & z_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & j_m & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & j_m & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & z_m & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & z_m & 0 & 0 & 0 & 0 \end{array}\right],$$

$$\Lambda[2(m+1) \times 4(m+1)]$$

$$= \left[\begin{array}{cccccccc||cccc} h & 0 & e & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & h & 0 & e & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & h & 0 & e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & h & 0 & e & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h & 0 & e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h & 0 & e \end{array}\right].$$

Under these equations

$$\mathbf{S} = (\mathbf{YY}') = \mathbf{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma\Phi\Gamma}'(\mathbf{I} - \mathbf{B}')^{-1}\mathbf{\Lambda}',$$

where the factor correlation matrix

$$\mathbf{\Phi}[4(m+2) \times 4(m+2)] = \left[\begin{array}{cccc||cccc} 1 & r & 0 & 0 & 0 & 0 & 0 & 0 \\ r & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline\hline 0 & 0 & 0 & 0 & 1 & r & 0 & 0 \\ 0 & 0 & 0 & 0 & r & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

The above gives a direct representation of the model in Figs. 1 and 2 and can be used to generate variance–covariance matrices under particular sets of parameter values. This enables us to explore the characteristic effects of this developmental model on the expected variance–covariance matrix under conditions of interest. However, for estimation purposes the model is not identified without further restrictions. In practice, we may fix $h$ and $e$ in $\mathbf{\Lambda}$ so that $h^2$ is the known or observed heritability at $i = 0$.

Since the $r$ in $\mathbf{\Phi}$ are given, we estimate only the parameters in $\mathbf{\Gamma}$ and $\mathbf{B}$ from the longitudinal data.

## GENERAL PROPERTIES OF THE MODEL AND ITS RESOLUTION

### The Change in Phenotypic Variance

Suppose that on any occasion there is an input from contemporary genetic and environmental influences, from either common factors or factors specific to that occasion, such that the resulting phenotypic variance is constant; for example, we may set $g_{c0} = g_{c1} = \ldots g_{ci}$, and so on, with

$$e^2(e_c^2 + e_s^2) + h^2(g_c^2 + g_s^2) = 1.$$

If we now introduce positive developmental transmission from the previous occasion, its consequence is an increase in this phenotypic variance. Eaves *et al.* (1986b) have shown how this variance increases toward an asymptotic value for their model. Setting $j_1 = j_2 = \ldots j_m = j$ and $z_1 = z_2 = \ldots z_m = z$, when contemporary influences are specific to each occasion, i.e., $g_c = e_c = 0$, then

$$V_{P_m} = h^2(1 + j^2 + \ldots j^{2m}) + e^2(1 + z^2 + \ldots z^{2m}),$$

simplifying to

$$V_{P_m} = h^2\left[\frac{1 - j^{2k}}{1 - j^2}\right] + e^2\left[\frac{1 - z^{2k}}{1 - z^2}\right], \qquad \text{where } k = m + 1.$$

With both general and specific influences we have

$$V_{P_m} = h^2[g_c^2 a_k + g_s^2 b_k] + e^2[e_c^2 c_k + e_s^2 d_k],$$

where

$$a_k = \left[\frac{1 - j^k}{1 - j}\right]^2, \qquad b_k = \frac{1 - j^{2k}}{1 - j^2},$$

$$c_k = \left[\frac{1 - z^k}{1 - z}\right]^2, \qquad d_k = \frac{1 - z^{2k}}{1 - z^2}.$$

Thus an important general result is that developmental transmission will give increasing phenotypic variance with age, up to an asymptotic value, such as that observed for adult systolic blood pressure (Hewitt *et al.*, 1988). Of course where measurements are age standardized or on an arbitrary scale, such information will not be directly available. Moreover, at equilibrium, but not before, the process will become a stationary process (Fredericksen and Rotondo, 1979).

The effects on covariances or correlations between relatives will depend on the relative predominance of environmental or genetic transmission. Under random mating and additive gene action the genetic covariance between individuals measured on occasions $k - 1$ and $l - 1$, respectively, $l \leqq k$, will be

$$\mathrm{Cov}_{P_k,P_l} = rh^2[g_c^2 a_{k,l} + g_s^2 b_{k,l}],$$

where

$$a_{k,l} = \frac{(1 - j^k)(1 - j^l)}{(1 - j)^2}, \qquad b_{k,1} = \frac{j^{k-l}(1 - j^{2l})}{1 - j^2},$$

and $r$ is the coefficient of relationship. If there is no genetic transmission ($j = 0$), but there is environmental transmission, then, of course, the correlation between relatives will decline with age.

### Phenotype-to-Phenotype Transmission

The model as we have outlined it does not specify direct measured phenotype-to-measured phenotype transmission. Eaves *et al.* (1986b) noted that such direct transmission is formally equivalent to setting $j = z$ in our model; formulating the model as we have facilitates parameter estimation and permits a test of the assumption of phenotype-to-phenotype transmission. The independent identification and estimation of all three developmental paths will be considered in another paper.

### The Effect of Transmission Along with a Constant Environment

Unlike genetic effects which contribute to both intra- and interindividual covariance, the environmental effects in our model contribute only to intraindividual covariance across occasions. One consequence of this is that if $j < z$, then test–retest correlations will increase during development, while correlations between relatives decrease. The intraindividual environment covariance between occasions $k - 1$ and $l - 1$, $l < k$, will be

$$\mathrm{Cov}_{Pk,Pl} = e^2[e_c{}^2 c_{k,l} + e_s{}^2 d_{k,l}],$$

where

$$c_{k,l} = \frac{(1 - z^k)(1 - z^l)}{(1 - z)^2}$$

and

$$d_{k,l} = \frac{z^{k-l}(1 - z^{2l})}{1 - z^2}.$$

Displayed in Table I are equilibrium values of $a_{k,l}(=a_{k,k})$ $d_{k,l}$ and $d_{k,k}$ for $k - l = 1$ together with illustrative value of $V_P$, $r_{\text{test,retest}}$, and $r_{\text{MZ}}$ for the case $h^2 = e^2$ and $g_c{}^2 = g_s{}^2 = e_c{}^2 = e_s{}^2$ and for $g_c{}^2 = e_s{}^2$, $g_s{}^2 = e_c{}^2 = 0$. The important conclusion from Table I is that when $z > j$ the test–retest stability can increase considerably as a consequence of transmitted constant environments, while the MZ correlation falls.

In reality there are few phenotypes whose test–retest stability correlation is as high as .8, while the correlation between identical twins is

**Table I.** Selected Equilibrium Values of Variance and Correlation Parameters

| | | | | | | | | $g_c^2 = g_s^2 = e_c^2 = e_s^2$ | | | $g_c^2 = e_s^2; g_s^2 = e_c^2 = 0$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $j$ | $z$ | $a_{k,l}$ | $b_{k,l}$ | $b_{k,k}$ | $c_{k,l}$ | $d_{k,l}$ | $d_{k,k}$ | $V_P$ | $r_{t,rt}$ | $r_{MZ}$ | $V_P$ | $r_{t,rt}$ | $r_{MZ}$ |
| .0 | .0 | 1.00 | .00 | 1.00 | 1.00 | .00 | 1.00 | 1.00 | .50 | .50 | 1.00 | .50 | .50 |
| | .2 | 1.00 | .00 | 1.00 | 1.56 | .21 | 1.04 | 1.15 | .60 | .44 | 1.02 | .59 | .49 |
| | .4 | 1.00 | .00 | 1.00 | 2.78 | .48 | 1.19 | 1.49 | .71 | .34 | 1.10 | .68 | .45 |
| | .6 | 1.00 | .00 | 1.00 | 6.25 | .94 | 1.56 | 2.45 | .84 | .20 | 1.28 | .76 | .39 |
| | .8 | 1.00 | .00 | 1.00 | 25.00 | 2.22 | 2.78 | 7.45 | .95 | .07 | 1.89 | .85 | .26 |
| .4 | .0 | 2.78 | .48 | 1.19 | 1.00 | .00 | 1.00 | 1.49 | .71 | .67 | 1.89 | .74 | .74 |
| | .4 | 2.78 | .48 | 1.19 | 2.78 | .48 | 1.19 | 1.99 | .82 | .50 | 1.99 | .82 | .70 |
| | .8 | 2.78 | .48 | 1.19 | 25.00 | 2.22 | 2.78 | 7.94 | .96 | .13 | 2.78 | .90 | .50 |
| .8 | .0 | 25.00 | 2.22 | 2.78 | 1.00 | .00 | 1.00 | 7.45 | .95 | .93 | 13.00 | .96 | .96 |
| | .4 | 25.00 | 2.22 | 2.78 | 2.78 | .48 | 1.19 | 7.94 | .96 | .87 | 13.10 | .97 | .95 |
| | .8 | 25.00 | 2.22 | 2.78 | 25.00 | 2.22 | 2.78 | 13.89 | .98 | .50 | 13.89 | .98 | .90 |

as low as .2, with the possible exception of hand preference. This suggests that if transmission of environmental influences rather than genetic influence predominates, then those environmental influences must occur on specific occasions rather than continuously. Even when $j = z$, i.e., transmission is effectively phenotypic without regard to its genetic or environmental cause, only if the MZ correlation is markedly less than the test–retest stability can $e_c$ take other than low or zero values. For traits influenced by a shared family environment the argument comes to hinge on the relative magnitude of the MZ and DZ correlations, which in the presence of a constant, shared, and developmentally transmitted environmental influence, will rapidly approach each other in magnitude. For traits such as IQ, where the reverse is the case, i.e., $r_{MZ}$ and $r_{DZ}$ diverge from birth to adolescence, environmental influences are seen to be occasion specific whether shared or individual, while genetic influences act continuously on the phenotype (Eaves *et al.*, 1986b).

In short, for most traits of interest, environmental effects will be found to be either developmentally transmitted but occasion specific or constant but not developmentally transmitted, but not both.

## No Unambiguous Resolution with Measurements on Three or Fewer Occasions

Our principal concern is discriminating between developmental transmission and common-factor accounts of developmental continuity. For individuals, with measurements on just three occasions we have three unique covariances and three variances. Even when the process is purely a transmission process, leading to a perfect simplex correlation matrix, a single common-factor model with distinct common-factor loadings for each occasion, along with specifics, will be formally equivalent (Joreskog, 1974). Similarly twin data from three occasions yield 12 unique variances and covariances which are fit perfectly by the 12 parameters of a genetic and environmental unconstrained common-factor model. Of course while such a model will have no predictive value and may be less parsimonious, we need to be aware that longitudinal studies with measurements on three or fewer occasions are insufficient to resolve even simple models of development, in the absence of further restrictive assumptions. Having said this, in practice a considerable amount of background knowledge of test–retest stabilities, typical values of correlations between relatives, and population variances is often available; putting this background information together with that from a study having measurements on only two occasions may prove useful in generating plausible hypotheses about development (Hewitt *et al.*, 1988).

## POWER CALCULATIONS

The procedure followed for computing the power to reject an incorrect model is similar to that set out by Martin *et al.* (1978) and Eaves *et al.* (1986). The robustness of the procedure has also been demonstrated in the context of structural equation modeling by Satorra and Saris (1985). First the true covariance matrix between the $P_i$ is generated for a particular set of parameter values. We then find the ML estimates for the parameters of the false model. For this purpose we have employed the LISREL VI estimation package (Joreskog and Sorbom, 1984), which minimizes the likelihood function.

$$F = \ln | \boldsymbol{\Sigma} | - \ln | \mathbf{S} | + tr(\mathbf{S}\,\boldsymbol{\Sigma}^{-1}) - k,$$

where $\mathbf{S}$ is the "observed" covariance matrix, $\boldsymbol{\Sigma}$ is the expected covariance matrix under the model being tested, and $k$ is the number of manifest variables. In our case $\mathbf{S}$ is the expected covariance matrix on the true model and $\boldsymbol{\Sigma}$ is the expected covariance matrix under the model being tested. If the observed variables follow a multivariate normal distribution and the sample sizes ($N$) are reasonably large, then $N - 1$ times the minimum value of $F$ is distributed as $\chi^2$ with $s - p$ degrees of freedom, where $s$ is the number of observed statistics and $p$ is the number of parameters. The $\chi^2$ may be used as a likelihood-ratio test of the adequacy of a constrained model, and we may reject the incorrect model on the basis of this test. Since the $\chi^2$ statistic is a function of sample size we may estimate the size of the study required to give sufficient power in the following way.

We determine $\chi^2_{s-p} = \lambda'$, which is a measure of the inadequacy of the false model based on an arbitrary sample size; it is convenient to take $n = 100$. Using tables based on the noncentral chi-square distribution (Pearson and Hartley, 1972), we may determine the value of $\lambda$ to ensure rejection of the false model at the $\alpha = 5\%$ level of significance in a given proportion, $\beta$, of the studies. The required sample size is given by

$$N = \lambda/\lambda' \times 100.$$

If the degrees of freedom for $\chi^2$ are greater than 100, then the available tables may be used with graphical extrapolation to give $\lambda$ to a good approximation. Alternatively, and more accurately, we may use the noncentral chi-square function (PROBCHI) implemented by SAS (SAS Institute, 1985), which yields

$$\beta = 1 - \text{PROBCHI}(\chi^2_{1-\alpha}, \text{df}, \lambda),$$

where $\beta$ is the power attained at the $100\alpha\%$ level of significance for given degrees of freedom and noncentrality parameters $\lambda$. It is a simple enough

matter to determine the required $\lambda$ for any specified values of $\beta$, $\alpha$, and df (Hewitt and Heath, 1988). Using this procedure we are able to explore the relationship between power, and hence the necessary sample sizes, and various aspects of the traits to be studied. For example, we can see what are the effects of differences in the fidelity of developmental transmission of the trait, the relative importance of transmission of environmental versus genetic influences, and the heritability of the trait.

## THE EFFECT OF INITIAL HERITABILITY ON POWER

To explore the general characteristics of the system we first generated the "true" covariance matrices for pairs of MZ and DZ twins. Rather than try all possible numbers of occasions we chose a number which was greater than three but reasonably accommodated by LISREL VI. Setting the number of occasions at six yielded a 12 × 12 covariance matrix of individual scores $P_{01}, P_{02}, P_{11}, P_{12}, \ldots, P_{m1}, P_{m2}$, where $m = 5$, for each zygosity.

We are initially interested in situations where $g_c = e_c = 0$, that is, the developmental continuity is a consequence of transmission in the absence of common-factor influences on the phenotype. We imposed the additional constraint that $g_{si} = e_{si} = 1, i = 0, \ldots, m$.

Our false model, however, assumes that $j = z = 0$ and attempts to account for continuity by the common-factor hypothesis. In estimating the parameters of the model we initially imposed the constraints that $g_{s0} = g_{s1} = \cdots = g_{sm}$, and $e_{s0} = e_{s1} = \cdots = e_{sm}$.

To identify the remaining parameters we set $h$ and $e$ to their true values. Under these conditions the power of a given twin study to reject the common-factor model depends on the magnitudes of $j$ and $z$ and the heritability of the phenotype.

Table II shows the $\lambda'$ values generated by LISREL VI for the inadequacy of the common-factor model fitted to the covariance matrix. Also given are the corresponding approximate sample sizes for 90% certainty of rejecting the model at the 5% level of significance, based on 152 df.

In many applications we cannot be sure whether the units of measurement are equivalent from occasion to occasion. Nevertheless, we model the unstandardized covariance matrix in order that the chi-square interpretation of the likelihood is valid (Joreskog and Sorbom, 1984). To accommodate this situation we have generated as a "worst case" the true model expected correlation matrix and modeled it as a covariance matrix. As can be seen from Table II there is only a small loss of power to detect transmission.

**Table II.** The Pattern of Consequences of Different $j$, $z$, and Heritability Values for Noncentrality Parameters ($\lambda'$) and Required Sample Sizes ($N$)[a]

| | | | Modeling covariance | | Modeling correlation | |
|---|---|---|---|---|---|---|
| $j$ | $z$ | $h^2$ | $\lambda'$ | $N$ | $\lambda'$ | $N$ |
| .1 | .1 | .1 | 12.8[b] | 458 | 12.8[b] | 458 |
| | | .4 | 12.8 | 458 | 12.8 | 458 |
| | | .8 | 12.8[b] | 458 | 12.8[b] | 458 |
| .4 | .4 | .1 | 183.9 | 32 | 181.5[b] | 32 |
| | | .4 | 183.9[b] | 32 | 181.5 | 32 |
| | | .8 | 183.9[b] | 32 | 181.5[b] | 32 |
| .4 | .8 | .1 | 630.7 | 9 | 548.1 | 11 |
| | | .4 | 496.8 | 12 | 442.6 | 14 |
| | | .8 | 367.6 | 16 | 342.3 | 17 |
| .8 | .4 | .1 | 218.5 | 27 | 212.7 | 28 |
| | | .4 | 324.4 | 18 | 307.3 | 20 |
| | | .8 | 471.8 | 12 | 439.2 | 14 |
| .8 | .8 | .1 | 691.7 | 9 | 596.7 | 10 |
| | | .4 | 691.9 | 9 | 596.7 | 10 |
| | | .8 | 691.8[b] | 9 | 596.7[b] | 10 |

[a] "True" model: $g_c^2 = e_c^2 = 0$; $g_s^2 = e_s^2 = 1$; $j$, $z$, $h^2$ constant as given. "False" model: $j = z = 0$; $g_c^2$, $e_c^2$, $g_s^2$, $e_s^2$, $h^2$ constant.
[b] Values derived, not computed by LISREL VI.

It is useful to note that the $\lambda'$ statistic is a linear function of the heritability of the trait. When $j > z$ and transmission is primarily genetic, power increases with the heritability of the trait. When $j < z$, traits with low heritabilities give rise to more powerful tests for rejecting the common factor hypothesis. When $j = z$, that is, the transmission is effectively phenotypic, the heritability of the trait is irrelevant.

The results show that in the simplest situation where transmission is the only cause of developmental continuity and has a moderate fidelity ($j$ and $z$ = .4), then studies with as few as 32 pairs each of MZ and DZ twins measured on six occasions have a 90% chance of rejecting the simplest common-factor hypothesis at the 5% level.

If the fidelity of transmission is poor ($j$ and $z$ about .1, say) then the necessary number of pairs of each zygosity rises to about 450 for the same power.

Allowing the common-factor loadings, $g_c$ and $e_c$, to take independent values on each occasion will, as we see below, reduce the noncentrality parameters by about 50% or more. However, the general properties of the effects on power of different $j$, $z$, and $h^2$ values for the true model will follow those shown in Table II.

## DISCRIMINATION OF TRANSMISSION AND COMMON-FACTOR HYPOTHESES WITH UNRESTRICTED MODELS

For a more detailed illustration we chose to explore two plausible situations in which the true model involves both developmental transmission and common-factor influences. The expected covariance matrix for the six occasions, $m = 0, 1, \ldots, 5$, was generated under the following conditions:

(I) $j = z = .4$, $g_c = 1$, $g_s = 0$, $e_c = 0$, $e_s = 1$, $h^2 = .4$;

(II) $j = .4$, $z = .8$, $g_c = 1$, $g_s = 0$, $e_c = 0$, $e_s = 1$, $h^2 = .4$.

For each case we considered either restricted, partially restricted, or unrestricted single common-factor models and restricted, partially restricted, or unrestricted transmission models; a restricted model has constant parameters, a partially restricted model allows the occasion specific paths to vary across occasi ons, and an unrestricted model lets all parameters vary across occasions. The required sample size for a given power to reject both the inappropriate common factor and the transmission model will be the larger of the two indicated in Table IV. In addition to varying the restrictiveness of the false models, we explored the effect of observing the phenotype on only the first three, four, or five occasions instead of six. As before, to allow for the possible absence of scale invariance across occasions, we standardized the covariance matrix, when fitting the factor models. However, such a procedure is inappropriate when fitting restricted transmission models since these require changes in variance.

We may note first that, as we have already said, a longitudinal twin study with measurements on only three occasions is insufficient by itself to reject an unrestricted common-factor hypothesis. Perhaps more interestingly, an unrestricted transmission model cannot be rejected given our true model however many occasions we have. However, the transmission parameters have to start off greater than unity and decline steadily as the growth process moves toward equilibrium. The only way to *predict* this change in transmission values is to suppose that the developmental process can be modeled by a restricted transmission and common-factor process together, which will give a more parsimonious account of equally good fit.

Turning to the sample sizes needed to reject restricted false models, the upper halves of Tables III and IV record an encouraging outcome. Even to reject an unrestricted common-factor hypothesis will require only about 100 pairs of each zygosity measured on four occasions. If the number of occasions can be increased to six, the required sample sizes fall to about 50 or fewer pairs of each zygosity, depending on the magnitude

**Table III.** Noncentrality Parameters ($\lambda'$) and Required Sample Sizes ($N$) for Case I

Case I. World: $j = z = .4$, $g_c = 1$, $g_s = 0$, $e_c = 0$, $e_s = 1$, $h^2 = .4$

| Model | Parameter | Number of occasions | | | |
|---|---|---|---|---|---|
| | | 6 | 5 | 4 | 3 |
| Restricted factor model: | | | | | |
| $j = z = 0$; $g_c$, $g_s$, $e_c$, $e_s$ const. | $\lambda'$ | 223.88 | 158.21 | 95.70 | 40.71 |
| | df | 152 | 106 | 68 | 38 |
| | $N$ | 26 | 32 | 44 | 82 |
| Partially restricted factor model: | | | | | |
| $j = z = 0$; $g_c$, $e_c$ const. | $\lambda'$ | 175.82 | 116.20 | 63.37 | 21.53 |
| | df | 142 | 98 | 62 | 34 |
| | $N$ | 32 | 42 | 64 | 142 |
| Unrestricted factor model: | | | | | |
| $j = z = 0$ | $\lambda'$ | 103.28 | 62.33 | 30.15 | .00 |
| | df | 132 | 90 | 56 | 30 |
| | $N$ | 54 | 77 | 129 | — |
| Restricted transmission model: | | | | | |
| $g_c = e_c = 0$; $j$, $z$, $g_s$, $e_s$ const. | $\lambda'$ | 116.99 | 88.56 | 61.10 | 35.58 |
| | df | 152 | 106 | 68 | 38 |
| | $N$ | 50 | 57 | 69 | 94 |
| Partially restricted transmission model: | | | | | |
| $g_c = e_c = 0$; $j$, $z$ const. | $\lambda'$ | 8.80 | 6.95 | 4.72 | 2.38 |
| | df | 142 | 98 | 62 | 34 |
| | $N$ | 647 | 704 | 856 | 1339 |
| Unrestricted transmission model: | | | | | |
| $g_c = e_c = 0$ | $\lambda'$ | .00 | .00 | .00 | .00 |
| | "df" | 134 | 92 | 58 | 32 |

of the transmission parameters. Generally for this purpose, increasing the number of occasions is more productive than increasing the number of subjects.

Against this picture of reasonable power to reject an inappropriate common-factor hypothesis, the power to reject transmission models, except of the most restricted type, is poor. It would appear that given both transmission and a common pleiotropic genetic factor, we are unlikely to be able to detect the common genetic factor statistically.

## DISCUSSION

Beyond simply describing developmental trends, progress in understanding the genetic and environmental causes of continuity and change requires formalized and testable hypotheses. These hypotheses should have explanatory power, that is, they should predict the quantitative re-

**Table IV.** Noncentrality Parameters ($\lambda'$) and Required Sample Sizes ($N$) for Case II

Case II. World: $j = .4$, $z = .8$, $g_c = 1$, $g_s = 0$, $e_c = 0$, $e_s = 1$, $h^2 = .4$

| Model | Parameter | Number of occasions | | | |
|---|---|---|---|---|---|
| | | 6 | 5 | 4 | 3 |
| Restricted factor model: | | | | | |
| $j = z = 0$; $g_c$, $g_s$, $e_c$, $e_s$ const. | $\lambda'$ | 598.19 | 386.70 | 209.09 | 76.32 |
| | df | 152 | 106 | 68 | 38 |
| | $N$ | 10 | 13 | 20 | 44 |
| Partially restricted factor model: | | | | | |
| $j = z = 0$; $g_c$, $e_c$ const. | $\lambda'$ | 465.38 | 277.58 | 130.61 | 29.74 |
| | df | 142 | 98 | 62 | 34 |
| | $N$ | 12 | 18 | 31 | 107 |
| Unrestricted factor model: | | | | | |
| $j = z = 0$ | $\lambda'$ | 266.70 | 149.28 | 64.58 | .00 |
| | df | 132 | 90 | 56 | 30 |
| | $N$ | 21 | 32 | 60 | — |
| Restricted transmission model: | | | | | |
| $g_c = e_c = 0$; $j$, $z$, $g_s$, $e_s$ const. | $\lambda'$ | 84.91 | 73.12 | 59.51 | 43.02 |
| | df | 152 | 106 | 68 | 38 |
| | $N$ | 69 | 69 | 71 | 77 |
| Partially restricted transmission model: | | | | | |
| $g_c = e_c = 0$; $j$, $z$ const. | $\lambda'$ | 12.72 | 10.77 | 8.10 | 4.41 |
| | df | 142 | 98 | 62 | 34 |
| | $N$ | 448 | 454 | 499 | 723 |
| Unrestricted transmission model: | | | | | |
| $g_c = e_c = 0$ | $\lambda'$ | .00 | .00 | .00 | .00 |
| | "df" | 134 | 92 | 58 | 32 |

lationships they seek to explain and not simply describe them using arbitrary mathematical functions of increasing complexity. The common-factor and transmission model described here and elsewhere (Eaves *et al.*, 1986b) sets out to predict changing intra- and interindividual variance–covariance patterns in terms of a small number of parameters. Its success should depend on the conformity of empirical reality to the model. If it does not, then the model is not testable. Where such a model is testable and appropriate, its explanatory and predictive power is considerable. The focus of this paper has been the power of longitudinal twin studies to resolve the contributions of these common-factor and transmission mechanisms during growth. Although we have imposed quite restrictive assumptions, there are a number of workable generalizations.

Under our assumptions, a hallmark of the superposition of developmental transmission on a system *otherwise in equilibrium* is a rise in phenotypic variance up to a new asymptotic value. If we are analyzing

standardized data, this will be detected as a decrease in the proportion of variance which is occasion specific. For preadolescent cognitive development (Eaves *et al.*, 1986b), adult blood pressure (Hewitt *et al.*, 1988), and Scholastic Aptitude Test scores (Joreskog, 1974; Hilton, 1969), for example, changes consistent with transmission superimposed on contemporary variance are observed. In the absence of this kind of change, transmission is not a useful hypothesis to account for developmental continuity, although some ad hoc provisions could permit a fit to the data. For example, we might reasonably allow the contemporary variance at $i = 0$ to take a value different from, and perhaps larger than, the subsequent contemporary variances. This uses additional degrees of freedom and weakens the developmental hypotheses. In practice, however, we may need to weaken our assumptions in this way to fit actual data.

A second result of general importance applies where environmental transmission, e.g., learning, predominates over transmission of genetic effects ($z > j$). In such situations the developmental pattern of test–retest and twin pair correlations which is usually observed arises under our model only when transmitted environmental effects are occasion specific rather than constant in their primary action. To give this some substance, suppose that an individual environmental influence such as the presence or absence of a particular experience (perhaps a vocabulary game) increased or decreased a child's vocabulary score during a particular month and that this increase or decrease is transmitted forward from month to month. If the games have the same ownership and each month ownership of the game contributes again to the vocabulary score, then the results will be increasing variance and increasing test–retest correlation but decreasing family resemblance if the games are randomly distributed to children initially or decreasing genetically based resemblance if the games are allocated to whole families who share them. In the latter case the difference between the MZ and the DZ correlation would diminish with increasing reliability of the test scores, while in the former case (which is studied in detail in Table I) MZ correlations will decline as test–retest reliabilities rise. Neither of these results is observed for human traits with the exception, as we have noted, of hand preference. This suggests that the effects of our environmental experience either are not transmitted forward (i.e., learned) or do not continue to exert an impact from month to month. Presumably the direct effect of a given environmental input is short-lived; a particular input is beneficial but is soon outgrown and exposure to the next critical environmental input is not highly correlated environmentally with exposure to the previous input. This specific conclusion is not new in behavior genetics but we may now assert it as a general result for traits where MZ and DZ correlations do not converge

and where MZ correlations do not decline relative to test–retest reliabilities.

As far as longitudinal twin study design is concerned, the power to detect genetic and environmental transmission will depend, among other things, on the relative magnitudes of the transmission parameters, the initial heritability, the sample sizes, and the number of occasions of measurement. As we would expect, higher initial heritabilities make it easier to detect genetic transmission but more difficult to detect environmental transmission. Nevertheless, unless the transmission is of extremely poor fidelity, the results of our power calculations suggest that we will have no difficulty in detecting it against an alternative common-factor model. Although confining ourselves to consideration of a single common factor or pleiotropic gene action throughout development might be considered too restrictive, merely increasing the number of common factors as the number of occasions of measurement rises can lead only to ad hoc description rather than explanation. Our real difficulty, however, is in rejecting plausible developmental transmission models, and we shall return to this elsewhere (Hewitt *et al.*, 1988).

In finding that within the range studied, additional occasions were generally more productive than additional subjects for resolving our developmental models, we noted that subject compliance and other practical considerations, such as the length of the study, necessitated a compromise here. A practical solution which is likely to be of considerable use is the overlapping cohort design (Heath and Eaves, 1986). We have been able to show that in a variety of situations interesting aspects of an a priori developmental model are testable and that the sample sizes required need not be unreasonably large. The exact situation depends on the true parameters and which aspects of the model are to be tested.

## ACKNOWLEDGMENTS

We are grateful to J. McArdle, D. Boomsma, and P. Molenar, who critically reviewed an early version of the manuscript. We have no doubt that they would still approach this problem rather differently, and they are in no way responsible for the present paper.

## REFERENCES

Boomsma, D. T., and Molenaar, C. M. (1987). Genetic analysis of repeated measures. I. Simplex models. *Behav. Genet.* **17**:111–124.

Eaves, L. J., Kendler, K. S., and Schulz, S. C. (1986a). The familial sporadic classification: Its power for the resolution of genetic and environmental etiological factors. *J. Psychiat. Res.* **20**:115–130.

Eaves, L. J., Long, J., and Heath, A. C. (1986b). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav. Genet.* **16**:143–162.

Fredericksen, C. H., and Rotondo, J. H. (1979). Time-series models and the study of longitudinal change. In Nesselroade, J. R., and Baltes, P. B. (eds.), *Longitudinal Research in the Study of Behavior and Development*, Academic Press, New York, pp. 111–153.

Harlan, W. R., Oberman, A., Mitchel, R. E., and Graybiel, A. (1973). A 30 year study of blood pressure in a white male cohort. In Onesti, G., Kim, K. E., and Moyer, J. H. (eds.), *Hypertension: Mechanisms and Management*, Grove and Stratton, New York.

Heath, A. C., and Eaves, L. J. (1986). Research strategies for the resolution of developmental genetic models. Abstract for Fifth International Congress on Twin Studies, Amsterdam.

Hewitt, J. K., and Heath, A. C. (1988). A note on computing the chi-square noncentrality parameter for power analyses. *Behav. Genet.* **18**:105–108.

Hewitt, J. K., Carroll, D., Sims, J., and Eaves, L. J. (1988). A developmental hypothesis for adult blood pressure. *Acta Genet. Med. Gemellol.* (in press).

Hilton, T. L. (1969). *Growth Study Annotated Bibliography*, Progress Report 69-11, Educational Testing Serives, Princeton, N.J.

Joreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In Krantz, D. H., Atkinson, R. C., Luce, R. D., and Suppes, P. (eds.), *Contemporary Developments in Mathematical Psychology, Vol. II. Measurement, Psychophysics, and Neural Information Processing*, W. H. Freeman, San Francisco.

Joreskog, K. G., and Sorbom, D. (1984). *LISREL VI User's Guide*, Scientific Software, Mooresville, Ind.

Martin, N. G., Eaves, L. J., Kearsey, M. J., and Davies, P. (1978). The power of the classical twin study. *Heredity* **40**:97–116.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behav. Genet.* **16**:163–200.

Nesselroade, J. R., and Baltes, P. B. (eds.) (1979). *Longitudinal Research in the Study of Behavior and Development*, Academic Press, New York.

Pearson, E. S., and Hartley, H. O. (eds.) (1972). *Biometrika Tables for Statisticians, Vol. 2*, Cambridge University Press, Cambridge.

Province, M. A., and Rao, D. C. (1985). A new model for the resolution of cultural and biological inheritance in the presence of temporal trends: Applications to systolic blood pressure. *Genet. Epidemiol.* **2**:363–374.

Rogosa, D. R., and Willett, J. B. (1985). Understanding correlates of change by modelling individual differences in growth. *Psychometrika* **50**:203–228.

SAS Institute (1985). *SAS User's Guide: Basics, Version 5 edition*, SAS Institute, Cary, N.C.

Sattorra, A., and Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika* **50**:83–90.

Sneiderman, C., Heyden, S., Heiss, G. Tyroler, H., and Hames, C. (1976). Predictors of blood pressure over a 16 year follow-up of 163 youths. *Circulation (Suppl. 2)* **54**:11–24.

Edited by N. G. Martin