

A NOTE ON HOLZINGER'S HERITABILITY COEFFICIENT h^2

PETER H. SCHÖNEMANN

*National Taiwan University
Taipei, Taiwan, R.O.C.**

It is shown that the still widely used heritability estimate h^2 developed by Holzinger (1937) is not valid because Holzinger's derivation of it was unsound: The variance component model he used to derive h^2 , together with his claim that it estimates the genetic variance ratio, imply the counterfactual assertion that dizygotic twins share no genes. While a competing coefficient, Nichols' HR , does indeed follow from the conventional variance component model, the necessary conditions it rests on lack empirical face validity.

1. Introduction

Heritability estimates intended to estimate the genetic variance ratio

$$v(g,e) := \text{var}(g) / [\text{var}(g) + \text{var}(e)] \quad (1.1)$$

— where g is a genetic random variable and e an environmental random variable which, possibly together with some other variable, are presumed to account for an observed variable y —are widely used in behavior genetics. Among other things, they provide the operational basis for Jensen's (1970) controversial claim that "The heritability of the IQ ... comes out to about 80 percent, the average value obtained from all relevant studies now reported" (p. H6325).

One of the first authors to attempt a psychometrically motivated index of heritability was Holzinger, a coauthor of one of the earliest systematic twin studies (Newman, Freeman, and Holzinger, 1937). He proposed the coefficient

$$h^2 := (\rho_M - \rho_D) / (1 - \rho_D) \quad (1.2)$$

as an index of heritability based on the intraclass correlations ρ_M for monozygotic twins raised together (MZs) and ρ_D for dizygotic twins raised together (DZs), in the mistaken belief that it equals the heritability ratio (1.1) under some plausible assumptions about the genetic variable g and the environmental variable e .

Holzinger's derivations (pp. 94-116) of this mistaken claim are not easy to follow because he left several critical covariance assumptions implicit. As a result, "the precise nature of the inadequacy of the [h^2] index have remained conceptually ob-

* This Note was completed while the author was on Sabbatical leave from the Department of Psychological Sciences, Purdue University, West Lafayette, IN 47906, USA. I wish to thank the Psychology Departments of the National Taiwan University and the National Chengzhi University for co-sponsoring my Visiting Research Professor appointment and the National Science Council for its financial support.

score" (Jensen, 1967, p. 149) for some 30 if not 50 years. However, his endresult, h^2 , has been criticized by several authors before. In particular, it has been noted repeatedly that Holzinger's h^2 (a) has no provision for measurement error, and (b) in effect ignores the between sums of squares, since it can be computed solely in terms of mean squares within for the MZs and DZs.

What has apparently not been noticed before is that Holzinger's model, together with the assumption that $h^2 = v(g,e)$, imply the counterfactual condition that dizygotic twins share no genes.

2. Twin data, intraclass correlations, and mean squares

If $Y := (y_{ik})$ is an $N \times 2$ table of scores for N twin pairs, and $\bar{y} := \sum_k \sum_i y_{ik} / 2N$ is the overall mean, then the conventional sums of squares are given by

$$\begin{aligned} SSB &= \sum_i (y_{i1} + y_{i2} - 2\bar{y})^2 / 2 \\ SSW &= \sum_i (y_{i1} - y_{i2})^2 / 2 \end{aligned} \quad (2.1)$$

i.e., by half the sums of squares of the row sums and row differences of the deviation score matrix around the grand mean, as was shown in more detail in (Schönemann, 1990). Note that the experimental units are twin pairs, not twins. Since SSB is a quadratic form of rank $N-1$, the mean square between is $MSB = SSB/(N-1)$. Similarly, SSW is a quadratic form of rank N so that the mean square within is $MSW = SSW/N$.

In the population, an intraclass correlation can be defined in terms of expected mean squares as

$$\rho := E(MSB - MSW) / E(MSB + MSW) \quad (2.2)$$

This definition is usually employed in twin studies and will be used throughout this note. Alternatively, one can define an intraclass correlation in terms of expected sums of squares. For any reasonable sample size N , the numerical difference between both definitions is negligible. A (not necessarily best) sample estimate r results on replacing the expected mean squares by their sample estimates. In Schönemann (1991), it was shown that a simple transform of r has exact central F -distribution so that, at least in principle, it is not necessary to invoke Fisher's z -transform for non-null tests.

3. Holzinger's derivation of h^2

Although Holzinger's derivation of his heritability ratio h^2 lacks stringency because he did not make all his assumptions explicit, they can be inferred from his conclusions.

He started out by postulating a variance component model for the MZs in terms of 5 latent variables, a genetic variable g , two environmental variables e_1 , e_2 , and two measurement variables d_1 , and d_2 :

$$\begin{aligned} \text{MZs: } y_1 - \mu &= g + e_1 + d_1, \\ y_2 - \mu &= g + e_2 + d_2, \end{aligned} \quad (3.1)$$

where

$$\mu = E(y_k), \quad E(g) = E(e_k) = E(d_k) = 0, \quad k=1,2 \quad (3.2)$$

and

$$\text{Var}(g, e_1, e_2, d_1, d_2) = \text{diag}[\text{var}(g), \text{var}(e), \text{var}(d), \text{var}(d)] \quad (3.3)$$

i.e., all latent variables are uncorrelated, and

$$\text{var}(e_k) = \text{var}(e), \quad \text{var}(d_k) = \text{var}(d), \quad k=1, 2. \quad (3.4)$$

At this early stage already, he planted the seed for later troubles: although Holzinger intended his model "for evaluating the relative influence of nature and nurture for fraternal twins reared together" (p.111), he assigned two different, uncorrelated environmental variables, e_1, e_2 (B, B' in his notation), to the two twins in each pair. As a result, $\text{var}(e)$ and $\text{var}(d)$ become confounded. Presumably for this reason, Holzinger discarded the measurement error variables d_k halfway through his derivations (p.113) by setting $\text{var}(d)=0$. He, therefore, in effect, worked with the variance component model

$$\text{MZs: } y_1 - \mu = g + e_1, \quad y_2 - \mu = g + e_2. \quad (3.5)$$

Since

$$y_1 + y_2 - 2\mu = 2g + e_1 + e_2, \quad y_1 - y_2 = e_1 - e_2 \quad (3.6)$$

and (3.1)-(3.4) imply

$$E[(y_1 + y_2 - 2\mu)^2] = 4\text{var}(g) + 2\text{var}(e), \quad E[(y_1 - y_2)^2] = 2\text{var}(e) \quad (3.7)$$

one finds

$$E[\sum_i (y_{i1} + y_{i2} - 2\mu)^2] = (N-1)[4\text{var}(g) + 2\text{var}(e)] = 2(N-1)E[MSB], \quad (3.8)$$

$$E[\sum_i (y_{i1} - y_{i2})^2] = 2N \cdot \text{var}(e) = 2N \cdot E[MSB] \quad (3.9)$$

by (2.1), so that the expected mean squares are

$$\begin{aligned} E(MSBM) &= 2\text{var}(g) + \text{var}(e), \\ E(MSWM) &= \text{var}(e). \end{aligned} \quad (3.10)$$

By (2.2), the intraclass correlation for the MZs is therefore

$$\rho_M = \text{var}(g) / [\text{var}(g) + \text{var}(e)]. \quad (3.11)$$

It thus emerges that, under his model, the genetic variance ratio $v(g, e)$ (1.1) is

simply given by the intraclass correlation of the MZs, a fact Holzinger overlooked. Instead, he went on to postulate a variance component model for the DZs to arrive at his definition of h^2 , (1.2). After dropping again the measurement error variables d_k , he described the dizygotic twins with

$$\text{DZs: } y_1 - \mu = g_1 + e_1, \quad y_2 - \mu = g_2 + e_2. \quad (3.12)$$

His derivations imply that he again assumed all variables on the right of the equality signs to be uncorrelated within twins and e_1, e_2 to be uncorrelated between twins. He also assumed that $\text{var}(g_k) = \text{var}(g)$, and that $\text{var}(g), \text{var}(e)$ are the same for both twin types. However, Holzinger did not explicitly state the "genetic correlation" between the two genetic variables g_1, g_2 . Suppose it is given by

$$\text{cor}(g_1, g_2) =: \rho_{12}. \quad (3.13)$$

Eqs. (3.12), (3.13) and the stated covariance assumptions then imply that the mean squares of the DZs are

$$\begin{aligned} E(\text{MSBD}) &= (1 + \rho_{12}) \cdot \text{var}(g) + \text{var}(e), \\ E(\text{MSWD}) &= (1 - \rho_{12}) \cdot \text{var}(g) + \text{var}(e) \end{aligned} \quad (3.14)$$

and their intraclass correlation is

$$\rho_D = \rho_{12} \text{var}(g) / [\text{var}(g) + \text{var}(e)]. \quad (3.15)$$

4. Counterfactual implication of Holzinger's derivations: $\rho_{12} = 0$

Theorem: The variance component model (3.4)-(3.5), (3.12) together with the assertion that $h^2 = v(g, e)$ imply that the genetic correlation ρ_{12} between g_1, g_2 is indeterminate or zero.

Proof: By (1.2), (3.11), and (3.15), Holzinger's h^2 can be expressed in terms of $\rho_M, \rho_D, \rho_{12}$ as

$$h^2 = (\rho_M - \rho_D) / (1 - \rho_D) = (1 - \rho_{12}) \text{var}(g) / [(1 - \rho_{12}) \text{var}(g) + \text{var}(e)]. \quad (4.1)$$

On solving the equality

$$h^2 = v(g, e) =: \text{var}(g) / [\text{var}(g) + \text{var}(e)] \quad (4.2)$$

for ρ_{12} , one arrives at

$$\rho_{12} \cdot \text{var}(g) \text{var}(e) = 0 \Leftrightarrow \text{var}(g) \text{var}(e) = 0 \text{ or } \rho_{12} = 0, \text{ qed.} \quad (4.3)$$

By the common elements correlation formula (e.g., Hogben, 1950, p. 365f), the condition $\rho_{12} = 0$ is equivalent to the counterfactual assertion that dizygotic twins share no genes. However according to standard genetic theory, such twins share half the genes on average, so that, in fact, $\rho_{12} = 1/2$.

In passing, note that (4.3) and $\text{var}(g) \cdot \text{var}(e) \neq 0$ imply that under Holzinger's

assumptions the intraclass correlation ρ_D (3.15) for dizygotic twins is identically zero and the denominator of h^2 (1.2) is identically 1.

5. Related heritability coefficients: Nichols' HR

Holzinger's coefficient h^2 is not the only statistic based on intraclass correlations which has been used in behavior genetics to estimate the genetic ratio $v(g, e)$, but it is the oldest and by far the most widely used heritability estimate. As a consequence, most assessments of the relative contributions of genetic and environmental causes to mental performance, including Jensen's 80% figure for IQ, are probably based on Holzinger's h^2 .

Nichols (1965) has proposed another heritability index HR based on intraclass correlations,

$$HR = 2(\rho_M - \rho_D) / \rho_M \quad (5.1)$$

which, in contrast to Holzinger's h^2 , can indeed be derived as the heritability ratio $v(g, e)$ under a suitably chosen variance components model. This model results from Holzinger's by making two changes: (a) e_1 must be set equal to e_2 , because both sets of twins are supposed to be reared in the same environment, and (b) ρ_{12} must be set to 1/2 to be consistent with standard genetic theory. One thus obtains the revised model

$$\begin{aligned} \text{MZs: } y_1 - \mu &= g + e + d_1, \quad y_2 - \mu = g + e + d_2, \\ \text{DZs: } y_1 - \mu &= g_1 + e + d_1, \quad y_2 - \mu = g_2 + e + d_2 \\ \text{cov}(g_1, g_2) &= \text{var}(g)/2, \end{aligned} \quad (5.2)$$

which implies the expected mean squares

$$\begin{aligned} E(MSBM) &= 2 \cdot \text{var}(g) + 2\text{var}(e) + \text{var}(d), \quad E(MSWM) = \text{var}(d) \\ E(MSBD) &= 1.5 \cdot \text{var}(g) + 2\text{var}(e) + \text{var}(d), \\ E(MSWD) &= .5 \cdot \text{var}(g) + \text{var}(d) \end{aligned} \quad (5.3)$$

and, thus, the intraclass correlations

$$\begin{aligned} \rho_M &= [\text{var}(g) + \text{var}(e)] / [\text{var}(g) + \text{var}(e) + \text{var}(d)] \\ \rho_D &= [.5 \cdot \text{var}(g) + \text{var}(e)] / [\text{var}(g) + \text{var}(e) + \text{var}(d)]. \end{aligned} \quad (5.4)$$

On substituting (5.4) into (5.1), one finds

$$HR = \text{var}(g, e), \quad (5.5)$$

so that Nichols' coefficient HR is indeed a valid deduction from the revised variance component model (5.2).

Empirically, one finds that HR is typically much larger than h^2 and often exceeds 1. For example, for the personality data reported by Osborne (1980, P. 127, 135), 74% of 80 HR are inadmissible, 60% exceeding 1. This suggests that the variance

component model needed to justify HR is often empirically incompatible with the data.

6. Discussion

As Henderson (1982) has observed, "The assumptions behind an analysis can sometimes have a major impact on the estimation of parameters" (p.413). It is, therefore, important that these assumptions be explicitly stated and empirically tested before drawing substantive inferences from the parameter estimates. Regrettably, this has not been the custom in human behavior genetics. This is regrettable, "Because human behavior studies are perceived to have implications for public policy", as Henderson (1982, P. 404) correctly noted. For example, between 1921 and 1964, 33,374 US citizens were sterilized on the grounds that they were "mentally retarded", i.e.. because they scored below 70 on "intelligence tests" (Robitscher, 1973, p. 123). Presumably, this was done because the authorities trusted the judgment of social scientists, biologists, and statisticians who told them that "intelligence" is 80 percent inherited.

Henderson wonders "where the earlier figure of 80% came from" (p. 409, loc. cit.). So far as I can tell, it came from twin studies employing heritability coefficients which have remained "conceptually obscure" for many decades, based on models which have never been adequately tested. According to Henderson, "Today an estimate of 50% seems more in vogue" (p. 411). These downward revised estimates are now obtained with "modelling procedures [which] use some form of maximum likelihood estimation or weighted least squares method" (p. 413). For most social scientists, these supposedly more sophisticated methods are conceptually even more obscure than the simple variance component model underlying Holzinger's coefficient h^2 .

The prepackaged programs used to obtain such MLEs and LSEs deliver global test of fit criteria which can mask serious and systematic violations of the underlying model. A better strategy would seem to be to study the assumptions of these models with sufficient care to be able to derive specific predictions which can be tested individually. To illustrate, a recent reanalysis of the twin data by Shields (1962) revealed that three types of specific predictions of a model proposed by Jinks and Fulker (1970) were violated by the data in more than 90% of the cases tested (Schönemann, 1990). Jinks and Fulker, relying on the overall least squares fit, had found no reason to reject this model. This is not surprising, because the sample sizes were much too small to reject it statistically. (See also Schönemann and Schönemann, 1991).

In the present case, we found that Holzinger's h^2 is theoretically unsound, while Nichols' HR, though theoretically sound, often conflicts empirically with the data. Taken together, these two results suggest, at the very least, that past pronouncements about relative contributions of nature and nurture to mental performance, including IQ, should be viewed with caution.

REFERENCES

- HENDERSON, N. D. (1982). Human behavior genetics. *Annual Review of Psychology*, 33, 403-440.
- HOGBEN, L. (1950). *Chance and choice by cardpack and chessboard*, Vol. I. London: Parrish.
- JENSEN, A. R. (1967). Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. *Proceedings of the National Academy of Sciences*, 58, 149-159.
- JENSEN, A. R. (1970). Statement of Dr. Arthur A. Jensen before the General Subcommittee on Education, House Education and Labor Committee. Washington, *Congressional Record*, I-237.
- JINKS, J. L. and FULKER, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, 73, 311-349.
- OSBORNE, T. (1980). *Twins; Black and white*. Athens, GA: Foundation for human understanding.
- NEWMAN, H., FREEMAN, F., and HOLZINGER, K. (1937). *Twins. A study of heredity and environment*. Chicago: University of Chicago Press.
- NICHOLS, R. C. (1965). The National Merit twin study. In G. S. Vandenberg (ed) *Methods and goals in human behavior genetics*. New York: Academic Press.
- ROBITSCHER, H. (1973). *Eugenic Sterilization*. Springfield: Thomas.
- SCHÖNEMANN, P. H. (1990). Environmental versus genetic variance component models for the identical twin data by Shields. *Cahiers de Psychologie Cognitive-European Bulletin of Cognitive Psychology*, 10, 451-473.
- SCHÖNEMANN, P. H. (1991). On non-null tests of intraclass correlations. *Chinese Journal of Psychology*, 33, 3-10.
- SCHÖNEMANN, P. H. and Schöнемann, R. D. (1991). We wondered where the errors went. *Behavior and Brain Sciences*, 14, 404-406.
- SHIELDS, J. (1962). *Monozygotic twins brought up apart and brought up together*. London: Oxford University Press.

初審收件：1992年8月3日

審查通過：1993年1月2日

對 Holzinger 遺傳係數 h^2 之評註

舒比德

國立台灣大學心理學系

本文指出 Holzinger (1973) 所發展而廣為使用的遺傳係數 h^2 並無確實理論根據，因其推演過程不嚴謹。Holzinger 所用以推導 h^2 的變異數成分模式和其認為 h^2 估計遺傳變異數比值的說法隱含了一個違反事實的論點，即異卵雙生子無任何基因相同。而另一係數，Nichols 的 HR ，雖由傳統的變異數成分模式推導所得，然其必要條件卻缺乏實徵上的表面效度。