Holzinger's H$_C$ Revised

Gordon Allen

Division of Biometry and Epidemiology, National Institute of Mental Health, US Department of Health, Education and Welfare, Rockville, Maryland

The formula $H'_C = 1 - C_{DZ}/C_{MZ}$ is suggested as a better summary of twin concordance data than the familiar Holzinger concordance formula, $H_C = (C_{MZ} - C_{DZ})/(1 - C_{DZ})$. The new formula better estimates degree of genetic determination, G, as calculated from a threshold model, but never exceeds unity, as G sometimes does. For high concordance rates, if $C_{MZ} + C_{DZ} > 1$, H_C may be more useful than either G or H'_C.

Key words: Heritability estimates, Twin concordance

The following formula for summarizing twin concordance data is attributed, perhaps erroneously, to K. J. Holzinger:

$$H_C = \frac{C_{MZ} - C_{DZ}}{1 - C_{DZ}} \tag{1}$$

Here, C_{MZ} and C_{DZ} are the proportions of concordant twins, not precisely defined, in the monozygotic (MZ) and dizygotic (DZ) classes, respectively. This is similar in form to Holzinger's 1929 index, H, for twin correlations in quantitative traits [1], with direct replacement of correlation coefficients by concordance rates. Neither of these formulas has a simple genetic meaning, and twin concordance bears no simple mathematical relation to correlation.

The most useful genetic working hypothesis to adopt with quantitative traits of unknown causation is one of polygenic control with completely additive effects. Under this model, one can meaningfully refer to heritability in the narrow sense and to degree of genetic determination. With human data, however, the model must be used with caution, because it requires random environmental effects within families, including DZ twin pairs.

Quantitative traits of unknown genesis can be reduced to the polygenic additive model by assuming a threshold phenomenon interacting with a continuum of liability. Smith [2] has provided a table which facilitates the translation of proband concordance rates into an index of genetic determination, G. A proband concordance rate is given by the ratio of

probands in concordant pairs to probands in concordant and discordant pairs together. Smith defines G as $2(r_{MZ} - r_{DZ})$. Applied within this model to concordance rates, r_{MZ} and r_{DZ} are the correlations implied by observable concordance rates.

Smith's model is not easily comprehended by nongeneticists, and a table that requires interpolation is less likely to be used than a formula. Moreover, the validity of the index depends upon so many assumptions that the necessary calculations may not be justified if a simpler index can be used as well. The range of values of Smith's G is unreasonable for many ordinary combinations of MZ and DZ concordance rates. Given a trait frequency of 10% and a DZ concordance rate of 28%, G becomes 1.0 at an MZ concordance rate of only 75%. With similar data but a DZ concordance rate of 20%, G has the impossible value, 1.3. At the other extreme, even at $C_{MZ} = 100\%$, G falls far short of unity for many moderate values of C_{DZ}. If r_{DZ} is more than half as great as r_{MZ}, Smith's formula will never indicate a heritability of unity, and in these cases, if the concordance rates are high, Holzinger's formula may more accurately represent the degree of genetic determination.

Even in the last-mentioned situation, however, H_C has no theoretical justification. At ratios of r_{MZ} to r_{DZ} of 2 or greater, H_C *always* underestimates G, and for lower ratios it usually does so.

A formula rather similar to Holzinger's original yields a better approximation of G in nearly all circumstances:

$$H'_C = \frac{C_{MZ} - C_{DZ}}{C_{MZ}} = 1 - \frac{C_{DZ}}{C_{MZ}} \tag{2}$$

Unlike G, this is always within the range of 0 to 1. H'_C stops short of 1, however, if there is any measurable concordance in DZ twins. The range can be extended to 1 by using Holzinger's unmodified index, Eq. (1), when the sum of the concordance rates exceeds unity, the value for which $H_C = H'_C$.

With the aid of Table 2 in Smith [2], H_C and H'_C can be plotted on the ordinate of a graph with corresponding values of G as the abscissa (Figure). A perfect estimator of G would occupy the 0 to 1 diagonal, and if $r_{MZ}/r_{DZ} \geq 2$, H'_C always intersects the diagonal at a low angle. The form of the locus of H'_C varies with the frequency of the trait in the population as well as the ratio of r_{MZ} to r_{DZ}. Given r_{MZ}/r_{DZ} ratios less than 2, G attains a value of 1.0 only for MZ concordance rates greater than 100%, while H'_C intersects G only for traits of rather high frequency, as detailed below.

One can also plot a band of values above and below H'_C, perhaps most appropriately $H'_C \pm pq$ (letting $p = H'_C$ and $q = 1 - p$). The intersections of such boundaries with the diagonal yield for each set of conditions a range of usefulness for H'_C as an estimator of G. In an actual case, however, the conditions are not known unless concordance rates are converted to correlations by use of Smith's table, so the ranges are only of academic interest.

The following Table shows seven points within the usual range of concordance data where H'_C intersects G, together with the range, in each case, of values of H'_C as close to G as $H'_C \pm pq$. The left value in each range, if not zero, is an overestimate of G; the right value, if not unity, is an underestimate. When the right-hand limit is given as unity, the maximum value of H'_C is less than G, but not by more than pq.

Also shown for the points of coincidence are the corresponding ratios, C_{MZ}/C_{DZ}, as a warning that this ratio is quite unlike the ratio of the correlations given in the left margin of the table. As an illustration of the actual concordance rates at coincidence of H'_C and G,
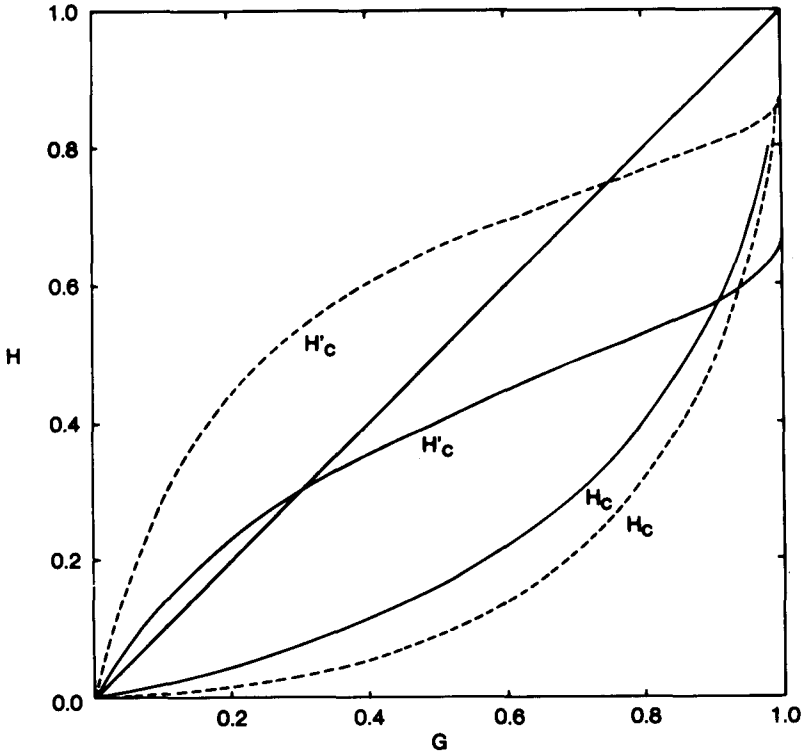
Figure. H_C *and* H'_C *plotted against corresponding values of Smith's G. Calculations assume* $G = r_{MZ} = 2r_{DZ}$. *Population frequency of the trait is 1.0% for the broken curves, 10% for the solid curves. The diagonal represents a perfect estimator of G.*

TABLE. Some Points of Coincidence (underlined) Between Smith's G and H'_C; *the Corresponding Range in Which* $|G - H'_C|$ *Is Less Than* $H'_C (1 - H'_C)$; *and (below) the Ratio of* C_{MZ}/C_{DZ} *at the Coincidence Point*

r_{MZ}/r_{DZ}	Frequency of the trait in the population		
	0.1%	1.0%	10%
3.0	0.93−0.94−1.00	0.73−0.82−0.86	0.00−0.35−0.56
	17	5.6	1.5
2.0	0.88−0.92−1.00	0.46−0.75−0.84	0.00−0.30−0.52
	12	4.0	1.4
1.5	( )−( )	0.00−( )−1.00	0.00−0.20−1.00
			1.2

$C_{MZ} = 32\%$ and $C_{DZ} = 8\%$ when the ratio of correlations is 2 and the trait frequency is 1%.

On the bottom row of the Table only one coincidence point is shown. At values of r_{MZ}/r_{DZ} as small as 1.5, the diagonal is cut off at $G = 0.667$, concordance then being maximal, $C_{MZ} = 100\%$. The intersection of H'_C with G is also cut off when it lies beyond this

point. At this ratio and a trait frequency of 0.1%, H'_C exceeds G by a quantity $\geqslant$ pq, throughout the possible range; consequently, no usefulness range is shown in the Table. At a trait frequency of 1.0%, H'_C exceeds G throughout the possible range by ½ pq. At a trait frequency of about 4.0%, H'_C intersects G just at the upper limit, 0.667, and at a trait frequency of 10% the intersection is already as low as G = 0.2. H'_C then has a maximum of 0.55, below that of G, but H_C surpasses H'_C at a value of 0.42 and surpasses G at a value of 0.625. Under these conditions, at the lower right corner of the Table, H'_C is always a good estimator of G, and where those indices both give implausibly low estimates of genetic determination, H_C becomes useful.

At all trait frequencies much less than 1.0%, the Holzinger coefficients have limited usefulness as estimators of G. However, rare traits are less likely to be polygenic than are common traits, so that for these traits G may be a poor reference, and no good genetic model for concordance is available. H'_C is not in fact proposed as an estimator of G; when a test of the polygenic model is desired, Smith's table or equivalent calculations should be used. Rather, H_C and H'_C provide a convenient index in their own right, being less arbitrary than H_C alone, being defined in terms of G by Smith and by the present data, and having a more useful range than G.

In summary, H_C and H'_C, when used appropriately, afford great simplicity of calculation, a range between 0 and 1 for all twin data in which C_{MZ} exceeds C_{DZ}, and a fairly consistent relation to the index G, in those circumstances where a polygenic mechanism is to be expected. It should be emphasized, however, that with twin data we are usually not estimating heritability, or even the degree of genetic determination. These indexes, H_C and H'_C, and even G, when applied to twin data, should be regarded as crude statements of the relative degree of *familial* determination.

REFERENCES

1. Holzinger KJ (1929): The relative effect of nature and nurture on twin differences. J Educ Psychol 20:241−248.
2. Smith C (1974): Concordance in twins: Methods and interpretation. Am J Hum Genet 26:454−466.

Correspondence: Dr. Gordon Allen, NIMH, 5600 Fishers Lane, Rockville, MD 20857, USA.