Acta Genet Med Gemellol 33:557-563 (1984)

Received 3 November 1983
Final 23 April 1984

Sense and Nonsense in Genetic Epidemiology: A Critique of the Statistical Model of Williams and Iyer

A.C. Heath, N.G. Martin, L.J. Eaves
Department of Human Genetics, Medical College of Virginia, Richmond

Abstract. The statistical model for the analysis of twin-family data of Williams and Iyer is examined. The model uses a large number of redundant parameters. It does not lead to quantitative predictions for new relationships. It allows for epistasis but not dominance. It makes assumptions about assortative mating which are inconsistent with any biologically plausible mechanism. It assumes that the environmental correlation between parent and offspring is due to the direct effect of the parental genotypes, not phenotypes, on offspring environment. Other models which avoid these problems are more appropriate for the analysis of extended twin-family data.

Key words: Genotype-environment models, Twin-family designs, Analysis of variance

INTRODUCTION

A statistical model for the analysis of genetic and environmental effects from data on pairs of adult twins and their spouses and offspring ('twin-family' data) has been discussed briefly by Crumpacker et al [7] and developed in more detail by Williams and Iyer [30]. This model has been used in analyses of Swedish twin family data on the familial transmission of personality [22]. Unfortunately the properties of their model are such that it is quite unsuitable for any form of data analysis. Without attempting a comprehensive review, we illustrate some of the more important deficiencies of the model here.

A good model of cultural and biological inheritance should have the following properties:

1) Parsimony – as few parameters as possible are used to encapsulate the assumptions of the model;

This is paper #222 from the Department of Human Genetics at the Medical College of Virginia. Research for this paper was supported by NIH Grants #GM30250 and #HL28922.

2) Quantitative prediction – the model can not only be shown to fit a given set of data, but also leads to quantitative predictions which can be tested using completely different data sets (eg, adoption data rather than twin data);
3) Strong theoretical basis – the model should be derived from assumptions about the underlying mechanisms of cultural and biological inheritance, rather than using atheoretical empirical parameters.

From our discussion of the treatment of gene action, assortative mating and environmental transmission in the paper of Williams and Iyer, it will be seen that these criteria are not satisfied by their model.

GENE ACTION

Fisher [11] first showed how it is possible to represent parsimoniously the contribution of additive gene action, dominance and epistasis to the correlation between relatives for a character which is determined by several or many loci of small effect. The approach to be used for a trait for which mating is random has been discussed extensively in elementary texts on quantitative genetics [eg, 10, 21]. The total variance in a character attributable to gene action (VG) may be expressed as a sum of components of variance representing the independent contributions of the additive effects of both genes at all loci (VA), the allelic interactions between additive effects (ie, the dominance deviations) at all loci (VD), and the nonallelic interactions (epistasis) between pairs of additive effects at all pairs of loci (VAA), between the additive effect of a gene at one locus and the dominance deviation at a second locus, for all pairs of loci (VAD), and between the dominance deviations at all pairs of loci (VDD). Expressions for these components of variance in terms of average gene effects at individual loci or pairs of loci have been given [18,20]. We may define further components representing the effects of trigenic and higher-order interactions [20], but the detection of such subtleties in man will never be possible. Further subtleties which may be allowed for include sex-limited gene action and sex-linkage [19,20].

Although the resolution of dominance and digenic interaction components of variance in humans is feasible in theory, a computer simulation study has shown that this would require enormous sample sizes to be realizable in practice [13]. Since studies of inbreeding depression have demonstrated directional dominance for some characters [1,28], it is usually assumed that any genetical non-additivity is due to dominance. In human studies, therefore, we may express the contribution of genetic factors to the correlation between relatives for a character for which mating is random in terms of only two parameters, VA and VD. A model which allows for dominant gene action without additive effects would be biological nonsense [20]. We may, therefore, fit models which include both parameters, or the parameter VA but not VD, but not models which include only the parameter VD. Quantitative predictions for correlations between new sets of relatives are easily derived from estimates of VA and VD obtained during data analysis, once the appropriate coefficients for these relationships have been derived (see eg, 6, 17).

Expectations for the genetic correlations between relatives for those relationships which constitute the twin-family study are given in the Table. For any given relationship, the coefficients for the parameters VAA, VDD, and VAD will be the square of the coefficient given for VA, the square of the coefficient for VD, or the product of the coefficients given for VA and VD in the Table. Even if we should wish to include epistasis in

our model, therefore, when using twin family data the parameters VD, VAD, and VDD will be completely confounded.

Table - Coefficients for Additive Genetic and Dominance Components of Variance under Random Mating

Relationship	VA	VD
MZ twins	1	1
DZ twins, full siblings	1/2	1/4
Parent-offspring, MZ twin-cotwin's offspring	1/2	0
MZ half-siblings, DZ twin-cotwin's offspring	1/4	0
First cousins	1/8	0
Spouses, affine relatives	0	0

The Williams-Iyer model (WIM) uses five separate parameters to represent the contribution of genetic factors to the resemblance of twins and their offspring. At least two of these parameters, and three if we ignore epistasis, would be made redundant by an appropriate reparameterization of the model. In addition to the parameter h^2, which corresponds to the additive genetic variance (VA) of classical quantitative genetics, separate parameters γ_0, γ_1, γ_2, and γ_3 are used to denote the contribution of nonadditive gene effects (dominance or epistasis) to the correlation between MZ twins, full siblings, half-siblings, and first cousins, respectively. These four latter parameters are atheoretical parameters which have no simple meaning in terms of gene action. We cannot, therefore, use estimates of these parameters to make quantitative predictions for new relationships. Under WIM, the coefficient for the nonadditive genetic parameter γ_2 is the same for parent and offspring as for DZ twins or siblings. As can be seen from the Table, the coefficient for the dominance component of variance VD is 1/4 for siblings, but zero for parents and offspring and other intergenerational blood relationships. Use of WIM, therefore, implies the assumption that any genetical nonadditivity is not due to dominance!

ASSORTATIVE MATING

Fisher [11] extended his treatment of the contribution of genetic factors to the correlation between relatives to allow for the effects of various types of assortative mating. Unfortunately, Fisher's treatment of assortative mating is notoriously difficult, and Williams and Iyer [30] have evidently misunderstood it. Linear regression theory allows a much simpler treatment of the models of assortative mating considered by Fisher, and allows simple extension of these models to incorporate environmental transmission as well as additive gene action and dominance [2,4,5,13,27]. We have discussed linear models of assortative mating in detail elsewhere [8,13] and so here we will only summarize briefly certain pertinent conclusions.

Two major classes of models of assortative mating have been developed, corresponding to the first and third models considered by Fisher in his original paper. (The second model considered by Fisher is equivalent to his third model under the simplifying assumption of no dominance implied by WIM). The first class of models, the "phenotypic" models, assumes that assortative mating is based on some aspect of the individual's phenotype, either the character measured in a particular study, or the same character corrected

for measurement error, or some other correlated variable such as the individual's educational level or socioeconomic status [13,16]. We may express an individual's phenotypic value as a function of the underlying genotypic and environmental values. An important implication of all the "phenotypic" models is that under assortative mating the cross-correlation between the genotypic value of one spouse and the environmental value of the other, or vice versa, will be non-zero. This will be true even when genotypic and environmental values are statistically independent within-individuals, under which circumstance there should be a direct correspondence between WIM and the linear regression theory models cited above. A second implication is that the expected correlation between the spouses of twins or between one twin and the cotwin's spouse should be a function of the true marital correlation for the character on which assortative mating is based and the expected value of the phenotypic correlation between twins for that character.

Williams and Iyer clearly believe that their model makes assumptions about assortative mating which are equivalent to those used by Fisher. They use such terms as "phenotypic assortment" (30, p. 13) and specifically compare their analyses of height data with the results which Fisher obtained using his first model. (Their claim that Fisher estimated that 27% of the additive genetic variation in height was attributable to assortative mating is incorrect. The correct figure is 17%). Under WIM, however, the cross-correlations between the additive genetic and environmental values of spouses are constrained to be equal to zero, which contradicts the prediction of Fisher's first model. Under WIM, too, the expected correlation between one twin and the cotwin's spouse is not equal to the product of the expected twin and marital correlations, nor is the expected correlation between the spouses of twins equal to the product of the expected twin and the square of the marital correlations. Both of these latter predictions are implied by Fisher's first model [3-5,13].

The second class of models of assortative mating considered in the literature, the 'social homogamy" models, assume that assortative mating is based either on sibling or parental phenotypes or on social background, where social background is a function of the phenotypes of the parents of spouses rather than the phenotypes of the spouses themselves [13]; or else on some aspect of the phenotypes of spouses which is determined by the same genetic or environmental factors as the character under investigation [3,24]. Under these models the expected correlations between one twin and the cotwin's spouse, and between the spouses of twins, are no longer simple functions of the expected twin and marital phenotypic correlations. Under these models, too, the expected cross-correlation between the additive genetic value of one spouse and the environmental value of the other can be zero. However, this can only occur when either the correlation between the additive genetic values of spouses or else the correlation between the familial environmental values of spouses is also zero. This constraint is not imposed by WIM, which is, therefore, inconsistent with these models. We can think of no other biologically plausible mechanism of assortative mating which could lead to the pattern of expected correlations predicted by WIM.

ENVIRONMENTAL TRANSMISSION

Wright [31] was the first to make allowance in the same model for the contribution of both genetic and environmental factors to the correlation between relatives, using path

analysis, a form of linear regression theory. Wright's work was later rediscovered and used extensively by Rao et al. [23-26] and, subsequently, by Loehlin [15] and many others. A variety of different mechanisms of environmental transmission from parent to offspring have been considered, including:

1) A direct effect of the parental genotypes on the environment of their offspring, which is not mediated through the parental phenotypes ("G-E" transmission; see eg, 12);
2) A direct effect of the childhood environments of the parents on the environment of their offspring, which is not mediated through the parental phenotypes ("E-E" transmission, see eg, 4, 5, 26);
3) A direct effect of the parental phenotypes on the environment of their offspring ("P-E" transmission; see eg, 9, 26).

As we have shown elsewhere [13], the collection of data on adult MZ and DZ twins and their spouses and offspring is one of the most powerful ways of resolving such competing hypotheses about environmental transmission. This is most easily seen if we make the simplifying assumption of random mating. Under pure G-E transmission, the expected correlation between an MZ twin and his cotwin's child will be identical to the expected correlation between parent and offspring, since the twins share a common genotype. This prediction will also be true under E-E transmission, since the twins also share a common childhood environment.

Under E-E transmission, but not G-E transmission, however, the contribution to the resemblance of a DZ twin and his cotwin's offspring will be the same as for parent and offspring. Finally, if there is any P-E transmission the expected correlation between parent and offspring will be greater than the expected correlation between an MZ twin and his cotwin's offspring, unless the phenotypic correlation between MZ twins is unity!

WIM uses a total of nine parameters to represent the contribution of environmental factors to the correlation between twins and their relatives. These are atheoretical parameters which are specific to one or at most two different relationships. Estimates of these parameters cannot, therefore, be used to make quantitative predictions for new relationships. Four of these parameters represent prenatal effects which are completely confounded with postnatal effects if only data on MZ and DZ twins and their offspring are available. Although WIM assumes that any excess of the mother-offspring correlation over the father offspring correlation is attributable to prenatal effects, there is no way of testing this somewhat unlikely assumption with this experimental design. WIM also assumes that environmental factors make identical contributions to the following pairs of correlations:

1) The correlation between parent and offspring and the correlation between an MZ twin and his cotwin's offspring;
2) For postnatal environmental factors only, the correlation between a DZ twin and his cotwin's offspring and the correlation between MZ half-siblings (cousins related through MZ twin parents).

Assumption (1) implies that there is no direct effect of parental phenotype on offspring environment, as we have seen above. Assumption (2) is bizarre but is presumably made because Williams and Iyer believe that, except in the case of siblings and parent and offspring, under G-E transmission the contribution of environmental factors to the resemblance of relatives of any type is solely determined by their degree of genetical relatedness! This will not in general be true. Under most biologically plausible circum-

stances the nine parameters required to represent environmental resemblance between relatives in WIM may be replaced by as few as two or three parameters which may not only be used in the analysis of extended twin data, but may also be used to make quantitative predictions for other relationships.

CONCLUSIONS

As we have argued in detail elsewhere [13], collection of data on pairs of adult MZ and DZ twins and their spouses and offspring, particularly when supplemented by data on the parents and parents-in-law of the twins, provides one of the most powerful experimental designs available for testing alternative hypotheses about gene action, environmental transmission and assortative mating. Expected correlations between twins and their spouses and offspring have been published for a variety of competing models [4,5,13,24]. All these models are based on theoretical biological and psychological considerations which can be used not only to explain existing findings, but also to predict novel results [29]. By contrast, WIM offers only an atheoretical description of existing findings, and, therefore is of little predictive use.

REFERENCES

1. Bashi J (1977): Effects of inbreeding on cognitive performance of Israeli Arab children. Nature 266:440-442.
2. Bulmer MG (1980): The Mathematical Theory of Quantitative Genetics. Oxford: Clarendon Press.
3. Cloninger CR (1980): Interpretation of intrinsic and extrinsic structural relations by path analysis: theory and applications to assortative mating. Genet Res 36:135-145.
4. Cloninger CR, Rice J, Reich T (1979a): Multifactorial inheritance with cultural transmission and assortative mating. II. A general model of combined polygenic and cultural inheritance. Am J Hum Genet 31:176-198.
5. Cloninger CR, Rice J, Reich T (1979b): Multifactorial inheritance with cultural transmission and assortative mating. III. Family structure and the analysis of experiments. Am J Hum Genet 31: 366-388.
6. Cockerham CC (1954): An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859-882.
7. Crumpacker DW, Cederlöf R, Friberg L, Kimberling WJ, Sörensen S, Vandenberg SG, Williams JS, McClearn GE, Grevér B, Iyer H, Krier MJ, Pedersen NL, Price RA, Roulette I (1979): A twin methodology for the study of genetic and environmental control of variation in human smoking behavior. Acta Genet Med Gemellol 28:173-195.
8. Eaves LJ, Heath AC (1981): On the detection of asymmetric assortative mating. Nature 289: 205-206.
9. Eaves LJ, Last K, Young PA, Martin NG (1978): Model-fitting approaches to the analysis of human behavior. Heredity 41:249-320.
10. Falconer DS (1960): Introduction to Quantitative Genetics. Edinburgh: Oliver and Boyd.
11. Fisher RA (1918): The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh 52:399-433.
12. Hayley CS, Last K (1981): The advantages of analyzing human variation using twins and twin half-sibs and cousins. Heredity 47:221-236.
13. Heath AC (1983): Human Quantitative Genetics: Some Issues and Applications. DPhil Thesis, University of Oxford.
14. Jinks JL, Fulker DW (1970): Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. Psychol Bull 73:311-349.

15. Loehlin JC (1978): Heredity-environment analyses of Jencks's IQ correlations. Behav Genet 8: 415-426.
16. Loehlin JC (1979): Combining data from different groups in human behavior genetics. In Royce JR, Mos L (eds): Theoretical Advances in Behavior Genetics. Alphen aan den Rijn: Sijthoff & Noordhoff.
17. Malecot G (1948): The Mathematics of Heredity (in French). Paris: Masson.
18. Mather K (1974): Non-allelic interaction in continuous variation of randomly breeding populations. Heredity 32:414-419.
19. Mather K, Jinks JL (1963): Correlations between relatives arising from sex-linked genes. Nature 198:314-315.
20. Mather K, Jinks JL (1971): Biometrical Genetics, 2nd Edn. London: Chapman and Hall.
21. Mather K, Jinks JL (1977): Introduction to Biometrical Genetics. London: Chapman & Hall.
22. Price RA, Vandenberg SG, Iyer H, Williams JS (1982): Components of variation in normal personality. J Pers Soc Psych 43:328-340.
23. Rao DC, Morton NE (1978): IQ as a paradigm in genetic epidemiology. In Morton NE, Chung CS (eds): Genetic Epidemiology. New York: Academic Press.
24. Rao DC, Morton NE, Cloninger CR (1979): Path analysis under generalized assortative mating. I. Theory. Genet Res 33:175-188.
25. Rao DC, Morton NE, Lalouel JM, Lew R (1982): Path analysis under generalized assortative mating. II. American IQ. Genet Res 39:187-198.
26. Rao DC, Morton NE, Yee S (1976): Resolution of cultural and biological inheritance by path analysis. Am J Hum Genet 23:228-242.
27. Rice J, Cloninger CR, Reich T (1978): Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models. Am J Hum Genet 30:618-643.
28. Schull WJ, Neel JV (1965): The Effects of Inbreeding on Japanese Children. New York:Harper & Rowe.
29. Urbach P (1974): Progress and degeneration in the IQ Debate I, II. Br J Phil Sci 25:99-135, 235-259.
30. Williams JS, Iyer H (1981): A statistical model and analysis for genetic and environmental effects in responses from twin-family studies. Acta Genet Med Gemellol 30:9-38.
31. Wright S (1931): Statistical methods in biology. J Amer Statist Assoc 26 (Supplement): 201-208.

Correspondence: Dr. A.C. Heath, Department of Human Genetics, Medical College of Virginia, P.O. Box 33, MCV Station, Richmond, VA 23298, USA.