Behav Genet (2009) 39:165–169
DOI 10.1007/s10519-008-9253-9

BRIEF COMMUNICATION

The Effect of Assumptions About Parental Assortative Mating and Genotype–Income Correlation on Estimates of Genotype–Environment Interaction in the National Merit Twin Study

John C. Loehlin · K. Paige Harden · Eric Turkheimer

Received: 21 July 2008 / Accepted: 12 December 2008 / Published online: 27 December 2008

Abstract In a previous paper (Harden et al. in Behav Gen 37:273–283, 2007) models of genotype–environment interaction were fitted to data from the National Merit Twin Study, resulting in evidence for an interaction: the heritability of National Merit Qualifying Test scores increased at higher levels of family income. The present paper investigates two assumptions made in the previous modeling. These were a lack of resemblance between parents for cognitive skill, and possible correlations between family income and a child's genes because of the contribution of parental genes to both. The assumptions were found not to seriously affect estimates of the interaction effect—heritability still increased with income—but they did make a difference for other parameter estimates from the modeling. One possible explanation of the observed interaction, decreasing levels of assortative mating at higher income levels, was examined and found not to be consistent with other evidence from the study. Another possible explanation, a greater freedom of members of DZ pairs at higher income levels to follow independent interests, remained plausible.

Keywords Genotype–environment interaction · Twins · Assortative mating · Genotype–environment correlation · Income · Cognitive skill · National Merit Twin Study

Edited by Irwin Waldman.

J. C. Loehlin (✉)
Department of Psychology, The University of Texas at Austin, 1 University Station A8000, Austin, TX 78712-0187, USA
e-mail: loehlin@psy.utexas.edu

K. P. Harden · E. Turkheimer
Department of Psychology, University of Virginia, Charlottesville, VA, USA

Introduction

In a previous paper (Harden et al. 2007), we found a statistically significant interactive effect of income on the heritability of scores on the National Merit Scholarship Qualifying Test (NMSQT). This served to generalize an earlier finding of an interactive effect of income on the heritability of IQ in low-income 7-year-old children (Turkheimer et al. 2003) to a different part of the age and ability ranges. In both cases, the heritability of cognitive skills increased with income.

In fitting models to the National Merit data, certain latent variables were assumed to be independent of one another. For example, shared environment (C) and income (I) were assumed to be uncorrelated, which amounts to defining C as those aspects of shared environment independent of income. The residual (E, unshared environment, error, etc.) was by definition uncorrelated with the other causal variables.

A possibly more debatable assumption (or definition of variables) was that the additive effects of genes (A) were uncorrelated with income. If parental cognitive skills are in part responsible for family income and are in part genetic, one would expect the genes transmitted to offspring to be correlated with family income. The alternative is to say that one is not talking about the heritability of cognitive skills as such, but about the heritability of cognitive skills that are uncorrelated with income. This is a bit awkward, since the main interest is presumably in cognitive skills as a whole.

Another debatable assumption made in the model fitting was that the genetic correlation of fraternal twins was .5, which implies that there was no assortative mating in their parents for the genes affecting cognitive skills.

The present note concerns the consequences of relaxing these two assumptions. Because we are manipulating certain parameters rather than solving for them, we need not address such interesting (and tricky) questions as whether one can simultaneously estimate both G × E interaction and GE correlation from twin data (Purcell 2002; Price and Jaffee 2008; Rathouz et al. 2008), or whether a one-generation study can separate out the genetic and environmental components of a shared family variable such as income (Turkheimer et al. 2005). Rather, we are simply asking: if we had made the two assumptions in question differently, how much effect (if any) would it have had on our conclusions?

Method

As in the previous paper, the data used were from 509 monozygotic (MZ) and 330 dizygotic (DZ) twin pairs from the National Merit Twin Study (Loehlin and Nichols 1976). A slightly simplified version of the model fitted by Harden et al. (2007) was used. It differed from that model in that it fitted directly to the NMSQT scores instead of treating them as latent variables estimated from the five subscales. Because the NMSQT scores are a simple sum of the subscale scores, this is analogous to using a composite score rather than a factor score. The other simplification was to convert the NMSQT scores to standard scores to permit easier interpretation of the parameter estimates. (Income and the latent variables were already standardized in the earlier model fitting.) As before, the interaction between additive genes and income was represented by the product of the latent variables A and I, as a variable A × I.

To allow for the possible effect of parental assortative mating, the genetic correlation between fraternal twins was set at values of .50 (random mating, used in the original model fitting), and then .55, .60, and .65, as reflecting the effects of weak to strong phenotypic assortative mating. Weighted mean spouse correlations for IQ from the literature have been summarized as .42 (Jensen 1978) and .33 (Bouchard and McGue 1981). A marital correlation of .33 in conjunction with a low estimate of adult IQ heritability (.4) would imply a sibling genetic correlation of about .57 under phenotypic assortative mating; a marital correlation of .42 in conjunction with a high heritability estimate (.8) would lead to a genetic correlation of about .67. (The calculation is $.5 + .5h^2r_p$, where h^2 is the heritability and r_p is the phenotypic spouse correlation.) The values .50 to .65 should reflect an adequate range of possibilities, especially considering that some restriction of variability of parental IQ is likely in this population. An additional assumption, to which we later return, is that if assortative mating occurs, it is constant across the income range.

To allow for the possible presence of a gene–income correlation, we set this correlation to the values .00 (as used in the original model fitting), and then .10, .20, and .30. The correlation of income with phenotypic cognitive skill in the US population has been estimated at about .35 (Jencks et al. 1972, p. 337). Correlation with genotypic skill would presumably be less than this, to a degree dependent on heritability. (In the model, a correlation of .40 or more will produce a negative estimate of the effect of income).

Models with the above specifications were fitted to the NMSQT data using M*plus*. Other conditions were as in the original model fitting.

Results

The results of fitting models incorporating different assumptions about the degree of assortative mating and gene–income correlation are shown in Table 1. The major effect of assuming greater assortative mating was to increase the estimate of the effects of genes, *a*, and decrease that for shared family environment, *c*. In the absence of gene–income correlation there was no effect of assortative mating on the path from income to NMSQT score (it remained at .252); in the presence of such correlation, the path value

Table 1 Model-fitting estimates under a range of assumptions regarding assortative mating and gene–income correlation

Assumptions		Estimates of path			
r_{GS}	r_{IA}	*a*	*c*	*i*	*x* (/SE)
.50	.00	.679	.593	.252	.061 (2.40)
	.10	.681	.595	.181	.055 (2.15)
	.20	.678	.610	.110	.049 (1.87)
	.30	.680	.627	.039	.041 (1.53)
.55	.00	.717	.549	.252	.058 (2.36)
	.10	.718	.550	.177	.053 (2.12)
	.20	.720	.562	.102	.047 (1.84)
	.30	.721	.583	.027	.040 (1.51)
.60	.00	.761	.484	.252	.055 (2.31)
	.10	.762	.488	.173	.049 (2.08)
	.20	.764	.503	.093	.044 (1.81)
	.30	.765	.529	.014	.038 (1.50)
.65	.00	.814	.389	.252	.051 (2.26)
	.10	.816	.394	.167	.046 (2.03)
	.20	.817	.415	.083	.041 (1.78)
	.30	.819	.451	−.002	.035 (1.48)

Note: Values of parameter *e* not shown, under all conditions $e = .347$. r_{GS} = genetic correlation between sibs; r_{IA} = correlation of income with additive genes. Paths: *a* = additive genes to NMSQT; *c* = family environment other than income to NMSQT; *i* = family income to NMSQT; *x* = product of additive genes and income to NMSQT; (/SE) = ratio of *x* to its standard error. Standard errors as provided by M*plus* based on maximum likelihood fitting shrank, but not much (.18–.17 at .10, .04–.00 at .30). The *x* path from the A × I product to NMSQT score also shrank a little under increasing assortative mating, but even less—from .06 to .05 at zero correlation, about half that much, from .041 to .035, at a correlation of .30.

The other assumption, an increasing gene–income correlation, had a marked effect on the income path, dropping it from .25 to near zero (.04 under random mating, .00 under strong assortment). If income is correlated with the genes, the observed *r* of .25 between income and NMSQT score is explainable by this fact, and requires no direct effect of income on NMSQT score. The other main effects, of genes and shared environment were affected less: the genes hardly at all—an increase of .005 at most; the shared environment, slightly more, but still only about .03–.06, or 6–16%.

What about the interactive effect of genes and income on NMSQT score, the main focus of the original article? Unlike the main effect of income, it did not disappear with increasing gene–income correlation, but it got slightly smaller. At lower levels of gene–income correlation, it was quite modest, .05 or .06, but it was statistically significant, greater than twice its estimated standard error. At the highest level examined, $r_{IA} = .30$, it was no longer statistically significant with these sample sizes, but there was no dramatic change—the size decrease was approximately linear with increasing correlation, and modest: in the case of random mating, the path value dropped from .06 to .04, and under strong assortative mating, from .05 to .035.

Initial discussion

Examination of two of the assumptions underlying the modeling of interactive effects of genes and income on NMSQT scores led to some qualification of the original conclusions. A reasonable presumption is that if NMSQT scores had been available for the parents of the twins in the study, the spouses would have been at least somewhat correlated on them, whereas our modeling assumed random mating. The consequence of this was that the modeling may have underestimated additive genetic effects and overestimated shared environmental ones. However, these were not of central concern in the paper, which was focused on the effects of gene–income interaction. Allowing for assortative mating would have decreased our estimate of the size of this interactive effect trivially, changing it from .061 to .051, but would not have affected its statistical significance.

Assuming no correlation between genes and income had a substantial effect on the estimate of the effects of income. If the gene–income correlation were as high as .30, the estimate of the direct effect of income on NMSQT score would have dropped close to zero. The effect of shared environment other than income would have increased a trifle; the effect of genes very slightly if at all. The gene–income interaction, our primary concern, was not changed dramatically in a quantitative sense. It showed a modest decrease in going from a zero gene–income correlation to one of .30; at the highest level of correlation, this component would no longer have been statistically significant, but there was nothing to suggest a sudden drop-off.

Any modeling necessarily involves simplifications, but it is often possible to assess the consequences of the simplifications. In our case, the assumption of random mating had relatively modest effects on the paths involving income, but would have been more serious if our focus had been on the relative magnitudes of *a* and *c*. The assumption that income and genes were uncorrelated had considerable consequences for estimating the direct effect of income on cognitive skills, but little effect on estimating the interactive effect of income and genes.

As mentioned earlier, our modeling of assortative mating assumed that it was constant across various levels of income. What if it were not?

Additional analyzes

For further analysis we break down the data according to three levels of income, allowing us to examine the interaction more directly, and to test two possible explanations of it, including the one raised in the preceding section: that assortative mating might vary with income.

The first columns of Table 2 present the correlations for MZs and DZs at different levels of family income. Family income was reported by a parent, who checked the appropriate box out of 7 on the questionnaire. We have combined the original income categories 1 and 2, 3 and 4, and 5–7 to obtain reasonable *N*s for the correlations. The middle columns of the table provide direct evidence of the

Table 2 Income levels, twin correlations, and assortative mating in the National Merit Twin Study

Income category	NMSQT		h^2 for r_{GS}		r_p
	r_{MZ}	r_{DZ}	.50	.65	
<$7499	.84 (188)	.66 (104)	.37	.52	.44 (285)
$7500–$14,999	.88 (205)	.63 (144)	.51	.73	.46 (347)
≥$15,000	.90 (82)	.57 (54)	.66	.94	.52 (136)

Note: Number of pairs in parentheses. Income = parental report of family income before taxes (in 1963 dollars); r_{MZ}, r_{DZ} = correlations of MZ and DZ twin pairs; NMSQT = National Merit Scholarship Qualifying Test total score; h^2 = heritability estimate; r_{GS} = genetic correlation between siblings; r_p = phenotypic correlation of parents' educational levels interaction in the form of estimates of heritability—under random mating, as in the original modeling, and under strong assortative mating (assumed constant across income levels). In both cases, it is clear that heritability increases with income in these data. Thus the original conclusion of an interaction of the effects of additive genes and income on cognitive skill, which was obtained by fitting an A × I term in an overall model, is supported by this more direct approach: the difference between the MZ and the DZ correlations gets larger at higher levels of income. Assuming that parental assortative mating is present raises all the heritability estimates, but they still increase from lower to higher income levels.

Inspection of the first columns of Table 2 suggests that the increase in heritability with income reflects a decrease in correlation among the DZ twins—the correlation between MZs remains high and even increases slightly across income levels. There might be various reasons for this difference—for example, economically better-off families might be more able to support the different interests of members of DZ pairs, whereas members of MZ pairs, because of their identical genotypes, pursue similar interests at all income levels. However, one possibility is that the decline in DZ correlation might reflect changes in the degree of parental assortment. Parental assortment will affect genetic resemblance only in DZ pairs, since MZ pairs are always genetically identical. If assortative mating were stronger at lower income levels, one would expect DZ *r*s to go down with increasing income, whereas MZ *r*s would be unaffected. An apparent A × I interaction could then be spurious—a change in assortative mating with income would result in a declining DZ correlation and a constant MZ correlation, and thus an increased estimate of h^2 despite no change in how the genes affect the trait.

Is such a change in assortative mating plausible? We do not have cognitive test scores for the parents, but we do have data on completed education. Spouse correlations for education at the three income levels are shown at the right in Table 2. Clearly, there is no evidence of a decline in assortative mating for education across income levels in these data—if anything, there is a slight increase. This does not absolutely rule out a decline across income levels in assortative mating for cognitive skills, but it surely renders it much less plausible. Indeed, if it is the case that assortative mating for IQ is in considerable part a byproduct of assortative mating for education (because potential spouses often meet as a side effect of the educational process), it becomes even less likely that assortative mating would have a drastically different pattern for education and cognitive skills.

What about the other option mentioned, the possibility that higher family income permits greater diversity in the activities and interests of DZ twin pairs? Table 3 reports one item of evidence relevant to this hypothesis; namely, correlations of twins on the intellectual orientation scale of the Vocational Preference Inventory. The MZs do not change much in resemblance—there is a slight decline, but not a statistically significant one. The DZs, however, show a marked decrease in resemblance with income. In fact, for the highest income category the correlation becomes negative, although the −.08 is based on a fairly small sample of 33 pairs, and would not differ significantly from zero, or from a low positive value, which is what a simple genetic hypothesis would predict.

Table 3 Twin correlations for intellectual interests in the National Merit Twin Study

Income category	Intellectual interests	
	r_{MZ}	r_{DZ}
<$7499	.38 (119)	.29 (56)
$7500–$14,999	.36 (131)	.22 (90)
≥$15,000	.33 (59)	−.08 (33)

Note: Number of pairs in parentheses. Income = parental report of family income before taxes (in 1963 dollars); r_{MZ}, r_{DZ} = correlations of MZ and DZ twin pairs; Intellectual interests = Intellectual orientation score from Vocational Preference Inventory

Concluding discussion

To summarize, our further scrutiny of the data of the National Merit Twin Study suggests that the existence of an A × I interaction was not an artifact of the specific assumptions made in the modeling about random mating and uncorrelated A and I, although other aspects of the results were affected by these assumptions. One source of the increase in heritability estimates was a decrease in DZ correlations as family income rose. One possible explanation—that the observed A × I was an artifact resulting from changing levels of assortative mating with income—was examined and found not to be plausible, given no decline for educational assortative mating with increasing income levels among the parents. Data on vocational interests suggested that the presence of diverging interests of DZ twins at higher income levels may well have been making a contribution. This would constitute a legitimate source of A × I interaction.

Is this the total story? Obviously not. For one thing, it would not account for an increase in the MZ correlation of NMSQT scores at higher incomes (if this increase should prove to be replicable). For another, cause-and-effect ambiguity remains: Is divergence in intellectual interests an effect or a cause of divergence in cognitive skills among DZ twins? Nevertheless, the present results do shed some light on the sensitivity of A × I modeling to assumptions, and thus have a bearing on further research in this area. It might be desirable for future investigators to incorporate assumption testing of this kind into their model fitting.

Acknowledgments We thank Gregory Carey and Wendy Johnson for helpful discussions about G $\times$ E interaction. A version of this paper was presented at a symposium at the Behavior Genetics Association meetings in Louisville, KY, June 28, 2008.

References

Bouchard TJ Jr, McGue M (1981) Familial studies of intelligence: a review. Science 212:1055–1059. doi:10.1126/science.7195071

Harden KP, Turkheimer E, Loehlin JC (2007) Genotype by environment interaction in adolescents' cognitive ability. Behav Genet 37:273–283. doi:10.1007/s10519-006-9113-4

Jencks C et al (1972) Inequality: a reassessment of the effect of family and schooling in America. Basic Books, New York

Jensen AR (1978) Genetic and behavioral effects of nonrandom mating. In: Osborne RH, Noble CE, Weyl N (eds) Human variation: the biopsychology of age, race and sex. Academic Press, New York, pp 51–105

Loehlin JC, Nichols RC (1976) Heredity, environment, and personality: A study of 850 sets of twins. University of Texas Press, Austin

Price TS, Jaffee SR (2008) Effects of family environment: gene–environment interaction and passive gene–environment correlation. Dev Psychol 44:305–315. doi:10.1037/0012-1649.44.2.305

Purcell S (2002) Variance components models for gene–environment interaction in twin analysis. Twin Res 5:554–571. doi:10.1375/136905202762342026

Rathouz PJ, Van Hulle CA, Rodgers JL, Waldman ID, Lahey BB (2008) Specification, testing, and interpretation of gene-by-measured-environment interaction models in the presence of gene–environment correlation. Behav Genet 38:301–315. doi:10.1007/s10519-008-9193-4

Turkheimer E, D'Onofrio BM, Maes HH, Eaves LJ (2005) Analysis and interpretation of twin studies including measures of the shared environment. Child Dev 76:1217–1233

Turkheimer E, Haley A, Waldron M, D'Onofrio BM, Gottesman II (2003) Socioeconomic status modifies heritability of IQ in young children. Psychol Sci 14:623–628. doi:10.1046/j.0956-7976.2003.psci_1475.x