Behav Genet (2016) 46:726–733
DOI 10.1007/s10519-016-9795-1

ORIGINAL RESEARCH

Minor Allele Frequency Changes the Nature of Genotype by Environment Interactions

Brad Verhulst[1] · **Michael C. Neale**[1]

Received: 23 September 2015 / Accepted: 11 April 2016 / Published online: 22 April 2016

Abstract In the classical twin study, phenotypic variation is often partitioned into additive genetic (*A*), common (*C*) and specific environment (*E*) components. From genetical theory, the outcome of genotype by environment interaction is expected to inflate *A* when the interacting factor is shared (i.e., *C*) between the members of a twin pair. We show that estimates of both *A* and *C* can be inflated. When the shared interacting factor changes the size of the difference between homozygotes' means, the expected sibling or DZ twin correlation is .5 if and only if the minor allele frequency (MAF) is .5; otherwise the expected DZ correlation is greater than this value, consistent (and confounded) with some additional effect of *C*. This result is considered in the light of the distribution of minor allele frequencies for polygenic traits. Also discussed is whether such interactions take place at the locus level or affect an aggregated biological structure or system. Interactions with structures or endophenotypes that result from the aggregated effects of many loci will generally emerge as part of the *A* estimate.

Keywords $G \times E$ interaction · Biometrical genetics · Bias · Minor allele frequency · Common environment · Additive genetic variance

Introduction

In the classical twin study, phenotypic variation is typically partitioned into additive genetic (*A*), common (*C*) and specific environment (*E*) components. Dominance genetic variation (*D*) is sometimes substituted for *C*, although power to detect it is relatively low (Martin and Eaves 1977). These models rely on several assumptions, of which two are key. One is that MZ and DZ twin pairs share environmental factors (that they do not elicit) to the same extent (Rose et al. 1990). A second assumption is that there is no genotype by environment interaction. It is this second assumption that is the focus of this article, particularly the consequences of its failure when the environmental factor in question is shared by twins.

To be explicit about the type of genetic variation being modeled, we distinguish between *A* and *D* and hereafter refer to genotype by environment (often denoted as $G \times E$) in terms of *A*, *C*, *D* and *E*, according to which components of variance are interacting. The primary focus of this article is on the interaction between *A* and *C*, which we term $A \times C$. Two levels of interaction effects are considered. The first is at the *locus* level, in which the difference between the means of the homozygotes is affected by the interacting *C* factor. The second is *aggregated* where multiple loci have been combined generate a structure, network or endophenotype, prior to the effect of the *C* factor on the expression of the phenotype. These two circumstances make different predictions about the resemblance between relatives, and therefore influence estimates from twin studies differently. It is important, therefore, to quantify the potential for the type of the interaction on parameter estimates.

From biometrical genetic theory (Mather and Jinks 1982) $A \times C$ interaction is predicted to inflate the estimate

Edited by Stacey Cherny.

✉ Brad Verhulst
brad.verhulst@gmail.com

[1] Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

of A in a classical twin study, among other research designs. We demonstrate that this prediction is correct when A consists of variance due to a single locus and the minor allele frequency is one half. For other minor allele frequencies (MAFs), the expected genetic correlation between DZ twins (or other sibling types) is greater than $\frac{1}{2}$ and approaches unity as the MAF approaches zero. At the locus level, therefore, $A \times C$ interaction will typically inflate the estimate of *both* A and C. These conclusions are demonstrated algebraically and illustrated graphically via simulations in the following sections.

Classical expectation

To examine the effect of MAF on the correlations between DZ twins, we start with the expected frequencies and mean values of pairs of siblings, as specified by Fisher (1918). Table 1 shows the expected frequencies of DZ twin or sibling pairs' genotypes at a single diallelic locus with frequencies p and $q = 1 - p$ for alleles D and d, respectively. The corresponding frequency table for MZ twin pairs, whose genotypes are identical, would have $p^2, 2pq$ and q^2 on the diagonal and zero elsewhere.

Following Fisher, but assuming additivity, we specify the phenotype means to be a, 0, and $-a$ for the *DD*, *Dd* and *dd* genotypes, respectively. The population mean is therefore simply $\mu = a(p^2 - q^2) = a(p - q)$. The population variance is calculated as the sum of the squared deviations from the mean, weighted by their population frequencies: $\sigma^2 = p^2(a - \mu)^2 + 2pq(-\mu)^2 + q^2(-a - \mu)^2$. Substituting $a(p - q)$ for μ and simplifying, the variance is $2pqa^2$. These quantities are those found in any standard text on biometrical genetics, once the simplifying assumption of no dominance is made (Mather and Jinks 1982; Neale and Cardon 1992).

Under this additive single locus model, the expected sibling (or DZ twin) covariance is the sum of the cross-products of each type of sib pair's deviations from the mean, weighted by their expected proportion in the population. The cross-products of sib pair deviations from the mean are shown in Table 2; these are weighted by the corresponding frequencies in Table 1 (element-wise matrix multiplication) and summed.

The sum of the element-wise product of Tables 1 and 2 (i.e., the cross-product of sibling's deviations from the mean weighted by their respective proportions in the population) gives the expected covariance between siblings or DZ twins, pqa^2. Dividing this quantity by the population variance of $2pqa^2$ yields the sibling genetic correlation of $\frac{1}{2}$, regardless of the values of the allele frequency p and the allelic effect a. The MZ correlation equals the MZ covariance divided by the population variance, and since these quantities are the same, their genetic correlation is unity.

Effect of a common environment moderator (M)

We now consider the effects of changing the quantity a due to the action of a binary environmental moderator (M) which is perfectly correlated within families. The population therefore consists of a mixture of two groups, whose proportions we denote f and $1 - f$, with $0 < f < 1$ (like p) bounded on the open interval from zero to one. Assuming that the allele frequencies are equal in the two groups, and denoting the two allelic effects as a and b, the mean of this heterogeneous population is:

$$\mu = (p^2 - q^2)af_1 + (p_2^2 - q_2^2)b(1 - f_1) \tag{1}$$

To compute the expected sibling correlation in the heterogeneous population, we follow the same general procedure of obtaining the mean, the population variance and the sibling covariance. The main differences here are: (i) all deviations are calculated from the grand mean of both groups (Eq. 1); and (ii) the proportions of each sib pair type are the product of the allele frequencies p and $q = 1 - p$, and f or $1 - f$, the proportion of pairs exposed to the environmental moderator. The variance, being the weighted sum of the squared deviations from the population mean, for both groups, is

$$\begin{aligned} &a^2\left(f^2(1 - 2p)^2 - 2f(1 - 2p)^2 + 2p^2 - 2p + 1\right) \\ &\quad - 2ab(f - 1)^2(1 - 2p)^2 + b^2(f - 1)^2(1 - 2p)^2 \end{aligned} \tag{2}$$

and the weighted sum of the crossproducts of sibling pair deviations is:

Table 1 Expected frequency of genotypes between siblings based upon allele frequencies

	DD	Dd	dd
DD	$p^4 + p^3q + \frac{1}{4}p^2q^2$	$p^3q + \frac{1}{2}p^2q^2$	$\frac{1}{4}p^2q^2$
Dd	$p^3q + \frac{1}{2}p^2q^2$	$p^3q + 3p^2q^2 + pq^3$	$\frac{1}{2}p^2q^2 + pq^3$
dd	$\frac{1}{4}p^2q^2$	$\frac{1}{2}p^2q^2 + pq^3$	$\frac{1}{4}p^2q^2 + pq^3 + q^4$

Table 2 Cross-products of sibling pairs' expected deviations from the population mean μ, for a diallelic locus

	DD	Dd	dd
DD	$(a - \mu)(a - \mu)$	$(-\mu)(a - \mu)$	$(-a - \mu)(a - \mu)$
Dd	$(a - \mu)(-\mu)$	$(-\mu)(-\mu)$	$(-a - \mu)(-\mu)$
dd	$(a - \mu)(-a - \mu)$	$(-\mu)(-a - \mu)$	$(-a - \mu)(-a - \mu)$

$$a^2\left(f^2(1-2p)^2 - 2f(1-2p)^2 + 3p^2 - 3p + 1\right) - 2ab(f-1)^2(1-2p)^2 + b^2(f-1)^2(1-2p)^2. \quad (3)$$

The sibling correlation is obtained by dividing Eq. 3 by Eq. 2. This can be simplified by completing the square $(a-b)^2$ for the term in a^2 to express it in terms of $(f-1)^2(1-2p)^2$. The resulting expression for sibling covariance in the population as a whole reduces to:

$$\frac{(f-1)^2(1-2p)^2(a-b)^2 - a^2(p^2-p)}{(f-1)^2(1-2p)^2(a-b)^2 - 2a^2(p^2-p)}. \quad (4)$$

This equation equals .5 whenever $(f-1)^2(1-2p)^2(a-b)^2 = 0$, which occurs when there is no moderation group in the population ($f = 1$), or no effect of the moderator ($a = b$), or when the major and minor allele frequencies are equal ($p = .5$). Deviations of equation 4 from 0.5 arise from the difference in the genotypic means of the exposed and unexposed groups. With a MAF of 0.5, there is no difference in means between the groups; both are zero. This is because, within each group, the proportions of the two homozygotes are equal and they have the same deviation from the mean.

Pre-aggregated allelic effects

The expected sibling correlation derived above differs from what would be expected if the effects of multiple loci were aggregated *prior* to the interactive effect of the shared environment. Biologically, this seems plausible if genetic factors create an intermediate structure or system [also known as an endophenotype (Kendler et al. 2010; Cannon and Keller 2006; Gottesman and Gould 2003)] which interacts with the environmental moderator. That is, the genetic factors would act jointly as a latent genetic factor. The statistical consequences of this type of $A \times C$ interaction have been described before (Purcell 2002) but are briefly reproduced here for comparison. They are consistent with polygenic biometrical genetic theory (Mather and Jinks 1982).

Under the classical twin model, gene by common environment interaction ($A \times C$) inflates only the additive genetic variance component. This prediction is correct regardless of pre-aggregation in the special case where the MAF is one-half at every trait relevant locus, although this is biologically implausible for any trait influenced by more than a small number of loci. However, when the effects of multiple loci are aggregated before interaction with a shared environmental factor, the expected sibling correlation can be derived in terms of linear regression coefficients. Let

$$P = aA + cC + eE \quad (5)$$

where a, c and e are the regressions of the phenotype P on the A, C and E additive genetic, common and shared environmental sources of variance. Assuming that the components A, C and E are independent, the phenotypic variance under this model is:

$$\text{var}\,(P) = a^2\,\text{var}\,(A) + c^2\,\text{var}\,(C) + e^2\,\text{var}\,(E). \quad (6)$$

Moderating the regression on A by a shared environmental moderator m would yield:

$$P = (a+m)A + cC + eE \quad (7)$$

which in turn would generate expected variance of P:

$$\text{var}\,(P) = (a^2 + m^2 + 2am)\,\text{var}\,(A) + c^2\,\text{var}\,(C) + e^2\,\text{var}\,(E), \quad (8)$$

because the A, C and E components are assumed to be independent. The predicted covariances of MZ and DZ twins would be:

$$\begin{aligned}\text{cov}\,(MZ) &= (a^2 + m^2 + 2am)\,\text{var}\,(A) + c^2\,\text{var}\,(C)\\ \text{cov}\,(DZ) &= .5(a^2 + m^2 + 2am)\,\text{var}\,(A) + c^2\,\text{var}\,(C).\end{aligned} \quad (9)$$

Clearly, the MZ and DZ covariances due to additive genetic effects remain in the ratio of 2:1, which indicates that the interaction would inflate only the estimate of the additive genetic variance component, by the amount $m^2 + 2am\,\text{var}\,(A)$.

Illustration and simulation

In this section we illustrate three $A \times C$ scenarios. First is the single locus $A \times C$, where a single locus interacts with a binary shared environment factor. This demonstration directly utilizes the algebra in Eqs. 3 to 4 above. The second illustration extends the single locus model to multiple loci, again by direct calculation. The third $A \times C$ scenario is where multiple loci aggregate to form an "endophenotype" prior to interaction with the environment. Here simulation is used to illustrate. All of the scripts that were used to plot the figures, calculate the correlations and simulate the data can be found in the (online) supplementary material.

Single locus $A \times C$

Methods

To explore the impact of minor allele frequency on the correlation between DZ twins, and hence the effects on the parameter estimates obtained from the classic twin study, we evaluate Eq. 4. The moderator is assumed to be binary, with 50% of the population exposed, i.e., $f = .5$. The additive genetic effect sizes are set at $a = 1$ for the unexposed and that in group 2 varies from $b = 1$ to $b = 2$. We also vary the minor allele frequency (which is equated across the two groups) from .01 to .5.

Results

Figure 1 graphs the results where $A \times C$ interaction changes the distance between the homozygote means under the additive genetic model. As expected, when the minor allele frequency is .5, the correlation between DZ twins is exactly $r = .5$, regardless of the difference in the effect size between groups. However, as the minor allele frequency decreases, the difference in the effect size between the two groups increases the DZ correlation. For minor allele frequencies greater than .3, the inflation of the DZ correlation is relatively modest; most twin studies would be underpowered to detect such differences, as the expected correlation would be only slightly greater than half that of MZ twin pairs. This would be especially true in the presence of moderate or large random environment variation, which would decrease both MZ and DZ correlations and reduce the statistical power to test differences between them. However, when the minor allele frequency is approximately .1, an effect size ratio of 3:2 (i.e., 1.5 on the graph) would often be readily detectable as the DZ correlation would be 60 % of that of the MZ pairs. This effect becomes much more pronounced with minor allele frequencies less than 0.1, with the DZ correlation rising to .9 × *rMZ* or greater. The asymptote for this effect (for very rare variants) would be for the DZ correlation to equal the MZ correlation, although the variance due to very rare alleles becomes negligible, unless the distance between the homozygote means is very great relative to the population variance.

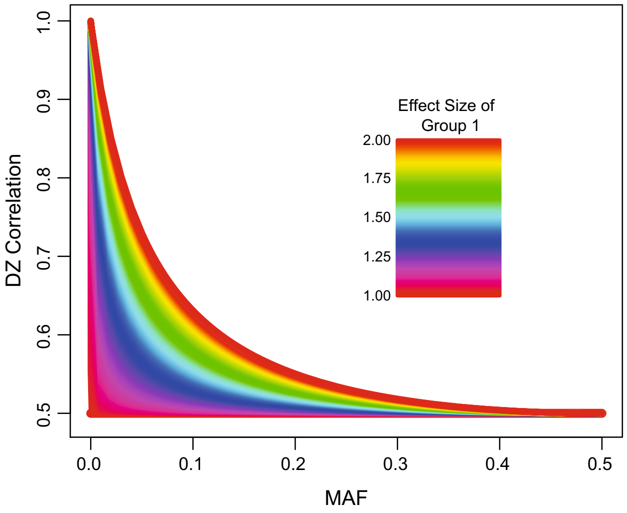

Fig. 1 Correlation between DZ twins or siblings as a function of (i) differences in relative proportion of two subpopulations; and (ii) differences in their minor allele frequency (MAF). The MAF is fixed at .5 in subpopulation 1, and varies from .01 to .5 in subpopulation 1

Multiple locus $A \times C$

Methods

To extend the single locus model to the multi-locus case, the same general framework was used; the variance and covariance were calculated for each locus, and then aggregated. Specifically, the effect size in group 1 was set to $a = 1$ and that of group 2 at $b = 2$. The groups were also set to be equally frequent ($f = .5$). Next, 20 minor allele frequencies for each set of loci were sampled from three allelic distributions: (A) a uniform distribution bounded by .01 and .50, (where minor allele frequencies are sampled at an equal probability across the MAF range.); (B) an exponential distribution ($\lambda e^{-\lambda x}$) with $\lambda = 1$ bounded by .01 and .50; this distribution slightly inflates the probability of sampling rare variants); and (C) an exponential distribution with a $\lambda = 10$ bounded by .01 and .50 (this distribution substantially inflates the probability of sampling rare variants). We then calculate and sum the variance and covariance of each locus, with the correlation equalling $\frac{\Sigma Cov_i}{\Sigma Var_i}$. This process was repeated 100,000 times to produce a distribution of sibling correlations.

Results

The first column of histograms in Fig. 2 shows the MAFs for three simulated scenarios. Scenario A has a flat, equiprobable distribution of MAFs, B has a negative exponential distribution with higher probability of lower MAFs. Scenario C is probably closest to reality with a higher negative exponential. All of the distributions are bounded by .01 and .50. The second column of histograms in Fig. 2 shows the results of the locus-by-locus $A \times C$ interaction. Just as in the single locus case described above, multiple loci $A \times C$ interaction increases the sibling or DZ twin genetic correlation above one half. The mean (SD) of the sibling genetic correlation for scenarios A, B and C are, respectively, .544 (.012), .557 (.014) and .60 (.023). Accordingly, if the environment interacts with a genome on a locus-by-locus basis, the aggregate effect will also inflate the sibling correlation, which would inflate the estimates of *C*, as well as *A*, in the classical twin study.

Endophenotype × *C*

Methods

Data for the endophenotype by environment interaction were generated in several stages. First, 20 diallelic loci genotypes for two parents were generated for each of the 10,000 families, such that each mother and father had 2

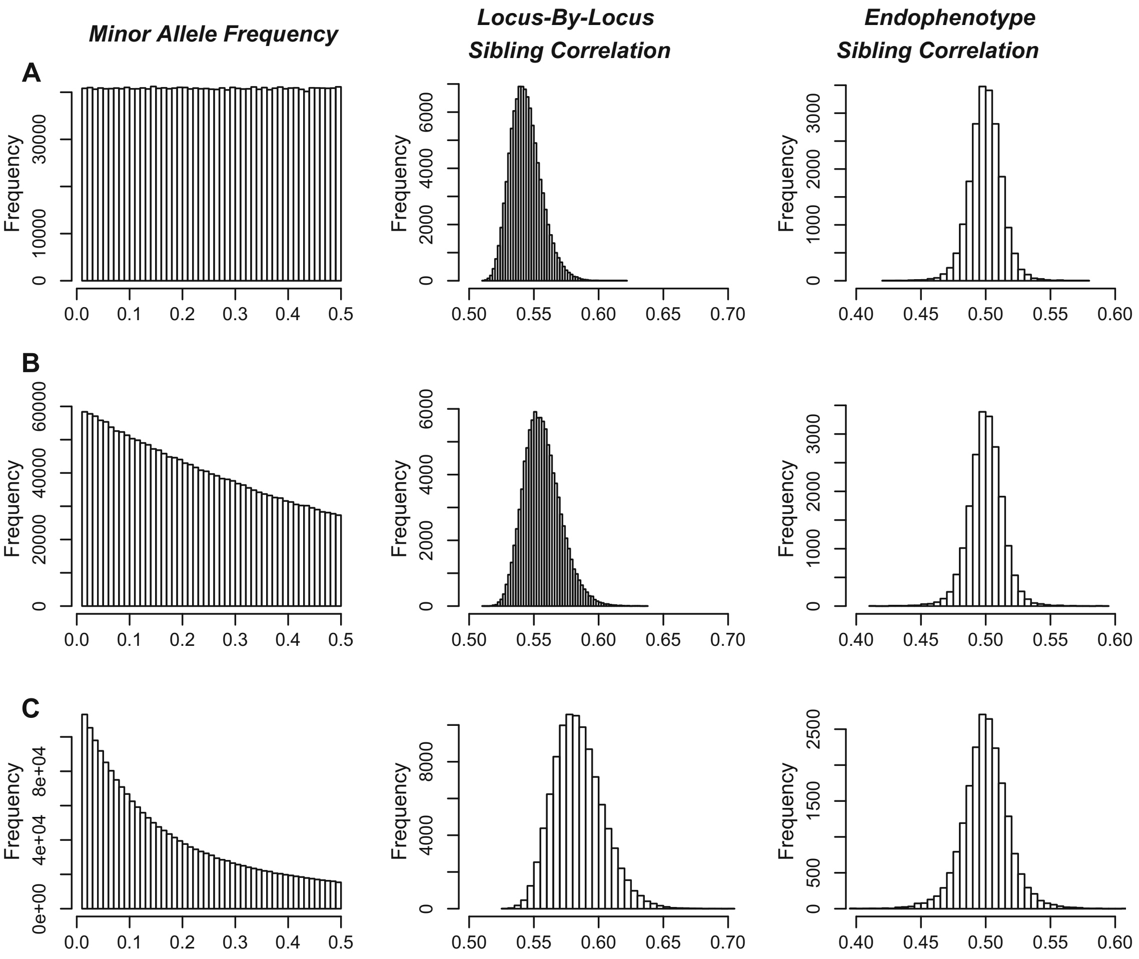

Fig. 2 Histograms of sibling correlations generated by two mechanisms (columns 2 and 3) for three different minor allele frequency distributions (**a**, **b** and **c**, shown in column 3). The two mechanisms are: column 2, an interacting shared factor that operating at the locus level by changing the distance between homozygotes; and on an aggregated 'endophenotype,' which is assembled as the sum of locus effects prior to interaction with the shared factor

"chromosomes" each. Minor allele frequencies were drawn from the same 3 distributions used in the Multiple Locus $A \times C$ simulation. Offspring genotypes were then generated by selecting one "chromosome" from each parent. This method preserves Hardy-Weinberg equilibrium. To generate the endophenotype, 20 β's were drawn from a random uniform distribution ranging from $-.02$ to .02, and multiplied by the genotypes of each sibling. The endophenotypes were then standardized. Finally, the endophenotypes were multiplied by a binary moderator that was perfectly correlated within families. Specifically, the endophenotypes for half the families were multiplied by 1, while the other families' endophenotypes were multiplied by 2. This process was repeated 20,000 times to generate a distribution of sibling correlations.

Results

The effects of a shared environment moderator on a pre-aggregated 'endophenotype' are shown in column 3 of Fig. 2. The results here are quite different, as the moderator has no effect on the sibling genetic correlation, which remains at 0.5, entirely consistent with biometrical theory that $A \times C$ interaction is confounded with the effects of A in the classical twin study. The aggregation of the effects of many (in this case 20) loci, prior to the variance changing effect of the moderator, eliminates the effects of MAF on the outcome. The mean sibling correlation in the simulations are 0.500, regardless of the MAF distribution. However, there is evidence of increased variance in the estimates as lower MAFs become more frequent. The standard deviations for scenarios A, B and C are .012, .013 and .018, respectively. This effect—also observed for the locus-by-locus effect—reflects the loss of precision of the sibling correlation when the variability at the locus is reduced at lower MAFs.

Discussion

The results presented above illustrate the consequences of four types of $A \times C$ interaction. The key novel results emerge from the first of these, which involves modeling $A \times C$ at the locus level. For simplicity, the interacting C component was assumed to be binary, and to change the distance between the homozygotes' means. This model is essentially symmetric, in that in both conditions the heterozygote has an expected mean of zero, which is the midpoint between the expected means of the two homozygotes. We showed algebraically the conditions under which the sibling or DZ twin genetic correlation would increase above the expected value of 0.5; this change increases the estimates of both C and A in the classical twin study. The ratio of $C : A$ inflation increases as the minor allele frequency decreases, such that rare variants that interact with a shared environment moderator would show only an increase in C.

This main result extends to the polygenic case as long as the interaction occurs at the level of the trait-relevant loci. Here again both A and C are inflated by $A \times C$ interaction. These changes occur irrespective of whether the environmental moderator has the same or opposite effects across multiple loci (increasing the difference between homozygotes at some loci, while decreasing it at others). Thus the effect is robust and $A \times C$ interaction has likely contributed to estimates of C in previously published studies. The relative increases of A and C depend on the distribution of the MAFs and on the interaction effect sizes at the loci. We considered several simple distributions of allele frequencies, and found, as expected, that a greater proportion of low MAFs led to a greater proportion of inflation of C. The distribution of MAFs in the population obviously depends on the trait being studied, and likely reflects the effects of selection. Traits that reduce reproductive fitness at one end of their continuum are likely to have a preponderance of low MAFs Fisher (1929), which would inflate C almost exclusively. Conversely, those under stabilizing selection (in which the intermediate phenotypes are fittest) would show little inflation of C relative to that of A, because the MAFs would be close to one half for most loci.

Until now, it was widely believed that the effects of $A \times C$ interaction are completely confounded with those of A in a classical twin study. This belief is also consistent with standard statistical theory for interaction terms, in which the coefficient of an interaction term is obtained as the product of the corresponding main effect terms. Thus $A \times C$ interaction is predicted to be confounded with (i.e., inflate) the estimate of additive genetic effects, a^2. By modeling the effect of a shared environment factor at the locus level, we demonstrate that the estimate of c^2 will also be inflated when the minor allele frequency is not equal to one half. For rare variants, $A \times C$ interaction will manifest almost entirely as an inflation of c^2 with little increase in the estimate of a^2.

These results do not hold when the allelic effects are pre-aggregated into a latent phenotype or 'endophenotype' prior to the effects of the moderator. That is, the moderator would *not* act on the loci directly, but instead on the pre-aggregated genetic trait by changing the magnitude of its effect on the phenotype, per Eq. 8. Given suitable knowledge as to which loci contribute to the phenotypic variance (such as might be gleaned from genome-wide association studies), their effect sizes and minor allele frequencies, along with how the genotypic means change as a function of the environmental moderator, it would be possible to predict how the sibling genetic correlation – and hence their phenotypic covariance – would change. This model predicts that the loci that interact with the environment do so in a manner that is proportional to their contribution to the mean of the trait. In practice, interaction at this level would inflate only the estimate of a^2.

In the second scenario, the effects of the moderator would *not* act on the loci directly, but would act on the pre-aggregated genetic trait by changing the magnitude of its effect on the phenotype, per Eq. 8. Given suitable knowledge as to which loci contribute to the phenotypic variance (such as might be gleaned from genome-wide association studies), their effect sizes and minor allele frequencies, along with how the genotypic means change as a function of the environmental moderator, it would be possible to predict how the sibling genetic correlation—and hence their phenotypic covariance-change. This scenario can be envisioned as an endophenotype interacting with an environmental variable. Importantly, this model predicts that the loci that interact with the environment do so in a manner that is proportional to their contribution to the mean of the trait.

That to some extent $A \times C$ would manifest as variance in C raises some interesting issues. First, while there is a notable lack of variation due to C for many traits in adults (Rowe 1994), this is not generally true of the same traits at younger ages (less than 18 years of age), in which common environmental factors often play a bigger role for behaviors (such as substance abuse). It seems possible that $A \times C$ contributes forms part of the estimates of both a^2 and c^2

during at these ages. It is precisely at these younger ages where siblings share more environmental factors, some of which may contribute to the main effects of the shared environment, others that interact with specific genetic loci, and still others do both. All of three may contribute to the estimates of the common environment's effects that are derived from classical twin models.

Alternatively, if the mechanism is at the locus level, then the increased c^2 from an $A \times C$ interaction might not be observed for several possible reasons. First, is that it is possible that there is less $A \times C$ occurring than is thought to exist. Thus we are not finding it because it does not exist. Second, it is possible that genetic dominance or epistasis mask the effects of non-additivity in the classical twin study. Specifically, both dominance and epistasis would deflate the DZ correlation relative to the MZ correlation, nullifying the effects of the $A \times C$ interaction. And third, the environment with which the genotype interacts is not shared between family members. Interactions with unshared environmental factors inflate the unique environmental variance component in twin models.

Limitations and extensions

A limitation of the current demonstration is that the we focus on the effect of $A \times C$ and in doing so ignore the potential main effects of A, C, E, or other components variance of the phenotype. These other sources of variation do not invalidate the results, but they reduce the amount of phenotypic variation that is due to the interaction. Also, non-additive genetic variation could mask the impact of $A \times C$ interaction. Suppose, for example, that the variation in the phenotype is due, in equal parts to additive (poly)-genetic variation, unique environmental variation and $A \times C$ variation with the interaction with a relatively rare allele (MAF $\approx$ .05) with a 2:1 ratio of effect sizes in two groups. In this case, the correlation between DZ twins would be a three part mixture of $r_a = .5$ from the additive genetic variance component, $r_e = 0$ from the unique environment and $r_{A\times C} = .8$ from the unique environment. The expected DZ correlation would be $r = .433$, while that of MZ twins would be $r = .666$ ($r_a = 1$, $r_e = 0$, and $r_{gxe} = 1$). Estimates from a classical twin study would approach $A = 2(r_{mz} - r_{DZ}) = .466$, $C = 2r_{DZ} - r_{mz} = .200$, $E = 1 - r_{mz} = .333$. Absent the interaction, the estimates would be expected to be $a^2 = .5$ and $e^2 = .5$.

There are several directions in which the current exploration could be continued. Most obvious is to change from a model of a purely shared environmental moderator to one that is either not shared (beyond by chance) between relatives, or some intermediate degree of correlation for the interacting environmental factor. The biometrical model predicts that $A \times E$ would be confounded with E in the classical twin study. This seems unlikely to remain the case, since the expected mean values of the groups (of which three would be distinct: concordant exposed, concordant unexposed and discordant) would be subject to some departure from the overall population mean. Similarly, one may expect that the effects of epistasis, i.e., loci interacting with each other, would also generate some effects confounded with C. These issues will be addressed in a subsequent article.

Additional possibilities exist by considering alternative effects of the shared environmental factor on each genotype. The model used here increased the distance between the homozygotes in a symmetrical fashion, and assumed a purely additive gene action at the locus. Should the factor influence only one or two of the genotypes, the allele frequency at which $A \times C$ interaction would appear entirely additive would deviate from the 0.5 value required for symmetry in the cases considered here. The possibilities for both additive and non-additive $G \times E$ interaction at a single locus were recently considered by Aliev et al. (2014). Epistatic interactions across loci will behave somewhat similarly. Zuk and colleagues (2012) noted that with epistatic interaction, there is a reduction in variance at low allele frequencies, and a change in the apparent variance component constitution of a trait. There is therefore scope for a broader treatment, involving a wide variety of types of $G \times E$, $E \times E$ and $G \times G$ interactions.

Conclusion

The results of this paper suggest that if an environmental factor shared between twins interacts with the genome at the locus level, then both the additive genetic and the common environmental variance components in a twin model will be inflated if the MAF deviates from .5. If the shared environmental variable interacts with an aggregated genetic variable (such as an endophenotype), then only the additive genetic variance component will be inflated.

Acknowledgments The authors are grateful for comments from Drs. Lindon Eaves, Mark Reimers and Greg Carey on the results described in this article, and from discussions arising from the simulations presented at the 2013 BGA Annual Meeting in Marseille. This research was supported by NIDA Grants R01DA-018673 and R25DA-26119.

Compliance with Ethical Standards

Conflict of Interest Brad Verhulst and Michael C. Neale declare that they have no conflicts of interest.

Human and Animal Rights and Informed consent This article does not contain any studies with human or animal participants performed by any of the authors.

References

Aliev F, Latendresse SJ, Bacanu S-A, Neale MC, Dick DM (2014) Testing for measured gene-environment interaction: problems with the use of cross-product terms and a regression model reparameterization solution. Behav Genet 44(2):165–181. doi:10.1007/s10519-014-9642-1

Cannon TD, Keller MC (2006) Endophenotypes in the genetic analyses of mental disorders. Annu Rev Clin Psychol 2:267–290. doi:10.1146/annurev.clinpsy.2.022305.095232

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb 52:399–433

Fisher RA (1929) The genetical theory of natural selection, 1st edn. Dover, Mineola

Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 160(4):636–645. doi:10.1176/appi.ajp.160.4.636

Kendler KS, Neale MC (2010) Endophenotype: a conceptual analysis. Mol Psychiatry 15(8):789–797. doi:10.1038/mp.2010.8

Martin NG, Eaves LJ (1977) The genetical analysis of covariance structure. Heredity 38:79–95

Mather K, Jinks JL (1982) Biometrical genetics, 3rd edn. Chapman and Hall, London

Neale MC, Cardon LR (1992) Methodology for genetic studies of twins and families. Kluwer Academic Publishers, Dordrecht

Purcell S (2002) Variance components models for gene-environment interaction in twin analysis. Twin Res 5(6):554–571. doi:10.1375/136905202762342026

Rose RJ, Kaprio J, Williams CJ, Viken R, Obremski K (1990) Social contact and sibling similarity: facts, issues, and red herrings. Behav Genet 20:763–778

Rowe DC (1994) The limits of family influence: genes, experience, and behavior. Guilford Press, New York

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci USA 109(4):1193–1198. doi:10.1073/pnas.1119675109