Published in final edited form as:

Behav Genet. 2015 July ; 45(4): 480–491. doi:10.1007/s10519-015-9710-1.

Comparing Alternative Biometric Models with and without Gene-by-Measured Environment Interaction in Behavior Genetic Designs: Statistical Operating Characteristics

Hao Zheng,

Department of Statistics, University of Wisconsin-Madison

Carol A. Van Hulle, and

Waisman Center, University of Wisconsin-Madison

Paul J. Rathouz*

Department of Biostatistics and Medical Informatics, University of Wisconsin School of Medicine and Public Health

Abstract

To extend Purcell's well known ACE model in testing gene by measured environment interactions (GxM) in behavior genetic designs, Rathouz et al. considered a broader class of models for quantifying and testing such interactions. Only a subgroup of these extended models have been investigated for their statistical operating characteristics by Van Hulle et al. due to lack of closed form likelihood. With an estimation procedure developed using numerical techniques in a companion paper, we study statistical operating characteristics of these extended models, especially those with non-linear effects. Type I error analysis shows the likelihood ratio test for GxM to be conservative in testing models extended from the bivariate Cholesky model, and to be liberal for models extended from the bivariate correlated factors model. Parameter estimation for all models is very good, with little bias exhibited for most models and parameters. Comparisons among alternative models under various simulated conditions show that it is relatively more difficult to confirm the existence of gene by environment interactions versus to detect non-linear effects which exclude such interactions.

INTRODUCTION

Non-linear latent variable models have been used extensively for estimating gene-by-measured environment interaction (GxM) in quantitative behavior genetic designs. Purcell (2002) proposed an important extension of the classic bivariate biometric model (Jinks & Fulker, 1970; Neale & Cardon, 1992) to allow quantifying and testing for GxM between a measured environment M and each of the classic biometric variance components, including additive genetic (A), and shared (C) and unshared (E) environmental influences impacting on a phenotype P of interest, while at the same time accounting for potential correlations between M and those same A , C , or E variance components. Rathouz, Van Hulle, Rodgers,

*Corresponding Author: Department of Biostatistics & Medical Informatics, University of Wisconsin School of Medicine and Public Health, K6/446 CSC, Box 4675, 600 Highland Avenue, Madison, WI 53792-4675, USA, 1.608.263.1706, rathouz@biostat.wisc.edu.

Waldman, and Lahey (2008) considered a broader class of models for quantifying and testing such interactions, and those models have been thoroughly reviewed and described in a companion article in this same issue of *Behavior Genetics* (Zheng & Rathouz, 2015).

In that article, we developed an estimation procedure (GxM) in **R** (R Core Team, 2013) using numerical techniques to compute and maximize the likelihood for the models from Rathouz et al. (2008). The full algorithm and investigation of the procedure's numerical performance are given therein; fitting accuracies reached acceptable levels with moderate numbers of numerical integration nodes. However, the statistical (versus the numerical) operating characteristics of those models have not yet been investigated.

The goals of this paper are to evaluate Type I error rates, parameter estimation, and performance of the Bayesian Information Criterion (BIC) for comparing amongst the models proposed earlier (Purcell, 2002; Rathouz et al., 2008). We focus primarily on those models which have not been quantitatively investigated before. In the next section, we briefly revisit the suite of models proposed by Rathouz et al. (2008) for testing and estimating GxM, followed by a summary of the operating characteristics for those models with a closed form solution (Van Hulle, Lahey, & Rathouz, 2013). We then present simulation studies of the statistical operating characteristics of these latter models. In the Discussion, we summarize these results and provide some guidance to the worker applying these models.

REVIEW OF BIOMETRIC MODELS FOR TESTING GXM

Here we present a very brief description of previously proposed models, complete explanations and interpretations of which are elsewhere (Rathouz et al., 2008; Zheng & Rathouz, 2015). The putatively moderating “environmental” variable is denoted by M and the ultimate phenotype of interest by P . Both are observable and are modeled together in a joint bivariate latent structural equation model. For simplicity, we present the models in terms of a single individual rather than a pair of related individuals (e.g., twins).

The classical biometric model for a single variable, in this case M , is given by

$$M = \mu_M + a_M A_M + c_M C_M + e_M E_M, \quad (1)$$

where μ_M is the intercept for M , and A_M , C_M and E_M are independent standard normal latent random variables representing additive genetic influences, and shared and unshared environmental influences on M , respectively.

Models for Testing GxM

The remainder of the models explore GxM interactions through various specifications for P . Purcell (2002) proposed an extension of the Cholesky model with GxM (CholGxM),

$$P = \mu_P + (a_C + \alpha_C M) A_M + (c_C + \kappa_C M) C_M + (e_C + \varepsilon_C M) E_M + (a_U + \alpha_U M) A_U + (c_U + \kappa_U M) C_U + (e_U + \varepsilon_U M) E_U, \quad (2)$$

where μ_P is the intercept for P , and A_U , C_U and E_U are again standard normal latent random variables, independent of each other and of the latent factors (A_M , C_M , E_M) influencing M .

Coefficients with the subscript C quantify additive genetic and environmental effects on P that are common with those on M , whereas those with U quantify additive effects that are unique to P . Coefficients a_U , c_U , and e_U are assumed non-negative. Greek coefficients α_C , κ_C , ε_C , α_U , κ_U and ε_U capture the interaction of moderator M with the genetic and environmental influences on P . When interaction effects are non-existent, Model (2) reduces to the classic bivariate Cholesky (Chol) model, specified as

$$P = \mu_P + a_C A_M + c_C C_M + e_C E_M + a_U A_U + c_U C_U + e_U E_U. \quad (3)$$

In contrast to specification of the Cholesky parametrization, in situations without a clear causal ordering, a “correlated factors” model is more appropriate (Johnson, 2007; Loehlin, 1996). Denoting the genetic and environmental influences on P by A_P , C_P , and E_P , we extend the correlated factors model to allow for GxM (CorrGxM), viz.,

$$P = \mu_P + (a_P + \alpha_P M) A_P + (c_P + \kappa_P M) C_P + (e_P + \varepsilon_P M) E_P, \\ r_a = \text{corr}(A_M, A_P), r_c = \text{corr}(C_M, C_P) \text{ and } r_e = \text{corr}(E_M, E_P). \quad (4)$$

Note that (4) directly quantifies the genetic (r_A) and environmental (r_C , r_E) correlations between M and P . In the absence of interaction effects, Models (2), (3) and (4) are equivalent.

Operating Characteristics for Models with GxM

Because algorithms for fitting and testing Models (2) to (4) are available in Mplus, in prior work (Van Hulle et al., 2013), we were able to examine the statistical operating characteristics of these models. We simulated $N=2000$ replicates of either $n=500$ or 2000 twin pairs under each of the alternative models, varying the degree of genetic and environmental influences. When comparing CholGxM to sub-models such as Chol or CorrGxM, false positive rates consistently fell short of the nominal Type I error rates and tests were therefore underpowered. In addition, fitting CorrGxM proved to be computationally unstable in Mplus, and tests of CorrGxM (versus CholGxM) as the null hypothesis performed poorly. In the present study, we further explore CorrGxM using numerical integration to calculate the log-likelihood instead of the closed form solution available in Mplus. Finally, in Van Hulle et al. (2013), when using the BIC to compare pairs of models, we were able to pick the correct model in the majority of cases when one of them was used to generate the data (i.e., the “true” model). Larger sample sizes (e.g., $N=2000$) were needed to distinguish non-linear effects from true interactions, though this should not prove to be an impediment to employing these models given the typical sample sizes in more recent twin and family studies.

Novel Non-linear Models: Gene-by-Gene and Other Latent Interactions

The Cholesky and correlated factors models can each be extended in a novel way—and capture novel genetic influences—which cannot be fitted using standard available software. These models do not contain GxM in the sense of (2) or in any other way. Rather, they posit that genetic and shared and non-shared environmental effects operate additively and

independently—but not linearly—in the model for P . The non-linear Cholesky (CholNonLin) model is specified as:

$$P = \mu_P + a_C A_M + c_C C_M + e_C E_M + \gamma_1 A_M^2 + \gamma_2 C_M^2 + \gamma_3 E_M^2 + a_U A_U + c_U C_U + e_U E_U + \delta_1 A_M A_U + \delta_2 C_M C_U + \delta_3 E_M E_U. \quad (5)$$

Analogously, we extend the classic bivariate correlated factors model to include non-linear but additive genetic and environmental influences, yielding a model (CorrNonLin) wherein the additive genetic effect A_M on M moderates the additive genetic effect A_P on P .

$$P = \mu_P + a_P A_P + c_P C_P + e_P E_P + \lambda_1 A_M A_P + \lambda_2 C_M C_P + \lambda_3 E_M E_P, \quad (6)$$

$$r_a = \text{corr}(A_M, A_P), \quad r_c = \text{corr}(C_M, C_P) \text{ and } r_e = \text{corr}(E_M, E_P).$$

A full description of the methods developed for fitting and testing these models can be found in Zheng & Rathouz, 2015. A main goal of the present paper is to study via simulation the statistical operating characteristics of our estimation procedure for Models (5) and (6), and to further illustrate the use of such models in ferreting out non-standard latent influences arising in the interplay between M and P .

TYPE I ERROR ANALYSIS WITH FINITE SAMPLE POINTS

Model Chol is nested in models CholGxM, CorrGxM, CholNonLin and CorrNonLin; model CorrGxM is nested in CholGxM; and model CorrNonLin is nested in CholNonLin. These models can be compared via likelihood ratio tests (LRT). Asymptotically, i.e., for large sample sizes, the LRT statistic will approach a chi-square distribution, so that using standard chi-square critical values will yield correct Type I error rates. In real-world conditions, however, especially for complex non-linear latent variable models such as these, asymptotic behavior may not emerge until sample sizes are quite large (Van Hulle et al., 2013). To examine the performance of Type I error rates with realistic finite sample sizes, we simulated data based on model Chol and compared the fit of each of the alternative models listed above to this null model. We also used these same simulated data to compare the fits of null models CorrGxM and CorrNonLin to alternatives CholGxM and CholNonLin respectively.

Design

Data from samples of twins, including both monozygotic (MZ) and dizygotic (DZ) twins, were simulated to carry out our analysis under two data generating mechanisms (DGMs), Chol_A and Chol_B, viz.,

$$M = \sqrt{0.45} A_M + \sqrt{0.10} C_M + \sqrt{0.45} E_M,$$

$$P = 0.5 A_M + 0.01 C_M + 0.1 E_M + \sqrt{0.9-0.5^2} A_U + \sqrt{0.2} C_U + \sqrt{0.9-0.1^2} E_U, \quad (\text{Chol\_A})$$

and

$$M = \sqrt{0.45} A_M + \sqrt{0.10} C_M + \sqrt{0.45} E_M, \\ P = 0.1 A_M + 0.01 C_M + 0.5 E_M + \sqrt{0.9-0.1^2} A_U + \sqrt{0.2} C_U + \sqrt{0.9-0.5^2} E_U, \quad (\text{Chol\_B})$$

wherein $\text{corr}(A_{M1}, A_{M2}) = 0.5$ for DZ twins and 1.0 for MZ twins, $\text{corr}(E_{M1}, E_{M2}) = 0$ and $C_{M1} = C_{M1}$ for all the twins. In both DGMs, the variance of M is 1.0 and the variance of P is 2.0001. The magnitudes of total genetic and total non-shared environmental influences on either M or P were set equal, while in contrast, the shared environmental influence was chosen to be mild. However, the magnitude of the genetic and environmental influences contributing to P varies between components common to M and P and those unique to P . The relative variance comparisons for Chol_A and Chol_B are shown in Table 1 with notations A and B. Table 1 also includes qualitative descriptions of DGMs for the simulations in following sections. For Chol_A, genetic influence common to M and P dominates the environment influence common to M and P in magnitude (0.5 vs. 0.1). Correspondingly, genetic influence unique to P is smaller than the environment influence unique to P . Chol_B reverses the relative strengths of the common and unique genetic and non-shared environmental influences. For simplicity, the magnitude of the shared environmental influence was set at a trivial value (0.01) in both DGMs.

We simulated $N=2000$ replicates for each of the following sample sizes: (1) 400 MZ twin pairs and 400 DZ twin pairs ($n=800$); (2) 1000 MZ twin pairs and 1000 DZ twin pairs ($n=2000$); (3) 4000 MZ twin pairs and 4000 DZ twin pairs ($n=8000$) (for Chol_A only, as these model fittings are quite computationally intensive). Whereas one can save considerable computational time using a closed form likelihood, such is not available for models CholNonLin and CorrNonLin. Therefore, to ensure that all model fits were comparable, all likelihoods were calculated numerically with 8 adaptive Gauss-Hermite quadrature (AGHQ) nodes for each dimension (see Zheng & Rathouz, 2015). In preliminary experiments (not shown), we also examined model fits using the closed form likelihoods, when available, and the outcomes were almost identical.

Results

These are non-linear structural equation models. As such, model fitting relies on the entire likelihood function, rather than primarily on the first two moments as in most classical methods. For this reason, we encountered varying degrees of difficulty in estimating the models. Estimation based on numerical integration resulted in convergence issues when the sample sizes were small, especially for model CorrNonLin. When fitting CorrNonLin to $n=800$ twin pairs from DGM Chol_A, we found that 10% of the 2000 model optimization replicates will not converge without manual intervention (e.g., trying various initial values). This percentage drops to 1% in fitting CorrNonLin to $n = 800$ twin pairs from DGM Chol_B. The convergence rates for the other models and DGMs are much higher than for CorrNonLin. When the sample size is increased to $n=2000$, model optimization converges without manual intervention for over 98% of replicates, with the exception of CorrNonLin, for which the corresponding figure is 96.5%. Rates of non-convergence continue to fall with increasing sample sizes (e.g. $n=8000$ twin pairs). Previously, we also encountered convergence issues when fitting CorrGxM using Mplus. In contrast, we did not have

difficulty in fitting CorrGxM using our numerical integration and optimization procedure in R.

For each pair of nested models, we recorded the number of times that the LRT statistic exceeded the chi-square critical values for $\alpha=0.10$, 0.05, and 0.01, and calculated the percentages of replicates exceeding the critical value out of the total replicates with convergence. Such percentages, referred to as empirical Type I error rates, are compared to nominal Type I error rates (0.10, 0.05, and 0.01) in Table 2. In this and subsequent performance results, we removed the replicates requiring manual intervention to achieve convergence.

Findings are parallel to our earlier work (Van Hulle et al., 2013) in that the rates of false positives are lower than the nominal Type I error rates for alternative models CholGxM and CholNonLin that are extensions of the Cholesky model. In contrast, the rates of false positives are higher than the nominal rates for alternatives CorrGxM and CorrNonLin that are extensions of the correlated factors model. In all cases, as the sample size increases, empirical Type I error rates approach their nominal values. Rather large sample sizes (e.g., $n=8000$) are needed, however, for this behavior to manifest. For example, model CholGxM falls short of the nominal 10% error rate when sample sizes are small (5% under $n=800$), but approaches the correct rate when sample sizes are much larger (8% under $n=8000$).

We note that, because the Cholesky model is a special case of both CorrGxM and CorrNonLin, the simulated data from Chol_A or Chol_B are valid simulations based on a CorrGxM or CorrNonLin model. For CorrGxM nested in CholGxM and CorrNonLin nested in CholNonLin, we further examined the empirical Type I error rates comparing these two pairs of models using simulated data from Chol_A and Chol_B. Results were consistent with our claim about conservativeness and liberalness of the fitting procedures for models based on the Cholesky or the correlated factors model. Specifically, the empirical Type I error rates for comparing CholGxM to CorrGxM and CholNonLin to CorrNonLin fall short of the nominal values. This is particularly true when comparing CholGxM to CorrGxM with a small sample size of $n=800$ pairs. As before, error rates approach the nominal values for larger sample sizes, but this asymptotic behavior is slow to manifest.

BIAS IN PARAMETER ESTIMATION

In a second set of numerical experiments, we generated data under a variety of mechanisms in order to examine the accuracy (bias) of parameter estimation for each of the four interaction and non-linear models in addition to the null Cholesky model, and to investigate the degree to which data distinguish between closely related alternative models. In this section, we describe the DGMs as well as the bias results.

Design

We simulated twin data jointly for M and P according to four different biologically plausible conditions (specified as A1, A2, B1 and B2 in Table 1), for each of the four models under investigation. In all 16 simulations, the data generating model for M was

$$M = \sqrt{0.45} A_M + \sqrt{0.10} C_M + \sqrt{0.45} E_M.$$

The DGMs for simulating P are listed in the columns labeled “true” in Tables 4–7; the logic for the chosen parameter values is given in the next section. Qualitatively, and building on DGMs Chol_A and Chol_B, the relative strengths of the genetic and non-shared environmental contributions to P vary. The strength of the genetic influences common to M and P is either high, with a correspondingly low effect of non-shared environmental factors common to M and P (conditions A1 and A2; upper section of Table 1) or the contribution of genetic factors common to M and P is low with a correspondingly high common non-shared environmental contribution (conditions B1 and B2 of Table 1). The total of each of the genetic and non-shared environmental effects, summed over common and unique components, are the same in all DGMs, as in Chol_A and Chol_B. The interaction effects are chosen to scale approximately with the main effects and follow the pattern given in the lower section of Table 1.

For each of the 16 conditions, we simulated $N=200$ replicates of $n=2000$ comprising 1000 MZ and 1000 DZ twin pairs. Likelihood calculation and model fitting is based on numerical integration with 8 AGHQ nodes at each dimension, as in Type I error simulation study. We focused on parameter bias, i.e., the difference between the mean of the simulated parameter estimates and the true values. Results are presented in Tables 3–7 for five model fitting specifications, under all DGMs which conform to each specification.

Results

The model fitting convergence rates for this entire batch of simulations are very good. For the three Cholesky-based models, the number of simulated data sets out of 200 replicates for which the fitting algorithm did not reach convergence is most often 0 and never exceeds 2 (1%) across the eight DGMs. For CorrGxM and CorrNonLin, the median of the non-convergence rates is 4 (2%), and the maximum is 7 (3.5%).

Bias in estimated parameters is minimal. Under the Cholesky model (Table 3), the difference between the mean parameter estimate and the true value decreases with increasing sample size, as expected, and the degree of bias is similar across the two DGM conditions. For all other specifications, bias is also generally very well controlled, indicating that our numerical procedures for computing and maximizing the likelihood are performing well.

The common exception to this pattern is that estimated shared environmental parameters, particularly c_C and c_U in Cholesky-based models Chol, CholGxM, and CholNonLin; and c_P and r_C in CorrGxM and CorrNonLin. These estimates deviate from their true values more than the estimated genetic or non-shared environmental parameters. To a lesser degree, estimates of non-linear shared environment coefficients κ_C , κ_U , κ_P , γ_2 , δ_2 , and λ_2 have more bias than estimates of other non-linear coefficients. As uncertainty in shared environmental parameters in behavior genetic models is a well-documented phenomenon, these results are not surprising.

DISTINGUISHING ALTERNATIVE MODELS

Rationale and hypotheses

In our 2008 paper (Rathouz et al., 2008), we hypothesized that some models containing GxM, in particular CholGxM Model (2) (Purcell, 2002), would be very difficult to distinguish with small to moderate sample sizes from other models that do not contain GxM. In some cases, our reason for this hypothesis is that we were able to show that models containing GxM were equivalent in mathematical form to models not containing such. In other cases, we hypothesized that such equivalence would arise in an approximate way, yielding to practical model unidentifiability in all but the most extreme cases of model separation or very large sample sizes. Some of those hypotheses were tested via simulations in Van Hulle et al. (2013), where we showed, for example, that data generated by a model for P containing $M \times M$ could appear to contain a GxM term such as $A_M \times M$ if the latter model is not compared against the former, and that even when such a comparison was made, it might be hard for data to distinguish between the two. The present study aims to extend those results taking into account the new CholNonLin Model (5) and CorrNonLin Model (6).

A major goal of these simulations is therefore to test our hypothesis that models with qualitatively differing biological interpretations about the relative roles of genes and environments are not always easily distinguished from one another. In one case—e.g., the CholGxM model—genes modify the influence of the environment M on the phenotype P ; in another—e.g., the CholNonLin model—genes and environments each operate *independently*, albeit curvilinearly, neither one moderating the effect of the others. To the degree that our hypothesis is proved true, it will call into question some of the conclusions about GxM that have been drawn in the domain area literature based on Model (2) because, in those analyses, Model (5) has not been considered as a competing explanation. Alternatively, in a more positive light, to the degree that data *can* distinguish between these competing models—and their attendant biological explanations—any conclusions about GxM will be empirically strengthened in cases where GxM models are compared to a wider class of non-linear models not containing GxM and shown to better explain the data.

Design

The foregoing simulation studies, already conducted to study Type I error rates and parameter bias, afford us the opportunity to investigate the degree to which data distinguish between closely related alternative models. That is, we can use the same DGMs and model fits as we did in those foregoing studies to test the degree to which true GxM interaction Models (2) and (4) can be distinguished from alternative non-linear Models (5) and (6), and vice versa.

Because we are interested in the degree to which the data distinguish between alternative non-nested models, we consider a situation wherein the analyst employs the BIC for model comparison and selection. We use the criteria (Raftery, 1995) that BIC differences (ΔBIC) of ± 2 , ± 6 , and ± 10 represent positive, strong, and very strong evidence in favor of the model with the lower BIC versus the comparator model. We wished to simulate situations wherein

there is very strong evidence ($\text{BIC} < -10$) of the true model over the null Cholesky model between 90% and 95% of the time (i.e., in 90% to 95% of replicates). (We note that under this criterion, a likelihood ratio test with a Type I error rate of 0.05 would lead to rejecting the Cholesky model in favor of the alternative very nearly 100% of the time.) In preliminary trial and error experiments, therefore, we identified coefficients for GxM interaction or non-linear terms to approximately achieve this criterion. For example, in the very first simulation, data were generated under specification CholGxM_A1, with non-linear genetic parameters a_U and a_C in particular chosen so that the BIC under model CholGxM would be over 10 units less than that under model Chol between 90% and 95% of the simulation replicates. The environmental parameters ε_U and ε_C play a much smaller role because the emphasis is placed on genetic influence under this condition. Simulated results yielded a rate of 90.5% (Table 8, Column 5) for the true CholGxM model being preferred over model Chol with very strong evidence.

For each simulation condition, we perform three model comparisons. For example, the first four DGMs are based on the CholGxM model. For this set of simulations we compared CholGxM to Chol to confirm that the rate at which CholGxM was preferred fell in the desired range. Considering CholGxM and CholNonLin as competitor models, we then compared CholNonLin to Chol. Finally, we directly compared CholGxM to CholNonLin to determine if GxM interactions could be distinguished from non-linear effects. We repeated this pattern for simulations based on the CholNonLin, CorrGxM, and CorrNonLin DGMs.

Results

Results of the model comparisons are shown in Table 8 (DGMs CholGxM and CholNonLin) and Table 9 (DGMs CorrGxM and CorrNonLin). The results are remarkably consistent across the range of scenarios we simulated. The general pattern is that the model for the true DGM is almost always very strongly preferred (by $\text{BIC} < -10$ criterion) over Chol (first row in each section of Tables 8 and 9), and that the comparator model is nearly as often very strongly preferred over Chol as is the true model (second row in tables).

Results on distinguishing between non-linear and true GxM interaction models are, as hypothesized, more variable and weaker (third row in tables). Examining the Cholesky-based models CholGxM and CholNonLin in Table 8, if we accept a small BIC difference as evidence in favor of one model over the other, the non-linear effects can be easily distinguished from GxM when the DGM includes non-linear effects (bottom half of Table 8). Unfortunately, GxM effects are not as easy to distinguish from non-linear effects when the DGM includes GxM (top half of Table 8). If we require a large BIC difference when deciding on model preference, and the true model includes GxM effects, then CholNonLin is equivocal to CholGxM under most conditions. Analogous results (Table 9) were obtained for comparing CorrNonLin and CorrGxM models.

We note that, whereas data were simulated under four different conditions, the relative strengths of genetic versus environmental interactions (or non-linearity) did not have a large impact on these results when the DGM contained non-linear effects (bottom half of Tables 8 and 9). When the DGM included GxM effects (top half of tables), it was much more

difficult to distinguish between GxM and non-linear effects when the ExM is low relative to AxM (conditions A1 and B1) than in the opposite setting (conditions A2 and B2).

DISCUSSION AND CONCLUSION

This paper continues the work of Rathouz et al. (2008) and Van Hulle et al. (2013) by developing alternative non-linear behavior genetic models for twin data as competitors to a now-classic model proposed by Purcell (2002). After the development of estimation procedures for all models from Rathouz et al. (2008) in a companion paper (Zheng & Rathouz, 2015), we focus here on the statistical operating characteristics of these new models. The Type I error analysis suggests conservative behavior for models based on the bivariate Cholesky behavior genetic model, and correspondingly liberal behavior for those based on the bivariate correlated factors models. Type I error rates do appear to converge to their nominal values as sample size grows, although very slowly; our simulations included sample sizes of $n = 8000$. Simulations of the bias in parameter recovery are very encouraging as, by and large, maximum likelihood estimates for parameters exhibit very little bias. Parameters involving common environmental influences C are a bit less stable.

We carried out systematic simulations to study the estimation and testing of alternative models with and without GxM. We found that, as hypothesized, it is often difficult or very difficult to distinguish GxM from other non-linear effects. Under our chosen DGM specifications, it is slightly easier to correctly identify non-linear models when GxM effects are absent than the other way around, although we do not know how specific that result is to those specifications. As in any simulation study, the number of cases we considered is limited, and other behaviors could emerge with different DGMs. In particular, we note that we considered “difficult” cases wherein the correct model is preferred over the simpler Cholesky model *very often*, but by no means *all* of the time. That is, evidence for the correct model is not overwhelming. In these cases, it is difficult to distinguish the true (e.g., GxM) model from the alternative (e.g., non-linear) model. However, it remains to be seen how well the data can distinguish between the two alternatives when evidence for the true model versus the Cholesky is overwhelmingly large.

Practical comments and guidance to researchers

In light of the simulation results in this paper, and our experience thus far with the CholNonLin and CorrNonLin models vis-à-vis the CholGxM and CorrGxM, we would like to provide some guidance to the practical researcher seeking to test for GxM in behavior genetic designs. To set that context, we first review the motivation for Models (5) and (6), and then provide some suggestions for applied work.

In Models (5) and (6), the A , C and E components continue to operate independently, a feature that is *fundamental* to the specification and interpretation of the classic univariate and multivariate ACE models (Neale & Cardon, 1992), including the Chol Model (3). The non-linearities are introduced within each of the A , C or E components and could arise because the shared components A_M , C_M , or E_M could operate on a different scale for P than they do for the manifest variable M , on which they operate linearly *by construction*. Note that none of the non-linearities involve solely the terms unique to P , A_P , C_P or E_P ; for

example, we do not propose models containing $A_P \times A_P$. This is because, by construction, A_P , C_P or E_P also operate linearly on P . Thus, compared to the GxM models, the non-linear models offer a different, competing, and equally valid *biological* explanation for the association of P to M . Whereas we recognize that the practical significance of that explanation may be less clear than one of GxM for purposes of disease prevention or intervention and/or for guiding future molecular studies, that does not make the models any more or less valid.

Some suggestions for future applied work include the following points. In addition, taking a much broader view, we also summarize *all* of our results to date on our research program on these GxM methods in a technical report on “Lessons Learned”, again with the researcher using these methods in applied work as the intended audience.¹

- First and foremost, the researcher should review the scientific evidence for testing for GxM in the context of a behavior genetic design, with a particular eye to considering whether that evidence suggests in any way that GxM could arise, and that other non-linearities are materially *less* likely. If that is the case, then a reasonable analytic approach may involve just testing GxM models against each other and simpler alternatives, avoiding non-linear models such as (5) and (6) altogether. The same would hold in the case of hypothesized non-linearities; although we do not know of cases where those models have been of primary interest. If such a case cannot be made, then we do not see a compelling argument for avoiding the non-linear models.
- Second, with the exception of comparing the Cholesky based models CholGxM or CholNonLin to those based on the correlated factors models, CorrGxM and CorrNonLin, the Type I error rates are somewhat but not terribly biased for moderate sample sizes in the range of $n=2000$; we feel that, except in the cases of borderline p -values, the researcher will by and large be able to stand by his or her conclusions in testing nested models. We feel the problem of comparing the Cholesky models to the correlated factor models is less critical because in most cases, broad conclusions (e.g., about the presence or absence of GxM) will be concordant between the two models.
- Third, again with modest sample sizes, parameter recovery is excellent; all researchers should be strongly encouraged to examine parameter estimates and to use them in interpretation of the tests of competing models.
- Fourth, we have no reason to believe that there are any major flaws in the statistical procedures including the use of the BIC to distinguish among competing models. Rather, any difficulty in distinguishing models reflects the underlying closeness of those models with respect to the manifest data which serve as the basis on which they are to be compared. Therefore, when competing models are difficult to distinguish in applied analyses, it should be stated as such in the conclusions from those analyses.

¹Available at <https://www.biostat.wisc.edu/~rathouz/Software/GxM/index.html>.

- Fifth, it is not clear that the GxM and the non-linear models will *always* be so difficult to distinguish. In our simulations, we chose especially challenging cases where the “power” in terms of BIC differences to detect the CholGxM model versus the Chol model (or, similarly, the CholNonLin versus the Chol model) was between 90% and 95%. In analyses where this power is much closer to 1.0, i.e., where evidence favoring CholGxM versus Chol is much stronger—and there are many such cases in the literature—we believe that the data will then have much greater power to distinguish between the GxM models and the non-linear alternatives. Such power could be realized by either greater sample sizes or by either GxM or non-linear models that are greater departures from the Cholesky baseline model. On this last point, our experience has been that, generally, average computational time increases no more than linearly with sample size. This is because each likelihood computation scales linearly with sample size, but larger sample sizes yield more information and hence require fewer iterations for the optimization algorithm to converge.
- Last, taking the foregoing points into account, the field has relatively little experience with these new non-linear models. Before final conclusions can be drawn, we need more experience in real applications with how well data will distinguish between competing hypotheses. Building on the prior point, we expect to learn that when there is even stronger evidence for GxM (or other non-linearities) versus the baseline Cholesky model than in the cases considered here, there may also often be reasonable evidence to distinguish between the two alternatives. Future studies are needed to see whether this is borne out.

Acknowledgments

This study was funded by the NIH grant R21 MH086099 from the National Institute for Mental Health.

References

- Jinks JL, Fulker DW. Comparison of the biometrical, MZ, and classical approaches to the analysis of the human behavior. *Psychological Bulletin*. 1970; 73(5):311–349. [PubMed: 5528333]
- Johnson W. Genetic and environmental influences on behavior: Capturing all the interplay. *Psychological Review*. 2007; 114(2):423–440. [PubMed: 17500633]
- Loehlin J. The Cholesky approach: A cautionary note. *Behavior Genetics*. 1996; 26(1):65–69.
- Neale, M.; Cardon, L. *Methodology for genetic studies of twins and families*. Vol. 67. Springer; 1992.
- Purcell S. Variance components models for gene-environment interaction in twin analysis. *Twin Research*. 2002; 5(6):554–571. [PubMed: 12573187]
- R Core Team. R: A language and environment for statistical computing [Computer software manual]. 2013. Retrieved from <http://www.R-project.org/>
- Raftery AE. Bayesian model selection in social research. *Sociological methodology*. 1995; 25:111–164.
- Rathouz P, Van Hulle C, Rodgers J, Waldman I, Lahey B. Specification, testing, and interpretation of gene-by-measured-environment interaction models in the presence of gene-environment correlation. *Behavior Genetics*. 2008; 38(3):301–315. [PubMed: 18293078]
- Van Hulle C, Lahey B, Rathouz P. Operating characteristics of alternative statistical methods for detecting gene-by-measured environment interaction in the presence of gene-environment correlation in twin and sibling studies. *Behavior Genetics*. 2013; 43(1):71–84. [PubMed: 23090766]

Zheng H, Rathouz PJ. Fitting procedures for novel gene-by-measured environment interaction models in behavior genetic designs. *Behavior Genetics*. 2015

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Qualitative descriptions of data generating mechanisms (DGMs)*.

Variance Component	DGM	
	A, A1, A2	B, B1, B2
Common Genetic	High	Low
Common Environment	Low	High
Unique Genetic	Low	High
Unique Environment	High	Low
	A1, B1	A2, B2
AxM or non-linear A	High	Low
ExM or non-linear E	Low	High

* Each DGM label is given as a suffix to DGMs listed in the text, e.g., Chol_A, CholGxM_B1, etc.

Table 2

Empirical Type I error rates for models testing for GxM or non-linear effects.

Model for H_0	Model for H_A	DGM	LRT df	Type I Error (%) $n = 800$			Type I Error (%) $n = 2000$			Type I Error (%) $n = 8000$		
				10	5	1	10	5	1	10	5	1
Chol	CholGxM	Chol_A	6	5.2	2.4	0.4	6.7	2.6	0.4	8.0	4.1	1.0
		Chol_B	6	5.7	2.3	0.5	6.0	2.9	0.5			
Chol	CorrGxM	Chol_A	3	11.5	5.5	1.0	12.1	6.5	1.2	11.3	5.7	1.4
		Chol_B	3	12.5	6.0	1.1	12.7	6.0	1.3			
Chol	CholNonLin	Chol_A	6	8.1	3.7	0.6	8.7	4.9	0.9	10.9	5.4	1.1
		Chol_B	6	7.7	3.7	0.8	8.9	4.6	0.7			
Chol	CorrNonLin	Chol_A	3	12.1	6.2	1.6	15.0	7.8	1.7	13.9	7.3	1.6
		Chol_B	3	12.6	6.8	1.3	13.6	7.1	1.5			
CorrGxM	CholGxM	Chol_A	3	1.9	0.5	0.1	3.0	1.0	0.2	7.1	3.2	0.7
		Chol_B	3	1.9	0.8	0.2	2.4	1.2	0.2			
CorrNonLin	CholNonLin	Chol_A	3	4.3	1.9	0.4	5.2	2.1	0.6	8.9	4.2	1.0
		Chol_B	3	4.8	1.8	0.5	5.3	2.3	0.4			

Each comparison is based on results of 2000 replicates.

Table 3

Bias of fitted parameters for model Chol under two DGMs. Bias is the difference between the average estimate and the true value ($\hat{a} - a$).

	Chol_A				Chol_B			
	bias		bias		bias		bias	
	true value	n = 800	n = 2000	n = 8000	true value	n = 800	n = 2000	n = 8000
μ_M	0	0.000	0.000	0.000	0	0.000	0.000	0.000
a_M	0.671	-0.007	-0.002	0.003	0.671	-0.007	-0.002	-0.002
c_M	0.316	-0.020	-0.012	-0.007	0.316	-0.020	-0.012	-0.012
e_M	0.671	-0.001	-0.002	-0.002	0.671	-0.001	-0.002	-0.002
μ_P	0	0.000	0.000	0.000	0	0.001	0.000	0.000
a_C	0.5	0.000	-0.001	0.003	0.1	-0.025	-0.012	-0.012
c_C	0.01	-0.005	-0.007	-0.011	0.01	0.003	-0.010	-0.010
e_C	0.1	0.000	-0.001	-0.003	0.5	0.003	0.002	0.002
a_U	0.806	-0.037	-0.010	0.000	0.943	-0.027	-0.007	-0.007
c_U	0.447	-0.189	-0.098	-0.024	0.447	-0.183	-0.096	-0.096
e_U	0.943	-0.004	-0.003	-0.002	0.806	-0.005	-0.004	-0.004

Table 4

Bias of fitted parameters for model CholGxM under six DGMs. Bias is the difference between the average estimate and the true value ($\hat{a} - a$).

	Chol_A		Chol_B		CholGxM_A1		CholGxM_A2		CholGxM_B1		CholGxM_B2	
	true	bias	true	bias	true	bias	true	bias	true	bias	true	bias
μ_M	0	0.000	0	0.000	0	-0.002	0	-0.003	0	-0.002	0	-0.003
a_M	0.671	-0.003	0.671	-0.003	0.671	-0.005	0.671	-0.005	0.671	-0.005	0.671	-0.005
c_M	0.316	-0.008	0.316	-0.008	0.316	-0.006	0.316	-0.007	0.316	-0.006	0.316	-0.006
e_M	0.671	-0.002	0.671	-0.002	0.671	-0.001	0.671	-0.001	0.671	-0.001	0.671	-0.001
μ_P	0	0.000	0	0.000	0	-0.001	0	-0.001	0	-0.001	0	-0.002
a_C	0.5	-0.003	0.1	-0.014	0.5	0.006	0.5	0.007	0.1	-0.004	0.1	-0.003
c_C	0.01	0.000	0.01	-0.003	0.01	-0.014	0.01	-0.016	0.01	-0.019	0.01	-0.019
e_C	0.1	-0.001	0.5	0.002	0.1	-0.004	0.1	-0.005	0.5	-0.001	0.5	-0.002
a_U	0.806	-0.017	0.943	-0.014	0.806	-0.017	0.806	-0.017	0.943	-0.010	0.943	-0.010
c_U	0.447	-0.077	0.447	-0.074	0.447	-0.083	0.447	-0.088	0.447	-0.074	0.447	-0.080
e_U	0.943	-0.006	0.806	-0.008	0.943	-0.006	0.943	-0.006	0.806	-0.007	0.806	-0.007
a_C	0	0.000	0	0.001	0.075	0.001	0.025	0.005	0.016	-0.003	0.005	0.005
a_U	0	0.000	0	0.000	0.121	-0.017	0.04	-0.002	0.151	-0.017	0.047	0.000
c_C	0	0.000	0	-0.001	0.01	-0.004	0.01	-0.009	0.01	-0.006	0.01	-0.009
c_U	0	-0.003	0	-0.002	0.01	0.013	0.01	-0.007	0.01	0.016	0.01	-0.005
e_C	0	0.000	0	-0.001	0.005	-0.001	0.013	-0.002	0.025	0.001	0.06	-0.002
e_U	0	0.000	0	0.000	0.047	0.002	0.123	-0.001	0.04	0.002	0.097	0.000

Chol_A and Chol_B are based on 2000 replicates while others are based on 200 replicates.

Table 5

Bias of fitted parameters for model CholNonLin under six DGMs. Bias is the difference between the average estimate and the true value ($\hat{a} - a$).

	Chol_A		Chol_B		CholNonLin_A1		CholNonLin_A2		CholNonLin_B1		CholNonLin_B2	
	true	bias	true	bias	true	bias	true	bias	true	bias	true	bias
μ_M	0	0.000	0	0.000	0	-0.003	0	-0.003	0	-0.003	0	-0.003
a_M	0.671	-0.005	0.671	-0.005	0.671	-0.005	0.671	-0.005	0.671	-0.003	0.671	-0.006
c_M	0.316	-0.004	0.316	-0.005	0.316	-0.007	0.316	-0.006	0.316	-0.007	0.316	-0.002
e_M	0.671	-0.001	0.671	-0.001	0.671	0.000	0.671	-0.001	0.671	-0.001	0.671	-0.001
μ_P	0	0.001	0	0.000	0	0.000	0	0.007	0	0.017	0	0.005
a_C	0.5	-0.002	0.1	-0.019	0.5	0.015	0.5	0.007	0.1	-0.007	0.1	-0.006
c_C	0.01	0.002	0.01	0.004	0.01	-0.032	0.01	-0.013	0.01	-0.011	0.01	-0.011
e_C	0.1	-0.001	0.5	0.003	0.1	-0.007	0.1	-0.006	0.5	-0.001	0.5	-0.003
a_U	0.806	-0.025	0.943	-0.020	0.806	-0.023	0.806	-0.026	0.943	-0.010	0.943	-0.015
c_U	0.447	-0.113	0.447	-0.115	0.447	-0.135	0.447	-0.130	0.447	-0.135	0.447	-0.124
e_U	0.943	-0.003	0.806	-0.004	0.943	-0.004	0.943	-0.003	0.806	-0.005	0.806	-0.005
γ_1	0	0.001	0	0.000	0.072	0.003	0.012	0.003	0.014	-0.003	0.002	0.001
γ_2	0	-0.001	0	0.001	0.01	-0.005	0.01	-0.012	0.01	-0.017	0.01	-0.007
γ_3	0	-0.001	0	-0.001	0.002	-0.001	0.01	0.000	0.012	0.001	0.055	0.000
γ_4	0	0.001	0	0.001	0.234	-0.013	0.04	-0.003	0.264	-0.017	0.047	-0.001
δ_2	0	-0.003	0	-0.005	0.01	0.010	0.01	-0.001	0.01	0.060	0.01	-0.007
δ_3	0	-0.001	0	-0.001	0.047	-0.001	0.193	-0.001	0.04	0.000	0.177	-0.001

Chol_A and Chol_B are based on 2000 replicates while others are based on 200 replicates.

Table 6

Bias of fitted parameters for model CorrGxM under six DGMs. Bias is the difference between the average estimate and the true value ($\hat{a} - a$).

	Chol_A		Chol_B		CorrGxM_A1		CorrGxM_A2		CorrGxM_B1		CorrGxM_B2	
	true	bias	true	bias	true	bias	true	bias	true	bias	true	bias
μ_M	0	0.000	0	0.000	0	-0.002	0	-0.002	0	-0.002	0	-0.003
a_M	0.671	-0.002	0.671	-0.002	0.671	-0.003	0.671	-0.003	0.671	-0.003	0.671	-0.004
c_M	0.316	-0.011	0.316	-0.010	0.316	-0.012	0.316	-0.012	0.316	-0.009	0.316	-0.007
e_M	0.671	-0.002	0.671	-0.002	0.671	-0.001	0.671	-0.001	0.671	-0.001	0.671	-0.001
μ_P	0	0.000	0	0.000	0	0.001	0	0.000	0	0.000	0	0.000
a_P	0.949	-0.002	0.949	-0.006	0.949	0.001	0.949	0.002	0.949	-0.001	0.949	-0.002
c_P	0.447	-0.027	0.447	-0.020	0.447	-0.030	0.447	-0.031	0.447	-0.016	0.447	-0.016
e_P	0.949	-0.005	0.949	-0.005	0.949	-0.005	0.949	-0.005	0.949	-0.006	0.949	-0.006
r_A	0.527	-0.010	0.105	-0.017	0.527	0.004	0.527	0.003	0.105	-0.005	0.105	-0.009
r_C	0.022	-0.002	0.022	-0.066	0.022	-0.046	0.022	-0.049	0.022	-0.103	0.022	-0.092
r_E	0.105	0.001	0.527	0.004	0.105	-0.004	0.105	-0.004	0.527	0.001	0.527	0.002
a_P	0	-0.001	0	0.000	0.114	-0.003	0.047	0.004	0.109	-0.013	0.047	-0.003
κ_P	0	0.002	0	-0.001	0.01	0.000	0.010	-0.016	0.01	0.020	0.01	0.002
e_P	0	0.000	0	0.000	0.047	0.000	0.095	-0.001	0.047	0.001	0.090	-0.001

Chol_A and Chol_B are based on 2000 replicates while others are based on 200 replicates.

Table 7

Bias of fitted parameters for model CorrNonLin under six DGMs. Bias is the difference between the average estimate and the true value ($\hat{a} - a$).

	Chol_A		Chol_B		CorrNonLin_A1		CorrNonLin_A2		CorrNonLin_B1		CorrNonLin_B2	
	true	bias	true	bias	true	bias	true	bias	true	bias	true	bias
μ_M	0	0.000	0	0.000	0	-0.002	0	-0.002	0	-0.003	0	-0.003
a_M	0.671	-0.003	0.671	-0.003	0.671	-0.003	0.671	-0.001	0.671	-0.001	0.671	-0.005
c_M	0.316	-0.011	0.316	-0.010	0.316	-0.013	0.316	-0.018	0.316	-0.013	0.316	-0.006
e_M	0.671	-0.002	0.671	-0.002	0.671	-0.001	0.671	-0.001	0.671	-0.001	0.671	-0.001
μ_P	0	0.000	0	0.000	0	-0.012	0	-0.002	0	0.012	0	0.004
a_P	0.949	-0.005	0.949	-0.007	0.949	-0.001	0.949	-0.001	0.949	0.001	0.949	-0.003
c_P	0.447	-0.068	0.447	-0.061	0.447	-0.083	0.447	-0.077	0.447	-0.071	0.447	-0.062
e_P	0.949	-0.003	0.949	-0.002	0.949	-0.003	0.949	-0.003	0.949	-0.004	0.949	-0.004
r_A	0.527	-0.008	0.105	-0.017	0.527	-0.004	0.527	-0.001	0.105	-0.001	0.105	-0.011
r_C	0.022	0.026	0.022	-0.044	0.022	0.005	0.022	-0.041	0.022	-0.095	0.022	-0.065
r_E	0.105	0.000	0.527	0.003	0.105	-0.003	0.105	-0.004	0.527	-0.000	0.527	0.000
λ_1	0	0.002	0	0.000	0.171	-0.005	0.047	-0.003	0.199	-0.011	0.047	-0.002
λ_2	0	-0.011	0	-0.006	0.01	0.009	0.01	0.001	0.01	0.034	0.01	-0.002
λ_3	0	-0.001	0	-0.001	0.047	-0.002	0.152	-0.002	0.047	-0.001	0.142	-0.002

Chol_A and Chol_B are based on 2000 replicates while others are based on 200 replicates.

Rates of model selection between model CholGxM and model CholNonLin at three BIC thresholds under Cholesky-based data generating mechanisms.

Table 8

Condition	DGM	H_0	H_1	H_1 is preferred (%)			H_0 & H_1 are equivocal (%)		
				BIC - 10	BIC - 6	BIC - 2	BIC ± 10	BIC ± 6	BIC ± 2
1	CholGxM_A1	Chol	CholGxM	90.5	92.5	95.0	9.0	7.0	3.0
		Chol	CholNonLin	88.4	91.5	95.5	11.1	8.0	1.5
		CholNonLin	CholGxM	0.5	8.1	34.8	98.5	87.4	47.5
2	CholGxM_A2	Chol	CholGxM	93.5	97.0	99.0	6.5	3.0	0.5
		Chol	CholNonLin	86.9	92.0	95.5	12.6	7.0	2.0
		CholNonLin	CholGxM	36.2	57.3	87.9	63.3	41.7	8.5
3	CholGxM_B1	Chol	CholGxM	94.5	97.5	98.5	5.5	2.0	0.5
		Chol	CholNonLin	93.5	97.5	98.5	6.5	2.0	0.5
		CholNonLin	CholGxM	1.5	8.6	40.9	98.0	87.9	38.4
4	CholGxM_B2	Chol	CholGxM	93.0	96.0	98.5	7.0	4.0	0
		Chol	CholNonLin	86.4	89.9	93.5	12.6	8.5	3.0
		CholNonLin	CholGxM	30.2	55.3	81.9	69.8	44.2	14.6
5	CholNonLin_A1	Chol	CholNonLin	93.0	96.0	97.0	7.0	4.0	2.5
		Chol	CholGxM	76.5	87.0	91.5	23.0	10.0	3.5
		CholGxM	CholNonLin	56.5	76.5	93.5	43.5	23.5	5.0
6	CholNonLin_A2	Chol	CholNonLin	91.0	96.5	98.0	9.0	3.5	0.5
		Chol	CholGxM	75.0	82.5	91.0	23.5	14.5	4.5
		CholGxM	CholNonLin	54.8	79.9	95.0	45.2	19.6	4.5
7	CholNonLin_B1	Chol	CholNonLin	90.9	95.5	98.0	9.1	3.5	0.5
		Chol	CholGxM	76.0	81.5	90.0	22.0	15.0	2.0
		CholGxM	CholNonLin	68.2	87.4	97.0	31.8	12.1	2.5
8	CholNonLin_B2	Chol	CholNonLin	93.5	97.5	98.5	6.5	2.5	0.5
		Chol	CholGxM	86.5	89.0	94.0	12.5	9.5	3.5
		CholGxM	CholNonLin	53.0	74.5	94.0	47.0	25.5	5.5

Table 9

Rates of model selection between model CorrGxM and model CorrNonLin at three BIC thresholds under correlated factors-based data generating mechanisms.

Condition	DCM	H_0	H_1	H_1 is preferred (%)			H_0 & H_1 are equivocal (%)		
				BIC - 10	BIC - 6	BIC - 2	BIC ±10	BIC ±6	BIC ±2
9	CorrGxM_A1	Chol	CorrGxM	92.8	97.4	99.5	7.2	2.1	0
		Chol	CorrNonLin	93.3	96.9	99.0	6.7	2.6	0.5
		CorrNonLin	CorrGxM	1.0	8.3	37.8	98.4	88.6	44.0
10	CorrGxM_A2	Chol	CorrGxM	93.8	96.9	99.0	6.2	3.1	1.0
		Chol	CorrNonLin	90.2	94.8	97.9	9.3	4.6	1.0
		CorrNonLin	CorrGxM	12.4	32.6	67.4	87.6	65.3	25.4
11	CorrGxM_B1	Chol	CorrGxM	93.4	96.5	98.5	6.6	3.5	1.0
		Chol	CorrNonLin	91.9	97.5	98.0	8.1	2.5	1.0
		CorrNonLin	CorrGxM	0	7.7	36.2	100	88.8	44.4
12	CorrGxM_B2	Chol	CorrGxM	95.4	98.0	98.5	4.6	2.0	1.0
		Chol	CorrNonLin	89.3	93.9	98.0	10.7	5.1	0.5
		CorrNonLin	CorrGxM	10.2	35.2	68.4	89.8	63.8	23.0
13	CholNonLin_A1	Chol	CorrNonLin	90.6	94.8	98.4	9.4	5.2	1.0
		Chol	CorrGxM	81.4	89.2	94.8	18.0	10.3	3.6
		CorrGxM	CorrNonLin	19.3	47.4	78.1	80.2	52.1	16.1
14	CholNonLin_A2	Chol	CorrNonLin	92.3	97.4	98.5	7.7	2.6	1.0
		Chol	CorrGxM	82.4	92.2	96.9	17.6	7.3	1.0
		CorrGxM	CorrNonLin	21.8	47.2	78.2	77.7	51.8	19.2
15	CholNonLin_B1	Chol	CorrNonLin	92.0	95.5	98.0	8.0	4.5	1.5
		Chol	CorrGxM	82.9	91.0	94.5	16.6	8.5	2.5
		CorrGxM	CorrNonLin	24.6	47.2	82.9	75.4	52.3	13.1
16	CholNonLin_B2	Chol	CorrNonLin	94.9	96.4	98.5	5.1	3.6	1.5
		Chol	CorrGxM	88.8	94.4	98.0	11.2	5.1	0.5

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Condition	DCM	H_0	H_1	H_1 is preferred (%)				H_0 & H_1 are equivocal (%)		
				BIC - 10	BIC - 6	BIC - 2	BIC ±10	BIC ±6	BIC ±2	
		CorrGxM	CorrNonLin	28.4	46.7	80.2	71.6	52.8	16.8	