Promoting Measured Genes and Measured Environments: On the Importance of Careful Statistical Analyses and Biological Relevance

Research testing the interaction between measured genes and measured environments in psychiatric disorders was promoted in a recent review by Moffitt et al. in the ARCHIVES. In presenting the emerging gene x environment interaction findings, Caspi et al. cite their finding of an interaction between the genetic variants of monoamine oxidase A conferring low enzymatic activity and childhood maltreatment to increase the risk for violent behavior. Moffitt et al. cite one replication of this finding by Foley et al. that was published in the ARCHIVES. A careful perusal of this latter study, which has now been cited 19 times, shows that it is flawed both in its analyses and interpretation. First, Foley et al. argued that the results obtained with logistic regression are likely to be more robust than results obtained within a linear regression framework. However, logistic regression has the disadvantage of collapsing ordinal or count response variables into a dichotomous variable, which may result in loss of information. Ordinal regression, negative binomial regression, or Poisson regression models are robust and more appropriate techniques when analyzing disorder symptom counts. Second, although Foley et al. favored logistic over linear regression to avoid false-positive interactions due to scaling artifacts (heteroscedasticity), they did not assess or report the fit of their model. The presence of zero or small cell counts in interaction terms (as evident from Table 2 in the Foley et al. article) may cause numerical problems in the modeling stage of the analysis. Using the raw data provided in the Foley et al. Table 2, we found that the logistic regression model presented has poor fit (Hosmer-Lemeshow test, \( \chi^2 = 8.9; P = .06 \)). However, model fit improved when we grouped categories 2, 3, and 4 of environmental adversity and used 3 (0, 1, and 2-4) instead of 5 categories (\( \chi^2 = 5.5; P = .14 \)). However, the interaction between monoamine oxidase A and environmental adversity was nonsignificant (\( P = .36, 2\text{-}sided \text{ test} \)). This is not surprising and could have been suspected simply by noticing that too many cells (7 [35%] of 20) in Table 2 had between 0 and 4 observations. It is also surprising to see the misleading Figure published in the highly reputed ARCHIVES, where a strong visual effect of interaction is in fact due to 1 observation made on a sample size of \( n=1 \) (1/1 = 100%). Finally, in this study, the monoamine oxidase A genotypes conferring low enzymatic activity are associated with a decreased risk of antisocial behavior whereas the same genotypes in combination with environmental adversity are associated with the opposite effect. In contradiction with the principal of parsimony, Foley et al. interpreted this observation as an important finding indicating the complicated nature of psychiatric genetics. This may simply reflect the lack of rigor in the application of statistical methods to complex psychiatric disorders.

In a more recently published study, Thapar and colleagues reported that the catechol-O-methyltransferase (COMT) Val/Val genotype is associated with increased symptoms of conduct disorder particularly in children with lower birth weight. Contrary to Foley et al., they opted to use multiple regression as their primary analysis. However, the Figure provided in the Thapar et al. article suggests that the distribution of the outcome variable is highly skewed (the majority of children do not have conduct disorders). Further, birth weight was not corrected for gestational age and we do not know whether it was checked for outliers or not. Unfortunately, we could not assess these critical issues given the lack of basic information such as demographic characteristics (along with their standard deviation) of the 3 genotype groups by birth weight and dispersion parameters of conduct disorder symptom scores for each of the different groups. All these arguments seriously call into question the validity of the linear regression model and the results of this study. Nevertheless, Thapar et al. applied logistic regression, which, even though robust to heteroscedasticity, does not address all the other concerns raised herein. Remarkably, all the significant results became only marginal when logistic regression was used.

Finally, we call into question the hypothesis advanced in the Thapar et al. study, which is based on the “links between COMT and prefrontal cortical functioning,” when a previous study by the same group on the same population concluded that the “Val158Met COMT genotype is not associated with neurocognitive performance (neurocognitive tests of prefrontal cognition).”

In conclusion, while it is important to investigate gene x environment interactions in psychiatric disorders, we underline the importance of rigorous application of statistical methods while avoiding potential bias, including review and publication biases, and necessitating biological relevance. Failure to do so may result in statistically significant results that may be biologically irrelevant and serve only to wrongly heighten expectations.

Ridha Joober, MD, PhD
Sarojini Sengupta, PhD
Norbert Schmitz, PhD
We appreciate the opportunity to respond to the letter by Drs Joober, Sengupta, and Schmitz. We were attempting to replicate a study conducted by Caspi et al and we therefore sought to replicate the methods and measures they used insofar as we could. We used logistic regression to model risk for conduct disorder because Caspi et al used logistic regression to model risk for conduct disorder. Logistic regression also has the advantage of being relatively robust to scaling artifacts (heteroscedasticity), which can be an issue when modeling interactions. We created a simple count variable of each measured adversity to obtain a quantitative measure of severity of exposure to family adversity. In practice, the architecture of familial adversity is likely to be complex and we would welcome efforts to address this issue in larger samples with greater power. The issue of statistical power is not a trivial one and our sample size was relatively modest (n = 514 male twins). We therefore published our data in a transparent way to permit evaluation and reanalysis. The reported interaction was not estimated on the basis of a single data point and 1-sided significance tests should be conducted when there is a clear directional hypothesis based on prior work. Ultimately, functional studies will be required to validate observed statistical associations, including statistical interactions, consistent with the presence of genotype × environment interactions. Until then, meta-analysis of published studies provides one way of attempting to summarize the strength of statistical evidence across studies of variable size.

Debra Foley, PhD

Brien Riley, PhD

Correspondence: Dr Foley, University of Melbourne, ORYGEN Research Centre, Applied Genetics Unit, 35 Poplar Rd, Parkville, Victoria 3052, Australia (dfoley@unimelb.edu.au).

In reply

We agree with Joober and colleagues about the importance of careful statistical analysis in all research but do not agree with their criticisms of our article. We would like to point out that the statistical issues raised about regression analysis are not specific to testing for gene × environment interaction. They apply to testing for statistical interactions regardless of whether a given interaction term involves social × social variables (eg, family conflict × sex), gene variant × gene variant, or interaction between any other set of quantitative variables. We are surprised that, having first noted that information can be lost when continuous variables are dichotomized and logistic regression used, Joober and colleagues then find it remarkable that the significant results in our article became only marginal when logistic regression was used. Do they expect information loss to lead to anything else? Far from a cause for concern, it is reassuring that the results remained significant when logistic regression was applied using the dichotomous variable of conduct disorder / no conduct disorder and unsurprising that the magnitude of the effect was less than that observed when we examined continuous data. Since the results are statistically significant under both analyses, one can be more confident that they are genuine and not merely artifacts of the violation of distributional assumptions, as implied by Joober and colleagues. Furthermore, transforming the symptom scores so that they approximately followed a normal distribution did not affect the significance of either the main effects of the predictor variables or their interactions.

Regarding their concern about gestational age, we can reassure Joober and colleagues that this is not an issue since we did not include preterm births in our sample. Dr Joober and colleagues question the hypothesis advanced, which is based on “the links between COMT [catechol-O-methyltransferase] and prefrontal cortical function,” on the grounds that we have previously failed to find a link between measures of cognitive functioning and COMT genotype. As Joober and colleagues are no doubt aware, lack of association between COMT and neurocognitive measures in one study is not equivalent to robust evidence against the existence of such a relationship. Our previous lack of association with neurocognitive measures simply means that none of the measures we have available (none are of prefrontal cortical functioning) are potential confounds or mediators of effects.

We have 2 other general points. Any finding obtained by a statistical analysis, regardless of the field of research, is, of course, probabilistic and thus requires replication. Meta analyses provide a good test of the robustness of effects, as has recently been shown for the monoamine oxidase A / maltreatment findings. Second, statistical interaction can-
not be equated with biological interaction; statistical interaction represents interplay between 2 (or more) passive quantitative elements, and biological interaction represents interplay between 2 (or more) active qualitative elements (the former does not inevitably equate to the latter). Finally, research on mental health problems and psychiatric disorder (whether epidemiological, etiological, or treatment studies) needs to be conceptually and statistically rigorous, with replication set as the gold standard by which effects are evaluated.

Anita Thapar, MD
Gordon Harold, PhD
Peter Holmans, PhD
Frances Rice, PhD
Kate Langley, PhD
Tom Fowler, PhD
John Aggleton, PhD
Marianne van den Bree, PhD
Michael Owen, MD
Michael O'Donovan, MD

Correspondence: Dr Thapar, Department of Psychological Medicine, Cardiff University School of Medicine, Heath Park, Cardiff CF4 4XN, Wales (thapar@cardiff.ac.uk).

Financial Disclosure: None reported.


Errors in Assessing DSM-IV Substance Use Disorders

In the June 2005 issue of the Archives, Kessler and colleagues\(^1\) reported results from the National Comorbidity Survey Replication (NCS-R), a 2001-2002 US psychiatric epidemiology survey of DSM-IV mental and substance use disorders.\(^2\)\(^3\)\(^4\)\(^5\) Included are prevalences and associated features, such as treatment of current (last 12 months) disorders. Given the fundamental nature of this study and the potential policy relevance, it is essential that rates estimated by the study be accurate. We are concerned about the accuracy of the rates reported for alcohol and drug abuse and dependence disorders. The specific problem is that the study purports to apply DSM-IV criteria, but the questionnaire used by Kessler and colleagues, the World Mental Health–Composite International Diagnostic Interview (WMH-CIDI),\(^2\) skips all respondents' past questions on DSM-IV dependence if they do not respond positively to questions on DSM-IV abuse, effectively using abuse as a screen for dependence. This same procedure has been used in the recent National Institute of Mental Health–sponsored NCS-A adolescent study; the NCS-2, a 10-year follow-up of the original NCS conducted in 1990-1992;\(^6\) and the National Study of African American Life, and the National Study of Latino and Asian Americans.\(^7\) Beyond the United States, the NCS-R methods have been used in epidemiological surveys under way in 27 other countries under the aegis of the World Health Organization (WHO) World Mental Health Survey Initiative.\(^8\)

Using abuse symptoms as a screen for dependence could be justified if DSM-IV required abuse symptoms for a dependence diagnosis. However, this is not the case. Alcohol abuse and alcohol dependence are conceptually and diagnostically separate and distinct entities.\(^9\) One is not a precursor or predictor of the other. In support of this distinction, prospective studies of representative samples have shown that most persons initially diagnosed as abusers do not progress to dependence 1 to 15 years later.\(^10\)\(^11\) Using abuse symptoms as a screen for current dependence might also be justifiable if virtually all dependence cases had abuse symptoms. For both alcohol dependence and drug dependence, this has been shown not to be the case, using data from the 2001-2002 National Institute on Alcohol Abuse and Alcoholism National Epidemiologic Survey on Alcohol and Related Conditions (NESARC).\(^15\)\(^16\) which used highly reliable, well-validated instruments\(^17\)\(^18\) in a nationally representative sample of 43,993 respondents in which all substance users were administered dependence items regardless of their responses on abuse. In the first of these studies, reported in the Archives, 34% of current alcohol dependence cases did not report abuse criteria, and those who did not were disproportionately women and minorities, traditionally underserved groups.\(^29\) The second study, of drug dependence and abuse, shows that 20% of current drug dependence cases did not meet drug abuse criteria, also disproportionately women.\(^30\) Consistent with this error, current rates of alcohol and drug dependence are markedly lower in the NCS-R (1.3% and 0.4%) than in the NESARC (3.8% and 0.6%).\(^16\)

The use of abuse as a screen for dependence also appears to have contributed to a major decrease in the prevalence of any current substance use disorder from the 1990-1992 NCS (11.3%),\(^6\) where the abuse skipping pattern was not used, to the NCS-R (3.8%), where it was used.\(^1\) Such decreases have not been found in other studies covering the same period.\(^31\)\(^32\) Dependence diagnoses made by the WMH-CIDI exhibit poor validity and low sensitivity when compared with clinical reappraisal.\(^2\)

Since dependence is more strongly associated with psychiatric comorbidity and treatment than abuse, underestimation of alcohol and drug dependence will produce biases and errors in reports on prevalence, sociodemographic and clinical correlates, comorbidity, and treatment use from the NCS-R and associated studies, as well as the WHO-WMH surveys.\(^14\)\(^15\)\(^16\) Furthermore, in future NCS-R and WHO-WMH estimates of the burden of mental illness, alcohol and drug use disorders will be underestimated as a cause of disability, which could lead to reductions in support and services for those with alcohol and drug dependence worldwide. The seriousness of this problem and the disservice to those with alcohol and