



Published in final edited form as:

*Alcohol Clin Exp Res.* 2013 March ; 37(3): 443–451. doi:10.1111/j.1530-0277.2012.01954.x.

## Assessment of a modified DSM-5 diagnosis of alcohol use disorder in a genetically informative population

Alexis C. Edwards, Ph.D.<sup>1</sup>, Nathan A. Gillespie, Ph.D., Steven H. Aggen, Ph.D., and Kenneth S. Kendler, M.D.

Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA

### Abstract

**Background**—Proposed changes to the upcoming DSM-5 include: i) combining criteria for DSM-IV alcohol abuse (AA) and alcohol dependence (AD) into one diagnostic category (alcohol use disorder, AUD); ii) exclusion of the “legal problems” (LP) criterion; and iii) addition of a “craving” criterion. Few published studies empirically assess the potential consequences of the proposed changes.

**Methods**—Using a population-based sample of twins assessed for lifetime AA/AD diagnoses, we explored phenotypic differences across DSM-IV and a modified DSM-5 diagnoses without craving due to its unavailability in the dataset. We used factor analysis and item response theory (IRT) to evaluate the potential consequences of excluding the LP criterion from AUD, and used twin modeling to examine genetic differences between DSM-IV and the modified DSM-5 diagnoses.

**Results**—The prevalence of AUD was slightly higher than that of DSM-IV diagnoses. Individuals meeting DSM-IV or DSM-5 criteria, but not both, exhibit fewer comorbid diagnoses than those meeting both sets of criteria. Individuals meeting only DSM-5 criteria were slightly less severely affected than those meeting only DSM-IV criteria. Factor analysis indicated that the LP criterion loading is the lowest of all symptoms; IRT analysis suggested that this criterion has low discriminatory power. The genetic correlation between DSM-IV and DSM-5 diagnoses was slightly but significantly lower than unity.

**Conclusions**—The proposed DSM-5 AUD criteria are unlikely to result in significant changes in prevalence of diagnosed alcohol problems. However, it is unclear whether the new criteria represent a more valid diagnosis: new cases are no more severely affected than DSM-IV-only cases. Given the psychometric properties of the LP, its exclusion should not negatively impact diagnostic validity. Similarly, the stable heritability across DSM-IV and DSM-5 diagnoses suggest that the proposed changes will not have substantial negative consequences in terms of familial influences, a key validator. The current results provide equivocal empirical support for the proposed DSM-5 changes for AUDs.

### Keywords

DSM-5; alcohol use disorder; diagnostic validity

---

<sup>1</sup>Corresponding author: Alexis C. Edwards, Ph.D. Virginia Institute for Psychiatric and Behavioral Genetics Department of Psychiatry Virginia Commonwealth University Box 980126 Richmond, VA 23298-0126 Phone: 804-828-8591 Fax: 804-828-1471 aedwards5@vcu.edu.

## Introduction

Current proposals for the upcoming fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) include several changes to the criteria used to assess alcohol problems. First, the current structure, which consists of two diagnoses (alcohol abuse [AA] and alcohol dependence [AD]), would be collapsed into a single construct, alcohol use disorder (AUD) (see below). Second, the AA criterion concerning legal problems would be removed. Third, a criterion addressing craving would be added. The rationales behind these changes are described in detail on the DSM-5 website ([www.dsm5.org](http://www.dsm5.org)). These proposed changes could potentially result in a qualitatively different diagnosis. Diagnoses derived using different criteria might also differ etiologically, which could have treatment and prevention implications. Empirical assessment of the proposed changes can help us understand substantive differences across the diagnoses.

Two recent studies have explored the potential impact of the proposed changes in diagnostic criteria (Agrawal et al., 2011; Mewton et al., 2011) using large, population-based samples. Mewton and colleagues compared the 12-month prevalence of alcohol use disorder (AUD) across DSM-IV (American Psychiatric Association, 2000) and DSM-5 diagnosis, and also investigated the psychometric properties of each criterion. Agrawal and colleagues also examined changes in 12-month prevalence across the diagnoses. Results varied slightly across samples: Mewton and colleagues reported a substantial increase in prevalence, while Agrawal et al. reported only a minor increase. Mewton et al. reported equivocal support for the inclusion of a craving criterion; Agrawal et al. found that its inclusion would not be disruptive in terms of prevalence, but did not assess the psychometric properties of the item. Mewton et al. reported IRT analyses that supported the exclusion of the legal problems criterion; Agrawal and colleagues reported that this change would have little impact on prevalence.

Other recent studies have explored individual components of the proposed diagnostic changes or used smaller/selected samples. Findings generally support the inclusion of an item addressing craving (Keyes et al., 2011; Casey et al., 2012) and indicate that the combination of current abuse and dependence criteria is appropriate, as these items typically load onto a single factor (Casey et al., 2012; Hasin et al., 2012). In addition, research suggests that, except in the case of illicit substances such as cannabis and cocaine (Gillespie et al., 2007), the criterion addressing legal problems is rarely endorsed and is inconsistently informative (Hasin et al., 2012; Hagman et al., 2011); thus, its exclusion is unlikely to adversely impact diagnosis.

None of the above studies were conducted on genetically informative samples, leaving unanswered the question of differences in genetic influences across diagnostic criteria. Familial aggregation is considered one key validator for psychiatric illness (Kendler, 1990; Robins and Guze, 1970): therefore, it is of interest whether the DSM-IV and DSM-5 diagnoses are similarly genetically influenced, both in terms of total heritability and potential qualitative differences in the underlying genetic factors. In other words, is a DSM-5 diagnosis informative of familial risk to the same extent that a DSM-IV diagnosis does? Do the two diagnoses reflect the same underlying genetic liability or is the new diagnosis influenced by novel genetic factors? To our knowledge, this question has not yet been addressed by studies exploring the ramifications of changes in diagnostic criteria.

In this report, we attempt to replicate the findings of recent reports (e.g., Agrawal et al., 2011; Mewton et al., 2011), evaluating the consequences of the proposed modification of diagnostic criteria from DSM-IV AA and AD to DSM-5 AUD, and using lifetime diagnosis rather than 12-month prevalence as the focal phenotype. Furthermore, we expand on those

analyses using a genetically informative sample of population-based twins. We employ four methods: i) a comparison of the clinical characteristics of three groups: individuals meeting DSM-IV but *not* DSM-5 criteria (D4-only), those meeting DSM-5 but *not* DSM-IV criteria (D5-only), and those meeting both DSM-IV and DSM-5 criteria (D4D5); ii) a factor analysis of AA and AD symptoms, with the primary objective of evaluating the factor loading for the legal problems symptom; iii) an IRT analysis, with the objective of evaluating the parameters of the legal problems symptom relative to other AA and AD symptoms; and iv) a genetic epidemiologic approach, wherein we test for genetic influences loading onto DSM-5 AUD but *not* DSM-IV AA/AD.

## Methods

### Sample

Participants were part of the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD), which has been previously described (Kendler and Prescott, 2006). The current sample includes same-sex female twins from the fourth wave of data collection for that subsample (N=1910; 1137 members of monozygotic [MZ] pairs; 773 members of dizygotic [DZ] pairs), and same-sex male twins and opposite sex twins from the second wave of data collection for that subsample (N=5544; 1643 members of male MZ pairs, 1250 members of male DZ pairs, and 2651 members of opposite-sex pairs). The sample is 56% male. Consistent with prior reports using this data (Kendler and Prescott, 2006), the modest number of lifetime abstainers from alcohol in this sample were considered to be unaffected.

Zygosity was determined using a combination of self-report measures, photographs, and genotyping (Kendler and Prescott, 2006). The project received human subjects approval from Virginia Commonwealth University and participants provided informed consent prior to all interviews.

### Measures

**DSM-IV alcohol abuse and alcohol dependence**—Participants were administered an extensive series of interview items adapted from the SCID interview (Spitzer and Williams, 1985), aimed at deriving the presence/absence of symptoms of DSM-IV AD and AA. A diagnostic algorithm was then established to derive a lifetime diagnosis of AA or AD based on these items and their temporal clustering.

**DSM-5 alcohol use disorder**—The proposed revisions for the DSM-5 include exclusion of the AA symptom “recurrent alcohol-related legal problems”, hereafter referred to as “legal problems” (LP). A new symptom will be added addressing craving (“craving or a strong desire or urge to use a specific substance”). Because craving was not measured, we were not able to derive a formal AUD diagnosis; readers should bear this in mind, as the term “AUD diagnosis” will be used throughout the manuscript for the sake of brevity. We were able to construct a variable that excluded the LP symptom and required the endorsement of at least two other symptoms. Since only 10 of the 11 possible symptoms were available, the prevalence of AUD reported here is likely a minor underestimate.

**Other psychiatric and substance use disorders**—Lifetime diagnoses for psychiatric and substance use disorders were derived using items from the SCID (Spitzer and Williams, 1985) as described previously (Kendler and Prescott, 2006).

## Statistical Analyses

**Factor analysis**—Previous studies have reported that the factor structure of AA and AD is best described by a single dimension (Feingold and Rounsaville, 1995; Langenbucher et al., 2004; Saha et al., 2006). This finding was replicated for the current data: the first eigenvalue is 8.407, followed by values well below 1 (the second value is 0.524). A confirmatory factor analysis was run in Mplus 5.2 (Muthen and Muthen, 1998-2007), using only one twin per pair (the first). All symptoms of AA and AD were modeled as binary variables. The standardized factor loadings onto each symptom are reported.

**Item response analysis.** An item response analysis was conducted in Mplus 5.2 (Muthen and Muthen, 1998-2007), using only one member of each twin pair. Each AA and AD symptom was included as a categorical variable. Two parameters, difficulty (also called “threshold” [Langenbucher et al., 2004]) and discrimination, were estimated using a 2 parameter IRT model using a probit metric. As discussed by Takane and de Leeuw (1987), there is a direct equivalence of the normal ogive item response model to the factor analysis model for binary data. Essentially, the item difficulty parameters correspond to the item thresholds, while the discrimination parameters correspond to the factor loadings. Difficulty refers to the location on the underlying latent continuous liability to alcohol problems at which the probability of endorsing each symptom is 50%. Discrimination measures the degree to which a given symptom distinguishes individuals with liability levels below and above that symptom's difficulty location. A symptom with a low discrimination parameter does not effectively distinguish such individuals.

**Twin modeling**—We conducted twin modeling in Mx (Neale et al., 2006) using full information maximum likelihood raw ordinal data methods. The use of ordinal data assumes that the observed categories represent an underlying normal distribution of liability, such that the proportions of the population lying between adjacent thresholds exactly matches the observed proportion of the sample in each category. In twin modeling, liability to phenotypes such as depression or AUD can be decomposed to three common latent sources of variance: additive genetic factors (A), shared environment (C), and unique environment (E). The C component represents environmental exposures and experiences that are shared by both members of a twin pair and contribute to twins' similarity irrespective of zygosity in a given phenotype. Environmental factors that are unique to one twin are accounted for by the E component; these factors reduce twin similarity for a given phenotype. The E component also includes random measurement error. Estimates of each of these variance components are calculated by comparing the phenotypic correlation between monozygotic twins, who share all their genes, to dizygotic twins, who share half of their genes on average identical by descent.

Genetic and environmental influences on AA and AD, including significant qualitative genetic sex differences, have been previously reported for this sample: Prescott and colleagues (Prescott et al., 1999) found evidence of non-overlapping genetic influences on AA and AD, and reported heritability estimates of 51%-66% across diagnoses and sexes. The analyses reported here are extensions of these previous findings: bivariate models were used to assess the significance of genetic factors influencing AUD *but not* AA/AD (represented by the dashed lines in Figure 1). Based on the prior findings (Prescott et al., 1999), the model allowed for qualitative genetic differences across the sexes. Model fitting was limited to testing the significance of common environmental effects (in keeping with Prescott et al., 1999) and the significance of genetic influences loading onto AUD but not onto AA/AD. The phenotypes used for these analyses were the AUD phenotype described above, and a variable that combines the AA and AD diagnoses. For the DSM-IV phenotype, any individuals who met diagnostic criteria for either AA *or* AD was coded as “1”; all others

were coded as “0”; for the DSM-5 phenotype, any individual who endorsed 2 or more AUD symptoms was coded as “1”; those endorsing 1 or 0 symptoms were coded as “0”.

## Results

### Preliminary analyses and descriptive statistics

The proportion of the sample endorsing each AA and AD criterion is reported in Table 1. Nearly one-third (30.6%, N=2281) of the sample met formal criteria for either AA (11.7%, N=875) or AD (18.9%, N=1406); most (88.8%, N=1248) AD cases also endorsed at least one AA symptom. A slightly larger proportion of the sample met lifetime DSM-5 AUD criteria (32.1%, N=2398). This increase disproportionately affects women, for whom the prevalence increases from 16.9% to 18.8%; for men, these proportions are higher but change less (from 41.4% to 42.7%). The prevalence difference is due to two groups: i) individuals who endorsed only 1 AA symptom, met criteria for AA, but do not meet criteria for AUD (D4-only); and ii) individuals who endorsed 2 AD symptoms and 0 AA symptoms, and therefore meet AUD criteria but not AA or AD criteria (D5-only). The proportions of the sample in each group, including D4D5 (those meeting both DSM-IV and DSM-5 criteria) and those who are unaffected under both sets of criteria, are depicted in Figure 2.

Of those meeting DSM-5 AUD criteria, 38.3% (N=919) could be classified as “moderate” AUD, with the rest being classified as “severe.” Of the 2281 individuals receiving a DSM-IV diagnosis, 7.6% (N=174, D4-only) would be excluded from a DSM-5 diagnosis. “Hazardous use” was the most commonly endorsed (39.7%) criterion among these individuals. New cases (N=291, D5-only) are disproportionately female (see Table 2). The most commonly endorsed AD symptoms among these individuals were “loss of control” (78.0%) and “trying to cut down” (50.5%).

### Comparison of clinical features across groups

We next explored clinical characteristics of the three groups described above (D4-only, D5-only, and D4D5), including the prevalence of psychiatric and substance use disorders that are commonly comorbid with AUD; comorbidity is frequently associated with severity (Boschloo et al., 2011; Chen et al., 2011). Table 2 provides details on these findings. Individuals in D4D5 had an earlier age-of-onset of symptoms relative to D5-only, but did not differ significantly from D4-only. D4D5 cases reported consuming more drinks in a 24-hour period than both D4-only and D5-only. The prevalence of nearly every psychiatric and substance use disorder examined was significantly higher within D4D5 relative to those in D4-only and D5-only.

We also directly compared these characteristics across D4-only and D5-only (Table 2, right-most column): if DSM-5 criteria represent a more valid diagnosis, we would expect a difference in clinical characteristics across these groups, with D5-only exhibiting more severe phenotypes. For most of the features examined, individuals in D4-only exhibited slightly – but not significantly – higher prevalence/severity (e.g., a younger age of onset of symptoms) than those in D5-only. Differences across the groups in the prevalence of cannabis abuse/dependence did reach significance. Furthermore, z-statistics provide some evidence that, with regards to externalizing disorders (e.g., conduct disorder, illicit drug diagnoses), D4-only is less different from D4D5 than is D5-only. These differences, however, do not meet significance criteria in the direct comparison.

### Confirmatory factor analysis

Previous studies have established that a single common factor model adequately describes the abuse and dependence criteria (Krueger et al., 2004; Mewton et al., 2011; Proudfoot et

al., 2006). Accordingly, we fit a model with a single latent factor. Model fit was quite good (CFI=0.995; TLI=0.998; RMSEA=0.031). Standardized factor loadings for each of the 11 AA and AD symptoms are included in Table 3. Although the factor loading for LP was strong, it had the lowest loading compared to all the other symptoms. Results presented represent the sexes pooled, as loadings did not qualitatively differ across the sexes (i.e., for both men and women, the factor loading of LP was lower than for any other symptom). A formal test of full measurement invariance indicated that sex differences exist (robust chi square difference test =38.6, df=7,  $p<0.001$ ), as expected given sex differences in thresholds for each item. Fit indices were nearly identical across the constrained and unconstrained models.

### Item response analysis

Table 3 describes the parameters of each AA and AD symptom; item characteristic curves are depicted in Figure 3. Consistent with its low endorsement rate, the difficulty parameter estimate for LP was relatively high (3<sup>rd</sup> of 11). However, the discrimination parameter estimate of LP was the lowest among all items, indicating that it is less effective at distinguishing individuals at different levels of liability. Several of the items' difficulty parameters cluster quite closely together (e.g., A4, D4, D5) but overall they cover a range of liabilities or vulnerability.

### Twin Modeling

In an attempt to maintain consistency with previous findings (Prescott et al. 1999), we first verified that the effects of common environmental factors could be set to 0. Dropping C did not produce a significant increase in model misfit (Table 4, Model 2). We next tested the significance of genetic effects specific to AUD: this test asks whether the genetic correlation between these phenotypes is no different from unity *versus* whether there are genetic factors that load onto AUD but do not load onto AA/AD. Setting the AUD-specific genetic effects (dashed lines in Figure 1, plus sex-specific genetic influences not depicted in the figure) to 0 significantly reduced model fit (Table 4, Model 3). Thus, the genetic correlation between DSM-IV and DSM-5 diagnoses is slightly, but significantly, less than unity (Table 5). Examination of the confidence intervals suggests that the decrease in fit is driven by the female portion of the sample, particularly the sex-specific genetic contributions to female phenotypic variance. Because the E term encompasses measurement error, in order to allow for the possibility of AUD-specific error we did not test whether the AUD-specific E path could be set to 0, though path estimates for AUD-specific E were quite low. Model 2, allowing for AUD-specific genetic influences, was selected as the best-fitting model. Final estimates of heritability and genetic and environmental correlations are presented in Table 5. We applied the same modeling approach to alternative phenotypic definitions of the diagnoses (e.g., modeling a DSM-IV dependence-only phenotype and modified AUD; coding AUD as 0, 1, 2 to capture the levels of severity that were initially proposed) and, while the overall fit of the model was worse than those reported here, the results did not change (available upon request).

### Discussion

The primary goal of these analyses was to investigate whether the proposed changes to alcohol-related diagnosis in the DSM-5 are supported from an empirical perspective. We approached this question in four ways: i) by comparing the clinical features of individuals in D4-only, D5-only, and D4D5; ii) conducting a confirmatory factor analysis to evaluate the factor loading of the LP criterion; iii) examining the parameters of the LP criterion from an item response analysis perspective; and iv) testing whether genetic influences underlying liability to a DSM-5 AUD diagnosis differ significantly from those underlying a DSM-IV

alcohol diagnosis (abuse or dependence). Items related to craving were not available for the current sample, making it impossible for us to construct a complete DSM-5 AUD diagnosis; however, as detailed below, this exclusion is unlikely to have introduced bias into our results. Results indicate that individuals in D4-only and D5-only are less severely affected than those in D4D5. The proposed changes will not have a pronounced effect on lifetime prevalence rates, and the exclusion of the LP item is unlikely to reduce the sensitivity or accuracy of a DSM-5 diagnosis. Results from the biometrical analysis suggest that there are modest, but significant, differences in the genetic liabilities to these diagnoses.

The lifetime prevalence of DSM-5 AUD is only slightly higher than that of DSM-IV abuse or dependence (32.1% vs. 30.7%, an increase of 5.1%). The modest increase in prevalence reported here is consistent with that reported by Agrawal et al. (2011), and less pronounced than that reported by Mewton and colleagues (2011). These differences could be partially attributable to the sample population (U.S. vs. Australian), or to differences in the assessment tools used in the different studies. We also note that those reports explored changes in 12-month prevalence of AUD, while the current report focuses on lifetime prevalence. Another possible contributor to the differences across studies is their treatment of items related to drunk driving. Agrawal and colleagues found that distinguishing between drunk driving and other types of hazardous use resulted in different prevalence changes between DSM-IV and DSM-5; conversely, drunk driving was not addressed in the interview used by Mewton and colleagues. Because of the structure of the interview used in the current study, distinctions between drunk driving and other types of hazardous use were not possible. Ultimately, our approach is more similar to that applied by Agrawal et al.; this could account for the comparable changes in prevalence reported here and in that study.

The prevalence reported here is likely an underestimate: some number of individuals who currently endorse only one criterion would certainly also endorse a craving criterion, thereby meeting the diagnostic threshold for AUD. However, a recent report on the impact of the changing diagnostic criteria in the NESARC found that only 4.1% of the individuals who endorsed only 1 of the DSM-IV AA or AD criteria (excluding legal problems) in the last 12 months would meet 12-month DSM-5 AUD criteria specifically because they also endorsed craving (Agrawal et al., 2011). Therefore, the underestimate of prevalence in the current report is likely very low. The change in diagnostic criteria will likely disproportionately affect women: 36.4% of the D5-only group (new cases) is female, compared to 24.3% of D4D5; in addition, the female-specific genetic correlation is lower than the sex-common genetic correlation. These sex differences are consistent with another recent report (Agrawal et al., 2011).

Members of D4-only were less likely to be diagnosed with other psychiatric (e.g., conduct disorder, major depression, etc.) and substance use disorders relative to those in D4D5, suggesting a less severe phenotype generally. However, the “loss” of these individuals is offset by the addition of D5-only, among whom other psychiatric/ substance use diagnoses are also less common relative to D4D5. Age of onset of AUD symptoms was older for D5-only than for D4-only and D4D5; both D4-only and D5-only reported lower “maximum drinks in 24 hours” than D5-only. For some disorders (e.g., opiate abuse/dependence diagnosis), prevalence is quite low, and these findings should be considered preliminary given the small sample sizes of D4-only and D5-only.

Direct comparisons between D4-only and D5-only, which are of greatest interest to evaluating the DSM-5 criteria, indicate that they are quite similar in terms of comorbidities: the only significant differences in clinical features between these groups was age of onset of AUD symptoms and the prevalence of cannabis use disorder. These differences suggest that D4-only has a slightly more severe phenotype. Thus, it is not clear that the proposed

diagnostic changes will result in a more accurate diagnosis: at best, one group of low-severity cases will be replaced by another; at worst, a group of individuals who exhibit more severe problems will be excluded from the DSM-5 diagnosis while less severely affected individuals will meet diagnostic criteria.

One benefit of the new guidelines is that individuals previously considered “diagnostic orphans” – those endorsing two dependence criteria but no abuse criteria – will be included in the new AUD diagnosis. Consistent with Agrawal et al. (2011), the most commonly endorsed dependence criteria among these new cases were “loss of control” and “trying to cut down”. In addition, those endorsing only a single abuse criterion (typically hazardous use), and who are presumed to have a low level of problems, will no longer be diagnosed with an alcohol problem. The exclusion of the LP criterion from the new guidelines does not substantially affect changes in prevalence (only N=62 individuals were given an AA diagnosis due only to their endorsement of that criterion). Though we were unable to evaluate the consequences of adding an item addressing craving, previous studies suggest this addition does not significantly inflate prevalence (e.g., Agrawal et al., 2011; Keyes et al., 2011). Given the modest change in prevalence under different DSM criteria, and particularly the quite minor effects of the “loss” of individuals who endorsed only the LP criterion, the consistency of the alcohol diagnoses from DSM-IV to DSM-5 is reasonable ( $\kappa=0.86$ ). However, this obscures potentially important qualitative differences among those meeting only DSM-IV or DSM-5 criteria.

Consistent with previous reports (Keyes et al., 2011; Mewton et al., 2011; Proudfoot et al., 2006), which examined 12-month rather than lifetime measures, the psychometric properties of the legal problems criterion were not robust: it had the lowest loading in the factor analysis, and the IRT analysis revealed that it has low discriminatory power but a relatively high difficulty parameter. However, this contrasts with the apparent utility of this criterion from the perspective of assessing familial risk to alcohol problems: another report (Kendler and Myers, in press) found that it was the most consistent predictor of familial risk.

Results from the biometric analyses indicate that total heritability is essentially constant across the diagnoses, and the genetic correlation between diagnoses approaches unity. Thus, the familial component of liability to alcohol problems is sufficiently captured by the new diagnosis and largely consistent with the previous criteria. However, we also found evidence of modest, but significant, diagnosis-specific genetic factors. There are significant differences in symptom endorsement rates between the diagnoses (Table 1), which results in a somewhat different profiles across the diagnoses; this could certainly correspond to a shift in relevant genetic influences. To ensure that the results were not an artifact of model structure, we reversed the order of the variables and re-ran the analysis; as expected since both variations are equivalent, the results were not qualitatively different. Hypotheses as to the nature of these diagnosis-specific genetic influences would be purely speculative based on the analyses reported here, though molecular genetic studies could be designed to address such questions.

In summary, these analyses provide equivocal support for the proposed changes to diagnostic criteria for DSM-5 AUD. The changes are unlikely to result in a dramatic difference in the prevalence of alcohol problems. The severity of alcohol problems in the subset of the sample with an unstable diagnostic status (i.e., those who would meet criteria for a DSM-5 diagnosis but were unaffected under DSM-IV guidelines, and vice versa) is low based on number of symptoms and patterns of comorbidity; thus, one “mild” phenotype is being exchanged for another. The factor and IRT analyses indicate that the LP criterion is not critically informative and its exclusion from the DSM-5 criteria is unlikely to be problematic. Regarding the biometrical comparison of the DSM-IV and DSM-5 diagnoses,



although there are significant AUD-specific genetic factors at play, the two diagnoses are very strongly genetically correlated. The new diagnostic criteria robustly index familial risk, fulfilling a critical validator of psychiatric illnesses (Robins and Guze, 1970). However, this is true of the DSM-IV criteria as well. Future analyses could be aimed at clarifying diagnosis-specific genetic influences, and could be informed by shifts in criteria endorsement (e.g., “new” cases frequently endorse “loss of control” and “trying to cut down”: might such symptoms be influenced by genes of a particular biological function?).

Taken together, do these findings suggest that the proposed DSM-V criteria represent an improvement over the DSM-IV criteria in terms of diagnostic validity? Results of the different assessments reported here do not clearly or consistently suggest that the new criteria are superior to the old with regards to identifying affected individuals. One advantage of the new criteria is their collapse into a single factor and the related inclusion of individuals who were formerly diagnostic orphans. However, these changes result in the exchange of one group of mildly affected individuals for another, with no indication that the newly included group is more severely affected or accurately diagnosed than the excluded group. The exclusion of the LP criterion could streamline clinical interviews, but there is no evidence that this criterion was resulting in *inappropriate* diagnoses. Finally, the genetic analyses suggest slight, but significant, differences in genetic influences underlying the two sets of criteria, though the total heritability is almost unchanged. There is no indication that this represents an improvement in the validity of the diagnosis. A study examining prognosis of the individuals in D4-only and D5-only could provide insight as to whether one group experiences a more negative outcome than the other, and could thus be considered more deserving of a diagnosis, but such data are not available for this sample. In light of these analyses, we would assert that the proposed changes do not represent an unequivocal improvement in diagnostic validity.

Beyond the impact of the proposed changes on prevalence and diagnostic validity, the effects will be felt by researchers working on datasets that lack items necessary to address craving, having been designed for previous iterations of the DSM. It is unclear whether the potential slight improvements in reliability and validity offset the probable costs to researchers whose data might be considered “out of date” (Kendler and Zachar, 2008). The ability of the research community to develop cumulative information about a diagnostic category is disrupted when criteria are changed. This might suggest a conservative approach to DSM revision, where change is made only in the presence of compelling data to support the superiority of the new definition off the older one that is to be replaced.

## Limitations

We note a number of limitations to the current report. First and foremost, we did not have data regarding craving and therefore could not construct a formal DSM-5 AUD diagnosis. As noted above, it is unlikely that this introduced substantial bias to our report of prevalence. However, it is possible that the inclusion of a craving item would change the results of the biometrical analyses; we hypothesize that the introduction of a new criterion (as opposed to the exclusion of the LP criterion, and slight changes in the endorsement frequencies for constant criteria) could lead to a lower genetic correlation between the DSM-IV and DSM-5 diagnoses. This possibility should be explored in genetically informative samples for which information on craving is available. We also note that the collinearity of the diagnostic phenotypes used in the twin modeling analyses is not ideal, and can result in unstable parameter estimates. In addition, for some of our analyses, we modeled the sexes jointly, though it is reasonable to expect that sex differences exist. For example, difficulty parameters for each symptom were higher for women than for men when examined separately. Importantly, qualitative differences between the sexes were limited; for example, while the difficulty parameter estimates differed, the ordering of the items based on

difficulty is nearly identical across sexes. Quantitative sex differences, including measurement invariance across sex, should be explored further, but were beyond the scope of the current analyses. Finally, this sample consisted exclusively of Caucasian individuals, and our results are not necessarily applicable to other ethnicities. These limitations do not diminish the relevance of our overall findings, which suggest that the proposed DSM-5 alcohol use disorder criteria do not represent a substantial improvement above the current criteria in terms of diagnostic validity.

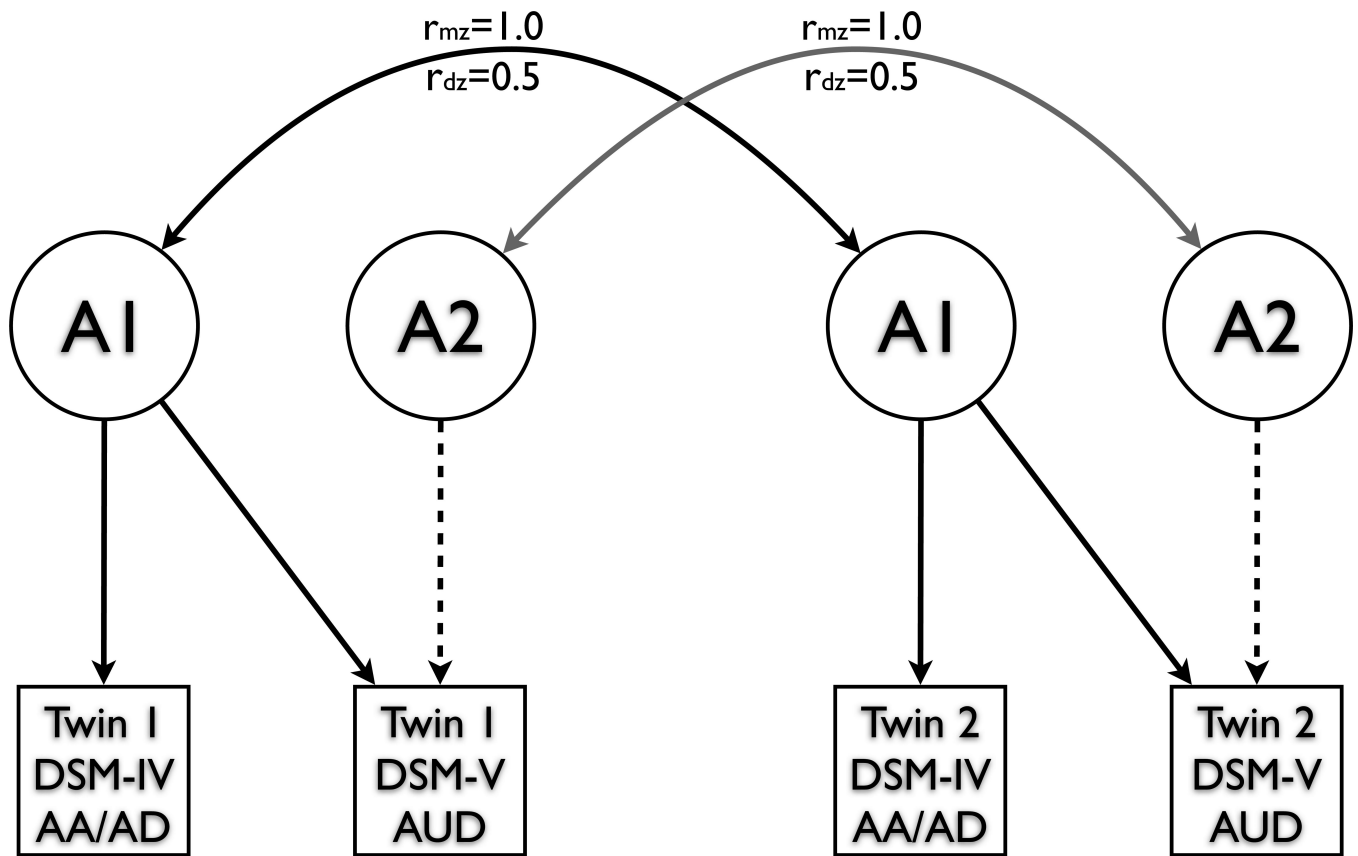
## Acknowledgments

These analyses were supported funding from the NIH (F32AA019849 [A.C.E.], R00DA023549 [N.A.G.], and P20AA017828 and R37AA011408 [K.S.K.]).

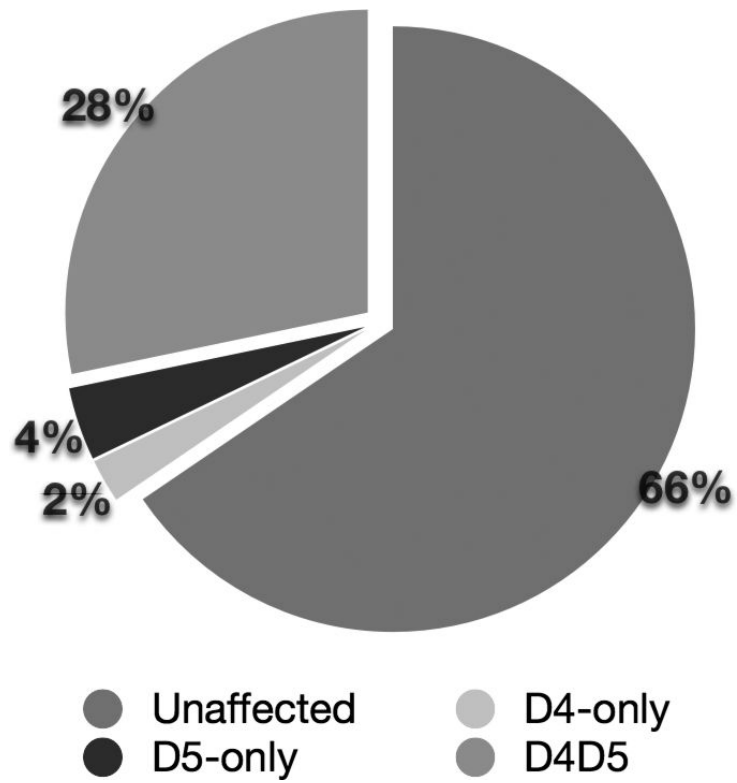
## References

- Agrawal A, Heath AC, Lynskey MT. DSM-IV to DSM-5: the impact of proposed revisions on diagnosis of alcohol use disorders. *Addiction*. 2011; 106:1935–1943. [PubMed: 21631621]
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)*. 4th ed.. American Psychiatric Publishing, Inc.; 2000.
- Boschloo L, van den Brink W, Penninx BW, Wall MM, Hasin DS. Alcohol-use disorder severity predicts first-incidence of depressive disorders. *Psychol Med*. 2011 Epub ahead of print.
- Casey M, Adamson G, Shevlin M, Mckinney A. The role of craving in AUDs: Dimensionality and Differential Functioning in the DSM-5. *Drug Alcohol Depend*. 2012 Epub ahead of print.
- Chen YC, Prescott CA, Walsh D, Patterson DG, Riley BP, Kendler KS, Kuo PH. Different phenotypic and genotypic presentations in alcohol dependence: age at onset matters. *J Stud Alcohol Drugs*. 2011; 72:752–762. [PubMed: 21906503]
- Feingold A, Rounsaville B. Construct validity of the abuse-dependence distinction as measured by DSM-IV criteria for different psychoactive substances. *Drug Alcohol Depend*. 1995; 39:99–109. [PubMed: 8529538]
- Gillespie NA, Neale MC, Prescott CA, Aggen SH, Kendler KS. Factor and item-response analysis DSM-IV criteria for abuse of and dependence on cannabis, cocaine, hallucinogens, sedatives, stimulants and opioids. *Addiction*. 2007; 102:920–30. [PubMed: 17523987]
- Hagman BT, Cohn AM. Toward DSM-V: mapping the alcohol use disorder continuum in college students. *Drug Alcohol Depend*. 2011; 118:202–8. [PubMed: 21514750]
- Hasin DS, Fenton MC, Beseler C, Park JY, Wall MM. Analyses related to the development of DSM-5 criteria for substance use related disorders: 2. Proposed DSM-5 criteria for alcohol, cannabis, cocaine and heroin disorders in 663 substance abuse patients. *Drug Alcohol Depend*. 2012; 122:28–37. [PubMed: 21963333]
- Kendler KS. Toward a scientific psychiatric nosology. Strengths and limitations. *Arch Gen Psychiatry*. 1990; 47:969–973. [PubMed: 2222134]
- Kendler KS, Myers J. Clinical indices of familial alcohol use disorder. *Alcohol Clin Exp Res*. 2012 In press.
- Kendler, KS.; Prescott, CA. *Genes, Environment, and Psychopathology*. 1 ed.. The Guilford Press; 2006.
- Kendler, KS.; Zachar, P. *The Incredible Insecurity of Psychiatric Nosology..* In: Kendler, KS.; Parnas, J., editors. *Philosophical Issues in Psychiatry*. The Johns Hopkins University Press; Baltimore, MD: 2008.
- Keyes KM, Krueger RF, Grant BF, Hasin DS. Alcohol craving and the dimensionality of alcohol disorders. *Psychol Med*. 2011; 41:629–640. [PubMed: 20459881]
- Krueger RF, Nichol PE, Hicks BM, Markon KE, Patrick CJ, Iacono WG, McGue M. Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychol Assess*. 2004; 16:107–119. [PubMed: 15222807]

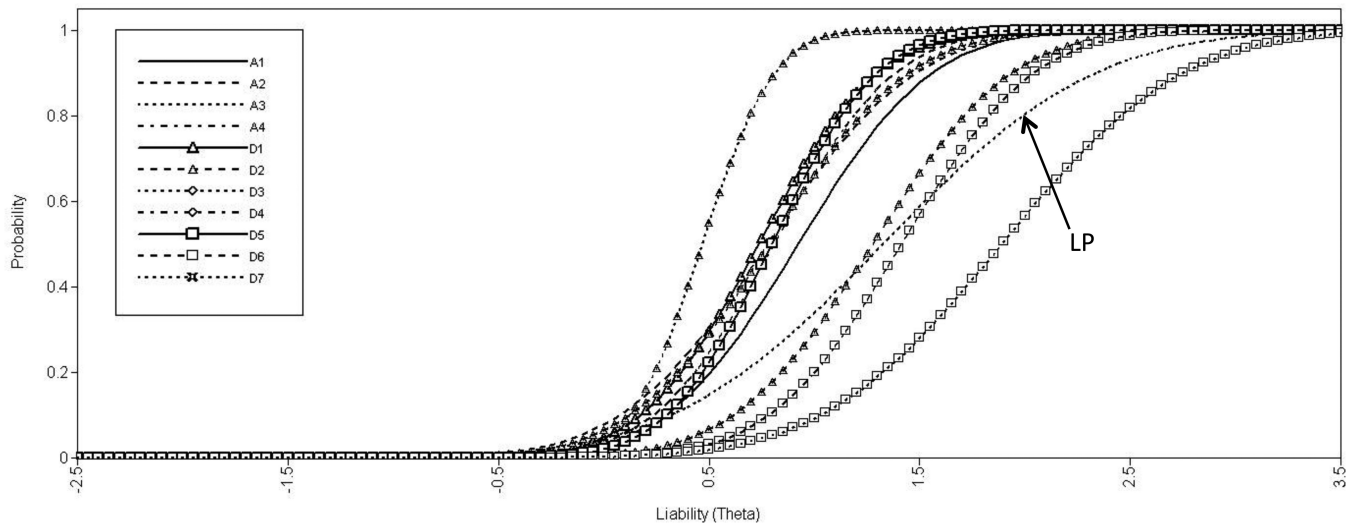
- Langenbucher JW, Labouvie E, Martin CS, Sanjuan PM, Bavly L, Kirisci L, Chung T. An application of item response theory analysis to alcohol, cannabis, and cocaine criteria in DSM-IV. *J Abnorm Psychol.* 2004; 113:72–80. [PubMed: 14992659]
- Mewton L, Slade T, McBride O, Grove R, Teesson M. An evaluation of the proposed DSM-5 alcohol use disorder criteria using Australian national data. 2011; 0965-2140 106:941–950.
- Muthen, LK.; Muthen, BO. *Mplus User's Guide*. Fifth Edition. Muthen & Muthen; Los Angeles, CA: 1998-2007.
- Neale, MC.; Boker, SM.; Xie, G.; Maes, HH. *Mx: Statistical Modeling*. 7th ed.. Dept. of Psychiatry, Virginia Commonwealth University; Richmond, VA: 2006.
- Prescott CA, Aggen SH, Kendler KS. Sex differences in the sources of genetic liability to alcohol abuse and dependence in a population-based sample of U.S. twins. *Alcohol Clin Exp Res.* 1999; 23:1136–1144. [PubMed: 10443978]
- Proudfoot H, Baillie AJ, Teesson M. The structure of alcohol dependence in the community. *Drug Alcohol Depend.* 2006; 81:21–26. [PubMed: 16005578]
- Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry.* 1970; 126:983–987. [PubMed: 5409569]
- Saha TD, Chou SP, Grant BF. Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychol Med.* 2006; 36:931–941. [PubMed: 16563205]
- Spitzer, RL.; Williams, JBW. *Structured Clinical Interview for DSM-III-R (SCID)*. Biometrics Research Department, New York State Psychiatric Institute; New York: 1985.
- Takane Y, de Leeuw J. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika.* 1987; 52:393–408.



**Figure 1.** Bivariate twin model. For simplicity, only genetic factors contributing to variance are shown; an identical structure exists for shared (C) and non-shared (E) environmental factors. Qualitative genetic sex effects are not included in this depiction. See methods for a description of the manifest variables. The significance of genetic influences specific to AUD can be assessed by testing whether the path estimates for the dashed lines can be set to 0. If this does not compromise model fit, then the genetic correlation between AA/AD and AUD is not significantly different from unity.  $r_{MZ}$ =genetic correlation of monozygotic twins;  $r_{DZ}$ =genetic correlation of dizygotic twins



**Figure 2.** Proportions of the total sample that are consistently unaffected or fall into D4-only, D5-only, and D4D5.



**Figure 3.**

Item characterization curves. Each of the 11 symptoms included in the DSMIV for AA and AD are included. LP (“A3” in the legend) is depicted by a dotted line without symbols. The position of the line across the x-axis corresponds to the item difficulty (1.526), which is relatively high; the slope of the line is related to its discriminatory power (1.101), which is lower than that of any other item. A1=failing obligations; A2=hazardous use; A3=legal problems; A4=use despite social problems; D1=tolerance; D2=withdrawal; D3=loss of control; D4=trying to cut down; D5=time spent; D6=activities given up; D7=use despite physical problems

Table 1

Symptom endorsement rates for the total sample, all those meeting DSM-IV diagnostic criteria, all those meeting DSM-V diagnostic criteria, those meeting *only* DSM-IV criteria, and those meeting *only* DSM-5 criteria.

Symptom	Description	Total sample	% Endorsing Symptom			
			DSM-IV cases <sup>1</sup>	DSM-5 cases <sup>2</sup>	D4-only	D5-only
A1	Failing obligations	15.0	47.9	45.2	12.1	4.5
A2	Hazardous use	18.7	59.3	54.2**	37.9	4.8
A3	Legal problems	10.7	32.4	28.7**	35.1	3.4
A4	Use despite social problems	18.0	57.1	53.6*	27.0	10.0
D1	Tolerance	19.5	58.7	60.2	5.2	38.8
D2	Withdrawal	8.9	28.1	27.7	0	7.6
D3	Loss of control	26.8	77.3	82.4**	8.0	78.0
D4	Trying to cut down	19.8	58.3	61.5*	0.6	50.5
D5	Time spent	17.7	55.1	55.1	0.6	22.3
D6	Activities given up	6.8	21.8	21.1	0	3.4
D7	Use despite physical problems	3.2	10.5	10.0	0	0.3

\* p 0.05

\*\* p<0.01 (chi square test, df=1); indicates significant difference in endorsement rate between DSM-IV and DSM-5 cases.

<sup>1</sup> Includes individuals in D4-only and D4D5.

<sup>2</sup> Includes individuals in D5-only and D4D5.

Table 2

Comparison of individuals with a DSM-IV diagnosis who would not meet DSM-5 criteria, and vice versa, with stably affected individuals.

Characteristic	%			Z statistic <sup>1</sup>		
	D4-only (N=174)	D5-only (N=291)	D4D5 (N=2107)	D4-only vs. D4D5	D5-only vs. D4D5	D4-only vs. D5-only
Male	74.1	63.6	75.7	-0.33	-4.07**	2.48*
Mean age on onset of symptoms	21.0 (SE=.39)	22.0 (SE=.31)	21.0 (SE=.12)	0.04	3.24**	-1.96*
Maximum drinks in 24 hours	13.3 (SE=.73)	11.8 (SE=.42)	16.7 (SE=.26)	-4.40**	-10.08**	0.52
Major depression	32.2	35.4	43.9	-2.95**	-2.71**	-0.91
Generalized anxiety disorder	19.5	22.7	31.4	-3.20**	-3.00**	-0.69
Conduct disorder	16.5	14.5	28.5	-3.22**	-4.64**	0.65
Antisocial personality disorder/adult antisocial behavior	10.1	8.6	26.0	-4.23**	-5.87**	1.00
Mean Fagerstrom Test for Nicotine Dependence Score <sup>2</sup>	4.3	4.2	4.9	-2.39*	-3.07**	0.35
Cocaine diagnosis <sup>3</sup>	4.6	2.4	14.1	-3.32**	-5.02**	1.25
Cannabis diagnosis	21.3	12.0	34.3	-3.34**	-7.16**	2.51*
Sedative diagnosis	2.9	2.1	7.8	-2.25**	-3.32**	0.61
Stimulant diagnosis	8.0	5.2	15.2	-2.32*	-4.25**	1.25
Opiate diagnosis	1.7	1.0	4.5	-1.65	-2.54*	0.63

<sup>1</sup> Coded such that a negative Z statistic indicates a lower prevalence in the first group listed

<sup>2</sup> Only applies to current or past regular smokers

<sup>3</sup> Drug-related diagnoses refer to abuse or dependence

\* p<0.05

\*\* p<0.01



**Table 3**

Standardized factor loadings and item response analysis parameter estimates.

Criterion	Description	Standardized Factor Loading	Item Difficulty	Item Discrimination
A1	Failing obligations	0.868	1.064	1.750
A2	Hazardous use	0.845	0.989	1.580
A3	Legal problems	0.740	1.526	1.101
A4	Use despite social problems	0.888	0.931	1.934
D1	Tolerance	0.893	0.848	1.980
D2	Withdrawal	0.860	1.472	1.684
D3	Loss of control	0.955	0.538	3.221
D4	Trying to cut down	0.859	0.906	1.680
D5	Time spent	0.912	0.920	2.228
D6	Activities given up	0.870	1.627	1.768
D7	Use despite physical problems	0.791	2.177	1.292

Table 4

Biometrical model fitting for AA/AD and AUD.

Model #	Description	Model Comparison	$\Delta \chi^2$	$\Delta df$	p-value	$\Delta AIC$
1	Full Model	n/a	(11835.23)	(14887)	n/a	(-17938.77)
2	Set C to 0	2 vs. 1	5.23	<sub>6</sub> <sup>1</sup>	0.51	-6.77
3	Set AUD-specific genetic effects to 0	3 vs. 2	18.09	<sub>3</sub> <sup>2</sup>	<0.01	12.09

<sup>1</sup>This includes path estimates for common environmental effects on both phenotypes, for both sexes.

<sup>2</sup>This includes path estimates for female-specific genetic effects as well as those common to both sexes.

**Table 5**

Estimates (95% confidence intervals) of genetic and environmental variance and correlations for DSM-IV AA/AD and modified DSM-5 AUD.

	Genetic Factors			Non-Shared Environmental Factors		
	AA/AD	AUD	$r_G$	AA/AD	AUD	$r_E$
Men	0.5566 (.5075-.5566) <sup>†</sup>	0.5770 (.5547-.6710)	0.9494 (.9407-.9494)	0.4460 (.4460-.4460)	0.4230 (.4230-.4230)	1.0000 (1.0000-1.0000)
Women	Sex-common 0.3876 (.3876-.3876)	0.4567 (.4567-.4567)	0.9994 (.9944-.9997)			
	Sex-specific 0.1066 (.0080-.1122)	0.0714 (.0449-.0714)	0.7814 (.7814-.7814)	0.5058 (.5058-.5058)	0.4719 (.4719-.4719)	0.9990 (.9982-.9990)

<sup>†</sup>In several cases, we were unable to obtain reliable confidence intervals for variance and correlation estimates.