



Decomposition of Mean Sex Differences in Alcohol Use Within a Genetic Factor Model

George B. Richardson¹ · Brian B. Boutwell^{2,3}

Received: 8 March 2019 / Accepted: 13 June 2020 / Published online: 17 June 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

A wealth of literature suggests that normative and heavy alcohol consumption continue to follow a historical pattern of greater prevalence among males as compared to females. Some prior research suggested that sex-specific factors might explain some of this gender gap. Generally speaking, though, more recent studies have indicated that the sources of differences for most complex traits, both genetic and environmental, are similar for males and females. To the best of our knowledge, however, no studies have tested whether genetic and environmental factors common to both sexes are *more often expressed* in males, on average, thereby accounting for some of the mean sex difference in alcohol use. The current study used nationally representative data from American twin respondents and a multiple group genetic factor model with a mean structure to address this gap in the literature. Results provide no evidence of sex differences in covariance structure and suggest that genetic and nonshared environmental influences common to both sexes largely explain why male alcohol use is more frequent and severe, on average, than is female use. In contrast, shared environmental influences seem to play a less important role. We discuss our findings in the context of the existing literature and chart out directions for future research.

Keywords Alcohol use · Sex limitation · Multiple group genetic factor model

Introduction

Key analyses of nationally representative data suggest that the consumption of alcohol in the United States has risen considerably across the last several decades. In 2002, for example, analysis of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) suggested that around 65% of the population consumed alcohol (Grant

et al. 2017). In the span of roughly a decade, that percentage rose, hovering around a national of prevalence of 72% in the NESARC III data (collected between 2012 and 2013; Grant et al. 2017). Connected to these findings, moreover, are results suggesting that the prevalence of consumption increased across various population groups, including socioeconomic status, gender, and race/ethnicity in that same time period. Importantly, these increases include not just normative alcohol ingestion, but also levels considered to be both high risk and clinically pathological (Grant et al. 2017). In 2002, roughly 9% of Americans could be considered high risk drinkers, and by 2012 that percentage had risen to 12%, with very similar patterns emerging for alcohol use disorder (AUD) (Grant et al. 2017).

Sex Differences in Alcohol Use

Despite the increase in alcohol use/abuse across various groupings in the United States population, current alcohol use, as well as alcohol abuse, continue to follow a historical pattern of greater prevalence among males as compared to females (Slutske 2005; Vink et al. 2012). In the NESARC, for instance, males surpassed females in

Edited by Tinca Polderman.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10519-020-10004-0>) contains supplementary material, which is available to authorized users.

✉ George B. Richardson
george.richardson@uc.edu

¹ School of Human Services, College of Education, Criminal Justice, and Human Services, University of Cincinnati, 460R Teachers-Dyer Complex, Cincinnati, OH, USA

² School of Applied Sciences, The University of Mississippi, Jackson, MS, USA

³ John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, MS, USA

their high risk drinking, both in 2002 (14%) and in 2012 (16%). The prevalence of females considered to be high-risk drinkers also increased across time (5% in 2002 to 9% in 2012), yet they never surpassed the level of male pathological drinking. Importantly, this “sex-differentiated” pattern was also observed for AUD. The overall prevalence (in the total population) of AUD rose from 8% in 2002 to 12% in the 2012–2013 NESARC, with males outpacing females in both datasets (12% versus 4% in 2002 and 16% versus 9% in 2012–2013) (Grant et al. 2017). The question lingering for some time now in both the clinical psychological literature as well as the public health literature relates to the potential underlying contributors to such differences.

To be sure, a variety of socio-cultural and historical factors aligned to play some role in the apparent changes across time in prevalence of consumption, as well as the gap that exists between male versus female use. At the same time in the general population, almost regardless of how you measure alcohol use (normative drinking to clinically dysfunctional levels of consumption), individual variation in the tendency to use and abuse alcohol is accounted for by both genetic and environmental influences (Polderman et al. 2015). Thus, genetically informative research designs that can better clarify sex differences in genetic and environmental sources of risk are needed. Fortunately, several studies have tested for sex limitation in genetic and environmental influence on consumption. To properly frame this emerging literature, we begin by briefly describing how researchers in behavioral genetics conceptualize and empirically distinguish among potential sources of sex differences in complex quantitative phenotypes like alcohol use.

Sources of Phenotypic Sex Differences

Two classes¹ of potential sources of sex differences in phenotypes like alcohol use are distinguished in behavioral genetics (see Neale and Cardon 1992; Viding et al. 2004; Neale et al. 2006). The first is qualitative sex differences in genetic and/or environmental influence (Neale and Cardon 1992). If qualitative differences exist, the sources of genetic and/or environmental variance in alcohol use are not the same between the sexes—certain factors influence use only in males or females. Qualitative differences can be due, for example, to sex-linked gene expression in sex chromosomes (i.e., the X and Y chromosomes) and/or autosomal chromosomes. Although autosomal genes are shared by the

sexes, factors like hormones and hormone fluctuation can affect their expression (Viding et al. 2004). Sex-specific gene expression may be initially detected as a significant sex difference in additive genetic variance and is investigated with greater power via designs that include opposite sex dizygotic (DZOS) pairs (Viding et al. 2004). To the extent that genetic and/or environmental influences are sex-specific, cross-twin correlations are smaller in DZOS pairs than same sex dizygotic (DZSS) pairs (Viding et al. 2004).

Currently, at least some prior research has suggested that sex-specific influences might explain some of the gender gap observed for drinking outcomes (Prescott et al. 1999). In their analysis of Virginia Twin Registry participants, for example, Prescott et al. (1999) reported evidence that while the genetic risk factors were largely similar for males and females, the overlap was not perfect and some genetic effects appeared to be sex-specific. However, the evidence on this last point is (at best) mixed, with additional research suggesting that increased exposure to environmental risk factors that differ across groups better explains the differences between males and females (Heath et al. 1997). More recently, Vink et al. (2012) found evidence of sex-specific influence for only 6 of 122 lifestyle, personality, psychiatric disorder, health, growth and developmental, and metabolic variables. These findings suggest the genetic architecture for most complex behavioral and psychological traits is similar in males and females. The authors did find evidence of sex-specific effects on alcohol use in adolescence, interestingly, but this finding did not extend to adulthood. Similar findings were reported by Verhulst et al. (2015), who conducted a meta-analysis of twin and adoption studies of genetic and environmental influence on AUD and found no evidence of qualitative sex limitation.

The second class of potential sources of sex differences is quantitative differences in genetic and/or environmental influence on a phenotype (Vink et al. 2012). The most commonly investigated quantitative differences exist when the same genetic or environmental factors are expressed in males and females, but the *magnitudes* of their effects vary between the sexes. A special case of this type of quantitative sex limitation is called scalar sex-limitation, which occurs when the total variance differs between the sexes but the proportions of variance explained by the ACE components in one sex are a scalar multiple of the variance components in the other. Put somewhat differently, the importance of A relative to C and E in males can be the same as in females, though A is greater in males. In this case heritability is the same in males and females, though the ACE variances are not.

Vink et al. (2012) addressed the possibility of quantitative sex limitation in addition to qualitative limitation. For adolescent alcohol use (“ever used alcohol” and “weekly alcohol”), the authors fit restricted models in which the ACE

¹ Selection bias is one additional potential source of apparent sex differences in frequency and severity of alcohol use. More males than females, for instance, might select to treatment and receive an AUD diagnosis. Here we do not consider this possibility further because we are attending to differences detected in nationally representative data.

components were constrained equal across the sexes. Given that they interpreted these more parsimonious models, it seems no evidence of inequality of the ACE component variances emerged during this period. The previously mentioned Verhulst et al. (2015) meta-analysis of twin and adoption studies provided a clearer picture of the extent of quantitative sex limitation. In that case, the authors failed to detect any quantitative heterogeneity across the sexes.

Another type of quantitative sex limitation occurs when sex-biased gene expression or environmental influence produces a difference only in the phenotypic mean. In this case, the genes contributing to a phenotype are expressed in both sexes, but more are expressed (penetrance is greater) on average in one sex. As described by Naqvi et al. (2019), “autosomal genes, operating identically in males and females to influence a trait, can be expressed more abundantly in one sex.” (p. 8) Or as described by Dolan et al. (1992), the difference in phenotypic means between the sexes may be due to a “difference in the location of the normal genetic and environmental distributions underlying the phenotypic individual differences.” (p. 319) This may occur in alcohol use, for example, if males are exposed to higher levels of hormones that affect gene expression, such that genetic influences common to both sexes are more penetrant in males (see Ngun et al. 2011). Effects of this sex-biased gene expression on alcohol use could be mediated by differences in reinforcement, stress-reactivity, or inhibitory self-control (Ngun et al. 2011; Sanchis-Segura and Becker 2016); or by differences in the function of organs involved in alcohol metabolism (e.g., the liver; Sanchis-Segura and Becker 2016).

The possibilities raised above can be addressed empirically via a multiple group genetic factor model that accommodates group differences in ACE component means and variances (Dolan et al. 1992; and for an extension to ordinal indicators, see Cho et al. 2009). Because Dolan and colleagues’ modeling approach relies on measurement invariance testing, here we briefly review how multiple group factor models are used to test instruments such as questionnaires for evidence of measurement invariance (e.g., see Millsap 2011; Wang et al. 2018). Measurement invariance testing is the statistical approach that enables researchers to determine whether a one-unit difference on a latent variable has the same metric within and between groups such as the sexes. As described by Wang et al. (2018), measurement invariance holds when at least three types of parameters in a latent variable model are not substantially different between groups.² The first and lowest level of invariance is configural

invariance, which holds when the number of latent variables and the pattern of loadings is the same across groups. Configural invariance allows researchers to discuss a construct meaningfully across groups but does not provide assurance that participants in different groups responded to the items in the same way. The next level of invariance testing is metric invariance, which holds when the factor loadings are equal across groups. When factor loadings are equal, this means that a one-unit difference on the latent variable in one group is equal to a one-unit difference in another. The third level of invariance is called scalar invariance or strong factorial invariance. Scalar invariance holds when both the item intercepts (or thresholds) and factor loadings are equal across groups. When scalar invariance holds, scores from different groups have the same unit of measurement as well as the same origin. Thus, differences in observed item means between groups can be interpreted as attributable to differences in the latent factor means. That is, a one-unit difference within groups on a latent variable has the same meaning as one-unit difference between groups (for more in depth discussion of measurement invariance, see Wang et al. 2018).

By adding a mean structure to the multiple group genetic factor model and extending invariance testing to phenotypic loadings on the genetic and environmental factors as well as their origins, Dolan et al. (1991, 1992) provided a method for determining whether phenotypic mean differences between males and females are attributable to genetic and environmental factors that subsume individual differences within each sex. This is possible if the loadings of alcohol use items on the additive genetic and environmental factors do not differ significantly by sex (i.e., metric invariance holds) and the origins (i.e., intercepts or thresholds) of the alcohol use item loadings on the genetic and environmental factors are also invariant by sex (i.e., scalar invariance holds). When scalar invariance holds, we can infer that our “yardstick” for measuring genetic risk for greater alcohol use, for example, is not different within and between males and females.

Below we illustrate configural, metric, and scalar invariance graphically to clarify how the average sex difference in alcohol use *could* be found attributable to genetic or environmental factors that are invariant by sex. Figure 1 shows that when only configural invariance holds, the genetic factor loading and alcohol use item intercept may each vary between the sexes. This would be consistent with qualitative sex limitation and quantitative sex limitation. When configural and metric invariance hold (but not scalar invariance), the intercepts may vary but the genetic factor loadings are the same. In this case, the genetic factor is on the same metric in both sexes, however, the mean sex difference in alcohol use cannot be attributed to greater average genetic risk (i.e.,

² Researchers can test still higher levels of invariance (e.g., invariance of factor variances and residuals) but typically only configural, metric, and scalar invariance are evaluated. This is partly because invariance at these levels ensures most inferences of interest are not compromised by bias, and partly because higher levels are often more difficult to achieve (for further discussion, see Wang et al. 2018).

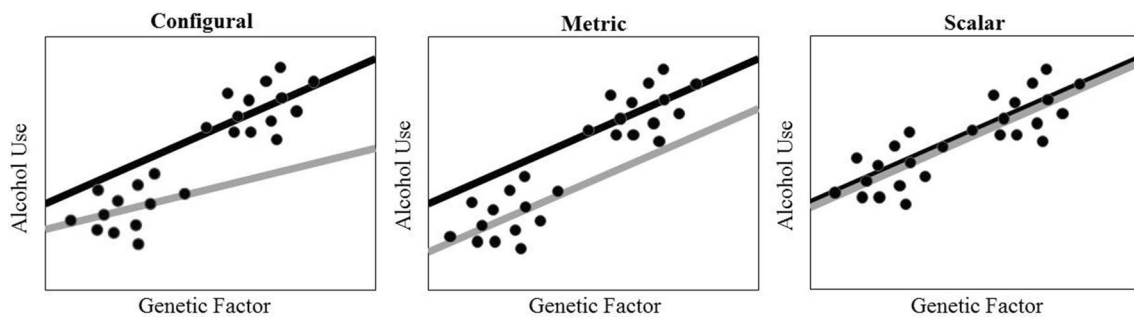


Fig. 1 Graphical depiction of three levels of invariance. Adapted from Wang et al. (2018). *Notes* Black lines are alcohol use loadings on the genetic factor for males and white lines are loadings for females. Female scores are clustered to the lower left and male scores to the upper right

to a higher mean on the genetic factor). This can be seen in the graphic, where increasing genetic risk in females (moving right along the gray line) is not sufficient to produce an alcohol use mean equivalent to males. Finally, when scalar invariance holds, the mean sex difference in alcohol use is indeed explained by the mean difference on the genetic factor. By moving toward greater genetic risk along the gray female loading, one crosses the male alcohol use mean.

To the best of our knowledge, this common cause explanation of alcohol use differences within and between the sexes has not yet been addressed empirically. For instance, Vink et al. (2012) seem to have found evidence that by adulthood the genetic and environmental influences on alcohol use are invariant by sex. However, the authors did not test whether the genetic and environmental factors common to both sexes explained the mean difference between them. This gap in the literature provides additional impetus for studies that can parse genetic from environmental effects using national data sources to further address this question, which remains a pressing and relevant public- and psychological-health concern.

Before delving into the current study, it is worth noting the gap highlighted above may extend beyond alcohol research. The addition of a mean structure to multiple group genetic factor models, to test whether sources of variation common to the sexes account for a mean difference between them, was not discussed in a recent review that covered potential sources of sex differences in antisocial behavior as well as research strategies for identifying them (Burt et al. 2019). In addition to shedding light on the origins of sex differences in alcohol use, the current study may aid investigators more broadly by serving as an additional case study in examining whether, and to what extent, genetic and environmental factors explain mean sex differences in various phenotypes that may also intersect with other public health outcomes.

The Current Study

The current study makes use of a representative dataset of American twin respondents in order to examine whether, and to what extent, genetic and environmental factors contribute to the gap between males and females for alcohol use. More specifically, we use multiple group genetic factor modeling (see Dolan et al. 1992; Cho et al. 2009) and data from the random sample of twins from the national survey of Midlife Development in the United States (MIDUS) to determine whether (a) there are sex-specific influences on alcohol use, (b) effects of genetic and environmental factors vary in magnitude by sex, and (c) average phenotypic sex differences in alcohol use can be attributed to mean differences in genetic and/or environmental risk.

Methods

Data

To further examine possible sources of sex difference in alcohol use, this study analyzed publicly available national data drawn from the MIDUS survey. Because our study involved analysis of a publicly available and de-identified dataset, the Institutional Review Board at the University of Cincinnati determined that it did not meet the regulatory criteria for research involving human subjects. The MIDUS survey collected data across a wide swath of individual-level factors, all possibly relevant for understanding a range of age-related health and wellbeing outcomes. Three rounds of longitudinal data are ultimately available to researchers (1995–1996, 2004–2006, and 2013). Important for the current analysis, an initial round of data were

collected from a national sample of twin pairs ascertained randomly via telephone based sampling ($n = 1914$ twins). Retention was moderately strong, with 78% of round 1 participants also taking part in round 2. For the current analysis, we used round 2 alcohol use data. The participants in the twin subsample of MIDUS were between the ages of 25 and 75 ($\bar{x} = 45$) and 45% were male. The racial/ethnic composition of the sample was 93.9% White, 4.2% African American, 0.6% Native American or Aleutian Islander/Eskimo, 0.3% Multiracial, and 0.9% Other. 87.7% graduated high school, 8.3% graduated from a two-year college or vocational school with an Associate's degree, 16.4% graduated college with Bachelor's degree, and 8.3% earned a graduate degree. For more information about the MIDUS samples, please see <https://www.midus.wisc.edu/midus1/index.php>.

Our initial sample consisted of $n = 1878$ twins. Following the lead of prior research in the area, we excluded 94 lifetime abstainers (i.e., those who answered “never had a drink” to the age had first drink of alcohol item) because initiation is strongly influenced by religious and cultural norms (Agrawal et al. 2012). Furthermore, in our analysis 110 cases were excluded by default because they were missing data on all the alcohol use items. Ultimately, our analytic sample consisted of $n = 1470$ twins (735 pairs). We used multiple group genetic factor modeling to analyze data from groups distinguished by zygosity and sex. The breakdown of our sample into these groups was: 136 MZ male (MZM), 103 DZ male (DZM), 145 MZ female (MZF), 154 DZ female (DZF), and 197 DZ opposite sex (DZOS) twin pairs. The dataset used in our analyses is posted in Open Science Framework: https://osf.io/2yrcq/?view_only=d26ba47508094871987e1bff347dc220

Instruments

Alcohol Use

We used eleven items in our attempt to comprehensively assess frequency and severity of past month, year, and lifetime alcohol use, including problem drinking. The past month items assessed (1) how often participants had at least one drink, (2) number of drinks they had on days when they drank, and (3) number of times they had five or more drinks on the same occasion. The lifetime use items assessed (4) how frequently participants had at least one drink during the period of life in which they drank most and (5) the number of drinks they typically consumed when drinking during this period. The past year items assessed (6) whether they had ever experienced emotional problems from drinking, (7) experienced desire/urge from drinking, (8) whether they had ever experienced

heavy drinking for one month or more, (9) whether they had to drink more to get the desired effects of the alcohol (tolerance), (10) the number of times they had consumed more alcohol than intended (loss of control), and (11) the number of times they had experienced effects of alcohol at work. Items 4–7 comprised the Michigan Alcohol Screening Test and were scored accordingly to form a binary indicator (0 = no alcohol problems and 1 = at least one alcohol problem) indicating whether participants screened positive an alcohol problem (see Selzer 1971). Item contents, response options, frequency distributions, and polychoric/tetrachoric correlations are presented in Supplementary Materials.

Analyses

To conduct our analyses, we employed the Cho et al. (2009) multiple group genetic factor model for the decomposition of phenotypic mean differences between groups when indicators are ordinal. Following Cho et al. (2009), we used the MPlus 8 software package to test our models, Robust Weighted Least Squares as the estimator (WLSMV; Muthén et al. 1997), and delta parameterization. All significance testing was conducted at the 0.05 alpha level.

Model specification

We specified a multiple group independent pathways model with a mean structure to achieve our research aims. This model is used when decomposing phenotypic means between groups because it is not possible to identify the ACE factor means in the nonreference group in the context of the common pathways models (Cho et al. 2009; Dolan et al. 1992). As a reminder, we employed multiple group genetic factor modeling because it allowed us to (a) test for sex-specific genetic and environmental influences via comparison of the DZSS and DZOS correlations, (b) test whether genetic and environmental factors were on the same metric within and between the sexes via measurement invariance testing, and (c) determine (assuming measurement invariance held to an acceptable degree) the relative importance of the ACE factors in accounting for mean phenotypic sex differences in alcohol use.

We identified males as the reference group and females as the nonreference group. Specifically, MZM, DZM, and DZOS male twins comprised the reference group; and MZF, DZF, and DZOS female twins comprised the nonreference group. We imposed the constraints described by Cho et al. (2009) to (minimally) identify the model.

Latent factor means and variances were set to zero and one, respectively, in the reference group; these parameters were then estimated in the nonreference group. Given that our observed variables had ordered categories, latent response variables were mapped onto observed variables with thresholds. Because we used delta parameterization, the variance of each observed variable was estimated via scale parameters, which were fixed to one in the reference group and estimated in the nonreference group.

Factor loadings³ were constrained equal across the reference and nonreference groups. The first two thresholds for the first three items were constrained equal across the groups. The rest of the thresholds were constrained equal within the reference (i.e., MZM, DZM, and DZOS male portions of the model) and nonreference (i.e., MZF, DZF, and DZOS female portions of the model) groups but were free to vary between them. As is typically the case in twin models, the model was symmetric and parameters were constrained equal between twins in a pair. To accommodate potential unique genetic and environmental influences on each alcohol use indicator, a between twin residual covariance for each indicator was estimated and allowed to vary across zygosity groups.

We incorporated information about zygosity into the MZM portion of the model, which was standardized, by constraining additive genetic and shared environmental covariances to one. For the DZM portion of the model, which was also standardized, we constrained the additive genetic and shared environmental covariances to 0.5 and one, respectively. We used nonlinear constraints to incorporate information about zygosity into the MZF, DZF, and DZOS portions of the model. This approach was necessary, as MPlus does not allow specification of correlations between non-standardized factors via WITH statements. Covariances between the additive genetic factors were constrained equal to the variance of A in the MZF group, to half of the variance of A in the DZF group, and to half of the square root of the variance of A in the female portion of the DZOS group. Covariances between shared environmental factors were constrained equal to the variance of C in the MZF and DZF groups and equal to the square root of the variance of C in the female portion of the DZOS group. These procedures yielded additive genetic correlations of 1, 0.5, and 0.5 in the MZF, DZF, and DZOS groups, respectively; and shared environmental correlations equal to one across these groups. Constraining the additive genetic and shared environmental

correlations to 0.5 and 1, respectively, in the DZOS group provided a test for qualitative sex limitation. MPlus syntax for specifying the minimally identified model is provided in Cho et al. (2009) and syntax for our final model is located in Supplementary Materials.

As a reminder, invariance testing allowed us to determine if the ACE factors had the same meaning across the sexes and whether sex differences in alcohol use indicators could be attributed to sex differences in genetic and environmental risk. Mean sex differences could be attributed to the ACE factors if there was no substantial evidence of misfit for the model including minimally identifying constraints (as described above). If this model was acceptable, we proceeded to test for higher levels of invariance.

Model Fit

We considered the substantive meaningfulness of the model and regarded Tucker-Lewis indices (TLI) greater than 0.95 (Hu and Bentler 1999; Byrne 2001), along with root mean square error of approximation values of less than 0.05 (RMSEA; Browne and Cudeck 1993), as evidence of acceptable fit to the data. We considered differences in TLI that exceeded 0.05 ($\Delta\text{TLI} > 0.05$) as strong evidence of differences in fit between models (Little 1997).

Results

Multiple Group Genetic Factor Model

We attempted to test our multiple group genetic factor model with a mean structure and received warning messages drawing our attention to several problems. First, there was a nearly perfect sample correlation (i.e., $r = 0.99$) between ‘number of times they had five or more drinks on the same occasion’ (past 30 days) and ‘number of times they had experienced effects of alcohol at work’ (past year), indicating these items were statistically redundant. Second, the alcohol screening test item had nearly perfect correlations with several variables. Finally, there were some inconsistent categories across the groups for four items—number of drinks they had on days when they drank (past 30 days), number of times they had five or more drinks on the same occasion (past 30 days), the number of drinks they typically consumed when drinking during this period (lifetime use), and the number of times they had consumed more alcohol than intended (past year). This occurred because the items were positively skewed and had sparse counts in their tails. We resolved the first problem by removing ‘number of times they had experienced effects of alcohol at work’ (past year) and the alcohol screening test item. We retained the ‘number of times they consumed more alcohol than intended’ item

³ Mplus fixes the first loading on each factor to one by default. We freed these loadings so they could be estimated and constrained them equal across the groups. Mplus constrains intercepts to zero by default when the WLSMV estimator is used and information about them is captured by thresholds. Observed variable residuals are not estimated under delta parameterization.

Table 1 Model fit information

Model	χ^2	<i>df</i>	<i>p</i>	TLI	RMSEA
Minimally identifying constraints	586.208	478	< 0.001	0.986	0.039 (0.027–0.050)
ACE variances and covariances fixed	571.593	463	< 0.001	0.985	0.040 (0.028–0.050)
Full scalar invariance	641.036	495	< 0.001	0.981	0.045 (0.034–0.054)
ACE variances equal	643.636	498	< 0.001	0.982	0.045 (0.034–0.054)
Scale parameters equal	647.182	504	< 0.001	0.982	0.044 (0.033–0.054)

df degrees of freedom, *TLI* Tucker-Lewis index, *RMSEA* root mean square error of approximation

because it had more categories than the other two items. We resolved the inconsistent categories across groups by bucketing responses to ‘number of drinks they had on days when they drank’ that exceeded four into a four or more drinks category, bucketing responses to ‘number of times they had five or more drinks on the same occasion’ that were equal to or exceeded three into a three or more drinks category, bucketing responses to ‘number of drinks they typically consumed when drinking during this period’ that were equal to or exceeded seven into a seven or more drinks category, and bucketing responses to ‘number of times they had consumed more alcohol than intended’ that exceeded three into a three or more times category.

After carrying out the procedures described above, we tested the hypothesized independent pathways model with minimally identifying constraints and fit to the data was excellent (e.g., $TLI = 0.986$; for complete fit information, see Table 1). Although the evidence of excellent fit suggested an absence of qualitative sex limitation, we wished to further investigate the possibility of qualitative sex limitation by examining modification indices associated with the genetic and shared environmental covariances in the DZOS group. Modification indices quantify the degree that fit will improve if specific restrictions on a model are removed. In this case, we were interested in identifying any strain introduced by the restrictions on the DZOS correlations. Because modification indices are not available in MPlus when nonlinear constraints are used, in a second step we removed these constraints and fixed the estimated ACE variances and covariances to the values observed in our initial minimally identified model. As expected, these constraints had virtually no effect on model fit ($\Delta TLI = -0.001$). We examined modification indices and none were relatively large (> 10) for the DZOS group, suggesting that indeed the DZOS correlations did not differ substantially from the DZSS correlations. Next, we returned to our minimally identified model with nonlinear constraints and tested for full scalar invariance by constraining all thresholds equal between the sexes. The fit of this model was virtually identical to the model with minimally identifying constraints (e.g., $\Delta TLI = -0.005$). As a reminder, scalar invariance suggests that mean sex differences on the alcohol use indicators can be attributed to mean differences between these groups in genetic and

environmental risk. Thus, we concluded that the same genetic and environmental sources of variance appeared to account for the variance within each sex as well the mean difference between them.

Given the evidence consistent with scalar invariance, we proceeded to test for higher levels of invariance by first constraining the factor variances equal across the sexes. We accomplished this by fixing the female ACE variances to 1 (as a reminder, the male variances were already fixed to 1); fixing the additive genetic covariances in the MZF, DZF, and DZOS groups to 1, 0.5, and 0.5, respectively; and fixing the shared environmental covariances in these groups to 1. Again, model fit was virtually identical to the minimally identified model (e.g., $\Delta TLI = 0.001$). We then constrained all scale parameters equal across the sexes to test them for invariance. These constraints did not degrade model fit relative to the minimally identified model (e.g., $\Delta TLI = 0.000$). Taken together, our findings provided no evidence of sex differences in covariance structure.

Following Cho et al. (2009), we next used R^2 estimates to evaluate the relative importance of additive genetic, shared environmental, and nonshared environmental factors in explaining the within sex variance in indicators of alcohol use. R^2 values were identical for males and females and could be computed as squared factor loadings because loadings were equivalent and latent response and factor variances were both fixed to one. We also calculated the phenotypic latent response mean differences ($\Delta \bar{x}_p$) across the indicators and these could be interpreted as Cohen’s d (Cohen 2013) because the model was standardized. Finally, we used the products of ACE mean differences and factor loadings ($\Delta x_{A-E} * \lambda$) to determine the mean difference of each latent response variable due to the ACE factors (i.e., the relative importance of ACE mean differences in generating the phenotypic latent response differences). Table 2 displays abbreviated item contents, means, factor loadings, R^2 estimates, and estimates of phenotypic mean differences due to ACE mean differences.

All genetic and nonshared environmental factor loadings were statistically significant. R^2 estimates indicated that genetic factors played the dominant role in explaining the within sex variance in the indicators of past 30 days and past year alcohol use (h^2 estimates ranged from 44 to

Table 2 Parameter estimates

Item	$\Delta \bar{x}_p$	Additive genetic			Shared environment			Nonshared environment		
		λ	R^2	$\Delta \bar{x}_A * \lambda$	λ	R^2	$\Delta \bar{x}_C * \lambda$	λ	R^2	$\Delta \bar{x}_E * \lambda$
<i>Past 30 days</i>										
1. Frequency had at least one drink	-0.287	0.919*	0.845	-0.198	-0.535*	0.286	0.155	0.308*	0.095	-0.244
2. # Of drinks on days drank	-0.356	0.840*	0.706	-0.181	0.272*	0.074	-0.079	0.122*	0.015	-0.097
3. # Of times they had 5+ drinks on the same occasion	-0.564	0.838*	0.702	-0.180	0.432*	0.187	-0.125	0.327*	0.107	-0.259
<i>Lifetime</i>										
4. How often at least one drink during period in which they drank most	-0.577	0.378*	0.143	-0.081	-0.143*	0.020	-0.041	0.679*	0.461	-0.537
5. # Of drinks when drinking during this period	-0.600	0.137*	0.019	-0.029	0.224*	0.050	-0.065	0.640*	0.410	-0.506
<i>Past year</i>										
6. # Of times more alcohol than intended	-0.445	0.662*	0.438	-0.142	0.110	0.012	-0.032	0.343*	0.118	-0.271
		95% CI	95% CI		95% CI	95% CI		95% CI	95% CI	95% CI
Factor mean		-0.215	-0.385	-0.044	-0.289	-0.468	-0.111	-0.791	-0.945	-0.637

* $p < .05$

$\Delta \bar{x}_p$ phenotypic latent response mean difference, λ factor loading, $\Delta \bar{x}_{A-E}$ ACE mean differences

All parameter estimates from scale parameter invariance model. Factor means are fixed to zero in the reference group (males) and estimated in the non-reference group (females) as differences from zero. Factor variances not displayed because they were fixed to one in this model

84%), whereas nonshared environmental factors played the dominant role in explaining the variance in the indicators of lifetime use (e^2 estimates ranged from 41 to 46%; h^2 estimates ranged from 2 to 14%). The shared environment loading for the indicator of past year alcohol use was not statistically significant and only two of six c^2 estimates exceeded 10% (c^2 ranged from 0.01 to 29%). The valences of the shared environmental effects on the indicators were inconsistent—whereas shared environmental factors were associated with greater numbers of drinks consumed when drinking in the past 30 days, greater numbers of times they had 5 + drinks on the same occasion in the past 30 days, and greater numbers of drinks typically consumed during the life period they drank most; they were associated with lower frequencies with which participants had at least one drink in the past 30 days as well as lower frequencies with which they had at least one drink during life period in which they drank most. Taken together, these results suggest shared environment played a relatively small and inconsistent role in explaining the variance in alcohol use.

Across the alcohol use indicators, mean phenotypic sex differences ranged from small to medium and the latent response means were uniformly lower in females than males. The ACE means were also significantly lower in females than males. The genetic and shared environment factor mean differences were small, whereas the nonshared factor mean difference was large. As a reminder, the relative importance of ACE mean differences in explaining the phenotypic latent response differences is determined via the products of the factor mean differences and the factor loadings ($\Delta \bar{x}_{A-E} * \lambda$). As shown in Table 2, genetic, shared environmental, and nonshared environmental factors were roughly similar in their importance as sources of the sex differences in two of the three past 30 days alcohol use indicators. Generally speaking, environmental factors were about half as important as genetic factors in explaining the observed sex difference in number of drinks on days when drinking. Shared environmental factors were of trivial importance in explaining sex differences in the lifetime and past year alcohol use indicators. Nonshared factors, alternatively, played the dominant role in explaining the sex differences in the lifetime use indicators. Both genetic and nonshared environmental factors explained substantial portions of the sex difference in the past year alcohol use indicator, with nonshared factors accounting for more of this difference. With the exception of number of drinks on days when drinking, nonshared factors accounted for (often slightly) larger portions of the phenotypic sex differences across the alcohol use indicators than did genetic and shared environmental factors.

Discussion

Psychiatric and clinical psychological research has demonstrated a consistent pattern in which males are more likely to experience alcohol over-use, and AUD, than are females (Slutske 2005; Vink et al. 2012). What seems less clear to this point, though, are the potential sources of this particular sex difference. As we mentioned earlier, prior research articulated two broad classes in which possible sources of sex differences for alcohol use/overuse would be expected to fall: (1) qualitative sex differences in genetic and/or environmental influence (i.e., sex-specific influences), and (2) quantitative differences in genetic and/or environmental influence on a particular phenotype. In the current study, we followed previous research in testing for qualitative and quantitative sex limitation in alcohol consumption. In addition, we tested for a type of quantitative sex limitation that to our knowledge had not been addressed in the alcohol literature; that the phenotypic mean difference between males and females might be attributable to genetic and environmental factors that subsume individual differences within each sex.

To examine the possibilities identified above, we used data from the MIDUS twin sample, and our results suggested several findings of note. First, and consistent with prior meta-analytic evidence of modest effects of shared environment among adults aged 18 to 65 (Polderman et al. 2015: $c^2_{\text{males}} = 17\%$, $c^2_{\text{females}} = 18\%$), the shared environment seemed to play a modest role in explaining past 30 days alcohol use variance within each sex (c^2 estimates ranged from 7 to 29%). Shared environment did not play an important role in explaining variance in the indicators of lifetime or past year alcohol use. These findings highlight the need to compare independent and common pathways models, because the roles of genetic and environmental factors may be heterogenous across alcohol use indicators.

Substantial heritability estimates emerged across most use indicators in the study (h^2 estimates for past 30 days and past year use ranged from 44 to 85%). The genetic effects on the lifetime use indicators, interestingly, produced h^2 estimates of only 2 and 5%. The past 30 days and past year heritability estimates observed in the current study are larger than recent meta-analytic findings for adults aged 18 to 65 (Polderman et al. 2015: $h^2_{\text{males}} = 39\%$; $h^2_{\text{females}} = 32\%$), while the lifetime estimates are smaller. The meta-analytic estimates of h^2 fall in between the former and the latter because, possibly, because most studies composited alcohol use indicators or used common pathway models, whereas the current study examined independent pathways. Nonshared environment played a smaller role in male and female alcohol use than genetic factors across all alcohol use indicators (e^2 estimates did not exceed 15%) except the two assessing lifetime use (e^2 estimates ranged from 41 to 46%).

Factor loadings and thresholds appeared invariant across the sexes. This finding of invariance, combined with a lack of evidence that the DZOS factor correlations were smaller than the DZSS correlations, suggests qualitative sex limitation may not be important in the etiology of alcohol use frequency and severity. These findings are consistent with prior research that tested for sex limitation in alcohol use (Vink et al. 2012; Verhulst et al. 2015). Scale parameters and additive genetic, shared environmental, and nonshared environmental variances were also not significantly different between the sexes. Taken together, our findings provide no evidence of sex differences in covariance structure and appear inconsistent with prior evidence that variances in males tend to exceed variances in females (e.g., Cross et al. 2011; Copping and Richardson 2019). However, they are consistent with some past research on sex limitation in alcohol use (Verhulst et al. 2015).

Finally, we attended to another type of potential quantitative sex limitation, which was that sex-biased gene expression or environmental influence might produce differences only in the phenotypic means. Consistent with this possibility, our findings suggest that genetic risk factors common to both sexes may be more abundantly expressed in males, thus *partly* explaining the increased male vulnerability to alcohol use in adulthood. Notably, we found that nonshared factors played a somewhat greater role, overall, than did genetic factors in explaining the sex differences in the past 30 days and past year alcohol use means. Nonshared factors, furthermore, played the dominant role in explaining the mean sex differences in the lifetime use indicators. Ultimately, then, our findings dovetail with and extend prior research on this topic in several key ways.

Recall that Vink et al. (2012) reported evidence that by adulthood, the genetic and environmental influences on alcohol use are invariant by sex. Relatively unknown at the time, however, was whether the genetic and environmental factors which were common to both sexes also explained the mean difference between them. Our analyses directly tested this possibility and did suggest that greater genetic risk in males explained substantial portions of the phenotypic mean sex differences across the past 30 days and past year alcohol use indicators. In other words, while the same allelic variants may contribute to an increased risk for frequent and severe alcohol use in both males and females, more of these variants might be expressed in males on average, thus pushing their average liability further upward. Moving forward, this suggests a new avenue for future research aimed at examining why genetic risk seems important in increased male vulnerability to problems related to alcohol use and abuse. Alternatively, it also raises the possibility that genetic influences might operate to insulate females from some of the risk of frequent and severe use, yet the reasons for these

putative protective factors remains unclear at the current juncture.

Equally important, our findings revealed that greater non-shared risk in males also explained substantial portions of the sex differences in past 30 days and past year use, as well as the vast majority of the phenotypic mean sex differences in the lifetime use indicators. This, coupled with the role of genetic factors, sets the stage for several important questions that need to be addressed. For instance, what aspects of the “unique” environments of males could be implicated in our findings? One possibility might entail differential exposure to socializing agents such as peer groups, encountering stressful life events (in which alcohol is used as a coping mechanism), or a range of other possible factors that will require more explicit empirical examination (see generally, Scholte et al. 2008). Returning briefly to the findings on genetic influences, these factors might involve differential aspects surrounding alcohol metabolism, liver functioning, functioning in the central and peripheral nervous system broadly, along with other possible avenues in which alcohol digestion is implicated (see generally, Thomasson 2002; Scholte et al. 2008).

It is also worth reiterating that shared environment was less important, overall, in accounting for the sex differences in alcohol use. However, shared environmental factors did appear to play a non-trivial role in explaining the sex differences in two of the past 30 days indicators. In particular, shared environmental factors seemed to decrease the difference in the average frequencies with which males and females had at least one drink, while apparently increasing the sex difference in number of times they had five or more drinks on the same occasion. Specific aspects of shared environments which might account for these opposing effects remains unknown at the current juncture. Although our results in some respects leave many important questions unresolved, they nonetheless serve to clarify which avenues of future research are likely to bear the most translational fruit on the issue of pathological alcohol use and abuse, and the differences in risk which seem to exist between males and females.

Limitations in the current study are certainly worth noting, as there were a couple of key points on this front. First, the alcohol consumption measures used are based on self-report data, which are subject to limitations of recall. Moreover, under-reporting of drinking could have impacted the parameter estimates observed. While anonymous and confidential surveys are the norm, and possess many desirable qualities, scholars should nevertheless cautiously interpret our results.

Second, some researchers have suggested that perhaps twins are non-representative of singletons. If true, concern regarding the external validity of our findings might be warranted despite our use of a random national sample.

Importantly, this issue is largely an empirical one, and Barnes and Boutwell (2013), analyzing a different data source, found little evidence of differences between twins and singletons in a large national sample of Americans. Moreover, Schwabe et al. (2017) conducted a comparison of twins and a sample of the entire Dutch population ($n=893,127$) and found that twin-based estimates were not an artifact of self-selection or due to differences between twins and singletons. More directly relevant to the current study, we conducted an analysis comparing the MIDUS random samples of twins and singletons on the alcohol use items examined in the current study (see Supplementary Materials). We also found little evidence that twins differed substantially from non-twin participants in the MIDUS sample.

Yet another important limitation is that while our findings do not suggest that shared environmental factors explained significant variance in the indicator of past year alcohol use, shared environmental effects are often relatively small (Polderman et al. 2015) and this effect might be detected in future studies with greater power. Relatedly, Vink et al. (2012) noted that although their findings suggested the genetic architecture of complex traits is similar across the sexes, genome wide association studies (GWAS) with much larger samples are still worth conducting given that individual differences in complex traits almost universally reflect the cumulative small effects of many genes (Chabris et al. 2015; Polderman et al. 2015). Thus, some genetic effects could be expressed only in one sex but go undetected given our sample size.

If many genes were sex limited, however, this would likely be detected as sex differences in the DZSS and DZOS correlations as well as the genetic effects. Ultimately, no reason emerged here to suggest that genetic factors should be on a different metric between the sexes, but more work is needed. Our current view is that the same genetic factors underlying alcohol use seem to be expressed in both sexes, although more may be disproportionately expressed in males, on average. The same appears true of nonshared sources of alcohol use. In short, results here suggest—for what seems to be the first time—that genetic and nonshared environmental influences which are invariant between the sexes largely explain why male alcohol use is more frequent and severe, on average, than is female use.

Compliance with Ethical Standards

Conflict of interest George B. Richardson and Brian B. Boutwell declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This study involved analysis of a publicly available and de-identified dataset, the Institutional Review Board at the University of Cincinnati determined that

it did not meet the regulatory criteria for research involving human subjects.

References

- Agrawal A, Freedman ND, Cheng Y-C et al (2012) Measuring alcohol consumption for genomic meta-analyses of alcohol intake: opportunities and challenges. *Am J Clin Nutr* 95:539–547
- Barnes J, Boutwell BB (2013) A demonstration of the generalizability of twin-based research on antisocial behavior. *Behav Genet* 43:120–131
- Browne MW, Cudeck R (1993) Alternative ways of assessing model fit. *Sage focus editions* 154:136–136
- Burt SA, Slawinski BL, Carsten EE et al (2019) How should we understand the absence of sex differences in the genetic and environmental origins of antisocial behavior? *Psychol Med* 49(10):1600–1607
- Byrne B (2001) *Multivariate applications book series. Structural equation modeling with AMOS: basic concepts, applications, and programming*. Lawrence Erlbaum Associates Publishers, Mahwah
- Chabris CF, Lee JJ, Cesarini D et al (2015) The fourth law of behavior genetics. *Curr Dir Psychol Sci* 24:304–312
- Cho SB, Wood PK, Heath AC (2009) Decomposing group differences of latent means of ordered categorical variables within a genetic factor model. *Behav Genet* 39:101–122
- Cohen J (2013) *Statistical power analysis for the behavioral sciences*. Academic Press, Cambridge
- Copping LT, Richardson GB (2019) Studying sex differences in psychosocial life history indicators. *Evol Psychol Sci* 6:47–59
- Cross CP, Copping LT, Campbell A (2011) Sex differences in impulsivity: a meta-analysis. *Psychol Bull* 137:97
- Dolan CV, Molenaar PC, Boomsma DI (1992) Decomposition of multivariate phenotypic means in multigroup genetic covariance structure analysis. *Behav Genet* 22:319–335
- Dolan CV, Molenaar PC, Boomsma DI (1991) Simultaneous genetic analysis of longitudinal means and covariance structure in the simplex model using twin data. *Behav Genet* 21:49–65
- Grant BF, Chou SP, Saha TD et al (2017) Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001–2002 to 2012–2013: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *JAMA Psychiatry* 74:911–923
- Heath AC, Bucholz K, Madden P et al (1997) Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychol Med* 27:1381–1396
- Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 6:1–55
- Little TD (1997) Mean and covariance structures (MACS) analyses of cross-cultural data: practical and theoretical issues. *Multivar Behav Res* 32:53–76
- Millsap E (2011) *Statistical methods for studying measurement invariance*. Taylor & Francis, Abingdon
- Muthén B, du Toit S, Spisic, D (1997) Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished technical report
- Naqvi S, Godfrey AK, Hughes JF et al (2019) Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* 365:eaaw7317
- Neale M, Cardon L (1992) *Methodology for genetic studies of twins and families*. Springer Science & Business Media, Berlin

- Neale MC, Røysamb E, Jacobson K (2006) Multivariate genetic analysis of sex limitation and G×E interaction. *Twin Res Hum Genet* 9:481–489
- Ngun TC, Ghahramani N, Sánchez FJ et al (2011) The genetics of sex differences in brain and behavior. *Front Neuroendocrinol* 32:227–246
- Polderman TJ, Benyamin B, De Leeuw CA et al (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 47:702
- Prescott CA, Aggen SH, Kendler KS (1999) Sex differences in the sources of genetic liability to alcohol abuse and dependence in a population-based sample of US twins. *Alcoholism* 23:1136–1144
- Sanchis-Segura C, Becker JB (2016) Why we should consider sex (and study sex differences) in addiction research. *Addict Biol* 21:995–1006
- Scholte RH, Poelen EA, Willemsen G et al (2008) Relative risks of adolescent and young adult alcohol use: the role of drinking fathers, mothers, siblings, and friends. *Addict Behav* 33:1–14
- Schwabe I, Janss L, Van Den Berg SM (2017) Can we validate the results of twin studies? A census-based study on the heritability of educational achievement. *Front Genet* 8:160
- Selzer ML (1971) The Michigan Alcoholism Screening Test: the quest for a new diagnostic instrument. *Am J Psychiatry* 127:1653–1658
- Slutske WS (2005) Alcohol use disorders among US college students and their non-college-attending peers. *Arch Gen Psychiatry* 62:321–327
- Thomasson HR (2002) Gender differences in alcohol metabolism. Recent developments in alcoholism. Springer, Berlin, pp 163–179
- Verhulst B, Neale MC, Kendler KS (2015) The heritability of alcohol use disorders: a meta-analysis of twin and adoption studies. *Psychol Med* 45:1061–1072
- Viding E, Spinath FM, Price TS et al (2004) Genetic and environmental influence on language impairment in 4-year-old same-sex and opposite-sex twins. *J Child Psychol Psychiatry* 45:315–325
- Vink JM, Bartels M, Van Beijsterveldt TC et al (2012) Sex differences in genetic architecture of complex phenotypes? *PLoS ONE* 7:e47371
- Wang S, Chen C-C, Dai C-L, Richardson GB (2018) A call for, and beginner's guide to, measurement invariance testing in evolutionary psychology. *Evol Psychol Sci* 4:166–178

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.