



Published in final edited form as:

*Alcohol Clin Exp Res.* 2011 September ; 35(9): 1716–1722. doi:10.1111/j.1530-0277.2011.01518.x.

## Alcohol Dependence in Men: Reliability and Heritability

Eivind Ystrom, Ted Reichborn-Kjennerud, Steven H. Aggen, and Kenneth S. Kendler

Department of Adult Mental Health (EY, TR-K), Norwegian Institute of Public Health, Oslo, Norway; Institute of Psychiatry (TR-K), University of Oslo, Oslo, Norway; Department of Epidemiology (TR-K), Columbia University, New York, New York; Virginia Institute for Psychiatric and Behavioral Genetics (SHA, KSK), and Department of Psychiatry (SHA, KSK), Virginia Commonwealth University, Richmond, Virginia

### Abstract

**Background**—The assessment of a Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV) life-time history of alcohol dependence (LTH-AD) has been found to be moderately reliable and substantially heritable. However, in studies of the heritability of LTH-AD, measurement error could not be discriminated from the true unique environmental effects. The aims of this study were to: (i) estimate the reliability of LTH-AD in a population based sample, (ii) identify characteristics of LTH-AD predicting a reliable diagnosis, (iii) investigate the heritability of LTH-AD as a function of diagnostic confidence, and (iv) to estimate the genetic and environmental influences on LTH-AD correcting for measurement error.

**Methods**—An unselected sample of 4,203 male twins was interviewed twice approximately 1-year apart assessing DSM-IV LTH-AD over the same period of life. Logistic regression was used to identify clinical features that predict a reliable diagnosis LTH-AD. Genetic and environmental influences on reliable LTH-AD were examined using structural equation models.

**Results**—Reliability of the diagnosis of LTH-AD was moderate ( $\kappa = 0.54$ ) and was predicted by the number of AD symptoms, treatment seeking, duration of most severe episode, and a great deal of time spent to obtain, use, or recover from alcohol use (DSM-IV AD criterion #5). Using an index of caseness, heritability of LTH-AD increased as a function of diagnostic confidence. Accounting for errors of measurement in a multivariate twin model, the heritability of LTH-AD increased from 55 to 71%.

**Conclusions**—Reliably diagnosed LTH-AD can be predicted by characteristics relevant to the disorder. LTH-AD appears to be a moderately reliable disorder of high heritability.

### Keywords

Alcohol Dependence; Diagnosis; Reliability; Twins; Heritability

---

In many epidemiologic studies, life-time history (LTH) of psychiatric or substance use disorders is assessed on 1 occasion. However, when people in a nonclinical population report on their past experiences of such symptoms or disorders on different occasions, they are often inconsistent in their responses (Culverhouse et al., 2005; Rice et al., 1986, 1992). A well-known consequence of unreliability is that associations are underestimated. In psychiatric genetic studies, the concordance between family members can, for example, be underestimated when reports from only a single interview are used. As nonconcordance

between twins is attributed to environmental influences unique to the individual, heritability estimates and the impact of environmental influences shared by family members are underestimated. Furthermore, in twin studies based on single assessments, measurement error cannot be separated from “true” environmental effects unique to each twin.

Compared with other common psychiatric disorders, the reliability of Diagnostic and Statistical Manual of Mental Disorders (DSM) alcohol dependence (AD) (American Psychiatric Association 1987, 1994) has been found to be relatively high (Chatterji et al., 1997; Demallie et al., 1995; Grant et al., 1995; Rice et al., 1986, 1992; Slutske et al., 1998; Wittchen, 1994; Young-Wolff et al., 2009). Although a high symptom count and treatment seeking predicts consistent reporting of AD in clinical samples (Culverhouse et al., 2005; Hasin et al., 1997), the predictors of unreliability for AD in the general population have received less attention. To our knowledge, no previous study has investigated a range of clinically relevant characteristics as predictors of reliability for DSM-IV LTH-AD (American Psychiatric Association 1994).

AD has been found to be heritable in a variety of different populations and genetic effects account for 40 to 60% of the variance in AD (Goldman et al., 2005; Heath et al., 1997; Kendler et al., 2003, 2007; Prescott and Kendler, 1999; Prescott et al., 1999; True et al., 1999). The impact of the effects of measurement error on these heritability estimates is, however, unknown.

We used a population-based sample of male–male twins assessed for LTH-AD on 2 different occasions to further clarify issues relating to reliability and heritability. First, we estimated the reliability of LTH-AD. Second, we investigated which clinical features of LTH-AD predicted a reliable diagnosis across the 2 occasions. Third, we examined the extent to which the heritability of LTH-AD is a function of diagnostic confidence. Finally, we estimated the heritability of LTH-AD with and without measurement error.

## MATERIALS AND METHODS

### Sample

The twins participating in this study came from the Mid Atlantic Twin Registry, and were interviewed as part of the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (Kendler and Prescott, 2006). Data were collected on male–male and male–female twin pairs (MM/MF) at 2 separate interview waves. To be eligible for participation, twins had to be Caucasian and born between 1940 and 1974. Of the 9,415 twins who were eligible at the first wave, 6,812 (72.4%) completed interviews primarily by telephone. Approximately 1 year later, the twins who had participated in the first wave were contacted for a follow-up interview. The median time gap between the 2 interviews was 15.3 months (mean = 19.0; SD = 8.7). Telephone and face-to-face interviews were complete for 5,621, or 82.5%, of those eligible. At the time of the first interview (1993 to 1996), the subjects had a mean age of 35.1 years (SD = 9.1; range = 18 to 56), and had in average 13.4 years (SD = 2.6) of education. (See Kendler and Prescott, 2006, for a more detailed description of the sample.)

A total of 5,599 twins had complete data on AD at both time points. From the first to the second wave, 10 subjects had new onsets of LTH-AD. As our aim was to use 2 measurement points to assess the disorder during the same life period, these cases were removed from further analyses (see below). The final sample for this study thus comprised 5,589 subjects, including 4,203 men and 1,386 women. The women in the sample were from opposite sex dizygotic (DZ) pairs and could therefore not be analyzed in twin modeling.

Only 4,203 men were therefore retained for further analyses. The twin pairs comprised 693 monozygotic (MZ) pairs, 477 DZ pairs, and 231 individuals without their co-twin.

Verbal informed consent was obtained from all participants prior to the first interview carried out by phone and written consent prior to the second face-to-face interview.

### Assessment Procedures

DSM-IV LTH-AD was assessed using the Structured Clinical Interview for DSM-IV (SCID) (American Psychiatric Association 1994; Spitzer and Williams, 1985), administered by clinically trained interviewers. Given that the aim of the study was to investigate the reliability of LTH of DSM-IV AD at time 1, this was the target of assessment at both time 1 and time 2.

Among randomly selected subjects re-interviewed 2 to 8 weeks after their initial interview, the test–retest reliability for AD was  $\kappa = 0.72$  (95% CI = 0.61, 0.82) ( $N = 382$ ) (Kendler and Prescott, 2006).

As described previously, zygosity was assigned by a combination of self-report measures, photographs, and DNA polymorphisms (Kendler et al., 2000).

### Statistical Analyses

Degree of agreement across assessment occasions was indexed by Cohen's Kappa (Landis and Koch, 1977) and the tetrachoric correlation coefficient.

To identify predictors of reliably diagnosed LTH-AD, we used logistic regression. Beginning with the 1,040 male subjects who met the DSM-IV criteria for LTH-AD at time 1, we examined the ability of clinical features of AD assessed at time 1 to predict consistent reporting of an LTH-AD at time 2. As there was a period of at least 1 year between the 2 assessments, any new cases that occurred during this period were excluded from the analysis. This was performed to ensure that the reporting of LTH-AD symptom criteria spanned the same lifetime period at both interviews. In the stepwise logistic regression analyses, using SPSS 17 (SPSS for Windows [Computer program], 2008), we applied a  $p$ -value of 0.05 for variables to be entered into analyses and 0.10 for them to be excluded. Due to dependence within twin pairs, we ran the solution found in the stepwise procedure again using Mplus 5.2 (Muthén and Muthén, 2007) using the “complex sample” feature to obtain correct standard errors.

Covariates that significantly predicted reliably diagnosed LTH-AD in the bivariate analyses were used to create an “index of caseness” of AD according to the procedure formulated by Rice and colleagues (1992). We divided each continuous variable by its maximum value, creating values from 0 to 1, and constructed the index of caseness by averaging all the covariates bivariately predicting reliably diagnosed LTH-AD. A person having the maximum score on this variable would have the maximum value on all significant predictors of reliably diagnosed AD.

We designated individuals with the highest values on this variable as “true cases,” assigning them the value 1. The rest of the cases in the sample would then, according to their covariates, have indices of caseness ranging between 0 and 1.

In the current study, a range of twin models were utilized to estimate the heritability of LTH-AD in men (Neale and Maes, 2004). In all models, individual differences in the common liability underlying the reporting of AD are assumed to have 3 different etiological sources: additive genetic (A) comprising all genetic effects that contribute additively, shared

environment (C) comprising all environmental exposures that contribute to similarity, and individual-specific or nonshared environmental factors (E) comprising all environmental factors contributing to differences between the twins, including measurement error. As MZ twins share all their genes, they are assumed to share all of their A factors. As DZ twins on average share half of their genes identical by descent, the expectation is that additive genetic factors will contribute about half as much to phenotypic resemblance compared with MZ twins. Both MZ and DZ twins are assumed to share none of their E factors and all of their C factors. The Mx statistical program, which uses a raw data maximum likelihood estimation approach, was used to fit the models to the data (Neale et al., 2006). Corresponding to the principle of parsimony, models with fewer parameters are preferable if they do not result in a significant deterioration of fit. A useful index of parsimony and fit is Akaike's information criterion (i.e.,  $\Delta\chi^2 - 2\Delta df$ ) (Akaike, 1987).

In the first series of twin analyses, we applied a standard univariate twin model where the variance in LTH-AD is accounted for by A, C, and E, to the index of caseness. In these models, the phenotype was modified as a function of the index of caseness. If an individual had an index of caseness of 0.30, he would be regarded as having LTH-AD if the cutoff (i.e., point at the continuous score where cases above the point is regarded as having AD) was below 0.20, and not having LTH-AD if the cutoff was set at 0.40. We ran 10 consecutive univariate twin models across 10 increasing levels on the index of caseness (equally sized groups with cut-points at: 0.13, 0.14, 0.15, 0.26, 0.27, 0.29, 0.40, 0.48, and 0.64).

The standard univariate twin model can be extended to include several phenotypes and several genetic and environmental factors. In the second analysis, we used a bivariate model (Cholesky) to examine the assumption that both assessments of the LTH-AD reflect the same underlying liability, that is, that the second assessment was not influenced by occasion-specific familial factors. This was performed to justify the use of the third type of models, the so-called measurement models (see Fig. 1). Here, the underlying assumption is that each assessment of LTH-AD serves as a separate assessment of the same phenotype. Thus, the covariance among individuals between the 2 assessments can be modeled as a latent common factor of reliable LTH-AD liability. Assessments at both time 1 and time 2 are assumed to be imperfect indicators of a true latent liability of AD (Fig. 1A). The  $\lambda_1$  and  $\lambda_2$  denote the linear regressions between the latent liability to AD and the assessments (i.e., the square root of the test-retest reliability). The other 2 paths influencing the LTH-AD,  $\kappa_1$  and  $\kappa_2$ , denote the effects of the residual components that include random error in the LTH-AD assessments. The latent phenotypic factor, that is, reliable liability to LTH-AD, is modeled as a single variable in a twin model. It is also influenced by additive genetic (path  $\lambda_a$ ), shared environmental (path  $\lambda_c$ ), and individual-specific environmental factors (path  $\lambda_e$ ).

There are 3 crucial differences between this model and a univariate twin model. First, it provides an estimate of the reliability of LTH-AD assessments. Second, it distinguishes error of measurement from true individual-specific environmental influences. Third, as the model explicitly partitions errors of measurement from total phenotypic variation, it provides estimates of additive genetic and common environmental effects correcting for the effects of measurement error.

## RESULTS

### Reliability of LTH-AD

The frequency of LTH-AD assessed at time 1 and time 2 was  $24.9 \pm 0.67$  and  $25.7 \pm 0.67\%$ , respectively, in twins that had data for both assessments (4,203 men). As can be seen in Table 1, 695 ( $67 \pm 1.5\%$ ) of 1,040 men who met criteria for an LTH-AD diagnosis at time 1, also reported meeting the criteria for the same time period at time 2. Among the 2,772 men

who did not meet criteria for LTH-AD at time 1, 383 ( $14 \pm 0.67\%$ ) men did when assessed again at time 2. The degree of agreement between the 2 assessments was moderate ( $\kappa = 0.54 \pm 0.015$ ; tetrachoric  $r = 0.77 \pm 0.014$ ).

### Reliably Diagnosed LTH-AD Predicted by Clinical Characteristics

In the first column of Table 1, the clinical characteristics of the subsample reporting LTH-AD at time 1 are presented. In the next column the characteristics of those who also reported LTH-AD at time 2 are given. For example, 67% of all subjects meeting an LTH-AD diagnosis at time 1 also report the same at time 2. However, among those whose most severe episode lasted more than 60 months, 80% were in agreement at both time points.

Table 2 shows results for the logistic regression analyses predicting reliably diagnosed LTH-AD from time 1 covariates. Reliably diagnosed LTH-AD was predicted in descending order of effect size by: (1) number of symptoms, (2) treatment, (3) impaired social or work activities (criterion #6), (4) withdrawal symptoms (criterion #2), (5) a great deal of time spent to obtain, use, or recover from alcohol use (criterion #5), (6) duration of most severe episode, (7) use despite physical consequences (criterion #7), (8) tolerance (criterion #1), and (9) being currently abstinent.

These analyses were repeated using stepwise logistic regression (Table 2; far right column), reducing the number of significant clinical features. Reliably diagnosed LTH-AD was then significantly predicted by number of symptoms, treatment, a great deal of time spent to obtain, use, or recover from alcohol use (criterion #5), and duration of most severe episode.

### Heritability of AD as a Function of Diagnostic Confidence

Using all significant predictors of diagnostic reliability of LTH-AD, we created an index of caseness operationalized as a continuous measure of diagnostic confidence for all individuals assessed at time 1. Using cutoffs at increasing levels on this score, we calculated the heritability of LTH-AD at different levels of diagnostic confidence. For the 10 models fit according to this index, the best-fitting model included only additive genetics and individual-specific environmental factors. As depicted in Fig. 2, heritability of AD increased as diagnostic confidence increased ( $\beta = 1.22$ ;  $R^2 = 0.80$ ;  $p < 0.01$ ). At 3 lowest levels of diagnostic confidence, heritability of LTH-AD was estimated to be  $<10\%$ , whereas at the highest level diagnostic confidence the heritability of LTH-AD was 82%.

### Heritability of LTH-AD Correcting for Error of Measurement

A measurement model was used to produce heritability estimates corrected for error of measurement (Fig. 1A). A basic assumption is that there can be no significant occasion-specific familial etiology for LTH-AD measured at time 2 if both assessments are to be conceived as independent measures of the same latent phenotype. We therefore estimated the unique additive genetic and shared environmental effects at time 2 LTH-AD (Model 1, Table 3), and found both to be 0, thus confirming our assumption (Model 2, Table 3). We then fitted a measurement model that specified error of measurement ( $\kappa_1$ ;  $\kappa_2$ ) separately (Fig. 1A). In the best-fitting model, the common environmental path ( $\lambda_c$ ) could be set to 0, and both factor loadings ( $\lambda_1$ ;  $\lambda_2$ ) and error terms ( $\kappa_1$ ;  $\kappa_2$ ) for the 2 time points could be set to equality (Model 5, Table 3).

When correcting for measurement error ( $\kappa_1$ ;  $\kappa_2$ ) at both time points, the heritability of LTH-AD was estimated to be 71% ( $\lambda_a^2$ ); whereas individual specific environment accounted for the remaining 29% ( $\lambda_e^2$ ) (Fig. 1B). When error of measurement at both time points was included in the individual specific environmental factor, the heritability of LTH-AD assessed at both time points was reduced to 55%. The square of the pathways between the

latent factor and the observed measurements ( $\lambda_1; \lambda_2$ ), estimated to be 0.77, gives an indication of the reliability of LTH-AD.

## DISCUSSION

In this study, we addressed the reliability of LTH-AD in a community and its implications for genetic analysis. We proposed 4 aims which we address in turn.

### Reliability of LTH-AD

We estimated the reliability of LTH-AD to be moderate ( $\kappa = 0.54$ ). Furthermore, the short-term reliability of AD in the current sample was estimated to be  $\kappa = 0.72$  (Kendler and Prescott, 2006). This is in accord with previous studies that have found AD to be among the more reliable common psychiatric disorders (Rice et al., 1986, 1992; Wittchen, 1994), with good short-term reliability ( $\kappa = 0.71$  to  $0.75$ ) (Chatterji et al., 1997; Grant et al., 1995; Young-Wolff et al., 2009) and moderate long-term reliability ( $\kappa = 0.56$  to  $0.61$ ) (Demallie et al., 1995; Slutske et al., 1998), and indicates that LTH-AD can be measured in population-based samples with a moderate reliability.

### Predicting Reliably of LTH-AD

Four features emerged as significant and unique predictors of reliability: number of symptoms; treatment seeking; a great deal of time spent to obtain, use, or recover from alcohol use; and duration of most severe episode.

Consistent with our findings, an earlier study using a clinical sample found that both a high symptom count and treatment seeking predicted AD diagnoses (Culverhouse et al., 2005). In general, reliability for psychiatric disorders is higher in clinical samples, explaining why reliability was higher among the subjects who had been in treatment for their AD. Higher reliability found in clinical samples may arise because more severe disorders are more easily recollected. Reliably diagnosed LTH of major depression has earlier been found to be predicted by disease severeness, treatment seeking (Foley et al., 1998; Kendler et al., 1993, 2001), and duration of episode (Kendler et al., 2001). Given that persons who report LTH-AD consistently on different occasions are more likely to be true cases, the covariates found are of importance to both clinical practice and clinical studies when ascertaining the diagnostic status of patients and subjects.

### Reliability and Heritability of LTH-AD

The heritability of LTH-AD increased with increasing values of diagnostic reliability as reflected in our index of caseness. This finding corroborates the notion that unreliable measures of LTH-AD will attenuate estimates of heritability in twin and family studies. Furthermore, creating a continuous index of caseness for a patient or clinical research subject, instead of solely using a dichotomous diagnosis, makes it possible to extract more diagnostically relevant information about each individual case.

### Heritability of LTH-AD and Accounting for Error of Measurement

We used a measurement model with 2 assessments to estimate the heritability of AD, while accounting for error of measurement, and also estimated the reliability of LTH-AD measured at each time point. Several noteworthy findings came out of fitting this model: First, as found in other studies examining AD in men, shared environmental influences were not statistically significant (Prescott and Kendler, 1999; Prescott et al., 1999; True et al., 1999). Second, when studying a single diagnostic assessment of LTH-AD in men, *c.* 23% of the variance in liability in fact reflects unreliability. Third, in the measurement model, forcing the factor loadings and error terms at times 1 and 2 to be equal did not significantly

degrade the model-data fit. This suggests that the reliability at the first assessment, performed about 1 year earlier, was approximately the same as at the second assessment. Fourth, this series of models, to the best of our knowledge, provides the first estimates of “true” individual-specific etiology of LTH-AD in men. Moreover, we found that 51% of what would in a standard twin model be attributed to individual-specific environment was in fact error of measurement, whereas 49% was true individual-specific environment. Finally, when accounting for unreliability in the twin model, LTH-AD in men was found to have a heritability point estimate of 71%. This revised estimate for the heritability of AD puts it in the range of the heritability reported for more severe psychopathology such as bipolar disorders (Smoller and Finn, 2003) and schizophrenia (Sullivan et al., 2003).

### Limitations

The results from the current study should be interpreted in the light of 3 limitations. First, the analyses are limited to men, and therefore cannot be extended to women. Second, the sample comprises only native Virginians of Caucasian origin. The results should therefore be generalized with caution to other male populations. Last, the unstable influences on the occasion-specific measures of LTH-AD and error of measurement, are not necessarily unimportant factors (Carey and Gottesman, 1978). They could comprise aspects valid to a general clinical evaluation, like failing memory, poor communication with the interviewer, or distracting events that happened the day of the interview. Therefore, the factors conceptualized as error of measurement in this study could be of importance in other studies.

### Implications

In a setting where only 1 clinical interview is available, the clinical characteristics predicting a reliable diagnosis can be used to create a continuous index of caseness. This approach to increase reliability could be relevant to any study of population-based cases of AD, including molecular analyses.

### Acknowledgments

This work was supported in part by grants AA-R37-011408 and AA-P20-017828 from the U.S. National Institutes of Health.

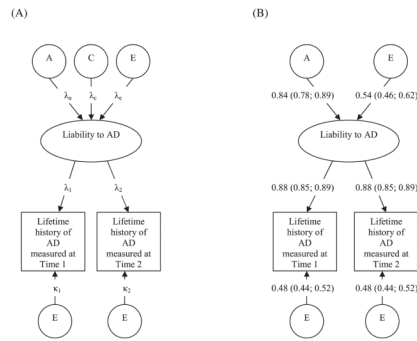
### References

- Akaike H. Factor-analysis and Aic. *Psychometrika*. 1987; 52:317–332.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)*, Revised. 3. APA; Washington, DC: 1987.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4. APA; Washington, DC: 1994.
- Carey G, Gottesman II. Reliability and validity in binary ratings: areas of common misunderstanding in diagnosis and symptom ratings. *Arch Gen Psychiatry*. 1978; 35:1454–1459. [PubMed: 727882]
- Chatterji S, Saunders JB, Vrašti R, Grant BF, Hasin D, Mager D. Reliability of the alcohol and drug modules of the Alcohol Use Disorder and Associated Disabilities Interview Schedule—Alcohol/Drug-Revised (AUDADIS-ADR): an international comparison. *Drug Alcohol Depend*. 1997; 47:171–185. [PubMed: 9306043]
- Culverhouse R, Bucholz KK, Crowe RR, Hesselbrock V, Nurnberger JI, Porjesz B, Schuckit MA, Reich T, Bierut LJ. Long-term stability of alcohol and other substance dependence diagnoses and habitual smoking—an evaluation after 5 years. *Arch Gen Psychiatry*. 2005; 62:753–760. [PubMed: 15997016]
- Demallie DA, Cottler LB, Compton WM III. Alcohol abuse and dependence: consistency in reporting of symptoms over ten years. *Addiction*. 1995; 90:615–625. [PubMed: 7795498]

- Foley DL, Neale MC, Kendler KS. Reliability of a lifetime history of major depression: implications for heritability and co-morbidity. *Psychol Med.* 1998; 28:857–870. [PubMed: 9723141]
- Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. *Nat Rev Genet.* 2005; 6:521–532. [PubMed: 15995696]
- Grant BF, Harford TC, Dawson DA, Chou PS, Pickering RP. The Alcohol-Use Disorder and Associated Disabilities Interview Schedule (AUDADIS)—reliability of alcohol and drug modules in a general-population sample. *Drug Alcohol Depend.* 1995; 39:37–44. [PubMed: 7587973]
- Hasin D, Carpenter KM, McCloud S, Smith M, Grant BF. The alcohol use disorder and associated disabilities interview schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample. *Drug Alcohol Depend.* 1997; 44:133–141. [PubMed: 9088785]
- Heath AC, Bucholz KK, Madden PAF, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DJ, Dunne MP, Whitfield JB, Martin NG. Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. *Psychol Med.* 1997; 27:1381–1396. [PubMed: 9403910]
- Kendler KS, Gardner CO, Prescott CA. Are there sex differences in the reliability of a lifetime history of major depression and its predictors? *Psychol Med.* 2001; 31:617–625. [PubMed: 11352364]
- Kendler KS, Karkowski LM, Neale MC, Prescott CA. Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. *Arch Gen Psychiatry.* 2000; 57:260–261.
- Kendler KS, Myers J, Prescott CA. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Arch Gen Psychiatry.* 2007; 64:1313–1320. [PubMed: 17984400]
- Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. The lifetime history of major depression in women—reliability of diagnosis and heritability. *Arch Gen Psychiatry.* 1993; 50:863–870. [PubMed: 8215812]
- Kendler, KS.; Prescott, CA. *Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders.* Guilford Press; New York, NY: 2006.
- Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry.* 2003; 60:929–937. [PubMed: 12963675]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
- Muthén, LK.; Muthén, BO. *Mplus User's Guide.* 5. Muthén & Muthén; Los Angeles, CA: 2007.
- Neale, MC.; Boxer, SM.; Xie, G.; Maes, HH. *Mx: Statistical Modeling.* 7. Department of Psychiatry, Virginia Commonwealth University; Richmond, VA: 2006.
- Neale, MC.; Maes, HHM. *Methodology for Genetic Studies of Twins and Families.* Kluwer Academic Publishers B.V; Dordrecht: 2004.
- 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 US twins. *Alcohol Clin Exp Res.* 1999; 23:1136–1144. [PubMed: 10443978]
- Prescott CA, Kendler KS. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *Am J Psychiatry.* 1999; 156:34–40. [PubMed: 9892295]
- Rice JP, McDonald-Scott P, Endicott J, Coryell W, Grove WM, Keller MB, Altis D. The stability of diagnosis with an application to bipolar II disorder. *Psychiatry Res.* 1986; 19:285–296. [PubMed: 3809326]
- Rice JP, Rochberg N, Endicott J, Lavori PW, Miller C. Stability of psychiatric diagnoses—an application to the affective disorders. *Arch Gen Psychiatry.* 1992; 49:824–830. [PubMed: 1417436]
- Slutske WS, True WR, Scherrer JF, Goldberg J, Bucholz KK, Heath AC, Henderson WG, Eisen SA, Lyons MJ, Tsuang MT. Long-term reliability and validity of alcoholism diagnoses and symptoms in a large national telephone interview survey. *Alcohol Clin Exp Res.* 1998; 22:553–558. [PubMed: 9622431]

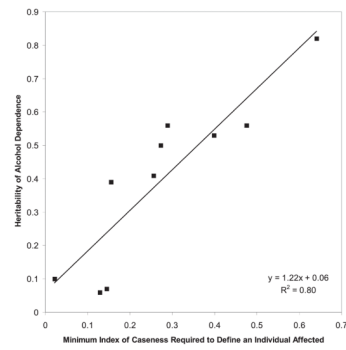


- Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet.* 2003; 123C:48–58. [PubMed: 14601036]
- Spitzer, RL.; Williams, JBW. Structured Clinical Interview for DSM-III-R (SCID). Biometrics Department, New York State Psychiatric Institute; New York, NY: 1985.
- SPSS for Windows [Computer program]. SPSS Inc; Chicago, IL: 2008. Rel. 17.0.0
- Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry.* 2003; 60:1187–1192. [PubMed: 14662550]
- True WR, Xian H, Scherrer JF, Madden PAF, Bucholz KK, Heath AC, Eisen SA, Lyons MJ, Goldberg J, Tsuang M. Common genetic vulnerability for nicotine and alcohol dependence in men. *Arch Gen Psychiatry.* 1999; 56:655–661. [PubMed: 10401514]
- Wittchen HU. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatr Res.* 1994; 28:57–84. [PubMed: 8064641]
- Young-Wolff KC, Kendler KS, Sintov ND, Prescott CA. Mood-related drinking motives mediate the familial association between major depression and alcohol dependence. *Alcohol Clin Exp Res.* 2009; 33:1476–1486. [PubMed: 19426164]

**Fig. 1.**

(A) A twin model for the heritability of liability to a life-time history (LTH) of alcohol dependence (AD) including occasion-specific influences on the recall and diagnosis of an LTH of AD (i.e., measurement error). In this model, it is assumed that there is a true, latent, liability to an LTH of AD, which is indexed by 2 assessments covering the same life period, measured at time 1 and time 2. The paths  $\lambda_1$  and  $\lambda_2$  illustrate the degree to which these assessments reflect the true liability to LTH of AD and is the square root of the test-retest reliability. The other path influencing LTH of AD,  $\kappa_1$  and  $\kappa_2$ , illustrate occasion-specific influences on each assessment of LTH of AD. The model is constrained such that

$\lambda_1^2 + \kappa_1^2 = 1.0$ . The liability to LTH-AD is modeled in a standard twin design with the sources of variance divided between additive (A), common (C), and individual-specific (E) environmental factors. By definition, the “common” environmental components are perfectly correlated in all twins, whereas the “individual-specific” environment is uncorrelated. Additive genetic factors are perfectly correlated in monozygotic twins and correlated 0.50 in dizygotic twins. The paths  $\lambda_a$ ,  $\lambda_c$ , and  $\lambda_e$ , denote the paths from these factors. The paths illustrate standardized regression coefficients, so that the proportion of variance in the dependent variables accounted for by the independent variables is equal to the square of the connecting path. For example, heritability of reliably diagnosed LTH-AD equals  $\lambda_a^2$ . (B) Parameter estimates with 95% confidence intervals from the best-fitting model (Model 4, Table 3). No evidence was found for different reliability in the diagnoses of LTH-AD at time 1 and time 2. No significant effect of common environmental influences on liability to LTH-AD was found. Heritability of reliable liability to LTH-AD =  $0.84^2 = 71\%$ . Conversely, the individual-specific environmentality of reliable liability to LTH-AD =  $0.54^2 = 29\%$ .



**Fig. 2.** Heritability of life-time history of alcohol dependence as a function of diagnostic confidence.  $p < 0.01$ . The regression line depicts this relationship. In all the models, only additive genetic and individual-specific factors were needed.

**Table 1**

Predictors of reliably diagnosed Life-Time History (LTH) of Alcohol Dependence (AD) in Males

	LTH-AD at time 1 (N)	LTH-AD at time 1 and time 2 (N [% ± SE])
AD	1,040	695 (67 ± 1.5)
Age of onset		
<18 years	228 (22 ± 2.7)	160 (70 ± 3.0)
18 to 20 years	399 (38 ± 2.4)	253 (63 ± 2.4)
21 to 24 years	222 (21 ± 2.8)	143 (64 ± 3.2)
>24 years	191 (18 ± 2.8)	139 (73 ± 3.2)
Duration of most severe episode		
<12 months	277 (27 ± 2.7)	165 (60 ± 2.9)
12 to 24 months	219 (21 ± 2.8)	136 (62 ± 3.3)
24 to 60 months	292 (28 ± 2.6)	193 (66 ± 2.8)
>60 months	252 (24 ± 2.7)	201 (80 ± 2.5)
Received treatment	307 (30 ± 2.6)	260 (85 ± 2.1)
Currently abstinent	219 (21 ± 2.8)	165 (75 ± 2.9)
No. of endorsed criteria		
3	445 (43 ± 2.3)	242 (45 ± 2.4)
4	272 (26 ± 2.7)	183 (67 ± 2.8)
5	155 (15 ± 2.9)	116 (75 ± 3.5)
6	104 (10 ± 2.9)	92 (88 ± 3.1)
7	64 (6 ± 3.0)	62 (97 ± 2.2)
Specific criteria		
#1 Tolerance	682 (66 ± 1.8)	588 (86 ± 1.3)
#2 Withdrawal symptoms	224 (22 ± 2.7)	219 (98 ± 1.0)
#3 Drinking more than intended	762 (73 ± 1.6)	622 (82 ± 1.4)
#4 Unsuccessful control of use	672 (65 ± 1.8)	581 (86 ± 1.3)
#5 A great deal of time spent to obtain, use, or recover from alcohol use	647 (62 ± 1.9)	599 (93 ± 1.0)
#6 Impaired social or work activities	320 (31 ± 2.6)	306 (96 ± 1.1)
#7 Use despite physical consequences	121 (12 ± 2.9)	120 (99 ± 0.8)

**Table 2**  
 Prediction of Reliably Diagnosed Life-Time History of Alcohol Dependence by Clinical Covariates in Males

	Covariates assessed one at a time				Covariates assessed by stepwise procedure			
	$\beta^a$	$\chi^2$	p	OR (95% CI)	$\beta^a$	$\chi^2$	p	OR (95% CI)
No. of symptoms	0.60	82.97	0.00	1.82 (1.59 to 2.08)	0.42	34.04	0.00	1.52 (1.31 to 1.76)
Treatment	1.33	62.70	0.00	3.79 (2.69 to 5.34)	0.85	21.96	0.00	2.34 (1.62 to 3.39)
Duration of most severe (years)	0.10	26.56	0.00	1.11 (1.06 to 1.15)	0.05	6.89	0.01	1.06 (1.02 to 1.10)
Age of onset (years)	0.00	0.08	0.77	1.00 (0.98 to 1.03)				
Currently abstinent	0.52	9.07	0.00	1.68 (1.20 to 2.35)				
#1 Tolerance	0.53	9.70	0.00	1.69 (1.21 to 2.36)				
#2 Withdrawal symptoms	0.88	29.58	0.00	2.42 (1.75 to 3.34)				
#3 Drinking more than intended	-0.01	0.00	0.96	0.99 (0.61 to 1.61)				
#4 Unsuccessful control of use	0.15	1.37	0.24	1.17 (0.90 to 1.51)				
#5 A great deal of time spent to obtain, use, or recover from alcohol use	0.96	29.33	0.00	2.61 (1.83 to 3.72)	0.53	7.58	0.01	1.70 (1.16 to 2.47)
#6 Impaired social or work activities	0.94	44.83	0.00	2.55 (1.93 to 3.37)				
#7 Use despite physical consequences	0.61	16.49	0.00	1.84 (1.37 to 2.48)				

<sup>a</sup>Logistic regression coefficient.

**Table 3**

Results of Bivariate Model Fitting for Life-Time History of Alcohol Dependence

Model	$\Delta df$	$\Delta\chi^2$	$p$	$\Delta AIC$
1. Bivariate ACE (Cholesky)	11	11.02	0.44	-10.98
2. Bivariate ACE; A and C specific to time 2 constrained to 0	13	11.02	0.61	-14.98
3. Measurement model ACE	16	11.31	0.79	-20.69
4. Measurement model AE	17	12.43	0.77	-21.57
5. Measurement model AE $\lambda_1 = \lambda_2; \kappa_1 = \kappa_2$	19	12.45	0.87	-25.55 <sup>a</sup>

<sup>a</sup>Best-fitting model.Saturated model fit:  $\chi^2 -4,625.939$ ; df 4,898.