Using a factor mixture modeling approach in alcohol dependence in a general population sample

Po-Hsiu Kuo\textsuperscript{a,b,∗}, Steven H. Aggen\textsuperscript{b}, Carol A. Prescott\textsuperscript{c}, Kenneth S. Kendler\textsuperscript{b}, Michael C. Neale\textsuperscript{b}

\textsuperscript{a} Institute of Clinical Medicine, College of Medicine, National Cheng Kung University, Taiwan
\textsuperscript{b} Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond VA, USA
\textsuperscript{c} Department of Psychology, University of Southern California, Los Angeles, CA, USA

Received 6 April 2007; received in revised form 13 March 2008; accepted 26 April 2008
Available online 30 June 2008

Abstract

Alcohol dependence (AD) is a complex and heterogeneous disorder. The identification of more homogeneous subgroups of individuals with drinking problems and the refinement of the diagnostic criteria are inter-related research goals. They have the potential to improve our knowledge of etiology and treatment effects, and to assist in the identification of risk factors or specific genetic factors. Mixture modeling has advantages over traditional modeling that focuses on either the dimensional or categorical latent structure. The mixture modeling combines both latent class and latent trait models, but has not been widely applied in substance use research. The goal of the present study is to assess whether the AD criteria in the population could be better characterized by a continuous dimension, a few discrete subgroups, or a combination of the two. More than seven thousand participants were recruited from the population-based Virginia Twin Registry, and were interviewed to obtain DSM-IV (Diagnostic and Statistical Manual of Mental Disorder, version IV) symptoms and diagnosis of AD. We applied factor analysis, latent class analysis, and factor mixture models for symptom items based on the DSM-IV criteria.

Our results showed that a mixture model with 1 factor and 3 classes for both genders fit well. The 3 classes were a non-problem drinking group and severe and moderate drinking problem groups. By contrast, models constrained to conform to DSM-IV diagnostic criteria were rejected by model fitting indices providing empirical evidence for heterogeneity in the AD diagnosis. Classification analysis showed different characteristics across subgroups, including alcohol-caused behavioral problems, comorbid disorders, age at onset for alcohol-related milestones, and personality. Clinically, the expanded classification of AD may aid in identifying suitable treatments, interventions and additional sources of comorbidity based on these more homogenous subgroups of alcohol use problems.

Keywords: Latent trait; Latent class; Mixture model; Diagnosis

1. Introduction

Alcohol abuse or dependence (AAD) is a clinically heterogeneous condition, with affected individuals varying in age of onset, clinical presentation, comorbid psychopathology, and severity. The diagnosis of AAD is usually based on subjective self-report symptoms or behaviors, as opposed to biological markers or known indices of pathophysiology. According to the DSM-IV (Diagnostic and Statistical Manual of Mental Disorder, version IV, American Psychiatric Association, 1994), individuals who meet three or more of seven criteria at any time in a year period are diagnosed as having alcohol dependence (AD). One of the assumptions behind this sum-score based diagnosis system is that each symptom provides the same amount of information as to whether an individual is affected. This assumption is typically not tested and is probably unrealistic. A further problem with the requirement of any combination of three out of seven symptoms is that those meeting criteria may be quite heterogeneous. Such variation between subjects with the same classification may hamper attempts to identify risk factors,
syndromes in a clinical sample (Allen et al., 1993). Using all structure of alcohol problems and whether there are distinct groups. was no correlation between the symptoms within either of the these symptoms covary in the combined population, even if there consequently endorsed than in a second group, it would appear that group both withdrawal and tolerance symptoms are more fre-

More subgroups. These subgroups differ in the means or vari-

ation arises solely because the population consists of two or

Usually, LCA assumes that symptom covariation in the popu-

'factors') are assumed to be normally distributed. Their influ-

ence on two or more symptoms generates covariation between the symptoms, and it is these patterns of symptom covariation that suggest the presence of latent factors.

An alternative statistical framework is that of latent class anal-

ysis (LCA), which is a person-centered approach that can be used to explore the spectrum of severity based on an individual’s profile of symptoms (Lazarsfeld and Henry, 1968; Moustaki, 1996). Usually, LCA assumes that symptom covariation in the population arises solely because the population consists of two or more subgroups. These subgroups differ in the means or variances of at least two of the symptoms. For example, if in one group both withdrawal and tolerance symptoms are more fre-

quently endorsed than in a second group, it would appear that these symptoms covary in the combined population, even if there was no correlation between the symptoms within either of the groups.

In the past, several studies have investigated the dimensional structure of alcohol problems and whether there are distinct classes of individuals with respect to heavy alcohol use. Factor analysis suggested a unidimensional structure of the AD syndromes in a clinical sample (Allen et al., 1993). Using all DSM-IV alcohol diagnostic criteria, some studies supported a unidimensional structure for alcohol problems using a Rasch model and latent trait modeling (Krueger et al., 2004; Kahler and Strong, 2006; Proudfoot et al., 2006; Saha et al., 2006), while others suggested a two-dimensional model (i.e. the abuse and dependence items fall into different factors). Krueger et al. (2004) conducted a comprehensive examination of 110 alcohol problems. Although they suggested that alcohol problems can be organized along a continuum severity, the performance of DSM-III-R abuse and dependence was quite different when comparing the relative shapes of the severity indexes by the set of items mea-

sured in a specific diagnostic category. The abuse items followed a bi-model distribution with items found in both the mild and severe ends of the continuum, whereas the dependence items tended to be located at the severe end of the continuum. Nelson et al. (1999) findings supported a 2-factor solution in a mixed population and clinical sample for dependence and abuse. Using large-scale data from the National Longitudinal Survey, Harford and Muthen (2001) found a two-dimensional structure with one main factor on AD; they also provided further support for the validity of AD in general population samples. While it is still in debate whether abuse should be included in the same contin-

uum as dependence, AD appears to be a more reliable unitary construct. In addition, two prospective studies used general pop-

ulation samples have indicated that the course of AD differs from that of abuse (Hasin et al., 1990, 1997), and implied that alcohol dependence and abuse are different in nature. Because alcohol abuse and dependence also have different features in terms of their clinical presentation, etiology and demographic character-

istics (Hasin et al., 2007), we focus on only alcohol dependence for the present study.

On the other hand, results from the LCA studies consistently showed a severity-based grouping in both adult and adoles-

cent samples. A 4-class solution has been identified for both genders among adult relatives of alcoholic probands, in which classes were distinguished primarily by increasing probability of endorsement (Bucholz et al., 1996). Using a large Australian twin sample, Lynskey et al. (2005) results revealed a severity-based 4-class solution in females and a 5-class solution in males based on eleven DSM-IV dependence and abuse symptoms. For adolescent samples, a 5-class severity-based solution was reported for female twins (Bucholz et al., 2000) and a 3-class solution represented increasing severity of alcohol problems among adolescents in addiction treatment program (Chung and Martin, 2001).

Conceptually, FA and LCA provide different explanations of the covariation among symptoms. Historically, they have been applied independently and typically for different purposes. Yet it is possible to compare the fit of these models using omnibus model fitting indices, such as log-likelihood or parsimony-based indices such as Akaike’s Information Criterion (AIC, Akaike, 1974) or Bayesian Information Criterion (BIC, Schwarz, 1978) that take into account the number of parameters estimated in the models. For example, LCA and latent-trait response models were applied separately on symptoms of major depression (MD) based on DSM-III-R criteria in a population-based twin sample (Aggen et al., 2005; Sullivan et al., 2002). Results from LCA suggested seven latent classes that had interpretable profiles corresponding to typical MD, atypical MD, and minor depressive states (Sullivan et al., 2002). The factor and item response model applied to the same data identified a unidimensional scale of depression liability (Aggen et al., 2005). These studies illustrate an important point – that the same set of symptoms can be used to estimate both their defining properties of a unidimensional continuous scale of liability or unobserved class membership structures – neither of which would be possible if symptoms had been aggregated into an affected vs. unaffected binary diagnostic variable.

In the last decade, efforts to identify unobserved heterogeneity have led to the development of advanced structural equation models. Only recently have the two concepts (continuous dimensionality and categorical subtypes) been unified in a single combined model (Dolan and Van der Maas, 1998; Muthen and Shedden, 1999; Yung, 1997), usually called a ‘Factor Mixture Model (FMM)’. FMM’s are a variant of finite mixture models (Everitt, 1988; Jedidi et al., 1997; McLachlan and Peel, 2000) which consist of a limited number of mixture components (the latent class model is also a finite mixture model). The FMM features two types of latent variables, namely a latent categor-
ical variable, and one or more continuous factors within each class. It can be used as a tool to address the discrimination between latent classes on the one hand, and continuous dimensions on the other (i.e. if there exist underlying dimensions for the symptoms covariation, whether there are heterogeneous subgroups in the general population). The hope is that these models can provide superior measurements and classification of complex phenotypes. One of the implications of the availability of these new techniques is to provide empirical findings that will help inform decisions to refine the DSM criteria and possible subtypes. Nevertheless, only a few studies have employed the mixture models to study problematic behaviors and substance use. With cross-sectional data, mixture models were applied to tobacco dependence data and they fitted data better than either latent-trait or LCA models (Muthen and Asparouhov, 2006; Muthen, 2006). Different kinds of combined mixture models have also been applied in longitudinal studies to explore subgroups with different growth trajectories. For instance, growth mixture modeling was applied to antisocial behavior and heavy drinking (Muthen and Muthen, 2000) and applied to assess intervention effects of reducing aggressive behavior (Muthen et al., 2002).

The present study links the variable-centered (latent trait) and person-centered (latent class) approaches in the substance use literature for AD. Our goal is to assess whether the factor structure, latent class, or finite mixture models provide better explanations of the patterns of symptoms co-occurrence for AD in a large population-based sample. In addition, a series of exogenous variables were also included in the analysis to validate whether there are distinguishable subgroups of problematic alcohol use.

2. Methods

2.1. Sample and assessment procedures

Subjects in this study were participants in the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders and were recruited from the population-based Virginia Twin Registry (now part of the Mid-Atlantic Twin Registry). Longitudinal data were collected from female–female twin pairs (FF) who participated in up to four personal interviews and from male–male and male–female twin pairs (MM/MF) who participated in one or two waves of interviews. The current report is based on individuals with completed interviews, including 3325 females with a mean age when last assessed of 36.5 years (range 20–62) and 4217 males with mean age of 37.0 years (range 20–58). Details of sample ascertainment and characteristics are presented elsewhere (Rendler et al., 2004).

The DSM-IV symptoms and diagnosis of AD were assessed using adapted versions of standard structured interviews, SCID (the Structured Clinical Interview for DSM-III-R) (Bucholz et al., 1994; Spitzer and Williams, 1985). The individual symptoms of AD include tolerance (TOL), withdrawal symptoms (WD), drinking more than intended (MORE), unsuccessful attempts to cut down on use (CUT), excessive time related to alcohol (TIME), impaired social or work activities (IMP), and use despite physical consequences (PHY). Lifetime DSM-IV AD history was assessed at waves 4 for the FF sample (FF4) and waves 2 for the MM/MF sample (MM/MF2). The interview used to assess the DSM-IV AD criteria contained a stem item. Individuals who answered no to “have you ever taken a drink in your lifetime” skipped out the rest of AD assessment section. In total, there were 226 individuals reporting they never had taken a drink in their lifetime. For these abstainers, we assumed all DSM-IV symptom criteria were negative and coded them 0. Therefore, abstainers were not distinguished from individuals who did drink but never developed any alcohol problems meeting the DSM-IV criteria.1

Randomly selected sub-samples from FF4 (N = 192) and MM/MF2 (N = 195) were interviewed a second time within 2–8 weeks of their original interview. The intraclass test–retest correlations indicate excellent reliability for AD (0.96 among FF and 0.83 among MM/MF twins) (Kuo et al., 2006) and good reliability at the symptom level (correlations ranged from 0.67–0.84 in dependence symptoms). In addition to diagnostic information of AD and other psychiatric disorders, we also collected a battery of self-report questions in the interview, such as demographic variables and personality questionnaires.

2.2. Statistical models

FMM models are suitable for multivariate data, which could be dichotomous, ordinal, or continuous indicators. Within each latent class there may exist one or more continuous factors that account for “residual” covariation among the indicators. Note that the FA model and LCA are special cases of the FMM. If the factor variance within each class is set to be zero, observed variables are independent given class (i.e., the assumption of local independence holds) and the FMM reduces to the latent class model because the factors can be dropped from the model. On the other hand, if only 1 class is specified, then the FMM model is reduced to a common factor model because the latent class variable can be dropped.

Fig. 1 displays the analytic diagram of our modeling framework. It includes LCA models with 2–5-class solutions (lca#_c models), factor models with 1 and 2 factors (factor# models), and factor mixture models with 1 factor and 2–4 classes (fmm1_#_f models).2 To compare the model fitting with the DSM-IV diagnostic system, we also included models that grouped people into diagnosis/non-diagnosis AD categories based on DSM-IV symptoms counts with 1 or 2 classes in each group (details see the methods section). LCA: latent class analysis; FMM: finite mixture model; FA: factor analysis.

---

1 When included, the 226 abstainers were all found to occupy the non-problem drinking class (their class-membership probabilities were all >0.99). Excluding them from the analyses did not change the overall model fitting results. Alternatively, one can treat them as missing data in the analysis.

2 The mixture models allow for multidimensional structure. Because of previously suggested unitary construct of AD, we only consider mixture models with 1 factor in our analytic diagram.
(group,2 model) given the flexibility to model heterogeneity within diagnostic and non-diagnostic group as each group allows having 2 classes.

The three main types of model were fitted to dichotomous AD symptoms based on DSM-IV criteria in our sample. These analyses were conducted separately by gender using the Mx program (Neale et al., 2003). The models displayed in the analytic diagram were compared using several goodness-of-fit indexes. In the case of latent class models, simulations showed that in some scenarios, BIC and adjusted BIC (especially in complex structure where items can have different endorsement probabilities for more than one latent classes) outperforms AIC (Nylund et al., 2007). Therefore, AIC,3 sample-size adjusted BIC (sBIC), and log-likelihood ratio test were used to compare across different models or among nested models. First, the more parsimonious models were identified for males and females separately. We then constrained parameters across genders in these models to test for gender invariance. For factor and mixture models, we constrained factor loadings and item response probabilities; for latent class models, we constrained item response probabilities across genders.

Finally, if the model fitting results support the existence of subgroups of alcohol use, we examine sets of exogenous variables to further characterize the subgroups. These measures included alcohol-caused behavioral problems (BP), comorbid disorders, age at onset for alcohol-related milestones, and personality. For BP, we have complete measures of four problems (BP1: legal problems or traffic accidents; BP2: increasing chance of being injured; BP3: drunk or hangover; BP4: social activity impairment). We also included several potentially comorbid disorders: major depression (MD), generalized anxiety disorder (GAD), any phobia (Phobia), conduct disorder (CD), antisocial personality disorder (APD), any other illicit drug abuse (ADA), any other illicit drug dependence (ADD). In terms of alcohol-related milestones, we measured the subjects’ reported age at which they: (i) first got drunk; (ii) drank regularly; (iii) had the first AD symptom; and (iv) had sufficient symptoms to meet criteria for an AD diagnosis. For personality, we used revised short-form of neuroticism (N), extroversion (E), and novelty seeking (NS).

One thing to note is that we analyzed all the above models using the Virginia twin sample but we did not correct for correlation between twins, because doing so would greatly increase the complexity of the models (see more details in discussion). In general, it is more reasonable to correct for clustering when the average group size is moderate to large; in the present case the group size is mostly two (31.1% one twin, 68.5% both twins, and 0.3% triplets and quadruplets) so the consequences for the model-fitting results are likely to be trivial (Rebollo et al., 2006).

3. Results

In our sample, 26.1% of men and 10.1% of women met criteria for DSM-IV lifetime AD diagnosis. Consistent with this prevalence difference, males have significantly (p < 0.001) higher endorsement probability for every AD symptom than females (see Fig. 2a). However, among those who meet AD diagnostic criteria, the symptom endorsement probability is relatively similar across genders (Fig. 2b). Men endorsed the TOL and WD symptoms more frequently, and women endorsed the CUT symptom more frequently (p < 0.01). Of all seven symptoms, MORE is the most frequently endorsed (more than 90% of those meeting diagnostic criteria reported drinking more than intended in both genders) and WD, IMP, and PHY are the low to moderate endorsed symptoms.

The results of model fitting, −2 times log-likelihood (−2LL), AIC, and sBIC for each model in Fig. 1 are listed in Table 1 and the −2LL are also shown in Fig. 3a and b. Because AIC has a linear relationship with −2LL and is a function of the number of parameter freely estimated in the model (p), two AIC contour lines were drawn in Fig. 3 to assist with the comparison of model fits. The first AIC line passes model factor1 (e.g. with a p of 15 for factor1 model the intercept of the AIC line for females equals to 7601.235 + 30), and the second passes model fmm1f3c1 to make most of our models fall in between these two AIC contour lines. The model that is close or falls below the second AIC line has lower −2LL and fits data well. For both genders, the DSM-IV diagnostic models (group,2 and group,2) fit our data poorly and the −2LLs are far above the AIC contour lines.

The LCA models with three to 5 classes perform better in general than the FA models, which contain several the 2nd or 3rd best-fitting models based on AIC and sBIC selection (see model fitting results in Table 1). The index AIC tends to choose model with one more class than that using sBIC, which is consistent with the findings in Nylund and colleagues. According to indexes AIC and sBIC, the best-fitting models for both genders are FMM models; FMM with 1 factor and 2 classes is the best-fitting model for both genders using sBIC, but 1 factor 3 classes model is for males using AIC. Using −2LR (−2 times log-likelihood ratio) test, the FMM 1-factor 2-class model did not differ significantly in fit from 3 classes model in females (p = 0.15), but was significantly worse than the 3-class model in males (p = 0.007). According to the model fitting results, we...
Table 1
Model fitting results for models in Analytic diagram (see Fig. 1) by gender

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th></th>
<th></th>
<th></th>
<th>Male</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>−2LL</td>
<td>AIC</td>
<td>sBIC</td>
<td></td>
<td>−2LR</td>
<td>AIC</td>
<td>sBIC</td>
</tr>
<tr>
<td>lca_{2c}</td>
<td>16</td>
<td>7798.053</td>
<td>−38721.947</td>
<td>−53457.487</td>
<td>19717.387</td>
<td>−39290.613</td>
<td>−66398.815</td>
<td></td>
</tr>
<tr>
<td>lca_{3c}</td>
<td>24</td>
<td>7551.881</td>
<td>−38952.119</td>
<td>−53560.848</td>
<td>18938.960</td>
<td>−40053.040</td>
<td>−66767.372</td>
<td></td>
</tr>
<tr>
<td>lca_{4c}</td>
<td>32</td>
<td>7523.628</td>
<td>−38964.372</td>
<td>−53535.464</td>
<td>18905.866</td>
<td>−40070.134</td>
<td>−66756.723</td>
<td></td>
</tr>
<tr>
<td>lca_{5c}</td>
<td>40</td>
<td>7512.241</td>
<td>−38969.759</td>
<td>−53531.045</td>
<td>18875.082</td>
<td>−40084.918</td>
<td>−66751.726</td>
<td></td>
</tr>
<tr>
<td>fmm_{1f2c}</td>
<td>30</td>
<td>7500.969</td>
<td>−38991.031</td>
<td>−53571.507*</td>
<td>18885.487</td>
<td>−40094.513</td>
<td>−66778.579*</td>
<td></td>
</tr>
<tr>
<td>fmm_{1f3c}</td>
<td>45</td>
<td>7480.234</td>
<td>−38981.766</td>
<td>−53544.886</td>
<td>18853.908</td>
<td>−40096.092*</td>
<td>−66755.599</td>
<td></td>
</tr>
<tr>
<td>fmm_{1f4c}</td>
<td>60</td>
<td>7477.397</td>
<td>−38954.603</td>
<td>−53509.316</td>
<td>18839.074</td>
<td>−40080.926</td>
<td>−66724.246</td>
<td></td>
</tr>
<tr>
<td>factor1</td>
<td>15</td>
<td>7601.235</td>
<td>−38918.765</td>
<td>−53555.896</td>
<td>19234.329</td>
<td>−39773.671</td>
<td>−66640.343</td>
<td></td>
</tr>
<tr>
<td>factor2</td>
<td>20</td>
<td>7560.952</td>
<td>−38949.048</td>
<td>−53563.708</td>
<td>19148.048</td>
<td>−39849.952</td>
<td>−66670.560</td>
<td></td>
</tr>
<tr>
<td>group_{c1}</td>
<td>16</td>
<td>8704.830</td>
<td>−37815.170</td>
<td>−53004.099</td>
<td>20779.159</td>
<td>−38228.841</td>
<td>−65867.928</td>
<td></td>
</tr>
<tr>
<td>group_{c2}</td>
<td>32</td>
<td>7798.848</td>
<td>−38689.152</td>
<td>−53417.635</td>
<td>19570.052</td>
<td>−39405.948</td>
<td>−66431.127</td>
<td></td>
</tr>
</tbody>
</table>

Notes: p – numbers of parameter estimated; LL – log-likelihood; AIC – Akaike’s Information Criterion; sBIC – sample size adjusted Bayesian Information Criterion. Numbers in bold represent the best three models and asterisk (*) represents the best model based on the index selection.

examined gender invariance among several models to assist for final model selection, including lca_{3c}, lca_{4c}, fmm_{1f2c}, and fmm_{1f3c}. The LCA models with constrained item response probability across genders significantly worsen model fitting ($P < 0.0001$); while both FMM models exhibit gender invariance in factor loadings but not in item response probability. In addition, both FMM models with gender invariance in factor loadings have better fit indexes results than those of LCA models using AIC and sBIC. Therefore, the final model is chosen between the FMM 1 factor 2- or 3-class models.

To make comparison between the FMM 1 factor 2- or 3-class models (with gender invariance in factor loadings) more meaningful and interpretable, we plot item response probabilities for each class in these two models by gender (Fig. 4). The fmm_{1f2c} model has very similar patterns of item response probabilities for both genders (Fig. 4a), however, it basically forms one non-drinking and one drinking group and therefore does not identify different patterns of alcohol use. For fmm_{1f3c} model (Fig. 4b), there were 2 classes of drinking group and the item response patterns display qualitative differences between the two drinking groups. In general, model-fitting should not focus exclusively on goodness-of-fit criteria, but also on the interpretability of results (Barrett, 2007; Burnham and Anderson, 2002).

We therefore chose the FMM 1-factor and 3-class model with gender invariance in factor loadings as the final model to compare parameter estimates across genders. Results indicate one severe and one moderate dependence drinking group (classes 1 and 2), and one non-problem drinking group (class 3). The class membership probabilities for classes 1–3 are 5.2%, 12.3%, 82.5% in females and 8.4%, 24.4%, 67.3% in males. In this model, residual correlations among symptoms have been taken into account in each class by including the factor structure. The unidimensional factor structure seems to serve reasonably well. Constraining the factor loadings to be equal across genders in

![Fig. 3. (a and b) The −2 times log-likelihood (−2LL) for each model described in the analytic diagram by gender (see note in Fig. 1). Notes: Two AIC contour lines were added, which pass through model “factor1” and model “fmm_{1f3c}”.](image-url)
Fig. 4. (a and b) Symptom profiles of model (a) fmm₁f₂c and (b) fmm₁f₃c.

Notes: The endorsement probability for each symptom is based on the model with constrained factor loadings in classes 1 and 3 across genders and the factor loadings in the moderate dependence group in females were specified to be proportional to those in males. The individual symptoms of AD: tolerance (TOL), withdrawal symptoms (WD), drinking more than intended (MORE), unsuccessful attempts to cut down on use (CUT), excessive time related to alcohol (TIME), impaired social or work activities (IMP), and use despite physical consequences (PHY).

The class symptom profiles varied somewhat across genders (see Fig. 4a and b). For men, they seem to form severity-based subgroups in that almost every symptom was endorsed more frequently in class 1 than class 2 other than TOL. Also the severe dependence group has great impairment (IMP) compared to the moderate dependence group. By contrast, in women the main difference between these 2 classes was that WD and IMP were endorsed more frequently in class 1. Comparing the symptom profiles for each class across genders, results showed that TIME and IMP were more frequently endorsed by men than women in the severe dependence group, whereas in the moderate dependence group, TOL was the symptom more frequently endorsed by men.

3.1. Validation of classes via comparison of exogenous variables

Modeling results suggested that three latent classes could be reliably identified from the DSM AD symptoms endorsement patterns. We next sought to validate these subgroups by examining a number of putative variables that might predict class membership. These included alcohol-related behavioral problems; comorbid psychiatric disorders; age at onset for alcohol-related events; and personality measures (see Fig. 5a).

Among four alcohol-caused behavioral problems in our measure (BP1: legal problems or traffic accidents; BP2: increasing chance of being injured; BP3: drunk or hangover; BP4: social activity impairment), there was a consistent trend of increasing endorsement probability with increasing severity level of alcohol dependence. It is noteworthy that even within the non-problem drinking group, 12% reported ever having legal problems or traffic accidents and 20% reported an increased chance of being injured because of drinking. In addition, both severe and moderate dependence groups had significantly higher (2–3-fold) maximum drinks in a single day than did the non-problem drinking group.

We found that the lifetime prevalence of every comorbid disorder we examined was the lowest in the non-problem drinking group, and significantly greater in moderate and severe dependence groups. Rates of comorbidity were especially high for GAD (22%), CD (38%), APD (12%), ADA (55%), and ADD (35%) in the severe dependence group, compared to 7%, 9%, 1%, 11%, and 3%, respectively, in the non-problem drinking group. This observation is consistent with results from the National Comorbidity Survey, in which high comorbidity between drinking problems and other psychiatric disorders was also found (Kessler et al., 1997).

Three age at onset milestones for alcohol-related problems were considered: (i) first got drunk; (ii) drank regularly; (iii) had the first AD symptom; and (iv) had sufficient symptoms to meet criteria for an AD diagnosis. For each stage, the severe depen-
Concomitant drinking group has the earliest age at onset, followed by moderate and non-problem drinking groups (see Fig. 5b). The time difference from drinking regularly to AD diagnosis was longer in the non-problem drinking group (16.3 years) than in the severe dependence group (2.7 years). The non-problem drinking group (the majority of our sample) exhibits slow development of alcohol use. In addition, since MD has very high comorbidity rate with AD (65%, 44%, 33% in classes 1–3), we also reported age at onset of MD. The mean age at onset of MD differed between groups: individuals in the dependence severity groups had MD onset significantly ($P < 0.0001$) earlier (25–26 years) than in the non-problem drinking group (31 years).

We also examined personality scores based on revised short-form of neuroticism (N), extroversion (E), and novelty seeking (NS) in different classes. For NS and E, both traits were higher in the two dependence severity groups than in non-problem drinking group ($P < 0.05$) but there was no significant difference between the two dependence severity groups. However, for N (with a maximum sum score of 12), there was a significant difference with dependence severity (both global and pair-wise comparisons, $P < 0.0001$, with moderate Cohen’s $d$ effect sizes of 0.23–0.62), with a mean score of 5 in the severe, 3.6 in the moderate, and 2.9 in the non-problem drinking group.

4. Discussion

The primary goal of this study was to assess whether the alcohol dependence criteria in the population could be better characterized by a continuous dimension, a few discrete subgroups, or a combination of the two. Although both factor and latent class models have been previously used to explain the population aggregation of AD symptoms (Allen et al., 1993), no clear structure has emerged. A FA approach seeks to understand the underlying continuous dimensions for symptoms, which is an important initial stage of developing constructs in the use or abuse of either a single substance (Harford and Muthen, 2001; Hasin et al., 1997; Lennox et al., 1996) or of a variety of substance disorders (Nelson et al., 1999) but is more difficult to provide direct clinical applications, such as treatment response and evaluation. While LCA models frequently support a severity-based classification in alcohol problems (Bucholz et al., 1996; Lynskey et al., 2005) or cannabis use (Grant et al., 2006), the symptom profiles from these analyses appear largely parallel, and involve relatively small differences in item endorsement probability. Such minor differentiation between classes can make it difficult to distinguish 1 class from another, and also problematic to assign class membership to individuals based on their item response pattern. A factor mixture model, which incorporates dimensions within classes, can add to the interpretation and understanding of the differences between classes. Unaccounted for covariation within the class structure can be modeled by the continuous latent variables, and the smaller number of classes is more likely to reflect qualitatively heterogeneous groups in the population. For instance, subgroups obtained from classification analysis have different characteristics and may respond differently in treatment or prognosis (see examples in Muthen et al., 2002 using growth mixture modeling).

In the present study, we illustrated that factor mixture models that combine categorical and continuous latent variables are a promising tool for this kind of phenotypic analyses. Model fitting results showed that a pure FA model was not satisfactory and that the LCA models fit better in general. This indicates a continuous latent construct could not explain the overall covariation among alcohol symptoms and the heterogeneous nature of alcohol problems in the general population. However, the mixture model with 1 factor and 3 classes model fit our data well, with one major non-drinking dependence group and two various severity levels of dependence groups. This finding is consistent with a previous study applied mixture models in tobacco dependence (Muthen and Asparouhov, 2006), which also found the FMM fits their data the best. In general, for heterogeneous behavioral traits, the mixture model seems to fit data better than either of the conventional models in both cases of cross-sectional or longitudinal data (see the examples of tobacco dependence and heavy drinking; Muthen and Asparouhov, 2006; Muthen and Muthen, 2000). Other than model-based classification analyses, Cloninger (1987) proposed a neurobiological learning model.
that split alcohol use disorders into two subtypes (types I and II). This dichotomy has been studied with respect to several aspects of alcohol problems. Results suggest that the two subtypes differ on dependence severity, age-at-onset, or treatment response. Although this scheme provides a way to subgroup individuals with AD, it is largely dependent on existing personality constructs (type I alcoholism had high harm avoidance and low novelty seeking personality). Compared with model-based phenotypic analysis, mapping the current DSM criteria onto the type II/II dichotomy has been more difficult. Future research may be better served by testing for latent classes directly from criteria level data.

In our sample, 96.1% of individuals in severe dependence group met diagnosis of AD, compared to 76.4% in moderate dependence group and 4.8% in non-problem drinking group; the average numbers of AD symptoms were 5.3, 3.1, and 0.2, respectively. Although many individuals in the moderate dependence group endorsed symptoms like TOL, MORE, CUT, and TIME (see Fig. 4a and b), a quarter of them did not meet the present DSM-IV AD diagnosis. From a clinical point of view, failure to detect these sub-clinical individuals or so-called ‘diagnostic orphans’, who may be in an earlier stage of development of drinking dependence problems, may hamper an effective treatment or intervention (Note: The moderate dependence group has relatively younger average age, 3 years younger than severe dependence group). In addition, although model-fitting results supported a unidimensional factor structure for AD symptoms, factor loadings were not consistent across 3 classes. The factor loadings were high for all symptoms in the severe dependence and non-problem drinking groups, but were low for several symptoms in the moderate dependence group (<0.2 for TOL, CUT, and TIME among females and <0.3 for TOL and MORE among males). The TOL criterion factor loading was especially low in both genders, implying that, relative to the other AD criteria; TOL was less endorsed and poorly discriminated in the moderate dependence group.

The mixture of AD and non-AD individuals in the moderate drinking group may explain why the two models based on DSM-IV diagnostic system fit the data very poorly. In addition, we fitted mixture models with the DSM-IV diagnostic system. Including dimensionality (one latent factor) into the DSM-IV framed models improves model fit (data not shown) but is still much worse than the freely estimated mixture 1-factor 3-class models. That finding in turn provides empirical evidence for the heterogeneity in AD diagnosis and suggests the reconsideration of the diagnostic system to classify according to, e.g., the 3-class system found here. Presumably individuals in severe or moderate drinking groups may vary in their clinical course, treatment response, or prognosis. In a predictive model, given a certain symptom profile, individuals could be best predicted to fall into one of the classes. From a clinical point of view, this refinement of grouping approach could help clinical workers to provide a suitable treatment or intervention for other comorbid disorders in more homogenous subgroups. We are developing a small web-based application that gives the likelihood for particular patterns of responses. This will provide a way to classify individuals based on different combinations of symptoms and also report the corresponding features of those exogenous variables that were tested in our study, including alcohol-related behavioral problems, comorbid psychiatric disorders, age at onset for alcohol-related events, and personality features.

The class membership probability for the non-problem drinking group was higher for females (82.6%) than males (67.4%), while class probabilities for other problematic drinking groups (classes 1 and 2) in males was nearly twice of those in females, which is consistent with the well-recognized gender differences in the prevalence of drinking problems (Green et al., 2004; Kessler et al., 1994). However, in other respects males and females were remarkably similar. First, the symptom endorsement probabilities among individuals diagnosed with AD are almost the same in males and females (Fig. 2b). Second, the age at onset of AD distributions are similar with a peak of 18–20 years in this sample for both genders (Kuo et al., 2006). Third, we calculated the class membership agreement within twin pairs. The agreement did not differ between females and males (weighted $\kappa = 0.21$, 95% CI = 0.13–0.29 for female twins and 0.17, 95% CI = 0.12–0.22 for male twins). However, the agreement was substantially greater in monozygotic (MZ) compared to dizygotic (DZ) twins (weighted $\kappa = 0.31$, 95% CI = 0.25–0.37 for MZ and 0.13, 95% CI = 0.09–0.17 for DZ twins). If restricted to two alcohol problem groups, the tetrachoric correlations between twins were 0.53 for MZ and 0.02 for DZ twins, implying a moderate genetic component for class assignment of alcohol problematic use. Additional differences were found between groups. Those classified in the more severe alcohol problem groups reported more alcohol-caused behavioral problems, had higher risk of having comorbid psychiatric disorders, exhibited a faster development of alcohol problems from regular drinking to AD diagnosis, and higher score of personality trait neuroticism.

The results presented in this report should be considered in the light of three potential limitations. First, the results are based on Caucasian twins born in Virginia and may not generalize to individuals from other ethnic backgrounds or geographical regions. Second, as we mentioned in the method section, we did not directly model the correlation between twins in order to avoid further complicating the models. We expect the consequences of non-independence for the model-fitting results to be negligible due to the small group size (i.e., here two for the twins). We re-fit our models to data where one twin was randomly selected from each pair ($N = 4460$) and found similar model fitting results. According to AIC and sBIC, the $fmm-1f3c$ and $fmm-1f2c$ models fit the single record data the best, respectively. This concurs with our general experience that confidence intervals are only modestly affected when models fit to non-independent (paired) data but that parameter point estimates typically are not. Third, we used outcome measures based solely on DSM-IV AD binary diagnostic symptoms; there may be other measures or instruments that could capture the distribution of alcohol problems better but were not included in the present study.

Finally, further work is needed. A larger sample (e.g. a national representative sample) could provide more power to distinguish and help choose among models. Including more symptoms/items and using ordinal or continuous measures which provide greater information to sub-classify and quantify.
individuals’ patterns of drinking behavior would also be useful. Longitudinal data would provide an opportunity to examine the degree to which individual’s transition between classes, or change on the dimension within a class, or both. For instance, subgroups obtained from classification analysis have different characteristics and may respond differently in treatment or prognosis (see examples in Muthen et al., 2002 using growth mixture modeling).

Conflict of interest

None declared.

Acknowledgements

The authors are grateful to the twin participants. The Virginia Twin Registry, now part of the Mid-Atlantic Twin Registry directed by L. Corey and L. Eaves, provided subject access. Patsy Waring, Frank Butera, Sarah Powers, Barbara Brooke and Lisa Halberstadt supervised data collection, Helen Wang provided computational assistance, and Indrani Ray provided database management.

Role of funding source: This work was funded by U.S. National Institutes of Health grants R01-MH-65322, R01-AA-11408 and DA-011287 for data analysis and original data collections. The NIMH had no further role in the writing of the report, or in the decision to submit the paper for publication. Preparation of this manuscript was supported by a Young Investigator award from the National Alliance for Research on Schizophrenia and Depression to Po-Hsiu Kuo.

Contributors: P.H. Kuo conducted analyses and wrote this manuscript. S.H. Aggen, C.A. Prescott and K.S. Kendler helped for data collection, management, as well as provided useful discussion. M.C. Neale assisted for framing statistical methods and study design. All authors contributed to and have approved the final manuscript.

References


Muthen, B., 2006. Should substance use disorders be considered as categorical or dimensional? Addiction 101, 6–16.


