Are the Symptoms of Cannabis Use Disorder Best Accounted for by Dimensional, Categorical, or Factor Mixture Models? A Comparison of Male and Female Young Adults

Nathan A. Gillespie  
Virginia Commonwealth University and Queensland Institute of Medical Research

Michael C. Neale  
Virginia Commonwealth University

Lisa N. Legrand, William G. Iacono, and Matt McGue  
University of Minnesota

Despite the consensus that criteria for cannabis abuse and dependence and symptoms of withdrawal are best explained by a single latent liability, it remains unknown whether alternative models provide a better explanation of these criteria. A series of latent factor, latent class, and hybrid factor mixture models were fitted to data from 872 recent cannabis users from the Minnesota Twin Family Study who completed Diagnostic and Statistical Manual of Mental Disorders (3rd ed., revised, and 4th ed.) diagnostic criteria for cannabis abuse, dependence, and symptoms of withdrawal. Despite theoretical appeal, results did not support latent class or factor mixture modeling. Instead, symptoms of abuse, dependence, and withdrawal were better summarized by a single latent factor Cannabis Use Disorder (CUD) for male and female young adults. An almost 2-fold sex difference in item endorsement was best explained by a single mean difference on the CUD factor, indicating that young men have a greater latent liability toward expressing CUD.

Keywords: cannabis, abuse, dependence, withdrawal, heterogeneity, latent class, latent factor, factor mixture

Cannabis is the most widely used illicit drug in developed countries including the United States (Dennis, Babor, Roebuck, & Donaldson, 2002; Hall, Johnston, & Donnelly, 1999). Population-based estimates of lifetime cannabis use in the United States between 1990 and 2004 range from 41.2% to 55.9% (Agrawal, Neale, Prescott, & Kendler, 2004b; Degenhardt et al., 2008; Kendler, Jacobson, Prescott, & Neale, 2003; Kendler, Karkowski, Neale, & Prescott, 2000; Substance Abuse and Mental Health Services Administration, 1998; von Sydow et al., 2001). Estimates of the lifetime prevalence of abuse range from 5.5% to 8.4%, whereas lifetime dependence estimates range from 1.3% to 2.2% (Agrawal & Lynskey, 2007; Stinson et al., 2005; von Sydow et al., 2001). The separation of the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM–IV) diagnostic criteria into the two distinct categorical outcomes of abuse and dependence (American Psychiatric Association, 1980, 1987, 1994; Edwards, Arif, & Hadgson, 1981) has been challenged in recent years. A number of reports examining cannabis and other substances using clinical and population-based samples have shown that the physiological, behavioral, and cognitive components of substance abuse and dependence, including withdrawal, are more consistent with a single psychometric factor model (Feingold & Rounsaville, 1995a, 1995b; Gillespie, Neale, Prescott, Aggen, & Kendler, 2007; Hartman et al., 2008; Langenbucher et al., 2004; Lynskey & Agrawal, 2007; Nelson, Rehm, Ustun, Grant, & Chatterji, 1999; Teesson, Lynskey, Manor, & Baillie, 2002). Consequently, the DSM–V is set to merge the abuse–dependence criteria along with symptoms of withdrawal into a single cannabis use disorder (CUD).

Despite this consensus, there is an absence in the cannabis literature in which alternative explanations concerning the coaggregation of symptoms have been tested and compared. Conceptually, latent factor analysis (LFA) and latent class analysis (LCA) represent competing hypotheses regarding symptom or item covariation. Historically, they have been applied independently and for...
different purposes. In reality, they are not mutually exclusive because ordered discrete distributions of latent classes can be used to approximate a latent trait (Vermunt & Hagenaars, 2004) in the same way a histogram approximates a continuous distribution (Schmitt, 2006).

LFA (Spearman, 1904) predicts that correlations between diagnostic criteria can be explained in terms of a reduced number of unobserved variables called latent factors. This method has proved useful for developing, validating, and calibrating drug abuse and dependence within and across substances (Harford & Muthén, 2001; Hasin et al., 1997; Kirisci et al., 2009; Lennox, Zarkin, & Bray, 1996; Nelson et al., 1999), including cannabis (Feingold & Rounsaville, 1995a, 1995b; Gillespie et al., 2007; Hartman et al., 2008; Langenbucker et al., 2004; Lynskey & Agrawal, 2007; Nelson et al., 1999; Teesson et al., 2002). The implicit conclusion drawn by cannabis researchers who have reported best-fitting single-factor structures assumes that drug-using populations are homogeneous. This is premature. For instance, there are significant cohort differences in the prevalence of use when measured in the United States between 1934 and 1974 (Kendler, Gardner, Jacobson, Neale, & Prescott, 2005), including higher rates of use, abuse, and dependence reported among males (Agrawal & Lynskey, 2007; Agrawal, Neale, et al., 2004b; Kendler et al., 2005; von Sydow et al., 2001). In addition to the need to determine whether these cohort and sex differences are qualitative versus quantitative in origin (Degenhardt, Lynskey, & Hall, 2000; Johnson & Gerstein, 1998; Muller & Gmel, 2002), the trajectory from cannabis initiation to regular use, abuse, and dependence could conceivably vary between subjects (phenotypic heterogeneity) or the same cannabis use disorder may result from different sets of risk factors (etiological heterogeneity; Ellickson, Martino, & Collins, 2004; Kandel & Chen, 2000; Windle & Wiesner, 2004).

An immediate alternative comparison ought to include LCA, which stringently assumes no covariance between diagnostic criteria. A population is predicted to consist of a small number of classes that differ in their means or variances (or item thresholds and item response probabilities) on at least two or more criteria. This approach has been used to identify severity spectrums for alcohol (Bucholz et al., 1996; Lynskey et al., 2005) and cannabis use (Grant et al., 2006). It also has been applied to evaluate drug treatment response and evaluation (Grant et al., 2006). Although theoretically plausible, a limitation of LCA is that very minor differences between classes can make it difficult to distinguish one class from the next. Moreover, because of the local independence assumption (Lazarsfeld & Henry, 1968), LCA, unlike LFA, does not distinguish individual within-class differences in severity (Muthén, 2006).

Hybrid latent class and factor models can circumvent the limitations of LFA and LCA, which historically have been applied independently, by providing a useful bridge between the two modeling traditions in the analysis of diagnostic criteria (Dolan & van der Maas, 1998; Everitt, 1988; Jedidi, Jagpal, & Desarbo, 1997; McLachlan & Peel, 2000; Muthén, 2006; Muthén & Shedden, 1999; Yung, 1997). This approach can be used to identify simultaneously categorical latent classes within a population as well as continuous variation or individual differences within classes. In factor mixture models (FMMs), the assumption of no covariance between items (i.e., item independence) is relaxed to allow for covariance between items but only within each class. In principle, FMMs may fit the data better than either LFA or LCA models; if so, they may provide a superior measurement and classification for complex phenotypes (see Helzer, van den Brink, & Guth, 2006). Simulations have demonstrated that it is possible to discriminate between latent class and factorial models using categorical data (Lubke & Neale, 2006, 2008). Recent evidence has shown that FMMs provide a better fit to DSM–IV criteria for prescription opioid abuse and dependence in adults than standard categorical diagnostics (Wu, Woody, Yang, Pan, & Blazer, 2010). Similarly, an FMM appears to provide a good, sex-invariant fit to DSM–IV diagnostic criteria for alcohol abuse and dependence (Kuo, Aggen, Prescott, Kendler, & Neale, 2008) and heavy drinking (Muthén & Muthén, 2000).

Our aim was to provide the first omnibus comparison between competing, alternative explanations of Diagnostic and Statistical Manual of Mental Disorders (3rd ed., revised; DSM–III–R) and DSM–IV cannabis diagnostic criteria to derive the best-fitting CUD phenotype from existing, widely used assessment items. Despite the fact that the genetic and environmental risks for cannabis use appear to be identical across sex (Kendler et al., 2005), rates of cannabis use and abuse are higher among males (Coffey et al., 2002; Kessler et al., 1994). It is unclear whether this sex difference is attributable to differences in item endorsement or reflects a greater latent liability among males indicated by latent mean or variance differences. Therefore, our second aim was to determine using tests of measurement invariance the source of any observed sex differences.

Method

Participants

Described in detail elsewhere (Iacono, Carlson, Taylor, Elkins, & McGue, 1999), subjects for this study came from the Minnesota Twin Family Study, which is a longitudinal study of two twin cohorts and their parents. Birth records and public databases were used to identify twins who were born in Minnesota between 1975 and 1984 and 1988–1994. These twins are predominantly Caucaussians (98%) whose average socioeconomic status level (Hollingshead, 1975) corresponds to parental occupations such as clerical, sales industry, and small-business owners. Only twins living within 1 day’s driving distance to Minneapolis, who could be accompanied by a biological parent, and who were not mentally or physically handicapped were eligible for participation. Cohort 1 was first assessed at age 11 with follow-ups at ages 14, 17, 20, 24, and 29 years. Cohort 2 was first assessed at age 17 with follow-ups at ages 20, 24, and 29 years. These follow-ups were timed to coincide with major transitions in the lives of adolescents and young adults. Participation rates have generally exceeded 90% across all follow-up assessments. We chose to use data from the third and first follow-ups of Cohorts 1 and 2, respectively, when twins were 20 years of age. The mean ages of onset and first symptoms of cannabis abuse were 18 and 19 years (Gillespie, Neale, & Kendler, 2009). The assessment was ideal, as twins had passed through the peak risk period while still being young enough to accurately respond to diagnostic criteria. The mean age at assessment was 21.7 years (6 = 0.01). There were 570 women and 392 men, or 42.7% and 62.1% of the sample, respectively, who reported lifetime cannabis use. Among them, there were 506
women (88.9%) and 366 men (93.3%) who reported using cannabis at least once within the past 4 years. Participants were on average over 5 years past the mean age of cannabis initiation (μ = 16.2 years, δ = 2.0), but they were still young enough to avoid recall bias associated with retrospective reporting at older ages (Pickles et al., 1998).

Measures and Reliability

**DSM-III-R** substance use disorder criteria were assessed using a modified version of the expanded Substance Abuse Module developed by Robins, Babor, and Cottler (1987) as a supplement to the World Health Organization’s Composite International Diagnostic Interview (Robins et al., 1988).

Interviewers possessed a bachelor’s or master’s degree in psychology or a related field, completed an extensive training program, and were blind to the clinical status of other family members. Coded interviews were reviewed by pairs of individuals with advanced clinical training (supervised by a PhD clinical psychologist), also blind to other family diagnoses. At study intake, assessments were lifetime. Follow-up assessments covered the period since the last assessment, and when aggregated with each other along with the intake assessment, provided an opportunity to obtain lifetime assessments at the time of the age 20 follow-up (Leckman, Sholomskas, Thompson, Belanger, & Weissman, 1982). Information regarding onset and offset of specific symptoms and episodes, crucial in a developmental study such as ours, was also coded. We have examined reliability of our consensus procedure by disorder, informant, subject age, and instrument. Interviews of approximately 600 participants were reviewed by two independent, blinded pairs and produced kappas >0.74, with kappas for all substance use disorders (alcohol, nicotine, cannabis, amphetamine abuse or dependence) exceeding 0.91 (Iacono et al., 1999).

The 11 symptoms shown in Table 1 assessed four abuse (listed first) and seven dependence criteria, each coded on a 2-point yes–no scale. Although our data were based on subjects who reported using cannabis at least once since the last assessment 4 years prior, administration of the diagnostic criteria was contingent on five or more reported uses since the last assessment. Subjects who reported using cannabis one through four times were coded as zero for each item. To correspond to the first **DSM-IV** substance abuse criterion, Item 1 is a manifestation of either “used often when doing something important” or “stayed away from school or missed appointments because of use.” Likewise, endorsement of Item 6, withdrawal, is a manifestation of either “felt sick when cutting down or stopped use” or “after not using cannabis, used to prevent sickness” to correspond with the **DSM-IV** substance dependence withdrawal criteria.

**Statistical Analyses**

We fitted LFA and LCA as well as FMMs directly to the raw ordinal data using the Mx software package (Neale et al., 2006). Raw ordinal data methods use both complete and incomplete data, thereby avoiding listwise deletion, which has the added advantage of increasing the accuracy of the item thresholds and improving the covariance estimates. In LFA, the observed items are specified as being caused by the latent factor with factor loadings \( f_i \),. Residual (nonfactor or specific) effects denoted as \( R_i \), are also included, and these generate additional variation in the observed scores. More than one common factor can be specified to allow for the possibility of more latent dimensions that can influence the observed items. Specifically, we used marginal maximum likelihood (MML) to model the cannabis diagnostic criteria and to improve computation efficiency. MML (Bock & Aitken, 1981) provides a fast means of calculating model parameters by integrating the latent factor distribution using 10-point quadrature (Neale,

<table>
<thead>
<tr>
<th>Item</th>
<th>Endorsements (%)</th>
<th>Item correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Used often when doing something . . . important/fail to fulfill obligations⁴</td>
<td>14.3</td>
<td>6.8</td>
</tr>
<tr>
<td>2. Hazardous use</td>
<td>44.9</td>
<td>19.0</td>
</tr>
<tr>
<td>3. Consequences: legal</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>4. Consequences: social</td>
<td>9.6</td>
<td>3.8</td>
</tr>
<tr>
<td>5. Need for larger amounts/doses (tolerance)</td>
<td>18.5</td>
<td>10.6</td>
</tr>
<tr>
<td>6. Withdrawal symptoms: feeling sick when cutting down/stopping⁴</td>
<td>10.7</td>
<td>4.6</td>
</tr>
<tr>
<td>7. Used more or longer than thought/planned</td>
<td>9.3</td>
<td>4.8</td>
</tr>
<tr>
<td>8. Loss of control: unable to stop/desire to stop, tried to cut down/stay using it</td>
<td>11.6</td>
<td>4.8</td>
</tr>
<tr>
<td>9. Spend time taking/using it, recovering from it, or doing whatever</td>
<td>16.2</td>
<td>8.6</td>
</tr>
<tr>
<td>10. Used instead of work/hobbies</td>
<td>5.8</td>
<td>2.4</td>
</tr>
<tr>
<td>11. Consequences: physical and psychological</td>
<td>11.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Note. Endorsements and item correlations based on 506 women and 366 men who reported using cannabis at least once within the past 4 years. Polychoric correlations for women appear above the diagonal; those for men appear below the diagonal.

⁴ Item endorsements represent the standard normal cumulative distribution based on marginal maximum likelihood (MML) threshold estimates. The estimated proportion of individuals who endorse an item when calculated under MML will differ from the sample statistic because the location of item’s threshold is sensitive to the correlation between diagnostic criteria. Based on the central limit theorem, polychoric correlations assume that underlying each ordinal response there is a continuously and normally distributed scale of liability, which is multifactorial, and that the joint distribution of this scale with liability scales underlying other ordinal items is bivariate normal. Manifestation of either “used often when doing something important” or “stayed away from school or missed appointments because of use” to correspond with **Diagnostic and Statistical Manual of Mental Disorders** (4th ed.) substance abuse criterion A1. Withdrawal as manifested by either “felt sick when cutting down or stopped use” or “after not using cannabis, used to prevent sickness” to correspond with the **Diagnostic and Statistical Manual of Mental Disorders** (4th ed.) substance dependence withdrawal criteria.
In MML, the likelihood function is calculated as the product of the conditional probability for each item at a series of abscissae along the latent trait distribution. This product of one-dimensional integrals can be used because, when conditioned on the factor, the items are independent. The Gaussian weighted sum of these conditional single-item responses approximates the overall likelihood for an observed vector of item scores.

The second approach for characterizing individual differences and item responses was LCA (Lazarsfeld, 1950). LCA assumes that any covariation between observed variables arises because the population consists of at least two subgroups, and these subgroups have different means for at least two of the measures. This situation is illustrated for two items in the scatter plots shown in Figure 1. Within each class, there is no correlation between Items X and Y; hence, the scatter plots of each group are approximately circular. With a large mean difference, as in Figure 1a, the distinction between the groups is quite obvious and clearly does not conform to a bivariate normal distribution. However, when the group mean differences are smaller (see Figure 1b), the difference between this plot and that of a single bivariate normal distribution (see Figure 1c) is much less obvious—both to the eye and to the statistical software designed to detect it.

The third approach, FMM, assumes that a population may consist of several latent classes, any or all of which may vary on one or more underlying continua. An illustrative scatter plot of two variables that might be generated by an FMM is shown in Figure 1d. Here, the two classes differ not only in their means, but also in their covariances. The class with the higher means on both Items X and Y shows marked positive correlation, whereas the observations within the other class are essentially uncorrelated. Clearly, with multivariate data, a very broad array of possible models exists within this framework. Latent trait models are special cases with one or more factors but only a single class. Latent class models have at least two classes but no factors.

Our current analytic approach assumes that members of twin pairs are independent. Failure to take account of statistical non-independence of the twins is not expected to change the parameter estimates but may alter their confidence intervals. Nonindependence of observations is rarely a problem when the group size is small; in the case of twin data, group size is at most two.

Model Comparisons

Latent factor, latent class, and factor mixture models were fitted to the ordinal data and compared using omnibus fit indices. In addition to the traditional Akaike information criterion (AIC; Akaike, 1987), we used the Bayesian information criterion (BIC) and sample-size adjusted BIC (sBIC; Schwarz, 1978) indices to identify the best-fitting models. Our decision to include the BIC and sBIC was based on simulations (Nylund, Asparouhov, & Muthén, 2007) that have shown how the BIC and sBIC (Schwarz, 1978) can outperform the AIC in complex structures in which symptoms have different endorsement probabilities for more than one latent class. Although the power to choose between competing latent class and factor models can vary according to the item response format, simulations based on data from the Virginia Twin Registry have shown that it is possible to choose between these models with categorical data (Lubke & Neale, 2008). With regard to the number of classes and factors, we have examined one, two, and three classes and one, two, and three factors in the factor mixture model.
to the power to choose between competing categorical models, this
depends on class separation and within-class sample size (Lubke &
have shown that sample sizes as small as 75 subjects within a class
can lead to incorrect model selection even if the distance between
classes is so small that heterogeneity would be hard to detect.
Although bootstrap likelihood ratio tests may provide a better
discrimination between models with different numbers of latent
classes (Nylund et al., 2007), this approach was not adopted
because of the heavy computational burden. Instead, the best-
fitting model was the one in which two of three or all three indices
showed the best fit.

For tests of sex differences, we conducted a hierarchical se-
quence of model comparisons using the definition variable option
in Mx (Neale, Boker, Xie, & Maes, 2006). In addition to chi-
square tests, the AIC, BIC, and sBIC were again used to determine
the best-fitting sex effects models. First, depending on the best
solution (LFA, LCA, or FMM), a baseline model representing
complete sex measurement invariance (no differences) was fitted
to the combined male and female data. Next, models with a sex
effect on (a) the latent factor mean (or categorical latent factor
variable), (b) factor variance, (c) factor loadings, and (d) item
thresholds (cutoff points) were then compared with the base. If
comparisons between the baseline and sex effect models produced
a significant misfit (significant misfit equals improved fit), then
sex effects were deemed significant. Where there was no signifi-
cant misfit, sex differences were deemed nonsignificant.

The lowest, most negative $-2 \times \log$ likelihood value is indica-
tive of the best-fitting model, whereas the lowest, most negative
AIC, BIC, and sBIC values are indicative of model parsimony.
This distinction is important because in maximum likelihood es-
imation log likelihoods can be improved by simply “overfitting”
or increasing the number of parameters. Indices of parsimony
penalize models with increasing numbers of parameters, thereby
providing an index of each model’s efficiency in terms of explain-
ing observed data patterns. Our decision to include the BIC and
sBIC was also based on simulations (Nylund et al., 2007) that have
shown how the BIC and sBIC (Schwarz, 1978) outperform the
AIC in complex structures in which symptoms have different
dendorsement probabilities for more than one latent class.

Results

Item Endorsements, Phenotypic Correlations, and
Eigenvalues

Table 1 includes the item endorsements for each diagnostic
criterion from which can be estimated the number of recent users
who did not endorse a diagnostic symptom. For men and women
alike, the most and least frequently endorsed items were “using
cannabis in hazardous situations” and “legal problems” arising
from use, respectively. Among men, the percentages of recent
users who endorsed none, one, two or three, and four or more
diagnostic criteria were 49.7%, 18.0%, 14.8%, and 17.5%, respec-
tively. Among women, the percentages of recent users who en-
 dorsed none, one, two or three, and four or more diagnostic criteria
were 74.5%, 10.6%, 7.6%, and 7.4%, respectively.

Polychoric correlations among the items also appear in Table 1.
Except for Item 8, “tried to cut down...stop using it,” nearly all
correlations ranged from moderate to high. The first three eigen-
values for men and women, respectively, were 8.11, 0.91, 0.58 and
8.18, 3.28, 0.42, suggesting either a one- or two-factor dimensional
structure to the data.

Model Comparisons

As shown in Table 2, for men, a one-factor two-class FMM
provided a good fit under the AIC. However, across AIC, BIC, and

<p>| Table 2: Comparison of Empirical, Latent Factor Analysis (LFA), Latent Class Analysis (LCA), and Factor Mixture Models (FMMs) |</p>
<table>
<thead>
<tr>
<th>Analysis</th>
<th>–2LL</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>sBIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFA–1 dimension</td>
<td>2190.71</td>
<td>3975</td>
<td>−5759.29</td>
<td>−10631.13</td>
<td>−4330.58</td>
</tr>
<tr>
<td>LFA–2 dimensions</td>
<td>2173.59</td>
<td>3964</td>
<td>−5754.41</td>
<td>−10612.23</td>
<td>−4324.12</td>
</tr>
<tr>
<td>LFA–2 correlated dimensions</td>
<td>2169.76</td>
<td>3963</td>
<td>−5756.24</td>
<td>−10611.19</td>
<td>−4324.68</td>
</tr>
<tr>
<td>LFA–3 dimensions</td>
<td>2163.34</td>
<td>3953</td>
<td>−5742.66</td>
<td>−10584.89</td>
<td>−4314.23</td>
</tr>
<tr>
<td>LCA–1 class</td>
<td>2904.99</td>
<td>3987</td>
<td>−5069.01</td>
<td>−10314.41</td>
<td>−3989.82</td>
</tr>
<tr>
<td>LCA–2 classes</td>
<td>2240.87</td>
<td>3974</td>
<td>−5707.14</td>
<td>−10608.10</td>
<td>−4304.14</td>
</tr>
<tr>
<td>LCA–3 classes</td>
<td>2177.00</td>
<td>3962</td>
<td>−5747.00</td>
<td>−10604.62</td>
<td>−4319.69</td>
</tr>
<tr>
<td>FMM–1 factor, 2 classes</td>
<td>2158.71</td>
<td>3952</td>
<td>−5745.29</td>
<td>−10584.25</td>
<td>−4315.18</td>
</tr>
<tr>
<td>FMM–1 factor, 3 classes</td>
<td>2112.80</td>
<td>3929</td>
<td>−5745.20</td>
<td>−10539.32</td>
<td>−4306.74</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFA–1 dimension</td>
<td>1712.69</td>
<td>5497</td>
<td>−9281.31</td>
<td>−16235.48</td>
<td>−7511.52</td>
</tr>
<tr>
<td>LFA–2 dimensions</td>
<td>1692.32</td>
<td>5486</td>
<td>−9279.68</td>
<td>−16211.46</td>
<td>−7504.97</td>
</tr>
<tr>
<td>LFA–2 correlated dimensions</td>
<td>1713.97</td>
<td>5485</td>
<td>−9256.03</td>
<td>−16197.53</td>
<td>−7492.62</td>
</tr>
<tr>
<td>LFA–3 dimensions</td>
<td>1804.71</td>
<td>5475</td>
<td>−9055.29</td>
<td>−16076.06</td>
<td>−7387.03</td>
</tr>
<tr>
<td>LCA–1 class</td>
<td>2427.05</td>
<td>5509</td>
<td>−8590.95</td>
<td>−15913.61</td>
<td>−7172.61</td>
</tr>
<tr>
<td>LCA–2 classes</td>
<td>1730.88</td>
<td>5496</td>
<td>−9241.12</td>
<td>−16213.27</td>
<td>−7490.91</td>
</tr>
<tr>
<td>LCA–3 classes</td>
<td>1690.96</td>
<td>5484</td>
<td>−9277.04</td>
<td>−16205.92</td>
<td>−7502.60</td>
</tr>
<tr>
<td>FMM–1 factor, 2 classes</td>
<td>1672.12</td>
<td>5474</td>
<td>−9275.88</td>
<td>−16184.25</td>
<td>−7496.80</td>
</tr>
<tr>
<td>FMM–1 factor, 3 classes</td>
<td>1658.92</td>
<td>5451</td>
<td>−9243.08</td>
<td>−16119.34</td>
<td>−7468.39</td>
</tr>
</tbody>
</table>

Note. –2LL = $-2 \times \log$ likelihood; AIC = Akaike information criterion; BIC = Bayesian information criterion; sBIC = sample-size adjusted BIC. LFA modeling also included one two-factor oblique solution. Under each fit index, the two best-fitting models appear in bold type.
sBIC indices, the single-factor LFA solution consistently outperformed all other models in terms of parsimony. For women, none of the FMMs provided a good fit to the data. The uncorrelated two-factor solution performed well under the AIC and sBIC. However, again it was the single-factor LFA solution that consistently outperformed all other models across all three parsimony indices. Factor loadings for the best-fitting single-factor model appear in Table 3. Apart from the legal consequences item, all of the loadings on the single-factor solution Cannabis Use Disorder were high for men and women.

**Modeling Sex Effects**

Tests for measurement invariance to determine the nature and location of sex differences, if present, appear in Table 4. The single latent factor model (Model 1) was used as the baseline comparison. This was fitted to the combined male and female data without any sex effects included. Models 2, 3, 4, 5, and 6 represent sex effects on the latent factor variance, latent factor mean, latent factor loadings and thresholds, latent factor loadings and item thresholds, respectively. Models 3, 4, and 6 significantly reduced the overall “misfit.” The significant misfit seen in Model 4 appears not to have been driven by sex effects on the factor loadings (Model 5) but rather by sex differences on the item thresholds (Model 6). Subsequent univariate analyses revealed that misfit in Model 6 was driven by one item: hazardous use. Although Models 4 and 6 each provided a marginally better fit to the data compared with Model 3, in terms of parsimony, Model 3 with fewer parameters consistently outperformed Models 4 and 6 in terms of larger negative AIC, BIC, and sBIC values. The best-fitting model included a single sex effect on the latent mean. This model predicts that the latent liability to CUD is significantly lower among women.

**Discussion**

Omnibus comparisons between latent factor, latent class, and factor mixture models revealed that a single latent factor best explained the patterns of association between the diagnostic criteria for cannabis abuse, dependence, and symptoms of withdrawal. Despite their theoretical appeal, we found little consistent evidence to support population heterogeneity among cannabis users; neither latent class nor hybrid factor mixture models improved model fit. Our sample included subjects, who despite reporting cannabis use, endorsed no symptoms. Therefore, the extent to which these individuals represent a distinct mixture or class of asymptomatic individuals was not borne out in terms of a better fitting latent class or factor mixture model in which one might expect to retrieve an abstinent or very low use class. Instead, our findings are commensurate with a growing number of reports with similar results (Baillie & Teesson, 2010; Feingold & Rounsaville, 1995a, 1995b; Gillespie et al., 2007; Hartman et al., 2008; Langenbucher et al., 2004; Lynskey & Agrawal, 2007; Nelson et al., 1999; Teesson et al., 2002) as well as with the proposed changes in the forthcoming *DSM–V* with regard to substance-related disorders.

This study and the comparisons between competing hypotheses illustrate a more general but important point: The same set of symptoms can be used to determine the defining properties of criterion in terms of dimensional continuous scales of liability, unobserved class membership structures, or hybrid classifications, all of which would not be possible if the symptoms had been aggregated into an affected versus unaffected binary diagnostic variable.

We know of one other published report comparing the fit of latent factor, latent class, and factor mixture models to cannabis data. Baillie and Teesson (2010) also found no improvement in fit for latent class or factor mixture models when compared with a single latent factor model for *DSM–IV* cannabis criteria measured on large population-based Australia sample. This emerging consensus supports the proposed removal of the abuse–dependence distinction, as well as the reintroduction of cannabis withdrawal symptoms, in the forthcoming *DSM–V*.

Contrary to the previous uncertainty surrounding the clinical significance of withdrawal symptoms (American Psychiatric Association, 1994), there is now enough evidence from laboratory and clinical studies (Budney, Hughes, Moore, & Vandrey, 2004; Budney, Moore, Vandrey, & Hughes, 2003; Lichtman & Martin, 2002; Lichtman et al., 2005) showing that withdrawal diagnostic criteria are empirically valid. Although we did not assess symptoms unique to cannabis withdrawal and although inclusion of the withdrawal items in our analyses was not contingent on symptoms causing significant occupational or social impairment (as proposed under *DSM–V*), we nevertheless found that the withdrawal item loaded convincingly onto the CUD factor. In withdrawal-gate models, the symptoms of withdrawal are central to diagnoses of dependence, and as such their factor loadings are expected to load among the highest (Langenbucher et al., 2000). However, the size of the withdrawal factor loadings for men and women suggests that

<table>
<thead>
<tr>
<th>Diagnostic criterion</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Used often when doing something . . . important/fail to fulfill obligations</td>
<td>.82</td>
<td>.90</td>
</tr>
<tr>
<td>2. Hazardous use</td>
<td>.83</td>
<td>.83</td>
</tr>
<tr>
<td>3. Consequences: legal</td>
<td>.65</td>
<td>.57</td>
</tr>
<tr>
<td>4. Consequences: social</td>
<td>.76</td>
<td>.84</td>
</tr>
<tr>
<td>5. Need for larger amounts/doses (tolerance)</td>
<td>.85</td>
<td>.92</td>
</tr>
<tr>
<td>6. Withdrawal symptoms: feeling sick when cutting down/stoping</td>
<td>.78</td>
<td>.87</td>
</tr>
<tr>
<td>7. Used more or longer than thought/planned</td>
<td>.82</td>
<td>.75</td>
</tr>
<tr>
<td>8. Loss of control: unable to stop/desire to stop, tried to cut down/using it</td>
<td>.75</td>
<td>.89</td>
</tr>
<tr>
<td>9. Spend time taking/using it, recovering from it, or doing whatever</td>
<td>.87</td>
<td>.88</td>
</tr>
<tr>
<td>10. Used instead of work/hobbies</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>11. Consequences: physical and psychological</td>
<td>.77</td>
<td>.77</td>
</tr>
</tbody>
</table>
Tests of Measurement Invariance Across Sex for the Minnesota Cannabis Abuse and Dependence Items

<table>
<thead>
<tr>
<th>Model</th>
<th>$-2\text{LL}$</th>
<th>df</th>
<th>$\Delta-2\text{LL}$</th>
<th>$\Delta df$</th>
<th>$p$</th>
<th>Parameter</th>
<th>AIC</th>
<th>BIC</th>
<th>sBIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Complete invariance</td>
<td>3988.42</td>
<td>9495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$-15001.58$</td>
<td>$-30128.29$</td>
<td>$-15051.40$</td>
</tr>
<tr>
<td>2. Sex on LF variance</td>
<td>3985.00</td>
<td>9494</td>
<td>$3.42$</td>
<td>$1$</td>
<td>$0.06$</td>
<td></td>
<td>$-15003.00$</td>
<td>$-30126.61$</td>
<td>$-15051.32$</td>
</tr>
<tr>
<td>3. Sex on LF mean</td>
<td>3940.06</td>
<td>9494</td>
<td>$48.36$</td>
<td>$1$</td>
<td>***</td>
<td>$23$</td>
<td>$-15047.94$</td>
<td>$-30149.08$</td>
<td>$-15073.79$</td>
</tr>
<tr>
<td>4. Sex on LF loadings and thresholds</td>
<td>3903.40</td>
<td>9473</td>
<td>$85.02$</td>
<td>$11$</td>
<td>***</td>
<td>$44$</td>
<td>$-15042.60$</td>
<td>$-30096.37$</td>
<td>$15054.42$</td>
</tr>
<tr>
<td>5. Sex on LF loadings</td>
<td>3974.60</td>
<td>9484</td>
<td>$13.82$</td>
<td>$11$</td>
<td>$0.24$</td>
<td>$33$</td>
<td>$-14993.41$</td>
<td>$-30097.98$</td>
<td>$15038.57$</td>
</tr>
<tr>
<td>6. Sex on item thresholds</td>
<td>3917.29</td>
<td>9484</td>
<td>$71.13$</td>
<td>$11$</td>
<td>***</td>
<td>$33$</td>
<td>$-15050.71$</td>
<td>$-30126.64$</td>
<td>$15067.22$</td>
</tr>
</tbody>
</table>

Note. $-2\text{LL} = -2 \times \log$ likelihood; $\text{AIC} = \text{Akaike information criterion}; \text{BIC} = \text{Bayesian information criterion}; \text{sBIC} = \text{sample-size adjusted BIC}; \text{LF} = \text{latent factor}, \text{p} < 0.001, \Delta-2\text{LL} = \text{change in} -2\text{LL}, \text{which is asymptotically distributed as a chi-square with the degrees of freedom equaling the difference between the fully saturated model (Model 1) with the each of nested submodels (Model 2–5). The baseline model (1), which included no sex effects, was compared with models with covariate sex effects on the (2) latent factor mean, (3) latent factor variance, (4) latent factor loadings, and (5) all item thresholds. For each fit index, best fitting model is in bold type. *** $p < 0.001$.

this item discriminates no better or worse than other criteria. Overall, our results are thus supportive of findings identifying cannabis withdrawal symptoms while at the same suggesting that they form part of a larger, quantitative CUD factor comprising symptomatic withdrawal, abuse, and dependence criteria.

Compared with women, men in our sample were almost twice as likely to endorse the withdrawal, abuse, and dependence criterion. Although this pattern is well established (Blanco et al., 2008; Degenhardt, Cheng, & Anthony, 2007), this is the first empirical report to test the location of sex differences among cannabis users. Although Model 3 with its single sex effect on the latent factor mean consistently outperformed other models, one item in Model 6, “hazardous use,” appears to be driving measurement noninvariance; men were more likely to have endorsed using cannabis in situations that increased their chances of getting hurt, for example, driving; using guns, knives, or machinery; crossing traffic; climbing; or swimming.

Several reports have speculated that a combination of societal and cultural factors, along with differences in biological responses and patterns of comorbidity with other psychiatric disorders, are responsible for the difference (Brady & Randall, 1999; Greenfield, Back, Lawson, & Brady, 2010; Tuchman, 2010). Concerning specific risk factors, one report (Agrawal, Jacobson, Prescott, & Kendler, 2004) argued that the stronger association between the personality dimension of novelty seeking and cannabis use in males is attributable to differences in the magnitude of the same genetic and environmental factors influencing personality and cannabis use across sex. Four twin studies have shown that sex differences in the genetic and environmental risks for cannabis use, abuse, and dependence also are more likely to be quantitative rather than qualitative (Agrawal, Neale, Jacobson, Prescott, & Kendler, 2005; Agrawal, Neale, Prescott, & Kendler, 2004a; Kendler et al., 2000; Kendler & Prescott, 1998). Our tests not only corroborate this conclusion but also demonstrate that the differences in item endorsements are better explained by a single mean difference on the latent factor, indicating that males have a greater overall liability toward expressing CUD.

Our best-fitting model predicts that covariation between the abuse and dependence items is better explained by a single latent factor that is likely to be a combination of both genetic and environmental risks. This conclusion is supported by genetically informative designs that have explored the genetic and environmental etiology of cannabis use, abuse, and dependence (Gillespie, Kendler, & Neale, 2011; Gillespie et al., 2009; Kendler et al., 2000; Kendler & Prescott, 1998). Nevertheless, our model fitting was by no means exhaustive. Item response theory (IRT; Lord, 1952, 1980), multiple indicators, multiple causes (Finch, 2005), Mantel–Hansel (Holland & Thayer, 1988), logistic regression (Swaminathan & Roger, 1990), and the likelihood ratio test (Thissen, Steinberg, & Wainer, 1993), among other methods, represent various means for detecting differential item functioning (DIF). As discussed by Takane and de Leeuw (1987), there is a direct equivalence of the normal ogive item response model and factor analysis model for binary data that we have fitted. Therefore, in terms of IRT, item difficulty parameters correspond to our item thresholds (which can be derived from item endorsements in Table 1), and the discrimination parameters correspond to the observed factor loadings in Table 3. Examining these properties and detecting DIF can improve scale construction through the elimination of items that are redundant or that fail to discriminate adequately. For example, the lower factor loadings, especially among women, for legal consequences (i.e., “Did your use of cannabis lead to trouble with police?”) suggests that this infrequently endorsed item is a poor discriminator. Indeed, based on results from a number of reports, some including formal tests of DIF (Gillespie et al., 2007; Hartman et al., 2008; Hasin, Paykin, Meydan, & Grant, 2000; Langenbucher et al., 2004; Lynskey & Agrawal, 2007; Schuckit et al., 1999; Teesson et al., 2002), the same conclusion can be made, thereby empirically justifying the removal of legal consequences from the proposed DSM–V.

Limitations

Our findings must be interpreted in the context of three important limitations. First, although the sample is representative of the state of Minnesota when subjects were born, our data were restricted to Caucasian male and female twins and cannot be extrapolated to ethnic minorities. Nevertheless, our estimates of use including the prevalence of item endorsements are very similar to those found in other large, population-based twin samples (Gillespie et al., 2007), which in turn are typical of the general population of North American Caucasians in rates of psychopathologic conditions, including illicit substance use (Kendler et al., 2000). Second, the mean age at assessment was 5 years past the...
mean age of cannabis initiation. Although this is an ideal assessment age for studying cannabis use, abuse, and dependence because most twins will have passed through the peak risk period while still being young enough to accurately respond to diagnostic criteria, the results cannot preclude the possibility of different latent factor or class structures at older ages performing better. However, the fact that a unidimensional factor structure is more often found in large population-based studies with varying sample ages lends support to these findings. Third, the models fitted were not exhaustive. Our results imply that the linear coefficients (i.e., intercepts and factor loadings) that link the observed items to the latent factor are fixed or common for all subjects, which may be too restrictive if subjects habitually respond to questionnaires idiosyncratically (Maydeu-Olivares & Coffman, 2006). Therefore, to test for idiosyncratic reporting, we incorporated into the intercept of the model an additional latent “random intercept factor” (RIF) described by Maydeu-Olivares and Coffman (2006). Normally, factor loadings from the RIF are constrained to 1 and the RIF variance estimated. However, because we were modeling ordinal data and latent factor variance is by default 1, we constrained the factor loadings for the random intercept factor to be equal. For men, the addition of the RIF did not yield a better fit (AIC = –5759.21, BIC = –10634.14, and sBIC = –4330.17), whereas for women, the RIF model provided a marginally better fit across all three indices (AIC = –9287.31, BIC = –16236.37, and sBIC = –7514.00). Although this suggests that women may have an idiosyncratic response pattern, all factor loadings remained significant, with an average change of only .02 across items.

Conclusion

Results from an omnibus comparison between theoretically plausible and alternative measurement models revealed that the diagnostic criteria for cannabis abuse, dependence, and symptoms of withdrawal were best explained by a single latent factor we labeled Cannabis Use Disorder. This was true for both men and women. The almost twofold sex difference in item endorsement was best explained by a single mean difference on the latent factor. Men appear to have a greater latent liability toward expressing CUD. Our next goal will be to explore the genetic and environmental etiology of this CUD phenotype using biometrical genetic analyses. Derivation of the best-fitting genetic and environmental model based on CUD criteria will enable estimation of individual latent trait scores. Therefore, given genetically informative data such as ours, it will be possible to identify maximally heritable factor scores, which in turn will increase the power to detect quantitative trait loci signals for CUD as part of anticipated genome-wide association analyses.

References


Hollingshead, A. B. (1975). Four factor index of social status. Unpublished manuscript, Yale University, New Haven, CT.


of the validity and significance of cannabis withdrawal syndrome. Hand-

www.psychometrika.org/journal/online/MN07.pdf

Lord, F. M. (1980). Applications of item response theory to practical

Specific Parameters. Structural Equation Modeling, 14, 26–47.

ioral Research, 41, 499–532.

Lubke, G., & Neale, M. C. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of

assessments of illicit drug abuse and dependence: Results from the National Epidemiologic Survey on Alcohol and Related Conditions
(NESARC). Psychological Medicine, 37, 1345–1355.

Lynskey, M. T., Nelson, E. C., Neuman, R. J., Bucholz, K. K., Madden,
Australian twins. Twin Research and Human Genetics, 8, 574–584.


NY: Wiley.

Muller, S., & Gmel, G. (2002). Changes in the age of onset of cannabis use:
Results of the 2nd Swiss Health Survey 1997. Sozial- Und Preventiv-
medizin, 47, 14–23.

Muthén, B. (2006). Should substance use disorders be considered as
categorical or dimensional? Addiction, 101(Suppl. 1), 6–16.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and
variable-centered analyses: Growth mixture modeling with latent trajec-
tory classes. Alcoholism: Clinical and Experimental Research, 24, 882–
891.


tment of Psychiatry.

Factor structures for DSM–IV substance disorder criteria endorsed by
alcohol, cannabis, cocaine and opiate users: Results from the WHO Reliability and Validity Study. Addiction, 94, 843–855.

Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the
number of classes in latent class analysis and growth mixture modeling.

(1998). Genetic “clocks” and “soft” events: A twin model for pubertal
development and other recalled sequences of developmental milestones,
transitions, or ages at onset. Behavior Genetics, 28, 243–253.

tional diagnostic interview: Expanded substance abuse module. St.
Louis, MO: Washington University Department of Psychiatry.

Robins, L. N., Wing, J., Wittchen, H. U., Helzer, J. E., Babor, T. F., Burke,