

**Comparing Classification vs. Continuum Models of the  
Structure of Substance Dependence and Abuse**

**A THESIS**

**SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF MINNESOTA**

**BY**

**Scott Ian Vrieze**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
Master of Arts**

**December, 2009**

© Scott Ian Vrieze 2009  
ALL RIGHTS RESERVED

# Comparing Classification vs. Continuum Models of the Structure of Substance Dependence and Abuse

by Scott Ian Vrieze

## ABSTRACT

Substance use disorders are classified as categorical disorders by prominent nosologies [1]. A bevy of structural equation models have suggested dimensional solutions to drug dependence and abuse criteria. However, it is well known that factor models can fit categorical structures, and class models can fit dimensional structures. Recent research has thus compared relative fits of both latent class and trait models, and in some cases mixtures of latent traits. Results have been inconsistent, in large part due to the level of analysis. We attend to both problems by fitting models in a large sample with high base rates of alcohol, marijuana, cocaine, and stimulant disorders, allowing us to fit complex models both at fine and coarse levels of analysis (e.g., only alcohol items versus all items from a variety of drug classes). In general, dependence and abuse items from different drug classes can be modeled with drug-specific factors, one per drug. When more complex models are fit only to alcohol items, the best fitting model is a mixture of latent traits that maps closely onto DSM-IV-TR [1] nosology.

# Contents

<b>Abstract</b>	<b>i</b>
<b>List of Tables</b>	<b>iii</b>
<b>List of Figures</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Method</b>	<b>16</b>
<b>3 Results</b>	<b>23</b>
<b>4 Discussion</b>	<b>29</b>
<b>References</b>	<b>37</b>

# List of Tables

2.1	Items Retained after Removal of Items with Correlations $> .95$ . . . . .	21
3.1	Analysis of Dependence and Abuse Criteria. . . . .	24
3.2	Parameters for the Four-Trait Latent Trait Model. . . . .	25

# List of Figures

4.1	Alcohol Item Response Probability Given Class Membership. . . . .	31
4.2	Item Characteristic Curves. . . . .	34

# Chapter 1

## Introduction

This article contributes to the broader discussion of mental disease structure. A long-standing issue in personality and psychopathology concerns how best to model mental disease manifestations and processes. Personality psychology has traditionally made use of factor analysis and personality was taken by many to consist of traits, each of which was expressed as a continuously varying latent factor. Of course, appropriately conducted factor analyses can return well-behaving factor structures, even if the personality construct is strongly taxonic [2] [3].

Tradition in psychopathology research is that mental diseases are categorical. The DSM-IV encyclopedic list of mental illness is obviously categorical, but the supervising committee explicitly states on page xxxi that the categories do not necessarily represent naturally-occurring categories (i.e., taxa) [1]. Regardless, the DSM-IV was created in a medical *Zeitgeist* that used non-overlapping categories. Medical taxa have not persevered as categories for administrative and clinical convenience, or because categorical disease structures are merely traditional. Taxa such as the common cold, measles, polio, etc., exist as categories because there is impressive evidence that they are naturally distinct disease entities. The apparently verisimilar taxonic theories in medicine are

predicated on well-established theories in biology and biochemistry, such that pathophysiology is known, and category-based prognoses can sometimes be made with high accuracy.

Understanding the nature of a mental illness, whether taxonic or continuous, is central to treatment decisions, methods, and prognoses. If a disease is strongly taxonic, then there may exist a discrete disease mechanism or threshold effect. Once the pathophysiology is characterized, treatment methods can be tailored specifically for the taxon. On the other hand, if a disease is continuous, then pathophysiology is likely continuous itself or multiply determined by several to many discrete causes. A full understanding the etiologies of alcoholism and other substance use problems may be far off, but current research can assist theoreticians in postulating disease theories, which are then submitted to falsifying tests.

Psychology, for many reasons (see [4] for a list of 20), has not progressed like medicine, and our adoption of categorical disease models emanates more from tradition than from defensible theory, or from strong research findings. The present article is an attempt to contribute to the data-analytic evidence for or against the existence of taxa within psychoactive substance use. These theories are cast in the form of statistical models, models are compared with one another, and the model that best predicts certain features in the data is considered increasingly corroborated.

Evidence for taxa vs. continua will help guide and substantiate etiological research. Assume a disease is clearly taxonic at the syndromal level. Then a putative etiological factor that happens to be irreconcilable with the syndrome's taxonic nature (e.g., an additive, polygenic genetic etiology) should be treated with skepticism. On the other hand, the additive genetic etiology theory might be revised to include a threshold effect that translates the additive polygenic cause into the taxonic syndrome. Example: Huntington's chorea was originally considered a two-allele, autosomal dominant gene



disease, but we have learned that what counts is the number of trinucleotide (CAG) repeats at the Huntington locus, with a quite sharp threshold effect at about 28.

The basic approach is: 1) Entertain two or more competing theories about the causal nature and/or phenotypic manifestations of the subject matter in question—here we consider maladaptive alcohol, marijuana, cocaine, and stimulant use. The theories are admittedly simple, given the likely complex nature of substance use and addiction. However, we target only a single characteristic of substance use: whether it is best considered a taxonic or continuous syndrome. The term “syndrome” refers to any of several data structures. We distinguish two for present purposes: (a) a continuous or quasi-continuous structure such that all indicator variables assess a single unitary construct (e.g., latent factor, item response theory model latent variable); and (b) a taxonic structure such that all indicator variables intercorrelate in the mixture of taxon and complement class due to the very fact of admixture.

2) Write statistical models that describe, as clearly as possible, each of the competing theories. Here we use three different, but related, models: a latent trait model, latent class models, and a mixture of traits model.

Latent trait models are also known as factor models or item response theory (IRT) models. They posit continuous latent variables (called *traits* in personality), and use individual difference variations on latent traits to explain covariance among manifest, categorical variables. Usually, such models assume the observed joint probability density function (p.d.f.) of the indicator variables within latent classes is a product of the individual indicator variable marginal p.d.f.’s (i.e., local independence). Latent trait models are very similar to confirmatory factor models, except in the latter, continuous latent variables are used to explain covariance among continuous indicator variables. If alcoholism is continuous, for example, it means that some degree (possibly zero) of an alcoholism-related syndrome is possessed by every individual in society. Some

individuals may have low alcoholism (e.g., teetotalers), some may have a moderate degree of alcoholism (e.g., problem drinkers), and some may be severely alcoholic<sup>1</sup>.

Latent class models, on the other hand, posit discrete latent variables (called *classes*), and use class admixture to explain covariance among manifest variables. For example, if there exists a latent class (i.e., a *taxon*, or instead a *complement class*) of alcoholics, then individuals belonging to their class are not ordered by disease severity in a way that affects the p.d.f.s of the indicator variables. Instead, class members are assumed to be homogeneous with respect to the indicator variable level probability values. The latent class model can represent the theory that alcoholism is a taxonic disease, so that individuals fall in either the taxon or complement class. Alcoholics would then be a category (or set of categories) of individuals afflicted with the disease of alcoholism, as opposed to the extreme end of a continuous trait of drinking behavior and its consequences.

More recently, a newer model has emerged called variously *mixtures of factor analyzers* [5] and *factor mixture models* [6] [7]. We are here dealing specifically with *mixture models of latent traits*, a more descriptive term.

A general mixture model of latent traits has features of both latent trait and latent class models. A simple data example: fitting a single latent trait model to disease symptoms of an admixture of males and females can lead to corroborating false models, as the true disease models may differ strikingly between males and females for unknown reasons (e.g., genetic differences, differences in psychosocial stressors, etc.). Such inappropriate model fitting can result in misleading conclusions about the overall disease structure; see [8] for interesting examples. Clearly, when fitting factor or item response theory (IRT) models in two distinct populations it is best to model each population separately, and only treat the populations as parts of a single latent class if closely

---

<sup>1</sup> Of course, the latent trait of alcoholism is not necessarily what laity would call “alcoholism”. Rather, it is a statistically-identified construct that, face validly at least, appears to be usefully related to what the common person might call alcoholism.

similarly models fit best in each class.

Mixtures of traits run along the same theme except the populations (e.g., men and women) are not known, and must be modeled. Hence, the mixture of traits (factor mixture model) posits that there are several taxa, each of which is composed of a latent trait. There are thus two or more categories of people (e.g., social drinkers and alcoholics), and individuals in each class can be ordered along some trait such as disease severity. For example, according to this model and as seemingly observed in society, some individuals with alcoholism drink more than others and suffer more negative consequences from it, despite the fact that they are all alcoholic. Similarly, some normals drink more than others, despite the fact that none of them suffer from alcoholism.

3) The third step is to adjudicate between the models, taking the best-fitting model to be the most verisimilar one. We discuss below the different ways in which model selection can be accomplished. Each method has benefits and drawbacks, and the different methods can give discrepant answers—making the choice of methods a controversial one.

Fitting trait and class models to psychopathology indicators is now becoming increasingly common. For example, Krueger and colleagues [9] fit a unidimensional IRT model to a large number of alcohol-problem symptoms and signs in a community-based sample of middle-aged men. They concluded that the model fit well enough to be useful. However, no fit statistics were given, and the model was not compared to other, competing models. Proudfoot et al. [10] took Krueger et al.'s results one step further, and compared one- and two-trait models in a community sample of Australians. The single-trait model provided the best fit. Here two competing models were compared, but the study failed to address the possibility of class structure in alcoholism. Other researchers have conducted similar studies of alcohol use, all of which have found either

single or two-dimensional trait models [11] [12] [13] [14] [15] [16].

Some have argued and provided evidence that *all* chemical dependence and abuse are dimensional, as opposed to categorical. Specifically, those investigations propose that maladaptive drug use (not just alcohol use) is an expression of a single quasi-continuous externalizing liability. The theory is often described as “dimensional” in the literature, because the latent liability of externalizing is mathematized as a continuous latent factor, and subjects can have latent scores anywhere within its range [17]. Consistent with previous authors, we refer to this idea as the *externalizing spectrum* theory of externalizing psychopathology. According to the theory, the development of externalizing phenotype (acting-out, antisocial, out-of-control, high impulsivity) largely depends on one’s position along the externalizing continuum. A particular phenotypic manifestation (e.g., cocaine addiction) of this externalizing liability depends on other influences, either genetic or environmental. For example, an individual with high externalizing who has access only to cocaine might be more prone to develop cocaine problems, whereas if he had access only to alcohol he might develop alcohol problems instead. An individual with very little externalizing would be unlikely to develop either, regardless of his exposure to various substances.

Rigorous evidence for the dimensional nature of the externalizing spectrum theory has been published. The present research complements that reported in recent articles [17] [18] [6] [19]. We examine each article in some detail, as they illustrate the current approach.

Krueger and Colleagues [17] examined the latent structure of externalizing psychopathology in the parents of a large community twin sample. Latent class and latent trait models were fit to diagnoses of adult antisocial behavior, conduct disorder, alcohol dependence, marijuana dependence, and a meta-category of “drug dependence.” If a subject met criteria for heroin, amphetamine, cocaine, inhalant, opioid, PCP, or

sedative dependence, they were considered to have drug dependence.

Two-, three-, four-, and five-class latent class models were fit to the diagnostic sign data. The fits of each model were compared to the fit of latent trait models, most notably a normally-distributed single latent trait model. Parameter estimation was accomplished according to maximum likelihood; the best-fitting model was taken to have the greatest verisimilitude. It is well known that maximum likelihood produces a biased estimate of fit [20], and bias increases with increasing number of model parameters. Corrections typically take the following form:

$$\log(L) - f(M),$$

where  $\log(L)$  is the logarithm of the likelihood and  $f(M)$  is some function of the model complexity. Krueger et al. [17] used the Bayesian Information Criterion (BIC [21]), which modifies  $\log(L)$  as follows:

$$BIC = -2\log(L) + \log(N)K, \tag{1.1}$$

where  $K$  is the number of model parameters, and  $N$  is the sample size. Another commonly used correction is the Akaike Information Criterion (AIC [22]). The AIC modifies the  $\log(L)$  as:

$$AIC = -2\log(L) + 2K. \tag{1.2}$$

Both corrections multiply  $\log(L)$  by a constant,  $-2$ , so that lower values of the BIC or the AIC indicate better fit.

Krueger and Colleagues [17] found that the unidimensional, normally distributed, latent trait model possessed the best relative fit according to the BIC. The AIC selected as the best model a 4-class model.

The same group of investigators have replicated in [23] results in [17] in a nationally representative dataset from the National Epidemiologic Survey on Alcohol and Related

Conditions (NESARC). They fitted latent class and latent trait models separately to lifetime and current diagnoses of antisocial personality disorder, nicotine dependence, alcohol dependence, marijuana dependence, cocaine dependence, and other substance dependence. Again, the BIC selected a unidimensional latent trait model. This result held whether lifetime diagnoses or current diagnoses were considered; it was the best model whether the entire sample, or only male or female subsamples, were considered. AIC results were again different. In the overall sample a 6-class latent class model was indicated. In the female sample the AIC selected the 5-class latent class model; for the males a 5-variable discrete latent trait model (essentially an ordered class model) was selected. The AIC-BIC discrepancies are difficult to reconcile.

Clearly, the choice of fit statistic is central to selecting the most corroborated theory. The BIC was used in both [17] and [18]. The use of the AIC and the BIC in distinguishing between categorical and dimensional theories of psychopathology has been explored in detail elsewhere [24]. For purposes of this article we highlight some important considerations.

First, we question whether the one true model is ever under consideration, at least among currently available models. The true variation in dependent variables is composed of many effects, some large, some small. It may even be the case that the truth cannot in principle be fully modeled (cf. [25] [26]). For example, consider a purely physical phenomenon: the water level of a some river under a certain bridge. To fully model the true water level, the model would account completely for all possible contributory sources. To be the true model it must account for all rain that falls within the watershed, how much runs off into the river, how much evaporates, and what amount seeps into ground water. It would have to be sensitive to golf course sprinklers and driveway car washes, etc. The point: the causal input to something as straightforward as water level is so extremely complex that we doubt it could ever be fully modeled without help

from Laplace’s demon. Furthermore, it is impractical to fully model it; an approximating model would suffice for all purposes to which the water level calculation would be made—an approximating model would be “good enough for government work,” or better. Psychological phenomena, especially across the human population, are likely this complex and, in particular, psychoactive chemical use will never be fully modeled.

We thus presume at the outset that the true model is *not* in the candidate set. Such research thus does not select the true model, but rather the best model, where “best” invokes some notion of verisimilitude. Degree of verisimilitude is a philosophically elusive concept, as is degree of corroboration of a theory. See [27] for discussion of the former, and [28] for the latter. See [29] for the requirements of an adequate theory of confirmation [28] or corroboration [27].

Analytical derivations of the AIC and the BIC are little help in determining which one should be preferred. Each has its own theoretical basis, mathematical properties, benefits, and drawbacks.

The AIC was derived from the Kullback-Liebler information (K-L, [30]), a measure of distributional divergence in information theory [31]. Akaike [22] showed that the maximum likelihood statistic  $\log(L)$  was a biased estimate of K-L divergence, with the bias equaling (roughly) the number of estimated model parameters. As seen in (1.2), the AIC increasingly penalizes  $\log(L)$  for decreasing model parsimony. The model with the least Kullback-Liebler divergence will have the lowest AIC [32]. However, the AIC is not consistent. When the true model is under consideration AIC will tend to select higher order models, rather than the true model [33]<sup>2</sup>.

---

<sup>2</sup> We assume throughout this paper, perhaps unadvisedly, that the number of model parameters in the true model is finite (although perhaps large). When the number of true model parameters is infinite or increases with  $N$ , then the AIC is asymptotically consistent, and the BIC is not. It is perhaps indefensible to assume the data-generating model is fixed as sample size increases. Small effects will not affect every individual in society. Even with truly random sampling (which of course is never realized in clinical psychological science) there will remain effects unobserved until the sample size is increased. The previously “true” model will no longer be true. Model parameters would have to be increased to account for these effects. The argument is trivial if one assumes statistical inference with non-random

The BIC is quite different. Instead of K-L divergence, it estimates a Bayes factor [34] [35] [32]. Bayes factors are odds ratios that give the posterior odds of a model being correct given the dataset, assuming a particular prior probability for that model (c.f., [36]). Different priors result in different Bayes factors. As an estimate of the Bayes factor, the BIC also requires prior probabilities. Depending on the priors invoked, the BIC is a better or worse estimate of the Bayes Factor. For example, if the prior distribution of parameters is normally distributed  $\sim N(\hat{\mu}, \frac{N}{1}\hat{\Sigma})$  where  $\hat{\mu}$  and  $\hat{\Sigma}$  are the dataset MLE's of  $\mu$  and  $\Sigma$ . In this case, the BIC estimates the Bayes factor with a.s. decreasing error  $O(N^{1/2})$  [35]. If other priors are assumed then the BIC can lead to errors of nondecreasing order  $O(1)$ . In the normally-distributed case the priors' variance is multiplied by the sample size to render the priors extremely weak, as if they were determined by observing a single measurement of one individual.

Equal priors with very little weight can be a good thing. It represents a state of ignorance about each theory and the data. The posterior probabilities obtained with the BIC are then those of a naive researcher basing their conclusions on the data alone. Whatever the costs of endorsing any of BIC's possible implicit priors, it is essential the applied researcher understand that a particular prior is adopted which may not be defensible (or instead, not agreeable to the researcher), and which plays a non-negligible role in calculating posterior probabilities with the BIC [37].

It was shown in [38] that the BIC is a.s. consistent in multiple regression; it is guaranteed to select the true model as the number of observations grows large, assuming the true model is in the candidate model set. However, as the true model is very arguably never under consideration in our field, the proof in [38], and related proofs, are moot for practical purposes in psychology.

---

sampling.



When the true model is not in the candidate model set the AIC’s asymptotic properties might become attractive to the applied researcher. When the candidate set does not contain the true model the AIC selects the best K-L approximating model. The AIC is thus asymptotically efficient for K-L divergence [39]—a desirable property if the candidate model set never contains the true model. Asymptotic efficiency of the AIC extends to mean squared error of prediction. Under certain conditions (e.g., regression model with normal errors), the AIC is equivalent to cross-validation [40], and minimizes MSE in prediction of new data points [34]. The BIC has not been proved to be asymptotically efficient, and its properties are unclear when the true model is not in the candidate set. Of course, all these results only hold in the limit, and finite sample behavior is difficult to determine.

To complicate matters further, the AIC also assumes that the true model is in the candidate model set. The AIC’s dependence lies in its correction of the  $\log(L)$ . To wit,  $K$  is only the proper penalty in (1.2) if the true model is under consideration. Indeed, the AIC’s correction of  $K$  is a specific instance of a more general correction given in the Takeuchi Information Criterion (see, e.g., [41] [32] [25], for comprehensive discussions). Despite this, some have cogently argued that the AIC remains viable because, while  $K$  may be biased, it has no variability, whereas the TIC has less bias and more variability [42].

Simulation studies must be considered to understand BIC and AIC behavior in finite sample sizes. Markon and Krueger [18] have investigated the likelihood with which the BIC will recover the data-generating model (which was among the candidate model sets in their simulations). Simulations in [18] were conducted in two parts, one with the data-generating model’s latent variables as discrete and the other normally distributed. In the discrete case, they considered data-generating models with two, three, four, and five discrete values (ordered classes). Two candidate models were considered, a continuously

distributed latent trait model and a discretely distributed latent trait model with the same number of ordered classes as the data-generating model. The results clearly showed that the BIC selects the “true” model nearly 100% of the time when the true model is simple (i.e., has two discrete latent classes). As the model complexity increases, and the number of model parameters increases, and the BIC became increasingly unhelpful in selecting the correct model. When the data-generating model was a five-valued latent trait model the BIC selected the five-valued latent trait model 19% of the time. As the sample size became smaller and the mean difference between classes decreased, the BIC performed less well. With 10 binary measurement items and a sample size of 2,000 the BIC selected the data-generating five-valued ordered latent class model zero times out of 500.

The BIC performed much better when the data generating model was a continuous normally distributed latent trait model (which can be specified with even fewer parameters than the two-valued ordered latent class model). The BIC selected the continuous normally distributed latent trait model nearly 100% of the time in the entire (but limited) parameter space studied in [18]. It is noteworthy that in [17] only five binary measurement items were used in a sample size just shy of 3,000. The simulation reported in [18] seems to suggest that the methods in [17] would have returned a single-factor solution with probability 0.89 to  $\sim 1.0$ , whether the true structure was unidimensional or had several ordered latent classes.

Others examined the relative performance of the BIC and the AIC in selecting the number of latent classes in latent class analysis, factor mixture modeling, and growth mixture models [43]. Data-generating latent class models had either three or four latent classes; modeled either eight, 10, or 15 items; and had equal or unequal latent class proportions. Not all combinations of these data properties were investigated. In sum, they found that the BIC more accurately selected the correct number of latent classes

than the AIC, and often with 30-40% greater accuracy. When erring, the AIC tended to overestimate the number of latent classes. Unfortunately, they did not examine selection between models differing in number of latent classes but alike in all other respects (e.g., equal in number of items, latent class proportions, etc.). It is thus impossible to conclude whether the AIC or the BIC tends to perform better when the latent structure is simple versus complex.

Focusing on Equations (1.2) and (1.1), it is clear that as the sample size  $N$  grows larger the BIC tends to prefer models with fewer parameters, *ceteris paribus* (see, e.g., [44]). One might expect the BIC to outperform the AIC when the data-generating model contains a few large effects, and the AIC to outperform the BIC when the data-generating model contains both large effects and smaller tapering effects, as the AIC is prone to select models with more parameters necessary to capture the small tapering effects.

Researchers have investigated this scenario [31]. They generated data from simple models (few large effects) and complex models (large effects and tapering effects). In every case the candidate model set included a range of simple and complex models. They found that the AIC outperformed the BIC when the data-generating model was complex. The BIC, on the other hand, outperformed the AIC when the data-generating model was simple. This is the same trend observed in other simulations [18].

Except for those models considered in [31], we have only discussed simulations with latent trait or latent class models. Given advances in structural equation modeling, other models are now available to the applied researcher (cf. [7]). We now examine the possible alternative role of mixture models of latent traits.

In the same NESARC sample studied by Markon and Krueger [18], Muthén [6] fit latent trait, latent class, and mixtures of latent traits, to alcohol abuse and dependence criteria. Item thresholds and factor variances were allowed to vary across the two classes.

Models were compared with the BIC. He reported that the best-fitting model was a two class mixture of latent traits, with a single trait undergirding each class. The AIC was not reported, but also supports the two-class, single factor model <sup>3</sup> .

Kuo and colleagues [19] have extended these findings. They compared models of alcohol dependence criteria in the Virginia Twin Registry, a community sample of twins. They fit latent class models with one, two, three, four, and five latent classes; factor models with one and two factors; and mixtures of traits with two classes/one trait and three classes/one trait. The best fitting model for both genders was a mixture of latent traits with two classes and a single factor. However, it was concluded that the three-class/one-factor model was to be preferred because it was less parametrically variant across genders, but also because the investigators claimed theory suggests alcoholism to be a complex disorder consisting of many effects, and that the three class model could better account for these putative effects.<sup>4</sup>

In light of [6] and [19] it is clear that comparing only latent trait with latent class models is not sufficient, because this limited model set misses a wide range of possibly more verisimilar models. Their findings also stress how the items under analysis can affect interpretive conclusions. One may obtain different results if many different types of externalizing psychopathology are modeled separately, instead of jointly.

Our present approach overcomes these problems with two main analyses. First it applies latent trait models, latent class models, and mixtures of latent traits. Second, the DSM-IV-TR [1] chemical dependence and abuse diagnostic criteria for a variety of substances are modeled jointly in a sample of inmates. Third, we model alcohol

---

<sup>3</sup> Note that the numbers in Table 1 of [7] contain a typesetting error. The decimals in the  $\log(L)$ , BIC, and ABIC columns should not be considered decimals, but commas. For example, 25.583 should actually read 25,583. Not knowing this, the table values appear incorrect.

<sup>4</sup> It is noted that the results reported in Table 1 of [19] are inconsistent. The AIC and BIC values given do not follow from the sample size, log likelihood, and number of parameters. Further, the formula given in [19] for the AIC is incorrect, as it rewards the AIC for increasing model complexity (instead of penalizing for it). The BIC is not given in [19]. The author was contacted about this issue in September, 2009, and no response has yet been received.

items separately, in an attempt to illustrate how different models may provide better fit depending on the level of analysis.

Analyzing an inmate sample has certain advantages, despite being unrepresentative of the community as a whole. Foremost among them are large base rates of substance use. These base rates generate substantial base rates of substance use. Relatively subtle effects are more pronounced in this case, and can be more reliably modeled than in a community sample. For example, the rate of cocaine dependence in the community is far below the 11% observed in the present report.

## Chapter 2

# Method

**The Participants** The data utilized in the analyses were drawn from the systematic evaluation of inmates for substance use disorders by counselors in the Minnesota correctional system using the SUDDS-IV (Substance Use Disorder Diagnostic Schedule-IV) during approximately a two year period. During this time, a total of 7,682 adult inmates between the ages of 18 and 65 (Mean = 30.99; SD = 9.20) were evaluated. The sample consisted of 6,882 males and 801 females. The majority of the inmates were white (51.5%) with African Americans being the largest minority (30.4%). Native Americans comprised 8.3% of the sample, and Hispanics accounted for 6.3%. No other minority group accounted for more than 2% of the sample. Over a third (35.4%) of the sample had not graduated from high school, and only 12.1% had any education beyond high school. Over two-thirds (67.4%) had never married and only 12.3% were married at the time of the evaluations. The remainder were separated, divorced, or widowed.

**Models** In traditional latent class analysis [5] the  $K$  variables are presumed independent given latent class membership with  $G$  latent classes (i.e., local independence).

Formally,

$$f(\mathbf{x}_n) = \sum_{g=1}^G \pi_g f(\mathbf{x}_n | \boldsymbol{\theta}_g), \quad (2.1)$$

where the  $g$ th component density is

$$f(\mathbf{x}_n | \boldsymbol{\theta}_g) = \prod_{k=1}^K \theta_{kg}^{x_{kn}} (1 - \theta_{kg})^{1-x_{kn}} \quad (2.2)$$

$\mathbf{x}_n = (x_{1n}, \dots, x_{kn})'$  is a single subject's scores on the  $K$  variables,  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_g)'$  the vector of  $G$  latent class proportions, and  $\boldsymbol{\theta}_g = (\theta_{1g}, \dots, \theta_{Kg})'$  are the latent class model parameters (i.e.,  $\theta_{kg} = \Pr[x_{kn} = 1 | x_{kn} \in g]$ , the probability that  $x_{kn}$  is one given that it belongs to the  $m$ th class). Hence, a latent class model here has  $G \times K + (G - 1)$  parameters.

Equation (2.1) is a simple mixture equation. Given knowledge of class membership (i.e., parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\pi}$ ), the observed data vector  $\mathbf{x}_n$  is simply the weighted sum of the  $G$  individual component distributions. We see in (2.2) that, given knowledge of  $\boldsymbol{\theta}$ , the joint distribution of the observed variables can be expressed as a product of the marginal p.d.f.s. This property is known as local independence. In essence, it presumes that the observed variables are statistically independent once one has knowledge of latent class membership—i.e., the latent classes account for all inter-relationships in the observed data.

Latent class analysis models the relationship between discrete indicator variables and discrete latent variables. Latent trait analysis models the relationship between discrete indicator variables and *continuous* latent variables. We seek to represent the observed covariance matrix  $\boldsymbol{\Sigma}$  as a function of an  $M \times M$  matrix of latent trait local and global dependencies, as follows:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda} \boldsymbol{\Psi} \boldsymbol{\Lambda}' + \mathbf{R}, \quad (2.3)$$

where  $\boldsymbol{\Psi}$  is a  $M \times M$  matrix of latent trait correlations,  $\boldsymbol{\Lambda}$  a  $K \times M$  matrix of loadings and  $\mathbf{R}$  a diagonal error matrix. This model, of course, only accounts for the

covariance matrix  $\Sigma$ , and not the mean structure. In the present analyses, the indicator variables were scaled to have mean of zero.

The latent trait model also assumes local independence. Given knowledge of the latent trait densities distributed as the Gaussian  $\sim N(\boldsymbol{\mu}, \Sigma^*)$ , the joint probability density function of the observed variables  $\boldsymbol{x}$  is a product of the marginal densities, or,

$$f(\boldsymbol{x}_n | \boldsymbol{\mu}, \Sigma) = \prod_{k=1}^K f(x_{kn} | \boldsymbol{\mu}, \Sigma). \quad (2.4)$$

Our particular latent trait model used dichotomous indicator variables. Each indicator variable was allowed to have its own latent response distribution. Specifically, it was assumed that a latent normal distribution, dichotomized at some suitable cutting score, was responsible for each observed dichotomous indicator variable. To accomplish this, one estimates threshold parameters  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_K)$  that indicate where the latent normal distribution was dichotomized to generate the observed indicator variable [45]. For a binary indicator variable  $K$  there is a single threshold parameter  $-\infty < \tau_k < \infty$ , bisecting the normal distribution such that individuals scoring greater than  $\tau_K$  on the  $K$ th indicator variable are assigned a 1 for  $K$ , and those scoring less than  $\tau$  are assigned a zero. Estimation in our factor models was conducted on the polychoric correlation matrices as determined by the threshold parameter vector  $\boldsymbol{\tau}$ .

A mixture model of latent traits takes precisely the same general form as (2.1), but with an expanded parameter structure:

$$f(\boldsymbol{x}_n | \boldsymbol{\xi}) = \sum_{g=1}^G \pi_g f(\boldsymbol{x}_n | \boldsymbol{\mu}_g, \Sigma_g) \quad (2.5)$$

In this case the parameter structure  $\boldsymbol{\xi}$  contains parameter vector  $\boldsymbol{\pi}$  for the latent class structure as well as  $\boldsymbol{\Lambda}_g$ ,  $\boldsymbol{\Psi}_g$ , and  $\boldsymbol{\mu}_g$  (a zero vector) for the  $M$  within-class latent traits. Each component density  $f(\boldsymbol{x}_n | \boldsymbol{\mu}_g, \Sigma_g)$  is of course structured in terms of equation (2.3) with  $\boldsymbol{\mu} = 0$  for simplicity. Latent traits were taken to be normally



distributed with mean vector  $\boldsymbol{\mu}^*$  and covariance matrix  $\boldsymbol{\Sigma}^*$  for ease of analysis; any probability density function could in principle be chosen.

Factor means and loadings were intentionally made invariant across the mixing distributions, i.e.,  $\boldsymbol{\Lambda}_a = \boldsymbol{\Lambda}_b$  and  $\mu_a^* = \mu_b^* = 0$  for all  $a, b \in g$ . Equating means and loadings is not necessary, but assists in dealing with model identification. This is a common, complex problem in factor mixture modeling (e.g., [46]).

Two crucial inherent limitations of the traditional latent trait model are (a) the assumed linear relationship  $\boldsymbol{\Lambda}$  between the latent and indicator probability distributions, as well as (b) the local independence assumption. Mixtures of latent trait models do not suffer these limitations, and can adequately fit a wider variety of data structures.

Latent class and latent trait models (among others) are submodels of the more general mixture of latent traits. If the number of latent classes is one, then a mixture of latent traits specializes to the latent trait model. Similarly, if the subjects show no individual differences within latent classes, then the mixture model of latent traits becomes a simple latent class model. One can think of the mixture model of latent traits as a latent class model with the assumption of local independence relaxed. For example, in substance use research the within-class latent traits could be interpreted as severity and/or extra-syndromal item clusterings. For example, if alcoholism is taxonic, then an individual either is an alcoholic or is not. Under a mixture model of latent traits, however, there would be individual variation within alcoholics on various dimensions.

Regardless of the model we employed, parameter estimates were obtained through maximum likelihood estimation via the Expectation-Maximization (E-M) algorithm [47] [5] [46]. All analyses were carried out with Mplus software [7]

**The Indicator Variables** Initially under consideration were 11 abuse and dependence criteria from five drug classes: alcohol, marijuana, cocaine, stimulants, and heroin.

Diagnostic criteria related to other substances of abuse did not occur frequently enough in our study population to make analyzing these criteria worthwhile. Each criterion variable was originally measured on a several-valued ordinal scale indicating the sum of each criterion item (each abuse/dependence criterion item was measured by several sub-items, which were summed to create the abuse/dependence criterion item score). Response distributions on these several-valued criterion items were quite positively skewed; some items had almost all (e.g., 97%) of their probability mass functions located at zero. Higher ratings (2 through 5) were fairly evenly distributed. To cope with the highly skewed nature of the items, we dichotomized all of them. Zero values were unchanged but any rating greater than zero was given a value of one.

Detailed descriptions of the SUDDS-IV instrument can be found in [48] [49] [50].

**Analysis of Alcohol, Marijuana, Cocaine, and Stimulant Criteria** Exploratory analyses were conducted prior to modeling. Variable histograms and correlations were examined. After examining the  $11 \times 11$  item polychoric correlation matrix, we considered redundant and excluded one item (the second) from each pair of items with correlations  $> .95$ . Retained items are listed in Table 1. All heroin use items were correlated  $> .95$  with each other, and were thus all removed from the analysis.

To all these diagnostic criteria we fit exploratory factor models using maximum likelihood estimation with robust standard errors [51] calculated by the EM algorithm, followed by promax rotation. The first six eigenvalues of the tetrachoric correlation matrix were 11.33, 7.54, 4.30, 3.82, 0.70, and 0.51, strongly suggesting a four-factor solution. Fit statistics also supported this conclusion. Root Mean Square Error of Approximation (RMSEA) for the one-, two-, three-, and four-factor solutions were .28, .17, .11, and .04, respectively. Root Mean Square Residual for the same models was:

Item	Alcohol	Marijuana	Cocaine	Stimulant
<i>Dependence</i>				
Tolerance	✓	✓	✓	✓
Withdrawal	✓	✓	✓	✓
Extended/Unplanned Use	✓	✓	✓	✓
Unable to Quit	✓	✓		
Time Spent Using	✓	✓		
Sacrifice of Activities	✓	✓	✓	✓
Contraindications to Use	✓	✓		
<i>Abuse</i>				
Role Fulfillment		✓		
Dangerous Use	✓	✓		
Legal Problems	✓	✓	✓	✓
Interpersonal Conflicts	✓	✓		

Table 2.1: Items Retained after Removal of Items with Correlations  $> .95$ .

.50, .27, .16, and .03, respectively. Inspection of the rotated factor structure showed a strikingly tight clustering among diagnostic criteria from the same drug class. In addition, the factors were largely uncorrelated. These results were replicated using a weighted least squares estimator in lieu of the maximum likelihood with robust standard error estimator.

Next, we fit a series of models involving latent classes, latent traits, and mixtures of latent traits. All parameters were estimated by the same maximum likelihood estimator as was used in exploratory analyses.

We fit latent class models with two to five latent classes. Traditional psychodiagnosis would suggest that the data contains five latent classes: one normal class, and four

classes of drug dependence, one for each drug investigated. In addition, the scree test suggested four factors, prior to rotation, and oblique rotation will, if anything, decrease the number of (rotated) factors needed to account for the correlation matrix. Hence, we chose the maximum number of latent classes to be five.

In the confirmatory factor model, factor variances were fixed at one, factor means at zero, and all loadings freely estimable. For the mixture of traits model, we allowed item thresholds and factor variances to vary across classes, but factor dimensionality, means, and loadings were assumed to be identical across classes. Of course, these assumptions may not all be correct. A mixture of traits model may have different factor loading and factor intercorrelation parameters in different classes. Indeed, classes may even have structurally different solutions, such as one class with four factors and another with five. However, parameter estimation becomes numerically more problematic as more parameters are freely estimated. Mixtures of traits with different loadings per class were fit, but this resulted in convergence problems in nearly every fit.

For model selection, we used the BIC and AIC, discussed at length in the introduction. The model with the lowest fit statistics was taken as probably being the most verisimilar model.

**Analysis of Alcohol Criteria Only** Finally, we investigated the structure of alcohol use only, to understand how different levels of analysis can lead to diverse structural conclusions. We investigated the latent structure of alcohol dependence and abuse items only. We fit latent class, latent trait, and mixtures of traits to these dichotomous items. As with our larger analysis, the best fitting model as quantified by the AIC and the BIC were taken to be most verisimilar.

## Chapter 3

# Results

**Alcohol, Marijuana, Cocaine, Stimulants** According to DSM-IV-TR criteria, as measured by the SUDDS interview, the following rates of lifetime dependence and abuse were observed in the present sample: 29% and 36% for alcohol dependence and abuse; 18% and 20% for marijuana dependence and abuse; 11% and 12% for cocaine dependence and abuse; and 11% and 12% for stimulant dependence and abuse. Hierarchical exclusion rules between abuse and dependence were ignored in arriving at these rates.

Results of the model comparisons are given in Table 2 for the alcohol, marijuana, cocaine, and stimulant items. The four-factor solution is clearly the best fitting model, as estimated by the BIC and AIC. Factors in the four-factor solution were fit according to drug class; the first factor to alcohol criteria, the second to marijuana criteria, and so on. All cross-drug loadings were fixed at zero. Factors were minimally correlated. The factor correlation matrix was:  $r(\text{Alcohol, Marijuana}) = .26$ ,  $r(\text{Alcohol, Cocaine}) = .22$ ,  $r(\text{Alcohol, Stimulants}) = -.245$ ,  $r(\text{Marijuana, Cocaine}) = -.02$ ,  $r(\text{Marijuana, Stimulants}) = .18$ ,  $r(\text{Cocaine, Stimulants}) = -.23$ . For each drug

Model	$\log(L)$	$K$	AIC	BIC
Latent Class				
2	-87750.535	63	175627.070	176064.708
3	-78786.155	95	157762.311	158422.241
4	-74918.006	127	150090.011	150972.234
5	-71944.791	159	144207.582	145312.097
Latent Trait				
4	-63338.087	68	126812.174	127284.545
Mixt. of Traits				
2C 1T	-71580.987	95	143351.975	144011.905
3C 1T	-68270.399	125	136790.797	137659.127
4C 1T	-65397.055	158	131110.110	132207.679
5C 1T	-64138.749	189	128655.499	129968.413

Table 3.1: Analysis of Dependence and Abuse Criteria.

$\log(L)$  is the log of the likelihood,  $K$  the number of free parameters.

class we also summed all items retained in Table 1, and correlated them. The resulting correlations between these raw scores were  $r(\text{Alcohol, Marijuana}) = .19$ ,  $r(\text{Alcohol, Cocaine}) = .11$ ,  $r(\text{Alcohol, Stimulants}) = -.09$ ,  $r(\text{Marijuana, Cocaine}) = .03$ ,  $r(\text{Marijuana, Stimulants}) = .11$ ,  $r(\text{Cocaine, Stimulants}) = -.04$ .

Under the restrictive assumption of equal prior probabilities—based on a single observed data point—for all models in Table 2, one can compute the Bayes posterior probability favoring one model over another [34]. This probability equals the difference in the BIC between the two models, raised to the Euler constant  $e$ . For example, the

Table 3.2: Parameters for the Four-Trait Latent Trait Model.

Item	T1 (SE)	T2 (SE)	T3 (SE)	T4(SE)	Thresholds
<i>Alcohol</i>					
(D) Tolerance	.850				.439
(D) Withdrawal	.927				.856
(D) Unplanned/Extended Use	.929				.595
(D) Unable to Quit	.950				.732
(D) Excess Time Spent Using	.958				.587
(D) Sacrifice of Activities	.956				.717
(D) Health Contraindications	.941				.536
(A) Dangerous Use	.898				.376
(A) Legal Problems	.821				.635
(A) Interpersonal Difficulties	.951				.418
<i>Cannabis</i>					
(D) Tolerance		.869			.950
(D) Withdrawal		.911			1.378
(D) Unplanned/Extended Use		.912			1.032
(D) Unable to Quit		.938			1.069
(D) Excess Time Spent Using		.900			.661
(D) Sacrifice of Activities		.968			1.146
(D) Health Contraindications		.944			.934
(A) Role Fulfillment		.963			1.079
(A) Dangerous Use		.892			.788
(A) Legal Problems		.750			1.170
(A) Interpersonal Difficulties		.913			.928

Table 3.2 Continued

Item	T1 (SE)	T2 (SE)	T3 (SE)	T4(SE)	Thresholds
<i>Cocaine</i>					
(D) Tolerance			.962		1.084
(D) Withdrawal			.967		1.328
(D) Unplanned/Extended Use			.986		1.024
(D) Sacrifice of Activities			.989		1.091
(A) Legal Problems			.876		1.186
<i>Stimulant</i>					
(D) Tolerance				.969	1.121
(D) Withdrawal				.971	1.373
(D) Unplanned/Extended Use				.973	1.198
(D) Sacrifice of Activities				.984	1.228
(A) Legal Problems				.919	1.143

posterior probability that the four-factor model is more verisimilar than the next best model (with five classes and one factor) is  $e^{2299.4}$ , a very large number.

Parameters for the four-factor latent trait model are given in Table 3. As expected, the threshold parameters indicate that more subjects endorse alcohol use problems than problems with other drugs.

**Alcohol Only** Despite clear evidence for a four-factor solution when all four drug classes were considered, a different result was obtained when only alcohol dependence and abuse criteria were entered into the model. Comparative model fits are listed in Table 4. The best fitting model according to the BIC and the AIC was a mixture of latent traits with three classes and a single latent trait underlying each one, replicating



previous results from findings reported in [19]. Furthermore, the general pattern of AIC/BIC findings is also similar to that reported in [19]. Mixture models of latent traits, and four- and five-class latent class models, all fit better than the unidimensional latent factor model.

Model	$\log(L)$	$K$	AIC	BIC
Latent Class				
2	-30200.032	23	60446.064	60605.836
3	-27837.183	35	55744.365	55987.497
4	-27452.476	47	54998.951	55325.443
5	-27327.459	59	54772.918	55182.769
Latent Trait				
1	-27726.449	22	55496.898	55649.724
2	-27420.866	34	54909.732	55145.917
Mixt. of Traits				
2C 1T	-27463.850	33	54993.699	55222.938
3C 1T	-27201.590	46	54495.181	54814.726

Table 3.3: Latent Structure of All SUDDS Alcohol Dependence and Abuse Criteria.

Note:  $\log(L)$  is the log of the likelihood,  $K$  the number of free parameters.

## Chapter 4

# Discussion

Overall, the best model depended on the level of analysis. When all items were analyzed, a four-trait model emerged. However, when models were fit only to alcohol dependence/abuse, the data supported a three class, single trait model. This finding is not only scientifically interesting in and of itself, and generally consistent with the DSM-IV's current structure, but has methodological implications for future modeling research.

The externalizing spectrum theory predicts that alcohol, marijuana, cocaine, and stimulant problems all stem from a unidimensional externalizing liability. This does not appear to occur in high-risk populations such as prison inmates. While we did not find evidence of a single externalizing spectrum *per se*, the present findings do suggest that each chemical use disorder does cohere as a dimensional syndrome, at least at a DSM-IV level of analysis.

Individuals can range anywhere from zero maladaptive use to extreme maladaptive use, and do so independently for each drug class studied. However, it should be noted that the item thresholds were all quite high (greater than zero, in any event), and no item measured the lower end of the dimension. The dimensions may then be best

considered problem-behavior dimensions, that may or may not extend into normalcy. Future research would do well to construct datasets that contain items measuring the low end of the dimension. For example, Krueger and colleagues in [9] did well to ask the subjects whether they had ever been intoxicated. Few items would measure any lower on the alcohol continuum; no such item was available in the present analysis.

Of course, our result depends on the level of analysis. When models were fit exclusively to alcohol items a very different result was obtained. Maladaptive alcohol use was not best modeled with a single latent trait but rather with three classes, all undergirded with a single latent trait. The result confirms previous modeling research [19] that obtained the single-trait, three-class model for alcohol dependence criteria in a sample of participants from the Virginia Twin Registry.

Classes identified with the three-class model map somewhat surprisingly onto the current constructs of alcohol abuse and dependence. Figure 1 displays the relationship between the alcohol items and the latent classes.

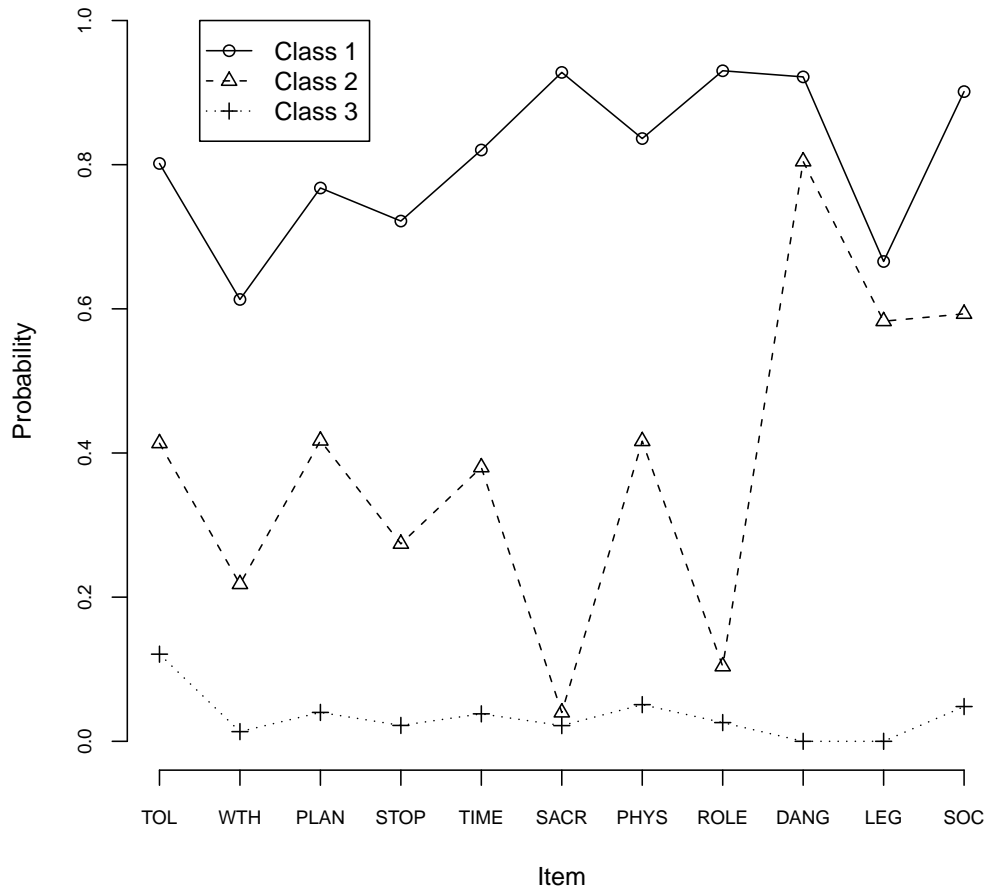


Figure 4.1: Alcohol Item Response Probability Given Class Membership.

Alcohol items are denoted as follows. Dependence items: tolerance (TOL), withdrawal (WTH), unplanned use over longer periods than originally intended (PLAN), unsuccessful efforts to cut down/control use (STOP), much time spent using the substance, getting it, etc. (TIME), a sacrifice of other activities in order to use (SACR), continued use despite physical and mental contraindications (PHYS). Abuse items: failure to fulfill role obligations (ROLE), use in dangerous situations (DANG), substance-related legal problems (LEG), continued use despite it causing problems in social/interpersonal relationships (SOC).

Class 3 is composed of normals. As can be seen in Figure 1, these individuals are very unlikely to endorse any alcohol problems. Class 3 comprises the majority of participants (4,541, 59%), based on posterior probability of class membership.

Class 2 is smaller (1,376, 18%) and reflects to some extent a class of alcohol abusers. The difference between this class and the DSM-IV-TR's putative alcohol abuse class is that individuals in Class 2 are quite likely to endorse alcohol dependence criteria (e.g., 40% endorse developing tolerance) except for role fulfillment problems and sacrificing other activities in order to drink. While it remains to be investigated, in our clinical experience many individuals with moderate drinking problems do not see themselves as sacrificing activities to drink. Instead, their hobbies and social lives happen to involve the consumption of copious amounts of alcohol.

Class 1 is comprised of 1,765 (23%) participants, and reflects a DSM-related alcohol dependence category. Individuals in this class endorse with high probability all alcohol dependence and abuse items, suggesting that they have a high drive to use alcohol at great expense to other life priorities, and that this alcohol use causes them social problems.

While our fitted model is nominally the same as that obtained in [19] the parameter estimates are somewhat different. First, they fit models only to alcohol dependence items; we entertained all alcohol and abuse items. They found a 3-class, 1-factor, solution for the seven dependence items alone. Our latent structure clearly contains an abuse-like group, and those group members have high probability of endorsing abuse items. In this way our results are incomparable.

It may very well be the case that the second group in [19] was an abuse-like group, albeit poorly modeled due to the lack of abuse items in their analysis. The abuse group in [19] has probability of responding to an alcohol dependence item as somewhere between the normals and the dependents. In no small part, the graph shapes in Figure 1 mirror

the shapes reported in [19] although our abuse group was somewhat less likely to endorse dependence items than theirs. One notable difference is that the two pathological classes in [19] endorsed the Phys item with low frequency (i.e., continued drinking despite physical or mental contraindications). Specific discrepancies of this sort may be due to sample composition, as [19] used a community sample with less severe substance use whereas our sample was drawn from prison inmates.

Item characteristic curves (ICCs) for the latent trait(s) undergirding each of the three latent classes are displayed in Figure 2. The general shape of these ICCs is consistent with expectations. For Class 3, the normal class, all dependence and abuse items are consistently more difficult (i.e., lie to the right of) than they are for the abuse class and, in turn, the items are easier for Class 1 (dependents) than they are for any other class. Note also that while the item difficulty (location on the x-axis) differs for each item across classes, the discriminations (slopes) do not. This is no coincidence, since item discriminations were fixed to be equal during model specification in order to ease estimation as discussed in the Method section.

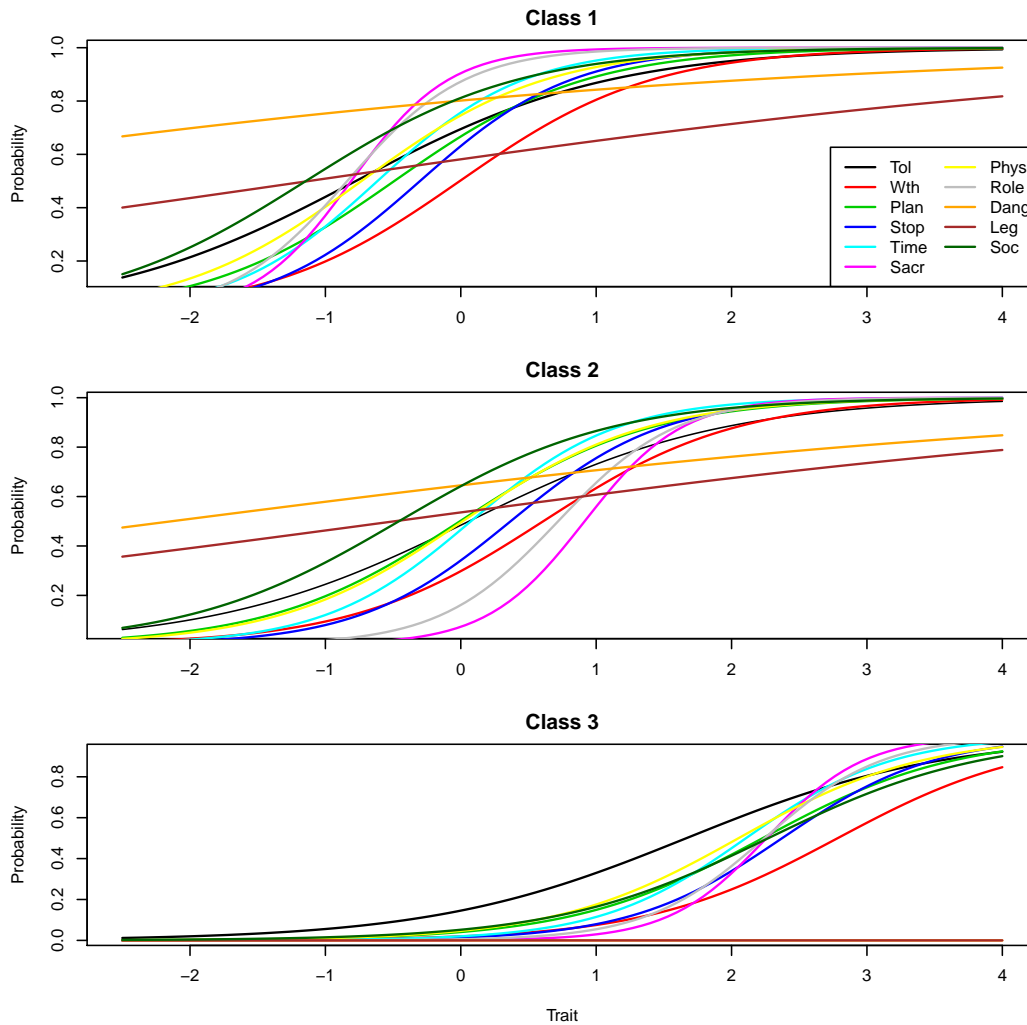


Figure 4.2: Item Characteristic Curves.

DANG and LEG ICCs in Class 3 have difficulty parameters beyond the right margin of the graph, and thus are not visible except for their trailing left tails, visible along the bottom of the graph.



Apparent in Figure 2 is that two items have extremely poor discriminations: Dangerous Use (DANG) and substance-related Legal Problems (LEG). These items appear to give very little information about the latent trait. There are at least two possible reasons for this. 1) It may be that a second latent trait underlies these two items, one related more to lack of fear, thrill-seeking, or criminal behavior than to substance use *per se*. 2) DANG and LEG are poorly measured by the SUDDS. The latter interpretation was not supported in the four-drug analysis, as standard errors reported in Table 2 were comparable for the DANG and LEG items, compared to other items. What is more, factor loadings in the four-drug analysis were consistently lower for DANG and LEG, suggesting perhaps that these two items are poorer relative measures of the latent trait. This effect may have been exacerbated in the mixture of latent traits model thus resulting in the very flat ICCs in Figure 2. Other models were fitted to deal with this issue; however, neither allowing loadings to vary across classes or introducing another latent trait resulted in improved model fit or interpretability.

Most applied structural modeling research, as far as we can determine, is largely exploratory, and entertains rather simple models. There is often no rationale for testing an extremely complex model (for example, we did not test a 4-class, 2-factor model for this reason). When models are fitted to dozens of variables the largest effects will dominate the model results. In our CFA of all drug items the largest effects appear to have been within-drug-class endorsement of items. However, when models of equal complexity are fit to a smaller number of variables, more subtle effects can be discovered (if they exist) and different models selected, as occurred when we restricted model fitting to the alcohol items only.

While unrepresentative of the general population, the present dataset is well-adapted for structural analysis of substance use for several reasons. 1) If dependencies on individual drugs are discrete illnesses then collapsing drugs (as done in previous research)

destroys the unique effects. Of course, we must grant that the stimulant drug class under investigation here is actually a meta-class of individual stimulant drugs, and that future researchers may wish to disentangle the different stimulants. 2) Rates of alcohol and marijuana use problems are greater in our sample than in a community sample. Any individual drug effects were more pronounced, allowing us to capture small, and large but rare, effects more easily. 3) With over 7,000 participants the sample size was quite large, allowing us to capture subtleties in the data, if they exist.

On the basis of the present results, it appears that the DSM appropriately distinguishes between alcohol, marijuana, cocaine, and stimulant use problems. These drug problems should not be collapsed into a meta-category of drug use; this obscures their independent ontology. While alcohol use problems may be distinct from marijuana use problems, each set contains its own structure. Alcohol use problems may have a three class structure comprised of normals, individuals who endorse some dependence-type problems but are most likely to endorse major life-role dysfunction (the DSM-termed “abusers”), and those that are likely to endorse every problem (so-called “dependents”).

# References

- [1] American Psychiatric Association. *Diagnostic and statistical manual for mental disorders*. Author, Washington, D.C., fourth edition, 1987.
- [2] N.G. Waller and P.E. Meehl. *Multivariate taxometric procedures: Distinguishing types from continua*. Sage, Thousand Oaks, CA, 1998.
- [3] W.M. Grove. Dimensional analyses of taxonic data. *Psychological Reports*, 101:617–635, 2007.
- [4] P.E. Meehl. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46:806–834, 1978.
- [5] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley-Interscience, New York, 2000.
- [6] B.O. Muthén. Should substance use disorders be considered as categorical or dimensional? *Addiction*, 101 (Suppl 1):6–16, 2006.
- [7] G.R. Hancock and K.M. Samuelsen, editors. *Latent Variable Mixture Models*. Information Age Publishing, Inc., Charlotte, NC, 2008.

- [8] B.O. Muthén. Latent variable modeling in heterogeneous populations. *Psychometrika*, 54:557–585, 1989.
- [9] R.F. Krueger, P.E. Nichol, B.M. Hicks, K.E. Markon, C.J. Patrick, W.G. Iacono, and M. McGue. Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment*, 16:107–119, 2004.
- [10] H. Proudfoot, Andrew J. Baillie, and Maree Teesson. The structure of alcohol dependence in the community. *Drug and Alcohol Dependence*, 81:21–26, 2006.
- [11] T.C. Harford and B.O. Muthén. The dimensionality of alcohol abuse and dependence: A multivariate analysis of DSM-IV symptom items in the national longitudinal survey of youth. *Journal of Studies on Alcohol*, 62:150–157, 2001.
- [12] B.O. Muthén, B. Grant, and D. Hasin. The dimensionality of alcohol abuse and dependence: Factor analysis of DSM-III-R and proposed DSM-IV criteria in the 1988 National Health Interview Survey. *Addiction*, 88:1079–1090, 1993a.
- [13] B.O. Muthén, D. Hasin, and K.S. Wisnicki. Factor analysis of ICD-10 symptom items in the 1988 National Health Interview Survey on Alcohol Dependence. *Addiction*, 88:1071–1077, 1993b.
- [14] D.S. Hasin, B.O. Muthén, K.S. Wisnicki, and B. Grant. Validity of the bi-axial dependence concept: A test in the us general population. *Addiction*, 89:573–579, 1994.
- [15] C.B. Nelson, J. Rehm, T.B. Ustun, B. Grant, and S. Chatterji. 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, 1999.

- [16] A. Feingold and B. Rounsaville. Construct validity of the dependence syndrome as measured by DSM-IV for different psychoactive substances. *Addiction*, 90:1661–1669, 1995.
- [17] R.F. Krueger, K.E. Markon, C.J. Patrick, and W.G. Iacono. Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, 114:537–550, 2005.
- [18] K.E. Markon and R.F. Krueger. Information-theoretic latent distribution modeling: Distinguishing discrete and continuous latent variable models. *Psychological Methods*, 11:228–243, 2006.
- [19] P. Kuo, S.H. Aggen, C.A. Prescott, K.S. Kendler, and M.C. Neale. Using a factor mixture modeling approach in alcohol dependence in a general population sample. *Drug and Alcohol Dependence*, 98:105–114, 2008.
- [20] A.E. Gelfand and D.K. Dey. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, 56:501–514, 1994.
- [21] G.E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.
- [22] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- [23] K.E. Markon and R.F. Krueger. Categorical and continuous models of liability to externalizing disorders. *Archives of General Psychiatry*, 62:1352–1359, 2005.
- [24] T. Millon, R.F. Krueger, and E. Simonsen, editors. *Contemporary directions in psychopathology: Toward the DSM-V*. Guilford Press, New York, in press.

- [25] K.P. Burnham and D.R. Anderson. *Model selection and multimodel inference: A practical-theoretic approach*. Springer-Verlag, New York, NY, 2003.
- [26] J. Rissanen. Stochastic complexity. *Journal of the Royal Statistical Society, Series B*, 9:223–239, 1987.
- [27] K.R. Popper. *The logic of scientific discovery*. Routledge, London, 1959.
- [28] Rudolf Carnap. *Logical foundations of probability*. University of Chicago Press, Chicago, IL, second edition, 1962.
- [29] C.G. Hempel. Studies in the logic of confirmation. *Mind*, 54:1–26, 97–121, 1945.
- [30] S. Kullback and R.A. Liebler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.
- [31] K.P. Burnham and D.R. Anderson. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33:261–304, 2005.
- [32] G.C. Chow. A comparison of the information and posterior probability criteria for model selection. *Journal of Econometrics*, 16:21–33, 1981.
- [33] M. Stone. Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society, Series B*, 41:276–278, 1979.
- [34] J. Kuha. AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods Research*, 33:188–229, 2004.
- [35] R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.
- [36] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, Boca Raton, FL, 2004.

- [37] David L. Weakliem. A critique of the Bayesian Information Criterion for model selection. *Sociological Methods & Research*, 27:359–397, 1999.
- [38] R. Nishii. Asymptotic properties of criteria for selection of variables in multiple regression. *Annals of Statistics*, 12:758–765, 1984.
- [39] R. Shibata. Asymptotic mean efficiency of a selection of regression variables. *Annals of the Institute of Statistical Mathematics*, 35:415–423, 1983.
- [40] M. Stone. An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion. *Journal of the Royal Statistical Society, Series B*, 9:44–47, 1977.
- [41] K. Takeuchi. Distribution of informational statistics and a criterion of model fitting. *Mathematical Sciences*, 153:12–18, 1976.
- [42] J.C. Willems, editor. *From Data to Model*. Springer-Verlag, London, UK, 1989.
- [43] K.L. Nylund, T. Asparouhov, and B.O. Muthén. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14:535–569, 2007.
- [44] X. Guyon and J. Yao. On the underfitting and overfitting sets of models chosen by order selection criteria. *Journal of Multivariate Analysis*, 70:221–249, 1997.
- [45] U. Olsson. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44:443–460, 1979.
- [46] Y. Yung. Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62:297–330, 1997.
- [47] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

- [48] N.G. Hoffmann and P.A. Harrison. *SUDDS-IV: Substance Use Disorder Diagnostic Schedule-IV*. New Standards, Inc., St. Paul, MN, 1995.
- [49] P.A. Harrison and N.G. Hoffmann. *SUDDS-IV Administration Guide*. Evinco Clinical Assessments, Smithfield, RI, 1995, 2001.
- [50] N.G. Hoffmann and T.D. Hoffmann. Construct validity for alcohol dependence as indicated by the SUDDS-IV. *Substance Use & Misuse*, 38:293–306, 2003.
- [51] L.K. Muthén and B.O. Muthén. *Mplus User's Guide*. Muthén & Muthén, Los Angeles, CA, fifth edition, 1998-2007.