Addiction
RESEARCH REPORT
doi:10.1111/add.12187

Measurement invariance of DSM-IV alcohol, marijuana and cocaine dependence between community-sampled and clinically overselected studies

Jaime Derringer[1], Robert F. Krueger[2], Danielle M. Dick[3], Arpana Agrawal[4], Kathleen K. Bucholz[4], Tatiana Foroud[5], Richard A. Grucza[4], Michie N. Hesselbrock[6], Victor Hesselbrock[6], John Kramer[7], John I. Nurnberger Jr[5], Marc Schuckit[8], Laura J. Bierut[4], William G. Iacono[2] & Matt McGue[2]

University of Colorado, Boulder, CO, USA,[1] University of Minnesota, Minneapolis, MN, USA,[2] Virginia Commonwealth University, Richmond, VA, USA,[3] Washington University in St. Louis, St Louis, MO, USA,[4] Indiana University, Indianapolis, IN, USA,[5] University of Connecticut, West Hartford, CT, USA,[6] University of Iowa, Iowa City, IA, USA[7] and University of California, San Diego, CA, USA[8]

ABSTRACT

Aims To examine whether DSM-IV symptoms of substance dependence are psychometrically equivalent between existing community-sampled and clinically overselected studies. **Participants** A total of 2476 adult twins born in Minnesota and 4121 unrelated adult participants from a case–control study of alcohol dependence. **Measurements** Life-time DSM-IV alcohol, marijuana and cocaine dependence symptoms and ever use of each substance. **Design** We fitted a hierarchical model to the data, in which ever use and dependence symptoms for each substance were indicators of alcohol, marijuana or cocaine dependence which were, in turn, indicators of a multi-substance dependence factor. We then tested the model for measurement invariance across participant groups, defined by study source and participant sex. **Findings** The hierarchical model fitted well among males and females within each sample [comparative fit index (CFI) > 0.96, Tucker–Lewis index (TLI) > 0.95 and root mean square error of approximation (RMSEA) < 0.04 for all], and a multi-group model demonstrated that model parameters were equivalent across sample- and sex-defined groups (ΔCFI = 0.002 between constrained and unconstrained models). Differences between groups in symptom endorsement rates could be expressed solely as mean differences in the multi-substance dependence factor. **Conclusions** Life-time substance dependence symptoms fitted a dimensional model well. Although clinically overselected participants endorsed more dependence symptoms, on average, than community-sampled participants, the pattern of symptom endorsement was similar across groups. From a measurement perspective, DSM-IV criteria are equally appropriate for describing substance dependence across different sampling methods.

Keywords Item response theory, sampling comparison, sex differences, substance dependence.

Correspondence to: Jaime Derringer, Institute for Behavioral Genetics, University of Colorado, Boulder, CO 80309-0447, USA.
E-mail: jaime.derringer@colorado.edu
Submitted 9 July 2012; initial review completed 24 August 2012; final version accepted 6 March 2013

INTRODUCTION

The structure of substance use and substance use disorders (SUDs—including abuse and dependence) as defined by the Diagnostic and Statistical Manuals (DSMs) [1], both for individual and multiple substances—has been modeled in a range of populations using a variety of measurement techniques. In this study we aim to test whether the structure of substance dependence criteria endorsement across multiple substances is similar between different sampling methods (i.e. community-sampled twins versus clinically overselected individuals) as well as between sexes. Applying diagnostic criteria across a variety of research settings assumes implicitly that these criteria have equivalent measurement properties (such as rank order of symptom endorsement frequencies) when applied to either clinical or epidemiological samples. This is not a minor issue, given that DSM criteria are designed explicitly for application within clinical rather than epidemiological settings, but

often studied in general population and research-based samples.

Many previous studies identified a unidimensional structure of SUDs for individual substances. These findings are robust enough that a unidimensional structure of SUDs will be implemented in the DSM-5 [2]. DSM-5 eliminates the abuse/dependence distinction and describes SUDs as a continuum, by including severity specifiers such as moderate versus severe, rather than simply a categorical diagnosis [3,4]. Previous research strongly supports a unidimensional structure of DSM-IV SUD symptoms within specific substances, including alcohol [5–9], marijuana [8–11] and cocaine [9–11]. Non-DSM measures of problems with alcohol [12–16] and marijuana [17,18] also show evidence of unidimensionality. While many of these studies relied upon self-report inventories, unidimensionality of substance use and problems has also been reported for a combination of self-report problem inventories, with confirmation of use by biological markers of substance use (e.g. saliva or urine samples) for marijuana, cocaine and other illicit substances [19]. This wealth of prior research supports the unidimensionality of substance use and problems within individual substances. We turn now to considering models of multi-substance use, as well as how these models compare across different demographic groups and sampling schemes.

Structure of multiple SUDs

SUD symptoms for individual substances are found consistently to be unidimensional. Further, problems with multiple substances may also represent a single continuum. SUDs tend to correlate positively across a wide range of specific substances. That is, problems with any given substance often predict a common liability to SUDs in general, rather than liability to problems with only a single, specific substance. Certain risk factors contribute to risk of problems with a range of substances. For example, early-adolescent characteristics (such as aggression and delinquency) predict increased rates of SUDs in early adulthood, but do not differentiate between substances. That is, we may expect that a person exposed to these risk factors is at an increased risk of any SUD, but we have limited ability to predict which specific substance(s) will become problematic [20].

The consistently high correlations among SUDs and presence of at least some common risk factors suggest that multiple substances may form a single underlying multi-substance dependence continuum. In a study by Kirisci and colleagues [21], multiple SUD diagnoses fitted a unidimensional model in adult men (oversampled for SUDs), their wives and their adolescent sons. A single dimension also captures ever use of multiple substances, for both males and females, in adolescent [22] and adult samples [23]. These studies demonstrate that the unidimensional structure of multiple substance problems is similar across ages and sexes, although they do not address concerns of generalizing the structure of substance dependence across samples drawn from different populations.

Studies comparing sampling methods

While previous studies examined the issues of age and sex differences (or similarities) in the factor structure of SUDs, there have been far fewer investigations comparing the factor structure of SUDs across multiple sampling schemes. Studies comparing the structure of non-substance characteristics between population and clinical samples found that the same factor structure applies when examining measures of intelligence [24], broad neuropsychological batteries [25] and death distress [26]. Here, clinic-based and population samples are not distinguishable by their measurement structures but, rather, simply differ in their average trait levels. Conversely, a measure of alexithymia (difficulty understanding emotions) displays a slightly different factor structure when comparing clinical and volunteer samples [27].

Few studies have examined how the factor structure of substance use or related problems generalizes across differing sampling schemes. In a comparison of substance use behaviors between volunteer and random samples, volunteers reported a greater incidence of psychosocial risk factors (such as lower average socio-economic status and IQ) and higher rates of SUDs for 'hard' drugs, but the two groups were not distinguishable on the basis of overall number or severity of SUDs. Further, volunteers demonstrated less social desirability bias in self-reports [28]. Comparing drug use problems between Swedish heavy drug users and a population sample, a drug problems inventory fitted a three-factor solution best among the heavy drug-using sample (including substance treatment in-patients and individuals who were either incarcerated or on probation), but a two-factor solution in the population sample. However, the authors of this particular study noted that the low prevalence of substance-related problems in the population sample suggest that factor results may be unreliable in that sample [29].

The current study

In the current analyses, we examine three questions relevant to the conceptualization of substance dependence criteria across multiple substances, in the context of comparing different substances across sexes and type of sample. First, we test whether a hierarchical model describes substance dependence criteria adequately for alcohol, marijuana and cocaine, where substance

dependence symptoms (as well as ever use of each substance) are indicators of individual substance factors which, in turn, are indicators of a multi-substance dependence factor. Secondly, we examine whether this model demonstrates similar model-fit properties when comparing males and females in two independent studies, designed as either community-sampled or clinically overselected for substance dependence. Finally, we test whether differences between samples can be attributed solely to mean differences in a multi-substance dependence latent trait rather than differing factor structures or substance-specific differences.

Applying a factor model to these data to test measurement invariance across sex- and sample-defined groups enables us to test explicitly the assumption that these DSM dependence criteria apply equally well (from a psychometric perspective) to epidemiological and clinically overselected samples. Because research often seeks to generalize findings between different populations, we must establish that DSM criteria measure the same constructs in the same way across different samples. This has not, to our knowledge, been established for DSM substance dependence criteria across clinical and community samples and is the primary aim of the current research.

METHODS

Participants

Minnesota Twin Family Study (MTFS)

The Minnesota Twin Family Study (MTFS) [30,31] includes twins born in Minnesota and recruited to participate in longitudinal assessments. Two cohorts (a younger cohort who entered the study at age 11 years, and an older cohort who began at age 17 years) participated in repeated assessments approximately every 3–4 years around ages 11, 14, 17, 21 and 25 years. We identified all twins assessed for substance dependence symptoms through their 25-year-old follow-up visit. The MTFS sample ($n = 2476$) had a mean age of 25.00 [standard deviation (SD) = 0.89, range = 22–29], was 52.5% female and primarily Caucasian and was representative demographically of the Minnesota population. Through their age 25 assessment, 11.5% of MTFS participants met DSM-IV criteria for alcohol dependence at some point in their life-time, 7.7% met criteria for marijuana dependence and 1.6% met criteria for cocaine dependence.

Study of Addiction: Genetics and Environment (SAGE)

The SAGE [32] is a case–control sample of unrelated individuals overselected for substance dependence. Specifically, the SAGE sample was derived from three primary studies of alcohol, nicotine and cocaine dependence, such that the final sample included approximately 50% alcohol-dependent cases and 50% non-dependent controls. All controls were community-ascertained and did not meet criteria for dependence on any substance. Alcohol-dependent cases drawn from the primary study of nicotine dependence were community-ascertained, while alcohol-dependent cases drawn from the primary studies of alcohol and cocaine dependence were ascertained clinically. All alcohol-dependent cases qualified for inclusion in SAGE, regardless of dependence status on other substances. The SAGE study ($n = 4121$) is 54.3% female, 67.3% Caucasian and 32.5% African American, with 3.4% participants reporting Hispanic ancestry. The SAGE sample has a mean age of 39.03 (SD = 9.10, range = 18–77), with 47.2% participants meeting DSM-IV criteria for alcohol dependence, 18.3% meeting criteria for marijuana dependence and 27.4% meeting criteria for cocaine dependence. All data from SAGE utilized for the current study are publicly available via the database of Genotypes and Phenotypes (dbGaP; phs000092.v1.p1).

Measures

We selected three substances (alcohol, marijuana and cocaine) for analysis because they represent a range of availability, severity and legal statuses, and both the MTFS and SAGE studies assessed these substances in the same manner. For each substance, both studies assessed ever use of the substance and endorsement of each of the seven DSM-IV dependence criteria [in SAGE via the Semi-Structured Assessment for the Genetics of Alcoholism II (SSAGA-II) [33,34]; and in the MTFS via the Composite International Diagnostic Interview Substance–Abuse Module (CIDI-SAM) [35]]. For the MTFS, life-time ever use and dependence symptom endorsement was established by aggregating across all longitudinal assessments through the age 25 assessment. That is, ever use or any substance dependence symptom was considered 'present' if the participant met the criterion at any assessment. Although not pathological, we include the ever use criterion to differentiate individuals exposed to a substance who have no dependence symptoms from those who were never exposed. Substance abuse symptoms were not included in the present analyses, as equivalent abuse criteria were not assessed in both the SAGE and MTFS samples.

Analyses

Based on substantial prior research indicating that (i) substance dependence symptoms are unidimensional within each substance and (ii) multiple substances are indicators of a single latent trait of multi-substance

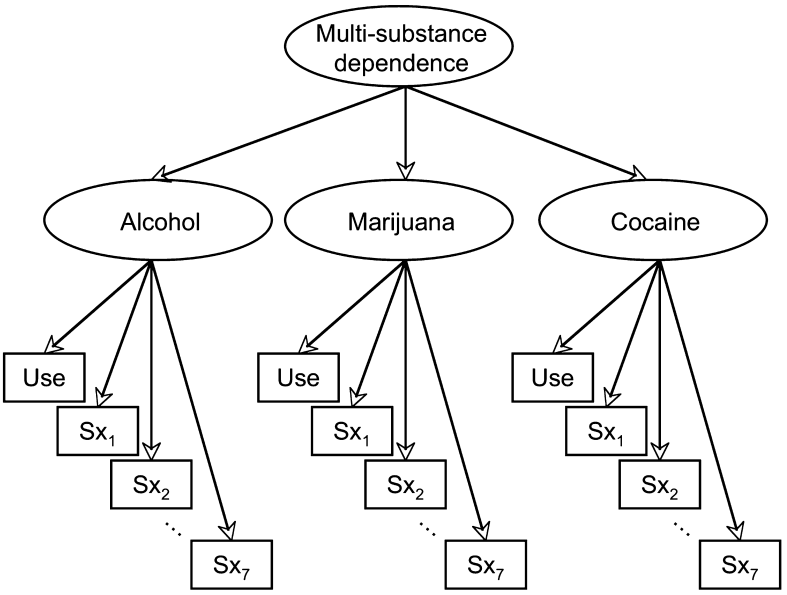

Figure 1 Hierarchical model, in which ever use (Use) and dependence symptoms (denoted Sx_1–Sx_7) are indicators of dependence on each substance, which are in turn indicators of the higher-order multi-substance dependence latent trait

dependence, we fitted a hierarchical confirmatory factor model (depicted in Fig. 1). Within this model, substance-specific dependence symptoms (along with ever use of the substance) were indicators of a factor specific to their respective substance (i.e. alcohol, marijuana or cocaine). The estimation of a continuous factor model of symptoms, rather than restriction to dichotomous diagnoses, makes full use of the available information. Those substance-specific factors were, in turn, indicators of a higher-order multi-substance dependence factor. In modeling dichotomous substance criteria as indicators of a continuous latent trait we estimate threshold parameters, in addition to the loading parameters estimated for both categorical and continuous indicators. Thresholds represent the standardized latent trait level (Z-score) at which the probability of an individual endorsing that criterion is 50%; therefore, higher thresholds represent less frequently endorsed criteria.

We examined model fit in three scenarios. In the first, we modeled all data concurrently, regardless of group membership, where 'group' is defined by sex (male versus female) and study source (MTFS, community-sampled; versus SAGE, clinically overselected). Secondly, we fitted the model separately in each group. From these initial models we evaluated whether the proposed hierarchical factor model fitted the data appropriately within each of the sex- and sample-defined groups, prior to formally testing measurement invariance in a multi-group model.

Next, we examined a multi-group model in which group differences were captured either by differences in criterion parameters (in the unconstrained model) or by mean differences in the multi-substance factor (in the constrained model). In the first multi-group model, the 'unconstrained' model, thresholds (for the categorical observed criteria) and factor loadings were estimated freely within each group, while residual variances and factor means were fixed (at 1 and 0, respectively) in all groups. In the second multi-group model, the 'constrained' model, thresholds (for the categorical observed criteria) and factor loadings were constrained equally across all groups, while residual variances and the multi-substance latent trait mean were fixed (at 1 and 0, respectively) in only the reference group (MTFS males) and estimated freely in all other groups. The comparison of these two models allowed us to evaluate measurement invariance between the different sex and sampling groups. Because only the means (and residual variances) vary among groups in the constrained model it is more parsimonious, but assumes that the substance dependence criteria exhibit measurement invariance across groups. That is, the structure of the substance dependence 'measure' is assumed to be identical across groups, and any group differences in symptom endorsement rates are captured simply by mean differences in the multi-substance dependence latent trait.

We estimated all models in Mplus [36] using a weighted least-squares estimator with theta parameterization. We used clustering (based on family membership) to account for the non-independence of the twin observations within the MTFS sample. We considered absolute model fit in terms of root mean square error of approximation (RMSEA) (with values less than 0.06 indicating a well-fitting model), as well as comparative fit index (CFI) and Tucker–Lewis index (TLI) (with values greater than 0.95 indicating a well-fitting model [37]). We evaluated relative model fit by the difference in CFI values. If the difference in CFI values between constrained and uncon-

strained models was less than 0.01, or if the fit indices were substantially better in the simpler constrained model, we retain the null hypothesis that the factor structure is the same between groups [38].

RESULTS

Table 1 provides the endorsement rates of ever use and the seven dependence symptoms for alcohol, marijuana and cocaine in subgroups defined by sex and sample (that is, MTFS males, MTFS females, SAGE males and SAGE females). For both the MTFS and SAGE samples males endorsed criteria a median of twice as frequently as females, and this pattern was consistent across substances. SAGE participants endorsed dependence symptoms a median of three times as frequently as MTFS participants (comparing across samples within sexes). Between-sample differences in symptom endorsement rates varied more greatly by substance (compared to the relative consistency across substances observed when examining endorsement rates by sex). In particular, cocaine symptoms showed the greatest difference between SAGE and MTFS samples due, in part, to similar disparities in endorsement rates of having ever used cocaine.

Table 2 provides fit statistics for each model applied to the data (or subsets thereof), and correlations among all substance dependence criteria are provided in the Supporting information, Tables S1 and S2. The hierarchical model presented in Fig. 1 fitted well when modeling all data as a single sample. Similarly, model fit was good within each of the individual groups (although the ever use criterion for cocaine was dropped from the model for SAGE males, due to multi-collinearity between this and other criteria that prevented the model from converging on a solution).

When applying the hierarchical model in a multigroup context, both the unconstrained model (in which criterion loadings and thresholds were estimated separately for each group) and the constrained model (in

Table 1 Ever use and dependence symptom endorsement rates by sex- and study-defined groups.

Substance	*Criterion*	*MTFS males (n = 1177)*	*MTFS females (n = 1299)*	*SAGE males (n = 1882)*	*SAGE females (n = 2239)*
Alcohol					
	Ever use	94.3%	88.8%	99.7%	99.6%
	Tolerance	52.9%	21.2%	61.1%	35.0%
	Withdrawal	17.3%	7.0%	33.5%	17.1%
	Using larger amounts or for longer than intended	39.4%	18.5%	72.5%	50.7%
	Persistent desire or unable to cut down	16.4%	8.8%	59.6%	34.7%
	Great amount of time spent obtaining/using/recovering	8.3%	3.0%	37.8%	20.0%
	Activities given up/reduced	4.7%	2.6%	39.4%	18.9%
	Continued use knowing causes physical/psychological problems	17.6%	8.5%	56.9%	36.1%
Marijuana					
	Ever use	60.0%	49.4%	81.8%	73.4%
	Tolerance	14.6%	6.8%	27.6%	11.2%
	Withdrawal	10.0%	4.5%	18.7%	8.5%
	Using larger amounts or for longer than intended	10.9%	4.5%	24.7%	11.9%
	Persistent desire or unable to cut down	11.7%	5.0%	26.1%	12.8%
	Great amount of time spent obtaining/using/recovering	14.5%	5.9%	29.0%	11.4%
	Activities given up/reduced	6.0%	2.3%	19.1%	6.9%
	Continued use knowing causes physical/psychological problems	13.5%	6.5%	21.9%	11.4%
Cocaine					
	Ever use	16.7%	8.7%	55.0%	34.7%
	Tolerance	1.9%	0.5%	29.9%	18.7%
	Withdrawal	2.4%	1.1%	31.4%	19.9%
	Using larger amounts or for longer than intended	2.3%	1.0%	34.5%	20.9%
	Persistent desire or unable to cut down	2.0%	1.0%	33.3%	20.5%
	Great amount of time spent obtaining/using/recovering	2.3%	0.5%	29.8%	18.9%
	Activities given up/reduced	0.7%	0.3%	29.2%	17.9%
	Continued use knowing causes physical/psychological problems	2.3%	1.0%	28.9%	18.5%

MTFS = Minnesota Twin Family Study; SAGE = Study of Addiction: Genetics and Environment.

Table 2 Model fit statistics for the hierarchical model; well-fitting models are those with values of comparative fit index (CFI) and Tucker–Lewis index (TLI) greater than 0.95 and root mean square error of approximation (RMSEA) less than 0.06.

Model	*n*	*CFI*	*TLI*	*RMSEA*	*No. free*
Total sample	6597	0.999	0.999	0.022	51
MTFS males	1177	0.976	0.974	0.033	51
MTFS females	1299	0.961	0.957	0.038	51
SAGE males	1882	0.999	0.999	0.025	49[a]
SAGE females	2239	0.998	0.998	0.021	51
Multi-group	6597				
Unconstrained model		0.998	0.997	0.023	204
Constrained model		0.996	0.996	0.027	135

No. free: number of parameters estimated freely within the model. [a]One criterion (cocaine 'ever use') was removed to identify the single-group model among SAGE males, due to multicollinearlity issues (thus the number of parameters is reduced by one loading and one threshold). MTFS = Minnesota Twin Family Study; SAGE = Study of Addiction: Genetics and Environment.

which group means for multi-substance dependence varied among groups, while holding constant the criterion loadings and thresholds) fitted the data well, as indicated by low RMSEA and high CFI and TLI. The difference in CFI between the constrained and unconstrained models was 0.002 (which is less than the recommended maximum difference cut-off of 0.01 for establishing measurement invariance [38]). Therefore, we conclude that mean differences in the multi-substance dependence latent trait describe adequately criterion-level differences between the groups and proceed with interpreting the results of the constrained model.

Table 3 provides model parameters from the preferred constrained model. Figure 2 illustrates the criterion loadings and thresholds for each individual substance as criterion information curves (see Embretson & Reise [39] for a thorough description of the Item Response Theory framework that defines these plots). Within Fig. 2, criteria that relate more strongly to the substance-specific factor (i.e. with greater loadings) display higher peaks, while criteria endorsed typically by individuals with a more severe level of the substance-specific latent trait are located further to the right (i.e. with higher thresholds, where the *x*-axis is a Z-score metric).

Figure 2 demonstrates visually that, as expected based on endorsement rates, ever use or endorsement of dependence symptoms for cocaine are more severe indicators of substance involvement (with ever use located approximately 2.0 SD above the mean, and dependence symptoms located between 3.0 and 4.0 SD above the mean), compared to either alcohol or marijuana (ever use of either of which is common, even among the community-sampled MTFS males and females). Similarly, as alcohol use and endorsement of dependence symptoms is far more common than endorsement of similar criteria for marijuana across all groups, these plots illustrate that individuals endorsing the same criteria represent a less extreme (that is, more normative) level of alcohol involvement (for which dependence symptoms are located between 0.5 and 2.5 SD above the mean) compared to individuals endorsing similar criteria for marijuana (for which dependence symptoms are located between 2.5 and 3.5 SD above the mean). These patterns reflect the 'dependence liability' of each substance (e.g. alcohol dependence is 'easier' to achieve than marijuana dependence). They do not capture differences in harm or treatment potential that may also vary between substances.

Within the preferred constrained model, mean differences in the multi-substance factor capture entirely the differences in individual substance criteria endorsement between groups. The mean multi-substance latent trait level was –0.57 for MTFS females, 0.00 for MTFS males (fixed within the model as the reference group), 0.51 for SAGE females and 1.30 for SAGE males. These mean differences in the multi-substance factor mirror group differences in criteria endorsement rates (which were consistent across substances, see Table 1). The standardized loadings of all three substance-specific factors on the multi-substance factor were high (ranging from 0.72 to 0.89), indicating that the multi-substance factor accounts for a majority of the variance in ever use and dependence symptom endorsement rates.

DISCUSSION

Based on evidence from the existing substance dependence literature, we identified a hierarchical factor model that provided a good fit to alcohol, marijuana and cocaine dependence symptom data in both a community-sampled and a clinically overselected study. Measurement invariance of the model across samples and between sexes demonstrates that patterns of differences in criterion endorsement rates between these groups are explained primarily by mean differences in a higher-order multi-substance dependence trait, rather than being specific to

Table 3 Criterion loadings (where discrimination is the unstandardized loading) and thresholds [the standardized (Z-score) latent trait level at which an individual is 50% likely to endorse that criterion].

Substance	*Criterion*	*Discrimination*	*Threshold*	*Loading*[a]
Alcohol				
	Ever use	0.29	−6.68	0.38
	Tolerance	0.79	0.90	0.75
	Withdrawal	0.79	2.16	0.75
	Using larger amounts or for longer than intended	0.88	0.59	0.79
	Persistent desire or unable to cut down	0.89	1.33	0.79
	Great amount of time spent obtaining/using/recovering	1.34	2.09	0.89
	Activities given up/reduced	1.72	2.12	0.93
	Continued use knowing causes physical/psychological problems	0.90	1.35	0.79
Marijuana				
	Ever use	0.92	−0.27	0.90
	Tolerance	0.92	2.79	0.90
	Withdrawal	0.97	3.22	0.91
	Using larger amounts or for longer than intended	0.81	2.91	0.87
	Persistent desire or unable to cut down	0.68	2.98	0.83
	Great amount of time spent obtaining/using/recovering	1.38	2.69	0.95
	Activities given up/reduced	0.85	3.40	0.88
	Continued use knowing causes physical/psychological problems	0.74	2.96	0.85
Cocaine				
	Ever use	1.20	2.03	0.89
	Tolerance	1.71	3.49	0.94
	Withdrawal	2.10	3.33	0.96
	Using larger amounts or for longer than intended	1.61	3.24	0.94
	Persistent desire or unable to cut down	1.88	3.29	0.95
	Great amount of time spent obtaining/using/recovering	1.80	3.44	0.95
	Activities given up/reduced	1.46	3.57	0.92
	Continued use knowing causes physical/psychological problems	1.67	3.49	0.94
Multi-substance				
	Alcohol	–	–	0.72
	Marijuana	–	–	0.89
	Cocaine	–	–	0.79

[a]Standardized to Minnesota Twin Family Study (MTFS) males as the reference group.

individual substances or symptom parameters. Our findings support the psychometric validity of combining population and clinical samples in explorations of substance dependence etiology and outcomes.

Limitations

These findings should be considered in light of several limitations, which may impact the generalizability of our

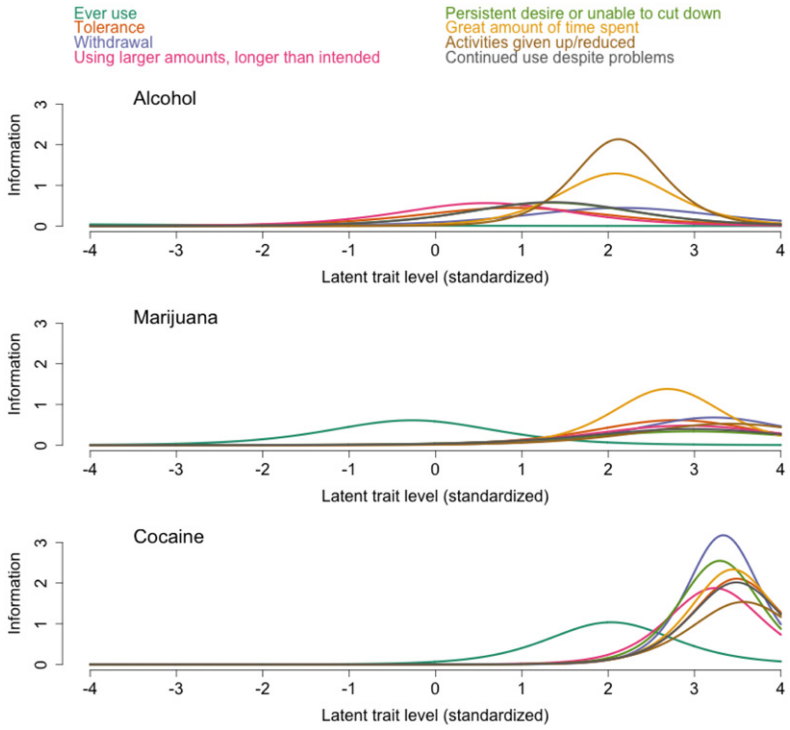

Figure 2 Criterion information curves for each substance (derived from parameters shown in Table 3). The peak height of each criterion's curve represents the relative loading of that criterion. The horizontal location of the peak is the threshold, or the *Z*-score latent trait level at which the likelihood of an individual endorsing that criterion is 50%. Information was calculated as $I(\theta) = (1.7*\alpha_i)^2 P_i(\theta)(1-P_i(\theta))$, where α_i is the normal-metric criterion discrimination (as given in Table 3) and $P_i(\theta)$ is the probability of an individual with latent trait level θ endorsing that criterion [39]

conclusions. Neither of these samples were representative of the US population as a whole, and our model does not include the effects of either ethnicity or age, both of which were represented more diversely in our clinically overselected sample. Previous studies report similar factor structures across ages [21], as well as among African American, European American and Mexican American individuals for alcohol [40] and cocaine dependence [41]. Our finding of measurement invariance in substance dependence across samples in the current study, despite these additional potential confounds not being modeled explicitly, strengthens our conclusion, based on this and previous studies, that the factor structure of substance dependence is consistent across samples with varying demographic characteristics.

We also note that the current findings refer to life-time history of substance use and dependence symptoms and so may not generalize to models of recent substance problems, where comorbidity among substances or even symptoms within a substance may be reduced. Although we include the ever use criterion as a broad indicator of whether or not an individual has tried a substance, there remains a gap in information available in the current study between ever use of a substance and the development of dependence symptoms. Further, there are probably additional key aspects of addiction, e.g. treatment-relevant factors such as social support, that the DSM-IV dependence criteria do not capture.

CONCLUSIONS

The current study provides strong support for the conceptualization of substance dependence as a dimensional construct, both in terms of individual substances and at a higher-order level encompassing a range of multiple substances. Factor loadings indicate that the higher-order multi-substance dependence factor explained 50–80% of the observed variance in each of the substance-specific dependence factors. The construct of multi-substance dependence has been shown previously and consistently to be an indicator of the broader externalizing spectrum, encompassing a range of substance use and misuse measures, non-substance disinhibitory behaviors (including criteria for conduct disorder and antisocial personality disorder), personality constructs (such as impulsivity) and aggression [42]. Although there is substantial overlap between general externalizing and substance-related problems, there remains variance that is unique to substance problems. Environmental aspects may account for some of the unique variation in substance problems, including availability of specific substances or cultural influences on what is considered normative versus problematic use [43].

The current data are unable to address the important question of predictive validity or utility of the multi-substance dependence construct. Future studies should examine whether issues such as clinical course and treatment outcomes may be informed by consistent versus

inconsistent patterns of involvement across multiple substances [44,45]. The finding of measurement invariance across sampling schemes supports the assumption that DSM criteria are psychometrically equally appropriate for describing substance dependence in epidemiological and clinical samples. Although sample means vary, the measurement properties of DSM-IV substance dependence criteria remain fundamentally the same.

Declaration of interests

J.D. was supported by NIH grants DA029377 and MH016880. A.A. was supported by NIH grants DA23668 and DA25886.

Acknowledgements

Funding support for the Minnesota Twin Family Study (MTFS) was provided by NIH grants AA09367 and DA05147. Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01 HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of data sets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392) and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse and the NIH contract 'High throughput genotyping for studying the genetic contributions to human disease' (HHSN268200782096C).

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, text revision (DSM-IV-TR). Washington, DC: American Psychiatric Association; 2000.
2. Hasin D. S. Introduction to DSM-5 criteria linked papers in drug and alcohol dependence. *Drug Alcohol Depend* 2012; **122**: 20–1.
3. Helzer J. E., Bucholz K. K., Gossop M. A dimensional option for the diagnosis of substance dependence in DSM-V. *Int J Methods Psychiatr Res* 2007; **16**: S24–33.
4. O'Brien C. Addiction and dependence in DSM-V. *Addiction* 2011; **106**: 866–7.
5. Dawson D. A., Saha T. D., Grant B. F. A multidimensional assessment of the validity and utility of alcohol use disorder severity as determined by item response theory models. *Drug Alcohol Depend* 2010; **107**: 31–8.
6. Harford T. C., Yi H., Faden V. B., Chen C. M. The dimensionality of DSM-IV alcohol use disorders among adolescent and adult drinkers and symptom patterns by age, gender, and race/ethnicity. *Alcohol Clin Exp Res* 2009; **33**: 1–11.
7. Saha T. D., Chou S. P., Grant B. F. Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychol Med* 2006; **36**: 931–41.
8. Martin C. S., Chung T., Kirisci L., Langenbucher J. W. Item response theory analysis of diagnostic criteria for alcohol and cannabis use disorders in adolescents: implications for DSM-V. *J Abnorm Psychol* 2006; **115**: 807–14.
9. Langenbucher J. W., Labouvie E., Martin C. S., Sanjuan P. M., Bavly L., Kirisci L. *et al*. An application of item response theory analysis to alcohol, cannabis, and cocaine criteria in DSM-IV. *J Abnorm Psychol* 2004; **113**: 72–80.
10. Gillespie N. A., Neale M. C., Prescott C. A., Aggen S. H., Kendler K. S. Factor and item-response analysis DSM-IV criteria for abuse of and dependence on cannabis, cocaine, hallucinogens, sedatives, stimulants and opioids. *Addiction* 2007; **102**: 920–30.
11. Lynskey M. T., Agrawal A. Psychometric properties of DSM assessments of illicit drug abuse and dependence: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Psychol Med* 2007; **37**: 1345–55.
12. Kahler C. W., Strong D. R., Kayaki J., Ramsey S. E., Brown R. A. An item response analysis of the alcohol dependence scale in treatment-seeking alcoholics. *J Stud Alcohol* 2003; **64**: 127–36.
13. Kahler C. W., Strong D. R., Read J. P., Palfai T. P., Wood M. D. Mapping the continuum of alcohol problems in college students: a Rasch model analysis. *Psychol Addict Behav* 2004; **18**: 322–33.
14. Krueger R. F., Nichol P. E., Hicks B. M., Markon K. E., Patrick C. J., Iacono W. G. *et al*. Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychol Assess* 2004; **16**: 107–19.
15. Morey L., Hopwood C. J. An IRT-based measure of alcohol trait severity and the role of traitedness in trait validity: a reanalysis of Project MATCH data. *Drug Alcohol Depend* 2009; **105**: 177–84.
16. Neal D. J., Corbin W. R., Fromme K. Measurement of alcohol-related consequences among high school and college students: application of item response models to the Rutgers Alcohol Problems Index. *Psychol Assess* 2006; **18**: 402–14.
17. Annaheim B., Scotto T. J., Gmel G. Revising the Cannabis Use Disorders Identification Test (CUDIT) by means of Item Response Theory. *Int J Methods Psychiatr Res* 2010; **19**: 142–55.
18. Piontek D., Kraus L., Klempova D. Short scales to assess cannabis-related problems: a review of psychometric properties. *Subst Abuse Treat Prev Policy* 2008; **3**: 25. doi: 10.1186/1747-597X-3-25.
19. Lennox R., Dennis M. L., Scott C. K., Funk R. Combining psychometric and biometric measures of substance use. *Drug Alcohol Depend* 2006; **83**: 95–103.
20. Hayatbakhsh M. R., Najman J. M., Jamrozik K., Al Mamun A., Bor W., Alati R. Adolescent problem behaviours predicting DSM-IV diagnoses of multiple substance use disorder:

findings of a prospective birth cohort study. *Soc Psychiatry Psychiatr Epidemiol* 2008; **43**: 356–63.
21. Kirisci L., Tarter R. E., Vanyukov M., Martin C., Mezzich A., Brown S. Application of item response theory to quantify substance use disorder severity. *Addict Behav* 2006; **31**: 1035–49.
22. Derringer J., Krueger R. F., Iacono W. G., McGue M. Modeling the impact of age and sex on a dimension of poly-substance use in adolescence: a longitudinal study from 11- to 17-years-old. *Drug Alcohol Depend* 2010; **110**: 193–9.
23. Kirisci L., Vanyukov M., Dunn M., Tarter R. Item Response Theory models of substance use: an index based on 10 drug categories. *Psychol Addict Behav* 2002; **16**: 290–8.
24. Burton D. B., Ryan J. J., Axelrod B. N., Schellenberger T. A confirmatory factor analysis of the WAIS-III in a clinical sample with crossvalidation in the standardization sample. *Arch Clin Neuropsychol* 2002; **17**: 371–87.
25. Dickinson D., Ragland J. D., Calkins M. E., Gold J. M., Gur R. C. A comparison of cognitive structure in schizophrenia patients and healthy controls using confirmatory factor analysis. *Schizophr Res* 2006; **85**: 20–9.
26. Abdel-Khalek A. M. A general factor of death distress in seven clinical and non-clinical groups. *Death Stud* 2004; **28**: 889–98.
27. Müller J., Bühner M., Ellgring H. Is there a reliable factorial structure in the 20-item Toronto Alexithymia Scale? A comparison of factor models in clinical and normal adult sample. *J Psychosom Res* 2003; **55**: 561–8.
28. Reynolds M. D., Tarter R. E., Kirisci L. Characteristics of men with substance use disorder consequent to illicit drug use: comparison of a random sample and volunteers. *Drug Alcohol Depend* 2004; **75**: 241–51.
29. Berman A. H., Bergman H., Palmstierna T., Schlyter F. Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. *Eur Addict Res* 2005; **11**: 22–31.
30. Iacono W. G., McGue M. Minnesota Twin Family Study. *Twin Res Hum Genet* 2002; **5**: 482–7.
31. Iacono W. G., McGue M., Krueger R. F. Minnesota Center for Twin and Family Research. *Twin Res Hum Genet* 2006; **9**: 978–84.
32. Bierut L. J., Madden P. A. F., Breslau N., Johnson E. O., Hatsukami D., Pomerleau O. F. *et al*. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007; **16**: 24–35.
33. Bucholz K. K., Cadoret R., Cloninger C. R., Dinwiddie S. H., Hesselbrock V. M., Nurnberger J. I. Jr *et al*. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability for the SSAGA. *J Stud Alcohol* 1994; **55**: 149–58.
34. Hesselbrock M., Easton C., Bucholz K. K., Schuckit M., Hesselbrock V. A validity study of the SSAGA—a comparison with the SCAN. *Addiction* 1999; **94**: 1361–70.
35. Cottler L. B. Composite International Diagnostic Interview—Substance Abuse Module (SAM). St Louis, MO: Department of Psychiatry, Washington University School of Medicine. 2000.
36. Muthén L. K., Muthén B. O. *Mplus User's Guide*, 6th edn. Los Angeles, CA: Muthén & Muthén; 1998–2010.
37. Hu L., Bentler P. M. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999; **6**: 1–55.
38. Cheung G. W., Rensvold R. B. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Modeling* 2002; **9**: 233–55.
39. Embretson S. E., Reise S. P. *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
40. Caetano R., Schafer J. DSM-IV alcohol dependence in a treatment sample of white, black, and Mexican-American men. *Alcohol Clin Exp Res* 1996; **20**: 384–90.
41. Schafer J., Caetano R. The DSM-IV construct of cocaine dependence in a treatment sample of black, Mexican American, and white men. *Psychol Assess* 1996; **8**: 304–11.
42. Krueger R. F., South S. C. Externalizing disorders: cluster 5 of the proposed meta-structure for DSM-V and ICD-11. *Psychol Med* 2009; **39**: 2061–70.
43. Degenhardt L., Chiu W. T., Sampson N., Kessler R. C., Anthony J. C., Angermeyer M. *et al*. Toward a global view of alcohol, tobacco, cannabis, and cocaine use: findings from the WHO World Mental Health Surveys. *PLoS Med* 2008; **5**: e141.
44. Chung T., Martin C. S., Clark D. B. Concurrent change in alcohol and drug problems among treated adolescents over three years. *J Stud Alcohol Drugs* 2008; **69**: 420–9.
45. Kahler C. W., Strong D. R., Papandonatos G. D., Colby S. M., Clark M. A., Boergers J. *et al*. Cigarette smoking and the lifetime alcohol involvement continuum. *Drug Alcohol Depend* 2008; **93**: 111–20.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1 Spearman's correlations among substance use and dependence symptoms. Correlations in the Minnesota Twin Family Study (MTFS) male sample ($n = 1177$) are below the diagonal; correlations for MTFS females ($n = 1299$) are above the diagonal. Correlations greater than 0.5 are shown in bold type.

Table S2 Spearman's correlations among substance use and dependence symptoms. Correlations in the Study of Addiction: Genetics and Environment (SAGE) male sample ($n = 1882$) are below the diagonal; correlations for SAGE females ($n = 2239$) are above the diagonal. Correlations greater than 0.5 are shown in bold type.