Original Investigation

Exploring the role of a nicotine quantity–frequency use criterion in the classification of nicotine dependence and the stability of a nicotine dependence continuum over time

Orla McBride, Ph.D.,[1,2] *David R. Strong, Ph.D.,*[3] *& Christopher W. Kahler, Ph.D.*[4]

[1] *School of Psychology, University of Ulster Magee Campus, Co. Londonderry, Northern Ireland*
[2] *Department of Epidemiology, Michigan State University, East Lansing, MI*
[3] *The Warren Albert Medical School of Brown University, Providence, RI*
[4] *Center for Alcohol and Addiction Studies, Brown University, Providence, RI*

Corresponding Author: Orla McBride, Ph.D., *School of Psychology, University of Ulster at Magee Campus, Northland Road, Co. Londonderry BT48 7JL, Northern Ireland. Telephone: +44 (0) 2871375367; Fax: +44 (0) 2871375315; E-mail: o.mcbride@ulster.ac.uk*

Received September 9, 2009; accepted November 20, 2009

Abstract

Introduction: This study investigated (a) the utility of a cigarette quantity–frequency (QF) use criterion as an indicator for nicotine dependence (ND) and (b) the stability of the ND continuum of severity over time.

Method: Data from individuals who smoked cigarettes in the year prior to both time points of the National Epidemiologic Survey on Alcohol and Related Conditions were analyzed (n = 6,185). The Alcohol Use Disorder and Associated Disabilities Interview Schedule *DSM-IV* Version (AUDADIS-IV) assessed for *DSM-IV* ND and nicotine use. Three QF criteria were created to represent daily consumption of ≥5 cigarettes, ≥10 cigarettes, or ≥20 cigarettes. Confirmatory factor analysis and item response theory analysis were used to explore the latent structure of ND. Differential item functioning (DIF) analysis investigated the stability of the ND continuum over time.

Results: A one-factor model, representing the *DSM-IV* conceptualization of ND, was an acceptable fit to the data at both time points. The inclusion of QF criteria decreased the fit of the one-factor model of ND. DIF in the severity and discrimination parameters of the diagnostic criteria was evident across the time points of the survey.

Discussion: Although QF of cigarette use is related to ND, it appears to be a separate construct. Researchers using the AUDADIS-IV should be aware that the characteristics of the *DSM-IV* ND criteria do vary slightly across time, even though the changes appear to be relatively small and of minor clinical or practical significance.

Introduction

Tobacco use is a major cause of premature and preventable death in the United States (Centers for Disease Control and Prevention, 2008). It is estimated that of the 24.8% of American adults who are current smokers, 12.8% are nicotine dependent (Grant, Hasin, Chou, Stinson, & Dawson, 2004). A concise definition of nicotine dependence (ND) and accurate assessments of the disorder over time are critical for assessing the effectiveness of prevention, intervention, and treatment efforts designed to reduce levels of nicotine use and dependence (Lecrubier, 2008).

The concept of ND has been defined broadly in the literature. Several questionnaires exist that purport to measure ND, for example, the Nicotine Dependence Severity Scale (Shiffman, Waters, & Hickcox, 2004) and the Fagerström Test for Nicotine Dependence (Heatherton, Kozlowski, Frecker, & Fagerström, 1991; see Piper, McCarthy, & Baker, 2006 for review). These instruments may offer advantages in operationalization and a theoretical assessment of dependence that is based upon nicotine specifically rather than traditional psychiatric classification systems, such as the *DSM-IV* (American Psychiatric Association, 1994), which cover all drugs (Hendricks, Prochaska, Humfleet, & Hall, 2008). Many researchers, however, consider the *DSM-IV* to be a "gold standard" for the classification of substance use disorders (SUD), including ND (Colby, Tiffany, Shiffman, & Niaura, 2000; Saunders & Cottler, 2007). The *DSM-IV* outlines seven diagnostic criteria to conceptualize ND based on the theoretical work of Edwards and Gross (1976): TOLERANCE, WITHDRAWAL, using larger amounts or for longer than intended (LARGER), more than once trying to stop or cut down use (CUT DOWN), spending a great deal of time using (TIME

doi: 10.1093/ntr/ntp196
Advance Access published on January 15, 2010

SPENT), giving up or cutting down on important activities (GIVE UP), and continued use despite physical health or psychological problems (CONTINUED USE). According to the *DSM-IV*, a diagnosis of ND is warranted when an individual experiences any three or more of the seven diagnostic criteria.

Several studies have investigated the performance of the *DSM-IV* ND criteria. Using item response theory (IRT) analysis, Strong, Kahler, Ramsey, and Brown (2003) analyzed a set of seven symptom questions administered in the National Comorbidity Survey (Kessler et al., 1994), which served as a proxy for the *DSM-IV* ND criteria. Results revealed that symptoms tapped an underlying continuum of ND and distinguished among three levels of severity, namely mild, moderate, and severe dependence. TOLERANCE and LARGER were useful indicators at the mild end of the continuum, whereas GIVE UP and CONTINUED USE were indicative of more severe ND. Similar findings have also been reported elsewhere (Muthén & Asparouhov, 2006; Strong, Kahler, Colby, Griesler, & Kandel, 2009).

The *DSM-IV* does not include an indicator of drug use as a diagnostic criterion for ND or any other drug use disorder. It has been suggested recently that the ability of the *DSM* criteria to discriminate among individuals with different levels of ND severity may be enhanced with the addition of a tobacco quantity–frequency (QF) use indicator (Hendricks et al., 2008). This is consistent with research demonstrating that a QF criterion may be an important element of both alcohol and cannabis use disorder continuums (Compton, Saha, Conway, & Grant, 2009; Saha, Stinson, & Grant, 2007). Previous research documenting the association between nicotine use and ND has produced inconsistent findings. For example, the number of cigarettes smoked daily is reported to be a strong predictor of treatment outcome (Breslau & Johnson, 2000; Hendricks et al.), suggesting that cigarette consumption may be a sensitive indicator of ND. Other studies, however, have revealed moderately high levels of ND among some low-use smokers and relative low levels of ND among some regular cigarette smokers (Dierker et al., 2007; Donny & Dierker, 2007; Donny, Griffin, Shiffman, & Sayette, 2008; Shiffman, 1989). These findings suggest that the relationship between nicotine use and dependence may be weaker than expected. One useful method to explore this relationship further is to introduce a smoking QF indicator into a factor model of ND. The strength of the factor loading and goodness of the model fit can determine whether such a criterion is a useful symptom of dependence or a qualitatively and quantitatively distinct construct. Although this method has proven fruitful for studies investigating this issue for other drugs (Compton et al.; Saha et al.), similar studies for ND have not been conducted.

In addition to exploring the relationship between nicotine use and the construct of ND, researchers are also interested in investigating the course of ND and whether specific interventions can reduce levels of dependence over time (Fiore, Hatsukami, & Baker, 2002; Foulds, Burke, Steinberg, Williams, & Ziedonis, 2004). Instruments used to measure ND in these instances must have sound psychometric properties. One fundamental, but often overlooked, aspect of psychometrics is the issue of measurement invariance over time (Reise, Widaman, & Pugh, 1993). If the psychometric properties of a measurement of ND change over time, this can affect the quality of longitudinal survey data. Specifically, differences in instrument scores over time can be attributed to differences in level of the underlying construct only if the measurement occasions are psychometrically equivalent or invariant (Byrne, Shavelson, & Muthén, 1989; Meredith & Horn, 2001). Despite the importance of this issue, no longitudinal study to date has explored how the *DSM-IV* ND criteria function over time in the general population. Thus, even though the *DSM-IV* criteria appear to map onto a continuum of ND severity, it is unknown whether these criteria function similarly over time. It is important that the relationship between the *DSM-IV* criteria and the construct of ND remains stable over time, even though individuals who are assessed at different time points may change in their level of severity (Strong et al., 2007; Wood, Kerr, & Brink, 2006).

This study was devised to address the gaps in the literature described above. Specifically, data from the longitudinal National Epidemiologic Survey on Alcohol and Related Conditions (NESARC; Grant & Kaplan, 2005) were analyzed to explore (a) the utility of a cigarette QF use criterion for evaluating a continuum of ND severity and (b) the stability of a measurement of *DSM-IV* ND continuum over a 3-year period.

Method

Survey

The NESARC is a nationally representative longitudinal household survey of civilian noninstitutionalized adults, living in the United States (Grant & Kaplan, 2005; Grant, Kaplan, Shepard, & Moore, 2003). The baseline survey (herein referred to as "Time 1") was conducted during 2001–2002. Face-to-face interviews were conducted with 43,093 individuals (81.2% response rate; Grant, Dawson, et al., 2003). Approximately 39,959 participants were eligible to be reinterviewed (e.g., had not died, left the country, become incapacitated, institutionalized) at the follow-up survey, which was conducted during 2004–2005 (herein referred to as "Time 2"). A total of 34,653 individuals were reinterviewed, representing a response rate of 86.7%. The overall response rate for the NESARC was 70.2% (Grant & Kaplan). The NESARC used the Census 2000–2001 Supplementary Survey (C2SS) as a sampling frame. For both surveys, the data were weighted to account for elements of the survey design, including primary sampling unit (PSU) selection probabilities, within-PSU selection probabilities, and nonresponse in the C2SS. Time 2 data were also weighted to account for nonresponse in relation to sociodemographic variables and Time 1 lifetime psychiatric disorders. Weighted data were adjusted to be representative of the U.S. general population for a variety of sociodemographic variables. Comprehensive details on the sampling frame, interviewer training, and field quality control are available elsewhere (Grant & Kaplan).

Measure

Twenty-two binary symptom questions from the Alcohol Use Disorder and Associated Disabilities Interview Schedule *DSM-IV* Version (AUDADIS-IV; Grant, Dawson, & Hasin, 2001) operationalized the *DSM-IV* ND criteria for each survey. All individuals who reported ever smoking 100+ cigarettes or 50+ cigars, using a pipe 50+ times, snuff 20+ times, or chewing tobacco 20+ times were asked whether they had experienced any of the symptom questions for ND during the last 12 months and/or prior to the last 12 months. The reliability and validity of the ND criteria in the AUDADIS-IV are good in the general population, with intraclass correlations (ICC) ranging from .75 to .76 and the κ value at .63 (Grant, Dawson, et al., 2003).

Three binary coded QF criteria, similar to the other diagnostic criteria, were created to represent the daily use of ≥5 cigarettes, ≥10 cigarettes, and ≥20 cigarettes. The nicotine use module in the AUDADIS-IV also has good reliability (ICC = .74–.84; Grant, Dawson, et al., 2003).

Sample

All participants in the NESARC were classified as current tobacco users, ex–tobacco users, or lifetime nonsmokers at both time points. Current and ex–tobacco users were categorized further into smoking status by specific tobacco product (i.e., cigarettes, cigars, pipe, snuff, or chewing tobacco): (a) smoked or used product in the last 12 months and (b) smoked or used product prior to the last 12 months. For the purposes of this analysis, only data from current cigarette users, defined as smoking cigarettes in the year prior to Time 1 and Time 2, were analyzed (n = 6,185).

Details reported in this section are based on weighted data at Time 1. The mean age of cigarette smoking onset was 15.70 years (SD = 4.358). Approximately 91% of smokers in this sample reported smoking cigarettes on a daily basis, and the mean usual quantity consumed was 17.19 cigarettes (SD = 12.112). The mean age of the sample was 44.53 years (SD = 17.45). The majority of the sample were male (51.2%), married or cohabiting (61.4%), in full-time employment (54%), living in urban areas (52.2%), had a high school or some college education (49.5%), and earned ≤ $20,000 per annum. Individuals were classified as White (70.8%), Black (11.2%), Hispanic (12.5%), American Indian (1.8%), and Asian (3.8%). A comparison of Time 2 smokers, who were followed up at Time 2 compared with those who were not, revealed that the latter group were more likely to be male, Hispanic, or Asian (compared with White); never married or widowed (compared with married/cohabiting); and unemployed.

Statistical analyses

Confirmatory factor analysis

Initially, four confirmatory one-factor models were specified and estimated using Mplus version 5.1 (Muthén & Muthén, 1998–2007). Model 1 represented the *DSM-IV* conceptualization of ND. Models 2–4 extended Model 1 by including a different QF criterion in each model: Model 2 (≥5 cigarettes/day), Model 3 (≥10 cigarettes/day), and Model 4 (≥20 cigarettes/day). Several goodness-of-fit indices were used to compare the factor models: the chi-square, the comparative fit index (CFI; Bentler, 1990), the Tucker–Lewis index (TLI; Tucker & Lewis, 1973), and the root mean square error of approximation (RMSEA; Steiger, 1990). Hu and Bentler (1999) recommend that a good model fit be indicated by a nonsignificant chi-square test, TLI, CFI values of ≥0.95, and RMSEA value of ≤0.05. The best fitting and most conceptually sound model would be used in all subsequent analyses.

Fit to the IRT model

Mplus computes IRT parameters (Birnbaum, 1968) by default for a one-factor CFA model. The two important parameters in an IRT model are (a) item severity and (b) item discrimination. The severity of an item is the point along the latent continuum of dependence at which the item has a 50% probability of being endorsed (Kahler & Strong, 2006). The severity of an item provides an estimation of the degree of dependence severity that is required for a specific item to be experienced (Saha et al., 2007). The discrimination of an item explains how rapidly the probability of observing the item changes across increasing levels of the latent continuum of ND. The discrimination indicates the degree of precision with which an item can distinguish between participants above and below an item's severity threshold (Hartman et al., 2008).

Two IRT models were tested to explore the association between the observed responses to the diagnostic criteria and the underlying continuum of ND severity: (a) a one-parameter logistic or "Rasch" model (Rasch, 1960) and (b) a two-parameter logistic model (Lord, 1980). The Rasch model proposes that each item has a similar ability to discriminate among individuals in a sample. The discrimination parameters are constrained to be equal for all items, whereas severity parameters are estimated for each item. The two-parameter model relaxes the assumptions of the one-parameter model, whereby the discrimination and severity parameters are estimated for each item underlying the latent continuum of ND (Langenbucher et al., 2004). The one-parameter model is more parsimonious when compared with the two-parameter model. A chi-square difference test can help determine which model is a better explanation of the data. The above analyses employed a robust weighted least squares estimator. The data were weighted, clustered on PSUs, and stratified appropriately to allow generalizability to the U.S. population.

Differential item functioning

Differences in the estimates of the criteria characteristics obtained from Time 1 and Time 2 were evaluated to assess the stability of the IRT model over time. A differential item functioning (DIF) approach that permits the use of model-based evaluations that utilize information about the measurement properties of the set of diagnostic criteria simultaneously across the time points to generate a posterior distribution of ND severity (cf. Strong et al., 2009) was conducted using the software IRTLRDIF version 2.0 (Thissen, 2001). A likelihood ratio test was used to provide a significance test for the null hypothesis that the criterion parameters do not differ across the two time points of the survey (Thissen, Pommerich, Billeaud, & Williams, 1995). Analyses were conducted iteratively to determine which criteria function differently across the time points and which criteria are DIF free (cf. Strong et al., 2009). To explore for DIF, the discrimination and severity estimates for each time point are constrained to be equal across all seven criteria (Model A). For each criterion, another model is estimated that permits the discrimination and severity estimates for that criterion to differ across the time points and constrains the discrimination and severity estimates of all the remaining criteria to be equal (Model B). The difference in the log-likelihood values (ll) of Models A and B [$G^2 = -2$ (ll Model A − ll Model B)] provides an omnibus test (df = 2) of whether there is DIF for the discrimination, severity, or both parameters for this criteria (Hussong, Flora, Curran, Chassin, & Zucker, 2008; Strong et al., 2009). Significant values can be followed up by additional tests (1 df) to investigate whether the DIF is present in the discrimination or severity estimates. When conducting DIF analyses involving 1 df tests, it is necessary to control for the statistical possibility of making false conclusions using the Benjamini and Hochberg procedure (Benjamini & Hochberg, 1995; Thissen, Steinberg, & Kuang, 2002).

Given the large sample sizes, small differences in severity between the time points could be statistically significant but may not be conceptually meaningful or of sufficient magnitude to affect the interpretation of scores over time (cf. Strong et al., 2009). Thus, it was decided a priori that only differences of ≥0.25 in symptom severity, which can be interpreted as one quarter of the "standard unit difference between the value of the (underlying) trait necessary to have a 50–50 chance of responding positively in one group compared to another" (pp. 405–6; Steinberg & Thissen, 2006), would be considered as clinically meaningful. In the absence of a similar metric to consider differences among the discrimination parameters, the item response curves (IRCs) of statistically significant discrimination parameters were visually inspected for clinical significance (Steinberg & Thissen).

Results

Descriptive statistics for *DSM-IV* ND diagnostic criteria

Table 1 presents the endorsement rates for the *DSM-IV* ND diagnostic criteria for the sample at Time 1 and Time 2. WITHDRAWAL was the most commonly endorsed criterion at both time points, whereas GIVE UP was the least commonly endorsed criterion.

Unidimensionality

Table 2 displays the standardized factor loadings and model fit indices for the CFA.

The significant chi-square values suggest that none of the factor models were good fitting models. Bollen (1989), however, noted that the chi-square statistic is highly sensitive to large sample sizes and may overestimate the lack of fit of a structural model. Inspection of the other fit indices revealed that Model 1 was the best fitting model across the two surveys, with moderate–strong, positive, and statistically significant factor loadings. Although the factor loadings for the QF criterion in Model 2 (≥10 cigarettes/day) at the two time points were moderate, the factor loadings for the QF criteria in the other models were poor. Importantly, the inclusion of a QF criterion in Models 2–4 resulted in a decreased model fit compared with Model 1. Therefore, the QF criteria were not considered in additional analyses.

IRT model selection

Model 1 was estimated as a one-parameter and a two-parameter IRT model. The chi-square difference test revealed that the two-parameter model was a superior explanation of the data (Time 1: χ^2 diff = 107.677, *df* = 4, *p* < .000 and Time 2: χ^2 diff = 84.330, *df* = 4, *p* < .000). Examination of the IRCs from both models (not presented) revealed that there was some crossover among the curves in the two-parameter model, suggesting that there was meaningful variation among the discrimination parameters.

DIF across the NESARC time points

Table 3 displays the severity and parameter estimates for the diagnostic criteria across the surveys. Positive values reflect those criteria likely to be endorsed among individuals with higher than average levels of ND. Alternatively, negative values represent the criteria that are more likely to be endorsed among respondents with lower than average levels of ND. GIVE UP and WITHDRAWAL reflected the highest and lowest levels of ND, respectively.

Severity parameters

Four diagnostic criteria exceeded the a priori criteria for clinical and statistical significance. Given the same level of ND severity, respondents were less likely to endorse WITHDRAWAL and TOLERANCE at Time 2 compared with Time 1. Alternatively, GIVE UP and CONTINUED USE were more likely to be endorsed at Time 2 compared with Time 1. The severity estimates for the two time points were strongly related (Spearman rank-order correlation coefficient = .964, *p* < .001), indicating high overall stability. The rank order of the severity estimates for all the criteria, except for WITHDRAWAL and CUT DOWN, remained stable across the time points. WITHDRAWAL was the least severe criterion at Time 1 and second least severe at Time 2. CUT DOWN was the second least severe criterion in Time 1 and the least severe at Time 2.

Discrimination parameters

Figure 1 represents the IRCs for the diagnostic criteria across the surveys. For WITHDRAWAL, the DIF was nonuniform, meaning that statistically significant DIF was also evident in the

Table 1. Positive endorsement rates for *DSM-IV* ND and quantity–frequency use criteria among current smokers in the NESARC (*n* = 6,185)

	Time 1 (2001–2002 survey)		Time 2 (2004–2005 survey)	
DSM-IV ND criteria	*N* (unweighted)	% (weighted)	*N*	%
WITHDRAWAL	4,633	77.2	4,818	79.6
TOLERANCE	1,075	17.8	711	10.9
GIVE UP	543	9.3	655	10.1
CUTDOWN	4,010	65.7	4,716	75.7
TIME SPENT	1,426	24.6	1,250	19.9
CONTINUED USE	3,289	55.4	4,139	67.7
LARGER/LONGER	1,616	27.2	1,898	30.8

Note. ND = nicotine dependence; NESARC = National Epidemiologic Survey on Alcohol and Related Conditions.

Table 2. Standardized factor loadings for *DSM-IV* nicotine dependence and QF use criteria at Time 1 and Time 2 (*n* = 6,185)

	Standardized factor loadings							
	Time 1 (2001–2002 survey)				Time 2 (2004–2005 survey)			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
WITHDRAWAL	0.684	0.696	0.763	0.727	0.586	0.621	0.741	0.609
TOLERANCE	0.774	0.769	0.762	0.770	0.714	0.708	0.684	0.711
GIVE UP	0.524	0.525	0.527	0.534	0.440	0.432	0.424	0.457
CUT DOWN	0.585	0.582	0.551	0.559	0.521	0.511	0.465	0.491
TIME SPENT	0.671	0.672	0.676	0.684	0.662	0.663	0.657	0.683
CONTINUED USE	0.691	0.692	0.681	0.690	0.602	0.599	0.596	0.601
LARGER/LONGER	0.803	0.800	0.779	0.780	0.728	0.720	0.680	0.709
QF (≥5 cigarettes/day)	—	0.249	—	—	—	0.351	—	—
QF (≥10 cigarettes/day)	—	—	0.443	—	—	—	0.508	—
QF (≥20 cigarettes/day)	—	—	—	0.357	—	—	—	0.292
Fit indices								
Chi-square	151.022	226.738	458.044	378.975	167.457	251.674	509.132	308.632
df	12	17	16	17	12	17	16	17
p	.0000	.0000	.0000	.000	.0000	.0000	.0000	.0000
CFI	0.972	0.957	0.919	0.931	0.941	0.913	0.843	0.895
TLI	0.967	0.954	0.909	0.927	0.931	0.908	0.833	0.889
RMSEA	0.032	0.033	0.050	0.044	0.040	0.042	0.062	0.047

Note. All factor loadings significant at $p < .001$. CFI = comparative fit index; QF = quantity–frequency; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index.

discrimination parameter. Two other statistically significant differences were evident: TIME SPENT and LARGER were more discriminating at Time 2 compared with Time 1.

Test information curve

The IRCs can be summed to produce an information curve for the full scale, which is referred to as the test information curve (TIC). The TIC represents the relative precision of the scale across different levels of the trait continuum (Fraley, Waller, & Brennan, 2000). The TIC is presented in the lower right-hand corner of Figure 1. For both time points, the continuum provided most information for individuals with moderate levels of dependence but less information for individuals with mild or severe dependence. The higher discrimination values for the criteria in Time 2 compared with Time 1 are reflected in the more “peaked” TIC for that time point.

DIF between age groups

Additional analyses were conducted to uncover a possible explanation for the measurement noninvariance in the ND criteria

Table 3. Differential item functioning, and the severity and discrimination parameters, of *DSM-IV* nicotine dependence criteria across the 2001–2002 and the 2004–2005 NESARC time points

		Severity parameter			Discrimination parameter			Time 2	
DSM-IV criteria	G^2 ($df = 2$)	Time 1	Time 2	Difference	Time 1	Time 2	Difference	Mean	*SD*
WITHDRAWAL	70.4	−1.00	−0.63	**0.37***	1.65	2.03	0.38*	0.21	0.70
TOLERANCE	147.5	1.17	1.45	**0.28***	2.25	2.63	0.38	0.22	0.78
GIVE UP	12.6	2.56	2.21	**-0.35***	1.09	1.19	0.10	0.18	0.74
CUT DOWN	59.6	−0.63	−0.79	−0.16*	1.30	1.53	0.23	0.17	0.75
TIME SPENT	64.7	1.07	1.10	0.03*	1.58	2.17	0.59*	0.20	0.73
CONTINUED USE	110.9	−0.11	−0.37	**−0.26***	1.63	1.85	0.22	0.15	0.75
LARGER	39.0	0.80	0.62	−0.18*	2.27	3.04	0.77*	0.16	0.71

Note. The G^2 test with 2 *df* evaluates differences between the waves in both severity and discrimination parameters. Differences between the waves on either parameter are evaluated using 1 *df* tests. *p* Values for *df* 1 tests were adjusted using the Benjamini–Hochberg procedure. Severity parameters that (a) represent a statistically significant difference between waves and (b) exceed our effect size criteria (0.25) are shown in bold. NESARC = National Epidemiologic Survey on Alcohol and Related Conditions.

**p* < .05.

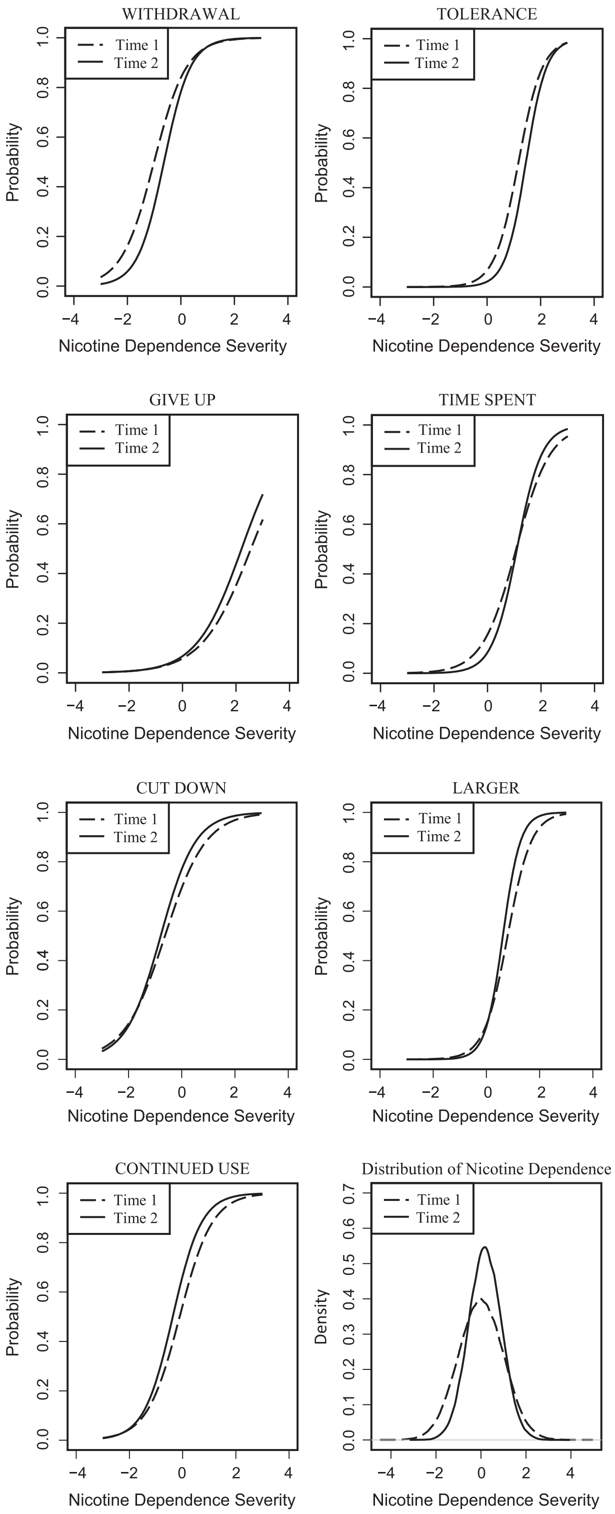

Figure 1. Plot of the item characteristic curves and total information curve for the seven *DSM-IV* nicotine dependence criteria for both National Epidemiologic Survey on Alcohol and Related Conditions time points.

across the surveys. One plausible hypothesis is that the noninvariance in the criteria over time could be a function of the age of the participants. Specifically, the individual changes in levels of ND that would be expected over time might be more concentrated in younger compared with older individuals. Age, therefore, may be related to individual changes in levels of ND, even though changes in levels of ND should not impact the relative severity or discrimination of the diagnostic criteria. To investigate this issue, data from Time 1 were used to explore for DIF in the ND criteria among younger adults (18–29 years) compared with older adults (≥30 years). DIF between the age groups was evident for several of the criteria (see Table 4). Given the same level of ND, younger adults were less likely than older adults to endorse WITHDRAWAL, GIVE UP, and CONTINUED USE. Conversely, older individuals were more likely to endorse CUT DOWN and TIME SPENT. The differences between the two age groups were in the small to medium effect size range, except GIVE UP, which had a very large effect size. Only LARGER exhibited DIF in the discrimination parameter, meaning that this criterion was more discriminating among older compared with younger adults.

Discussion

This study formally tested the utility of a cigarette QF indicator as a diagnostic criterion for ND and assessed the stability of an ND continuum of severity, as measured by the AUDADIS-IV, over a 3-year period. Consistent with previous research (Strong et al., 2003, 2009), the *DSM-IV* ND diagnostic criteria tapped into a continuum of severity at both time points of the NESARC. Of particular interest was the finding that none of the cigarette QF criteria were useful indicators of the underlying continuum. This suggests that QF of cigarette use do not increase the precision of the AUDADIS-IV as an instrument for assessing ND. This finding is not entirely unexpected, given the poor concordance between the FTND, which assesses other facets of ND such as smoking heaviness, and the *DSM-IV* (Hughes, 2006; Moolchan et al., 2002). The current findings somewhat contradict research advocating the inclusion of a QF diagnostic criterion in the future classification of SUD (Compton et al., 2009; Saha et al., 2007). They are consistent, however, with other empirical studies that have cautioned against the introduction of such an indicator in the *DSM-V*. For example, Beseler, Shmulewitz, Aharonovich, and Hasin (2009) reported that a "binge" drinking QF criterion was not a strong indicator of an alcohol use disorder (AUD) continuum of severity nor did its inclusion provide a superior fit to the data over a one-factor model of AUD based solely on the *DSM-IV*. Keyes, Geier, Grant, and Hasin (2009) noted that including an alcohol QF as an additional criterion for alcohol dependence warrants serious careful consideration because it would greatly increase the prevalence of alcohol dependence in the general population. The current findings contribute to this on-going debate and support the consensus that additional research is required prior to the introduction of a QF indicator as a diagnostic criterion for SUD in the future.

Importantly, the ordering of the severity and discrimination estimates for the ND criteria largely remained stable across the surveys. GIVE UP was consistently located at the most severe end of the continuum, whereas WITHDRAWAL and CUTDOWN fluctuated between being the least severe and the second least severe indicator of the continuum across the surveys. LARGER was consistently the most discriminating criterion, whereas GIVE UP was the least. These findings indicate that researchers and clinicians can be relatively confident with regard to the

Table 4. Differential item functioning, and the severity and discrimination parameters, of *DSM-IV* nicotine dependence criteria across the older (30+ years) and younger adults (18–29 years) using 2001–2002 NESARC data

		Severity parameter			Discrimination parameter			Young (18–29)	
DSM-IV criteria	G^2 ($df = 2$)	Old (30+)	Young (18–29)	Difference	Old (30+)	Young (18–29)	Difference	Mean	*SD*
WITHDRAWAL	20.1	−0.99	−0.70	**−0.29***	1.64	1.74	−0.10	0.21	1.03
TOLERANCE	7.0	1.24	1.22	0.02*	2.33	1.94	0.39	0.15	1.09
GIVE UP	26.9	2.42	3.61	**−1.19***	1.12	0.84	0.28	0.18	1.11
CUT DOWN	15.4	−0.55	−0.86	**0.31***	1.32	1.16	0.16	0.13	1.12
TIME SPENT	45.6	1.23	0.83	**0.40***	1.55	1.63	−0.08	0.12	1.07
CONTINUED USE	40.8	−0.14	0.16	**−0.30***	1.72	1.47	0.25	0.21	1.08
LARGER	9.4	0.83	0.88	−0.05	2.46	1.81	0.65*	0.16	1.11

Note. The G^2 test with 2 *df* evaluates differences between the age groups in both severity and discrimination parameters. Differences between the age groups on either parameter are evaluated using 1 *df* tests. *p* Values for *df* 1 tests were adjusted using the Benjamini–Hochberg procedure. Severity parameters that (a) represent a statistically significant difference between the age groups and (b) exceed our effect size criteria (0.25) are shown in bold. NESARC = National Epidemiologic Survey on Alcohol and Related Conditions.

**p* < .05.

severity of ND criteria measured by the AUDADIS-IV at different assessments. Nevertheless, the presence of clinically significant DIF for four of the diagnostic criteria (i.e., WITHDRAWAL, TOLERANCE, GIVE UP, and CONTINUED USE) over the 3-year period suggests that smokers' responses to the AUDADIS-IV questions operationalizing these criteria change upon readministration. Although this finding may raise some concerns, it is important to recognize that identifying significant DIF at the criterion level does not necessarily translate into practical differences in scale scores (Flora, Curran, Hussong, & Edwards, 2008). For example, WITHDRAWAL and TOLERANCE had higher severity parameters at Time 1 compared with Time 2, whereas the opposite was true for GIVE UP and CONTINUED USE. It is likely that these effects would cancel each other out to some extent in the creation of scale scores (Flora et al.). Notwithstanding this, attempts were made to investigate the potential source of the measurement noninvariance across the two time points. Although this can be a notoriously difficult process, it was possible that the age of the participants may have been a significant influencing factor in this study. There was evidence of measurement noninvariance for a few of the criteria across the two age groups, which suggests that age alone cannot account for the shift in the diagnostic criteria across the NESARC surveys. Other unknown factors (e.g., response shift; Schwartz & Sprangers, 1999) also may have contributed to changes in the manner with which respondents interacted with the instrument over the 3-year period. Measurement noninvariance in longitudinal research is a complex issue. Researchers and clinicians need to be aware of it and attempt to control it if they are to obtain an accurate index of ND over time. A variety of options exist for overcoming measurement noninvariance (cf. Cheung & Rensvold, 1998). For example, it may be more appropriate to rely on factor scores for the instrument, which are weighted sums of the item scores rather than composite scores (Hofmans, Pepermans, & Loix, 2009).

The major strength of this study was the use of sophisticated statistical techniques to analyze data from a large longitudinal household survey, which had a high response rate. The duration between the two surveys was relatively short, reducing the degree of recall error often found in surveys with longer periods of retrospective recall.

Despite these strengths, several limitations must be considered when interpreting the above findings. As with any longitudinal survey, the impact of censoring (i.e., individuals being withdrawn from observation because of death, institutionalization, etc.) must be considered. Although the current results are generalizable to the U.S. general population, they are limited to individuals who participated in both time points of the NESARC and reported smoking cigarettes in the year prior to both interviews. Individuals who consumed nicotine via other methods (e.g., pipe, cigars; 9%) were excluded from the analysis, and therefore, results cannot be assumed to generalize to all tobacco users. As a final point, it is noteworthy that participants completed the same items at two time points, which might increase the degree to which items have similar parameter estimates (discrimination and severity) over time and thus reduce power to detect DIF. The current analysis does not allow for the control of influences from individuals across this timeframe and presumes that there is sufficient variability both in individual levels of ND and in how they respond to questions over assessments to minimize impact on the item parameter estimates at the follow-up assessment. However, having responses from the same people over time also has substantial advantages because changes in item functioning over time can be seen as resulting only from changes in the relative severity of the items within the population; comparisons between two samples separated by time would reflect not only changes associated with time of assessment but also differences due to cohort effects or other between-sample differences.

In conclusion, the *DSM-IV* ND criteria provide reasonable coverage of an ND continuum; however, a cigarette QF index does not enhance the precision of this assessment. Changes in the performance of the *DSM-IV* ND criteria across the NESARC surveys appear to be relatively small and of minor clinical or practical significance. Nevertheless, researchers should be aware that the characteristics of the *DSM-IV* ND criteria, as assessed by the AUDADIS-IV, vary slightly across time.

Funding

The NESARC was sponsored and conducted by the National Institute on Alcohol Abuse and Alcoholism, with supplemental support from the National Institute on Drug Abuse. No other source of funding is declared.

Declaration of Interests

None declared.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Author: Washington, DC.

Benjamini, Y., & Hochberg, Y. (1995). Controlling false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, *57*, 289–300.

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238–246.

Beseler, C. L., Shmulewitz, D., Aharonovich, E., & Hasin, D. (2009). DSM-IV alcohol abuse and dependence: An IRT analysis in Israeli household residents. *Alcoholism: Clinical and Experimental Research*, *33*(Suppl.), 66A.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479), Reading, MA: Addison-Wesley.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Breslau, N., & Johnson, E. O. (2000). Predicting smoking cessation and major depression in nicotine-dependent smokers. *American Journal of Public Health*, *90*, 1122–1127.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Centers for Disease Control and Prevention. (2008). *Smoking-attributable mortality, years of potential life lost, and productivity loss—United States 2000–2004*. Retrieved 18 August 2009, from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5745a3.htm

Cheung, G. W., & Rensvold, R. B. (1998). Cross-cultural comparisons using non-invariant measurement items. *Applied Behavioral Science Review*, *6*, 93–110.

Colby, S. M., Tiffany, S. T., Shiffman, S., & Niaura, R. S. (2000). Measuring nicotine dependence among youth: A review of available approaches and instruments. *Drug and Alcohol Dependence*, *59*, S23–S29.

Compton, W. M., Saha, T. D., Conway, K. P., & Grant, B. F. (2009). The role of cannabis use within a dimensional approach to cannabis use disorders. *Drug and Alcohol Dependence*, *100*, 221–227.

Dierker, L. C., Donny, E., Tiffany, S., Colby, S. M., Perrine, N., & Clayton, R. R. (2007). The association between cigarette smoking and DSM-IV nicotine dependence among first year college students. *Drug and Alcohol Dependence*, *86*, 106–114.

Donny, E. C., & Dierker, L. C. (2007). The absence of DSM-IV nicotine dependence in moderate-to-heavy daily smokers. *Drug and Alcohol Dependence*, *89*, 93–96.

Donny, E. C., Griffin, K. M., Shiffman, S., & Sayette, M. A. (2008). The relationship between cigarette use, nicotine dependence, and craving in laboratory volunteers. *Nicotine & Tobacco Research*, *10*, 934–942.

Edwards, G., & Gross, M. M. (1976). Alcohol dependence: Provisional description of a clinical syndrome. *British Medical Journal*, *1*, 1058–1061.

Fiore, M. C., Hatsukami, D. K., & Baker, T. B. (2002). Effective tobacco dependence treatment. *Journal of the American Medical Association*, *288*, 1768–1771.

Flora, D. B., Curran, P. J., Hussong, A. M., & Edwards, M. C. (2008). Incorporating measurement non-equivalence in a cross-study latent growth curve analysis. *Structural Equation Modeling*, *15*, 676–704.

Foulds, J., Burke, M., Steinberg, M., Williams, J. M., & Ziedonis, D. M. (2004). Advances in pharmacotherapy for tobacco dependence. *Expert Opinion in Emerging Drugs*, *9*, 39–53.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*, 350–365.

Grant, B. F., Dawson, D., & Hasin, D. S. (2001). *The alcohol use disorder and associated disabilities interview schedule DSM-IV version (AUDADIS-IV)*. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Grant, B. F., Dawson, D., Stinson, F. S., Chou, P. S., Kay, W., & Pickering, R. (2003). The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug and Alcohol Dependence*, *71*, 7–16.

Grant, B. F., Hasin, D. S., Chou, P., Stinson, F. S., & Dawson, D. (2004). Nicotine Dependence and Psychiatric Disorders in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry*, *61*, 1107–1115.

Grant, B. F., & Kaplan, K. D. (2005). *Source and accuracy statement for the wave 2 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)*. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.

Grant, B. F., Kaplan, K., Shepard, J., & Moore, T. (2003). *Source and accuracy statement for wave 1 of the 2001–2002 National*

Epidemiologic Survey on Alcohol and Related Conditions. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Hartman, C., Gelhorn, H., Crowley, T. J., Sakai, J., Stallings, M. C., Young, S. E., et al. (2008). An item response theory analysis of DSM-IV marijuana abuse and dependence criteria in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*, 165–173.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K. O. (1991). The Fagerstrom Test for Nicotine Dependence: A revision of the Fagerström Tolerance Questionnaire. *Addiction*, *86*, 1119–1127.

Hendricks, P. S., Prochaska, J. J., Humfleet, G. L., & Hall, S. M. (2008). Evaluating the validities of different DSM-IV-based conceptual constructs of tobacco dependence. *Addiction*, *103*, 1215–1223.

Hofmans, J., Pepermans, R., & Loix, E. (2009). Measurement invariance matters: A case made for the ORTOFIN. *Journal of Economic Psychology*, *30*, 667–674.

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariate structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Hughes, J. R. (2006). Should criteria for drug dependence differ across drugs? *Addiction*, *101*(Suppl.), 134–141.

Hussong, A. M., Flora, D. B., Curran, P. J., Chassin, L. A., & Zucker, R. A. (2008). Defining risk heterogeneity for internalizing symptoms among children of alcoholic parents. *Developmental Psychopathology*, *20*, 165–193.

Kahler, C. W., & Strong, D. R. (2006). A Rasch model analysis of DSM-IV alcohol abuse and dependence in the National Epidemiologic Survey on Alcohol and Related Conditions. *Alcoholism: Clinical and Experimental Research*, *30*, 1165–1175.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, C. B., Hughes, M., et al. (1994). Lifetime and 12-months prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry*, *51*, 8–19.

Keyes, K. M., Geier, T., Grant, B. F., & Hasin, D. S. (2009). Influence of a drinking quantity and frequency measure on the prevalence and demographic correlates of DSM-IV alcohol dependence. *Alcoholism: Clinical and Experimental Research*, *33*, 761–771.

Langenbucher, J., Labouvie, E., Sanjuan, P., Kirisci, L., Bavly, L., Martin, C., & Chung, T. (2004). An application of item response theory analysis to alcohol, cannabis and cocaine criteria in DSM-IV. *Journal of Abnormal Psychology*, *113*, 72–80.

Lecrubier, Y. (2008). Refinement of the diagnosis and disease classification in psychiatry. *European Archives of Psychiatry and Clinical Neuroscience*, *258*, 6–11.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change. Decade of behavior* (pp. 203–240). Washington, DC: American Psychological Association.

Moolchan, E. T., Radzius, A., Epstein, D. H., Uhl, G., Gorelick, D. A., Lud Cadet, J., et al. (2002). The Fagerström Test for Nicotine Dependence and the Diagnostic Interview Schedule: Do they diagnose the same smokers? *Addictive Behaviors*, *27*, 101–113.

Muthén, B. O., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*, *31*, 1050–1066.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (4th ed.). Los Angeles: Muthén and Muthén.

Piper, M. E., McCarthy, D. E., & Baker, T. B. (2006). Assessing tobacco dependence: A guide to measure evaluation and selection. *Nicotine & Tobacco Research*, *8*, 339–351.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test*. Copenhagen, Denmark: Denmark's Paedagogiske Institute.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552–566.

Saha, T. D., Stinson, F. S., & Grant, B. F. (2007). The role of alcohol consumption in future classification of alcohol use disorders. *Drug and Alcohol Dependence*, *89*, 82–92.

Saunders, J. B., & Cottler, L. B. (2007). The development of the diagnostic and statistical manual of mental disorders version V substance use disorder section: Establishing the research framework. *Current Opinion in Psychiatry*, *20*, 208–212.

Schwartz, C. E., & Sprangers, M. A. G. (1999). Methodological approaches for assessing response shift in longitudinal health related quality-of-life research. *Social Science & Medicine*, *48*, 1531–1548.

Shiffman, S. (1989). Tobacco "chippers"—individual differences in tobacco dependence. *Psychopharmacology*, *97*, 539–547.

Shiffman, S., Waters, A. J., & Hickcox, M. (2004). The Nicotine Dependence Syndrome Scale: A multidimensional measure of nicotine dependence. *Nicotine & Tobacco Research*, *6*, 327–348.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*, *11*, 402–415.

Strong, D. R., Kahler, C. W., Abrantes, A. M., MacPherson, L., Myers, M. G., Ramsey, S. E., et al. (2007). Nicotine dependence symptoms among adolescents with psychiatric disorders: Using a Rasch model to evaluate symptom expression over time. *Nicotine & Tobacco Research*, *9*, 557–569.

Strong, D. R., Kahler, C. W., Colby, S. M., Griesler, P. C., & Kandel, D. (2009). Linking measures of nicotine dependence to a common latent continuum. *Drug and Alcohol Dependence*, *99*, 296–308.

Strong, D. R., Kahler, C. W., Ramsey, S. E., & Brown, R. A. (2003). Finding order in the DSM-IV nicotine dependence syndrome: A Rasch analysis. *Drug and Alcohol Dependence*, *72*, 151–162.

Thissen, D. (2001). *Irtlrdif v.20b: software for the computation of the statistics involved in item response theory likelihood ratio tests for differential item functioning*. Chapel Hill, NC: University of North Carolina. Unpublished manuscript.

Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measures*, *19*, 39–49.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamimi–Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, *27*, 77–83.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10.

Wood, M. J., Kerr, J. C., & Brink, P. J. (2006). *Basic steps in planning nursing research: from question to proposal* (6th ed.). Sudbury, MA: Jones and Bartlett.