Published in final edited form as:
Psychol Med. 2010 September ; 40(9): 1549–1558. doi:10.1017/S0033291709992005.

Item response modeling of DSM-IV mania symptoms in two representative U.S. epidemiological samples

Arpana Agrawal[1], **John I. Nurnberger Jr.**[2], and **Michael T. Lynskey**[1]

[1]Washington University School of Medicine, 660 S. Euclid CB 8134, St. Louis, MO 63110

[2]Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis, Indiana 46202-2873

Abstract

Background—There is considerable debate surrounding the effective measurement of DSM-IV symptoms used to assess manic disorders in epidemiological samples.

Methods—Using two nationally representative datasets, the National Epidemiological Survey of Alcohol and Related Conditions (NESARC, N=43,093 at Wave 1, N=34,653 at 3-year follow-up) and the National Comorbidity Survey-Replication (NCS-R, N=9,282), we examined the psychometric properties of symptoms used to assess DSM-IV mania. The predictive utility of the mania factor score was tested using the 3-year follow-up data in NESARC.

Results—Criterion B symptoms were unidimensional (single factor) in both samples. The symptoms assessing flight of ideas, distractibility and increased goal-directed activities had high factor loadings (0.70–0.93) with moderate rates of endorsement thus providing good discrimination between individuals with and without mania. The symptom assessing grandiosity performed less well in both samples. The quantitative mania factor score was a good predictor of more severe disorders at 3-year follow-up in the NESARC sample, even after controlling for a past history of DSM-IV diagnosis of manic disorder.

Conclusion—These analyses suggest that questions based on some DSM symptoms effectively discriminate between individuals at high and low liability to mania, while others do not. A quantitative mania factor score may aid in predicting recurrence for patients with a history of mania. Methods for assessing mania using structured interviews in the absence of clinical assessment require further refinement.

INTRODUCTION

Bipolar disorder (BD) is a debilitating mental illness affecting between 1–3% of the general population (Fountoulakis, 2008; Kessler *et al.* 2005; Merikangas *et al.* 2007). BD is associated with significant morbidity and mortality, including elevated risks of suicide (Fajutrao *et al.* 2009), as well as with a host of serious medical problems (e.g. cardiovascular illness) (Kupfer, 2005). Each of the disorders falling under the BD spectrum, particularly BD-I, are characterized by periods of elevated, expansive mood coupled with excitation, psychomotor agitation, increased risk-taking and goal-directed activities, or manic/hypomanic episodes, and in a preponderance of instances, by intervening episodes of low, depressive mood (American Psychiatric Association, 1994).

Please address correspondence to Washington University School of Medicine, Dept. of Psychiatry, 660 S. Euclid, CB 8134, Saint Louis, MO 63110, USA, Ph: 314-286-1778, Fax: 314-286-2213, arpana@wustl.edu.

Declaration of Interest: NONE

Challenges associated with the clinical diagnosis of BD include clinical course (i.e. depressive episodes preceding later mania/hypomania, thus delaying appropriate diagnosis), and patient denial of hypomania (which may be viewed as relief from depressive symptomatology and not as an impairment), (Fountoulakis, 2008). In a clinical setting, physicians are well equipped to investigate the possibility of a BD diagnosis since they have the opportunity to follow patients over time. In cross-sectional epidemiological studies, however, researchers rely on the psychometric properties of one-time assessments of manic and depressive episodes and therefore, the evaluation and psychometric performance of DSM-IV symptoms for these episodes is of critical importance.

Studies have examined the quality of criteria and the factorial structure underlying DSM-IV major depressive episodes (Aggen *et al.* 2005; Muthen, 1989; Reiser, 1989) but none have focused on manic episodes. One goal of such an investigation is to test whether there is a single dimension of liability underlying multiple symptoms for mania and whether this continuum affords an increment in information content over and above a diagnostic (i.e. binary) measure of affection status. Also, as demonstrated by numerous studies of substance abuse and dependence, item response analysis can identify symptoms that work poorly (i.e. have low factor loadings and are infrequently or too commonly endorsed) and those that work well (i.e. with high factor loadings and moderate levels of endorsement) at assessing liability. A study by Aggen and colleagues (2005) used a population-based sample of twins and found support for a unidimensional continuum underlying the DSM-IV criteria used to diagnose major depressive disorder with individual criteria performing well.

We are not aware of any study that conducts a similar analysis of the psychometric properties of the symptoms constituting manic episodes – a cornerstone of BD. Therefore, in the current study, we used data from two independent samples representative of U.S. adults – the National Epidemiological Survey of Alcohol and Related Conditions (NESARC) and the National Comorbidity Survey-Replication (NCS-R) to examine:

- **a.** Whether a unidimensional liability distribution (i.e. a one-factor solution) underlies DSM-IV mania criterion B symptoms in both samples;
- **b.** The discrimination, threshold, and total information (defined as the product of discrimination and threshold and representative of measurement precision) provided by the 7 mania criterion B symptoms in each sample;
- **c.** Whether a continuous measure of mania provides superior prediction of Manic/Hypomanic Disorder at the 3-year follow-up interview, when controlling for prior diagnoses of manic/hypomanic and major depressive disorder.

METHODS

Samples

Two epidemiological samples representative of U.S. adult populations were used:

- **a.** National Epidemiological Survey of Alcohol and Related Conditions (NESARC): NESARC is a nationally representative sample of 43,093 participants aged 18–99 years (at Wave 1). Comprehensive details regarding the survey design and sample characteristics are available elsewhere (Grant *et al.* 2003b). Wave 1 was collected during 2001–2002 by the U.S. Bureau of the Census on behalf of the National Institute on Alcohol Abuse and Alcoholism and the sample includes data from adult, non-institutionalized U.S. citizens and non-citizens (including Alaska and Hawaii). Approximately 57% of the sample is female and 19% of the sample is Hispanic (76% Caucasian), with an over-sampling for non-Hispanic Black households and for young adults aged 18–24 years. After complete description of

the study to the subjects, informed consent was obtained. Statements regarding the strict confidentiality of respondent privacy are available at http://niaaa.census.gov/confidentiality.html. The Alcohol Use Disorders and Associated Disabilities Schedule (AUDADIS-IV) was used to collect interview data from all individuals. The reliability of assessments from the AUDADIS-IV is good and these have been discussed in detail previously (Grant *et al.* 2003a; Ruan *et al.* 2008). However lifetime prevalence of mania is substantially higher than estimated in other studies (see below), and this may indicate difficulty in distinguishing bipolar subtypes in a non-clinical context. A 3-year follow-up interview was completed on 34,653 of these participants – diagnostic measures of manic and hypomanic disorder were collected at 3-year follow-up. The study showed a response rate of 86.7% (Ruan *et al.* 2008) or an effective sample size of 34,653, with exclusions due to death, deportation and mental or physical impairment. The cumulative response rate at Wave 2 was 70.2% and this compares favorably with many cross-sectional studies.

b. The National Comorbidity Survey-Replication (NCS-R): The NCS-R is an independent sample of 9,282 English-speaking U.S. adults (Kessler *et al.* 2005) interviewed in 2001–2003. The sample is 55% female. All 9,282 participants were administered Part I of the interviews (which included the core assessment). Part II was administered to all those who met lifetime criteria for any disorder in Part I and also to a probability subsample of the population (N=5,692). Informed consent was obtained from all subjects after the study protocol was explained – further details regarding the study protocol may be found in other publications (Kessler *et al.* 2005; Merikangas *et al.* 2007). Part I included assessments of mania and hypomania using the World Health Organization's Composite International Diagnostic Interview (CIDI) (Kessler and Ustun, 2004); 50 respondents who met criteria for a mood disorder diagnosis on the CIDI were clinically reappraised using the Structured Clinical Interview for DSM-IV (SCID) (Spitzer *et al.* 1987), showing very high concordance between the two instruments for a bipolar spectrum diagnosis (Kessler *et al.* 2006), but some differences when bipolar subtype (BDI, BDII, BDNOS) was considered.

Measures

Mania items from both interviews were used to assess the 7 DSM-IV criterion B symptoms. Each item (i.e. individual questions constituting a criterion B symptom) and criterion B symptom was individually assessed for its psychometric properties. Much like the DSM-IV, both interviews used an initial criterion (criterion A) referring to a period of abnormally and persistently elevated, expansive or irritable mood to identify those at risk for further endorsement of the mania items (which were converted to symptoms).

In NESARC, criterion A (1+ weeks of abnormally and persistently elevated, expansive or irritable mood) was assessed using 3 items (each experienced for 1+ weeks) that queried (a) excitement/elation that seemed not normal, (b) excitement/elation that made others concerned for respondent, and (c) irritability/annoyance that led to shouting/breaking things/fighting. Only those subjects who endorsed experiencing (a) or (b) or (c), (a total of 5148 individuals) were then queried using 13 items about their mania symptomatology.

In NCS-R, individuals were asked screening questions about excitement/restlessness as well as irritability/grouchiness that was excessive and persistent (N=2074). In follow-up questions, those who responded positively to the screening questions were asked whether they started arguments/fights/hit people during an episode of high mood and whether the episode lasted 4 days or longer. Those that responded positively to this question (N=1258)

were queried regarding 15 mania items. Of these individuals, 863 individuals reported experiencing an episode that lasted 1 week or longer or being hospitalized and their responses to the 15 mania items were used to construct DSM-IV symptoms.

The use of a skip-out question poses challenges for the generalizability of item response parameters to the population. Those who do not satisfy criterion A are structurally missing on criterion B, yet by not incorporating the information contained within criterion A, threshold parameters arising from analyses on the criterion B items (i.e. the 7 DSM-IV mania symptoms) are not population representative. Hence, we created ordinal measures to represent criterion A: In NESARC, the 3 questions (a), (b) and (c) were summed; in NCS-R, there were 2 distinct skip-outs, therefore 2 ordinal measures were created. The first was a three-level ordinal measure created by summing across responses to the first set of screening items while the second was a 4-level ordinal measure created by summing across the second set of screening items. As responses on the criterion B items could now be examined in the multiple non-zero levels of the screening questions, we could model criterion A jointly with items comprising the 7 DSM-IV symptoms that form criterion B.

Construction of DSM-IV Criterion B Mania Symptoms—The 7 DSM-IV symptoms (see Table 1) that are specified under criterion B were assessed using a series of items from the NESARC and NCS:

(1) Grandiosity: (a) Felt unusually important or felt had special gifts/ powers (1 item, NESARC);
(b) Too much self-confidence *or* associated with celebrities (2 items, NCS);
(2) Less sleep: Needed much less sleep than usual (1 item, NESARC, NCS);
(3) Talkativeness: (a) More talkative than usual *or* talked too fast (2 items, NESARC);
(b) Talk a lot more than usual (1 item, NCS)
(4) Flight of ideas: (a) Couldn’t keep track of thoughts *or* hard to follow thoughts (2 items,
NESARC);
(b) Thoughts jumped and raced (1 item, NCS);
(5) Distractibility: (a) Had trouble concentrating (1 item, NESARC);
(b) Constantly changed plans *or* hard to keep mind on tasks (2 items, NCS);
(6) Goal-directed activity: (a) Increased activity at home/work *or* more sexually active *or*
physically restless *or* restless/fidgety (4 items, NESARC);
(b) Take on large amounts of work *or* overly friendly *or* talking/acting
in unusual ways (e.g. tell embarrassing secrets) *or* restless/fidgety
(4 items, NCS);
(7) Activities with painful consequences: (a) Did things (foolish decisions, buying things, driving
recklessly) that could cause trouble *or* did things
later regretted (2 items, NESARC);

(b) Involved in foolish schemes *or* get into financial trouble *or* do risky things *or* sexual indiscretions (4 items, NCS).

Item response models were fitted to the 7 DSM-IV criterion B symptoms and also to the individual items (13 in NESARC and 15 in NCS-R) used to create the symptoms. As described above, criterion A, was included in the analyses as an ordinal measure.

Manic and Hypomanic Disorders—The lifetime prevalence of DSM-IV mania was 3.6% in NESARC (3.3% for BD-I); the prevalence figure for "bipolar disorder" was 4.4% in NCS-R (1.0% for BD-I) (Merikangas et al, 2007). The corresponding lifetime prevalence for DSM-IV hypomanic disorder in NESARC and NCS-R was 2.4% and 1.1% with the latter, lower estimate being for a hierarchical diagnosis. In the NESARC, test-retest reliability of BD-I, estimated for a subset of NESARC participants, was 0.59 (Grant *et al.* 2008). Assessment using the Short Form 12v2 mental disability scores yielded significant disability and social/occupational impairment in those diagnosed with BD-1 (published in Grant et al., 2008). In the NCS-R, a probability sample of 50 subjects, including 10 subjects, each, with BD-I, BD-II, sub-threshold BD and those endorsing a stem question on mania/hypomania, were re-interviewed using the lifetime non-patient version of the Structured Clinical Interview for DSM-IV (SCID). Prevalence of BD-I in the reappraisal was estimated at 1.1%. Concordance across assessments ranged from 0.50 (BD-II) to 0.94 (any BD spectrum disorder). A test of the utility of the screening questions (excitement/restlessness as well as irritability/grouchiness) suggested high sensitivity (0.72–0.96) however positive predictive values ranged from 0.32–0.52 (Kessler *et al.* 2006).

Statistical Analyses

Exploratory factor analyses (EFA)—EFA were conducted using the maximum likelihood estimator in MPlus (Muthen and Muthen, 2007). When there was evidence for a single factor solution, item response modeling was conducted.

Item response models—One-factor confirmatory factor analyses that allow the computation of item response parameters (Birnbaum, 1968) were conducted separately in NESARC and NCS-R using the MPlus (v5) software (Muthen and Muthen, 2004) using the maximum-likelihood estimator with robust standard errors (MLR) which is particularly well suited for complex survey designs. Symptom difficulty (or threshold in the context of psychiatric symptoms) and discrimination (or factor loading that shows how well a symptom correlates with the underlying construct that it is used to measure) were computed using a 2 parameter (2P) logistic model with logit function $L = 1.7*a(\theta - b)$, where

a = discrimination, or the ability of a symptom to distinguish individuals with high liability from individuals with low liability; this parameter is also represented by factor loadings.

b = threshold, or the location along the underlying distribution where the symptom functions; this parameter is termed *difficulty* in traditional item response models.

θ = the liability distribution for the disorder.

Parameters from the confirmatory factor model can be easily converted to discrimination and difficulty parameters (MacIntosh and Hashim, 2003; Muthen, 1985; Takane and de Leeuw, 1987). An alternative, more parsimonious model, the 1 parameter (Rasch; 1960; 1966) model is also possible, where all symptoms are assumed to discriminate equally. We

tested this model in both samples by constraining the factor loadings (i.e. equal discriminations) across symptoms - this led to a serious deterioration of model fit ($p<.0001$) and hence the 2P model was selected for analysis.

The primary item response models focused on the 7 DSM-IV criterion B symptoms – formal tests of differential symptom functioning were also conducted on the symptoms across NESARC and NCS-R to examine whether discrimination and threshold parameters were statistically different across the 2 samples. We also factor analyzed the individual items that constituted the 7 criterion B symptoms to determine their behavior in a series of secondary analyses.

Association between factor score and manic/hypomanic episodes at 3-year follow-up—A factor score, which represents an individual's liability to mania, was created within the factor analysis on the wave 1 screening item(s) and 7 DSM-IV criterion B symptoms – higher factor scores represent increased risk for mania. In the NESARC alone, data from a 3-year follow-up interview were used to also create the mania factor score at 3-year follow-up and to examine whether the factor scores representing the underlying liability to mania at wave 1 were a better predictor of the factor score at 3-year follow-up and of subsequent manic and hypomanic disorder when compared to diagnoses of mania/ hypomania at wave 1. A series of linear and logistic regression models (SAS Institute, 1999) were fitted, with the mania factor score at 3-year follow-up, manic and hypomanic disorder diagnosed at 3-year follow-up as the outcomes, and the wave 1 factor score as well as age at wave 1 and diagnosis at wave 1 (mania, hypomania, and/or major depressive disorder) as predictors.

RESULTS

Mania symptoms

The frequency of the 7 DSM-IV criterion B symptoms and the items that were used to create the symptoms in NESARC and NCS-R is presented in Table 1. While there were marked differences in the weighted prevalence of individual symptoms across the 2 samples, the most commonly endorsed symptom in both was an increase in goal-directed activities (69–87%) while the least commonly endorsed symptom was inflated self-esteem or grandiosity (15–38%).

Item Response Modeling

EFA revealed that underlying the 7 DSM-IV dependence symptoms was a single factor (CFI/TLI > 0.85). Next, we used confirmatory factor models to compute factor loadings and thresholds, which were also used to compute item characteristic curves. Loglikelihood ratio chi-square statistics from the confirmatory models were not significant, further confirming unidimensionality (p values ranging from 0.77–0.99, AIC= 81715.6 & 23980.6; BIC=81817.1 & 24137.6). Factor loadings from this one factor model are presented in Table 1. Factor loadings for the 7 DSM-IV symptoms ranged from 0.65 (grandiosity, painful consequences) to 0.90 (goal-directed activities) in NESARC while NCS-R showed comparable factor loadings ranging from 0.52 – 0.93. While some factor loadings appeared to differ across samples, after allowing for differing thresholds (i.e. endorsement rates) across the samples, only the factor loadings for flight of ideas and for goal-directed activities were statistically different across NESARC and NCS-R. In both samples, the screening questions were jointly modeled with the symptoms and these screening items had fairly high factor loadings suggesting that they fall on the same continuum and are highly correlated with the symptoms that are conditional on them.

Confirmatory factor analyses were also conducted on the individual items used to create the 7 DSM symptoms – likelihood ratio chi-square statistics were not significant suggesting that the unidimensional model fit the data well. Factor loadings for the individual items were high in both samples (Table 1) confirming that utilization of these individual items to create symptoms did not influence the factorial structure underlying the mania assessment.

Item characteristic curves (ICC) for the NESARC and NCS-R assessments of the 7 criterion B symptoms are presented in Figures 1A and 1B respectively. The steepness of each curve represents discrimination while its position the x-axis represents difficulty. Consistent with Table 1, in both samples, the symptom assessing increase in goal-directed activities had the lowest liability threshold while the symptom assessing grandiosity had a high liability threshold (see Fig 1A and 1B). Also, as denoted by the higher factor loadings, the symptom assessing flight of ideas was highly discriminating (with only moderate threshold) across NESARC and NCS-R. In NESARC, increase in goal-directed activities, while a frequently endorsed symptom was also highly discriminating – its discrimination was somewhat lower in NCS-R.

Information from ICC for the 7 criterion B symptoms were summed to create the test characteristic curve (TCC) for the 7 DSM-IV symptoms (Figure 2). The TCC for the 2 samples look highly comparable across the samples. Criterion B for manic disorder is met when 3 or more of the 7 criterion B symptoms are endorsed – those with factor scores exceeding 1.2 on the NESARC and 1.4 on the NCS-R would likely satisfy Criterion B, which is a fairly low threshold, underscoring the probability that epidemiological interviews tend to 'cast a wider net' in terms of their assessment of liability to mania.

The total information curves (TIC), representing measurement precision for both samples are shown in Figure 3. TIC represent the total (additive) information from all 7 symptoms, where, for symptom *i*, the information at liability level θ, is $I_i(\theta) = a_i^2 P_i(\theta)[1-P_i(\theta)]$, where $P_i(\theta)=1/[1+\exp(-a_i(\theta-b_i))]$ and a_i and b_i refer to the discrimination and threshold, respectively, of item *i*. TIC are shown (Figure 3) with and without screening questions – when screening items were included, the NESARC mania symptoms performed with considerable precision but over a fairly narrow range of liability while the NCS-R suggests somewhat lower precision. TIC without screening items is also shown – these curves are lower. Thus, variations in screening items used in each study may have contributed to total information.

Association between factor score and mania/hypomania at 3-year follow-up

The mania factor was computed within the factor model using the screening items and 7 DSM-IV criterion B symptoms created from items assessing manic symptoms in the 3-year period of follow-up – this factor was significantly correlated with the wave 1 factor score (r=0.26). Factor loadings of the 7 symptoms ranged between 0.62 to 0.89 and the screening item also loaded well (0.86). Thresholds at 3-year follow-up were modestly higher than at wave 1 (for instance, the threshold for goal-directed activities was 0.95 at wave 1 and 1.05 at 3-year follow-up). Of the 41,885 individuals with valid factor scores at the first wave of NESARC, 33,745 also had data at the 3-year follow-up. Of these, 819 and 549 met criteria for manic and hypomanic disorder respectively since their wave 1 interview (i.e. new onsets). Scores on the underlying mania factor (while controlling for age) were excellent predictors of, not only the 3-year follow-up mania factor score ($\beta=0.22$, $p < .0001$), but also of diagnosis of manic (O.R. = 2.54, $p < .0001$) and hypomanic (O.R. = 2.24, $p < .0001$) disorder at 3-year follow-up. Lifetime diagnoses of manic/hypomanic disorder (combined to reflect either disorder at wave 1) and major depressive disorder were associated with increased risk for meeting NESARC criteria for manic disorder at follow-up (O.R. ranging from 1.92–6.09). However, when the factor scores were added to this model (Model 5), the

effect of the diagnostic measure of wave 1 lifetime manic/hypomanic disorder was eliminated. This was largely replicated for hypomanic disorder as well. Intriguingly, for hypomanic disorder, when a factor score that excluded the screening item was included, a substantial decline in predictive utility was noted, demonstrating that individuals who may not endorse any of the 7 symptoms but pass through a mania screen may be at risk for hypomanic disorder.

DISCUSSION

We sought to characterize the symptoms used to assess manic disorders in two large scale epidemiological samples representative of the U.S. population. Item response analyses revealed a unidimensional liability to mania criterion B symptoms – certain symptoms, such as flight of ideas, were highly discriminating and displayed moderate thresholds while others, such as grandiosity were found to have less utility. The quantitative liability measure was a good predictor of subsequent symptoms of manic disorder.

Factor analyses of mania

While no study has conducted an item response analysis of the DSM-IV symptoms for mania in epidemiological surveys, a number of investigators have examined the factorial nature of mania scales in patient populations. For instance, Cassidy et al (Cassidy *et al.* 1998) used data on items assessing mania and dysphoria in 237 BD patients to reveal 5 factors. Similarly, Picardi et al (Picardi *et al.* 2008) conducted a factor analysis of the 34-item Brief Psychiatric Rating Scale in an inpatient sample to identify 4 factors including a 'mania' factor indexed by elevated mood, psychomotor agitation and distractibility. Cassano and colleagues (Cassano *et al.* 2009) used data on 617 patients diagnosed with bipolar spectrum disorders and reported that underlying 68 items assessing manic-hypomanic (and some vegetative function) features were 5 factors indexing aspects of pure and mixed mania. While these studies have been largely informative in subtyping manic features in patient populations where item endorsement is higher, they do not address the core issue of whether items and symptoms assessing mania in population-based samples work reasonably well in distinguishing individuals at high versus low risk for BD. Our analyses suggest that while several items, such as flight of ideas, distractibility and increased goal-directed activities are good indices of liability to manic disorder, others such as grandiosity discriminated poorly and may require re-working in order to be optimized for use in epidemiological samples.

Screening items and item response modeling

Analyses were done to compare models including and excluding the screening item(s). Factor loadings were generally higher when the screening items were included – furthermore, the convergence in factor loadings for individual symptoms across NESARC and NCS-R dramatically increased upon jointly modeling the screening items with the criterion B symptoms. It is also noteworthy that certain symptoms, such as less sleep than usual, were assessed in both samples similarly (i.e. with a single item) - while analyzing them in a population-based manner (i.e. via inclusion of the screening items) did not produce statistically significant differences in their factor loadings, when modeled in sub-samples screened for mania, their liability threshold and discrimination were different across samples. This implies that the screens for mania in NESARC and NCS-R may have been somewhat different in the sub-samples they selected such that rates of endorsement and factor loadings of the 7 DSM-IV symptoms, subsequent to the screen, were quite different across the samples. This is also reflected in the total information curves, which imply increased measurement precision In NCS-R which utilizes multiple screening items.

Association between factor score and mania/hypomania at 3 –year follow-up

The factor score representing liability to mania at wave 1 was also associated with DSM-IV manic and hypomanic disorder at 3-year follow-up. For manic and hypomanic disorder, when the wave 1 mania factor score was included in the model, a diagnosis of manic/ hypomanic disorder at wave 1 did not retain significant predictive utility, suggesting that the wide range of liability captured by the mania factor score is a superior index of continued liability to manic and hypomanic disorders than prior diagnoses.

Mania in epidemiological surveys

An important question remains regarding whether mania assessed using the AUDADIS and CIDI capture similar individuals and whether these individuals would be diagnosed as 'affected' if a clinical assessment were to be made. For instance, while the prevalence of BD (BDI and II) was originally estimated to be similar in NCS-R (3.9%) and NESARC (5.7%), and somewhat greater than a prior assessment of mania using the CIDI in a Swedish cohort of older twins (2.6%) (Soldani *et al.* 2005), it is notable that subsequent clinical reappraisal in NCS-R reduced the rate of BD-I to 1.0%, which is more in keeping with classical estimates of BD-I while the NESARC rates for BD-I remain close to 3% as such a clinical re-appraisal was not conducted in the NESARC. The reduction of NCS-R rates may reflect refined clinical assessments of impairment due to mania, which may not have been adequately captured by either epidemiological interview. It seems likely that NESARC-defined mania includes conditions that would generally be regarded as milder forms.

This raises the question: should mania assessments be included in non-clinical interviews? The goal of this study was to demonstrate the measurement properties of DSM-IV mania symptoms in population-based epidemiological samples – a preponderance of subjects in such samples are non-manic and thus the functionality of such instruments will be somewhat limited by the prevalence of the syndrome itself. Mania assessments as well as other major psychopathology (e.g. Schizophrenia) lend themselves better to patient populations and their performance in general population settings may be hindered by multiple factors (e.g. interviewer's inability to verify level of impairment, confounds with other disorders or with substance use). However, in both samples, as a set of DSM-IV symptoms, mania items appear to have some utility. Additionally, as shown by a high concordance for BD in general, but more modest concordance for BD-I between the NCS-R CIDI assessment and the clinical-reappraisal interview, a subset of subjects qualifying as 'manic' in epidemiological surveys truly represent clinical cases while others are likely sub-threshold or unaffected – reducing such false-positives may be challenging in epidemiological surveys and researchers may find greater utility in using quantitative assessments of mania, which capture of range of variation in risk, instead of using diagnoses.

Limitations

There are some important limitations to this study: First, neither epidemiological sample (for practical purposes) utilized an interview designed for the assessment of severe psychopathology nor were any of the interviews administered by interviewers with clinical expertise in diagnosis of BD – for instance, DSM-IV mania requires 'marked impairment' – if this is not carefully assessed, diagnosis is often less precise. While the goal of this paper was to examine the utility of general epidemiological interviews, access to data on the same individuals interviewed with a clinical interview schedule would have allowed for psychometric comparability. Second, 3-year follow-up data in NESARC used an interval instrument (i.e. questions were asked for the 3-year duration between the interviews) – thus, our associations between the factor score and the factor score at 3-year follow-up as well as mania and hypomania at 3-year follow-up represent a prediction of recurrence but not an assessment of diagnostic stability per se. Third, these are samples of U.S. adults that may not

generalize to other populations. Fourth, some of the younger participants in both studies may not have passed the age of risk for onset of manic disorders. Finally, while the DSM-IV symptoms are the 'gold standard' for diagnosis of manic disorders, it would have been intriguing to compare the psychometric properties of DSM symptoms with those stemming from other mania/hypomania scales – these were not available.

Conclusion

The finding of a single liability dimension underlying mania symptoms adds to a growing body of literature that has begun to view psychiatric disorders in a dimensional framework – this may have significant utility in studies where diagnostic dichotomies reduce power (e.g. gene- finding efforts). Caution, however, is needed in their interpretation. Concerns still exist surrounding the interpretation of mania constructs in epidemiological samples to confirm the extent to which these assessments are congruent with clinical evaluations of manic disorders.

Acknowledgments

Funding: Research reported here is supported by DA23668 (A.A.), DA18660 (M.T.L.), DA18267 (M.T.L.), MH68009 (J.I.N.) and a grant to J.I.N. from the Indiana Division of Mental Health and Addictions.

Reference List

Aggen SH, Neale MC, Kendler KS. DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. Psychologicl Medicine 2005;35:475–487.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association; 1994.

Birnbaum, A. Some latent trait models. In: Lord, FM.; Norvick, MR., editors. Statistical Theory of Mental Test Scores. Reading, MA: Addison-Wesley; 1968. p. 397-472.

Cassano GB, Mula M, Rucci P, Miniati M, Frank E, Kupfer DJ, Oppo A, Calugi S, Maggi L, Gibbons R, Fagiolini A. The structure of lifetime manic-hypomanic spectrum. J Affect Disord 2009;112:59–70. [PubMed: 18541309]

Cassidy F, Forest K, Murry E, Carroll BJ. A factor analysis of the signs and symptoms of mania. Arch Gen Psychiatry 1998;55:27–32. [PubMed: 9435757]

Fajutrao LB, Locklear JC, Priaulx J, Heyes A. A systematic review of the evidence of the burden of bipolar disorder in Europe. Clin Pract. Epidemol. Ment Health 2009;5:3.

Fountoulakis KN. The contemporary face of bipolar illness: complex diagnostic and therapeutic challenges. CNS Spectr 2008;13:763–769. [PubMed: 18849896]

Grant BF, Chou SP, Goldstein RB, Huang B, Stinson FS, Saha TD, Smith SM, Dawson DA, Pulay AJ, Pickering RP, Ruan WJ. Prevalence, correlates, disability, and comorbidity of DSM-IV borderline personality disorder: results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 2008;69:533–545. [PubMed: 18426259]

Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. 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 Alcohol Depend 2003a;71:7–16. [PubMed: 12821201]

Grant, BF.; Kaplan, K.; Shepard, J.; Moore, T. Source and Accuracy Statement for Wave 1 of the 2001–2002 National Epedemiological Survey on Alcohol and Related Conditions. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2003b.

Kessler RC, Akiskal HS, Angst J, Guyer M, Hirschfeld RM, Merikangas KR, Stang PE. Validity of the assessment of bipolar spectrum disorders in the WHO CIDI 3.0. J Affect Disord 2006;96:259–269. [PubMed: 16997383]

Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005;62:617–627. [PubMed: 15939839]

Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res 2004;13:93–121. [PubMed: 15297906]

Kupfer DJ. The increasing medical burden in bipolar disorder. Journal of the American Medical Association 2005;293:2528–2530. [PubMed: 15914754]

MacIntosh R, Hashim S. Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. Applied Psychological Measurements 2003;27:372–379.

Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, Kessler RC. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch Gen Psychiatry 2007;64:543–552. [PubMed: 17485606]

Muthen BO. A method for studying the homogeneity of test items with respect to other relevant variables. Journal of Educational Statistics 1985;10:121–132.

Muthen BO. Dichotomous factor analysis of symptom data. Sociological Methods and Research 1989;18:19–65.

Muthen LK, Muthen B. MPlus: The comprehensive Modeling Program for Applied Researchers, Version 3. 2004

Picardi A, Battisti F, de GG, Morosini P, Norcio B, Bracco R, Biondi M. Symptom structure of acute mania: a factor study of the 24-item Brief Psychiatric Rating Scale in a national sample of patients hospitalized for a manic episode. J Affect Disord 2008;108:183–189. [PubMed: 18029028]

Reiser M. An application of item-reponse model to psychiatric epidemiologu. Sociological Methods and Research 1989;18:66–103.

Ruan WJ, Goldstein RB, Chou SP, Smith SM, Saha TD, Pickering RP, Dawson DA, Huang B, Stinson FS, Grant BF. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of new psychiatric diagnostic modules and risk factors in a general population sample 5. Drug Alcohol Depend 2008;92:27–36. [PubMed: 17706375]

SAS Institute. SAS User Guide, Version 8.2. Cary, NC: SAS Institute Inc.; 1999.

Soldani F, Sullivan PF, Pedersen NL. Mania in the Swedish Twin Registry: criterion validity and prevalence. Aust N Z J Psychiatry 2005;39:235–243. [PubMed: 15777359]

Spitzer, RL.; Williams, JB.; Gibbon, J. Structured Clinical Interview for DSM-III-R: Patient Version (SCID-P). New York: New York State Psychiatric Institute; 1987.

Takane Y, de Leeuw J. On the relationship between item response theory and factor analysis of discretized variables. Psychometrika 1987;52:393–408.

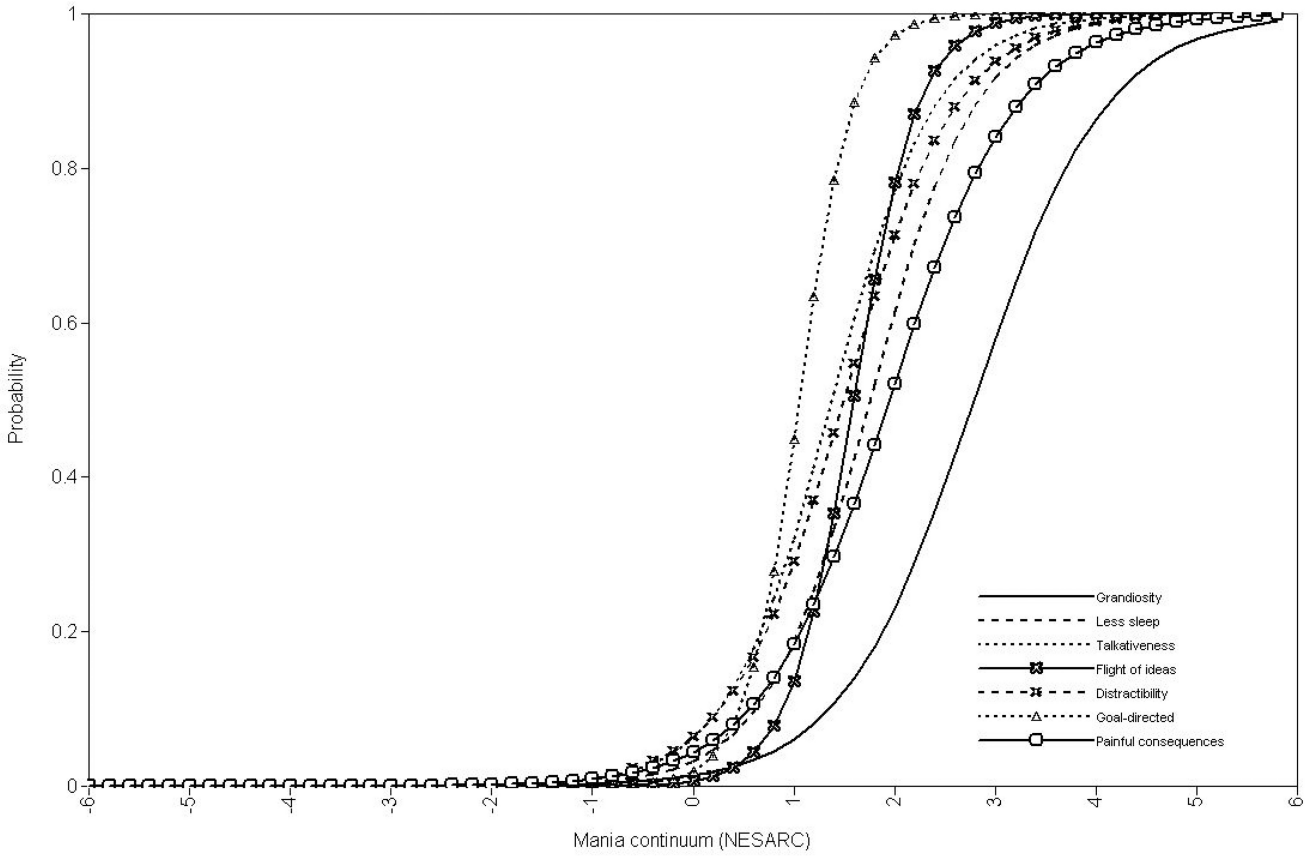

Figure 1A

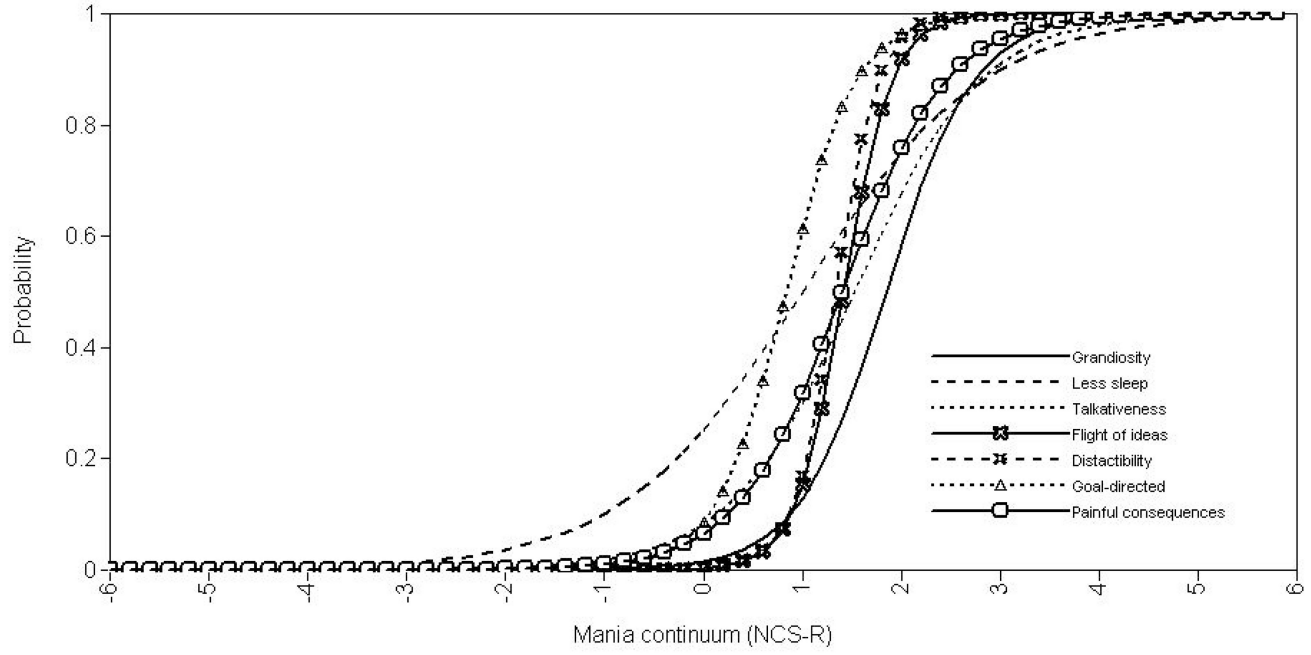

Figure 1B

Figure 1.

Figure 1A. Item characteristic curves for mania symptoms in 5,148 individuals reporting 1+ weeks of elevated mood, restlessness irritability in the NESARC

Figure 1B. Item characteristic curves for mania symptoms in 863 individuals reporting 1+ weeks of elevated mood, restlessness or irritability in the NCS-R.

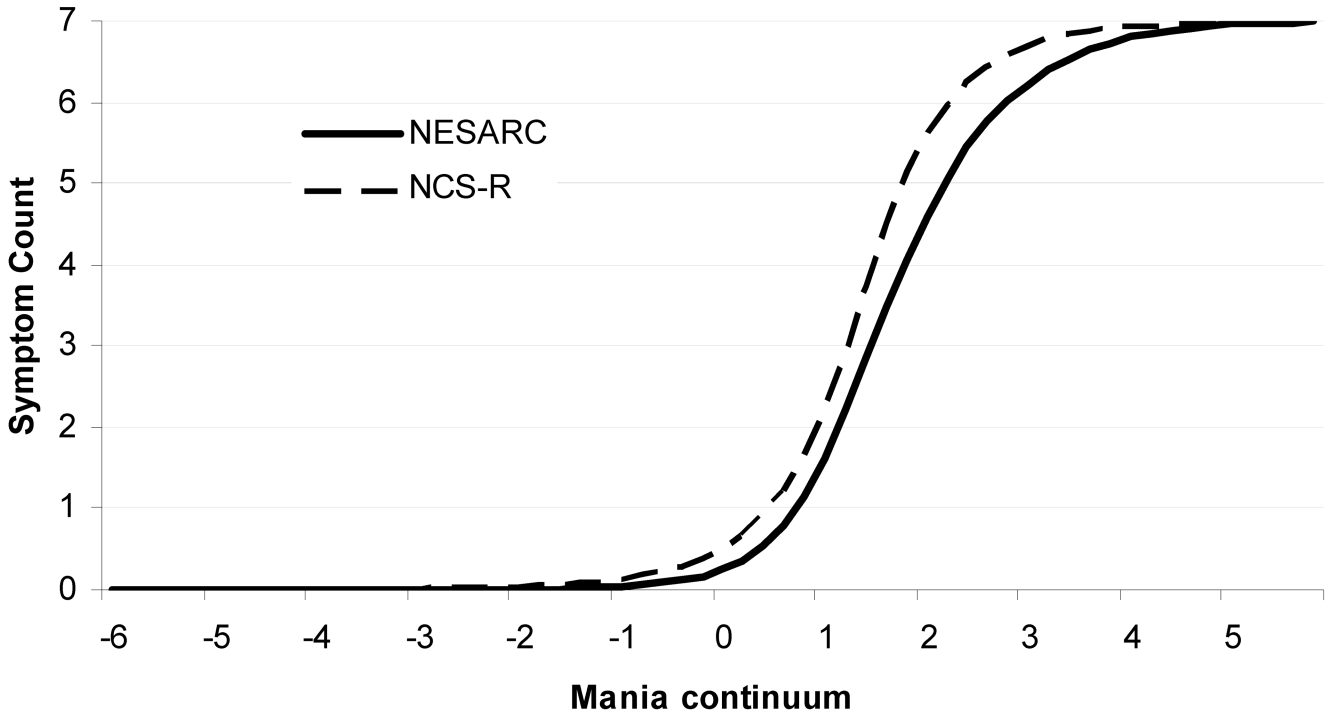

Figure 2.
Test characteristic curves showing the relationship between symptom-endorsement and the liability to mania.

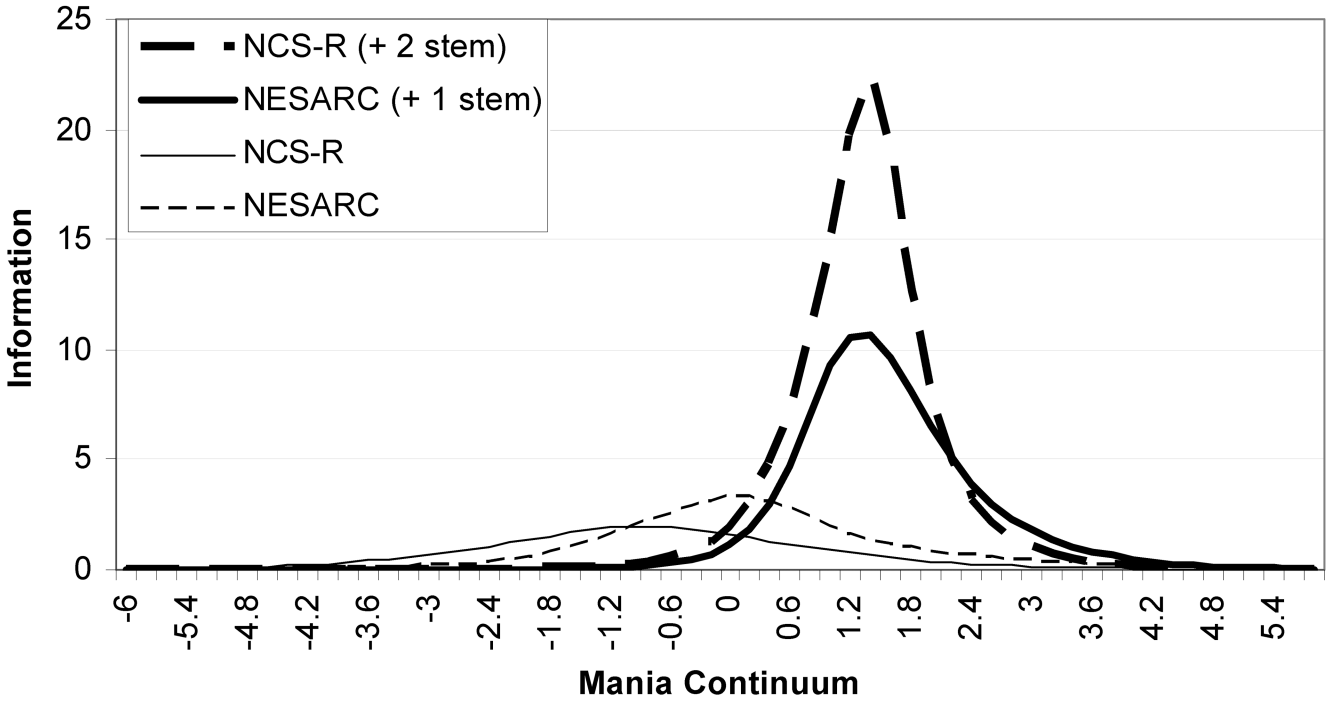

Figure 3.
Total information curves for the 7 mania symptoms with and without the screening items in NESARC and NCS-R. Note that NCS-R used 2 screening items while NESARC used 1.

TABLE 1

Frequency, standardized factor loadings and thresholds of the screening items and of individual mania symptoms and the items comprising them in eligible individuals in NESARC and NCS-R respectively.

Symptom	Items	Prevalence	Factor Loading	Threshold
NESARC (N=41,885)				
Stem/Screen	-	1=8.0; 2=2.7, 3=1.5	0.88	1=1.14; 2=1.74; 3=2.19
Grandiosity		**1.9**	**0.65**	**1.80**
	Felt unusually important	-	0.55	1.66
Less sleep		**4.8**	**0.73**	**1.28**
	Needed less sleep than usual	-	0.60	1.04
Talkativeness		**6.4**	**0.73**	**1.01**
	More talkative	5.9	0.52	0.75
	Talked too fast	3.2	0.76	1.55
Flight of ideas		**5.2**	**0.87***	**1.38**
	Lost track of thoughts	4.8	0.97	1.50
	Hard to follow thoughts	4.0	0.97	1.60
Distractibility		**5.9**	**0.70**	**1.05**
	Trouble concentrating	-	0.78	1.10
Goal-directed activities		**8.5**	**0.90***	**0.95**
	Increased activity	5.9	0.60	0.87
	More sexually active	1.6	0.56	1.73
	Physically restless	3.2	0.75	1.57
	Fidgety	5.7	0.83	1.18
Painful consequences		**4.1**	**0.66**	**1.28**
	Did reckless things	3.0	0.68	1.49
	Did things that later regretted	2.8	0.61	1.45
NCS-R (N=9,282)				
Stem/Screen 1	-	1=16.6; 2=5.7	0.89	0.78
Stem/Screen 2		1=4.8; 2=1.8; 3=2.4, 4=0.3	0.95	1.62
Grandiosity		**5.2**	**0.78**	**1.36**
	Too much confidence	4.9	0.77	1.34
	Associate with celebrity	0.7	0.87	2.35
Less sleep		**8.4**	**0.52**	**0.51**
	Less sleep than usual	-	0.59	0.62
Talkativeness		**7.5**	**0.66**	**1.00**
	Talked a lot more	-	0.61	0.79

Symptom	Items	Prevalence	Factor Loading	Threshold
		NESARC (N=41,885)		
Flight of ideas		**8.7**	**0.92***	**1.30**
	Thoughts jumped/raced	-	0.84	0.99
Distractibility		**9.3**	**0.93**	**1.25**
	Constantly changed plans	6.5	0.86	1.26
	Hard to keep mind on tasks	7.6	0.79	1.07
Goal-directed activities		**11.8**	**0.84***	**0.70**
	Unrealistic goals	6.1	0.70	1.11
	Overly friendly	6.2	0.44	0.70
	Talk/act unusual	5.7	0.70	1.16
	Restless/fidgety	6.7	0.69	1.04
Painful consequences		**7.9**	**0.72**	**1.02**
	Foolish schemes	1.5	0.69	1.90
	Financial trouble	4.8	0.73	1.36
	Do risky things	5.2	0.77	1.32
	Sexual indiscretions	2.8	0.64	1.54

* Denotes factor loading was statistically different across samples.