Published in final edited form as:
Int J Methods Psychiatr Res. 2010 September ; 19(3): 156–166. doi:10.1002/mpr.318.

Measurement non-invariance of DSM-IV narcissistic personality disorder criteria across age and sex in a population-based sample of Norwegian twins

THOMAS S. KUBARYCH, **STEVEN H. AGGEN**, **KENNETH S. KENDLER**, **SVEN TORGERSEN**, **TED REICHBORN-KJENNERUD**, and **MICHAEL C. NEALE**
Department of Psychiatry, Virginia Commonwealth University, Richmond, VA

Abstract

We investigated measurement non-invariance of DSM-IV narcissistic personality disorder (NPD) criteria across age and sex in a population-based cohort sample of 2794 Norwegian twins. Age had a statistically significant effect on the factor mean for NPD. Sex had a statistically significant effect on the factor mean and variance. Controlling for these factor level effects, item-level analysis indicated that the criteria were functioning differently across age and sex. After correcting for measurement differences at the item level, the latent factor mean effect for age was no longer statistically significant. The mean difference for sex remained statistically significant after correcting for item threshold effects. The results indicate that DSM-IV NPD criteria perform differently in males and females and across age. Differences in diagnostic rates across groups may not be valid without correcting for measurement non-invariance.

Keywords

narcissistic personality disorder; twins; population-based sample; item response theory; measurement non-variance

Introduction

Narcissistic personality disorder (NPD) was introduced into the Diagnostic and Statistical Manual of Mental Disorders (DSM) in its third edition (DSM-III, American Psychiatric Association, 1980) in an effort to capture, within a set of operationalized criteria, the construct commonly used in psychoanalysis. Clinicians diagnosed NPD at twice the rate of patients who actually met the DSM-III threshold for NPD (Morey and Ochoa, 1989; Cain *et al.*, 2008), and NPD remains much more frequently diagnosed in clinical populations than suggested by epidemiological prevalence rates. Torgersen (2005) reports NPD to be a rare personality disorder in samples from the general population. Subsequent editions of the DSM have revised the criteria for NPD. NPD is linked to other personality disorders, including borderline personality disorder (Kernberg, 1975), and antisocial personality disorder (Kernberg, 1989), and normal personality traits such as extraversion and (low) agreeableness (Costa and Widiger, 2002). It has recently been included, with psychopathy

Correspondence Thomas S. Kubarych, Department of Psychiatry, Virginia Commonwealth University, P.O. Box 980126, Richmond, VA 23298-0126. Telephone 804-828-8135 Fax 804-828-1471 tkubarych@vcu.edu or tkubarych@gmail.com.

Declaration of interest statement The authors have no competing interests.

and Machiavellianism, as one of the `dark triad' personalities (Jacobwitz and Egan, 2006; Paulhus and Williams, 2002; Kubarych, 2005a, 2005b).

NPD remains one of the least studied personality disorders. Fossati *et al.* (2005) reported on the latent structure of DSM-IV (American Psychiatric Association, 2000) NPD criteria in a sample of 641 Italian clinical psychology and psychotherapy outpatients (61.3% women). Confirmatory and exploratory factor analyses showed that the nine criteria loaded on two factors, which correlated 0.77 and corresponded to the widely described difference between `overt' and `covert' variants of narcissism. The first factor consisted of six items (DSM-IV criteria 1, 3, 5, 6, 7 and 9, see Table 1) and was identified with overt narcissism. The second factor consisted of three items (criteria 1, 4 and 8) and was identified with covert narcissism. Latent class analysis, however, suggested a three-class solution: NPD subjects; NPD non-subjects; and a class resembling the overt factor. The latent class base-rate estimate of NPD among this patient group was 18.1%. Taxonomic analyses provided evidence for a discrete latent structure of NPD. This study did not address whether measurement was invariant across sex. If the measurement of NPD differed between men and women, the results might be different. Measurement could also have differed across age or Axis I diagnosis.

Miller *et al.* (2008) tested the Fossati *et al.* two-factor model of DSM-IV NPD criteria against a one-factor model in two combined samples (total $n = 289$). Sample 1 ($n = 151$, 68% women, 86.8% white) was composed of 70 psychiatric patients, 23 diabetic patients and 58 university faculty or staff. Sample 2 was composed of 138 psychiatric outpatients (76% women, 74% white). A one-factor model provided a more parsimonious fit to the data better than a two-factor model. The Miller *et al.* cohort, however, was heterogeneous across age, sex, sample, ethnicity, or psychopathology groups. Measurement could have varied across age, sex, sample, ethnicity, and psychopathology groups. Measurement non-invariance could account for the failure to replicate the two-factor solution.

Ronningstam *et al.* (1995) examined changes in pathological narcissism in a clinical sample of 20 patients diagnosed with NPD. Baseline scores on the Diagnostic Interview for Narcissism (DIN) were compared with scores three years later using *t*-tests and chi-squared statistics. Sixty percent of the patients had significantly decreased in pathological narcissism, particularly in interpersonal relations and reactiveness, but patients with higher baseline scores remained unchanged. The authors suggested that changes in pathological narcissism were related to achievements, new durable relationships and disillusionment. No item response studies of changes in DSM-IV criteria over age have been reported. Sensitivity to group differences, such as in different age groups or between men and women,is a core concern in psychiatric research. No previous studies have investigated whether or not measurement of DSM NPD criteria is invariant across age or sex. The present study aims to provide such data for NPD, as well as draw attention to the need for such studies in other phenotypes.

Sample and assessment methods

Participants were 2794 Norwegian twins (1022 male and 1772 females). The participants are a sub-sample from the Norwegian Institute of Public Health Twin Panel (NIPHTP, Harris *et al.*, 2002). Twins were identified through the Norwegian National Medical Birth Registry. The NIPHTP contains 15 370 like- and unlike-sexed twins born between 1967 and 1979. Age range at time of interview was 19 to 36 years. Two questionnaire studies have been conducted; in 1992 (twins born 1967–1974) and in 1998 (twins born 1967–1979). Tambs *et al.* (2009) examined recruitment and attrition bias in the sample, and found that poor health predicted non-participation in the first questionnaire. Monozygosity, female sex, having no children and high education predicted participation in the second questionnaire. Altogether,

N = 12 700 twins received the second questionnaire, and N = 8045 responded after one reminder (63%). Participants for the current study were recruited from 3221 twin pairs from the NPPHTP, who were requested to complete an extensive interview of Axis I and Axis II psychiatric disorders. The response rate was 44%.

A Norwegian version of the Structured Interview for DSM-IV Personality (SIDP-IV, Pfohl *et al.*, 1995) was used to assess personality disorders. The SIDP uses the `five year rule,' meaning that behaviors, cognitions and feelings that have predominated for most of the last five years are considered to be representative of the individual's long-term personality functioning. In this interview, each DSM-IV criterion is scored as 0, absent; 1, subthreshold; 2, present; 3, strongly present. In the present study, it was necessary to collapse the top two categories because of sparse data. Interviewers (mostly psychology students in the final part of their training and experienced psychiatric nurses) were trained by professionals (one psychiatrist and two psychologists) with extensive previous experience with the instrument. The interviews, largely conducted face-to-face, were carried out between June 1999 and May 2004. For practical reasons, 231 interviews (8.3%) were done over the telephone. Each twin in a pair was interviewed by different interviewers. The non-independence of twin pairs was modeled by fitting separate models for each member of the twin pairs.

Statistical analysis

Models were fit using Mx (Neale *et al.*, 2002). Measurement differences across age, sex and age by sex interaction were assessed using single-group item-factor analysis (Wirth and Edwards, 2007) with covariates (Neale *et al.*, 2006; Kubarych *et al.*, 2008). A path diagram of the model used is depicted in Figure 1. In Figure 1, boxes (□) represent observed variables (the NPD criteria); solid-line circles (◯) represent factors; triangles (▵) represent unit constants for estimating means; diamonds (◇) represent covariates (age, sex, and age × sex interaction); broken line circles represent special nodes used to estimate the covariate moderation effects; single-headed arrows indicate linear regression effects and double-headed arrows represent variances and covariances. In the top section of Figure 1, labeled `1', covariate effects for the factor mean (B→) and variance (D→) are shown using the definition variables and special nodes (DF2). Sections `2a' and `2b' comprise the measurement portion of the model. `2a' identifies the NPD criteria factor loadings and their covariate moderation. Factor loadings are labeled $L_{\#}$ with their corresponding covariate moderation effects denoted $J_{\#}$. The J values provide estimates of the direction and magnitude of each covariate effect on each NPD criterion loading that differentially functions given the covariate effects at the factor level. Section `2b' shows how the threshold locations and their differential moderation ($m_{\#}$) are obtained. With ordinal data, covariate moderation parameters $k_{\#}$ estimate the changes in thresholds due to the effects of the covariates on the factor mean. Separate MZ and DZ correlations ($r_{\mathrm{MZ}}/r_{\mathrm{DZ}}$) are allowed for the twin1-twin2 common NPD factors (F1 and F2). Specific variances ($r_{\#}$) for each criterion are obtained by formulae calculation. Correlations across twins for residuals for the same criteria are also estimated. Parameter labels with subscripts (e.g. the B_1, D_1, L_{i}, J_{i}, and K_{i}) indicate parameters constrained to be equal across twin 1 and twin 2. Labels without subscripts can take different values; for example, Cov1 and Cov2 indicate that a covariate age can take on different values for members of a twin pair.

Rather than dichotomizing age into young versus old, age was rescaled based on the 19 to 36 sample range to a range between 0 and 1 and modeled as a covariate. Similarly, rather than treating males and females as separate groups as in a multi-group analysis, males and females were treated as a single group with sex modeled as a second covariate. Age-by-sex interaction (computed by multiplying age and sex) was also studied as a covariate. The covariates (age, sex, or age × sex interaction) may affect: (i) the mean of the factor (i.e. is the mean of the latent variable different in the males versus the females? Is the mean

different in younger versus older subjects?); (ii) the variance of the factor; (iii) the factor loadings (i.e. the regressions of the items on the factor); and (iv) the means of the individual items. Differences in the factor mean and variance across covariates are`genuine' if and only if measurement is invariant across the groups being compared.

To determine if measurement is invariant, one looks at differences in the factor loadings or in the item thresholds, which are changes in the functioning of the measurement instrument. One compares the fit of a model that specifies the effects of a covariate (age or sex in this case) on the factor loadings against a model that specifies the same covariate effects on the latent variance. One also compares the fit of a model that specifies covariate effects on the item thresholds, to one that specifies covariate effects on the latent factor mean. The same comparisons are also made with age and sex both in the model, and with age, sex, and age × sex interaction all in the model. Details of the procedure can be found in Kubarych *et al.* (2008).

Statistical modeling makes use of various criteria for choosing between different models. We present three commonly used criteria. The first is minus two times the logarithm of the likelihood function ($-2\ln L$). This statistic is based on the `likelihood' or joint probability of the data for particular parameter values; taking the logarithm and multiplying by −2 yields a statistic that is useful for model comparison. Second, the difference between the $-2\ln L$ statistics of two nested models is, under certain regularity conditions, asymptotically distributed as chi-square, with degrees of freedom equal to the difference between the number of parameters in the two models (MacCallum, 1995). The last, Akaike's Information Criteria (AIC), is called an `information-theoretic' criterion because it emphasizes minimizing the amount of information required to express the data in the model, therefore favoring parsimonious representations of the data. Lower (more negative) values of information theoretic criteria such as AIC reflect a `better' balance of parsimony and explanatory power in models of the data (Akaike, 1987). AIC and $-2\ln L$ are indexes suitable for the comparison of nested models.

Bootstrapping is an empirical resampling method which uses the available data to obtain confidence intervals on parameter estimates. Cases from the given data set are resampled with replacement. Since cases are resampled with replacement from the full sample, the same case can appear more than once in the resampled dataset, and the composition of cases varies somewhat across the generated samples. Repeating the procedure a large number of times simulates drawing numerous samples from a population. The standard error of the generated samples is an estimate of the variability of the sample means around the population mean (Kline, 1998).

Results

DSM-IV diagnostic criteria for NPD, on which the current study is based, are given in Table 1, along with the frequencies and sample proportions for endorsing each criteria for males, females and total sample. We performed a one-factor confirmatory factor analysis of the nine criteria separately for males, females and total sample. The results are displayed in Table 2. The Mplus program and the robust weighted least square mean and variance adjusted estimator were used (Muthen and Muthen, 2004). We concluded that it was reasonable to treat the criteria as defining a unidimensional construct.

We first performed simultaneous tests of whether allowing the means and variances to be freely estimated across covariates resulted in better model fit compared to a baseline model where nothing was free to vary, and whether allowing thresholds and factor loadings to be freely estimated resulted in better model fit compared with the model with means and

variances freely estimate. These tests indicated statistically significant model improvements. To further isolate whether means or variances contributed more to model improvement, and whether thresholds or loadings contributed more to model improvement, we then tested for effects on means separately from effects on variances, and effects on thresholds separately from tests on loadings. The results are shown in Table 3. A full measurement invariance baseline model in which means, variances, thresholds and loadings are constrained to be equal in males and females and across age results in a $-2\ln L$ of 16 314.30 on 24 777 degrees of freedom, with an AIC of −33 239.70 (model 1 in Table 3). Using model 1 as a baseline for comparison, model 2 tests how much better the fit of the model to the data is when the age parameter is freely estimated. Model 2 improves the fit of the model to the data by 6.03 chi-squared units on one degree of freedom. There is also an improvement in AIC, indicating a more parsimonious model. The effect size is −0.27, indicating that there is a linear decrease in the mean with increasing age.

Model 3 in Table 3 performs the same test for sex. The improvement in model fit when the mean of the latent factor is no longer constrained to be equal across sex is much larger than was the case for age: 53.13 chi-squared units on one degree of freedom; AIC indicates that model 3 is more parsimonious than either model 1 or model 2. In model 4, both age and sex are included in the model at the same time, resulting in a two degree of freedom test. Model 4 improved model fit by 57.25 chi-squared units, and has a slightly lower (more negative) AIC than model 3, indicating that it yields a slightly better fit to the data. Model 5, which includes age, sex and age × sex interaction, improves model fit only slightly over model 4, and the slightly higher AIC indicates that model 4 provides the most parsimonious fit to the data. These results indicate differences at the *factor* level. It is at the criteria level that we have to assess measurement non-invariance (MNI).

Models 6–9, which examine the effect of covariates at the *criteria* level, test for measurement non-invariance. Model 6 is one in which, in addition to age being allowed to affect the mean of the latent factor as in model 2, the item thresholds are not constrained to be equal across age. This model can be compared against model 2, in which only the mean is free to vary across age. This comparison test is statistically significant, indicating that allowing thresholds to vary across age results in better fit to the data than merely allow the mean to vary across age. Measurement is therefore not invariant with respect to age. Similarly, model 7 allows thresholds to vary across sex. Comparing model 7 to model 3, results in a much larger (148.41 chi-squared units on 8df) improvement in model fit, and a much better AIC. Including the effects of both age and sex on the thresholds in the model (model 8) again results in better model fit AIC compared to only allowing the covariates to affect the latent mean. Finally, adding age × sex interaction makes a small improvement in model fit, but is less parsimonious than model 8 based on AIC. There are no statistically significant differences in factor loadings across covariates. Therefore we cannot conclude that the mean differences detected in models 2–5 at the factor level are real, because measurement at the criteria level varies across the covariates.

The next step is to determine which individual criteria thresholds are specifically affected and to what extent. We ran the same sequence of models as earlier, but tested each criterion individually rather than all the criteria together. The results for age are shown in Table 4. Measurement differs across age for three criteria: 1, 2 and 9. Allowing the threshold for criteria 1 (grandiose sense of self-importance), and criteria 2 (preoccupation with fantasies) the effect sizes (in *z*-score) are 0.40 and 0.57, respectively. A positive effect means that the threshold increases with age. A higher threshold indicates that it takes a greater amount of the latent variable for participants to endorse the item. The implication is not, therefore, that older individuals have less grandiose senses of self-importance or a less preoccupied with fantasies of unlimited success. The point is that measurement for these items do not function

in the same way for older versus younger individuals and that relative to the other criteria these items become progressively more `difficult' with age – higher levels of liability are required to endorse them. Similarly, the −0.46 effect on item 9 indicates that arrogant, haughty behaviors are not measuring NPD equivalently in older versus younger persons; the change in this criteria with age is in the opposite direction from those seen with items 1 and 2.

Table 5 shows the effects with regard to sex. Six of the nine criteria do not assess the latent trait equivalently in males and females. Note also that half of these effects are positive and the other half are negative. There is no consistent pattern of the thresholds being higher or lower in one sex versus the other. Controlling for the factor differences between the sexes, grandiose sense of self-importance, lack of empathy and arrogant, haughty behaviors are more readily endorsed by males, while requiring excessive admiration, entitlement and enviousness are more readily endorsed by women, regardless of standing on the latent trait. The results with both age and sex in the model at the same time are displayed in Table 6. Seven of the nine criteria exhibit MNI with respect to age, sex, or both, with the lack of empathy item having the biggest difference across sex. We also ran the same sequence of models with age, sex and age × sex interaction. The results are displayed in Table 7.

Finally, we can make the crucial test of whether or not the statistically significant factor mean differences in NPD with respect to age, sex and age × sex remain after adjusting for measurement non-invariance in thresholds across age and sex. The effects of age on the latent mean were no longer statistically significant ($\Delta\chi^2 = 0.61$ on 1df, $p = 0.436$). All other effects remained statistically significant.

Bootstraps

We obtained 476 bootstraps for the item thresholds; i.e. the data was resampled with replacement 476 times, allowing 476 estimates of the parameters from slightly varying samples in order to estimate confidence intervals. Results are presented in Figure 2. In Figure 2, the median (50th percentile) bootstrap estimate and 95% confidence intervals are drawn for each item, for both males (m) and females (f), with two time points (young and old) to represent the age effect (so there are four medians and confidence intervals per item). The medians and confidence intervals are drawn in green for the males and purple for the females. Lines connecting the medians for young versus old are drawn in the same color as the confidence intervals within both sexes, with slightly thicker line widths. Lines connecting the medians between young time points between males and females are drawn in dotted red lines, and lines connecting the age medians males and the females are drawn in dotted blue lines. Whereas the model fitting exercises with respect to the item thresholds in Tables 4–7 provide information about each item individually, bootstrapping provides information about all the items tested simultaneously, both on the items and on the latent mean and variance.

Results of bootstraps are consistent with the model fitting results. Note that all threshold estimates are very high – close to or above two standard deviations above the mean. Item 3 (believes he/she is special) and item 6 (is interpersonally exploitative) have no measurement non-invariance effects included; hence the lines between both age and sex are flat and overlaid. Item 1 (grandiose sense of self-importance) includes both age and sex adjusted thresholds; in Table 4, the effect sizes are 0.40 for age and 0.32 for sex. Item 2 (preoccupation with fantasies) has a threshold adjusted only for age effect (0.57 from Table 2); the positive sign means that older respondents have a higher threshold for endorsing this item. Item 4 (requires excessive admiration) has only a sex-adjusted threshold; the slope is negative corresponding to the −0.34 effect in Table 3. Item 5 (sense of entitlement) has

negative effects for both age (−0.26) and sex (−0.45). The effects of age on item 7 (lack of empathy) are negligible, but the effect of sex is strong, where women have a higher threshold. The sizes of the confidence intervals in the bootstrapping results are sensitive to the endorsement frequencies of each criterion, but this does not detract from the fact that the MNI effects were statistically significant in the model testing.

Discussion

In this paper, we present the first study to assess whether the DSM-IV criteria for NPD assess the construct equivalently across age and sex. It is critical in research to know whether differences between groups are real or reflect measurement artifacts. An apparently statistically significant age difference in latent means was no longer statistically significant after correcting for measurement non-invariance. Nevertheless, future research in psychiatry should test for measurement non-invariance before drawing conclusions about differences between groups. Women did have a statistically significantly lower mean for liability to NPD than men even after accounting for measurement non-invariance of the items.

We found that the nine DSM-IV NPD criteria were unidimensional in both men and women. This is consistent with the findings of Miller *et al.* (2007) but not with the findings of Fossati *et al.* (2005). Since our sample is population-based, our results are not directly comparable to the samples used in these two studies. The factor structure of these criteria appears to be sample dependent.

Most importantly, after controlling for effects of age, sex and age × sex interaction at the level of the latent factor, seven of the nine DSM NPD criteria were found to have measurement non-invariance across age, sex or both. This has important implications for clinicians. The grandiosity, preoccupation with fantasies, and lack of empathy criteria have statistically significantly higher thresholds in older persons than in younger persons. This means that it requires a higher level of the underlying trait for older subjects to endorse these items. Therefore older patients may not endorse these items even though their level of the latent trait equals or exceeds younger patients who do endorse them. Similarly, men require a higher level of the latent trait before they can be expected to endorse the item concerning need for admiration. Thus, clinicians may want to attend differently to these symptoms in older versus younger patients or in men versus women. Alternatively, new criteria could be developed for NPD or other personality disorders which minimize variation in item performance across sex and age.

Limitations

The present results pertain to the DSM-IV NPD criteria and the latent factor they define. The psychoanalytic construct of narcissism is difficult to operationalize in a standardized set of criteria, and the current criteria are not the only possible criteria. Although our sample is relatively large and epidemiologic, it is restricted to relatively Norwegian twins with a relatively modest age range. Our sample has also undergone attrition, but this does not appear to affect analyses of mental health related variables (Tambs *et al.*, 2009). Finally, age is confounded with cohort in this sample. The possibility that age effects are actually related to social trends in this period cannot be ruled out.

References

Akaike H. Factor analysis and AIC. Psychometrika. 1987; 52:317–332.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed.. American Psychiatric Association; 1980.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed.. American Psychiatric Association; 2000. text revision

Cain NM, Pincus AL, Ansell EB. Narcissism at the crossroads: phenotypic description of pathological narcissism across clinical theory, social personality psychology, and psychiatric diagnosis. Clinical Psychology Review. 2008; 28:638–656. [PubMed: 18029072]

Costa, PT.; Widiger, TA., editors. Personality Disorders and the Five-factor Model of Personality. American Psychological Association; 2002.

Fossati A, Beauchaine TP, Grazioli F, Carretta I, Cortinovis F, Maffei C. A latent structure analysis of Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Narcissistic Personality Disorder criteria. Comprehensive Psychiatry. 2005; 46:361–367. [PubMed: 16122536]

Harris JR, Magnus P, Tambs K. The Norwegian Institute of Public Health twin panel: a description of the sample and program of research. Twin Research. 2002; 5:415–423. [PubMed: 12613498]

Jacobwitz S, Egan V. The dark triad and normal personality traits. Personality and Individual Differences. 2006; 40:331–339.

Kernberg, O. Borderline Conditions and Pathological Narcissism. Jason Aronson; 1975.

Kernberg, O. The Psychiatric Clinics of North American 12(3): Narcissistic Personality Disorder. W.B. Saunders Company; 1989. The narcissistic personality disorder and the differential diagnosis of antisocial behavior.

Kline, RB. Principles and Practice of Structural Equation Modeling. The Guilford Press; 1998.

Kubarych TS. Self-deception and Peck's analysis of evil. Philosophy, Psychiatry and Psychology. 2005a; 12:247–255.

Kubarych TS. On studying evil. Philosophy, Psychiatry and Psychology. 2005b; 12:265–269.

Kubarych TS, Aggen SH, Hettema JM, Kendler KS, Neale MC. Assessment of generalized anxiety disorder symptoms in women in the National Comorbidity Survey and Virginia Twin Study of psychiatric and substance use disorders: a comparative study. Psychological Assessment. 2008; 20:206–216. [PubMed: 18778157]

Miller JD, Campbell WK, Pilkonis PA. Narcissistic personality disorder: relations with distress and functional impairment. Comprehensive Psychiatry. 2007; 48:170–177. [PubMed: 17292708]

Miller JD, Hoffman BJ, Campbell WK, Pilkonis PA. An examination of the factor structure of Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, narcissistic personality disorder criteria: one or two factors. Comprehensive Psychiatry. 2008; 49:141–145. [PubMed: 18243885]

Morey LC, Ochoa ES. An investigation of adherence to diagnostic criteria: clinical diagnosis of the DSM-III personality disorders. Journal of Personality Disorders. 1989; 3:180–192.

Neale, MC.; Boker, SM.; Xie, G.; Maes, HH. Mx: Statistical Modeling. Department of Psychiatry; Medical College of Virginia; Virginia Commonwealth University; Richmond, VA: 2002.

Neale MC, Aggen SH, Maes HH, Kubarych TS, Schmitt JE. Methodological issues in the assessment of substance use phenotypes. Addictive Behaviors. 2006; 31:1010–1034. [PubMed: 16723188]

Paulhus DL, Williams KM. The dark triad of personality: narcissisn, Machiavellianism and psychopathy. Journal of Personality. 2002; 36:556–563.

Pfohl, B.; Blum, N.; Zimmerman, M. Structured Interview for DSM-IV Personality (SIDP-IV). Pamphlet; Department of Psychiatry; University of Iowa; Iowa City, IA: 1995.

Ronningstam E, Gunderson J, Lyons M. Changes in pathological narcissism. American Journal of Psychiatry. 1995; 152:253–257. [PubMed: 7840360]

Tambs K, Rønning T, Prescott CA, Kendler KS, Reichborn-Kjennerud T, Torgersen S, Harris JR. The Norwegian Institute of Public Health Twin Study of Mental Health: examining recruitment and attrition bias. Twin Research and Human Genetics. 2009; 12:158–168. [PubMed: 19335186]

Torgersen, S. Epidemiology. In: Oldham, JM.; Skodal, AE.; Bender, DS., editors. The American Psychiatric Publishing Textbook of Personality Disorders. American Psychiatric Publishing; 2005.

Wirth RJ, Edwards MC. Item factor analysis: current approaches and future directions. Psychological Methods. 2007; 12:58–79. [PubMed: 17402812]

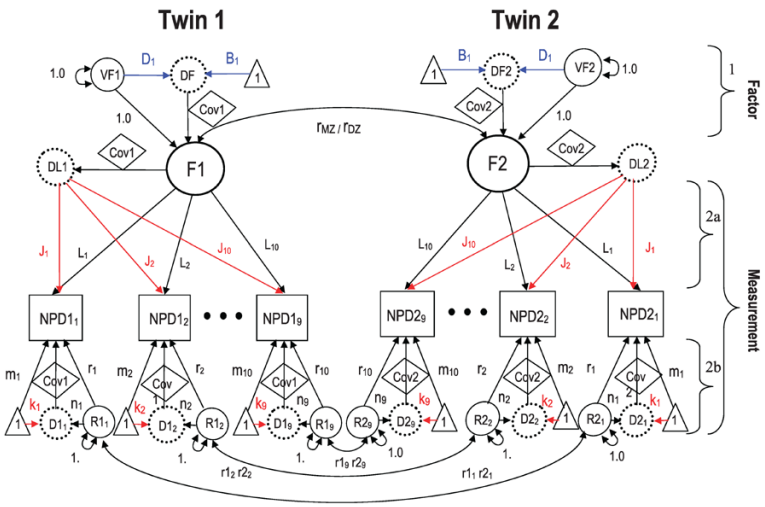

Figure 1.

Path diagram for single group item-factor model with covariates for twins. (☐) = observed variables, (◯) = unobserved variables (factors), (△) = unit constants for estimating means and threshold covariate effects, (◇) = definition variables for incorporating covariate effects (e.g. Cov), broken line circles = special nodes used to estimate the covariate moderation effects (e.g. DF and DL), (®) = linear regression effects, (blue = factor level effects, red = differential criterion level moderation effects), (↔) = variances and covariances with `1.0' indicating fixed values, VF = factor variance, r_{MZ}/r_{DZ} = estimated monozygotic (MZ) and dizygotic (DZ) twin1/twin2 factor correlations, $r1_i r2_i$ = twin1/twin2 correlations between same NPD criterion residuals.

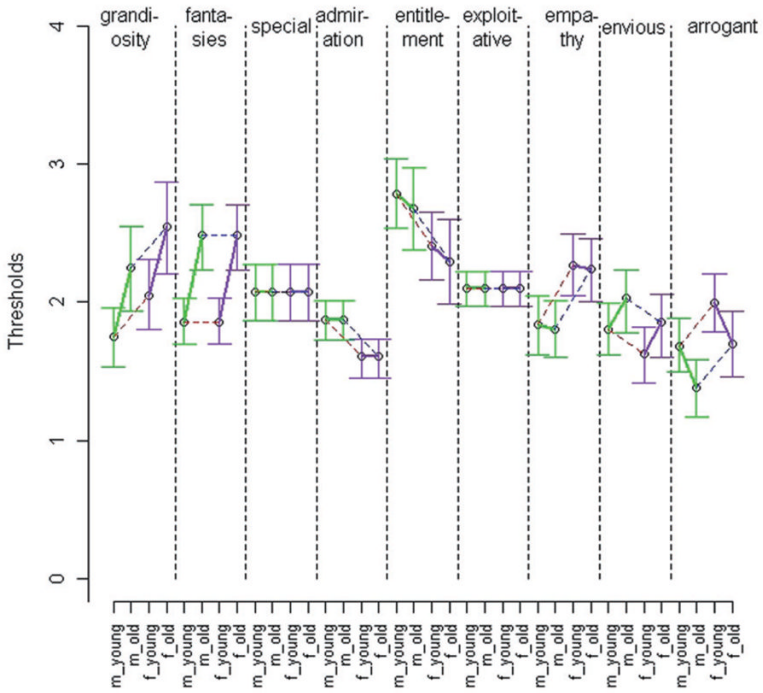

Figure 2.

Median parameter estimates and 95% confidence intervals derived from 730 samples for thresholds of DSM-IV NPD symptom criteria in males (m) and females (f). Note: Medians and confidence intervals drawn in green for males and purple for females. Lines connecting medians for young versus old are drawn in the same color as the confidence intervals within both sexes, with slightly thicker line widths. Lines connecting medians between young time points between males and females are drawn in dotted red lines, and lines connecting the age medians older males and older females are drawn in dotted blue lines.

Table 1

Summary of interviewer ratings used to categorize the degree of presence of each of the nine DSM-IV NPD criteria in male and female Norwegian twins

		Rating categories					
		Frequencies			Proportions		
Response in category		**0**	**1**	**2**	**0**	**1**	**2**
DSM-IV NPD criteria	Sex						
1. Has grandiose sense of self-importance	Male	992	80	20	90.22	7.83	1.96
	Female	1717	45	10	96.90	2.54	0.56
2. Is preoccupied with fantasies of unlimited success	Male	903	104	15	88.36	10.18	1.47
	Female	1639	121	12	92.49	6.83	0.68
3. Believes he or she is 'special' and unique	Male	984	19	19	96.28	1.86	1.86
	Female	1729	27	16	97.57	1.52	0.90
4. Requires excessive admiration	Male	835	160	27	81.70	15.66	2.64
	Female	1421	291	69	80.19	16.42	3.39
5. Has sense of entitlement	Male	983	35	3	96.28	3.43	0.29
	Female	1686	78	8	95.15	4.40	0.45
6. Is interpersonally exploitative	Male	876	130	16	85.71	12.72	1.57
	Female	1616	144	12	91.20	8.13	0.68
7. Lacks empathy	Male	862	123	37	84.34	12.04	3.62
	Female	1670	90	12	94.24	5.08	0.68
8. Is often envious of others or believes that others are envious of him or her	Male	868	128	26	84.93	12.52	2.54
	Female	1523	205	44	85.95	11.57	2.48
9. Shows arrogant, haughty behaviors or attitudes	Male	821	135	66	80.33	13.21	6.46
	Female	1616	123	33	91.20	6.94	1.86
Totals	Male	8124	914	229	88.32	9.94	2.49
	Female	14617	1124	216	91.65	7.05	1.35

Note: 0 = not at all, 1 = sub-threshold 2 = present. N = 1022 (males), 1772 (females).

Table 2

Fit statistics for one-factory confirmatory factor analysis of nine DSM-IV NPD symptoms in Norwegian twins

Fit index	Males	Females
Comparative Fit Index	0.969	0.967
Tucker–Lewis Index	0.967	0.967
Root Mean Square Error of Approximation	0.026	0.026
Weighted Root Mean Square Residual	0.766	0.851

N = 1022 (males), 1772 (females).

Table 3

Global model comparisons for testing age, sex, age and sex, and age, sex and age-by-sex interaction on latent mean and variance of DSM -IV NPD factor and thresholds and factor loadings of DSM-IV NPD criteria

Model	−2ln*L*	df	Δ χ^2	Δ df	*p*	AIC	Effect size		
(1) Full measurement invariance (baseline)	16 314.30	24 777	–	–	–	−33 239.70	–	–	–
Effects on latent mean									
(2) Age (versus 1)	16 308.27	24 776	6.03	1	0.014	−33 239.73	−0.27	–	–
(3) Sex (versus 1)	16 261.17	24 776	53.13	1	0.000	−33 290.83	–	−0.39	–
(4) Age and sex (versus 1)	16 257.05	24 775	57.25	2	0.000	−33 292.94	−0.23	−0.39	–
(5) Age, sex and age × sex (versus 1)	16 256.33	24 774	57.97	3	0.000	−33 291.67	−0.34	−0.49	0.19
Effects on thresholds									
(6) Age (versus 2)	16 274.63	24 768	33.68	8	0.000	−33 275.37	–	–	–
(7) Sex (versus 3)	16 112.76	24 768	148.41	8	0.000	−33 423.24	–	–	–
(8) Age and sex (versus 4)	16 073.42	24 759	183.64	16	0.000	−33 444.58	–	–	–
(9) Age, sex and age × sex (versus 5)	16 066.26	24 750	190.07	24	0.000	−33 433.74	–	–	–
Effects on latent variance									
(10) Age (versus 1)	16 313.82	24 776	0.46	1	0.483	−33 238.18	0.06		
(11) Sex (versus 1)	16 301.52	24 776	12.78	1	0.000	−33 250.48	–	0.17	
(12) Age and sex (versus 1)	16 301.10	24 775	13.20	2	0.001	−33 248.90	0.06	0.17	
(13) Age, sex and interaction (versus 1)	16 300.84	24 774	13.46	3	0.004	−33 247.16	−0.02	0.20	0.12
Effects on factor loadings									
(14) Age (versus 10)	16 310.49	24 768	3.36	8	0.910	−33 225.51	–	–	–
(15) Sex (versus 11)	16 293.55	24 768	7.99	8	0.434	−33 242.45	–	–	–
(16) Age and sex (versus 12)	16 290.28	24 759	10.83	16	0.820	−33 227.72	–	–	–
(17) Age xsex interaction (versus 13)	16 284.62	24 750	16.22	24	0.880	−33 215.38	–	–	–

Effect sizes in *z*-scores.

Table 4

DSM-IV NPD item-by-item tests for effects of age on item thresholds

	−2ln*L*	Δ χ^2	Δ df	*p*	AIC	Effect on threshold
Model						
Age, sex and interaction on latent mean and variance (baseline)	16 250.29	64.01	6	0.000	−33 291.71	
Item free						
Has grandiose sense of self-importance	16 244.32	5.96	1	0.015	−33 295.67	0.40
Is preoccupied with fantasies of unlimited success	16 234.79	15.49	1	0.000	−33 305.21	0.57
Believes he or she is 'special' and unique	16 250.32	−0.03	1	NA	−33 289.68	0.01
Requires excessive admiration	16 249.94	0.35	1	0.555	−33 290.06	−0.07
Has sense of entitlement, i.e. unreasonable expectations of favorable treatment	16 247.93	2.36	1	0.124	−33 292.07	−0.28
Is interpersonally exploitative, i.e. takes advantage of others	16 248.95	1.33	1	0.248	−33 291.05	−0.16
Lacks empathy: is unwilling to recognize/identify with feelings of others	16 249.69	0.59	1	0.441	−33 290.30	−0.11
Is often envious of others or believes that others are envious of him or her	16 248.96	1.33	1	0.249	−33 291.04	0.15
Shows arrogant, haughty behaviors or attitudes	16 239.27	11.02	1	0.001	−33 300.73	−0.46

N = 2794 (1772 female, 1022 male). Females coded as one. Effect sizes in *z*-scores. NA, not available.

Table 5

DSM-IV NPD item-by-item tests for effects of sex on item thresholds

	−2ln*L*	Δ χ^2	Δ df	*p*	AIC	Sex effect on threshold
Model						
Age, sex and age × sex on latent mean and variance (baseline)	16 250.29	64.01	6	0.000	−33 291.71	
Item free						
Has grandiose sense of self-importance	16 233.83	16.46	1	0.000	−33 306.17	0.32
Is preoccupied with fantasies of unlimited success	16 250.35	−0.06	1	NA	−33 289.65	0.00
Believes he or she is 'special' and unique	16 248.15	2.13	1	0.144	−33 291.85	−0.14
Requires excessive admiration	16 217.86	32.45	1	0.000	−33 322.14	−0.34
Has sense of entitlement, i.e. unreasonable expectations of favorable treatment	16 224.14	26.14	1	0.000	−33 315.86	−0.45
Is interpersonally exploitative, i.e. takes advantage of others	16 249.72	0.57	1	0.451	−33 290.28	0.05
Lacks empathy: is unwilling to recognize/identify with feelings of others	16 210.49	39.80	1	0.000	−33 329.51	0.44
Is often envious of others or believes that others are envious of him or her	16 236.52	13.77	1	0.000	−33 303.48	−0.24
Shows arrogant, haughty behaviors or attitudes	16 221.51	28.72	1	0.000	−33 318.49	0.35

N = 2794 (1772 female, 1022 male). Females coded as one. Effect sizes in *z*-scores. NA, not available.

Table 6

DSM-IV NPD item-by-item tests for effects of age and sex on item thresholds

	−2ln*L*	Δ χ^2	Δ df	*p*	AIC	Age threshold	Sex threshold
Model							
Age, sex and age × sex on latent mean and variance (baseline)	16 250.29	64.01	6	0.000			
Item free							
Has grandiose sense of self-importance	16 228.06	22.22	2	0.000	−33 309.93	0.40	0.32
Is preoccupied with fantasies of unlimited success	16 234.82	15.47	2	0.000	−33 303.18	0.57	0.00
Believes he or she is 'special' and unique	16 248.21	2.07	2	0.355	−33 289.78	0.03	−0.14
Requires excessive admiration	16 217.30	32.99	2	0.000	−33 320.70	−0.09	−0.34
Has sense of entitlement, i.e. unreasonable expectations of favorable treatment	16 221.97	28.32	2	0.000	−33 316.03	−0.26	−0.45
Is interpersonally exploitative, i.e. takes advantage of others	16 248.41	1.87	2	0.392	−33 289.58	−0.16	0.05
Lacks empathy: is unwilling to recognize/identify with feelings of others	16 210.00	40.28	2	0.000	−33 327.99	−0.10	0.44
Is often envious of others or believes that others are envious of him or her	16 235.30	14.99	2	0.001	−33 302.70	0.14	−0.24
Shows arrogant, haughty behaviors or attitudes	16 210.84	39.45	2	0.000	−33 327.16	−0.46	0.34

N = 2794 (1772 female, 1022 male). Females coded as one. Effect sizes in *z*-scores.

Table 7

DSM-IV NPD item-by-item tests for effects of age, sex and interaction together on item thresholds

	−2ln*L*	Δ χ^2	Δ df	*p*	AIC	Age mean Age threshold	Sex mean Sex threshold	Interation mean Interaction threshold
Model								
Age, sex, and interaction on latent mean and variance (baseline)	16 250.29	64.01	6	0.000	−33 291.71			
Item free								
Has grandiose sense of self-importance	16 226.50	23.78	3	0.000	−33 309.50	0.60	0.52	−0.41
Is preoccupied with fantasies of unlimited success	16 234.69	15.60	3	0.001	−33 301.31	0.45	−0.10	0.19
Believes he or she is 'special' and unique	16 248.23	2.06	3	0.560	−33 287.77	0.03	−0.13	−0.04
Requires excessive admiration	16 216.97	33.31	3	0.000	−33 319.03	−0.19	−0.42	0.15
Has sense of entitlement, i.e. unreasonable expectations of favorable treatment	16 221.03	29.25	3	0.000	−33 314.96	−0.04	−0.28	−0.32
Is interpersonally exploitative, i.e. takes advantage of others	16 247.49	2.80	3	0.424	−33 288.51	−0.32	−0.10	0.28
Lacks empathy: is unwilling to recognize/ identify with feelings of others	16 208.48	41.80	3	0.000	−33 327.52	0.08	0.65	−0.38
Is often envious of others or believes that others are envious of him or her	16 235.12	15.17	3	0.002	−33 300.88	0.08	−0.29	0.11
Shows arrogant, haughty behaviors or attitudes	16 210.59	39.70	3	0.000	−33 325.41	−0.41	0.40	−0.10

N = 2794 (1772 female, 1022 male). Females coded as one. Effect sizes in *z*-scores.