The Position of Anxiety Disorders in Structural Models of Mental Disorders

Hans-Ulrich Wittchen, PhD*, Katja Beesdo, PhD, Andrew T. Gloster, PhD

Keywords: Diagnostic classification, Structural models, Internalizing, Externalizing, Anxiety disorders

“Comorbidity” among mental disorders is commonly observed in both clinical and epidemiological samples. The robustness of this observation is rarely questioned; however, what is at issue is its meaning. Is comorbidity “noise” – nuisance covariance that researchers should eliminate by seeking “pure” cases for their studies – or a “signal” – an indication that current diagnostic systems are lacking in parsimony and are not “carving nature at its joints?” (Krueger, p. 921).1

With these words, Krueger1 started a discussion on the structure of mental disorders, which suggested that a 3-factor model of common mental disorders existed in the community. These common factors were labeled “anxious-misery,” “fear” (constituting facets of a higher-order internalizing factor), and “externalizing.” Along with similar evidence from personality research and psychometric explorations2–4 and selective evidence from genetic and psychopharmacologic studies,5–10 Krueger1 suggested that this model might not only be phenotypically relevant, but might actually improve our understanding of core processes underlying psychopathology. Since then, this suggestion has become an influential, yet also controversial topic in the scientific community, and has received attention particularly in the context of the current revision process of the Manual of Mental Disorders (Fifth Edition) (DSM-V) and the International Classification of Diseases, 11th Revision (ICD-11).11,12

Focusing on anxiety disorders, this article critically discusses the methods and findings of this work, calls into question the model’s developmental stability and utility for clinical use and clinical research, and challenges the wide-ranging implications that have been linked to the findings of this type of exploration. This critical appraisal is intended to flag several significant concerns about the method. In particular, the concerns center around the tendency to attach wide-ranging implications (eg, in terms of clinical research, clinical practice, public health, diagnostic nomenclature) to the undoubtedly interesting statistical explorations.11–13

INTERNALIZING AND EXTERNALIZING AS CORE DIMENSIONS OF THE STRUCTURE OF MENTAL DISORDERS

The Approach

Krueger1 used the epidemiological diagnostic data from the US National Comorbidity Survey (NCS)14 to analyze patterns of comorbidity among DSM-III-R mental disorders. Ten diagnoses were selected for this analysis from a wider set of all diagnoses in the NCS, namely: major depressive episode (MDE), dysthymia (DYS), panic disorder (PD), agoraphobia (AGPH), social phobia (SOP), simple phobia (SIP), generalized anxiety disorder (GAD), alcohol dependence (AD), drug dependence (DD), and antisocial personality disorder
All 10 disorders were submitted to confirmatory factor analysis (CFA), a formal statistical means of evaluating dimensional accounts of comorbidity among mental disorders. Four competing models, positing from 1 latent factor to 4 latent factors, were evaluated to determine their fit in the entire NCS sample and various subsamples (Fig. 1).

Krueger’s main findings were noteworthy. For the entire NCS sample, across sexes, and across random halves, a 3-factor model provided the best fit to the correlations among the 10 disorders. These factors were labeled “anxious-misery,” “fear,” and “externalizing.” Because of the high correlation between the anxious-misery and fear factors in the 3-factor model (0.73), these factors were described as meaningful expressions of subfactors of a higher-order “internalizing” factor. Consistent with this, the 2-factor internalizing-externalizing model was also found to provide the best fit in a treatment-seeking subsample of his exploration. Thus, at the highest level of the factor hierarchy, a broad internalizing factor (with anxious-misery and fear subfactors) and a broad externalizing factor explained the pattern of correlations among mental disorders.

The Conclusion

Krueger concluded that his findings provide evidence for (a) refocusing the study of common mental disorders on their common substrates, namely broad, higher-order internalizing and externalizing dimensions, and (b) that the model helps to organize observations of psychopathology by suggesting directions for future research in psychiatric epidemiology, psychopharmacology, and psychiatric genetics in several ways (Krueger, 1 pp 925–926): First, regarding psychiatric epidemiology, it is suggested that the model would organize common psychopathological variance into psychologically sensitive patterns labeled internalizing, such as pervasive anxiety and sadness (anxious-misery) and phobic avoidance (fear), and externalizing patterns involving antisocial behaviors (APD) and lifestyles (AD, DD).

Psychologically speaking, the model suggests that maladjustment can be expressed primarily inward, as anxious-misery or fear, or primarily outward, as antisocial, disruptive behavior. Second, according to the author’s conclusions, the model might suggest directions for psychopharmacology because it “predicts” the observed effectiveness of similar interventions for putatively different common mental disorders. For example, the selective serotonin reuptake inhibitors (SSRIs), although initially regarded as an antidepressant, have been found to be effective in treating other internalizing conditions, such as PD and DYS. The author claims that SSRIs may be effective in treating all these conditions because they influence a core internalizing process – perhaps the personality trait of neuroticism/negative emotionality, which has also been shown to be reduced by SSRI administration. Third, the author claimed that the model is consistent with genetic findings, because internalizing disorders share a considerable degree of genetic variance. As predicted by the model, MDE and GAD are closely linked, PD and the phobias are closely linked, and MDE and the phobias, although significantly linked, are linked to a lesser degree. In addition, APD and substance use disorders share significant genetic variance. Although it is admitted that multivariate and molecular genetic research framed in terms of the model is needed to verify this possibility, existing research is interpreted to suggest that the model may organize common psychopathological variance by shared genetic etiology.

BEYOND THE INTERNALIZING-EXTERNALIZING MODEL: ADDITIONAL IMPACT AND FAR-REACHING IMPLICATIONS FOR DSM-V AND ICD-11
Since the publication of the article by Krueger— and despite a range of critical concerns— this approach has received considerable attention and extensions. Positive replications of the internalizing-externalizing structure clearly outnumber failures, particularly if identical or only a slightly different set of diagnoses and identical methods were used (see Refs. 12,16). However, it should be stated up front that there were also several failures to replicate this structure for various reasons, and there are also some studies that revealed different structures, some of which suggest a considerably larger number of classes.16,17

In light of the seemingly attractive possibility of simplifying our current diagnostic classification structure, the DSM-V Task Force chose this approach as one of several guiding principles with which to explore the option of simplifying the classificatory structure of the future DSM and ICD.13 By broadening the model to cover a wider range of disorders, and ultimately the whole spectrum of mental disorders currently listed in the DSM,13,18 a new “metastructure of mental disorders” consisting of 5 “clusters” was suggested in the context of the DSM-V revision process (see special issue of Psychological Medicine, 2009, in press). In addition to an “externalizing” and “internalizing” (labeled “emotional”) cluster, 3 further groupings were suggested as part of this proposition, labeled “Neurocognitive,” “Neurodevelopmental” and “Psychotic.”

At first sight, this structural grouping of a metastructure appears considerably simpler than the current diagnostic classification structures used in the DSM Fourth Edition, Text Revision (DSM-IV-TR) (17 categories)18 and, in some parts, more convergent with the Tenth Revision of ICD (ICD-10) (10 categories).19 Responding to some dissatisfaction expressed with the current diagnostic classification structure and its mostly descriptive principles, the authors of the proposed metastructure11 made an impressive attempt to provide evidence for this position via a selective critical review of relevant research. In addition to extensive reference to CFA findings similar to Krueger’s original exploration,1 evidence was examined to address the question of whether current individual mental disorders sufficiently differ from each other, and whether the current more fine-grained distinction of specific mental disorders and their grouping in 10 major classes according to ICD-10 and 17 in DSM-IV-TR is justified. A priori criteria were chosen in the form of a wide range of “validators” grouped into so-called causal risk factors (ie, shared genetic risk factors, familiality, shared environmental risks, shared neural substrates, shared biomarkers, shared temperamental antecedents), and factors believed to reflect the clinical picture itself (shared abnormalities of cognitive and emotional processing, symptom similarity, rates of comorbidity, course, treatment response).11 The outcome of this admittedly selective exploration was interpreted as proof for the proposed 5 metastructure clusters. However, there was also a fairly large residual category of disorders not yet assigned. Particularly noteworthy examples for classificatory changes associated with the proposed metastructure involve: the suggestion to group anxiety, somatoform, and depressive disorders together under the term “emotional disorders”; the allocation of bipolar disorders to the “psychotic cluster”; and the formation of a broad externalizing cluster that comprises substance use disorders, some of the personality disorders, and impulse control disorders.

Several core motivations and promises are presented by the authors of the proposed new metastructure to justify their proposal.11

1. Increased utility to the degree that the current system has been found to be overly complicated and too complex for routine care, in particular by nonspecialists like primary care physicians. From this perspective, the proposed metastructure is also expected to provide
considerable simplification for research and a better adoption in routine care and public administration.

2. Increased validity and advantages for research, in the sense that proposed grouping would move classification of mental disorders away from one based purely on symptomatology and closer to current understanding of shared putative causal risk and clinical factors identified in research.

3. Increased homogeneity within clusters, in that the proposed metastructure would more appropriately highlight similarities rather than subtle differences between our current specific diagnoses.

FROM A FACTOR ANALYTIC EXPLORATION TO A CLASSIFICATORY METASTRUCTURE OF MENTAL DISORDERS?

Given this set of promises, the deeper meaning apparently ascribed to the metastructure, and the potentially far-reaching implications for research and practice, Wittchen and colleagues recently criticized and challenged these assumptions and predictions of the model in a commentary, highlighting a range of deficits and problems. The commentary emphasized that the claims and promises remain largely untested and unproven, and that evidence provided was selective, “biased toward” confirming the proposition. Further, doubts were raised about how methodologically sound and valid factor analytic approaches are in deriving a classification structure that is meaningful and sensitive to clinical issues. If the intent, motivation, and claims of the authors would have been more moderate, then the proposed metastructure would be a helpful step. For example, focusing primarily on the technical aspect of reducing the number of major groups of mental disorder without overly ambitious claims of a deeper meaning and increased clinical utility would have been useful. However, the goal of this proposal is much more ambitious, namely to develop a more parsimonious metastructure based on risk factors and clinical factors with the long-term goal of improving clinical practice, administration, research, and even training and education.

Wittchen and colleagues came to the conclusion that such a metastructure proposal might actually imply more risks for harm than benefits. For example, it may lead to an oversimplification of mental disorders, the deletion of disorders from the classification, premature adoption of putative common risk factors to the exclusion of other valid factors, diminished specificity with regard to recognition of established diagnoses, and the overemphasis of common treatment procedures that might in fact be very different. From a diagnostic perspective, the serious concern is that the broader, higher-order factor statistical approach might largely miss the subtle distinctions between disorders that constitute the essential difference for improved recognition and improved allocation of optimal treatment. Finally, from a technical perspective, the authors of this article are disquieted that numerous caveats and concerns regarding statistical and methodological limitations of higher-order factorial methods are not acknowledged. This disquiet also applies to concerns about the interpretation of results, the value and implications of these findings for general classificatory issues, and the propositions for a simplified metastructure of diagnostic classification systems based on these results.

A CRITICAL REAPPRAISAL OF THE „STRUCTURE OF MENTAL DISORDERS“

The new metastructure proposal that is meant to inform the structure of a clinical diagnostic classification system seems to be driven by the 2- and 3-factor solutions within the
internalizing and externalizing” model. The structure is largely based on epidemiologic data collected from diagnostic instruments and self-report data, and derived using higher-order factor analytic methods. The concerns expressed earlier prompted us to conduct a critical appraisal of these methods and findings.

With support from the DSM-V Task Force, we were commissioned to reexamine the assumptions, models, and findings of this line of research. Specifically, we examined the effects of using a considerably broader scope of diagnoses than originally included, examined the stability/instability of the factors in different age groups, and explored the value and limitations for specific diagnostic areas, with a particular emphasis on anxiety. In addition, we explored models and solutions that might be potentially more sensitive to clinical, developmental, and prospective-longitudinal epidemiological data and issues, to inform about an alternative and potentially more appropriate grouping of disorders. That is, we examined which disorders and to what degree sufficiently fit the proposed 3-cluster solution and which disorders do not. Although the proposal might have implications for various other diagnostic groups, emphasis for the purpose of this article was deliberately placed on anxiety disorders.

Method

The aims and methods of this comprehensive appraisal and corresponding findings have been described elsewhere in greater detail.21 In brief, we used 2 large epidemiological data sets: (a) the prospective-longitudinal Early Developmental Stages of Psychopathology (EDSP) study data22,23 with 4 waves of investigation (n = 3021, age range 14–34 years) and (b) the cross-sectional German National Health Interview and Examination Survey – Mental Health Supplement (GHS-MHS) (n = 4181, age range 18–65 years)24 to test the following 4 core aspects of the model.

1. Using CFA, we first attempted to replicate Krueger’s cross-sectional model (ie, factor solutions and model fit) using identical methods and conventions with respect to diagnoses, diagnostic assessment, and referent time frame of the assessment.

2. We examined the stability of factor solutions against minor and major additions to the diagnoses included in the model by examining the effects of adding additional diagnoses. Additional diagnoses were added in stepwise, grouped, and comprehensive manner. The additional diagnoses to those used in initial analyses were: group (a) hypomanic episodes (HME), manic episodes (MNE), separation anxiety disorder, obsessive compulsive disorder (OCD), posttraumatic stress disorder (PTSD), specific phobia subtypes; group (b) pain disorders, somatoform disorders (SSI4/6); group (c) psychotic disorders; group (d) eating disorders; and group (e) childhood disorders (attention deficit hyperactivity disorder [ADHD], oppositional defiant disorder [ODD], tics, elimination disorders).

3. We also examined the developmental stability of the model by examining whether the solution and model fits differ substantially by age group (ie, 14–17, 18–21, 22+). These tests were conducted for Krueger’s standard diagnostic set1 as well as a broader range of diagnoses.

4. Finally, we also examined the structure of DSM-IV anxiety disorders using exploratory factor analysis (EFA) to explore the existence of similarly acceptable solutions.

Model Specifications
Irrespective of the considerable range of general and specific statistical problems inherent in factor analytic approaches (i.e., violations of basic assumptions, which render all subsequent results suspect), we expected to replicate Krueger’s findings\(^1\) of 2-factor (internalizing versus externalizing) and 3-factor (anxious-misery, fear, externalizing) solutions only when using exactly the same restricted set of DSM diagnoses assessed cross-sectionally. In contrast, we did not expect to replicate the 2- and 3-factor solutions (i.e., insufficient model fit) when examining the sample with restricted age groups of young and older subjects. This method provides information as to whether the model is robust against developmental and age changes.

We also did not expect that the 2- or 3-factor model provides a sufficient model fit when adding other diagnoses to the model. The latter 2 cases provide information as to whether the model is robust against even minor modifications of the diagnoses included.

It should be noted that we do not separately address the considerably more general limitations of the factor analytic approach for classificatory diagnostic purposes, which have been expressed and discussed elsewhere in greater detail.\(^{15}\) However, in light of these severe limitations we emphasize that considerable caution is warranted when interpreting the implications of positive replications of the Krueger model\(^1\) on the one hand, and any alternative structure solution on the other.

**Replication: Can the 3-Factor Solution be Replicated if Exactly the Same Diagnoses are Used?**

In the first step of our examination,\(^{16}\) we used exactly the same conventions and diagnoses as in Krueger’s original work\(^1\) using cross-sectional (12-month) data. We found that, by and large, a 3-factor model fit the data in both samples (EDSP and GHS-MHS) used, separating fear, anxious-misery, and externalizing. However, Krueger\(^1\) suggested that anxious-misery and fear constitute second-order factors of a latent “higher” dimension called “internalizing.” This factoring could not be replicated consistently. Further, some of the factor loadings were much smaller than in Krueger’s work,\(^1\) which is particularly evident for the “fear” factor on specific phobia (0.43 vs 0.77). This result might indicate that “fear” is less strongly related to specific phobia in this replication sample. In general we found more heterogeneity between loadings of latent variables on diagnoses, probably due to the prospective-longitudinal nature of the EDSP data set used. However, with the factor loadings freely estimated, the Krueger\(^1\) model fits the data well.

In the lifetime data set, the model once again necessitated the omission of the higher-order factor “internalizing,” which resulted from the failure to differentiate the “fear” and “anxious-misery” factor as subordinate factors of internalizing. Overall, in both the EDSP and GHS-MHS samples the model fit was satisfactory only when the second-order latent factor “internalizing” was omitted (Table 1).

**Stability: How Robust is the 3-Factor Model Against the Additional Inclusion of Disorders?**

In the second step, additional disorders were assigned to the 3 assumed factors. The following diagnoses were added: group (a) HME, MNE, separation anxiety disorder, OCD, PTSD, specific phobia subtypes; group (b) pain disorders, somatoform disorders (SSI4/6); group (c) psychotic disorders; group (d) eating disorders; group (e) childhood disorders (ADHD, ODD, tics, elimination disorders). Assignment of diagnoses was explored by testing whether...
assignment of additional disorders to different factors led to different results. Models were tested using all diagnoses as well as sequentially adding them by group of disorders.

We found that the addition of anxiety diagnoses and HME and MNE to the Krueger\(^1\) model could not be fitted, or only just fitted when the factor „internalizing“ was omitted (Table 1). After adding the somatoform disorders the model fit became particularly poor. Adding psychotic disorder and childhood disorder further impaired the fit.

In conclusion, the model could not be appropriately fitted in any of the analyses when several or many additional diagnoses were added. This lack of fit suggests that the model is very sensitive to even minor changes and is thus not robust.

**Developmental Stability: How Robust is the 3-Factor Model when Examined Across Different Developmental Stages?**

In the third step, and based on considerable evidence that the comorbidity patterns in children and young adolescents differ considerably from those in older adults,\(^17,25–27\) we examined whether the factor solution and model fits substantially differ across the age span (1–13, 14–17, 18–21, 22+ years). These ages are considered a proxy for developmental stages ranging from childhood through adolescence to adulthood. Using the same diagnoses as in the original Krueger\(^1\) work, we found that the model fit was appropriate for some age groups, namely 14- to 17-year-olds and 18- to 21-year-olds, but does not fit in the younger or the older cohorts (Table 1). The age-related increase in comorbidity seems to yield an association structure that is neither consistent with nor in line with the 3-factor model. Where the model could be fit, its overall model fit was reasonable in some age groups only if the second-order factor „internalizing“ was ignored. To conclude, the model could not be consistently replicated in different age groups, despite some evidence for acceptable fits in some age groups. Thus, overall the model seems to be developmentally unstable.

**Is There a Better Factor Solution?**

Given that the 3-factor CFA solutions were neither robust nor developmentally stable, we also explored whether other meaningful and stable factor models could be identified when using EFA. Separate EFA were conducted in the total sample and different age groups as well as by including and excluding various groups of disorders. The scree test combined with parallel analyses suggested different numbers of factors across age spans, with up to 8 necessary factors. Moreover, the factor structures of the different models were not nested. That is, when more factors are extracted the factors did not simply add up to subfactors. Therefore, it seems that there is no one single model that would apply to all different age spans examined.

Nevertheless, the factor solutions may offer a range of clinically meaningful groupings, with some consistency across factor solutions. For example, the fairly consistent emergence of an externalizing-like factor (ie, substance dependence, ASP/CD), the frequent emergence of a panic/agoraphobia factor, a phobia factor (specific subtypes of phobia) and, less consistently and depending on age group considered, a psychotic factor, an ADHD/ODD/CD factor, and so forth. It is also noteworthy that some disorders such as OCD, psychotic, and hypomania/mania did not consistently reveal particularly high loading on one single factor but rather displayed moderate loadings on several factors. The factor structures became even more heterogeneous when adding childhood disorders and when running the analyses only among children and adolescents. For example, among the youngest cohort subjects a 4-factor solution emerged: The first factor loads high on MDE (0.86), dysthymia (0.51), OCD (0.50),
and eating disorders (0.48), the second on some specific phobias, the third on GAD, ADHD, ODD, and elimination disorders, and the fourth on panic, agoraphobia, separation anxiety, and some phobias.

An example of this exploration is presented in Fig. 2, namely a 6-factor solution from the total sample based on EFA with good fit values. Partly consistent with the 3-factor model, this model describes a first „externalizing“ factor (alcohol and drug dependence, conduct/ASP), and second factor that resembles the „anxious-misery“ factor (MDE, DYS, and GAD), albeit with the addition of OCD and eating disorders. Krueger’s1 „fear factor“ is, however, reflected by 3 factors, namely panic/agoraphobia, specific phobias (animal and environmental), and specific phobias (blood injury, situational, and other type). It is noteworthy that social phobia falls somehow in between factors 2 (anxious-misery) and 3 (panic and agoraphobia). In addition, we found a sixth factor that describes a clinically more heterogeneous group of disorders, namely separation anxiety, PTSD, pain and somatoform disorders, and psychotic disorder.

In conclusion, this exploration suggests various clinically meaningful patterns with good fit, which go substantially beyond the original Krueger1 structure. However, as with the Krueger structure, there is little consistency in findings when different age groups or different diagnostic groups are included or excluded. Further, it is noteworthy that a higher-order factor of „internalizing“ could be substantiated in any of the explorations. This finding suggests that even when reducing the association structure to a matrix of 2-way tetrachoric correlations (although there might be higher dimensional structures), there seems to be no simplified structure that is stable over and across age groups. Instead, there is strong evidence that there might be more than 3 dimensions behind expressions of psychopathology as assessed in this data set.

**Structures of DSM-IV Anxiety Disorders?**

Finally, we also subjected the anxiety disorders (including separation anxiety disorder and specific phobia subtypes) to an EFA to explore whether a meaningful model could be identified within anxiety disorders. The scree plots and parallel analyses consistently indicated 3-factor solutions with acceptable model fits overall and by age group considered. Despite some differences in the loadings and the grouping of specific disorders there are common findings. (a) Agoraphobia and PD were always found in the same factor. (b) GAD and social phobia were always found in the same factor. (c) In all 3-factor solutions from age 14 on (and in the total sample) GAD, social phobia, PTSD, agoraphobia, PD, OCD and separation anxiety where assigned, with some variations, to the same factor. (d) The specific phobia subtypes animal and natural environment are always assigned to the same factor; in the age spans 14 to 17 and 18 to 21 they form a factor on their own, in the age span 1 to 13 they cluster with GAD and social phobia, and in the age span 22 to 34 with the situational subtype. (e) The blood injection injury specific phobia subtype could not be clearly assigned to one single factor.

**DISCUSSION**

**Serious Limitation of the Internalizing-Externalizing Construct**

In the decade following Krueger’s initial landmark article1 there is still insufficient evidence for a meaningful, relatively simple structure of common mental disorders. The results of this article challenge the higher-order factor of „internalizing“ disorders with 2 subfactors of
„anxious-misery“ and „fear“ conditions. Overall, our findings strongly suggest that even sophisticated factor analytic methods are too vulnerable for the type of diagnostic and comorbidity data used in such explorations. We have demonstrated that such methods lead to findings that are neither sensitive for developmental variations in the expression of psychopathology, nor sufficiently robust against even minor modifications with regard to the type and range of diagnoses covered. Despite the existence of several replications of Krueger’s1 work,3 we and others17 were unable to replicate the model in the majority of analyses with sufficient confidence and integrity. This failure, along with further methodological constraints discussed later in this article, strongly suggests that it is unwise and inappropriate to draw strong conclusions based on such findings. This reservation applies in particular to diagnostic classificatory issues and the extension of this model in the context of a metastructure proposition for DSM-V and ICD-11 recently presented.11,12 These concerns further extend to the wide-ranging implications attached to the model, for example in terms of future research needs in psychiatric epidemiology, genetics, pharmacology, and intervention research. In particular, there is little evidence for the claim that this approach and the available explorative data justify the decision to refocus the study of common disorders on their common substrates, namely broad, higher-order internalizing and externalizing dimensions, simply because they might reflect „core psychopathological processes.“1

It is difficult to dispute that the internalizing-externalizing distinction is and has been a fertile phenotypic construct to study personality, temperament, and cross-sectional self-report on symptoms within a dimensional questionnaire assessment of patients from a psychometric and quantitative psychological perspective. As compared with categorical clinical diagnoses, the dimensional nature of latent constructs has also been shown to be advantageous with regard to identifying associations with validators, such as neurobiological and genetic variables. We also do not dispute that identifying dimensions that are common across disorders can be helpful in understanding some of the mechanisms involved in etiology, maintenance, and treatment of mental disorders.28

One should be seriously concerned, however, when this model and its findings are used as a basis for designing the structure of diagnostic classification, the development of diagnostic interviews, clinical decision making, and public health decisions. More generally speaking, one should be concerned about making far-reaching implications beyond the fact that this work represents a potentially useful phenotypic exploration. These concerns have been previously discussed in greater detail elsewhere15,20 and will not be presented here again. In brief, in terms of the data used the significant limitations refer to the use of categorical (instead of dimensional), threshold (instead of subthreshold and threshold) disorders based on structured diagnostic interviews, the very limited range of diagnoses covered, the lack of consideration to temporal and developmental issues, and the lack of data beyond subjective-verbal information from the respondent. In terms of the statistical method, limitations lie in the fact that confirmatory factor analytic methods typically allow for the derivation of quite different models, all of which might be justified by the data. There are simply too many unknown parameters (factor loadings, variances and covariances of residuals) in the presence of an insufficient number of equations. Basic problems (eg, the use of tetrachoric correlations, see Ref.29) of this sort force almost invariably a relatively arbitrary choice of models, most likely those that are easy to interpret or those that fit the researchers’ intention best. Therefore, as replication alone is insufficient and misleading, CFA may be used to reject or support a model but never to prove that a theory holds. In this respect, Krueger1 has provided some evidence that the model might hold for his restricted set of diagnoses and his particular set of analyses. Replications to date have also used a similar set and range of diagnoses.3 However, as shown in the data exploration of this article, the model evidently does not hold in general.
when other (even similar) diagnoses are taken into account, or even minor variations by age group are introduced. It is unlikely that the 2- and 3-factor models can be invalidated, but our exploration within the same limitations as in Krueger1 shows that there is a wide range of alternative, clinically meaningful solutions or typologies that might be regarded as equally well supported as the Krueger model. In particular, the „internalizing“ concept seems to be highly problematic, and only poorly reflects the structure of mental disorders and the structure of anxiety disorders in particular.

Of course, these restrictions should drastically limit the level of speculation attached to this type of exploration and the models derived. Yet, it is concerning to what degree the model continues to generate even further speculation and considerable extension not appropriately supported by data. Thus, beyond the possibility of identifying shared phenotypic expressions, extreme caution is warranted with regard to the potential clinical, cognitive-behavioral, genetic, and other neurobiological considerations, as well as with regard to suggested implications for interventions research, clinical practice and public health.

Is There an Alternative?

Having summarized the severe restrictions of the factor analytic approach for diagnostic classificatory issues and this type of investigations in general, the question of whether there is a better or alternative solution possible within such an approach has to be denied. These types of methods are not suitable for the definitive detection of latent dimensions, or the relationships between latent factors and their observed variables. This nonsuitability applies to the structure of anxiety disorders as well as the structure of mental disorders in general. In this respect, the presentation of our more comprehensive structural model (see Fig. 2) was meant only as an example that there are other possible structural models that can be justified with similar or even better values, and that may or may not be much more useful clinically. These models should in no way, however, be interpreted as proof of a better or even alternative model. Factor analytic models should simply not be used for this purpose. Yet, overall, we have little doubt that the data presented in our article suggest clearly that the assumption of the existence of more than 3 dimensions is much more likely to reflect expressions of psychopathology as the Krueger1 model.

The Metastructure and Implications for DSM-V and ICD-11?

Even though somewhat repetitious in argumentation, we now comment on the use or misuse of this model for the clinical classification of mental disorders in general. The fact that such factor analytic findings are currently seriously being considered to provide guidance for a DSM-V diagnostic classificatory metastructure11,12 is concerning.20 Within the metastructure proposal,11 the emotional cluster (almost identical to internalizing) lumps together all anxiety disorders along with depression and somatoform disorders under one umbrella. It is fair to state that such a proposition is actually not new and has in fact a long history. From the perspective of diagnostic classification, with its neurosis concept it resembles the ICD sixth to eighth revision, and is consistent with the old neurotic versus endogenous distinction as well as „disorders of temperament and personality“ in the case of the externalizing dimension. It is also largely consistent with quantitative psychology and psychometric evaluations of cross-sectional questionnaires and scales based on self-report, and the psychological tradition of trait and personality research. These approaches have always opted for restricted 1-factor (general distress), 2-dimensional (eg, positive and negative affect), or 3-dimensional structures (eg, tripartite model).30–35 It is further noteworthy that scales that have claimed to have more factors (SCL-90) have been shown to load strongly on one factor with little
discriminate value among the subscales. Thus, it seems fair to state that the level of differentiation that is reliably achievable by self-report measures is limited; its value lies in the dimensional description of facets of self-reported behavior, cognitive-emotional and circumscribed symptom elements, and in the dimensional characterization of perceived distress of a person.

Yet is this sufficient basis for revising diagnostic classification or even suggesting a new metastructure? Is correlational evidence that the assumed dimensions are more consistent with selected evidence from neurobiological, genetic, and treatment research than single categorical diagnosis a sufficient justification? Taking into account the serious limitations of the factor analytic approach, and if examined against the criteria for establishing diagnostic entities by Robins and Guze or its extensions (Hyman and colleagues, cited in Ref. 11), these questions must be denied. So what might be the additional benefit?

Could it help to decide on the deletion of groups of disorders or specific diagnoses from our research and clinical practice agenda? Would it help to improve current diagnostic practice by allowing the cluster labels to be used as a sufficient global diagnostic substitute, thus easing the burden of careful differential diagnostic workups? Will it reduce spurious comorbidity, by assuming that „relevant“ comorbidity only exists between the 5 clusters? All these claims have been recently stated by promoters of the new metastructure with more or less pronounced limitations. Given the very limited exploratory evidence for the clusters, its adoption seems premature. It is therefore disconcerting that Andrews and colleagues and others suggest the proposed metastructure as a basis for research on treatment, prognosis, and even practice. Yet there is neither explanation nor proof as to why one would expect the proposition to have greater utility and validity compared with the current DSM grouping.

**Harmful Effects?**

Leaving aside the substantial gaps regarding supporting data, might the lumping of anxiety, somatoform, and depressive disorders under the broad umbrella of „emotional cluster“ revive the stigmatizing and otherwise controversial concept of neurosis? The initial beauty and simplicity of grouping mental disorders into conditions that could be mainly characterized by constructs like „emotional dysregulation“ as in the emotional cluster, and „disinhibition“ as in the externalizing cluster, does not protect against the revival of potentially stigmatizing diagnostic group labels. Even worse, such an organization carries an increased risk of losing sight of research-based distinctions and considerably different intervention needs of the specific anxiety, depressive, somatoform, and other „emotional cluster“ disorders. It took 2 decades to communicate – albeit sometimes too subtly – that diagnosing specific mental disorder matters. For example, it is certainly true that the cross-sectional picture of a person with PD with past uncued panic attacks might not be that different from agoraphobia, severe specific phobia, GAD, or even depression with anxiety features when considering presenting symptoms, distress, and associated neurobiological and behavioral correlates. However doesn’t the occurrence of initial spontaneous panic attacks tell us something extremely useful with regard to prognosis as well as treatment? Isn’t there persuasive evidence that PD without agoraphobia is a quite different „animal“ from panic disorder with agoraphobia? How damaging will a broad cluster proposition be in this respect if a „deeper“ meaning is involved? One might also ask what progress might be impeded if these factors, or dimension, or proposed metastructures are incomplete or incorrect. Intended or not, the diagnostic system will always influence the field’s thinking about psychopathology and treatment. Is it therefore wise to adopt such a structural system with a „deeper meaning“ that has not been adequately tested? (Fig. 3)
To conclude, we currently see no solid evidence to adopt a new structure of mental disorders along the lines of factor analytic explorations and other statistical methods. In particular, we have deep concerns about adopting the 2-factor model of „internalizing“ and „externalizing“ or the 3-factor model of „anxious-misery,“ „fear,“ and „externalizing,“ or similar extensions of this sort. We do not see evidence of immediate or long-term advantages but have significant concerns about potentially critically harmful effects for basic and clinical research as well as treatment practice and public health issues. However, we share the strong feeling that a wider inclusion of dimensional measures for diagnosis and syndrome-specific core features of mental disorders may be fruitful. These measures could be conceptualized for single diagnoses as well as for cross-cutting diagnostics. Such alternative propositions to revising DSM-V have been recently elaborated\textsuperscript{28,40} (see Fig. 3 for an example), and seem to address the current needs in research and practice much more appropriately and straightforwardly.

\textsuperscript{28,40}

\textbf{Fig. 1.} Krueger’s 3-factor model based on 10 lifetime disorders (NCS-Data). (From Krueger RF. The structure of common mental disorders. Arch General Psychiatry 1999;56:921-6; with permission.)
Fig. 2. Result of the exploratory factor analysis (EFA) including anxiety, mood and other disorders (EDSP, total sample, 6 factors). GAD, generalized anxiety disorder; HME, hypomanic episodes; MDE, major depressive episode; MNE, manic episodes; OCD, obsessive compulsive disorder; PTSD, posttraumatic stress disorder; RMS, root mean square; SSR/RE, somatiform disorders.

By Anxiety Diagnosis

OSM-IV anxiety disorders | Specific and shared evaluation domains* according to diagnostic criteria
---|---
Panic disorder, w/o Agoraphobia | Anxiety neurotic features | Anticipatory anxiety | Avoidance behavior | Impairment | Distress | Duration
Panic disorder with Agoraphobia | | | | | |
Agoraphobia w/o history of panic | | | | | |
GAD | Specific phobias | Social phobias | OCD | PTSD | Acute stress disorder | Separation anxiety disorder

* Each domain can be rated continuously by severity, frequency, etc. Other possible domains may include biological factors (neurobiological, developmental, etc.).

Fig. 3. Diagnosis specific and cross-cutting assessment domains of anxiety disorders. (From Shear MK, Bjelland I, Beesdo K, Gloster AT, Wittchen H-U. Supplementary dimensional assessment in anxiety disorders. Int J Methods Psychol Res 2007;16(suppl 1):554; with permission.)
29. Kraemer HC. What is the ‘right’ statistical measure of twin concordance (or diagnostic reliability and validity)? Arch Gen Psychiatry 1997;54(12):1121–4.
40. Special Issue: Dimensional approaches to psychiatric classification: refining the research agenda for DSM-V. Int J Methods Psychiatr Res 2007;16(Suppl 1).