Using Hybrid Modeling to Determine the Latent Structure of Psychopathology

Diana J. Whalen

Department of Psychiatry, Washington University School of Medicine in St. Louis, St. Louis, Missouri.

Van Dam *et al.* (1) took an innovative, data-driven (as opposed to top-down, diagnostically driven) approach to elucidate psychiatric phenotypes and related differences in functional brain connectivity in a large sample of adults. This work attempted to clarify one of the major problems in both clinical practice and research—namely, high comorbidity of DSM diagnoses. The authors' intention was to move away from DSM-5 labels, and their results using a hierarchical clustering approach showed distinctions between individuals with internalizing and externalizing symptoms, similar to the approach advocated by the Hierarchical Taxonomy of Psychopathology consortium (R. Kotov, M.D., *et al.*, The Hierarchical Taxonomy of Psychopathology [HiTOP]: A dimensional alternative to traditional nosologies [unpublished data], 2016). The hierarchical clustering approach used by Van Dam *et al.* (1) illuminated groups and subgroups (referred to as "clusters") comprising typically and atypically functioning individuals that cut across DSM-5 disorders, as well as several functional connectivity differences between the two largest groups. It is important to recognize that the authors could have chosen other data-driven statistical approaches. The hierarchical clustering approach that was used relied on the assumption of an underlying "hierarchy," and also assumed that individuals fit into specific clusters based on their phenotype profile. Essentially, the approach still categorized individuals using pheno-type profile instead of DSM diagnosis. In using such an approach, the authors lost the ability to determine whether a categorical, dimensional, or true hybrid structure best fit the data. We describe several benefits of alternative data-driven modeling strategies below.

Data-driven approaches structure the statistical model that is being tested to be the best fit for the data rather than fitting the data into an existing model with specific assumptions about the underlying structure. Several types of data-driven approaches exist, including the clustering methods used by Van Dam *et al.* and combinations of latent variable models, such as factor mixture models. Approaches that combine facets of different types of latent variable models are often referred to as hybrid models (2). Hybrid approaches allow for the search of both traits/continua (e.g., symptoms of depression) and types/kinds (e.g., depressed vs. nondepressed individuals) in a single model (3,4). Continuous and categorical latent structures can be directly compared and model parameters can be relaxed/constrained to determine the degree to which those data represent continuous, categorical, or hybrid

Address correspondence to Diana J. Whalen, Ph.D., Washington University School of Medicine, Department of Psychiatry, 4444 Forest Park, Suite 2100, St. Louis, MO 63108; whalend@psychiatry.wustl.edu.

phenomena. Thus, hybrid models (e.g., factor mixture models or latent variable mixture models) combine the continuous aspects of latent trait models with the discrete aspects of latent class models. In addition, factor mixture models allow for a direct comparison of model fit, and thus it is possible to compare a hybrid model to the more categorical or dimensional modeling approaches to determine which provides the best fit and to document what additional information is gained/lost from each type of model (2–5). This approach also removes the relatively subjective decision making from the model selection, as statistical indices of model fit are available (e.g., Bayesian information criterion) and can be directly compared. In addition, factor mixture models can account for age and gender within the model. The authors were only able to control for these important covariates in their exploratory factor analysis.

One primary benefit to hybrid approaches, such as factor mixture models—in contrast to the hierarchical clustering approach used by Van Dam *et al.*—is that factor mixture models overcome the assumption of conditional independence within each class or cluster [see Miettunen *et al.* (2) and Masyn *et al.* (6) for graphical descriptions of different types of models]. In models that assume conditional independence, each individual is assigned to the group with highest probability of membership, and it is assumed that there are no differences between group members. On the other hand, using factor mixture models, within-group variation is anticipated and accounted for, such that there are distributions of severity within each group. To make this less abstract, imagine two groups of individuals: healthy and not healthy. Within the healthy group, there are individuals who are more healthy and less healthy, representing a continuous distribution. A distribution is also likely to be found in the not healthy group. This distinction may be important for the data presented in the current paper because the groups that were determined based on the clustering method conducted by the authors may actually have unique phenotype profiles that differ based on cluster membership. For instance, some clusters may not have a six-factor solution of symptom-based traits; some clusters may have fewer or more factors that provide the best fit within this particular group. Specifically, had the authors conducted a factor mixture model, it would have been possible to determine whether the factor structure between the "functional" and "nonfunctional" clusters significantly differed. Perhaps there are, for example, four unique factors in the functional cluster or six unique factors in the nonfunctional cluster.

As stated above, another major advantage of using a hybrid approach is determining whether the hybrid model is even necessary, depending on model fit. For example, although these hybrid models are intuitively appealing, the symptoms and traits assessed here may be more consistent with a purely dimensional as opposed to categorical or hybrid conceptualization. In fact, in the current study, the authors were left with two large clusters (functional vs. less functional) to conduct analyses with functional connectivity data. This may indicate that there are few meaningful differences between the smaller clusters and that such a complicated approach is not warranted or needed. The unique data that are available in this project would allow for this kind of determination, should the modeling strategy be altered to a hybrid approach. This may be particularly useful for the current sample, given that many of the participants were relatively healthy and free of psychiatric disorder. Should a hybrid model be found to have the best fit, unique distributions among healthy individuals would be able to be highlighted and compared.

In closing, the report by Van Dam *et al.* offers a significant step toward more data-driven approaches in determining the latent structure of psychopathology—an important challenge that has troubled researchers for decades. While there are advantages to the hierarchical clustering approach used by the authors, there are several other types of modeling available that can determine the structure of psychopathology in a truly hybrid fashion, combining the assets of discrete- and continuous-type models. We encourage the field to use such approaches to determine the underlying structure of the variables of interest. In addition, as suggested by Eaton *et al.* (4), all of our structural assumptions about variables are inherently linked to the way in which they were assessed. Therefore, it is difficult to argue for and model continuous structures when a discrete indictor was used in assessment and vice versa. An ongoing challenge in our field continues to be the accurate and meaningful assessment of psychopathology, and hybrid models further augment the need for such assessments.

Acknowledgments and Disclosures

References

1. Van Dam NT, O'Connor D, Marcelle ET, Ho EJ, Craddock RC, Tobe RH, et al. (2017): Data-driven phenotypic categorization for neurobiological analyses: Beyond DSM-5 labels. Biol Psychiatry 81:484–494. [PubMed: 27667698]

2. Miettunen J, Nordström T, Kaakinen M, Ahmed AO (2016): Latent variable mixture modeling in psychiatric research—A review and application. Psychol Med 46:457–467. [PubMed: 26526221]

3. Borsboom D, Rhemtulla M, Cramer AOJ, van der Maas HLJ, Scheffer M, Dolan CV (2016): Kinds versus continua: A review of psychometric approaches to uncover the structure of psychiatric constructs. Psychol Med 46:1567–1579. [PubMed: 26997244]

4. Eaton NR, Krueger RF, Docherty AR, Sponheim SR (2014): Toward a model-based approach to the clinical assessment of personality psychopathology. J Pers Assess 96:283–292. [PubMed: 24007309]

5. Eaton NR (2015): Latent variable and network models of comorbidity: Toward an empirically derived nosology. Soc Psychiatry Psychiatr Epidemiol 50:845–849. [PubMed: 25599937]

6. Masyn KE, Henderson CE, Greenbaum PE (2010): Exploring the latent structures of psychological constructs in social development using the dimensional-categorical spectrum. Soc Dev 19:470–493. [PubMed: 24489441]