

predict prognosis or how much treatment a patient might need.

A general factor has also been used in the self-reported personality pathology domain. Tellegen et al¹⁰ developed a Minnesota Multiphasic Personality Inventory scale to measure non-specific variance associated with unpleasant mood states, which they labeled demoralization. They observed that “this general factor appears to inflate correlations between attributes that are considered relatively independent”¹⁰. Therefore, they aimed to “remove from each clinical scale items primarily marking demoralization” in order to improve the ability to differentiate among patients¹⁰.

We share their sentiment about the importance of discriminant validity, and spec-

ulate that one potential clinical advantage of measuring a general factor is that the remaining scales might better highlight differences between patients. This, in turn, might help guide treatment modality. For example, without isolating a general factor, individuals with a broad symptom load often display elevated scores on a wide range of psychiatric scales. However, after isolating a general factor, such individuals might only display elevated scores on a smaller subset of scales, which might represent a suitable target for initial treatment.

Erik Pettersson¹, Henrik Larsson², Paul Lichtenstein¹

¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; ²Örebro University, Örebro, Sweden

1. Lahey BB, Moore TM, Kaczkurkin AN et al. *World Psychiatry* 2021;20:57-63.
2. Pettersson E, Larsson H, D’Onofrio BM et al. *World Psychiatry* 2020;19:206-13.
3. Caspi A, Houts RM, Belsky DW et al. *Clin Psychol Sci* 2014;2:119-37.
4. Brandes CM, Herzoff K, Smack AJ et al. *Clin Psychol Sci* 2019;7:1266-84.
5. Lahey BB, Van Hulle CA, Singh AL et al. *Arch Gen Psychiatry* 2011;68:181-9.
6. Ross J. *Psychol Rev* 1963;70:432-43.
7. Kozhuharova P, Dickson H, Tully J et al. *PLoS One* 2019;14:e0224455.
8. Latvala A, Kuja-Halkola R, Ruck C et al. *JAMA Psychiatry* 2016;73:1268-75.
9. Smith GT, Atkinson EA, Davis HA et al. *Annu Rev Clin Psychol* 2020;16:75-98.
10. Tellegen A, Ben-Porath YS, McNulty JL et al. The MMPI-2 restructured clinical (RC) scales: development, validation, and interpretation. Minneapolis: University of Minnesota Press, 2003.

DOI:10.1002/wps.20813

The p factor is the sum of its parts, for now

The general factor of psychopathology, or p factor, has received increasing attention over the last half decade, but questions remain about how to best conceptualize it^{1,2}.

Here, we use data from a large-scale survey (National Epidemiologic Survey on Alcohol and Related Conditions, NESARC), conducted in a nationally representative US sample, to demonstrate that, statistically, p is nearly identical to the sum of diagnoses it is estimated upon. The same holds for specific factors such as internalizing and externalizing, and our results are robust to various estimation methods. We discuss implications of this finding for the nature of these factors, and raise the question whether the sole reliance on reflective latent variable models used in the p factor literature is justified, given that nearly identical scores can be obtained by a much simpler statistical procedure that has fewer parameters and imposes fewer assumptions on the data.

Questions about the interpretation of p are important, because this factor carries at least two meanings. Statistically, p refers to a latent variable estimated on a covariance matrix of psychopathology symptoms or diagnoses in a given dataset. The field has utilized one specific class of models, the reflective latent variable model, in several flavors, such as the bifactor and

second-order factor models, that decompose variance somewhat differently³. No matter the specific model, general factors such as p necessarily emerge when data feature a positive manifold⁴. Statistically speaking, p is just a different way of stating that observed items are positively related.

The p factor’s second meaning is conceptual: what p represents. Conceptualizations of p vary widely, including severity/dysfunction and a general liability for psychopathology through non-specific genetic and environmental influences, disordered thought processes, and/or trait-like attributes (e.g., negative emotionality)^{1,2}. It is an open question how these conceptualizations of p fit the data and methods used in the field. If p represents liability, for example, it is unclear why models are estimated on data on symptoms and diagnoses rather than data on risk factors and etiology, such as early adversity, mentalization/reflective functioning, and attachment insecurity^{5,6}.

To shed light on the relation between the statistical and the conceptual p factor, we estimated two types of general factor models: the bifactor model (M1) and the higher-order factor model (M2). We repeated analyses for these general factor models’ specific factors (distress, fear and externalizing), as well as the correlated 3-factor (M3: distress, fear, externalizing)

and correlated 2-factor (M4: internalizing, externalizing) models. The rationale for estimating numerous models was to investigate the degree to which latent variables are generally more than the sum of their indicators and to rule out that results are due to one particular parameterization.

We utilized two waves of the NESARC dataset (W1: N=43,093, W2 follow-up: N=34,653; see <https://osf.io/yrrpw8> for details), which has commonly been used in the p factor literature⁷. Our main findings are as follows. First, in both waves, we identified high correlations between sum scores of all diagnoses and p, approaching unity for both M1 (range: 0.87-0.99) and M2 (range: 0.87-1.00). Second, domain-specific factors and their respective sum scores (e.g., externalizing factor with sum of externalizing diagnoses) were also highly related across all models: 0.82-0.94 for M1, 0.87-0.96 for M2, 0.78-1.00 for M3, and 0.82-0.96 for M4. Third, correlations between W1 and W2 latent variables were strikingly similar to those between W1 and W2 sum scores (e.g., M1 p factor vs. total sum score: 0.40 vs. 0.44). These findings hold regardless of whether the relations between latent factors and sum scores were estimated using factor scores or a single structural equation model.

In sum, we show that the p factor as well as domain-specific factors are identical or

nearly identical with the sum of diagnoses that go into these respective factors, and that results hold in both general factor and correlated factor models. We see three main implications of our findings.

First, we compare two types of models: a simple sum of indicators vs. a class of highly sophisticated structural equation models that estimate a large number of parameters and impose considerable assumptions on the data, such as hierarchies in which factors are organized, or relations among factors that are constrained to zero³. Both models produce nearly identical scores for participants. If replicated in other studies, this finding suggests that the use of reflective latent variable models should be considered more carefully: what are the specific benefits of this modeling framework for the p factor literature, and do they outweigh the potential costs, such as over-parameterization and stringent assumptions imposed on the data^{3,4}? Such deliberations will benefit from explicit goals to determine whether specific statistical models are adequate in the context of a given research question. In general, scientific progress is often hampered by overreliance on any particular type of model⁸, and thinking more clearly about conceptualizations of p may offer opportunities to diversify methodology.

Second, we provide evidence that p is simply a re-expression of the sum of diagnoses that individuals experience. This is not

surprising: about 70 years ago, Cattell⁹ described scores on the general factor as “essentially the sum of the scores”, and Lahey et al¹ acknowledge the p factor is a “weighted average” of items. Our results imply that p represents severity or comorbidity, not liability, much in the same way as the sum of flu symptoms provides a rough index for severity, not liability. Whether competing accounts of p offer better explanations, such as the idea that it represents liability, requires that models be estimated on variables that actually denote liability, rather than variables denoting severity and comorbidity.

Third, if p is a mere index of the data, this suggests that the meaning of p will only be invariant across studies inasmuch as the data that go into our models are invariant across studies.

Overall, data can be brought to bear on theories when statistical models impose assumptions on the data that are in line with the theories. The p factor literature has been largely atheoretical and primarily concerned with *description* of data – a crucial first step to establish phenomena that can then be explained. But let us not lose sight of the fact that p is an effect that needs to be explained (i.e., *explanandum*), not something that does the explaining (i.e., *explanans*). It necessarily emerges from a positive manifold, and tells us nothing about the mechanisms that generated the data^{4,6}. Further, if the goal is the description of data, it is unclear why

the reflective latent variable model that is solely relied upon in the literature should be the only model suited for this goal.

Thinking more clearly about theories of p^{1,2}, and spelling out these theories precisely, will help adjudicate between different conceptual accounts of p. Criticizing and modifying theories requires that we know exactly where they start and end. Clearer theories will then facilitate choosing appropriate statistical models that can in turn guide theory reform.

Eiko I. Fried¹, Ashley L. Greene²,
Nicholas R. Eaton²

¹Unit of Clinical Psychology, Faculty of Social and Behavioral Sciences, Leiden University, Leiden, The Netherlands;

²Department of Psychology, Stony Brook University, Stony Brook, NY, USA

E.I. Fried and A.L. Greene contributed equally to this work.

1. Lahey BB, Moore TM, Kaczkurkin AN et al. *World Psychiatry* 2021;20:57-63.
2. Caspi A, Moffitt TE. *Am J Psychiatry* 2018;175:831-44.
3. Greene AL, Eaton NR, Li K et al. *J Abnorm Psychol* 2019;128:740-64.
4. van Bork R, Epskamp S, Rhemtulla M et al. *Theory Psychol* 2017;27:759-73.
5. Bender DS. *J Pers Assess* 2019;101:356-9.
6. Fried EI. <https://doi.org/10.31234/osf.io/zg84s>.
7. Greene AL, Eaton NR. *Compr Psychiatry* 2017;72:74-82.
8. Vaidyanathan U, Vrieze SI, Iacono WG. *Psychol Inq* 2015;26:209-30.
9. Cattell RB. *Factor analysis: an introduction and manual for the psychologist and social scientist*. New York: Harper, 1952.

DOI:10.1002/wps.20814

Why hierarchical dimensional approaches to classification will fail to transform diagnosis in psychiatry

At the outset, I would like to stipulate that the current DSM and ICD approaches towards diagnostic classification are not perfect. Others have elaborated on the limitations of these categorical approaches towards diagnosis¹; so I do not repeat them here. I also stipulate that there are some advantages to a dimensional conceptualization of psychopathology over a categorical one. Nonetheless, I am fairly confident that an empirically derived dimensional classification will not replace the DSM-5/ICD-11 anytime soon, if ever.

Eight potential barriers to the integra-

tion into clinical practice of one such model, the Hierarchical Taxonomy of Psychopathology (HiTOP), have been identified². Among them, are the length of clinical evaluations, billing for clinical encounters, and incorporating the model into training. The implicit message is that clinicians will require some convincing. That is, clinicians are likely to resist such a seismic change unless a compelling case is made to support the adoption of a new approach towards assessment and diagnosis. While the supporters of dimensional approaches have identified some obstacles to be over-

come to transform the categorical system to a dimensional one², there are some further important obstacles that they have not addressed, which make such a transformation highly unlikely.

Recognizing that such a change will be a challenge, Lahey et al³ note that it will be essential to demonstrate that a hierarchical dimensional diagnostic approach improves patient outcomes. If patient outcomes are not demonstrably better, it will be difficult to convince the clinical community that it is worth the effort to learn a new diagnostic language.