

- MacArthur studies of successful aging. *Arch Intern Med* 1997; 157:2259–68.
7. McEwen BS, Stellar E. Stress and the individual. Mechanisms leading to disease. *Arch Intern Med* 1993;153:2093–101.
 8. Meimaridou E, Hughes CR, Kowalczyk J, Guasti L, Chapple JP, King PJ, Chan LF, Clark AJ, Metherell LA. Familial glucocorticoid deficiency: new genes and mechanisms. *Mol Cell Endocrinol* 2013;371:195–200.
 9. Nicholls DG, Fergusson SJ. *Bioenergetics*. 4 ed. Cambridge, MA: Academic Press; 2013.
 10. Picard M, Turnbull DM. Linking the metabolic state and mitochondrial DNA in chronic disease, health, and aging. *Diabetes* 2013;62:672–8.
 11. Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener AL, de Cabo R, Cohen P. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab* 2015;21:443–54.
 12. Zhang Q, Raoof M, Chen Y, Sumi Y, Sursal T, Junger W, Brohi K, Itagaki K, Hauser CJ. Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature* 2010;464:104–7.
 13. Lu B, Kwan K, Levine YA, Olofsson PS, Yang H, Li J, Joshi S, Wang H, Andersson U, Chavan SS, Tracey KJ. Alpha 7 nicotinic acetylcholine receptor signaling inhibits inflammasome activation by preventing mitochondrial DNA release. *Mol Med* 2014.
 14. Picard M, Juster RP, McEwen BS. Mitochondrial allostatic load puts the 'gluc' back in glucocorticoids. *Nat Rev Endocrinol* 2014;10:303–10.
 15. McEwen BS. Protective and damaging effects of stress mediators: central role of the brain. *Dialogues Clin Neurosci* 2006;8:367–81.
 16. Juster RP, Moskowitz DS, Lavoie J, D'Antono B. Sex-specific interaction effects of age, occupational status, and workplace stress on psychiatric symptoms and allostatic load among healthy Montreal workers. *Stress* 2013;16:616–29.
 17. Picard M. Pathways to aging: the mitochondrion at the intersection of biological and psychosocial sciences. *J Aging Res* 2011;2011:814096.
 18. Picard M, McManus MJ, Gray JD, Nasca C, Moffat C, Kopinski PK, Seifert EL, McEwen BS, Wallace DC. Mitochondrial functions modulate neuroendocrine, metabolic, inflammatory, and transcriptional responses to acute psychological stress. *Proc Natl Acad Sci U S A* 2015;112:E6614–23.

OPEN

Considering the Appropriateness of the Factor Analytic Operationalization of Allostatic Load

In a recent issue of *Psychosomatic Medicine*, Wiley et al. (1) made a valuable contribution to the discussion of the optimal measurement of allostatic load (AL). In the most comprehensive factor analytic investigation of AL to date, they found that a bifactor model with a general AL factor and seven physiological system factors fits better than a higher-order model in which the seven system factors loaded on the general AL factor. Similar models have been

applied by the author (T.B.) and others to operationalize AL (2–4). Here, we consider the primary theoretical assumptions underlying latent variable modeling, argue that the construct of AL is inconsistent with these assumptions, and propose alternate operationalizations of AL.

UNDERLYING CONSTRUCT (COMMON CAUSE)

A latent variable model is estimated based on the patterns of covariance in a set of variables. By including an AL general factor in a latent variable model, researchers are positing that an underlying construct is the common cause of the observed covariation in all of the modeled biological measures. Although the theoretical relation of the common cause or construct to the original variables differs in bifactor versus higher-order models, in either case, we must ask: What could this common factor be? Wiley et al. stated that the AL factor “[captures] the notion that there is an underlying process influencing multiple physiological systems” ((1): p. 4). However, the observation of a general factor estimated from inter-individual summary statistics (i.e., covariances) says little about what this process may actually be.

INDEPENDENCE CONDITIONAL ON THE LATENT TRAIT

A primary assumption of latent variable models is that once the effect of the latent factors has been accounted for, the measured variables—in this case, the biological measures—are independent. This is unlikely to be the case with AL measures. Levels of different biomarkers are linked causally to each other, rather than only through the common cause latent variable(s). For example, body mass index (BMI) has previously been used as a metabolic system AL biological measure (e.g., (2,5)). However, Mendelian randomization studies have found that increased BMI has a causal effect on levels of other metabolic biological measures as well as levels of AL biomarkers used to represent other physiological systems, such as blood pressure and inflammation (e.g., (6)). Thus, it is most likely that the biomarkers are not conditionally independent but are instead dynamically related in complex networks. Such networks can produce observed correlations between variables that have no common cause (7).

INTERCHANGEABILITY OF INDICATORS

A further assumption of the latent variable model is that the definition of the latent variable does not change when different sets of indicators are used (8). This holds because the indicators are affected by, but do not affect, the latent variable. Another key finding of Wiley et al. was that fitting models in which the biological measures from each of the 7 physiological systems were excluded caused no large changes in AL factor loadings (1). This method provides only a weak test of interchangeability. The stability of general intelligence factor loadings has long been a research

focus for intelligence researchers, so AL researchers may benefit from applying their approaches to this issue (e.g., (9,10)). For example, researchers could compute and correlate AL scores from different nonoverlapping multisystem sets of biological measures (8). The existence of diverse causal links between AL biological measures from different physiological systems suggests to us that the nature of what relates the biomarkers may change depending on which measures are included in the model. We predict that more thorough, more powerful tests of the stability of AL factor loadings will find that it does not hold.

FORMATIVE VERSUS REFLECTIVE INDICATORS

In the common factor model, the biological measures are reflective indicators, that is, they are manifested by a common cause latent variable. However, to the extent that the model assumptions are violated (previously discussed), the factor model is not appropriate. Thus, it may instead be profitable to consider the biological measures as formative indicators, that is, as variables that define the construct (8). This way of thinking about how the biological measures relate to AL is consistent with any number of weighted or sum scores. It is also consistent with AL theory, in that more severe, more widespread physiological dysregulation will relate to higher AL scores.

Alternatively, the associations between AL biological measures could be modeled using each measure individually, without the need for any single latent or observed summary. This could be done with network analysis, which has been used beneficially by researchers studying symptom networks in mental disorders (11). Allostatic load indicators can also be modeled separately without consideration of their associations. Consistent with this approach, *Psychosomatic Medicine* typically provides data of separate biological measures when articles report about complex phenomena such as AL and metabolic syndrome.

Aside from any issues regarding model assumptions, two further points warrant comment about the models presented by Wiley et al.

IMPROVED MODEL FIT FOR BIFACTOR APPROACH

The complex causal links between biological measures from different physiological systems also help to explain why the bifactor AL model fits better than the hierarchical AL model. The hierarchical model imposes “proportionality constraints” ((12): p. 115): the ratio of the AL general factor loadings to the system factor loadings is constrained to equality within the biological measures of each physiological system. Considering the diverse causal links between different AL biomarkers, both within and across systems, these proportionality constraints are likely to be violated. Furthermore, it has been shown that when the true

model contains “unmodelled complexity” ((13): p. 407) in the form of small correlated residuals and cross-loadings, or even modeled complexity in the form of correlated residuals across factors, fit indices and criteria may be biased in favor of the bifactor model. Consequently, the better fit of the bifactor model follows from AL theory and research, as well as from methodological findings, for reasons other than those Wiley et al. (1) focused on.

VARIANCE EXPLAINED BY PHYSIOLOGICAL DYSREGULATION FACTORS

Statistically, a desirable property of a general factor is that it accounts for most variance in the constituent indicator variables. In the study by Wiley et al., the AL factor explained only approximately 11% of the variance in the AL biological measures. Some of the physiological system-specific factors were also weak. For example, the hypothalamic-pituitary-adrenal axis and inflammation factors explained only approximately 9% and 16% of the variance in their respective biological measures. Note that weak factor saturation of physiological dysregulation factors has also been an issue in other samples (2,3).

PROPERTIES OF OPTIMAL SCORES FOR AL

Ideally, AL scores should be: 1) calculated using biological measures from various physiological systems; 2) consistently calculated across samples; and (3) closely related to criterion variables. Those who desire scores that are rooted in AL theory would prefer the AL scoring method that produces the scores most closely related to chronic/repeated perceived stress. For a pragmatist, the focus may not be on investigating how different physiological dysregulation scores relate to prior perceived stress but rather on finding the scores that most strongly predict important health outcomes such as cardiovascular disease and death. It may also be advantageous to have scores that explicitly represent the accumulation of the effects of repeated environmental challenges.

Our theoretical and methodological concerns with the factor analytic operationalization of AL suggest to us that factor scores will not prove to be the optimal AL scoring method. We therefore believe that further research is required to determine the optimal operationalization(s) of AL.

ZANDER CROOK, MSc

Department of Psychology
The University of Edinburgh, Edinburgh
United Kingdom

TOM BOOTH, PhD

Department of Psychology
The University of Edinburgh, Edinburgh
United Kingdom

Source of Funding and Conflicts of Interest: ZC is supported by a UK Economic and Social Research Council Advanced Quantitative Methods PhD award. Neither author reports a conflict of interest.

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

REFERENCES

1. Wiley JF, Gruenewald TL, Karlamangla AS, Seeman TE. Modeling multisystem physiological dysregulation. *Psychosom Med* 2016;78:290–301.
2. Booth T, Starr JM, Deary I. Modeling multisystem biological risk in later life: allostatic load in the Lothian Birth Cohort study 1936. *Am J Hum Biol* 2013;25: 538–43.
3. Seeman T, Gruenewald T, Karlamangla A, Sidney S, Liu K, McEwen B, Schwartz J. Modeling multisystem biological risk in young adults: the Coronary Artery Risk Development in Young Adults study. *Am J Hum Biol* 2010;22: 463–72.
4. Howard JT, Sparks PJ. Does allostatic load calculation method matter? Evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. *Am J Hum Biol* 2016;28:627–35. <http://dx.doi.org/10.1002/ajhb.22843>.
5. Crimmins EM, Johnston M, Hayward M, Seeman T. Age differences in allostatic load: an index of physiological dysregulation. *Exp Gerontol* 2003;38:731–4.
6. Millard LA, Davies NM, Timpson NJ, Tilling K, Flach PA, Smith G. MR-PheWAS: hypothesis prioritization among potential causal effects of body mass index on many outcomes, using Mendelian randomization. *Sci Rep* 2015;5: 16645. <http://dx.doi.org/10.1038/srep16645>.
7. van der Maas HLJ, Dolan CV, Grasman RPPP, Wicherts JM, Huizenga HM, Raijmakers MEJ. A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol Rev* 2006;113:842–61.
8. Bollen KA, Bauldry S. Three Cs in measurement models: causal indicators, composite indicators, and covariates. *Psychol Methods* 2011;16:265–84.
9. Johnson W, Bouchard TJ Jr., Krueger RF, McGue M, Gottesman II. Just one g: consistent results from three test batteries. *Intelligence* 2004;32:95–107.
10. Thorndike RL. Stability of factor loadings. *Pers Individ Dif* 1987;8:585–6.
11. Borsboom D, Cramer AOJ. Network analysis: an integrative approach to the structure of psychopathology. *Annu Rev Clin Psychol* 2013;9:91–121.
12. Yung Y-F, Thissen D, McLeod LD. On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika* 1999;64:113–28.
13. Murray AL, Johnson W. The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence* 2013;41: 407–22.

The Authors Reply: Pursuing the Optimal Operationalization of Allostatic Load

In their commentary on our paper (1), Crook and Booth (2) raise important questions regarding the theoretical and methodological appropriateness of applying factor analysis to model allostatic load (AL). They argue that factor scores are not the “optimal” measure of AL and conclude that more research is needed.

From a methodological perspective, Crook and Booth argue that the poor fit of the hierarchical model may be due to the proportionality constraint inherent in its structure. Whether the proportionality constraint or other features of the model are the sources, it does not change the result that the hierarchical model provided an inferior fit to the data than did the bifactor model. Considering the bifactor model, Crook and Booth suggested a stronger test of exchangeability by computing and correlating AL scores from nonoverlapping biomarkers (2). However, unlike intelligence research, studies on biomarkers tend to have relatively few indicators as a consequence of feasibility factors (e.g., available blood sample volume, participant burden, and cost). Separating already limited biomarker panels into two nonoverlapping sets will be too few to estimate the bifactor model in many studies. As advances in multiplexing reduce barriers to assaying numerous biomarkers (3), we look forward to studies that address this question.

Crook and Booth also note that, on average, the general factor explains approximately 11% of the variance in the biomarkers (2). In psychometric studies of carefully designed scales, 11% may be considered a small amount of variance, but this is expected for biomarkers for several reasons. First, except for the heart rate variability measures, each biomarker is distinct (e.g., interleukin 6 and tumor necrosis factor α are separate analytes with unique roles in the immune system and inflammatory processes). Second, biomarkers have circadian rhythms, which introduce variability due to the timing of assessment. Third, the timeframe reflected in each biomarker varies. For example, glycosylated hemoglobin approximately indexes the previous three months, overnight urinary epinephrine and norepinephrine index approximately 12 hours, and blood pressure is comparatively momentary. Fourth, the general factor accounted for variance over and above the covariates age and sex. Therefore, we expected the general factor to account for a modest amount of the variance in biomarkers. Furthermore, considered over and above the effects of age, 11% of the variance is not necessarily trivial. Finally, the overall model (i.e., general + system factors and covariates) accounted for an average of 55% of the variance in biomarkers, rising to 60% when excluding soluble intracellular adhesion molecule 1 and low-density lipoprotein. If researchers believe