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.

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

including additional variability from individual biomarkers yields a more optimal measure, principal component analysis can be used (e.g., (4,5)). Using principal component analysis, researchers can choose the tradeoff between dimension reduction and the desired amount of variance explained by extracting more components.

From a theoretical perspective, Crook and Booth (2) argue the complex interrelations among biomarkers to be inconsistent with the model indicated by a bifactor model. We agree that the relations among biological measures are complex. Indeed, there is ample evidence that these biomarkers and systems relate to each other and dynamically to themselves via positive and negative feedback loops (e.g., (6)). Crook and Booth (2) argue that not accounting for these interconnected pathways biases results to favor the bifactor model. It is certainly implausible that one latent factor causes all of the biomarkers measured in our study. However, the practical implications of whether the general factor from the bifactor model captures variability in biomarkers due to a shared latent factor or instead captures variability due to numerous and bidirectional pathways among biomarkers are less clear. In either case, the general factor indexes a phenotype that emerges across multiple systems.

Nevertheless, from a pragmatic perspective, what alternatives are available? One possibility is graphical models, such as complex network analysis (7). If there are no latent or selection variables, then inference regarding possible causal relations can be extracted from observational data using do-calculus (8). As with many statistical models, the challenge lies in meeting the assumptions: no latent or selection variables. These assumptions are violated unless all variables related to assignment are known, measured, and included, which is typically only plausible with random assignment.

Rather than speculate that graphical models or network analysis may be more appropriate than a factor model, we tested this approach in the MIDUS data. Specifically, we extracted residual scores for each biomarker adjusting for age and sex and calculated the correlation matrix for all biomarkers using full information maximum likelihood estimation. These correlations were then subjected to the PC algorithm (7) to explore *potential* underlying causal structures that *could* give rise to the data. The results are shown using a graphical model (9) in Figure S1 (Supplemental Digital Content 1, http://links.lww.com/PSYMED/A336). The potential underlying structure identified had biomarkers from related systems clustered together, although this is not surprising. Interestingly, the potential structure included cortisol, dehydroepiandrosterone sulfate, epinephrine, and norepinephrine as drivers of other biomarkers, but not vice versa. In comparison, although insulin resistance was upstream of triglycerides, waist-to-hip ratio, and E-selectin, it was also downstream from cortisol, epinephrine, and C-reactive protein. Thus, the potential structure identified is strikingly consistent with AL theory on primary versus secondary mediators (10). It is worth noting that several closely related measures of parasympathetic nervous system activity were included in this analysis (high- and low-frequency heart rate variability as well as the standard deviation and root mean square of successive differences of the interbeat intervals). Highly correlated variables do not pose analytical challenges for factor analyses and the graphical models used here but would not be advisable to include simultaneously in some statistical models, such as regression analyses where multicollinearity is a concern.

Although graphical models can model complex interrelations among biomarkers and systems, without strict assumptions, they cannot identify causal structures and they raise unique challenges. How would one operationalize a primary end point for a study or trial from such a complex network? Must 23 biomarkers be individually included to control for physiological dysregulation, and what consequences will this have for studies with modest sample sizes? In summary, we believe that the bifactor model of AL offers several practical advantages over alternate methods. When reporting results from latent factor scores, including supplementary analyses showing the results with individual biomarkers can help to clarify whether results are similar or vary across individual biomarkers, although caution is warranted, as examination of individual results for many biomarkers will substantially increase the chance of type 1 errors.

Biobehavioral research in psychosomatic medicine and related disciplines will benefit from further critical debate and research on novel measures of AL as well as statistical methods to quantify AL that appropriately capture our evolving understanding of the basic biology. Future methodological research that compares the strengths and weaknesses of alternate measures of AL in different contexts (e.g., small and large sample sizes, few or many biomarkers) would be especially useful to advance the appropriate use of AL. However, the greatest advancements in the pursuit of the optimal measure of AL may come from a deeper understanding of the processes that may give rise to AL, such as suggested in the letter by Picard et al (11), more sophisticated assessments (e.g., stress reactive and functional measures of biomarkers; (12)), functional measures of biomarkers), and more sophisticated designs (e.g., longitudinal and experimental studies).

JOSHUA F. WILEY, PhD
School of Psychological Sciences and Monash Institute of Cognitive and Clinical Neurosciences
Monash University
Victoria, Australia

TARA L. GRUENEWALD, PhD
Davis School of Gerontology
University of Southern California
Los Angeles, CA

ARUN S. KARLAMANGLA, MD, PhD
TERESA E. SEEMAN, PhD
Division of Geriatrics
David Geffen School of Medicine
University of California
Los Angeles, CA

Source of Funding and Conflicts of Interest: The authors declare no conflict of interest.

SDC Supplemental Content

REFERENCES

1. Wiley JF, Gruenewald TL, Karlamangla AS, Seeman TE. Modeling multisystem physiological dysregulation. Psychosom Med 2016;78:290–301.
2. Crook Z, Booth T. Considering the appropriateness of the factor analytic operationalization of allostatic load. Psychosom Med 2016;78:117–19.
3. Mills PJ, Peterson CT. Multiplexing and beyond in biobehavioral research. Psychosom Med 2016;78:642–5.
4. Cohen AA, Li Q, Milot E, Leroux M, Faucher S, Morissette-Thomas V, Legault V, Fried LP, Ferrucci L. Statistical distance as a measure of physiological dysregulation is largely robust to variation in its biomarker composition. PLoS One 2015;10:e0122541.
5. Wiley JF, Carrington MJ. A metabolic syndrome severity score: a tool to quantify cardio-metabolic risk factors. Prev Med 2016;88:189–95.
6. Biddinger SB, Kahn CR. From mice to men: insights into the insulin resistance syndromes. Annu Rev Physiol 2006;68: 123–58.
7. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P. Causal inference using graphical models with the R package pcalg. J Stat Softw 2012;47:1–26.
8. Pearl J. Causality: models, reasoning and inference, 2nd ed. Cambridge, UK: Cambridge University Press; 2009.
9. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: network visualizations of relationships in psychometric data. J Stat Softw 2012;48:1–18.
10. McEwen BS, Seeman T. Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. Ann N Y Acad Sci 1999;896: 30–47.
11. Picard M, Juster R-P, Sloan RP, McEwen BS. Mitochondrial nexus to allostatic load biomarkers. Psychosom Med 2016;78:114–17.
12. 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.