



Measurement: Interdisciplinary Research and Perspectives

ISSN: 1536-6367 (Print) 1536-6359 (Online) Journal homepage: www.tandfonline.com/journals/hmes20

Validating the Chronic Stress Indicator: A Data-Driven Framework for Integrating Physiological and Socio-Behavioral Measures

Matthew Hill, Emmanuel Obeng-Gyasi & Sayed A. Mostafa

To cite this article: Matthew Hill, Emmanuel Obeng-Gyasi & Sayed A. Mostafa (24 Apr 2026): Validating the Chronic Stress Indicator: A Data-Driven Framework for Integrating Physiological and Socio-Behavioral Measures, Measurement: Interdisciplinary Research and Perspectives, DOI: [10.1080/15366367.2026.2661304](https://doi.org/10.1080/15366367.2026.2661304)

To link to this article: <https://doi.org/10.1080/15366367.2026.2661304>



© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 24 Apr 2026.



[Submit your article to this journal](#)



Article views: 52




[View related articles](#)



[View Crossmark data](#)

Validating the Chronic Stress Indicator: A Data-Driven Framework for Integrating Physiological and Socio-Behavioral Measures

Matthew Hill^a, Emmanuel Obeng-Gyasi^b, and Sayed A. Mostafa ^a

^aDepartment of Mathematics & Statistics, North Carolina A&T State University; ^bDepartment of Built Environment, North Carolina A&T State University

ABSTRACT

Chronic stress contributes to cardiovascular disease, diabetes, and mental health disorders through its cumulative physiological toll, or allostatic load (AL). Building on our previously proposed Chronic Stress Indicator (CSI), this study aims to refine, validate, and compare the CSI using a data-driven framework that integrates physiological, socioeconomic, and behavioral factors. Using data from the MIDUS II biomarker project, a nationally representative sample of U.S. adults aged 34–84, we assessed the performance of the refined CSI and traditional AL indices in predicting multiple stress-related outcomes. Advanced statistical techniques, including the Boruta feature selection algorithm and factor analysis, were applied to optimize biomarker selection and weighting. Sensitivity analyses evaluated the robustness and reliability of each construction. Compared with the original CSI, the refined model incorporating socio-behavioral variables and data-driven weighting demonstrated improved predictive performance for short-term stress outcomes, while both traditional and extended models performed well for long-term outcomes. This work advances the measurement of chronic stress by validating and enhancing the CSI, providing a robust tool for identifying at-risk populations and guiding targeted interventions.

KEYWORDS


Allostatic load; composite indicators; item weights; MIDUS; uncertainty analysis

Introduction

Chronic stress represents a critical public health challenge, manifesting as cumulative physiological dysregulation that elevates risks for cardiovascular diseases, metabolic disorders, immune dysfunction, and mental health disorders (Duric et al., 2016). Despite extensive evidence linking chronic stress to adverse health outcomes, its measurement remains fragmented and inconsistent. The absence of a standardized, reliable measure that can be operationalized using variables commonly available in national health databases limits our ability to identify at-risk populations and design effective interventions (Lazarus, 1990). This lack of a unified approach of measuring stress has been highlighted throughout the literature (Juster et al., 2010). Several studies use varying numbers of sub-indicators which often times come from three or four different biomarker groups, which cause further inconsistencies in composite constructions (Li et al., 2024). Developing a robust, data driven approach to chronic stress measurement is essential because it allows for earlier intervention and more comprehensive analysis on chronic stress.

To address this gap, our prior work (Hill et al., 2024) introduced the Chronic Stress Indicator (CSI) – a data-driven, multidimensional composite that integrates biomarkers and sociodemographic

CONTACT Sayed A. Mostafa  sabdelmegeed@ncat.edu  Department of Mathematics & Statistics, North Carolina A&T State University, 1601 E Market St, Greensboro, NC 27411, United States

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15366367.2026.2661304>

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

factors to quantify chronic stress more comprehensively than traditional allostatic load (AL) measures. The CSI builds on the concept of allostatic load – the cumulative physiological burden resulting from repeated or chronic stress exposure (McEwen, 1998). Operationalizing AL involves converting the theoretical construct – the cumulative physiological “wear and tear” on the body resulting from chronic stress exposure – into a quantifiable index using measurable biomarkers across multiple biological systems, including cardiovascular, metabolic, and inflammatory domains. These biomarkers are aggregated to reflect the degree of multisystem physiological dysregulation. However, prior approaches have varied considerably in marker selection, weighting, and aggregation methods, leading to inconsistencies across studies.

Building upon our previously proposed CSI framework (Hill et al., 2024), this study focuses on extending and validating that approach through methodological refinement and comparative evaluation. Specifically, we enhance biomarker selection to better represent stress-responsive physiological systems and refine weighting strategies to improve the indicator’s sensitivity and specificity for detecting chronic stress. These refinements are accompanied by a systematic evaluation of the uncertainty introduced at each stage of the construction process. To ensure the indicator’s scalability and practical relevance, we integrate sociodemographic and behavioral factors available in large-scale health surveys, facilitating population-level applications. Multiple data-driven weighting strategies, such as Boruta feature selection and factor analysis, are implemented to assess robustness and interpretability, with the resulting CSI variants benchmarked against established operationalizations of AL. Validation is performed using data from the Midlife in the United States (MIDUS) II biomarker project, a nationally representative dataset that enables comprehensive assessment across multiple stress-related outcomes (Ryff et al., 2025). The CSI is evaluated in relation to both short-term and long-term stress measures, including the Perceived Stress Scale (PSS), Satisfaction with Life Scale (SWLS), and number of lifetime symptoms and chronic conditions (NOLSACC). The first two reflect immediate or subjective stress, whereas the latter capture cumulative life stress and health burden. These validation instruments provide external benchmarks for assessing the CSI’s ability to represent multidimensional stress processes. Testing multiple approaches for CSI construction provides benefits for both data science and public health. From a data science perspective, it enhances robustness, interpretability, and methodological transparency. From a public health standpoint, it supports precise risk stratification and targeted interventions for stress-related health outcomes. Ultimately, this work seeks to advance the standardization and reproducibility of chronic stress measurement, providing a scalable, data-driven framework for both research and practice.

Materials and methods

Overview of composite indicator construction

Composite indicators are statistical tools that combine multiple individual measures into a single summary index, allowing for the representation of complex and multidimensional phenomena. They are widely used across disciplines – including economics, social sciences, environmental studies, health, and education – to inform policy and guide data-driven decision-making.

The construction of a composite indicator generally follows several key steps (Nardo et al., 2008). Defining the objective is the initial step of constructing the composite indicator and it involves specifying the dimensions and sub-dimensions to be included to ensure that the indicator adequately represents the underlying concept. Selection of indicators is next, which involves using meaningful variables to capture different dimensions of the construct of interest. It is important to ensure that selected indicators are based on reliable and accessible data sources. Each indicator should align with the theoretical or conceptual framework underpinning the composite indicator. Literature reviews or expert consultations are often used to guide selection.

Data collection and standardization are done immediately following these steps. We must gather data for each selected indicator from credible sources, such as national statistics, international

databases, or population-based surveys. Address missing values through appropriate imputation or exclusion, depending on the degree and mechanism of missingness. Since some indicators may be measured on different scales, standardization helps to ensure comparability. Common approaches include min – max scaling, z-score standardized, or dichotomized as shown in Equations (1–3) respectively:

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)}, \quad (1)$$

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

$$X = \begin{cases} 1; & X \geq \mu \\ 0; & X < \mu \end{cases} \quad (3)$$

where X is the original value, μ is the mean, and σ is the standard deviation.

Weighting of Indicators comes after standardization and can be done equally or unequally. Equal weighting assigns identical weights to all indicators. This approach is simple but assumes all indicators contribute equally to the construct. Unequal weighting assigns different weights to indicators based on their relative importance. Weighting can be based on expert judgment, policy priorities, or theoretical relevance. Alternatively, weighting can be derived from statistical methods such as Principal Component Analysis (PCA), factor analysis, or the Analytic Hierarchy Process (AHP). PCA, for instance, is often used to extract components that explain the maximum variance in the data and to derive corresponding weights.

Aggregation of the composite involves combining the indicators to form the composite score. This is done through simple linear aggregation or geometric aggregation, which is shown in Equations (4) and (5), respectively:

$$I = \sum_{i=1}^n w_i X_i, \quad (4)$$

$$I = \prod_{i=1}^n X_i^{w_i} \quad (5)$$

where X_i are the sub-indicators, w_i are the corresponding weights, and n is equal to the number of sub-indicators. Next the composite is either, min-max scaled, z-score standardized, or dichotomized as shown in Equations (1–3), respectively. Min-max scaling rescales the composite score to a fixed range (e.g., 0–1 or 0–100) for interoperability. This normalization facilitates comparison across populations, time periods, or datasets.

Validation and sensitivity analysis requires an initial check of the internal consistency of the composite indicator using measures such as Cronbach's α or Coefficient ω which can be found in Equations (6) and (7):

$$\alpha = \frac{p\bar{r}}{1 + (p-1)\bar{r}}, \quad (6)$$

$$\omega = \frac{(\sum_{i=1}^p \lambda_i)^2}{(\sum_{i=1}^p \lambda_i)^2 + \sum_{i=1}^p \psi_i}, \quad (7)$$

where p is the number of sub-indicators, r is the average inter-correlation among them, λ_i are the factor loadings, and ψ_i are the uniqueness's of the sub-indicators (Nardo et al., 2008; Padilla & Divers, 2016). To test the composite, it is important to evaluate how variations in weighting, normalization, or indicator inclusion affect the composite results. Monte Carlo simulations are frequently employed for

this purpose. Assessing the validity of the composite indicator by comparing it with established benchmarks or outcomes theoretically related to the target construct is also critical.

Interpretation and reporting are the final steps and provide a clear explanation of what the composite indicator measures and how to interpret its values. This is done in conjunction with discussing potential limitations, including data quality, methodological assumptions, and sensitivity to weighting or scaling choices. Transparency in reporting strengthens credibility and reproducibility.

Traditional AL operationalization

The typical operationalization of chronic stress using AL follows the general framework described in (Guidi et al., 2021; Hill & Obeng-Gyasi, 2022; Mauss & Jarczok, 2021). Given a set of biomarkers $X = \{x_1, x_2, \dots, x_n\}$, AL is defined as in Equation (4), where w_1, w_2, \dots, w_n represent the corresponding weights for the biomarkers, and n denotes the number of sub-indicators. Some studies dichotomize this continuous measure as shown in Equation (3) (Carbone et al., 2022). Alternatively, studies have applied a fixed threshold of three or four for dichotomization (Guidi et al., 2021). In our earlier analyses, we observed that the mean of AL typically falls between 3 and 4, although this range varies across datasets (Hill & Obeng-Gyasi, 2022). While this operationalization is suitable for basic applications, it lacks the precision needed to fully capture chronic stress. Specifically, it focuses exclusively on biological markers and omits the socio-behavioral dimensions that substantially contribute to chronic stress. Moreover, it implicitly assumes equal weighting across all sub-indicators, such that $w_n = 1 \quad n \in \{1, \dots, p\}$.

In most cases, the biomarker variables x_n are measured on different scales and may not contribute equally to overall AL. Additionally, two individuals can share the same AL value despite having distinct stress profiles. One improvement to this approach involves standardizing or transforming x_n prior to aggregation, as expressed in Equation (2). This formulation improves comparability across biomarkers and partially addresses scale heterogeneity. However, our study seeks to further enhance chronic stress measurement through optimized weighting, variable selection, and sensitivity analysis—three elements recognized as essential in composite index construction (Becker et al., 2017; Freudenberg, 2003; Greco et al., 2019; Mariani & Ciommi, 2022).

Chronic stress indicator

To better understand the multifactorial nature of chronic stress and its influence on disease risk, the CSI integrates sociodemographic factors and physiological biomarkers to produce a continuous stress-risk score, reflecting variation across individuals from low to high chronic stress burden (Hill et al., 2024). This multidimensional design recognizes that stress arises not only from biological dysregulation but also from social and behavioral determinants that shape exposure and resilience (Cohen & Janicki-Deverts, 2012; Folkman & Moskowitz, 2000; Lavretsky & Newhouse, 2012; Rohleder, 2019).

The algorithm in Table 1 summarizes the general procedure for constructing the CSI, following the composite indicator development framework outlined by Nardo et al. (2008). Given a dataset containing both biomarker and socio-behavioral variables X , the algorithm computes CSI values through sequential steps involving data preprocessing, normalization, feature selection, weighting, and aggregation. These refinements enable a more standardized, data-driven measure of chronic stress suitable for population-level research and clinical applications.

To enhance interoperability and facilitate classification, binary versions of both AL and CSI were also created. For each construction, we used statistical quartiles as thresholds to distinguish between low- and high-risk individuals. Sub-indicators in the high-risk quartile were assigned a score of one, and those in the low-risk quartile a score of zero. Summing across sub-indicators yielded an overall AL or CSI score, with individuals scoring greater than or equal to the sample mean classified as high risk and all others as low risk. Using the outlined framework, we generated 32 distinct constructions of AL and CSI, each validated using regression models. One configuration employed an automated feature

Table 1. CSI construction algorithm.

Require: $X = \{x_1, x_2, \dots, x_p\}$ Items (e.g., biomarkers and sociodemographic factors)

(1) **Outliers: (Optional)**

for each $x_i \in X, i \in \{1, 2, \dots, p\}$, where p is the number of sub-indicators **do**

 Compute interquartile range (IQR): $IQR = Q_3 - Q_1$, Where Q_1 and Q_3 are the first and third quartiles respectively

 Identify potential outliers:

$$O_i = \begin{cases} x_i > Q_3 + (1.5 * IQR) \\ x_i < Q_1 - (1.5 * IQR) \end{cases}$$

 Remove O_i from dataset

end for

(2) **Standardization:**

a. **Z-Score Standardization:**

$$\tilde{X} = \left(\frac{x_i - \bar{x}_j}{\sigma_{x_j}} \right)$$

 where \tilde{X} is the standardized indicators, \bar{x}_j and σ_{x_j} are the mean and standard deviation respectively of each biomarker, or

b. **Min-Max Scaling:** $\tilde{X} = \frac{X - \min(X)}{\max(X) - \min(X)}$

(3) **Missing Data Management:**

a. Complete Case Analysis (CCA), or

b. Multivariate Imputation by Chained Equations (MICE)

(4) **Weighting:** Obtain weight matrix $W = \{w_i\}$, through equal or unequal weighting

a. **Equal Weighting:** $W = w_i = 1 \quad i \in \{1, 2, \dots, p\}$, or

b. **Unequal Weighting:** See tables 24 & 25 for algorithms of Boruta and factor analysis respectively

(5) **Aggregation:**

a. **Weighted Linear Aggregation (WLA)** $CSI = \sum_{i=1}^p w_i \tilde{x}_i$

(6) **Standardization of CSI:**

$$CSI' = \left(\frac{CSI - \overline{CSI}}{\sigma_{CSI}} \right)$$

 where CSI' is the standardized composite, \overline{CSI} and σ_{CSI} are the mean and standard deviation respectively of the CSI.

(7) **Dichotomization: (Optional)**

$$CSI_q = \begin{cases} 1; & CSI' \geq \overline{CSI'} \\ 0; & CSI' < \overline{CSI'} \end{cases}$$

Output: CSI_q or CSI'

selection approach based on a penalized factor model that incorporated an adaptive lasso penalty on factor loadings. This procedure effectively shrinks non-informative sub-indicators toward zero, yielding a parsimonious yet interpretable composite score. Additionally, penalized factor models were built for all of the imputed datasets. In 39 of the datasets, total cholesterol and income were removed from CSIAUTO. The remainder of the imputations removed tobacco use, income, and total cholesterol from CSIAUTO.

Table 2 summarizes the biomarker and sociodemographic variables included in the baseline AL and CSI constructions. Additional reconstructions incorporated alternative subsets of these factors to evaluate sensitivity and performance, given known variability in AL computation across studies (Duong et al., 2017). The naming conventions for the AL and CSI constructions are summarized in Table 3.

Statistical techniques for weighting the CSI

Previous research has emphasized the critical role of weighting in the construction of composite indicators (Greco et al., 2019). The development of the CSI utilized two primary statistical techniques to assign weights. These included Boruta feature selection and Factor Analysis (FA). We applied the

Table 2. Variables involved in AL/CSI constructions.

	CSI10	CSIAUTO	AL10	AL5
RMSSD		√		√
SBP	√	√	√	
DBP	√	√	√	√
TC	√		√	
HBA1C	√	√	√	√
BMI	√	√	√	
HDL		√	√	
TG		√	√	
Alb			√	
CLCR		√	√	
CRP		√	√	
LDL		√		√
WC		√		√
Education Level	√			
Poverty status	√			
Alcohol consumption	√	√		
Tobacco use	√	√		
Physical activity	√	√		

Table 3. Naming conventions for AL/CSI constructions.

Suffix	Description
.q	Dichotomized using the sample mean
.boruta	Weighted using Boruta algorithm
.FA	Weighted using factor analysis loadings
.FA_SCORES	Used component score as composite

Table 4. Boruta algorithm.

Require: $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_p\}$, Standardized Features from the Algorithm in Table 1.

1. **Initialize:** Mark all features as "Tentative".
2. **Iterative Comparison** (Repeat until all features are classified or max iterations reached):
 - a. **Copy & Shuffle:** Create shadow features S by randomly shuffling each tentative feature to generate a baseline distribution.
 - b. **Train Random Forest Model:** Combine \tilde{X} and S into one feature set.
 - i. Use DOFOR as the target variable.
 - ii. Obtain importance scores for both original and shadow features:

$$I_j = \frac{1}{n_{tree}} \sum_{t=1}^{n_{tree}} G_j^t,$$

where G_j^t denotes the decrease in Gini impurity attributed to variable j in tree t .

- c. **Compute Threshold and Evaluate Significance:** for each feature j , record a hit if $I_j > T$, where $T = \max(I_{shadow})$. Perform a two-sided binomial test on the cumulative hits.
 - i. Mark features with significantly higher hits than expected as confirmed or important and keep in the model.
 - ii. Mark features with significantly lower hits than expected as rejected or unimportant and remove from the model.
 - iii. Carry remaining tentative or undecided features to the next iteration.

3. Compute Weights:

$$w_j = \frac{\bar{I}_j}{\sum \bar{I}_j},$$

where \bar{I}_j is the min-max-scaled mean importance score.

4. **Output:** the weights for all confirmed features: $W = \{w_1, w_2, \dots, w_k\}$

Boruta algorithm, a wrapper-based feature selection method in *R* that uses a random forest classifier, to identify the most important variables for the CSI. The Boruta approach enabled the use of importance scores for weighting of our composites. By comparing the importance of each original feature with that of its randomly permuted shadow counterpart, Boruta effectively ranks biological and sociodemographic variables according to their contribution to chronic stress. Only those features exhibiting statistically significant importance are retained in the CSI. Table 4 shows the Boruta Algorithm. Our target variable for Boruta was DOFOR as the loss of someone has been shown to undermine an individual's physical and mental health (Umberson et al., 2017).

In parallel, factor analysis was used to derive weights by estimating factor loadings and component scores. These loadings represent the shared variance among observed variables and thus quantify the contribution of each indicator to the latent construct of chronic stress. The detailed factor analysis procedure is provided in Supplementary Table S1. Factor scores were standardized and aggregated to produce the final CSI score.

For comparison, AL was defined following prior research and our earlier work, integrating biomarkers across immune, cardiovascular, and metabolic systems. AL values were constructed through standardization and summation of individual biomarkers, consistent with established approaches in the literature.

Reliability analysis of CSI

Reliability of the CSI was assessed using Coefficient ω , which has been shown to provide a more accurate estimate of internal consistency than Cronbach's alpha, particularly when the assumption of tau-equivalence is violated. Since our sub-indicators are not equally weighted, ω is more appropriate for this context. The key assumptions of Coefficient Omega include unidimensionality, heterogeneous factor loadings, and congeneric items.

Study sample

This study aimed to compare and validate multiple CSIs using data from the MIDUS II Biomarker Project (Ryff et al., 2025). The MIDUS II study is a nationally representative sample of U.S. adults aged 34–84, collected between 2004 and 2009, with an initial sample size of 1,054 participants. It provides extensive biological and sociodemographic data suitable for the operationalization and validation of chronic stress measures. Participants were included if they had complete data for the biomarkers and sociodemographic variables selected for CSI construction and validation. Sampling weights were not applied in this analysis, as only one subsample contained weights, and their use would have substantially reduced the analytic sample. Furthermore, our primary goal was to assess the effect of internal sub-indicator weighting within the CSI framework.

The CSI was constructed using a combination of sociodemographic, behavioral, and biological indicators. Demographic and stress-related factors included age, gender, race/ethnicity, income level, physical activity, education, perceived stress scale (PSS), death of a friend or relative (DOFOR), satisfaction with life scale (SWLS), and the number of lifetime symptoms and chronic conditions (NOLSACC). These measures were obtained through self-administered questionnaires and phone interviews. Health-related behaviors included alcohol consumption and tobacco use.

Laboratory-based biomarkers included root mean square of successive differences (RMSSD), C-reactive protein (CRP), fibrinogen, glycated hemoglobin (HbA1c), high- and low-density lipoprotein cholesterol (HDL, LDL), triglycerides (TG), total cholesterol (TC), albumin (Alb), creatinine clearance (CLCR), body mass index (BMI), waist circumference (WC), systolic and diastolic blood pressure (SBP, DBP), and resting pulse. All blood samples were collected at one of three clinical research centers (University of Wisconsin, Georgetown University, and UCLA), following the standardized MIDUS II biomarker collection protocol (Ryff et al., 2025).

Table 5. Percent missing for variables of this study.

Variable Name	(%) Missing
Tobacco	38.61
Income	10.06
RMSSD	8.73
Alcohol	5.03
Glucose	1.42
CRP	1.33
Fibrinogen	1.33
HBA1C	1.33
HDL	1.04
LDL	1.04
TG	0.85
TC	0.85
Albumin	0.66
Race/Ethnicity	0.28
Education	0.28
PSS	0.28
SWLS	0.28
CLCR	0.28
Pulse	0.19
BMI	0.09
WC	0.09
SBP	0.09
DBP	0.09
Age	0
Sex	0
Physical Activity	0
DOFOR	0
NOLSACC	0

For simplicity, annual income was dichotomized at \$22,500—the approximate federal poverty threshold for the study period – coded as one for incomes at or below this value and zero otherwise (Fass, 2009). A full list of variables and their responses are provided in Supplementary Table S2.

Initial missing data treatment

Missing data were first handled using complete case analysis (CCA), also known as listwise deletion, which excludes participants with missing values in any variable used for analysis. Variables with the highest missingness were tobacco use (38.6%), income (10.1%), and RMSSD (8.73%), while all others had less than 6% missing data. Table 5 shows the percent missing for all variables of interest. After applying CCA, the analytic sample size was $n = 510$. While straightforward to implement, CCA assumes data are missing completely at random (MCAR); violation of this assumption can bias estimates. To assess robustness under the less restrictive missing at random (MAR) assumption, we later conducted sensitivity analyses using multiple imputation via chained equations (MICE).

Validation framework

Validation of the CSI constructions was performed using three established psychosocial measures: PSS, SWLS, and NOLSACC. The PSS was measured using the 10-item self-rated scale, with responses ranging from 1 (never) to 5 (very often), summed to produce the total PSS score (Cohen et al., 1983; Smith et al., 2014). The SWLS used a 7-point Likert scale, NOLSACC was treated as a continuous variable, and DOFOR was binary (yes/no).

Each CSI and AL configuration was evaluated using regression-based associations with these validation measures. Linear regression models were fitted for PSS, SWLS, and NOLSACC. For each configuration, explanatory power was quantified using R^2 :

$$R^2 = 1 - \left(\frac{SS_{RES}}{SS_{TOTAL}} \right),$$

where SS_{RES} is the residual sum of squares and SS_{TOTAL} is the total sum of squares. Configurations were assessed based on their R^2 values, standardized coefficients (β) and associated p -values, with higher R^2 , $|\beta|$, and smaller p -values indicating better model performance. All analyses were conducted in *R* version 4.4.4 (R Development Core Team, 2022) using the packages *Boruta*, *psych*, *penfa*, and *mice*. For the *Boruta* algorithm, we used 500 trees and 100 iterations, which are the default values in *R*. Future work should look at further optimizing the number of trees and iterations to allow for more stable importance scores. Factor analysis, regression modeling, and sensitivity testing were implemented in *R* with a significance threshold of 0.05.

Results

The summary statistics for our studies can be found in Supplementary Table S3. All continuous variables show mean and standard deviation in parentheses, while all categorical variables show proportions. Summary statistics for the MICE datasets were pooled using Rubin's rule as is customary in the literature (White et al., 2011).

Reliability results

We evaluated the internal consistency of each AL and CSI construction using Coefficient ω , which captures the degree to which sub-indicators measure a common latent construct, assumed here to represent chronic stress. As shown in Table 6, the CSI AUTO construction – derived via automatic statistical feature selection – achieved the highest reliability ($\omega = 0.45$). This was followed by AL10 and CSI10. In contrast, AL5 demonstrated the lowest reliability ($\omega = 0.32$), suggesting weaker coherence among its indicators.

Validation results

To evaluate predictive performance across different CSI and AL constructions, we fitted linear regression models using multiple validation outcomes. The initial models used complete case analysis (CCA) to handle missing data and did not incorporate survey weights. An alternative approach to handling missing data; namely, multiple imputation, is covered in the uncertainty analysis section below. The predictive validity of each construction was assessed against perceived stress (PSS), subjective well-being (SWLS), and long-term health burden (NOLSACC). All models were adjusted for age and sex. R^2 and p -value 95% confidence intervals were constructed using 500 bootstrap samples. Figures 1–3 show standardized coefficients (β), $\sqrt{p$ -values}, and R^2 as well as 95% confidence intervals across constructions for PSS, SWLS, and NOLSACC. Constructions are ordered by the number of sub-indicators they include and are labeled to indicate whether variable selection, binary transformation, or statistical weighting was applied. $\sqrt{p$ -values} were used to aid in the readability of the graphs.

For PSS, the highest-performing models were AL5.q.FA and AL5.q.FA_SCORES ($\beta = 0.3009$, p -value = 0.0010, $R^2 = 0.0679$), which combined five health biomarkers and were dichotomized. AL5.q.FA was weighed via communalities of a one-factor model, while AL5.q.FA_SCORES was constructed using the component scores from a one-factor model.

Table 6. Coefficient Omega for AL/CSI constructions.

Construction	CSI AUTO	AL10	CSI10	AL5
Omega (ω)	0.45	0.34	0.33	0.32

Standardized Coefficient Estimates – PSS, SWLS, NOLSACC

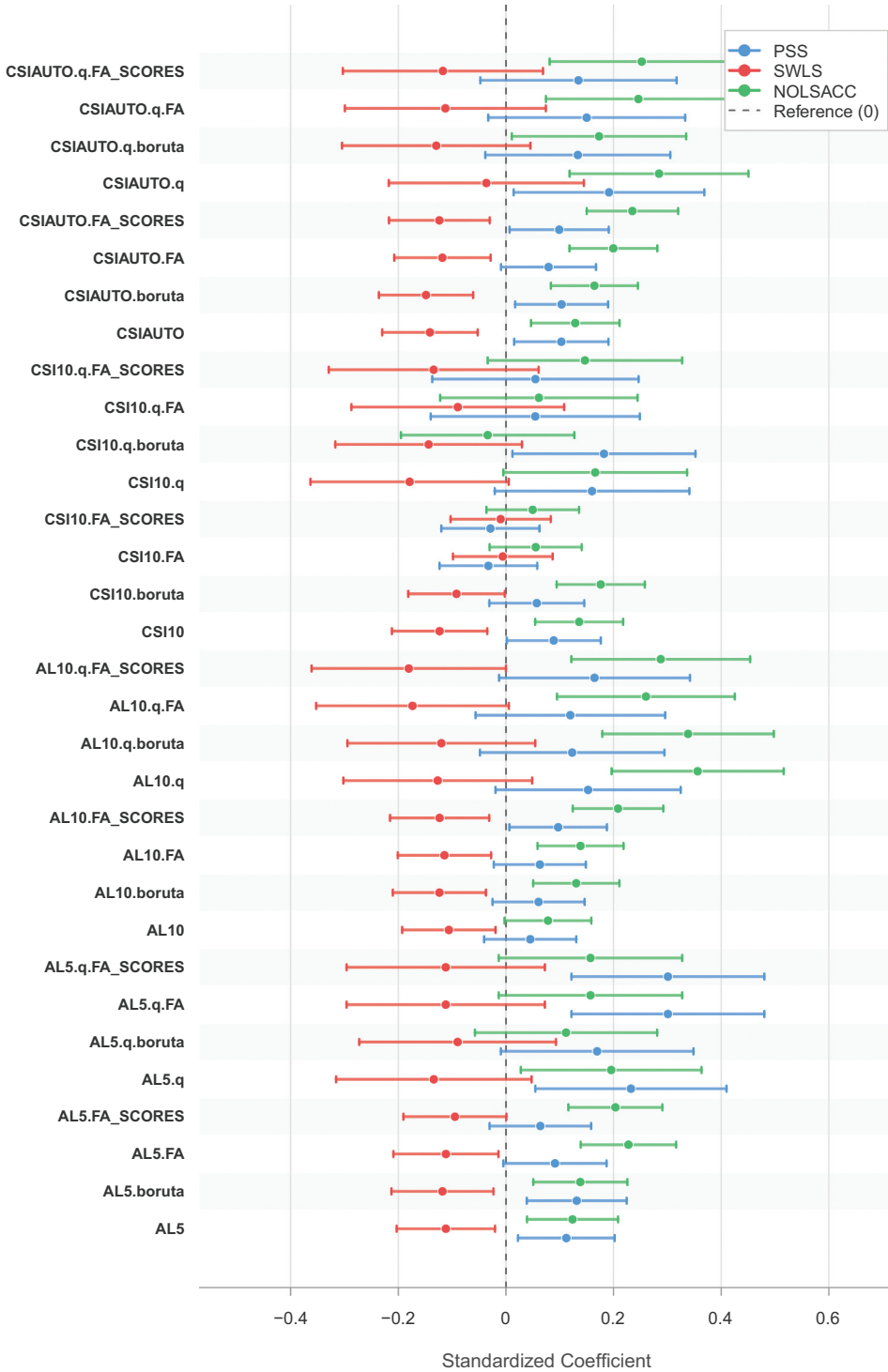


Figure 1. Standardized coefficients and 95% confidence intervals of al and CSI constructions across validation outcomes using CCA data.

Square Root of P-Values – PSS, SWLS, NOLSACC



Figure 2. Square root of p -value and 95% confidence intervals of al and CSI constructions across validation outcomes using CCA data.

R² Values – PSS, SWLS, NOLSACC

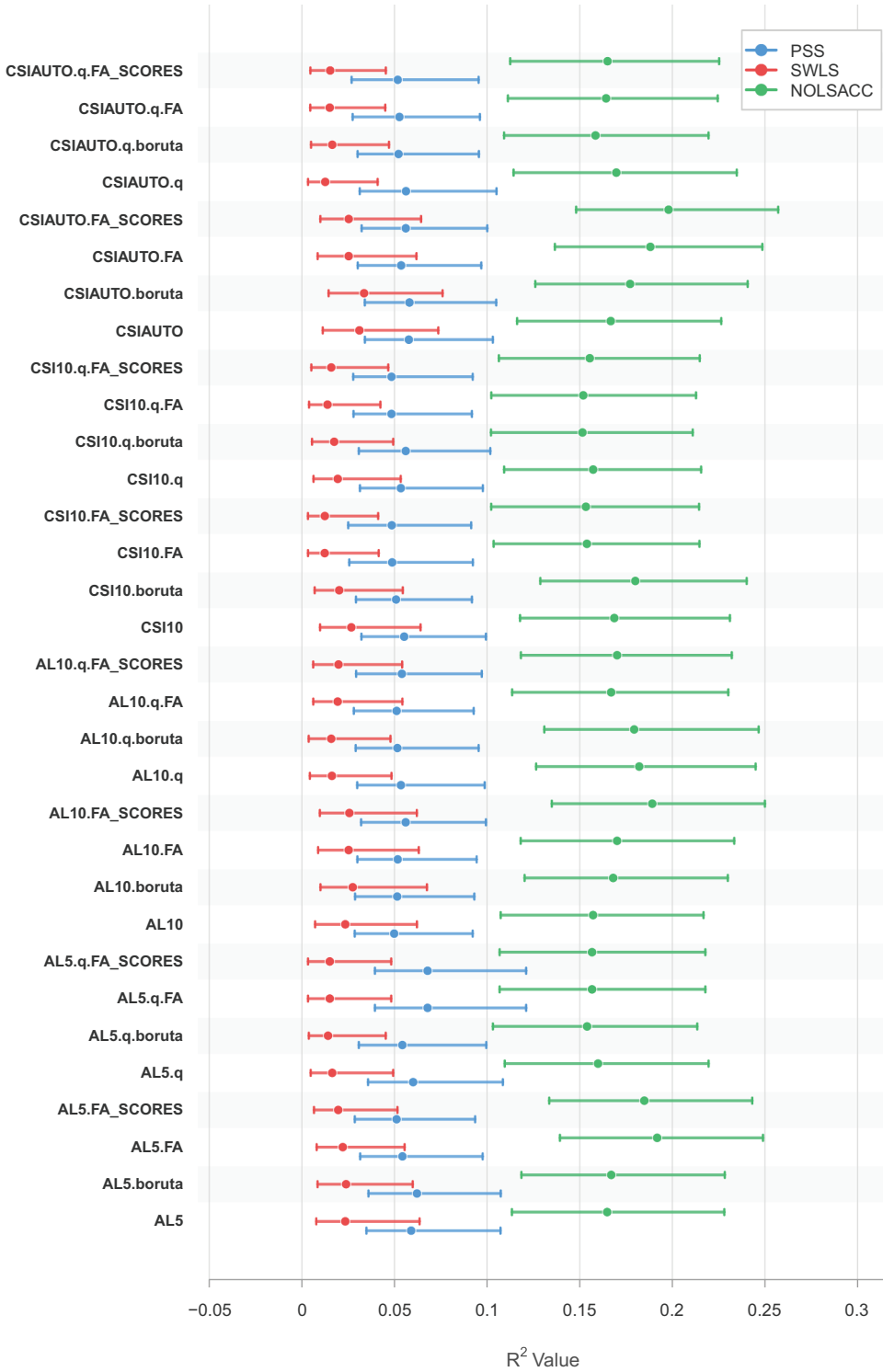


Figure 3. R² values and 95% confidence intervals of al and CSI constructions across validation outcomes using CCA data.

For SWLS, the top-performing model was CSIAUTO.boruta ($\beta = -0.1486$, $p - value = 0.0009$, $R^2 = 0.0336$), which applied penalized factor analysis for variable selection and derived weights from the Boruta algorithm. The second-best model, CSIAUTO ($\beta = -0.1413$, $p - value = 0.0019$, $R^2 = 0.0311$), used the same sub-indicators but was equally weighted. Both constructions incorporated variable selection through penalized factor analysis, suggesting that judicious variable selection can improve predictive accuracy. Most models showed negative coefficients for SWLS, which aligns with the inverse relationship between satisfaction and stress.

For NOLSACC, representing long-term health burden, CSIAUTO.FA_SCORES achieved the highest fit ($\beta = 0.2349$, $p - value < 0.0001$, $R^2 = 0.1980$). This model utilized variable selection and was constructed using component scores. The second-best model, AL5.FA ($\beta = 0.2275$, $p - value < 0.0001$, $R^2 = 0.1918$), consisted of five health biomarkers and was weighed using communalities from a one-factor model. These models produced the highest standardized coefficients and R^2 values as well as the lowest p -values. This is expected as the underlying biomarkers are typically used to diagnose many of the symptoms and chronic conditions measured in NOLSACC.

Uncertainty results

Before finalizing the CSI, we examined the latent structure of the sub-indicators using exploratory factor analysis (EFA). Factor loadings below 0.001 were excluded for clarity (see Supplementary Table S4). Factor 1 was significantly associated with Factor 4 ($p - value < 0.01$), and Factor 2 was correlated with Factor 3. The comparative fit index (CFI = 0.910) and root mean square error of approximation (RMSEA = 0.085) indicated acceptable model fit, supporting the aggregation of sub-indicators into a single composite measure.

We next assessed how sensitive our 32 AL/CSI constructions were to changes in preprocessing and data handling. Specifically, we examined how they shifted under different standardization methods, outlier treatments, and missing-data procedures. This analysis quantifies the robustness of the CSI to methodological choices, ensuring that its interpretation is not unduly influenced by preprocessing artifacts. Two data standardization methods were compared: z-score normalization, which scales variables to have a mean of 0 and a standard deviation of 1, and min - max scaling, which rescales values to a range of [0, 1]. We further assessed the effect of including versus excluding outliers. Our reference data was the data using CCA with z-score standardization.

We also investigated the uncertainty in missing data handling. Specifically, we compared complete case analysis (CCA) with multiple imputation by MICE. For MICE, predictive mean matching (PMM) was used for continuous variables and logistic regression for categorical ones. Following (White et al., 2011), we generated $m = 40$ imputed datasets (since the highest missing value rate in our data was about 38%) with 10 iterations per chain. Parameter estimates were combined using Rubin's Rules:

$$\tilde{S} = 1 \sum_{i=1}^m S_i,$$

where S_i denotes the statistic of interest (e.g., p -value or R^2) from the i^{th} imputation.

Uncertainty in standardization technique

Supplementary Figures S1-S3 show standardized coefficients (β), $\sqrt{p - values}$, and R^2 as well as 95% confidence intervals across constructions for PSS, SWLS, and NOLSACC. For these constructions, CCA data was used and all continuous sub-indicators were min-max scaled as opposed to z-score standardization. These conditions produced results that were broadly comparable to Figures 1–3, with the same general hierarchy of outcomes (NOLSACC > PSS > SWLS) and the same pattern of CSI composites outperforming AL composites. However, subtle differences in the absolute magnitude of R^2 values were noted for specific indicators. FA and FA_SCORES weighted variants within the CSI10

and CSIAUTO families exhibited slightly attenuated R^2 values relative to the reference, particularly for SWLS. This may reflect the sensitivity of factor-analytically derived weights to the distributional properties of the input variables, which differ between z-score and min-max normalization. The dichotomized indicators showed a similar pattern to the reference, and the AL-based composites remained comparably poor predictors of all three outcomes regardless of standardization approach, reinforcing the robustness of this finding across methodological choices.

Uncertainty in outlier removal

Next, we investigated the uncertainty in the inclusion/exclusion of outliers. Supplementary Figures S4-S6 show standardized coefficients (β), $\sqrt{p - \text{values}}$, and R^2 as well as 95% confidence intervals across constructions for PSS, SWLS, and NOLSACC. These conditions closely replicated the reference pattern in the standardized coefficients (β), $\sqrt{p - \text{values}}$, and R^2 as well as 95% confidence intervals across composite indicators and outcomes. SWLS retained its position as the outcome most strongly explained by CSI-based composites, and the overall range of R^2 values was consistent with that observed in the reference. The congruence between Supplementary Figures S1-S3 and S4-S6 suggest that outlier removal did not substantively distort the explanatory structure of the composites and that the z-score standardization approach preserves the relative ordering of indicator performance. Minor deviations were observed for NOLSACC, where CSI10.q and CSIAUTO.q variants showed marginally reduced R^2 compared to the reference, possibly reflecting reduced variability in the outcome after case exclusion.

Uncertainty in missing data handling

Additionally, we altered the handling of missing data to use MICE as opposed to CCA. Supplementary Figures S7-S9 show standardized coefficients (β), $\sqrt{p - \text{values}}$, and R^2 as well as 95% confidence intervals across constructions for PSS, SWLS, and NOLSACC. These models used MICE to handle missing data and z-score standardization. These conditions displayed the most notable departures from the reference pattern. Most prominently, the overall scale of R^2 values was compressed, with the upper bound reduced to approximately 0.20 compared to the 0.30 observed in the reference and complete case analyses. This attenuation is consistent with the well-documented phenomenon of variance inflation in imputed datasets, whereby the incorporation of imputation uncertainty into pooled estimates via Rubin's Rules tends to moderate point estimates, yielding more conservative assessments of explained variance. These findings also reflect the gain in precision from retaining all participants as opposed to CCA.

Taken together, the analytical conditions demonstrate a high degree of convergence in the performance of composite indicators, lending credibility to the substantive conclusions drawn from any individual approach. The consistent superiority of CSI-based composites, particularly those incorporating social determinants of health alongside physiological biomarkers, across all conditions and outcomes suggests that the broader conceptual scope of the CSI framework is a genuine advantage in predicting psychosocial wellbeing, rather than an artifact of a particular analytical choice. The most consequential methodological difference lies between the complete case analyses and the multiply imputed approach. The compression of R^2 values underscores the importance of selecting an analytical framework that is appropriate to the inferential goals of the study. Where the primary aim is conservative, population-level inference that accounts for missing data uncertainty, these results should be treated as the most methodologically rigorous. Where the aim is to characterize the maximum explanatory potential of the composites under complete data conditions, the CCA results – and their close alignment with the reference – provide a useful benchmark. The minimal differences between CCA datasets further suggest that the standardization strategy is a secondary methodological

concern relative to the choice of imputation approach, and that substantive conclusions regarding composite performance are robust to this choice. Future work may consider whether hybrid approaches, such as applying different aggregation techniques within a multiple imputation framework, yield further improvements in either explanatory power or distributional properties of the pooled estimates. This further reinforces the CSI's reliability as a robust, data-driven measure of chronic stress.

Discussion

This study extends the work of Hill et al. (2024), which originally introduced the CSI as a composite tool for comprehensive stress analysis. Building upon that foundation, the present research refines the CSI through methodological enhancements to improve overall validity of the composite. While prior studies – particularly those focusing on AL (Seeman et al., 1997) have centered primarily on physiological biomarkers, they often neglect socio-behavioral stressors that contribute substantially to chronic stress. The refined CSI addresses this limitation by systematically integrating sociodemographic and behavioral variables such as income, education, physical activity, and substance use, thereby operationalizing the biopsychosocial model of health at the population level.

Methodologically, this study advances the CSI framework by employing Boruta feature selection and factor analysis to optimize variable weighting, improving upon earlier summation-based methods by retaining only the most informative predictors (Breiman, 2001). Validation against established measures of stress and well-being (PSS, SWLS) further demonstrates this improved utility relative to traditional AL models. Notably, CSI and AL variants (e.g., CSIAUTO.boruta, AL5.q.FA) achieved superior model fits, highlighting the benefits of using boruta for feature importance and consistent with factor-based approaches observed in previous work (McCaffery et al., 2012). To facilitate adoption and adaptation of the CSI constructions examined in this study, we developed a publicly available R Shiny application that enables users to construct and compare multiple CSI variants using user-supplied data. The application is available at <http://mhcode.shinyapps.io/comp>.

The CSI offers a comprehensive and flexible tool for assessing chronic stress in both clinical and public health contexts. In healthcare settings, it can aid in identifying individuals at high risk for stress-related conditions, guiding interventions that address both physiological and social determinants of health. In public health, its inclusion of factors like poverty and physical activity enables data-driven planning and equitable resource allocation. Additionally, by leveraging feature selection and factor-analytic weighting, the CSI illustrates how machine learning can enhance composite health measures and improve predictive accuracy. Beyond individual risk assessment, the CSI supports efforts toward health equity by quantifying socioeconomic and behavioral stressors. This allows practitioners and policymakers to advocate for structural interventions such as improving education access, healthcare availability, and neighborhood conditions that reduce chronic stress at the population level.

Limitations

Despite the promise of the CSI framework studied in this paper, several challenges remain. Reliable computation of the CSI requires comprehensive biomarker and sociodemographic data, which are not uniformly available across healthcare systems. Continuous models, while statistically robust, may be less intuitive for practical interpretation. Moreover, variability in data collection and scaling could hinder cross-population comparability. To address these issues, simplified context-specific CSI variants and practitioner training in composite indicator interpretation are recommended. Integration into electronic health records (EHRs) would further facilitate automated computation and longitudinal tracking.

Future work

Future research should validate the CSI across diverse populations and expand its coverage to include additional contextual variables such as employment, neighborhood characteristics, and social support. Importantly, extending the CSI to longitudinal data would allow researchers to examine changes in cumulative stress and allostatic load over time, enabling stronger inferences about temporal dynamics and causal pathways. Applying machine learning techniques (e.g., support vector machines, neural networks) may further refine variable selection and weighting. Standardizing AL and CSI methodologies will be essential for comparability and replication. Additionally, this study underscores the critical role of sensitivity analysis in composite indicator construction, which warrants deeper exploration in future longitudinal applications.

Acknowledgments

The authors thank the staff of the Clinical Research Centers at the University of Wisconsin-Madison, UCLA, and Georgetown University for their support in conducting the MIDUS study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Research reported in this study was supported by the National Institute of General Medical Sciences of the National Institutes of Health under [Award Number R16GM149473]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Matthew Hill was supported by the North Carolina A&T State University Chancellor's Distinguished Fellowship, a Title III HBGI grant from the U.S. Department of Education.

ORCID

Sayed A. Mostafa  <http://orcid.org/0000-0002-8113-702X>

Open practice statement

The code and data used in all analyses within this study are publicly available at the following GitHub repository: <https://github.com/mkx14/Operationalization-and-Validation-of-Chronic-Stress.git>.

References

- Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, 80, 12–22. <https://doi.org/10.1016/j.ecolind.2017.03.056>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Carbone, J. T., Clift, J., & Alexander, N. (2022). Measuring allostatic load: Approaches and limitations to algorithm creation. *Journal of Psychosomatic Research*, 163, 111050. <https://doi.org/10.1016/j.jpsychores.2022.111050>
- Cohen, S., & Janicki-Deverts, D. (2012). Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009, 1. *Journal of Applied Social Psychology*, 42(6), 1320–1334. <https://doi.org/10.1111/j.1559-1816.2012.00900.x>
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4), 385–396. <https://doi.org/10.2307/2136404>
- Duong, M. T., Bingham, B. A., Aldana, P. C., Chung, S. T., & Sumner, A. E. (2017). Variation in the calculation of allostatic load score: 21 examples from NHANES. *Journal of Racial and Ethnic Health Disparities*, 4(3), 455–461. <https://doi.org/10.1007/s40615-016-0246-8>

- Duric, V., Clayton, S., Leong, M. L., & Yuan, L. L. (2016). Comorbidity factors and brain mechanisms linking chronic stress and systemic illness. *Neural Plasticity*, 2016, 5460732. <https://doi.org/10.1155/2016/5460732>
- Fass, S. (2009). *Measuring poverty in the United States* fact sheet. National Center for Children in Poverty, Mailman School of Public Health.
- Folkman, S., & Moskowitz, J. T. (2000). Stress, positive emotion, and coping. *Current Directions in Psychological Science*, 9(4), 115–118. <https://doi.org/10.1111/1467-8721.00073>
- Freudenberg, M. (2003). Composite indicators of country performance: A critical assessment. *Oecd science, technology and industry working papers*. <https://doi.org/10.1787/405566708255>
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141, 61–94. <https://doi.org/10.1007/s11205-017-1832-9>
- Guidi, J., Lucente, M., Sonino, N., & Fava, G. A. (2021). Allostatic load and its impact on health: A systematic review. *Psychotherapy and Psychosomatics*, 90(1), 11–27. <https://doi.org/10.1159/000510696>
- Hill, M., Mostafa, S., & Obeng-Gyasi, E. (2024). Chronic stress indicator: A novel tool for comprehensive stress analysis. *International Journal of Environmental Research and Public Health*, 21(3), 302. <https://doi.org/10.3390/ijerph21030302>
- Hill, M., & Obeng-Gyasi, E. (2022). The association of cytomegalovirus IgM and allostatic load. *Diseases*, 10(4). <https://doi.org/10.3390/diseases10040070>
- Juster, R.-P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience and Biobehavioral Reviews*, 35(1), 2–16. <https://doi.org/10.1016/j.neubiorev.2009.10.002>
- Lavretsky, H., & Newhouse, P. A. (2012). Stress, inflammation, and aging. *American Journal of Geriatric Psychiatry*, 20(9), 729–733. <https://doi.org/10.1097/JGP.0b013e31826573cf>
- Lazarus, R. S. (1990). Theory-based stress measurement. *Psychological Inquiry*, 1, 3–13. https://doi.org/10.1207/s15327965pli0101_1
- Li, C., Howard, S. P., Rogers, C. R., Andrzejak, S., Gilbert, K. L., Watts, K. J., Bevel, M. S., Moody, M. D., Langston, M. E., Doty, J. V., Toriola, A. T., Conwell, D., & Moore, J. X. (2024). Allostatic load, educational attainment and risk of cancer mortality among US men. *JAMA Network Open*, 7(12), e2449855–e2449855. <https://doi.org/10.1001/jamanetworkopen.2024.49855>
- Mariani, F., & Ciommi, M. (2022). Aggregating composite indicators through the geometric mean: A penalization approach. *Computation*, 10(4), 64. <https://doi.org/10.3390/computation10040064>
- Mauss, D., & Jarczok, M. N. (2021). The streamlined allostatic load index is associated with perceived stress in life-findings from the MIDUS study. *Stress*, 24(4), 404–412. <https://doi.org/10.1080/10253890.2020.1869935>
- McCaffery, J. M., Marsland, A. L., Strohacker, K., Muldoon, M. F., & Manuck, S. B. (2012). Factor structure underlying components of allostatic load. *PLOS ONE*, 7(10), e47246. <https://doi.org/10.1371/journal.pone.0047246>
- McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840(1), 33–44. <https://doi.org/10.1111/j.1749-6632.1998.tb09546.x>
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., Giovannini, E. (2008). Handbook on constructing composite indicators: Methodology and user guide, (2nd ed.). OECD Publishing. <https://publications.jrc.ec.europa.eu/repository/handle/JRC47008>
- Padilla, M. A., & Divers, J. (2016). A comparison of composite reliability estimators: Coefficient omega confidence intervals in the current literature. *Educational and Psychological Measurement*, 76(3), 436–453. <https://doi.org/10.1177/0013164415593776>
- R Development Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rohleder, N. (2019). Stress and inflammation - the need to address the gap in the transition between acute and chronic stress effects. *Psychoneuroendocrinology*, 105, 164–171. <https://doi.org/10.1016/j.psyneuen.2019.02.021>
- Ryff, C. D., Seeman, T., & Weinstein, M. (2025). Midlife in the United States (MIDUS 2): Biomarker project, 2004–2009 (ICPSR 29282; Version V11) [Data set]. Inter-university Consortium for Political and Social Research. Retrieved August 01, 2025, from <https://doi.org/10.3886/ICPSR29282.v11>
- Seeman, T. E., Singer, B. H., Rowe, J. W., Horwitz, R. I., & McEwen, B. S. (1997). Price of adaptation-allostatic load and its health consequences: MacArthur studies of successful aging. *Archives of Internal Medicine*, 157(19), 2259–2268. <https://doi.org/10.1001/archinte.1997.00440400111013>
- Smith, K. J., Rosenberg, D. L., & Timothy Haight, G. (2014). An assessment of the psychometric properties of the perceived stress scale-10 (PSS 10) with business and accounting students. *Accounting Perspectives*, 13(1), 29–59. <https://doi.org/10.1111/1911-3838.12023>
- Umberson, D., Olson, J. S., Crosnoe, R., Liu, H., Pudrovska, T., & Donnelly, R. (2017). Death of family members as an overlooked source of racial disadvantage in the United States. *Proceedings of the National Academy of Sciences*, 114(5), 915–920. <https://doi.org/10.1073/pnas.1605599114>
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.*, 30(4), 377–399. <https://doi.org/10.1002/sim.4067>