Informative title: Exploring the optimal allostatic load scoring method in women of reproductive age

Running title: The optimal allostatic load scoring method

Yang LI, PhD [1], Marie-Anne S. ROSEMBERG, PhD, RN [2], Vanessa K. DALTON, MD, MPH [3], Shawna J. LEE, PhD, MSW, MPP [4], Julia S. SENG, PhD, CNM, FAAN [2]

[1] Sinclair School of Nursing, University of Missouri, Columbia, MO 65211, US

[2] School of Nursing, University of Michigan, Ann Arbor, MI 48104, US

[3] Department of Obstetrics and Gynecology Medical School, University of Michigan, Ann Arbor, MI 48109, US

[4] School of Social Work, University of Michigan, Ann Arbor, MI 48109, US

Correspondence: Yang Li, S449 Sinclair School of Nursing, University of Missouri, Columbia, MO 65211. Telephone number: 734-927-2688. E-mail: liy5@missouri.edu

Author contributions:

Criteria	Author Initials
Made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data;	YL, JSS
Involved in drafting the manuscript or revising it critically for important intellectual content;	YL, JSS
Given final approval of the version to be published. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content;	MASR, VKD, SJL, JSS
Agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.	YL, MASR, VKD, SJL, JSS

Declarations of interest: No conflict of interest has been declared by the authors.

Acknowledgments: The authors would like to thank the Consulting for Statistics, Computing and Analytics Research (CSCAR) consultants for their support on the statistical analyses of the study, especially the CSCAR director – Dr. Kerby Shedden.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DR YANG LI (Orcid ID : 0000-0001-8901-3454)

Article type : Original Research

Exploring the optimal allostatic load scoring method in women of reproductive age

Abstract

Aims: To determine the optimal allostatic load scoring method.

Design: This is a secondary analysis of data on women of reproductive age from the 2001-2006 National Health and Nutrition Examination Survey.

Methods: We created allostatic load summary scores using five scoring methods including the count-based, Z-Score, logistic regression, factor analysis and grade of membership methods. Then we examined the predictive performance of each allostatic load summary measure in relation to three outcomes: general health status, diabetes and hypertension.

Results: We found the allostatic load summary measure by the logistic regression method had the highest predictive validity with respect to the three outcomes. The logistic regression method performed significantly better than the count-based and grade of membership methods for predicting diabetes as well as performed significantly better for predicting hypertension than all of the other methods. But the 5 scoring methods performed similarly for predicting poor health status.

Conclusion: We recommended the logistic regression method when the outcome information is available, otherwise the frequently used, simpler count-based method may be a good alternative.

Impact: The study compared different scoring methods and made recommendations for the optimal scoring approach. We found allostatic load summary measure by the logistic regression method had the strongest predictive validity with respect to general health status, diabetes and hypertension. The study may provide empirical evidence for future research to use the

recommended scoring approach to score allostatic load. The allostatic load index may serve as an "early warning" indicator for health risk.

KEYWORDS: allostatic load, scoring, women of reproductive age, nursing

1 INTRODUCTION

Allostatic Load (AL) refers to the accumulated multi-system physiologic dysfunction resulting from repeated, chronic stress that could ultimately lead to disease (McEwen, 1998). When stress (e.g., socioeconomic disadvantage, child abuse and neglect) occurs, there is a cascade of effects that begins with primary stress mediators such as cortisol from the hypothalamic–pituitary–adrenal (HPA) axis, a primary effect, which in turn leads to secondary and tertiary outcomes (Beckie, 2012). The AL theory depicts how chronic stress leads to diseases. As a holistic measure of physiological dysfunction, AL may provide a multi-systemic approach to understand mechanisms involved in the impacts of chronic stress on health.

AL is operationalized by combining physiological indictors from multiple systems (i.e., neuroendocrine, immune, metabolic and cardiovascular) into one single index. The index is a more sophisticated, comprehensive physiological measure than a single system-specific indicator. It could reduce the probability of a type I error by combining multi-system indicators into one single index, rather than analyzing each individual indicator separately (McDade, 2008). However, there is no commonly accepted, gold-standard way to operationalize AL because of its multifaceted nature. Many scoring methods have been used to create an AL index (ALI) in previous studies, including the count-based, Z-Score, canonical correlation, recursive partitioning and grade of membership (GOM) method. Controversies or challenges regarding AL scoring methods primarily arise from three issues: technique for calculating the index, weighting of respective indicators in the index and norming on a population. Thus, the scoring issue must be further considered before the concept of AL can be reliably and validly applied to research and clinical practice.

1.1 Background

1.1.1 The count-based method

The most frequently used scoring method is the count-based method. The ALI by this method is the sum of the number of indicators for which individuals fall into the risk quartile of the sample distribution (Seeman, Singer, Rowe, Horwitz, & McEwen, 1997). It is simple to calculate the overall index using the count-based method but dichotomizing each individual indicator would lose information regarding the potential variability in their contributions to overall risk and might decrease the statistical power in analyses (Seeman et al., 2008). This method also has the limitation of making the ALI sample-specific by dichotomizing indicators based on the risk quartile of the sample distribution. For all AL indicators, no current population norms in terms of age, race, sex, etc. have ever been derived. Thus, the sample-specific summary measure may not be meaningfully compared across samples. Furthermore, all physiological indicators count equally in the summary score. The relative importance of various physiological components to the overall score for predicting health outcomes is not considered. Some indictors may be more critical than others with regard to certain outcomes.

1.1.2 The Z-Score method

Another relatively simple scoring approach is the Z-Score method. In this approach, all indicators are individually standardized to a mean of zero and a standard deviation of one. The ALI is the sum of the standardized distances of each indicator from its respective mean. The formulation is based on a continuous, rather than a categorical, function of the biological measures (Vie, Hufthammer, Holmen, Meland, & Breidablik, 2014). Compared with the count-based method, the Z-Scored ALI could account for some variances (Hampson et al., 2009). But it is still sample-specific and fails to account for the weighting of each indicator in the summary measure.

1.1.3 Canonical correlation, recursive partitioning and GOM

Some AL studies applied other scoring methods that are more complex than a simple count or a Z score, such as canonical correlations, recursive partitioning and GOM. These alternative scoring approaches provide more complex scoring algorithms and incorporate more information of each individual indicator than the simple counting of high-risk cut-off points. They also allow for unequal weighting of various biological measures (Beckie, 2012).

Canonical correlation has been used to determine the best linear combination of AL indicators that is maximally correlated with the best linear combination of health outcomes (Karlamangla, Singer, McEwen, Rowe, & Seeman, 2002). An AL summary score can be constructed using the sets of AL indicators and their canonical weights in the best linear correlation. This approach permits unequal weights for each AL indicator, but it requires continuous variables and relies on the subsequent outcome information. Since the canonical weights are derived from and applied to the same sample, it makes the ALI too specific to the data used to derive it. This may magnify the predictive ability of the index, deplete its predictive power in other contexts and cause the endogeneity bias, in other words may not be generalized to other contexts (Seplaki, Goldman, Glei, & Weinstein, 2005).

Recursive partitioning is a technique that has been used to classify individuals into outcome risk categories. It can identify multiple combinations of physiological indicators and their value ranges to best differentiate among outcomes across individuals (Juster, McEwen, & Lupien, 2010). It can also be used to define AL categories (e.g., high, intermediate, low). Similar to the canonical correlation, this approach has the limitation of incorporating information on subsequent health outcomes (Seplaki et al., 2005).

The GOM method has been used to create N pre-defined pure profiles, which are the collections of response probabilities corresponding to each level of discrete indicators. Accordingly, N GOM (summing to one) scores are assigned to each individual, measuring the similarity of the set of a person's indicator values to each respective profile. The GOM score-based ALI is the sum of N-1 of the GOM scores (excluding the score for the reference/low risk profile), measuring dissimilarity to the low risk profile (Seplaki, Goldman, Weinstein, & Lin, 2006). The method does not incorporate information on subsequent health outcomes, but still categorizes each indicator into low, moderate, or high levels based on the sample distribution.

1.1.4 Factor analysis and multivariable logistic regression

Three prior studies used factor analysis to construct and evaluate structural models of AL reflecting the cumulative physiological burden across multiple systems (Booth, Starr, & Deary, 2013; Kubzansky, Kawachi, & Sparrow, 1999; Seeman et al., 2010). Parameter estimates obtained from factor analysis can be considered as the specific contributions of respective indicators to the summary score. Studies on creating other clinical index measures used some

other statistical techniques such as the multivariable logistic regression (Hughes et al., 2012; Lee, Lindquist, Segal, & Covinsky, 2006). The multivariable logistic regressions are fitted with all potential components as predictors and outcomes as response variables. Coefficients obtained from the regression models can be considered as weights for each component. Scores are allocated to each component based on those weights and summed up to a total index. But to our knowledge, no previous studies have used the factor analysis or logistic regression method to assign weights to each AL indicator.

1.1.5 Research gaps

Although previous studies used different scoring methods to create an ALI, there is not yet a gold-standard measure of AL that is valid across health outcomes. No studies have focused on comparing different scoring methods and determining the optimal AL scoring method, which represents a gap in the current research on chronic stress and health outcomes. Thus, how best to incorporate multiple physiological indicators into one single summary measure needs to be addressed. More research is needed to compare the predictive validities of different scoring methods and to determine which method is optimal to score AL before examining AL as a mediating pathway for the impact of chronic stress on health outcomes.

2 THE STUDY

2.1 Aims

This study aimed to determine the optimal AL scoring method by comparing several scoring methods within a single population dataset. Because age and gender would influence the AL summary score, the study focused on a more homogeneous female population – women of reproductive age from the 2001-2006 National Health and Nutrition Examination Survey (NHANES) database. We constructed the ALI using five scoring methods including the count-based, Z-Score, logistic regression, factor analysis and GOM methods. We then examined the predictive performance of each ALI with women of reproductive age in relation to 3 outcomes: self-reported general health status, diabetes and hypertension.

2.2 Design

This is a secondary analysis of data from the NHANES. NHANES is a cross-sectional study with a complex, multistage probability sampling design used to select a sample representative of the civilian non-institutionalized resident population of the United States, which has been conducted in 2-year cycles since 1999 (Curtin et al., 2012). In this study, we used data from the 2001 to 2006 cycles of NHANES to test the study aims. Data from 2007-2010 were used to replicate the main analyses and compare the results with the 2001-2006 data to evaluate the stability of the results. The data collected between 1999-2000 were not used because general health status was not queried during that 2-year cycle. Data collected after 2010 was not used because no C-reactive protein (CRP) has been measure since 2011.

2.3 Participants

Female participants with reproductive ages of 15-49 were included in the study. Women who were pregnant at the exam measured by the urine pregnancy test were excluded. A total of 5525 women were eligible for the study in the 2001-2006 NHANES data. But 1206 women (21.8%) had missing data on the three outcome variables (general health status, diabetes and hypertension). Thus, 4319 women were finally included for analysis to address the study aims. In the 2007-2010 NHANES data, a total of 3018 women were included to replicate the main analyses.

2.4 Variables and data sources

2.4.1 AL

The selected 10 indicators in this study were CRP, systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse, body mass index (BMI), total cholesterol (TC), high-density lipoprotein (HDL), triglycerides, glycohemoglobin and glucose. These indicators were frequently used in previous studies (Juster et al., 2010). Other indicators such as low-density lipoprotein, glucose, insulin, C-peptide and fibrinogen in the NHANES database were not included in the study because there is a large amount of missing data or some of those indicators were collected only in subsamples. Standard examination and laboratory procedures were described in the NHANES Examination and Laboratory Protocols (CDC & NCHS).

2.4.2 Outcomes

General health status was measured using 1 question asking whether participants' general health is excellent, very good, good, fair, or poor from the current health status questionnaire. In this study, it was dichotomized into two levels: "poor" and "fair, good, very good, or excellent". We used 1 item— "Other than during pregnancy, have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes?" from the diabetes questionnaire to determine diabetes being present or not. Participants who reported "Borderline" were considered as no diabetes. The question—"Have you ever been told by a doctor or other health professional that you had hypertension, also called high blood pressure?" from the blood pressure & cholesterol questionnaire was used to determine hypertension present or not.

2.4.3 Sociodemographic characteristics

Age, race, poverty income ratio (PIR), education and marital status from the demographic dataset were included in this study. We dichotomized race into two categories: non-Hispanic black and other races (e.g., Mexican American, other Hispanic, non-Hispanic white and others including multi-racial). PIR is an index for the ratio of family income to poverty threshold, ranging between 0-5.00. Education level was categorized into: Less than high school, high school diploma including GED and more than high school. Marital status was recoded as married/living with partner and widowed/divorced/separated/never married.

2.5 AL Scoring methods

The count-based, Z-Score, logistic regression, factor analysis and GOM methods were used to construct the ALI in this study. None of the five scoring methods incorporate outcome information in the calculation of the summary measure except the logistic regression method. The logistic regression, factor analysis and GOM methods considered the weighting issue. In this study, the canonical correlation approach was excluded because it requires continuous variables including outcome variables, while the outcome variables available in the NHANES database are categorical. The recursive partitioning technique was also not used because only AL categories (e.g., high, intermediate, low) can be defined and no total score can be constructed through this approach.

Among the 10 indicators, glucose and glycohemoglobin are direct clinical indicators for the diagnosis of diabetes and SBP and DBP are directly related with the diagnosis of

hypertension. An issue that arises is whether the associations between ALI and diabetes or hypertension reflect only or largely the impact of the 4 indicators or whether the other indicators have significant and independent relationships with these two outcomes. Thus, using the five scoring approaches, we also constructed tailored ALI without glucose and glycohemoglobin predicting diabetes and without SBP and DBP predicting hypertension. Results based on tailored ALI (8 indicators) were compared with the results of ALI (10 indicators) in a sensitivity analysis.

2.5.1 The Count-based method

A dichotomous high-risk score was computed for each indicator by assigning a score of 1 to participants whose scores were in the top risk quartile of the sample distribution (75th percentile for all indicators except HDL for which 25th percentile corresponds to high risk) and a score of 0 otherwise. An ALI was then constructed as the sum of the 10 dichotomous (0/1) indicator risk scores, yielding a possible score range of 0-10.

2.5.2 The Z-Score method

All 10 indicators were individually standardized to a mean of zero and a standard deviation of one. The HDL Z-Score was reversed so that high values reflect greater dysregulation. An ALI was then calculated by summing the Z-Scores of all indicators.

2.5.3 The logistic regression method

Multivariable logistic regressions were conducted with all 10 AL indicators as explanatory variables and the 3 outcome variables (i.e., general health status, diabetes and hypertension) as the response variable respectively. The standardized coefficients obtained from the models were used as the weights for each individual indicator. Indicators were first individually standardized to a mean of zero and a standard deviation of one. The Z-Scores were then multiplied by the coefficients for each individual indicator derived from the regression models. Using this method with the 3 outcome variables as the response variables respectively, 3 ALI were computed by summing the multiplied values for each indicator.

2.5.4 The factor analysis

We conducted the factor analysis using robust maximum likelihood estimation with the number of factors set as 1. Indicators were first individually standardized to a mean of zero and a standard deviation of one. The Z-Scores were multiplied by the factor loading for each individual indicator derived from the factor analysis. We then created the summation scores for ALI.

2.5.5 The GOM

Each indicator was divided into low and high risk for poor health based on the 75th percentile of the sample distribution except HDL for which 25th percentile was the risk quartile. The number of pure-type profiles was set in advance. Each pure-type profile is a collection of response probabilities corresponding to each level of the 10 discrete indicators. Our analyses showed that compared with 3 and 4 pure types, 5 pure-type profiles provide reasonable interpretability and summaries of the physiological functions. Detailed definitions for the 5 pure types can be seen in Supplemental Figures 1A and 1B. Accordingly, a set of 5 GOM scores for each individual that quantify the individual's similarity to each pure-type profile was created, ranging from 0-1 and summing to unity. Excluding the score measuring similarity to the low-risk, or reference, pure-type profile (the 5th profile), the other 4 GOM scores were summed to create a single GOM-based AL summary measure, reflecting dissimilarity to the low-risk profile. Detailed explanations for the GOM method can be found in previous studies (Seplaki et al., 2005; Seplaki, Goldman, Weinstein, & Lin, 2004; Seplaki et al., 2006).

2.6 Ethical considerations

The NHANES 2001–2010 were approved by the National Center for Health Statistics Research Ethics Review Board under protocols #98-12 and #2005-06 and Continuation of Protocol #2005-06. This secondary analysis of data was exempt from IRB review because it was done via the de-identified dataset.

2.7 Data analysis

Means, standard deviations, 25th/75th percentiles, frequencies and percentages were used to describe sociodemographic characteristics, the 3 outcome variables and the 10 AL indicators. The multiple imputation (MI) method (Rubin, 2004) was used to impute all missing data. We used chained equations and predictive mean matching with non-missing sociodemographic

variables and indicators as predictor variables. The imputations of the missing values are predicted values from these regression models, with the appropriate random error included. Since there is 17.6% of data missing, 10 imputed datasets were created. In each of the imputed datasets, we conducted all main analyses including constructing the ALI with different scoring approaches and validating the index. The overall estimate is the average of the estimates from each of the imputed datasets.

The distributional qualities, including range, mean, standard deviation, median, skew and kurtosis, were used to describe AL summary measures by each of the 5 scoring methods. The odds ratio (OR) by each method was computed through fitting binomial logistic regression models to estimate the strengths of the associations of each AL summary measure with general health status, diabetes and hypertension respectively. The three outcomes were included as the response variables respectively and each summary measure of AL was included as the explanatory variable. The covariates included age, race and PIR. All ALI scores by the five methods were standardized to a mean of zero and a standard deviation of one before fitting the regression models, so that the strengths of the (adjusted) associations between AL summary measures and outcomes can be compared across different scoring approaches. Additionally, the areas under the receiver operating characteristic (ROC) curve (AUC) were calculated to estimate the predictive validity of each AL summary measure for predicting the 3 outcomes. An AUC with successively higher values above 0.5 indicates increasing levels of predictive value (Hanley & McNeil, 1982).

To investigate the performance of different AL measures in an external sample, the process was subsequently repeated, conducting the same analyses in the NHANES 2007-2010 dataset. To make a recommendation of the optimal scoring method for clinical use purposes, we also evaluated each scoring method by qualitative comparisons in terms of strengths and weaknesses. Using the optimal scoring method, we calculated the cut-off points, sensitivities and specificities. All statistical analyses were performed using R Software Version 3.4.2 (R Core Team, 2017).

2.8 Validity and reliability

Data are collected and processed with standardized procedures and protocols developed and validated by the National Center for Health Statistics (NCHS) for all household interview,

clinical examinations and laboratory tests. This helps to assure that the data for this analysis are of high quality in terms of validity and reliability.

3 RRSULTS

3.1 The sample characteristics

The mean age of the sample was approximately 30 years and about 26% of women were non-Hispanic Black. Around 58% reported completing high school or less than high school education and 56.5% were married or living with partner. Only 1.7% reported poor health status, 3.1% had diabetes and 12.6% had hypertension (Table 1). Table 2 showed the descriptive statistics of each AL indicator.

3.2 The descriptive statistics of ALI

The ALIs constructed by the count-based and GOM method did not consist of negative values, while the ALIs ranged from a negative value to a positive value for the other 3 methods. The skew and kurtosis of ALIs using the count-based method, the logistic regression with general health and diabetes as the outcome and the GOM method were close to 0, indicating these indices are more normally distributed (Table 3). The skew and kurtosis of the tailored ALI using the count-based measure, the tailored ALI without glucose and glycohemoglobin using the logistic regression and the tailored ALI without SBP and DBP using the GOM were less than 1, suggesting the distributions of those indices were more normal (Supplemental Table 1). All distributions were unimodal except for the tailored ALI without glucose and glycohemoglobin using the GOM method (Supplemental Figure 2). Interestingly, the tailored ALI without glucose and glycohemoglobin using the GOM method presented a bimodal distribution with two peaks close to 0 and 1 respectively, which visually showed the cut-off point of the ALI for poor health risk (Supplemental Figure 3).

3.3 The predictive validities of ALI

The logistic regression method was most strongly associated with the 3 outcome measures, whether adjusted or not adjusted (Table 4). This remained the case when 2 indicators diagnostic for diabetes or hypertension were removed from the index (Supplemental Table 2). Using the factor analysis method, the associations of ALI with general health and hypertension were

smallest (OR=1.43, 95% CI=1.29-1.59; OR=1.84, 95% CI=1.67-2.03) and significantly smaller than the logistic regression method (OR=2.26, 95% CI=1.87-2.73; OR=2.88, 95% CI=2.60-3.19). But there were no significant differences in terms of the strengths of the associations among the count-based, Z-Score, logistic regression and GOM methods. The count-based measure was nearly as strongly related to the outcome measures as the logistic regression, adjusted or unadjusted, tailored or not. As expected, all ALIs with 10 indicators were more strongly associated with diabetes and hypertension compared with the tailored ALI without glucose and glycohemoglobin and the tailored ALI without SBP and DBP.

The 5 scoring methods had similar predictive performances with regard to general health (AUC=0.72-0.75). But the logistic regression method (AUC=0.92, 95% CI=0.88-0.95) had better predictive performance for predicting diabetes compared with the count-based (AUC=0.83, 95% CI=0.79-0.87) and GOM (AUC=0.82, 95% CI=0.78-0.86) methods and had the best performance for predicting hypertension (AUC=0.79, 95% CI=0.77-0.81) than the other 4 methods. The ALI by any method predicted diabetes and hypertension better than it predicted the subjective appraisal of overall health status (Table 5, Figure 1 and Supplemental Figure 5). The tailored ALI (excluding glucose and glycohemoglobin or SBP and DBP) by any methods had similar predictive validities in terms of diabetes and hypertension except that the logistic regression method predicted hypertension better than the GOM method. As expected, the tailored ALI by any method had worse predictive powers compared with the ALI with all 10 indicators included (Supplemental Table 3 and Supplemental Figures 4 & 6).

3.4 Parallel analyses

All analyses were conducted again using the NHANES 2007-2010 data, yielding approximately the same results. Similarly, the logistic regression method had the strongest associations with the outcome measures, whether adjusted or not adjusted, tailored or not. The count-based method was nearly as strongly associated with the outcome measures as the logistic regression, adjusted or unadjusted, tailored or not. The five scoring methods had similar predictive validities with regard to the three outcome measures. Similarly, the logistic regression method still had the best predictive performances, whether tailored or not.

4 DISCUSSION

This study constructed an ALI using five scoring approaches and assessed the predictive performances across different scoring approaches in women of reproductive age. We found the AL summary measure by the logistic regression method had the strongest predictive validity with respect to general health status, diabetes and hypertension. The logistic regression method performed significantly better than the count-based and GOM methods for predicting diabetes as well as performed significantly better for predicting hypertension than all of the other methods. But the 5 scoring methods performed similarly for predicting poor health status. Excluding the diagnostic indicators for diabetes and hypertension, the independent contributions of the other 8 indicators to the risk of diabetes and hypertension were demonstrated. Differences in the predictive performances in terms of diabetes and hypertension became smaller among the five scoring methods, but the logistic regression method still performed the best. The findings were duplicated using the 2007-2010 NHANES data, underscoring the robustness of the finding.

The predictive performances across different scoring methods in this study are similar, which is partially consistent with a study using data from a population-based sample of older Taiwanese to compare several count-based formulations as well as the Z-Score and GOM methods. All AL summary measures had similar predictive performances for predicting self-assessed health, impairments in activities of daily living and mobility, cognitive performance and depressive symptoms. The study recommended the count-based and Z-Score measures since the two methods are simple to compute and the GOM method is more complicated (Seplaki et al., 2005). Another study with a community sample of 470 participants from the Hawaii Personality and Health cohort also reported similar performances of the count-based and Z-Score methods for predicting self-rated health (Hampson et al., 2009). The differences among the 5 summary measures were not pronounced in this study, suggesting that the advantages of one method over another are relatively subtle.

The differences in the predictive performances between the logistic regression method and the other scoring methods for predicting diabetes and hypertension were larger than for predicting poor health status. In addition, the differences became smaller after excluding the four diagnostic indicators (glucose, glycohemoglobin, SBP and DBP) for diabetes and hypertension. Given that the logistic regression method accounts for the non-uniform contributions of distinct biological measures to health risk, the possible explanation for this finding is that large weightings were assigned to the 4 diagnostic indicators by the logistic regression method. The

finding suggests that the logistic regression method predicts better when some AL components have much stronger associations with specific health outcomes than the other AL components.

Each scoring approach has its own strengths and weaknesses (Table 6). The ALI by the logistic regression method had the best predictive performance compared with the other methods. But this method assigns scoring weights to each indicator based on information on subsequent outcomes. It is challenging to compare AL summary scores across different outcomes. And the logistic regression method may not be the optimal scoring method when the outcome information is unknown. For example, in the preliminary stage of a research project, only data on physiological indicators is available while data on the targeted outcome has not been collected. Also, the outcome is not needed for some studies that only focus on exploring some stressors in relation to AL levels.

Under the above conditions, the count-based method may be a good alternative. The predictive performance of the ALI by the count-based method for predicting general health status is similar to the other approaches and even for predicting diabetes and hypertension is similar to the other approaches except the logistic regression method. Additionally, after excluding the diagnostic indicators for diabetes and hypertension, the count-based method performed as well as the logistic regression method for predicting diabetes and hypertension. Compared with the other methods, the count-based method has its own strengths. It is the most frequently used method in prior AL studies. The AL summary score by this method is the number of indicators of risk for poor health, which is a real value and easy to interpret. It is simple to calculate, easy to understand and feasible to be applied in clinical practice. Therefore, if the outcome information is available, needed and consistent across different contexts, we recommended the logistic regression method; otherwise the count-based method may be a good alternative from the perspectives of predictive validity, feasibility and interpretability.

Using the count-based method, we calculated cut-off points, sensitivities and specificities of the ALI score (Supplemental Table 4). Although the count-based method had advantages in terms of predictive validity, feasibility and interpretability, it has the limitation of making the ALI sample-specific. A better way to address the limitation is to use the clinical risk cut-off points based on national standards instead of risk quartiles of the sample distribution to count the total number of indicators of risk for poor health. But further work on establishing population norms in terms of age, sex, race, etc. is needed. Especially, no current population norms for

pregnancy have ever been derived, which make it challenging to apply the AL theory to perinatal outcomes research.

4.1 Limitations

This study had some limitations. First, we focused on women of reproductive age. Because the dysregulated levels of each AL indicator are different in terms of age and gender, scoring AL in a more homogeneous female population may contribute to the reliability of our findings. But our findings may not be generalized to the male or elder population. Future research needs to replicate our analyses in different age- and gender-specific populations. Age- and gender-specific population norms for the ALI score by the optimal scoring method will be also needed. Second, data on indicators from the primary mediating neuroendocrine system are lacking in the NHANES database. The ALI was constructed without indicators from the neuroendocrine system, relying solely in the indicators of secondary dysregulations for the scoring, which may decrease the predictive validity and explanatory power of the total score on health outcomes. Third, because of the cross-sectional study design of the NHANES, data on the outcome variables and AL indicators were collected at the same time. This may also affect the predictive performances of ALI for predicting general health status, diabetes and hypertension. A prospective study using a full complement of physiological indicators to operationalize the AL and using different scoring approaches is needed to validate the recommendations made based on this secondary analysis. Lastly, since no commonly-accepted set of physiological indicators has been determined, incorporating different sets of physiological indicators into the ALI may influence which scoring method is optimal for use. Future research with different sets of physiological indicators is needed to validate our recommendations.

5 CONCLUSION

Our study advanced studies of AL by focusing on scoring methods with a nationally representative dataset and making recommendations for the optimal method to score AL. It provides empirical evidence for researchers to use the recommended scoring approach to score AL in their research. Our findings may also be useful for clinicians. The ALI can serve as a sign for risk of subclinical syndromes. Most of AL indicators such as BMI, blood pressure and pulse are routine clinical assessments and thus are feasible to be measured. The logistic regression

method can be used through computer software and the count-based ALI as an alternative measure can be easily calculated by hand. Therefore, the AL summary measure is easy and feasible for use as an "early warning" indicator for health risk across a variety of care settings.

REFERENCES

Beckie, T. M. (2012). A systematic review of allostatic load, health and health disparities. Biological Research for Nursing, 14(4), 311-346.

Booth, T., Starr, J. M., & Deary, I. (2013). Modeling multisystem biological risk in later life: allostatic load in the Lothian birth cohort study 1936. American Journal of Human Biology, 25(4), 538-543. doi:http://dx.doi.org/10.1002/ajhb.22406

Curtin, L. R., Mohadjer, L. K., Dohrmann, S. M., Montaquila, J. M., Kruszan-Moran, D., Mirel, L. B., . . . Johnson, C. L. (2012). The National Health and Nutrition Examination Survey: Sample Design, 1999-2006. Vital and Health Statistics, 2(155), 1-39.

Hampson, S. E., Goldberg, L. R., Vogt, T. M., Hillier, T. A., & Dubanoski, J. P. (2009). Using Physiological Dysregulation to Assess Global Health Status Associations with Self-rated Health and Health Behaviors. Journal of Health Psychology, 14(2), 232-241. doi:10.1177/1359105308100207

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36.

Hughes, D. A., Malmenas, M., Deegan, P. B., Elliott, P. M., Ginsberg, L., Hajioff, D., . . . Investigators, F. O. S. (2012). Fabry International Prognostic Index: a predictive severity score for Anderson-Fabry disease. Journal of Medical Genetics, 49(3), 212-220. doi:10.1136/jmedgenet-2011-100407

Juster, R. P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. Neuroscience & Biobehavioral Reviews, 35(1), 2-16. doi:10.1016/j.neubiorev.2009.10.002

Karlamangla, A. S., Singer, B. H., McEwen, B. S., Rowe, J. W., & Seeman, T. E. (2002). Allostatic load as a predictor of functional decline. MacArthur studies of successful aging. Journal of Clinical Epidemiology, 55(7), 696-710.

Kubzansky, L. D., Kawachi, I., & Sparrow, D. (1999). Socioeconomic status, hostility and risk factor clustering in the Normative Aging Study: any help from the concept of allostatic load? Annals of Behavioral Medicine, 21(4), 330-338.

Lee, S. J., Lindquist, K., Segal, M. R., & Covinsky, K. E. (2006). Development and validation of a prognostic index for 4-year mortality in older adults. JAMA, 295(7), 801-808.

McDade, T. W. (2008). Challenges and opportunities for integrative health research in the context of culture: A commentary on Gersten. Social Science & Medicine, 66(3), 520-524. doi:10.1016/j.socscimed.2007.09.005

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. The New England Journal of Medicine, 338(3), 171-179. doi:10.1056/NEJM199801153380307

Centers for Disease Control and Prevention, National Center for Health Statistics. National Health and Nutrition Examination and Laboratory Protocols [1999–2006]. Retrieved from http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm

Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys (Vol. 81): John Wiley & Sons.

Seeman, T., Gruenewald, T., Karlamangla, A., Sidney, S., Liu, K., McEwen, B., & Schwartz, J. (2010). Modeling multisystem biological risk in young adults: The Coronary Artery Risk Development in Young Adults Study. American Journal of Human Biology, 22(4), 463-472. doi:http://dx.doi.org/10.1002/ajhb.21018

Seeman, T., Merkin, S. S., Crimmins, E., Koretz, B., Charette, S., & Karlamangla, A. (2008). Education, income and ethnic differences in cumulative biological risk profiles in a national sample of US adults: NHANES III (1988-1994). Social Science & Medicine, 66(1), 72-87. doi:10.1016/j.socscimed.2007.08.027

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.

Seplaki, C. L., Goldman, N., Glei, D., & Weinstein, M. (2005). A comparative analysis of measurement approaches for physiological dysregulation in an older population. Experimental Gerontology, 40(5), 438-449. doi:10.1016/j.exger.2005.03.002

Seplaki, C. L., Goldman, N., Weinstein, M., & Lin, Y. H. (2004). How are biomarkers related to physical and mental well-being? Journals of Gerontology Series A-Biological Sciences & Medical Sciences, 59(3), 201-217.

Seplaki, C. L., Goldman, N., Weinstein, M., & Lin, Y. H. (2006). Measurement of cumulative physiological dysregulation in an older population. Demography, 43(1), 165-183.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. .

Vie, T. L., Hufthammer, K. O., Holmen, T. L., Meland, E., & Breidablik, H. J. (2014). Is self-rated health a stable and predictive factor for allostatic load in early adulthood? Findings from the Nord Trondelag Health Study (HUNT). Social Science & Medicine, 117, 1-9. doi:10.1016/j.socscimed.2014.07.019

TABLE 1 The descriptive statistics of sample sociodemographics and health outcomes (N=4319)

	N	M (SD)/%
Age	4319	29.58 (10.76)
Poverty income ratio	4112	2.40 (1.64)
Race	4319	
Mexican American	1038	24.03
Other Hispanic	179	4.14
Non-Hispanic White	1786	41.35
Non-Hispanic Black	1130	26.16
Other Race - Including Multi-Racial	186	4.31
Education level	4317	
Less than high school	1609	37.27
High school diploma including GED	897	20.78
More than high school	1811	41.95
Marital status	4318	
Married/living with partner	2439	56.48
Widowed/divorced/separated/never married	1879	43.52
General health	4319	
Excellent	484	11.21
Very good	1427	33.04
Good	1667	38.60
Fair	666	15.42
Poor	75	1.7
Diabetes	4319	
Yes	133	3.08
No	4186	96.92
Hypertension	4319	
Yes	546	12.64
No	3773	87.36

TABLE 2 The descriptive statistics of the 10 allostatic load indicators (N=4319)

	N	M (SD)	Percent 25th	Percent 75th
Pulse, beat per min	4225	76.13 (11.57)	68.0	84.0
SBP, mmHg	4183	111.87 (13.19)	103.0	118.0
DBP, mmHg	4023	67.78 (10.87)	61.0	75.0
BMI	4263	27.52 (7.48)	21.97	31.74
TC, mg/dL	4060	183.14 (38.19)	156.0	205.0
HDL, mg/dL	4060	55.67 (14.66)	45.0	64.0
CRP, mg/dL	4089	0.42 (0.78)	0.05	0.48
Glycohemoglobin, %	4116	5.27 (0.69)	5.0	5.4
Glucose, mg/dL	4056	88.39 (20.48)	80.0	91.0
Triglycerides, mg/dL	4056	104.81 (85.08)	58.0	126.0

Note. SBP: systolic blood pressure; DBP: diastolic blood pressure; BMI: body mass index; TC: total cholesterol; HDL: high-density lipoprotein; CRP: C-reactive protein.

TABLE 3 The descriptive statistics of allostatic load indices using the 5 scoring methods

	M (SD)	Median	Min-Max	Skew	Kurtosis
Count-based method	2.35 (2.03)	2	0-10	0.91	0.31
Z-Score method	0 (4.89)	-0.92	-10.53-38.11	1.62	5.52
Logistic regression					
General health as the outcome	0 (0.81)	-0.11	-2.48-4.36	0.76	1.08
Diabetes as the outcome	0 (1.41)	-0.23	-5.36-17.00	4.33	32.67
Hypertension as the outcome	0 (1.06)	-0.16	-4.28-5.56	0.86	1.44
Factor analysis	0 (0.62)	-0.12	-1.03-7.46	4.47	32.88
Grade of membership	0.30 (0.29)	0.22	0.02-0.94	0.71	-0.74

TABLE 4 The binary logistic regressions of allostatic load indices by the 5 scoring methods on general health, diabetes, and hypertension

	General Health		Diabetes		Hypertension	
	OR (95% CI)	Adjusted OR (95% CI)	OR (95% CI)	Adjusted OR (95% CI)	OR (95% CI)	Adjusted OR (95% CI)
Count-based method	2.11 (1.74-2.55)	1.68 (1.35-2.10)	3.15 (2.69-3.70)	2.67 (2.24-3.19)	2.32 (2.12-2.54)	1.90 (1.72-2.10)
Z-Score method	1.84 (1.59-2.14)	1.53 (1.29-1.82)	3.42 (2.91-4.03)	3.12 (2.62-3.71)	2.19 (1.99-2.41)	1.81 (1.64-2.00)
Logistic regression	2.26 (1.87-2.73)	1.86 (1.50-2.30)	4.10 (3.41-4.92)	3.68 (3.07-4.43)	2.88 (2.60-3.19)	2.34 (2.09-2.62)
Factor analysis	1.43 (1.29-1.59)	1.25 (1.11-1.42)	3.61 (3.05-4.29)	3.27 (2.75-3.90)	1.84 (1.67-2.03)	1.48 (1.35-1.62)
Grade of membership	2.06 (1.66-2.57)	1.62 (1.28-2.05)	3.27 (2.69-3.98)	2.71 (2.21-3.33)	2.04 (1.86-2.23)	1.71 (1.55-1.89)

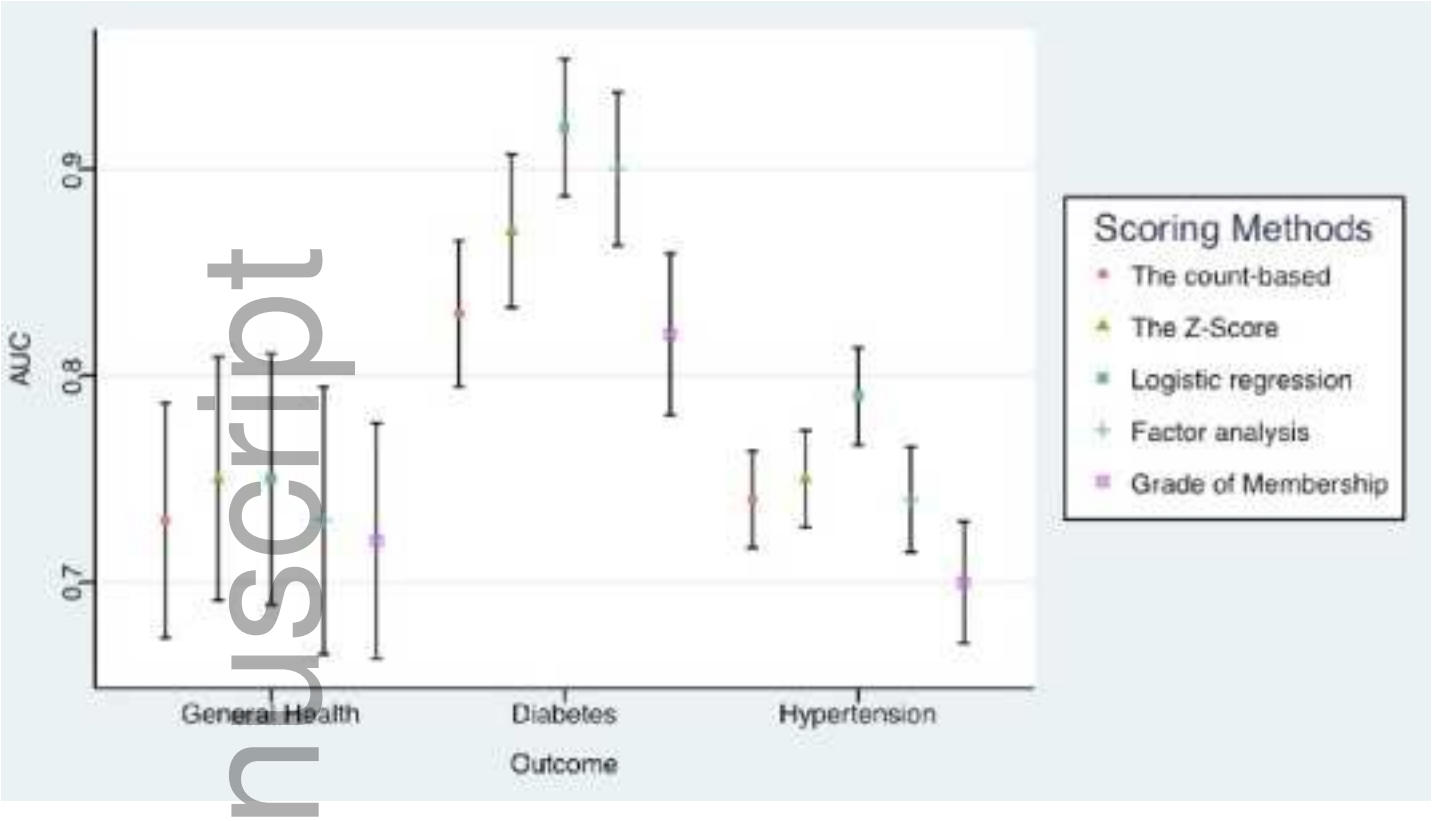
Note. Age, race, and poverty income ratio were adjusted for.

TABLE 5 The area under the ROC curve of allostatic load indices by the 5 scoring methods

	General Health		Diabetes		Hypertension	
	AUC	95% CI	AUC	95% CI	AUC	95% CI
Count-based method	0.73	0.68-0.79	0.83	0.79-0.87	0.74	0.72-0.77
Z-Score method	0.75	0.69-0.81	0.87	0.83-0.91	0.75	0.73-0.77
Logistic regression	0.75	0.69-0.81	0.92	0.88-0.95	0.79	0.77-0.81
Factor analysis	0.73	0.67-0.80	0.90	0.86-0.94	0.74	0.72-0.77
Grade of membership	0.72	0.66-0.78	0.82	0.78-0.86	0.70	0.67-0.73

TABLE 6 Evaluations of the 5 scoring methods

Scoring Methods	Strengths	Weaknesses
The count-based method	• Simple; • Most frequently used; • Use natural units (i.e., number of indicators within high risk quartiles).	• Discretizing variables loses information regarding the potential variability in their contribution in relation to overall risk; • Fails to consider the unequal weights of each indicator in the index.
The Z-Score method	• Simple; • The continuous function of biological measures makes maximal use of available variance.	• Fails to consider the unequal weights of each indicator in the index; • More difficult interpretation due to standardization and loss of natural units.
Logistic regression	• Allows for unequal weights for each indicator.	• Incorporates information on subsequent outcomes; • No prior AL studies have used it to assign weights to each indicator.
Factor analysis	• Allows for unequal weights for each indicator; • Does not incorporate information on subsequent outcomes.	• The number of factors could be subjectively determined if not set at 1 a priori; • No prior AL studies have used it to assign weights to each indicator.
Grade of membership	• Allows for unequal weights for each indicator; • Does not incorporate information on subsequent outcomes.	• The number of pure-type profiles is subjectively determined; • The method is challenging to produce.



jan_14014_f1.jpg