Application of a Classification Method for Studies of Allostatic Load

Sun Y. Jeon and Eric N. Reither

Department of Sociology, Utah State University, Logan, UT, USA

Abstract

Since initially studied in the MacArthur Study of Successful Aging, health researchers have used the concept of allostatic load (AL) to help understand the consequences of repeated psychological stress on physical health. In the majority of empirical studies, AL is operationalized through a 4-step procedure: (1) attain 8-15 biomarkers for the AL construct from a given dataset; (2) dichotomize each biomarker at the upper quartile (75^{th} percentile), assuming that 25% of the population is exposed to the risk factor; (3) assign one point for each biomarker that lies beyond the risk threshold; (4) sum the points across all biomarkers to acquire the total AL score. In this paper, using the Random Forest classification model, we show that this standard operationalization of AL can produce misleading AL scores through the arbitrary selection of particular biomarkers and risk threshold settings.

Keywords. Allostatic Load, Classification Tree, Random Forest

1 Introduction

Allostatic load (AL) is a useful concept to understand the consequences of chronic and/or repeated psychological stress on physical health. When an individual perceives external stressors, the body attempts to achieve stability (i.e., homeostasis) through various adaptation processes (i.e., allostasis). For example, the human body activates the sympathetic adrenal medullary system and the hypothalamic pituitary adrenocortical (HPA) axis in response to perceived stressors. This adaptation process is functional in the short term, but when it occurs repeatedly over long periods of time, the body loses its ability to return to its previous normal state (e.g., with respect to blood pressure). The trending of biomarkers into unhealthful ranges over the life course reflects the accumulation of AL. As shown in Table 1, indicators of elevated AL include both primary and secondary mediators and may eventually lead to chronic diseases such as cancer, cardiovascular diseases, brain or cognitive dysfunction, and immune deficiencies [1].

Initially developed through the MacArthur Study of Successful Aging in the 1990s, the concept of AL has been adopted by various studies using several different sources of data. Generally speaking, AL is operationalized through a 4-step process [2, 3, 4, 5]: (1) attain 8-15 biomarkers for the AL construct from a given dataset; (2) dichotomize each biomarker at the upper quartile (75^{th} percentile), assuming that 25% of the population is exposed to the risk factor; (3) assign one point for each biomarker that lies beyond the risk threshold; (4) sum the points across all biomarkers to acquire the total AL score.

In this paper, we demonstrate that the predominant operationalization of AL can produce misleading AL scores because biomarkers and risk thresholds are established through an arbitrary process. We further show that future research involving the AL construct should more carefully weigh which biomarkers and AL outcomes are included in the modeling process. Our critique of the main approach to measuring AL is based upon our analyses of this construct that rely extensively upon Random Forest methods of classification.

Table 1: Stages of AL

Primary Stage	Secondary Stage	Tertiary Stage
Meditors	**Outcomes**	**Outcomes**
DHEA-S	heart disease	Cancer
cortisol	Immune problem (flu/pneumonia)	Elevated mortality
Epinephrine	Cognitive decline	Other diseases
	Measures	
	systolic & diastolic BP	
	HDL & total cholesterol	
	glyc-hemoglobin	
	waist-hip ratio	

Note: Primary mediators refer to *"Chemical Messengers That Are Released as Part of Allostasis"*[1].

2 Method

2.1 Data & Measures

We used National Health and Nutrition Examination Survey (NHANES) 2005-2006. NHANES is the dataset that is frequently used in the recent AL studies since it provides information on various biomarkers and health outcomes that are measured not only by questionnaire but also by laboratory and examination.

We used data from the National Health and Nutrition Examination Survey (NHANES) 2005-2006. NHANES has been used frequently in recent AL studies because it provides information on a variety of important biomarkers and health outcomes that are measured not only by survey questionnaires but also by laboratory tests and physical examinations.

When AL was initially studied in the MacArthur Studies of Successful Aging, the data provided information on 10 biomarkers: DHEA-S, urinary cortisol excretion, urinary norepinephrine and epinephrine excretion levels, systolic and diastolic blood pressure (BP), waist-to-hip ratio, serum high-density lipoprotein (HDL) and total cholesterol, and glycosylated hemoglobin. While these 10 biomarkers provided a range of measures across various physical systems (e.g., cardio-respiratory function and metabolic status), there was no clear rationale for the inclusion and equal weighting of these specific biomarkers. Moreover, since the initial study, researchers have either excluded or replaced certain biomarkers, either by choice or necessity (e.g., due to variable availability in particular sources of data).

After reviewing the biomarkers that have been adopted by previous studies of AL [2, 3, 4, 5], we encompassed as many biomarkers as the NHANES 2005-2006 survey would allow. The 11 biomarkers used in our analysis were systolic and diastolic blood pressure, homocystein, c-reactive protein, albumin, glycosylated hemoglobin, HDL, ratio of HDL-to-total cholesterol, BMI, creatinine clearance, and triglycerides. Creatinine clearance was calculated using the Cockroft and Gault equation from the raw creatinine scores.

2.2 Random Forest Classification

This classification method is based on a binary decision process [6]. Figure 1 shows an example of a classification tree, fitted for NHANES 2005-2006 data on cancer. The response variable is the absence (group=0) or presence (group=1) of cancer. Predictor variables were the 11 AL biomarkers mentioned previously. At each node of the tree, every possible split on these biomarkers is considered; the split resulting in the most important classification is chosen first. For example, the creatinine clearance of 78.51 at the top of the Figure 1 is the first best split to distinguish people with cancer from others. This kind of classification method can be particularly useful because it highlights the most important biomarkers for specific diseases, and also reveals the level at which biomarkers are most predictive of these disease outcomes.

In Random Forest (RF) classification, each classification tree is fit to a bootstrap sample of the original dataset, and a random sample is independently drawn for each node of the tree. This procedure is likely to reduce error rates and improve the selectivity and sensitivity of the classification [7]. Taking advantage of these

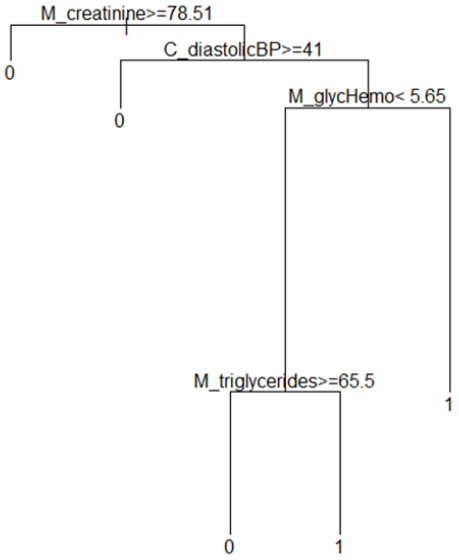


Figure 1: Tree diagram for cancer by AL biomarkers, NHANES 2005-2006

desirable statistical properties, we firstly fit RF for the NHANES 2005-2006 data. Predictors for each classification were our eleven AL biomarkers, and the response variables were four secondary or tertiary AL outcomes (cancer, heart disease, brain/cognitive dysfunction estimated by confusion/memory problems, and immune deficiencies measured by the presence of influenza and pneumonia). Second, we drew variable importance plots to check the explanatory power of each biomarker in the classification of people with the disease (i.e. the secondary/tertiary AL outcomes) from others. Third, we drew partial dependence plots, which permit us to identify the location where the probability of disease begins to increase for each biomarker. Our analysis was conducted under R 3.02 using the *randomForest* package [8].

3 Results

The percentages correctly classified (PCC) by fitting RF models to our data are 90.91% for cancer, 96.26% for heart disease, 95.14% for flu/pneumonia and 92.22% for confusion/memory problems. Figure 2 presents variable importance plots for biomarkers estimated by RF. The x-axis of the plot - "mean decrease in accuracy" - is the normalized difference of the classification accuracy when data for a particular biomarker are included. Higher values on the x-axis indicate that the biomarker is more important to the classification. The importance of specific biomarkers vary by disease, but some biomarkers such as creatinine clearance, BMI and cholesterol levels are consistently important across the four AL outcome measures. On the other hand, some biomarkers such as glycosylated hemoglobin and c-reactive protein (CRP) play only trivial roles for predicting the health outcomes included in our study. This procedure makes it possible to identify the "surplus" biomarkers that have negligible ability to predict secondary/tertiary health outcomes in AL analyses.

Figure 3 shows partial dependence plots, which graphically express the relationship between BMI and the predicted probability of the disease. (We drew the same set of plots for each biomarker; they are omitted here for space considerations). The red triangle on the x-axis represents the risk threshold set by the standard approach (i.e., 75^{th} percentile). Across all four AL outcomes, we can see that the quartile method sets the risk threshold at a point substantially higher than the point where the probability of disease actually begins to increase. For example, Figure 3-(b) shows that while the probability of presenting with heart disease begins to increase around BMI=25, the quartile method set the risk threshold at BMI=32.15. This leads to substantial differences in the number of people identified as at risk for health outcomes. When the risk threshold of BMI is 32.15, 402 persons in our sample are considered to be at the risk. By comparison, when the dependence plot sets the risk threshold (BMI=25), 1,113 persons are shown to be at elevated risk for heart disease. The plots for other biomarkers that

are omitted from this extended abstract indicate that the quartile method consistently sets risk at higher levels than where the probability of disease actually begins to increase.

4 Discussion

It is important to accurately operationalize AL in empirical research. Our findings suggest that simple indexes of AL may be misleading due to (1) the inclusion of biomarkers that have negligible ability to predict secondary and tertiary AL health outcomes, and (2) arbitrary and often excessive settings for risk thresholds. Our results by RF classification models suggest that some biomarkers do not play a critical role in predicting AL health outcomes. Therefore, assigning a point to each elevated biomarker without any consideration to its true importance can yield misleading AL scores. To avoid this problem, the inclusion or exclusion of biomarkers in AL studies should be determined empirically, using RF or other modeling techniques. Across repeated studies, this process may reveal that certain biomarkers are especially important to the AL construct. Conversely, if there is dramatic variation in the most relevant biomarkers across health outcomes and data sources, then it may be necessary to reconsider AL as a global indicator of stress accumulation and physiological deterioration.

Our findings also suggest that the quartile method of setting risk thresholds needs to be reconsidered. In our analyses, the top quartile of each biomarker misrepresented the point where the probability of disease began to increase. This occurs because the upper quartile is largely affected by the distribution of the population on the biomarker rather than being defined by actual disease risk. For example, the cut-off for BMI is likely to be set at a higher point in a population clustered around higher BMI levels (e.g. African Americans) than in other populations (e.g. Asian Americans), regardless of the underlying probabilities of disease. This arbitrary risk threshold heightens the odds of misclassifying research participants as at risk or not at risk for particular conditions based solely on underlying population distributions of biomarker values. Based on the results of our investigation, we suggest that future studies of the AL construct empirically test the appropriateness of biomarker risk thresholds, rather than presuming that secondary and tertiary AL outcomes will necessarily correspond to the top quartile of biomarker values.

References

[1] B. S. McEwen and T. Seeman, "Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load," *Annals of the New York Academy of Sciences*, vol. 896, no. 1, pp. 30–47, 1999.

[2] A. T. Geronimus, M. Hicken, D. Keene, and J. Bound, "weathering and age patterns of allostatic load scores among blacks and whites in the united states," *Journal Information*, vol. 96, no. 5, 2006.

[3] R. Kaestner, J. A. Pearson, D. Keene, and A. T. Geronimus, "Stress, allostatic load, and health of mexican immigrants*," *Social Science Quarterly*, vol. 90, no. 5, pp. 1089–1111, 2009.

[4] M. K. Peek, M. P. Cutchin, J. J. Salinas, K. M. Sheffield, K. Eschbach, R. P. Stowe, and J. S. Goodwin, "Allostatic load among non-hispanic whites, non-hispanic blacks, and people of mexican origin: effects of ethnicity, nativity, and acculturation," *American journal of public health*, vol. 100, no. 5, pp. 940–946, 2010.

[5] L. Chyu and D. M. Upchurch, "Racial and ethnic patterns of allostatic load among adult women in the united states: findings from the national health and nutrition examination survey 1999–2004," *Journal of Women's Health*, vol. 20, no. 4, pp. 575–583, 2011.

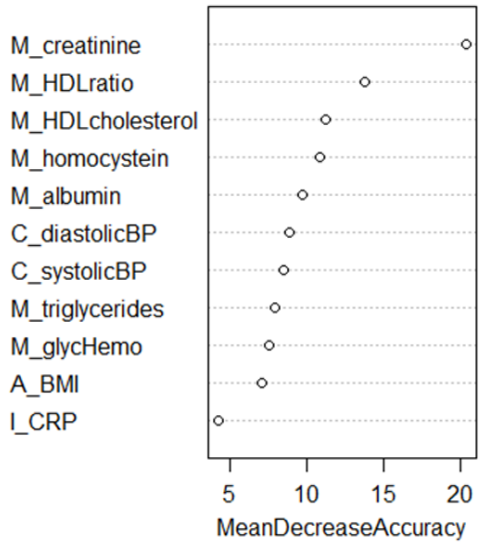
[6] A. Cutler and J. R. Stevens, "[23] random forests for microarrays," *Methods in enzymology*, vol. 411, pp. 422–432, 2006.

[7] D. R. Cutler, T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.

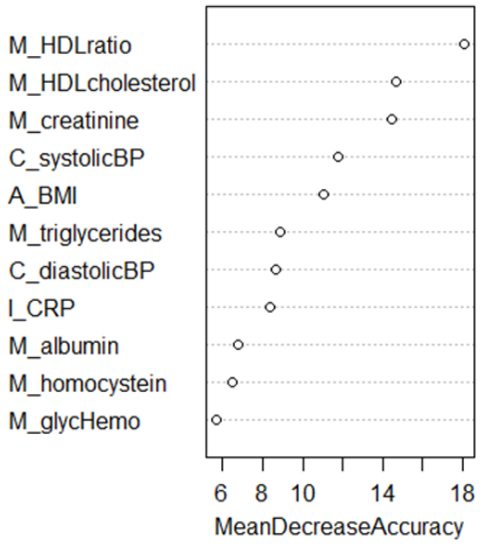
[8] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

Figure 2: Variable Importance Plots

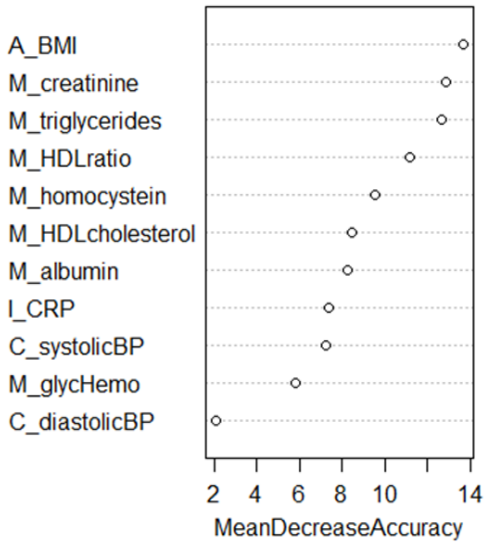
(a) Cancer



(b) Heart disease



(c) Flu/Pneumonia



(d) Memory problem

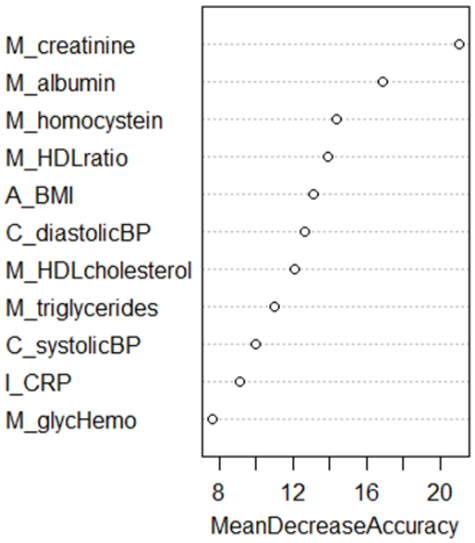
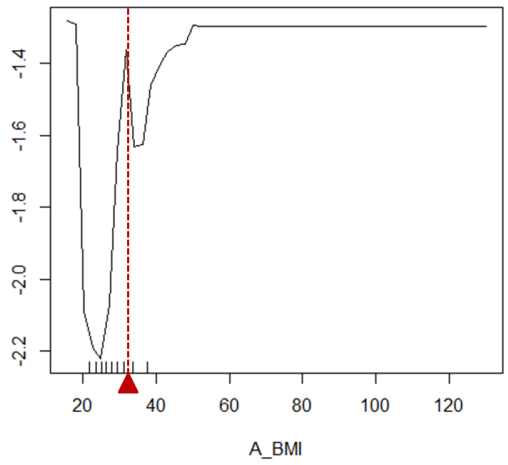
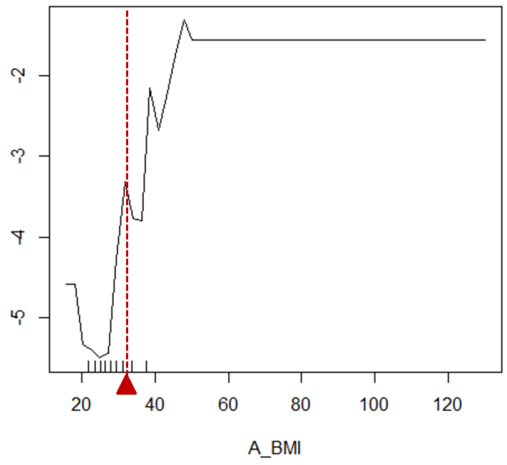


Figure 3: Partial dependence plots of BMI to predict the probability of AL outcomes

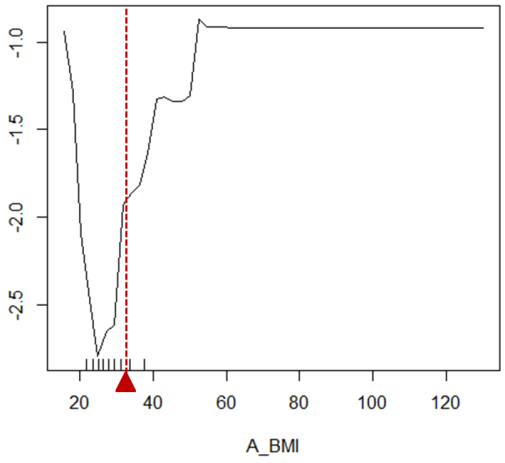
(a) Cancer



(b) Heart disease



(c) Flu/Pneumonia



(d) Memory problem