Send Orders for Reprints to reprints@benthamscience.ae

Current Aging Science, **2017**, *10*, 00-00

REVIEW ARTICLE

A Random Forest-Induced Distance-Based Measure of Physiological Dysregulation

Ghalib A. Bello[1,*] and Gerard Dumancas[2]

[1]*Department of Environmental Medicine & Public Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA;* [2]*Department of Mathematics & Physical Sciences, Louisiana State University of Alexandria, Alexandria, LA 71302, USA*

Abstract: ***Background*:** Aging involves gradual, multisystemic physiological dysregulation (PD) which increases risk of age-related comorbidities. Ability to quantify age-related PD could provide insights into biological mechanisms underlying the aging process. One approach to measuring PD exploits the fact that increasing PD manifests as a gradual deviation of physiological parameters away from normal levels. A recent geometric approach for quantifying PD uses Mahalanobis distance to measure the extent to which an individual's physiological parameters (measured via biomarkers from clinical blood biochemistry panels) deviate from normal levels. While useful, this approach has shortcomings that may impact its accuracy, primarily the incorrect assumption of multivariate normality among biomarkers, and identical weighting of biomarkers. Herein, we develop a more robust multivariate distance-based measure of PD.

***Methods*:** Proximity matrices induced by survival tree ensembles (Random Survival Forests) were used to compute a robust distance metric for quantifying how abnormal an individual's biomarker profile is. This approach requires no distributional assumptions and allows differential weighting of biomarkers based on association with mortality. Using receiver operating characteristic analysis and model fit statistics we compared performance of our measure to the standard approach based on Mahalanobis distance.

***Results & Conclusion*:** Our new metric showed statistically significant improvements in predicting mortality, health status and biological age, compared to the standard approach. Additional advantages offered by our method are the ability to handle missing values in biomarkers and to accommodate categorical risk factors. These results suggest our approach could provide greater precision in the evaluation of PD, which could enable better characterization of the extent and impact of degenerative processes resulting from aging.

ARTICLE HISTORY

Received: November 08, 2016
Revised: December 07, 2016
Accepted: January 15, 2017

DOI:
10.2174/1874609810666170117155756

Keywords: Physiological dysregulation, statistical distance, Mahalanobis distance, random survival forests, proximity matrix.

INTRODUCTION

In recent years, increased focus has been put on studying the cumulative toll that aging puts on an individual's health. Aging is known to be associated with dysregulation of multiple physiological systems: immune, cardiovascular, metabolism, *etc*. [1]. The term Physiological Dysregulation has been used as a generic descriptor for this set of biological processes. Recently, Cohen *et al*. [2] introduced a general approach for quantifying physiological dysregulation, opening the door for empirical studies of this construct and its relationships to various aspects of aging. Dysregulation in various physiological systems can be detected by measuring clinical biomarkers (laboratory tests), *e.g*. disruption in kidney function can be detected by measuring serum markers like creatinine, blood urea nitrogen, *etc*. Clinical biomarkers, as a whole, can provide a cumulative sense of how dysregulated one's physiology is [3]. Cohen *et al*. proposed a way of leveraging the cumulative information content in biomarkers by reducing these measures into a single score representing the degree of dysregulation. They demonstrated how an individual's 'biomarker profile'-a panel of lab test results obtained on an individual-can be aggregated into a univariate numeric score using a statistical distance-based metric [3-5]. The components comprising the biomarker profile typically span multiple physiological systems, *e.g*. components from the metabolic panel (electrolytes, proteins, enzymes), lipid panel (cholesterol, HDL, triglycerides), complete blood count, *etc*. They propose assessing physiological dysregulation by quantifying how aberrant an individual's biomarker profile is. This is done by comparing the levels of these biomarkers to those of a population average which is assumed to represent the physiological norm. The key idea, as explicated by the originators, is that the extent of physiological dysregulation is related to the degree to which an individual's biomarker profile diverges from that of a population norm. The hypothesis underlying this statistic is

*Address correspondence to this author at the Department of Environmental Medicine & Public Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; Email: ghalib.bello@mssm.edu

1874-6098/17 $58.00+.00

rooted in biological theory of homeostasis, the idea that there is a tendency for the body to maintain critical physiological parameters (*e.g*. blood glucose, blood metabolites, *etc*.) of its internal milieu within very narrow, specific ranges [2, 6]. A sense of what these 'normal' ranges look like can be captured by looking at the population average. Deviations from these averages point to a system that is poorly regulated or dysregulated. A highly dysregulated system reflects the deterioration of the body's ability to maintain physiological parameters within normal/healthy ranges [7].

To quantify the degree to which an individual's biomarker profile diverges from the population norm, we can estimate the probability of observing those biomarker values. The reasoning behind this approach is that lower probabilities correspond to rarer, and thus more divergent or extreme biomarker profiles. Let us represent a random biomarker profile by the random variable **X**, representing the measured levels of *p* physiological biomarkers. Here, **X** is a *p*-dimensional variable consisting of values of *p* different clinical biomarkers, *i.e*. $\mathbf{X} = (X_1, X_2,\ldots, X_p)$, where the X_i represent levels of different biomarkers (*e.g*. cholesterol, bilirubin, albumin, *etc*.). To estimate the rarity of a particular observed value of **X**, one intuitive direction involves the presumption that the distribution of **X** is multivariate normal, *i.e*.:

$$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1}$$

Here, $\boldsymbol{\mu}$ denotes the population average of **X** and $\boldsymbol{\Sigma}$ the variance-covariance matrix of **X**. Then **X**'s probability distribution function (which describes the probability of observing a biomarker profile $(X_1,\ldots X_p)$) is given by:

$$f_x(X_1,....,X_p) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\} \tag{2}$$

The only varying portion of the expression above is the quantity within the exponential term (since it depends on random variable **X**). So under assumption of multivariate normality, the probability of **X** is a simple function of the quantity in the exponential: $(\boldsymbol{X}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$. This quantity is referred to as the Mahalanobis distance (D_M) [8] and for a particular observed biomarker profile (a combination of levels of *p* biomarkers), the larger D_M is, the lower the probability of observing that profile. To see this more clearly, let us express the equation above in terms of D_M:

$$f_x(x_1,....,x_p) \propto \exp\left\{-\frac{1}{2}D_M^2(\boldsymbol{x})\right\} \tag{3}$$

This shows the inverse relationship between the probability of observing a biomarker profile value and the magnitude of distance metric D_M. D_M could therefore be seen a measure of how rare, or atypical an individual's biomarker profile is, or how much of an outlier their biomarker measures are. And to the extent that an individual's biomarker profile represents their underlying physiological function, D_M gives a sense of how abnormal or 'out of the ordinary' their physiological parameters are.

To use D_M in practice, sample-based estimates of the quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ must be used (we henceforth refer to these sample estimates as $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$). Cohen and coauthors were the first to propose Mahalanobis distance for quantifying the extent of physiological dysregulation [3]. Indeed, D_M is a natural and intuitive measure, which also has a straightforward relationship to the Euclidean distance, as is demonstrated below:

$$\begin{aligned} D_M^2(\boldsymbol{x}) &= (\boldsymbol{x}-\bar{\boldsymbol{x}})^T \boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}) = \left(\boldsymbol{S}^{-1/2}(\boldsymbol{x}-\bar{\boldsymbol{x}})\right)^T \left(\boldsymbol{S}^{-1/2}(\boldsymbol{x}-\bar{\boldsymbol{x}})\right) \\ &= \left\|\boldsymbol{S}^{-1/2}(\boldsymbol{x}-\bar{\boldsymbol{x}})\right\|_p^2 \end{aligned} \tag{4}$$

When there exist zero correlations among the biomarkers and they all have unit variance, $\boldsymbol{S}$ in the equation above reduces to the identity matrix, and the expression reduces to the familiar Euclidean distance-a simple geometric measure of the distance between the observed biomarker profile x and the population mean.

In practice, the biomarkers have varying scales and units of measurements, and hence different variances. Further, there is a complex correlation pattern among biomarkers. The Mahalanobis distance accounts for this variance-covariance structure by inclusion of the $\boldsymbol{S^{-1}}$ factor in the Euclidean distance metric. Geometrically, this is tantamount to transforming the data into uncorrelated, standardized (zero mean, unit standard deviation) data and computing the Euclidean distance for this transformed data. Hence the Mahalanobis distance provides a way to measure distances that takes into account the scale of the data.

The above shows how, starting from the assumption of multivariate normality, a simple measure can be derived which has the familiar structure of a distance metric and adjusts for the differing units/scales of the biomarkers and their underlying covariance structure. In a series of studies, Cohen and coworkers validated this measure by showing it was associated with mortality in multiple human populations. D_M was also found to provide valid and reliable estimates of physiological dysregulation in multiple datasets spanning different age ranges and geographical locations [2-4, 9, 10]. In addition, D_M appears to be highly correlated with age and frailty, and was found to be a strong predictor of chronic ailments like cardiovascular disease and diabetes [5]. Longitudinal versions of D_M have been used to study trajectories of physiological dysregulation in aging. More recently, a study by Milot and coworkers revealed substantial sex differences in the dynamic aging process, with males showing greater sensitivity to dysregulation (in terms of mortality risk) and females showing a faster pace of dysregulation [10, 11].

While enormously useful, the D_M approach is based on a couple of key assumptions which cannot, in most practical situations, be justified:

1. The distribution of biomarkers is multivariate normal.
2. Each variable contributes equally to determination of physiological dysregulation.

The first point is important to note; D_M assumes multivariate normality among the biomarkers. In other words, its accuracy as a measure of physiological dysregulation depends on how closely the joint distribution of biomarkers resembles a multivariate normal distribution. Cohen and co-

workers concede that this is rather unlikely in practice [3], and note that D_M is a rather conservative measure, and is not the most accurate measure of physiological dysregulation.

The second point presents a problem because the assumption is not borne out by empirical evidence. To illustrate this idea clearer, let us recast Equation (2) for the p=2 case where there are only 2 biomarkers (x_1 and x_2). For ease of interpretation, assume further that x_1 and x_2 have zero correlation (the point we intend to illustrate also applies when correlation is nonzero). For this special case, it can be shown that Equation (2) reduces to:

$$D_M(x) = \sqrt{\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2} \quad (5)$$

Note how each biomarker is weighted by its standard deviation. The standard deviation weighting is merely for the purpose of putting both biomarkers on the same scale. It has no biological relevance. In other words, we can imagine a scenario where marker x_1 has a stronger association with adverse health than marker x_2, so that abnormal values of x_1 are more life-threatening than equally abnormal values of x_2 (the term 'abnormal' in this specific context implies deviation from the mean). Thus, we see from this simple example that usage of the Mahalanobis approach to quantify physiological dysregulation carries the implicit assumption that, after variance-covariance adjustment, each biomarker is equally important in determining the extent of physiological dysregulation. Cohen *et al.* highlight this problem with their measure, noting that there is no incorporation of biological knowledge about the importance of being far from the mean/norm, for different variables [3]. Several studies have shown that not every biomarker contributes equally to physiological dysregulation, and that deviations from the norm on some biomarkers may carry more adverse health effects than for other biomarkers [12-14]. Further, factor analytic studies of allostatic load, a construct quantifying physiological dysregulation, have revealed differential factor loadings of biomarkers [15, 16]. The results of these studies suggest the need for measures of physiological dysregulation that do not assign equal weights to all biomarkers.

NONPARAMETRIC DISTANCE METRIC BASED ON IMPORTANCE/RELEVANCE WEIGHTING

In this paper, we introduce a novel approach that addresses these 2 issues. This approach does not rely on the assumption of multivariate normality and we will attempt to demonstrate that it represents a better and more accurate measure of physiological dysregulation. Second, our approach does not assume all biomarkers have equally strong relationships with overall physiological functioning. The approach focuses on identifying a subset of biomarkers with the strongest association with physiological dysregulation. Using mortality risk as a proxy for physiological dysregulation, biomarkers are differentially weighted with respect to their relevance. Biomarkers which are such that abnormal levels indicate a higher degree of physiological dysregulation are given greater weight in the metric. We call this general approach '*importance weighting*'. In geometric terms, this is equivalent to identifying the relevant subspace of biomarkers within which deviation from normal levels has the most substantive impact on overall health and life expectancy.

Beyond theoretical improvements that our method provides, it also has some practical advantages that make it a promising alternative to D_M in the quantification of physiological dysregulation. Calculation of D_M suffers from the following practical difficulties, all of which are addressed by our new metric: 1) there is no natural way to incorporate categorical variables in the calculation of D_M; 2) cannot deal with missing values in biomarker measures.

MATERIALS AND METHODS

Our method is based on the use of Random Survival Forests (ensembles of survival trees) [17] to induce a robust distance metric that does not rely on distributional assumptions. Below, we give brief descriptions of the construction of *survival trees* - the basic constituent of a random survival forest. This introduction to random survival forests is by no means meant to be exhaustive, and for the interested reader, complete theoretical and practical details about survival trees and random survival forests can be found in [17-21].

SURVIVAL TREES

Survival trees are a promising nonparametric technique for modeling right-censored time-to-event data. It has the ability to model complex nonlinear relationships and interactions, unlike its parametric and semi-parametric counterparts, *e.g.* Cox Proportional Hazards Models. A survival tree models the association between a set of explanatory variables and a time-to-event outcome, expressed as the time elapsed from a preset baseline to the occurrence of an event (death, in the present case), or the occurrence of censoring, whichever comes first. Survival trees also handle censoring in a natural and intuitive manner.

CONSTRUCTION

A survival tree is built by recursively partitioning a dataset using a series of binary splits determined by values of the individual predictor variables. Fig. (**1A**) gives a simple example of a survival tree created using 2 predictor variables (biomarkers)-A1c (glycohemoglobin) and C-Reactive Protein (CRP). In this illustrative example, we use just 2 variables for the sake of simplicity and note here that trees are typically built with more variables.

In Fig. **1A**, we see a series of binary splits that recursively divide the dataset into pairs of subsets. The first binary split (seen at the top of the tree) divides the population (n=5000) into 2 subsets - those with C-reactive protein above 0.08 mg/dL (n=3715) and those with C-reactive protein equal to or less than 0.08 mg/dL (n=1285). Each of these subsets is then further divided by subsequent binary splits, and so on. For each split, we refer to the variable used as the *splitting variable*, and the threshold value as the *split point*. For example, for the first split in Fig. **1A**, C-reactive protein would be the *splitting variable* and 0.08 would be the corresponding *split point*.

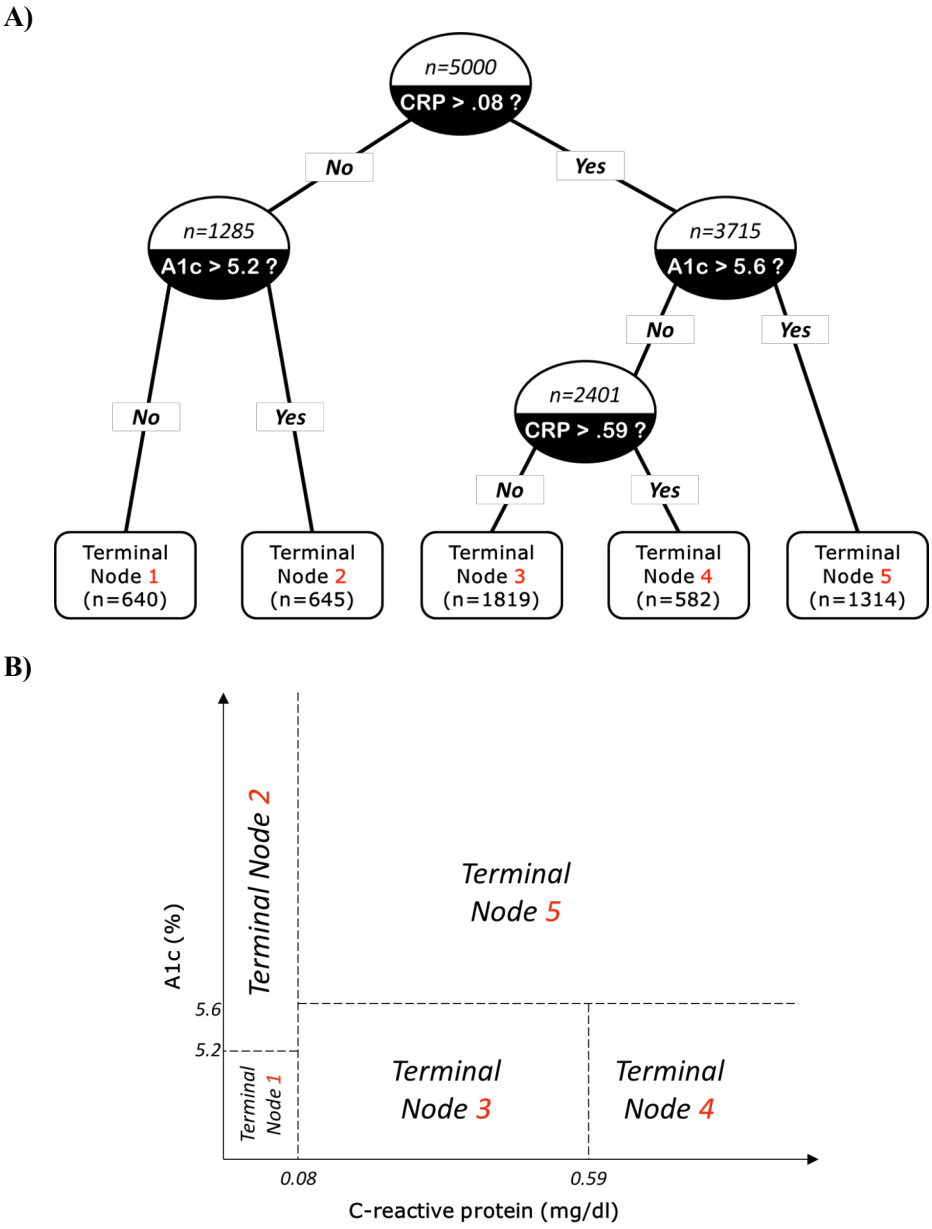

Fig. (1). A): Simple example of a survival tree modeling the relationship of biomarkers (A1c and C-reactive protein) with censored time-to-death outcome, among a sample of 5000 individuals. **B):** is a geometric depiction of the partitions generated by the tree (in Fig. **1A**) in the 2-dimensional space spanned by A1c and C-reactive protein values.

SPLIT CRITERION AND PRIORITIZATION OF HIGH-RELEVANCE VARIABLES

A natural question that arises is how the splitting variable and split points are chosen for each split. For example, in the first split of (Fig. **1A**), why is C-reactive protein (and not A1c) chosen as the splitting variable, and why is the value of 0.08 mg/dL (as opposed to any other value of C-reactive protein) chosen as the split point? Before each split is carried out, all variables (and all their corresponding split points) are compared, with the goal of finding the particular combination that does the best job of stratifying the subjects into 2 groups that are as different as possible with respect to the outcome (in the present case, survival). For example, in the first split of (Fig. **1A**), those with C-reactive protein > 0.08 mg/dL have substantially different survival outcomes than those with C-reactive protein ≤ 0.08 mg/dL, and this difference in survival trends was the maximum possible difference achievable after testing all variables and all possible split points. For survival trees, criterion functions based on statistical rules (*e.g.* log-rank test [22, 23]) are typically used for finding the optimal splits. Note that choosing splitting variables (and split points) in this manner guarantees that only variables with strong associations with the outcome (*i.e.* high-relevance variables) will be used in the tree. These variables will therefore control the architecture of the tree. Low-relevance variables with weak relationships with the outcome will tend to be rarely used.

TERMINATING TREE CONSTRUCTION

Subsequent splits follow the same heuristic described above, with each subset being further split into smaller subsets using the most optimal combinations of splitting variables and split points. Therefore constructing a tree essentially involves partitioning the dataset into smaller and smaller subgroups that are increasingly homogeneous with respect to the outcome. The hierarchical tree-like structure that emerges at the end of this recursive splitting process functions as a piecewise approximation of the underlying relationship between the biomarker variables and mortality, our surrogate measure of physiological dysregulation. The splitting procedure ends when certain criteria are met, *e.g*. subgroups are too small to be split further (see [17] for more details). In (Fig. **1A**), the final subgroups produced when the splitting process ends are referred to as *terminal nodes*.

GENERATING A DISTANCE MEASURE FROM SURVIVAL TREES AND RANDOM SURVIVAL FORESTS

A geometric interpretation of what the tree construction process does is that it partitions the space of predictor variables into a set of non-overlapping, homogeneous regions. For example, (Fig. **1B**) shows the partition that the survival tree in (Fig. **1A**) induces in the 2-dimensional space formed by all possible combinations of values of A1c and C-reactive protein. Since the partitions are defined by splitting values of A1c and CRP, individuals falling within the same region/partition can be thought of as having similar ranges of these variables. Conversely, individuals with very different values of these biomarkers will tend not to fall into the same tree-induced partition. This allows the definition of a distance metric that can quantify the similarity between pairs of individuals based on their biomarker measurements. Continuing with the example of the tree shown in (Fig. **1A**), we illustrate the idea of a survival tree-induced distance metric with the example below:

The table in (Fig. **2A**) lists the CRP and A1c measures of 4 sample individuals. Based on their values of these biomarkers, the tree in (Fig. **1A**) assigns them to certain terminal nodes or, equivalently, regions in the 2-dimensional space formed by all possible values of A1c and CRP. To summarize the similarities among these 4 patients, a 4 x 4 matrix can be defined (Fig. **2B**) in which each element (*i,j*) represents the 'similarity' between patient *i* and patient *j*. Here, 'similarity' is a simple indicator variable that evaluates to 1 if patients *i* and *j* fall into the same terminal node/region and 0 otherwise. This matrix is referred to as a *proximity matrix*. In the proximity matrix in (Fig. **2**), we see that all off-diagonal elements are 0 except for elements (1, 4) and (4, 1), indicating that patients 1 and 4 are the only pair who fall into the same region due to their similar values of A1c and CRP.

The example of tree construction we have provided above applies to a single tree. In practice, a large number of such trees are combined to make inference. This guarantees lower bias and lower variance (see [24] for more theoretical details). Using randomness-infusing techniques, a large number of survival trees with diverse architectures are generated. The resulting ensemble of survival trees is known as a Random Survival Forest (Ishwaran 2008) [17]. Each tree is able to partition the biomarker data based only on the most 'relevant' variables, *i.e*. those with the strongest association with mortality. Because of the random nature of tree construction, each tree likely will bin/partition the data differently. In doing so, each tree induces a proximity matrix (such as the one seen in Fig. **2**) for all pairs of individuals.

To calculate the statistical distance between a set of biomarker levels **x** and the population norm ($\mathbf{x}_{norm}$), we first need to define a population norm - a central value within the population that represents what a biologically 'normal' set of biomarker values looks like. For D_M, the multivariate mean is used. In our approach, we instead use the multivariate *marginal* median to define our hypothetical 'normative' profile, which we outline below as x_{norm}:

$$x_{norm} = \left(median(x_1), median(x_2), ..., median(x_p)\right)$$

$\mathbf{x}_{norm}$ is a multivariate quantity, each element of which represents the median of each of the p biomarkers. To calculate the distance between an individual's biomarker profile **x** and $\mathbf{x}_{norm}$, both **x** and $\mathbf{x}_{norm}$ are passed down all the trees in the ensemble and for each tree, the terminal node that each one landed in is noted. We then count the number of trees in which **x** and $\mathbf{x}_{norm}$ landed in the same terminal node. If **x** is close to the population norm, it will be in the same bin more often. If **x** deviates greatly, it will not. Notice that since the trees are constructed using mortality as an outcome, the bins are largely defined by the subset of biomarkers with strong associations with mortality (*i.e*. high-relevance biomarkers). For these biomarkers, deviations from the norm carry greater health risks than do equally large deviations for less impor-

A

Patient	CRP (mg/dl)	A1c (%)	Terminal Node
1	0.21	5.4	3
2	1.06	9.2	5
3	0.06	4.8	1
4	0.32	5.2	3

B

	Pt 1	Pt 2	Pt 3	Pt 4
Patient 1	1	0	0	1
Patient 2	0	1	0	0
Patient 3	0	0	1	0
Patient 4	1	0	0	1

Fig. (2). Generating a proximity matrix for a set of individuals.

tant biomarkers. This underscores the usefulness of our approach.

What follows is a formalization of the distance metric we have described above. Let B be the number of trees in an ensemble of trees. Further, denote each tree in the ensemble by T_{θ_b}, where θ_b is a parameter defining the series of splits that produced tree T_{θ_b}. Every tree T_{θ_b} has a set of terminal nodes L=1,….,L_b. For each terminal node *l*, let R_L be the region (in the biomarker space) defined by the terminal node. If $\mathbf{x}_i$ is the vector of values of the *p* biomarkers for patient *i* then our distance between $\mathbf{x}_i$ and the population norm ($\mathbf{x}_{norm}$) is given by:

$$D_{RF}^2(x) = 1 - \frac{1}{B}\sum_{b=1}^{B} I\left\{x \in R_{l(x_{norm}, T_{\theta_b})}\right\} \quad (6)$$

The notation $I\{\cdot\}$ is used to denote an indicator function which resolves to 1 if x falls within the same terminal node as $\mathbf{x}_{norm}$, and 0 otherwise. Similar to D_M, higher values of D_{RF} imply a greater degree of deviation of biomarker levels from the population norm (estimated by $\mathbf{x}_{norm}$), which reflects a failure of an individual's body to maintain physiological parameters within normal operating ranges.

Note that $\mathbf{x}_{norm}$ as defined in equation (3) can be regarded as the 'coordinate-wise median'. The coordinate-wise median is an easy-to-compute multivariate location estimator that has been shown to be more robust than the multivariate mean (which is the population norm used in the Mahalanobis Distance) [25]. The coordinate-wise median is particularly ideal as a measure of centrality for clinical biomarker levels since some biomarkers (*e.g.* C-reactive protein) have very skewed distributions, making the mean an inappropriate measure of multivariate center. An additional advantage is that the median as a measure of centrality is less sensitive to outliers.

PERFORMANCE COMPARISONS

Data

This study utilized clinical and biological data collected by the CDC on a large cohort of the US population. This cross-sectional study, known as the National Health and Nutrition Examination Survey (NHANES), was designed to assess the health and nutritional status of the US population. These files also provide information about death status and survival times (up to December 31, 2011) via linkage to the National Death Index. We used 33 biomarkers for the computation of D_{RF} and D_M, which are listed in Table **S1** (Appendix). The analysis was carried out in 2 stages. In the first stage, we used biomarker and mortality information from the 1999-2002 NHANES cycles to fit random survival forests (described above) in order to produce proximity matrices required for the computation of D_{RF}. In the second stage, we computed the D_M and D_{RF} metrics, then the association of these metrics with multiple surrogates of physiological dysregulation (self-assessed health, age and mortality) was compared. The goal of these comparisons is to determine whether D_{RF} demonstrates a stronger association to these measures than D_M does. Details of these comparisons are given below:

Self-Assessed Health

Studies have shown that self-assessed health status is a valid and reliable measure of overall health [26]. Participants in the NHANES surveys were asked to rate their health on a 5-point scale: Excellent, Very Good, Good, Fair or Poor. Multiple studies have shown this measure is reasonably accurate and has high test-retest reliability. For the purpose of our study, we dichotomized this measure into 2 categories defined by Excellent/Very Good/Good health and Fair/Poor health. We compared the predictive power of D_{RF} and D_M for this outcome using AUC.

Age

According to allostasis theory, physiological dysregulation increases gradually with age. Therefore, age is a valid and intuitive proxy for physiological dysregulation. Previous studies have shown that D_M is strongly associated with age, and have validated the paradigm of accumulating physiological deficits concomitant with aging. We compared the correlation of D_M and D_{RF} with age using Spearman's rank-based method. To adjust for potential confounders, we fit a full regression model for each distance metric, adjusting for sex and race/ethnicity. Each model had age as a dependent variable and sex, race/ethnicity and D_M or D_{RF} as independent variables. We compared adjusted R-squared measures for each of these models to determine which produced a better fit.

Computational Details of Random Survival Forest Generation and Calculation of D_{RF} Measures

Random Survival Forests were fitted using the R package *random Forest SRC* [27], with 15,000 trees and node size of 3. D_{RF} was computed as described above. First, $\mathbf{x}_{norm}$ was derived by computing the coordinate-wise median for the 33 biomarkers within the 1999-2002 NHANES data. Then this vector was passed through each of the trees, and the identity of the terminal node it landed in was recorded for each tree. Then for each individual in the sample, we record the proportion of the 15,000 trees in which their biomarker profile $\mathbf{x}$ landed in the same terminal node as $\mathbf{x}_{norm}$ and compute D_{RF} as defined in equation (6).

All analysis was carried out in R.

RESULTS

Summary statistics for demographics and health outcome measures for the study sample are summarized in Table **1**. We excluded individuals who were below the age of 18 at the time of participation in NHANES 1999-2002 surveys. This left 10,568 individuals, of which only n=8977 had non-missing values for the 33 biomarkers used to compute our metrics of physiological dysregulation.

IMPORTANCE-WEIGHTING

Random survival forests provide measures of the importance of each variable in predicting the outcome. Analogous to regression coefficients in parametric regression models,

Table 1. Summary statistics on analytical sample.

Variable		Mean	Median	SD	Range
Demographic variables					
Age		46.2	44	20.26	18-85
Sex	**Female**	52.5%	-	-	-
Race/Ethnicity	**White**	46.0%	-	-	-
	Black	20.1%	-	-	-
	Hispanic	31.0%	-	-	-
Health outcome variables					
Survival time (years)	**Censored**	10.9	11	-	4.8-12.75
	Uncensored	6	5.9	-	.08-12.58
5-year mortality	-	6.5%	-	-	-
Self-assessed health	**Excellent**	13.3	-	-	-
	Very Good	30.6	-	-	-
	Good	35.6	-	-	-
	Fair	17.2	-	-	-
	Poor	3.3	-	-	-

these measures give a sense of the 'effect size' of each explanatory variable with respect to the outcome variable. These measures allow the ranking of biomarkers according to predictive power for mortality, and identification of the highest- and lowest-importance variables. (Fig. **3**) below shows variable importance measures for each of the 33 biomarkers used to construct D_{RF}. Note that the plotted values are normalized to fall between 0 and 1. We observe that Creatinine, A1c and Red Blood Cell Distribution Width (RDW) have the highest importance measures. It is worth noting that multiple studies have found A1c and RDW to be significant predictors of all-cause mortality in general population cohorts across different countries [28-31]. These results provide insights into the differential weighting of biomarkers and the degree of influence they have on the value of the D_{RF} measure. We may conclude that the values of D_{RF} are largely determined by the values of the high importance biomarkers.

ASSOCIATION WITH AGE

Both D_M and D_{RF} were compared with respect to the strength of their association with age. The Spearman correlation with age was 0.13 for D_M and 0.5 for D_{RF}. Regression models of age adjusting for sex and race/ethnicity show that D_{RF} has a stronger association with age than D_M after adjusting for the potentially confounding influences of sex and race/ethnicity. Adjusted R-squared was 0.17 for the D_{RF} model and 0.06 for the D_M model.

SELF-ASSESSED HEALTH

D_{RF} showed higher predictive power for the dichotomized measure of self-assessed health (Excellent/Very Good/Good *vs*. Fair/Poor health). AUC for D_{RF} was 0.64 and 0.61 for D_M. Using DeLong's test, we compared this AUC and found the difference to be statistically significant (p=0.0035).

PREDICTION OF MORTALITY IN TRAINING AND VALIDATION SAMPLES

The D_{RF} metric was computed using proximity matrices derived from random survival forests trained on biomarker and survival data. Because of this, it is expected that this metric will outperform D_M in the prediction of mortality. This was confirmed via AUC measures for 5-year mortality. The analytical sample was classified into 2 groups, those who died within 5 years of NHANES participation date (the date their clinical biomarker data was collected), and those who survived beyond 5 years. Using this binary measure of 5-year mortality, we confirmed that D_{RF} yielded a statistically significantly ($p < .0001$) higher AUC (0.85) than D_M (AUC=0.69). These predictive accuracy comparisons are likely biased in favor of D_{RF} due to the fact that the comparisons are carried out on the same data used to train the survival tree ensemble from which the D_{RF} score was generated. We carried out a less biased comparison by using a different dataset, the NHANES data from 2003-2006. The NHANES is a strictly cross-sectional study and individuals are never selected to participate more than once. Hence, the NHANES 2003-2006 cohort is completely independent of the NHANES 1999-2002 cohort we used as our training data. Using this independent cohort, we found that D_{RF} yields an AUC of 0.8 for 5-year mortality, as compared to 0.71 for D_M. Aside from the dichotomized 5-year outcome, we compared mortality predictive accuracies of D_M and D_{RF} with

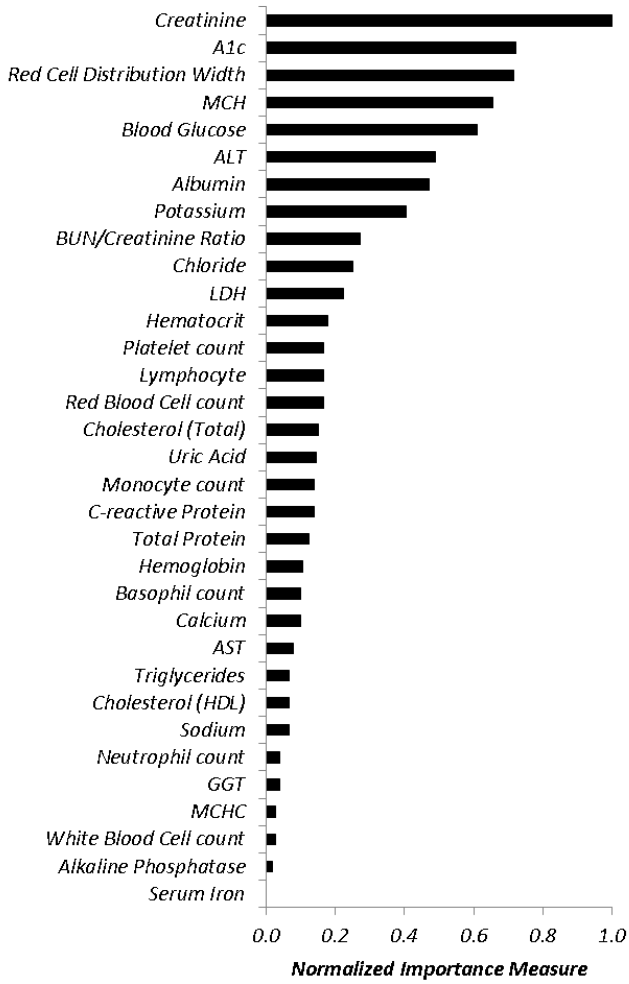

Fig. (3). Variable Importance Measures for biomarkers of Physiological Dysregulation.

Abbreviations: ALT=Alanine Aminotransferase; LDH=Lactate Dehydrogenase; AST=Aspartate Aminotransferase; GGT=Gamma-Glutamyl Transferase; MCHC=Mean Corpuscular Hemoglobin Concentration.

respect to the continuous survival outcome in the independent dataset (number of years to death or censoring). Harrell's Concordance Index [32] (known as Harrell's C) is a U-statistic similar to AUC and in fact may be considered a conceptual extension of AUC to right-censored survival outcomes. For metrics like D_M and D_{RF} wherein higher values signify greater mortality risk, Harrell's C is computed by estimating the probability that, for a randomly selected pair of individuals, the one with the higher value of the distance metric has shorter survival time (*i.e.* higher risk). Harrell's C has a similar interpretation to AUC, *i.e.* values closer to 1 suggest higher predictive accuracy. In the independent cohort, Harrell's C was significantly ($p < .0001$) higher for D_{RF} (0.79) than for D_M (0.7).

DISCUSSION

This study has introduced a new approach to computing distance-based measures of physiological dysregulation. D_{RF} addresses some of the shortcomings of the presently used D_M metric. By incorporating the use of mortality information in the computation of the D_{RF} distance metric, we account for the non-uniform contributions of biomarkers in the characterization of physiological integrity. We demonstrated using random forest variable importance measures that some biomarkers have significantly higher associations with mortality (and, by proxy, physiological dysregulation) than others.

The use of random survival forests as the machinery for generating our distance metric also confers a number of additional benefits. Random forests can handle different variable types, therefore categorical variables can easily be used in the computation of the D_{RF} metric. In contrast, the D_M metric is designed to handle continuous biomarker variables, and no natural, straightforward way exists for incorporating categorical variables in the metric. In addition, the D_M metric cannot handle missing values in biomarker variables. On the other hand, random forests provide an intuitive feature for handling missing values, referred to as surrogate splitting [33]. This means that D_{RF} can still be computed for individuals missing at least one of the biomarker variables used for measuring physiological dysregulation.

In addition to introducing a new type of distance metric for physiological dysregulation, our study provides empirical support for the approach of incorporating mortality in the derivation of physiological dysregulation measures. This has also been suggested in other studies (*e.g.* see [34]). We used random survival forests to fit biomarker data to mortality outcomes and the resulting distance metric shows strong associations to other surrogates of physiological dysregulation – self-rated health status and chronological age. We propose that the superior performance of D_{RF} in predicting these measures is due in large part to the fact that, unlike D_M, D_{RF} incorporates mortality information, which allows more 'important' biomarkers to be weighted more heavily in this metric. Since mortality is such a strong proxy measure for physiological dysregulation (and indeed, its ultimate byproduct), we contend that the biomarkers with the strongest relationships to mortality should be weighted more heavily in any measure of physiological dysregulation. However, a potential issue with using mortality as a calibration criterion is that D_{RF} as a measure of physiological dysregulation may be less sensitive to non-fatal types of dysregulation. One additional benefit of our approach is that it provides an objective, data-driven way to address the question of which set of biomarkers to use for quantifying physiological dysregulation. Cohen and coworkers [4, 5] have dedicated much effort to addressing the concern about which set of biomarkers are the most ideal to use. While their work revealed that the biomarker set used is largely irrelevant, we feel that our approach provides a simple, elegant and data-driven heuristic for addressing this problem. Our approach involves using all available biomarkers and simply letting the random survival forests prioritize more important/relevant biomarkers by upweighting their contributions and diminishing the influence of less relevant biomarkers. This way, we allow the data to make the choice of which biomarkers should drive the computation of the distance metric.

Despite its key improvements over D_M, our new approach has a few downsides. The calculation of D_{RF} requires fitting a random survival forest to the data. Besides being substantially more computationally intensive than the calculation of D_M, current implementations of random survival forests do

not account for left truncation. Also, our approach requires that mortality information be available. This type of information is not available in all datasets. In cases where mortality information is not available, we suggest using other surrogate measures of physiological dysregulation, *e.g*. self-assessed health. We are currently testing an 'unsupervised' method developed by Shi & Horvath (2006) [35] which does not require mortality or any outcome data. This technique involves creating synthetic data sampled from the joint marginal distribution of the biomarkers, and fitting a forest that distinguishes the synthetic data from the real data. This method has been applied successfully to multiple problems in the areas of cancer genomics [36] and Alzheimer's disease [37]. A second issue is a concern that is often raised about health indices developed by statistical models calibrating a set of predictive variables to an outcome/endpoint of interest [38]. It has been argued that there is an element of circularity introduced when an index calibrated using an outcome of interest is then used to carry out inference on that same outcome [5, 38]. D_{RF} was calibrated using mortality as an outcome, so its ability to predict mortality and other endpoints preceding (and associated with) mortality may largely be a reflection of its design. In contrast, an impressive feature of D_M that has been highlighted is that its computation is 'unsupervised', *i.e*. not derived by model-fitting or regressing biomarkers to a certain outcome/endpoint related to physiological dysregulation (*e.g*. mortality). Despite this, D_M shows remarkably strong associations with mortality, age and a host of other symptoms and endpoints known to be associated with physiological dysregulation.

Aside from the shortcomings of our technique, our study itself has a few limitations. Firstly, our metric was validated on only one clinical population – the NHANES. Future work will focus on comparing the performance of D_{RF} and D_M using biomarker data from other large population studies. Secondly, our study uses a cross-sectional design with biomarkers measured at just one time point. While we have demonstrated the superior performance of D_{RF} in this setting, it remains to be shown whether this trend holds for longitudinally measured values. Thirdly, only 2 outcome measures were used to validate D_{RF}–health status and age. While there is ample evidence for the relevance of these measures to physiological dysregulation, we note here that there exist other measures (*e.g*. telomere length, physical function, cognitive functioning) that can be used for these purposes. Future studies will focus on comparing the performance of D_{RF} and D_M on a more comprehensive set of outcome measures. Lastly, it must be noted that biomarker measures could potentially be affected by medication use among the cohort. However, since our study was based on a random, representative sample of the US population (NHANES), the effect of medication use on biomarker levels could be presumed to be negligible in this predominantly healthy and asymptomatic cohort.

An important insight about quantifying physiological dysregulation using distance-based measures is worth highlighting and discussing. As demonstrated in the introduction section, when a set of biomarkers follows a multivariate normal distribution, the Mahalanobis distance of a particular biomarker profile is inversely proportional to its probability of occurrence, *i.e*. its rarity. We can therefore think of the Mahalanobis distance as a measure of the 'outlyingness' of a profile – the extent to which it is an outlier, compared with the rest of the population. Thus we can conceptualize the task of quantifying the degree of physiological dysregulation as an exercise in measuring the degree to which an individual's biomarker profile is a population outlier. This paradigm is potentially useful because the statistical field of Outlier Detection (and the related field of Anomaly Detection in data mining) provides a wealth of techniques for quantifying 'outlyingness'. Decades of research in these areas have produced a diverse and expansive toolbox of methodologies that could be leveraged for deriving improved distance-based measures of physiological dysregulation. We are currently exploring the application of nonparametric outlier-detection techniques for quantifying physiological dysregulation, *e.g*. Isolation Forests [39], the Local Outlier Factor (LOF) algorithm [40], and Angle-Based Outlier Detection (ABOD) algorithms [41]. With suitable modifications, these algorithms can be adapted to produce continuous measures of 'degree of outlyingness' which can function as measures of physiological dysregulation.

We discuss here a few technical details for readers interested in utilizing the D_{RF} measure. In this study, we used the random survival forest implementation available in the R package *rfsrc* [27]. However, other implementations are available (*e.g*. the R package *cforest*). Random forests have relatively few tuning parameters and we have found that the D_{RF} measure is not highly sensitive to the choice of tuning parameters. We do however recommend the use of a sufficiently large tree ensemble (on the order of thousands) and to calibrate the minimum allowable terminal node size to a reasonable fraction of the total sample size. Note also that missing values can be handled in most random forest implementations, and built-in routines are available. One method uses an iterative tree-based imputation scheme that can handle mixed data types (*e.g*. continuous, categorical) [42]. Another approach (implemented in R package *cforest*) utilizes surrogate splitting. Recall that each non-terminal node in a constructed tree is split by a selected variable (the splitting variable). To handle possible cases with missing values on this variable, the rest of the variables in the dataset are compared at this node to determine how well they are able to mimic the primary split in the node. A good surrogate split is one that resembles the primary split as closely as possible, producing a similar or identical partitioning of the individuals in the node. Each candidate variable is ranked based on how well they function as surrogates for the original (primary) splitting variable, and this ranking information is stored for all the non-terminal nodes in the tree. At each of these nodes, a case/observation missing a value of the splitting variable will instead be split using the best surrogate variable for which it has a non-missing value [43].

We have demonstrated that D_{RF} is superior to D_M in terms of associations with proxy measures of physiological dysregulation. The results provide further validation for the novel notion of using mortality as a criterion measure for developing metrics of physiological dysregulation. With the burgeoning interest in exploring interventions for lifespan extension [44-46], there is a greater demand than ever for accurate and biologically valid measures of the physiological dysregulation that accompanies aging.

Appendix

Table S1. Summary statistics on biomarkers used to compute physiological dysregulation measures

Biomarker	Mean	SD	Median
Alanine Transaminase (ALT)	25.04	26.451	20
Albumin	4.2	0.407	4.2
Alkaline Phosphatase	72.2	26.856	68
Aspartate Transaminase (AST)	25.6	22.443	23
Basophil count	0.04	0.061	0
Blood Urea Nitrogen/Creatinine Ratio	13.8	4.980	13
Calcium	9.5	0.372	9.5
Chloride	103.7	2.733	104
Cholesterol (Total)	196.5	42.672	193
C-reactive protein	0.47	0.916	0.21
Creatinine	0.9	0.421	0.9
Gamma Glutamyl Transferase (GGT)	28.0	46.783	19
Glucose	98.2	33.568	91
Glycohemoglobin (A1c)	5.5	0.962	5.3
HDL Cholesterol	54.7	16.067	52
Hematocrit	42.3	4.578	42.4
Hemoglobin	14.3	1.565	14.3
Iron	84.7	35.951	80
Lactate Dehyrogenase (LDH)	128.7	31.902	125
Lymphocyte count	2.2	1.543	2
Mean Corpuscular Hemoglobin (MCH)	30.4	2.241	30.6
Mean Corpuscular Hemoglobin Concentration (MCHC)	33.7	0.850	33.8
Monocyte count	0.56	0.202	0.5
Neutrophil count	4.43	1.810	4.1
Platelet count	269.8	70.257	262
Potassium	4.0	0.342	4
Protein (total)	7.2	0.514	7.2
Red Blood Cell count	4.7	0.527	4.7
Red blood cell Distribution Width (RDW)	12.8	1.186	12.5
Sodium	139.0	2.259	139
Triglycerides	140.2	107.307	111
Uric Acid	5.3	1.415	5.2
White Blood Cell count	7.4	2.600	7.1

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

PATIENT'S CONSENT

Declared none.

REFERENCES

[1] Markides K. Encyclopedia of Health and Aging. Thousand Oaks, California: SAGE Publications 2007.

[2] Cohen AA, Milot E, Yong J, *et al*. A novel statistical approach shows evidence for multi-system physiological dysregulation during aging. Mech Ageing Dev 2013; 134(3-4): 110-7.

[3] Cohen A, E Milot, V Morissette-Thomas. Statistical distance: A promising measure of physiological dysregulation in biodemographic studies in XXVII International Union for the Scientific Study of Population 2013; Busan: South Korea.

[4] Cohen AA, Li Q, Milot E, *et al*. Statistical distance as a measure of physiological dysregulation is largely robust to variation in its biomarker composition. PLoS One 2015; 10(4): e0122541.

[5] Cohen AA, Milot E, Li Q, Legault V, Fried LP, Ferrucci L. Cross-population validation of statistical distance as a measure of physiological dysregulation during aging. Exp Gerontol 2014; 57: 203-10.

[6] Pear J. Physiological homeostasis and learning. In: Seel N, Ed. Encyclopedia of the sciences of learning. New York: Springer 2012; pp. 2628-31.

[7] Cohen A. Physiological and comparative evidence fails to confirm an adaptive role for aging in evolution. Curr Aging Sci 2015; 8(1): 14-23.

[8] Mahalanobis P. On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India 1936; 2(1): 49-55.

[9] Cohen A. Complex systems dynamics in aging: new evidence, continuing questions. Biogerontology 2016; 17(1): 205-20.

[10] Milot E, Morissette-Thomas V, Li Q, Fried LP, Ferrucci L, Cohen AA. Trajectories of physiological dysregulation predicts mortality and health outcomes in a consistent manner across three populations. Mech Ageing Dev 2014; 141-2: 56-63.

[11] Arbeev KG, Cohen AA, Arbeevaet LS, *et al*. Optimal versus realized trajectories of physiological dysregulation in aging and their relation to sex-specific mortality risk. Front Public Health 2016: 4; 3.

[12] Bello GA, Dumancas GD, Gennings C. Development and Validation of a Clinical risk-assessment tool predictive of all-cause mortality. Bioinform Biol Insights 2015; 9(3): 1-10.

[13] May HT, Anderson J, Muhlestein J, Lappe D, Ronnow B, Horneet B. Addition of albumin, bilirubin, WBC differential count, and other routinely collected laboratory tests to the Intermountain Risk Score improves the predictive ability of death among a general healthcare population. J Am Coll Cardiol 2012; 59(Suppl A): E1816.

[14] Karlamangla AS, Singer BH, McEwen BS, Rowe JW, Seeman TE. Allostatic load as a predictor of functional decline: MacArthur studies of successful aging. J Clin Epidemiol 2002; 55(7): 696-710.

[15] Howard J, Sparks P. Does allostatic load calculation method matter? evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. Am J Hum Biol 2016; 28: 627-35.

[16] McCaffery JM, Marsland AL, Strohacker K, Muldoon MF, Manuck SB. Factor structure underlying components of allostatic load. PLoS One 2012; 7(10): e47246.

[17] Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat 2008; 2: 841-60.

[18] Chen X, Ishwaran H. Pathway hunting by random survival forests. Bioinformatics 2012; 29: 99-105.

[19] Ishwaran H, U Kogalur. Consistency of random survival forests. Stat Prob Letters 2010; 80: 1056-64.

[20] Ishwaran H, Kogalur UB, Chen X, Minn AJ. Random survival forests for high-dimensional data. Stat Anal Data Mining 2011; 4: 115-32.

[21] Mogensen U, Ishwaran H, Gerds T. Evaluating random forests for survival analysis using prediction error curves. J Stat Softw 2012; 50(11): 1-23.

[22] LeBlanc M, Crowley J. Survival trees by goodness of split. J Am Statist Assoc 1993; 88: 457-67.

[23] Segal M. Regression trees for censored data. Biometrics 1988; 44: 35-47.

[24] Breiman L. Random forests. Mach Learn 2001; 45(1): 5-32.

[25] Rousseeuw P, Leroy A. Robust regression and outlier detection. Hoboken, New Jersey: John Wiley & Sons 2003.

[26] Zajacova A, Dowd J. Reliability of self-rated health in US adults. Am J Epidemiol 2011; 174(8): 977-83.

[27] Ishwaran H, Kogalur U. Random survival forests for R. R News 2007; 7(2): 25-31.

[28] Paprott R, Schaffrath RA, Busch MA, *et al*. Association between hemoglobin A1c and all-cause mortality: results of the mortality follow-up of the German National Health Interview and Examination Survey 1998. Diabetes Care 2015; 38(2): 249-56.

[29] Zhong GC, Ye MX, Cheng JH, Zhao Y, Gong JP. HbA1c and risks of all-cause and cause-specific death in subjects without known diabetes: a dose-response meta-analysis of prospective Cohort Studies. Sci Rep 2016; 6: 24071.

[30] Horne BD, Muhlestein JB, Bennett ST, *et al*. Association of the dispersion in red blood cell volume with mortality. Eur J Clin Invest 2015; 45(6): 541-9.

[31] Carson AP, Fox CS, McGuire DK, *et al*. Low hemoglobin A1c and risk of all-cause mortality among US adults without diabetes. Circ Cardiovascular Qual Outcomes 2010; 3(6): 661-7.

[32] Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982; 247(18): 2543-6.

[33] Feelders A. Handling missing data in trees: surrogate splits or statistical imputation? In: Zytkow J, Rauch J, Eds. Principles of data mining and knowledge discovery. Berlin: Springer-Verlag 1999.

[34] Arbeev K, Ukraintseva S, Yashin A. Dynamics of biomarkers in relation to aging and mortality. Mech Ageing Dev 2016; 156: 42-54.

[35] Shi T, Horvath S. Unsupervised learning with random forest predictors. J J Comput Graph Stat 2006; 15(1): 118-38.

[36] Seligson DB, Horvath S, Shi T, *et al*. Global histone modification patterns predict risk of prostate cancer recurrence. Nature 2005; 435: 1262-6.

[37] Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D; Alzheimer's Disease Neuroimaging Initiative. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. Neuroimage 2013; 65: 167-75.

[38] Singer B, Ryff C, Seeman T. Operationalizing Allostatic Load. In: Schuli J, Ed. Allostasis, homeostasis, and the costs of physiological adaptation. Cambridge, UK: Cambridge University Press 2004; pp. 113-49.

[39] Liu F, Ting K, Zhou ZH. Isolation-based anomaly detection. AACM Trans Knowl Discov Data 2012; 6(1).

[40] Breunig, M., *et al*., LOF: Identifying Density-based Local Outliers. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000. **29**(2): p. 93-104.

[41] Kriegel HP, Schubert M, Zimek A. Angle-based outlier detection in high-dimensional data. KDD 2008; 2008: 444-52.

[42] Stekhoven D and Buhlmann P. MissForest-non-parametric missing value imputation for mixed-type data. Bioinformatics 2012; 28(1): 112-8.

[43] Hapfelmeier A, Hothorn T, Ulm K. Recursive partitioning on incomplete data using surrogate decisions and multiple imputation. Comput Stat Data Anal 2012; 56: 1552-65.

[44] Mendonca GV, Pezarat-Correia P, Vaz JR, Silva L, Almeida ID, Heffernan KS. Impact of exercise training on physiological measures of physical fitness in the elderly. Curr Aging Sci 2016; 9(4): 240-59.

[45] Khizhkin EA, Ilukha VA, Vinogradova IA, *et al*. Physiological and biochemical mechanisms of lifespan regulation in rats kept under various light conditions. Curr Aging Sci 2017; 10(1): 49-55.

[46] Kulminski A, Svetlana VU, Irina VC, *et al*. Cumulative deficits and physiological indices as predictors of mortality and long life. J Gerontol A Biol Sci Med Sci 2008; 63(10): 1053-9.