Does the choice of Allostatic Load scoring algorithm matter for predicting age-related health outcomes?

Sinead McLoughlin[a,*], Rose Anne Kenny[a,b], Cathal McCrory[a]

[a] The Irish Longitudinal Study on Ageing (TILDA), Trinity Central, Trinity College Dublin, 152-160 Pearse Street, Dublin 2, Ireland
[b] Mercer's Institute for Successful Ageing, St James's Hospital, Dublin, Ireland

ABSTRACT

Allostatic Load (AL) is posited to provide a measure of cumulative physiological dysregulation across multiple biological systems and demonstrates promise as a sub-clinical marker of overall health. Despite the large heterogeneity of measures employed in the literature to represent AL, few studies have investigated the impact of different AL scoring systems in predicting health. This study uses data for 4477 participants aged 50+ years participating in the Irish Longitudinal Study on Ageing (TILDA) to compare the utility of 14 different scoring algorithms that have been used to operationalise AL (i.e. count-based high-risk quartiles, deciles, two-tailed cut-points, z-scores, system-weighted indices, clinical cut-points, sex-specific scores, and incorporating medication usage). Model fit was assessed using R^2, Bayesian Information Criterion (BIC), and the area under the Receiver Operating Characteristic curve (AUC). The measure incorporating medications predicted walking speed and SRH marginally better than others. In general, AL was not predictive of grip strength. Overall, the results suggest that the choice of AL scoring algorithm exerts a relatively modest influence in predicting a number of important health outcomes.

1. Introduction

Allostatic Load (AL) is posited to represent a sub-clinical measure of physiological wear and tear resulting from chronic exposure to life course stressors (McEwen and Stellar, 1993). In recent decades, the framework has contributed to an enhanced interdisciplinary understanding of how social, environmental, and psychological factors impact physiological functioning and shape health disparities (Beckie, 2012; Merkin et al., 2009; Upchurch et al., 2015). Despite this apparent utility, AL is beset by a number of methodological and conceptual difficulties that have hampered its potential clinical utility. These include: (1) the ongoing failure to agree a core set of biomarkers that define the construct, and (2) a plethora of different AL scoring algorithms, limiting our ability to compare results across studies. The former issue, concerning the heterogeneity of biomarkers used across studies, has been discussed at length in a number of recent reviews (Johnson et al., 2017; Juster et al., 2010), but the latter issue has received rather less attention in the literature, and is arguably just as important.

1.1. Different scoring systems for AL

Seeman et al. (1997) provided the first operational definition of AL using a high-functioning sample of adults aged 60 years and older. They employed 10 biomarkers across the neuroendocrine, immune, metabolic and cardiovascular systems, and summed the number of parameters for which an individual had values in the highest risk quartile, based on the sample distribution (Seeman et al., 1997). This measure predicted physical and cognitive decline within the MacArthur Studies of Successful Ageing in the US, and greater incidence of cardiovascular disease over a three-year follow-up. The authors concluded that AL might perform even better in general population studies. The original scoring algorithm for AL remains a popular method today, with the aforementioned systematic review (Johnson et al., 2017) reporting that 73 % of included studies utilised this approach, despite the potential loss of information from the full risk spectrum.

Lack of consensus around an agreed biomarker panel (Johnson et al., 2017; Juster et al., 2010; Szanton et al., 2005) and scoring system (Beckie, 2012; Seplaki et al., 2005) has led to a proliferation of different means for characterising AL. Considerable heterogeneity in AL calculation exists across studies utilising the MacArthur cohort alone, including: the original count of high-risk quartiles (Seeman et al., 1997, 2004; Weinstein et al., 2003), recursive partitioning (Gruenewald et al., 2006), and use of second-order terms (Karlamangla et al., 2006). Alternate AL formulations from other international studies include: a sum

---

* Corresponding author.
E-mail address: mclougs4@tcd.ie (S. McLoughlin).

of z-scores, which may be considered more informative as it retains the continuous properties of the data (Hawkley et al., 2011; Levine and Crimmins, 2014), or the use of decile cut-points which consider only the top end of the risk distribution. Two-tailed approaches (i.e. top and bottom quartiles or deciles) for some (Hwang et al., 2014; LeBron et al., 2019) or for all biomarkers (Seplaki et al., 2006) have also been employed to account for the fact that biological risk may be non-linear.

Canonical correlation, a technique that measures the associations between sets of inter-connected variables (Karlamangla et al., 2002), and recursive partitioning, a decision-tree technique (Gruenewald et al., 2006; Singer et al., 2004), have been criticised for incorporating information on health outcomes in their derivation, which can result in the over-fitting of models and limits replication across other datasets (Seplaki et al., 2005). Clinical cut-points have also been used (Allen et al., 2019; Borrell et al., 2010; Crimmins et al., 2009; Glei et al., 2013; Petrovic et al., 2016; Vasunilashorn et al., 2013); but, the lack of universally agreed values (e.g. cut-points for parasympathetic or immunological markers) potentially limits their application. In addition, such definitions of risk remove the potential utility of AL as a subclinical disease measure for identifying risk prior to the emergence of the clinical phenotype. More sophisticated attempts to develop AL measures include multivariate methods, such as grade of membership (GOM) which assesses how well an individual's scores on the set of biomarkers used to define AL correspond to a set of predefined archetypal profiles (Li et al., 2019; Seplaki et al., 2006). Interestingly, the predictive utility of AL measures calculated using GOM and count based methods are not dissimilar (Li et al., 2019). These dissimilarities, and also how the use of such complex algorithms would not be practical in clinical settings, may explain why more elaborate measures have not been adopted more widely.

1.2. Does the choice of AL scoring method matter for predicting health?

Methodological disparities in the scoring of the AL index across studies not only makes comparison of results challenging, but also hampers progress towards the use of AL as an early diagnostic screener and / or therapeutic target. This is an important theoretical and empirical matter that has not been subjected to the type of systematic investigation one might expect. Our review of the literature identified only two studies to date (Li et al., 2019; Seplaki et al., 2005) that have explicitly examined this issue.

Seplaki et al. (2005) compared the predictive utility of 9 different AL scoring algorithms (including count of high-risk quartiles/deciles, GOM, z-scores, and two-tailed quartiles/deciles, derived from two biomarker panels) in identifying poor health outcomes. Results were remarkably similar irrespective of the method used to calculate AL. They did, however, suggest that two-tailed formulations, or measures that retain the continuous properties of the biological variables, might be preferred going forward. A separate investigation by Li et al. (2019) involving the NHANES cohort compared the utility of 5 scoring algorithms (i.e. count of high-risk quartiles, z-scores, logistic regression, factor analysis, GOM) in predicting self-reported health, diabetes and hypertension. The best predictive performance was provided by a measure utilising the standardised coefficients from multivariable logistic regressions to weight individual biomarkers. However, similar to canonical correlation, health outcomes were used in the development of the scoring algorithm, which hampers replication across other populations. Li et al. (2019) recommended the original method of high-risk quartiles as a good alternative.

1.3. Should we incorporate sex differences in the calculation of the al index?

To date, despite evidence of sex differences within the stress response (Juster et al., 2019; Verma et al., 2011), and in the downstream biological parameters associated with the stress response (Freire et al.,

2020; Goldman et al., 2004; Santos-Lozada and Howard, 2018; Stoney et al., 1988; Yang and Kozloski, 2011), there is no consensus regarding the incorporation of sex-specific risk values in the calculation of AL measures. Seplaki et al. (2006) were the first to suggest deriving an AL score using sex-specific risk definitions, with a number of others adopting this approach since (Castagné et al., 2018; Christensen et al., 2018, 2019; Duru et al., 2012; Gustafsson et al., 2012; Robertson et al., 2015). Notwithstanding the potential import of biological sex differences, a recent review of NHANES revealed that only 1/21 AL studies incorporated sex-specific risk definitions (Duong et al., 2017), and to date, no study has compared the predictive utility of an AL measures including and excluding sex-specific risk definitions with alternate calculations.

1.4. Should we incorporate medications in the calculation of the AL Index?

Whether and how doctor-prescribed medication should be incorporated into the development of the AL index is another bone of contention (Lipowicz et al., 2014). Medications can act on the biological systems to reduce observed values, reducing risk of disease (Seeman et al., 2002, 2004). On the other hand, wear and tear may already have occurred if there is a need to medicate, and this excess risk should be accounted for when calculating the score. Varied approaches to the resolution of this problem exist including: scoring an individual at high risk if taking a medication which deflates values on the biomarker (Seeman et al., 2014); adding 'medication use' as a covariate in regression models (Piazza et al., 2018); or increasing individual biomarker values in an attempt at increasing prediction accuracy (e.g. systolic and diastolic blood pressure adjusted by adding 10mmHG and 5mmHG, respectively if taking anti-hypertensives) (Robertson and Watts, 2016). To date, no study has compared the predictive utility of AL measures including medication use compared with alternate formulations.

1.5. The present study

The present study builds upon the work of Seplaki et al. (2005) and Li et al. (2019) in several important ways. Firstly, we use a sample from a nationally representative cohort of community-dwelling older persons, whereas Seplaki et al. (2005) used a militaristic sample with heavy male bias, and Li et al. (2019) used an all-female sample. Secondly, we use a number of objective measures of physical health functioning (i.e. walking speed, grip strength) as the criterion variables in the analysis. Thirdly, we incorporate sex-specific and clinical risk definitions. Finally, we examine whether included medication usage improves the predictive accuracy of AL.

2. Methods

2.1. Sample

This analysis used data from the first wave of The Irish Longitudinal Study on Ageing (TILDA), a nationally representative prospective study of over 8175 persons aged 50 years and over living in the Republic of Ireland. Sampling involved a three-stage selection process with the Irish Geodirectory as the sampling frame, where residential addresses were divided into geographical clusters, of which 640 were selected based on area level socio-economic status and location. Additional information on the sampling frame and study design is available in detail elsewhere (Whelan and Savva, 2013). Data was collected through computer assisted personal interview (CAPI) carried out by a trained interviewer, a leave behind self-completion questionnaire (SCQ), and a comprehensive clinic-based, nurse-administered health assessment. These assessments included the collection of blood samples for biomarker assessment and a battery of cognitive and physical tests. Ethical approval for the study was provided by the Faculty of Health Science Research

Ethics board in Trinity College Dublin, while informed consent is obtained from all respondents during data collection. In total, 5894 people completed either the centre or home-based health assessment. Respondents who were missing on any of the biomarkers were excluded for this study, leaving a total analytical case base of 4,477. A flow chart summarising inclusion criterion for analysis is provided in Supplementary Fig. 1.

2.2. Physiological parameters

A total of 12 biomarkers across the cardiovascular, metabolic, renal and immune systems were used to construct the various AL indices examined in this study. Biomarkers were selected to reflect the most commonly chosen systems in the AL literature (Juster et al., 2010). Two of the biomarkers - pulse wave velocity (PWV) and cystatin C (Cysc) - are relatively novel with respect to AL formulation but their inclusion was theoretically motivated. For example, there is evidence to suggest that acute and chronic stress may induce arterial dysfunction as indexed using PWV (Logan et al., 2012; Matsumura et al., 2019). Similarly, Cysc may serve as a preferred/complementary biomarker of kidney functioning compared with creatinine in older cohorts as it is less affected by declining muscle mass (Canney et al., 2018). The component biomarkers and a description of their primary function is provided in Supplementary Table 1.

In terms of the cardiovascular system, the average of two measurements of seated Systolic Blood pressure (SBP), Diastolic Blood pressure (DBP) and Resting Heart Rate (RHR) were obtained separated by a 1-minute interval using an automatic digital oscillometric blood pressure monitor (OMRONTM, M10-IT). An average of two measurements between the carotid and femoral arteries was obtained using a Vicorder system, representing PWV (Skidmore Medical Ltd, Bristol, UK), a non-invasive gold standard method (Whelan and Savva, 2013). Metabolic system biomarkers included waist-hip ratio (WHR), body mass index (BMI) glycosylated haemoglobin (HbA1c), total cholesterol (TChol), and high-density lipoprotein (HDL). Weight was measured using a SECA electronic floor scales, and height was measured using a SECA 240 wall mounted measuring rod. BMI was calculated by dividing weight in kgs by height in metres squared. SECA measuring tapes were used to record WHR with measurements taken to the nearest millimetre. TChol was obtained from non-fasting blood samples, with venepuncture performed with a butterfly and a green 21-gauge Vacutainer needle and stored in EDTA tubes. A direct determination of HDL was performed using PEG-modified enzymes and dextran sulphate. HbA1c reflects blood sugar levels and metabolic functioning in the 8–12 weeks prior to blood sampling and were analysed by reverse-phase action exchange chromatography using an ADAMS A1c HA-8180 V analyser.

C-Reactive Protein (CRP) serves as our singular marker of immunological dysregulation as it was the only inflammatory marker available in the full TILDA cohort at baseline. It was measured in non-fasting blood serum on a Roche Cobas c 701 analyser using an immunoturbidimetric assay for the in vitro quantitative determination of CRP. Detection limits ranged from $0.3 - 350$ mg/l. The two renal markers included Creatinine (Crea) and Cysc, which were measured simultaneously from frozen plasma. Crea was measured using an enzymatic method traceable to isotope-dilution mass spectrometry (Roche Creatinine plus ver.2, Roche Diagnostics, Basel Switzerland), whilst Cysc was measured using a second-generation particle enhanced immunoturbidimetric assay (Roche Tina-quant™) on a Roche Cobas 701 analyzer. This assay has a measuring range of 0.40–6.80 mg/L and is traceable to the European reference standard material (ERM-DA471/IFCC). Log transformations were applied to CRP and PWV to account for right-skewed distributions.

2.3. Calculation of summary measures

High-risk quartiles (Classic Method): To re-create the original and most commonly employed scoring algorithm across studies (Seeman et al., 1997), dichotomous indicator variables were created for each biomarker. HDL was reverse coded such that higher values represented increased risk. Empirically defined high-risk thresholds were distinguished based on the distribution of that biomarker in the sample; "1" was assigned to values falling above the 75th percentile of the distribution for each marker, and "0" was assigned to values below this threshold. These biomarkers were summed to create an index, with higher scores reflecting higher AL burden. **Two-tailed high-risk quartiles:** Seplaki et al. (2005) suggested that including risk factors at both high and low ends of the risk continuum may be more informative than simply using high-risk quartiles. To re-create this score, a value of "1" was assigned to values above the 75th percentile and below the 25th percentile of the distribution, and a value of "0" was assigned for values that fell intermediate of these thresholds. A sum of these biomarkers represented two-tailed physiological dysregulation.

Following the above methodologies, two further algorithms were created employing decile cut-points, to measure risk at more extreme levels of the distribution. **High-risk Deciles:** "1" was assigned for values above the 90th percentile, "0" was assigned to all values below this threshold, and all biomarkers were then summed. **Two-tailed high-risk deciles:** "1" was assigned for values above the 90th percentile and below 10th percentile for the measure incorporating both ends of the distribution, and "0" assigned to all values that lay intermediate between these thresholds, and scores on the individual biomarkers were then summed.

Z-scores: To create a score that retained the continuous properties of each physiological variable, each biomarker was standardised to a mean of 0 and a standard deviation of 1, and then summed to generate an overall AL score expressed in standard deviation units (Daly et al., 2019; Hawkley et al., 2011; Levine and Crimmins, 2014). **System Weighted:** It is common in the AL literature for some biological systems to be more heavily represented than others, of which metabolic system biomarkers tend to be most populous (Juster et al., 2010). A system-weighted score was therefore created to reduce this bias. Firstly, biomarkers were dichotomised following the classic methodology (Seeman et al., 1997). Secondly, each dysregulated biomarker was weighted according to the total number of biomarkers per system. System risk indices were computed as the proportion of individual biomarkers for each system for which participant values fell into high-risk quartiles (Gruenewald et al., 2012). For example, in the present study, the cardiovascular system was comprised of four biomarkers (SBP, DBP, RHR, PWV). Therefore, the sum of dysregulated biomarkers within this system were divided by 4 and expressed as a proportion ranging from $0.0 - 1.0$. As CRP is the only immunological biomarker, no scaling occurred, hence "0" indicated no risk, whilst "1" indicated high risk. Finally, an AL score was computed as the sum of these weighted systems. This methodology has been used previously and ensures that each system is equally represented in the overall AL index (Gruenewald et al., 2012; Piazza et al., 2018; Read and Grundy, 2014).

Sex-specific cut-points: To account for variability in biomarker values according to sex, each of these 6 scoring algorithms were re-created using sex-specific cut-points (Table 2). Therefore, to create the sex-specific count of high-risk quartiles measure, individuals falling above the 75th percentile were identified using the sample distributions for men and women separately, and scores were then pooled.

High-risk quartiles (incorporating medications): To allow for the possibility that medication use was masking high values, prescription drugs were incorporated into an additional algorithm based on the classic methodology, with individuals taking medication automatically reclassified as high-risk across affected biomarkers. In TILDA, medication usage was ascertained as part of the household interview, where respondents were asked to retrieve the medicinal packaging of any

regularly taken medications and interviewers transcribed the brand/generic name of the prescription medication. The medications data was then coded in-house by pharmacologists according to Anatomic Therapeutic Classification (ATC) codes (Whelan and Savva, 2013). HbA1c was coded as high risk if individuals were currently prescribed medications of codes for insulin or analogues (A10), SBP was classified as high risk if individuals were taking anti-hypertensive medication (C02, C03, C09). RHR was recoded as high risk if taking beta-blockers (C07) or calcium blockers (C08) and finally, HDL was recoded as high risk if prescribed lipid modifying agents (C10).

Clinical: Finally, we derived a measure of AL using recognised risk definitions according to clinical guidelines (Allen et al., 2019; Borrell et al., 2010; Gruenewald et al., 2012; Prag and Richards, 2018). Clinical cut-points for each biomarker are provided in Supplementary Table 2.

2.4. Health outcomes

We utilised two objectively measured health outcomes – walking speed and grip strength (described below) - and to enable comparison with Seplaki et al. (2005), a general self-reported health measure was included.

2.4.1. Walking speed

In older adults, walking speed has been shown to be a good indicator of overall physical function (Viccaro et al., 2011), and a predictor for hospitalisation (Studenski et al., 2003) and higher morbidity (Guralnik et al., 2000). In TILDA, gait measurements were taken using a 4.88-metre computerised walkway with embedded pressure sensors (GAITRite, CIR Systems Inc., New York, NY). Participants completed two walks at their normal walking speed. The average of the two readings represents the overall walking speed expressed in centimetres travelled per second (cm/s).

2.4.2. Grip strength

Low grip strength has been robustly associated with functional decline and mortality in older adults (Bohannon, 2008) and serves as a proxy for the overall strength of the musculoskeletal system. Two measures were taken from the dominant hand using a Baseline® Hydraulic Hand Dynamometer, and the highest of the two readings in kilograms (kgs) per square inch represented the measure of grip strength used in the analysis.

2.4.3. Self-rated health

Self-rated health (SRH) is a commonly assessed item in health surveys (Manor et al., 2000), and can provide insight into the perceptions that an individual has of their own health status relevant to their peers. SRH has previously been associated with decreased physical performance (Perez-Zepeda et al., 2016) and increased mortality risk (Jylha, 2009). Previous studies have reported that lower SRH predicts higher AL (Vie et al., 2014; von Thiele et al., 2006). SRH was obtained during the CAPI with the item "In general, compared to other people your age, would you say your health is…" rated on a five-point scale of excellent, very good, good, fair, and poor.

3. Statistical analysis

All analyses were conducted in Stata (v.15; StataCorp, 2017), using TILDA dataset v.1.8.0. The outcome measures were regressed separately on each of the 14 AL scoring algorithms, adjusting for age and sex, using ordinary least squares regression with respect to walking speed and grip strength, and ordinal logistic regression for SRH. Overall model fit was assessed using the proportion of variance explained (R^2 for OLS, pseudo R^2 for ordinal logistic regression) and the Bayesian Information Criterion (BIC). An obvious advantage of using information measures is that one can compare the goodness of fit of non-nested models (Williams, 2019) as the different AL scoring algorithms are

essentially just reformulations of the same underlying set of 12 component biomarkers. The extent to which one model is preferred over another depends on the magnitude of the difference between the information measures. Higher values of R^2 and lower values of BIC indicate better model fit. BIC values were obtained using the FITSTAT package in Stata (Scott Long and Freese, 2014). Following guidelines proposed by Raftery (1995), we noted a BIC difference > 10 to indicate a better fitting model than the standard scoring algorithm. To reduce the large number of potential contrasts between the different AL scoring algorithms, we take as our reference, the overall (i.e. non-sex-specific) quartile-based risk score (hereinafter referred to as classic method) as it is the one that has been predominantly used in the AL literature.

In order to provide a comparison with Seplaki et al. (2005), binary indicator variables were created for each of the health outcomes using cut-points guided by previous literature. Grip strength was dichotomised at < 37 kg for men, and < 21 kg for women (Sallinen et al., 2010). Walking speed was dichotomised at < 120 cm/s (Studenski et al., 2011). SRH was recoded as "1" if respondents rated their health as fair/poor, "0" otherwise (Seplaki et al., 2005). Each of the binary health outcomes were treated as dependent variables in separate logistic regression models, controlling for age and sex. The area under the Receiver Operator Characteristic curve (AUC) was calculated for each health outcome for each of the 14 different AL scoring algorithms using the ROCTAB procedure (Pepe et al., 2009). Higher AUC values indicate greater classification accuracy. Comparison of AUC estimates to test the statistical significance of the difference between the classic AL algorithm and all alternative algorithms were investigated using the ROC-COMP procedure (Cleves, 2002). Finally, as age is the strongest predictor of health, we also examined which of the AL algorithms was most strongly correlated with chronological age.

4. Results

Table 1 reports sample characteristics. The mean age of the sample was 61.8 years and female respondents accounted for 53.7 % of the sample. Mean walking speed was 136.1 cm/s (SD = 20.4), and mean grip strength was 27.5 kgs (SD = 9.88). 89.0 % of the sample reported that they were in excellent/very good/good health. In total, 45.4 % of the sample were taking prescribed medications that potentially impacted values of biomarkers included in this study. Respondents included in the study were younger than those who were missing biomarker data, more educated and had, on average, a lower average count of chronic diseases. No sampling weights were applied however, as the primary focus of this study was to compare the predictive utility of various AL scoring algorithms and not to infer the relationship of AL and the health outcomes to the population. The relationship between

Table 1
Sample characteristics.

Variable	In sample (N = 4477)		Missing (N = 3697)	
	Mean	SD	Mean	SD
Sex (n, %)				
Male	2071	46.3	1673	45.3
Female	2406	53.7	2024	54.7
Age	61.8	8.4	66.3	10.8
Education (n, %)				
Primary	959	21.4	1545	41.8
Secondary	1882	42.1	1381	37.4
Tertiary	1635	36.5	768	20.8
Chronic diseases	1.6	1.4	1.8	1.5
Medications (n, %)	2033	45.4	697	18.85
Walking speed	136.1	20.4	132.5	22.0
Grip strength	27.5	9.88	25.6	9.59
SRH (n, %)				
Excellent, very good, good	3978	89.0	2922	79.2
Fair, poor	494	11.1	766	20.8

Table 2
Univariate descriptives of individual biomarkers by sex.

Biomarker	Mean	SD	Min	10th p	25th p	Median	75th p	90th p	Max	p
Systolic blood pressure	134.67	19.49	78.50	111.00	121.00	133.50	146.50	159.50	220.00	
Male	138.95	18.28	83.00	117.50	126.50	137.00	150.00	162.00	220.00	< 0.001
Female	130.99	19.75	78.50	107.50	117.00	129.00	143.00	158.00	208.50	
Diastolic blood pressure	82.30	11.06	51.50	68.50	75.00	82.00	89.50	96.50	132.00	
Male	83.72	10.88	53.50	70.50	76.50	83.50	90.50	97.00	132.00	< 0.001
Female	81.09	11.06	51.50	67.00	73.50	80.50	88.00	95.50	125.50	
Resting heart rate	68.78	11.37	33.00	55.50	60.50	67.50	75.50	83.50	135.50	
Male	67.81	12.00	33.00	54.00	59.50	66.50	74.50	83.50	135.50	< 0.001
Female	69.62	10.73	35.00	56.50	62.00	68.50	76.00	83.50	116.50	
Pulse wave velocity *	2.33	0.19	1.67	2.09	2.20	2.32	2.45	2.58	2.99	
Male	2.37	0.18	1.82	2.14	2.25	2.37	2.49	2.61	2.99	< 0.001
Female	2.29	0.19	1.67	2.05	2.15	2.28	2.42	2.54	2.94	
High density lipoprotein (reversed)	−1.56	0.44	−3.82	−2.13	−1.80	−1.49	−1.23	−1.07	−0.48	
Male	−1.36	0.34	−3.50	−1.80	−1.54	−1.31	−1.11	−0.99	−0.63	< 0.001
Female	−1.72	0.44	−3.82	−2.30	−1.98	−1.66	−1.41	−1.22	−0.48	
Cholesterol	5.15	1.05	1.80	3.80	4.50	5.10	5.80	6.50	12.20	
Male	4.88	1.05	1.80	3.50	4.20	4.90	5.60	6.20	9.10	< 0.001
Female	5.39	1.00	2.20	4.10	4.70	5.30	6.00	6.70	12.20	
Glycosylated haemoglobin	33.06	5.45	16.00	28.00	30.00	32.00	35.00	38.00	94.00	
Male	33.47	5.72	17.00	28.00	30.00	32.00	35.00	39.00	88.00	< 0.001
Female	32.72	5.18	16.00	28.00	30.00	32.00	34.00	37.00	94.00	
BMI	28.51	4.82	17.08	23.00	25.24	27.99	31.12	34.58	61.86	
Male	29.10	4.33	18.33	24.16	26.28	28.69	31.45	36.50	59.28	< 0.001
Female	28.00	5.15	17.08	22.33	24.47	27.24	30.69	34.66	61.86	
Waist hip ratio	0.90	0.09	0.63	0.79	0.84	0.90	0.96	1.01	1.29	
Male	0.96	0.07	0.74	0.77	0.92	0.96	1.00	1.07	1.28	< 0.001
Female	0.85	0.07	0.63	0.77	0.81	0.85	0.90	0.94	1.29	
Cystatin C	0.97	0.21	0.46	0.77	0.84	0.93	1.05	1.18	3.95	
Male	0.99	0.22	0.56	0.79	0.86	0.95	1.06	1.19	3.95	< 0.001
Female	0.95	0.20	0.46	0.76	0.83	0.92	1.03	1.18	2.97	
Creatinine	79.21	19.90	40.00	59.00	66.00	77.00	88.00	101.00	486.00	
Male	89.09	20.21	49.00	71.00	78.00	86.00	96.00	109.00	486.00	< 0.001
Female	70.71	15.13	40.00	56.00	61.00	68.00	77.00	87.00	231.00	
C-Reactive protein *	0.69	0.77	0.00	0.00	0.00	0.47	1.17	1.79	6.28	
Male	0.65	0.75	0.00	0.00	0.00	0.43	1.12	1.68	5.73	0.003
Female	0.72	0.78	0.00	0.00	0.00	0.51	1.22	1.84	6.28	

Total sample n = 4477; male = 2071, female = 2406. p = significance level determined from two sample t-test, comparisons of mean values within biomarkers between male and female respondents. *log transformed.

age and each of the AL scoring algorithms was assessed using Spearman's correlations, reported in Supplementary Table 3. The measure incorporating medications was most strongly correlated with age (r = 0.34), whilst the two-tailed measures revealed weak correlations with age (e.g. overall quartile (r = 0.05), overall decile based (r = 0.06)).

Univariate statistics and percentile cut-points for each of the 12 component biomarkers defining the AL construct are presented in Table 2. Two-sample t-tests revealed marked differences in biomarker values according to sex. Specifically, male respondents were characterised by significantly higher mean values for 9/12 biomarkers, whilst female respondents had significantly higher mean values for the latter 3 (RHR, CRP, Total Cholesterol). Supplementary Table 4 presents univariate statistics in respect of each of the 14 AL scoring algorithms used in the present study. Tables 3–5 report the model fit statistics for each AL scoring algorithm and the difference in model fit relative to the classic scoring algorithm for walking speed, grip strength and SRH, respectively. In this context, higher values for R^2 and lower values for BIC indicate a better fitting model compared with the standard model. In general, the results indicate that the choice of scoring algorithm has relatively modest effects on the proportion of variance explained in the health outcomes.

4.1. Walking speed

Looking first at the results for walking speed, Table 3 shows that the proportion of variance explained by the classic AL scoring algorithm (including age and sex) was 20.3 %, compared with 17.8 % for the worst (overall two-tailed quartiles and overall two-tailed deciles) and

Table 3
Model Fit Statistics for different Allostatic Load Scoring Algorithms predicting Walking Speed.

	R^2	DIFF R^2	BIC	DIFF BIC
High-risk quartiles				
Overall	0.203***	*REF*	38344.70	*REF*
Sex-specific	0.209***	0.006	38310.73	−33.97[a]
Two-tailed quartiles				
Overall	0.178***	−0.025	38479.84	135.14
Sex-specific	0.182***	−0.021	38449.62	113.30
High-risk deciles				
Overall	0.205***	0.002	38329.40	−10.51[a]
Sex-specific	0.207***	0.004	38322.91	−20.94[a]
Two-tailed deciles				
Overall	0.178***	−0.025	38467.57	136.23
Sex-specific	0.180***	−0.023	38458.66	124.48
System weighted				
Overall	0.207***	0.004	38304.19	−22.17[a]
Sex-specific	0.208***	0.005	38328.90	−27.40[a]
Continuous				
Overall	0.204***	0.001	38337.43	−7.27
Sex-specific	0.205***	0.002	38331.71	−12.99[a]
Medications	0.212***	0.009	38296.41	−48.29[a]
Clinical	0.199***	−0.004	38362.26	17.56

Note: Adjusted for age and sex. R^2 = proportion of variance explained. R^2 Diff = gain in R^2 compared with the classic AL scoring algorithm. BIC = Bayesian Information Criterion. BIC Diff = Difference in BIC compared with the classic AL scoring algorithm. *p < 0.05, **p < 0.01, *** p < 0.001.

[a] Substantial decrease in BIC compared with the classic AL scoring algorithm.

Table 4
Model Fit Statistics for different Allostatic Load Scoring Algorithms predicting Grip Strength.

	R^2	DIFF R^2	BIC	DIFF BIC
High-risk quartiles				
Overall	0.6078	*REF*	28968.53	*REF*
Sex-specific	0.6077	−0.0001	28968.90	0.37
Two-tailed quartiles				
Overall	0.6078	0.0000	28967.96	−0.57
Sex-specific	0.6078	0.0000	28968.06	−0.46
High-risk deciles				
Overall	0.6077	−0.0001	28969.34	0.82
Sex-specific	0.6076	−0.0002	28969.86	1.33
Two-tailed deciles				
Overall	0.6084**	0.0006	28960.84	−7.68
Sex-specific	0.6080*	0.0002	28965.64	−2.89
System weighted				
Overall	0.6077	−0.0001	28969.43	0.90
Sex-specific	0.6076	−0.0002	28969.74	1.21
Continuous				
Overall	0.6082*	0.0004	28963.277	−5.25
Sex-specific	0.6082**	0.0004	28963.083	−5.44
Medications	0.6077	−0.0001	28969.45	0.92
Clinical	0.6081*	0.0003	28964.77	−3.76

Note: Adjusted for age and sex. R^2 = proportion of variance explained. R^2 Diff = gain in R^2 compared with the classic AL scoring algorithm. BIC = Bayesian Information Criterion. BIC Diff = Difference in BIC compared with the classic AL scoring algorithm. *p < 0.05, **p < 0.01, *** p < 0.001.
[a]Substantial decrease in BIC compared with the classic AL scoring algorithm.

Table 5
Model Fit Statistics for different Allostatic Load Scoring Algorithms predicting self-rated health.

	R^2	DIFF R^2	BIC	DIFF BIC
High-risk quartiles				
Overall	0.011***	*REF*	12156.79	*REF*
Sex-specific	0.013***	0.002	12133.15	−23.65[a]
Two-tailed quartiles				
Overall	0.002***	−0.009	12256.37	99.58
Sex-specific	0.003***	−0.008	12248.24	91.44
High-risk deciles				
Overall	0.011***	0.000	12149.79	−7.00
Sex-specific	0.011***	0.000	12158.46	1.67
Two-tailed deciles				
Overall	0.002***	−0.009	12262.80	106.00
Sex-specific	0.002***	−0.009	12262.81	106.02
System weighted				
Overall	0.012***	0.001	12139.15	−17.64[a]
Sex-specific	0.011***	0.000	12149.24	−7.56
Continuous				
Overall	0.010***	0.001	12163.989	7.20
Sex-specific	0.010***	0.001	12158.346	1.55
Medications	0.019***	0.008	12052.43	−104.37[a]
Clinical	0.008***	−0.003	12183.01	26.21

Note: Adjusted for age and sex. R^2 = proportion of variance explained. R^2 Diff = gain in R^2 compared with the classic AL scoring algorithm. BIC = Bayesian Information Criterion. BIC Diff = Difference in BIC compared with the classic AL scoring algorithm. *p < 0.05, **p < 0.01, *** p < 0.001.
[a] Substantial decrease in BIC compared with the classic AL scoring algorithm.

21.2 % for the best (high-risk quartiles incorporating medications) fitting models. The sex-specific high-risk quartiles, high-risk decile measures, weighted measures, sex-specific z-scores, and the algorithm incorporating medications led to a strong improvement in model fit relative to the classic AL scoring algorithm, according to the BIC guidelines proposed by Raftery (1995). However, it should be acknowledged that the gain in R^2 was typically very modest and amounted to, at most, +0.9 %. The measures calibrated according to sex performed marginally better when predicting walking speed than

the overall measures. The two-tailed measures by contrast performed poorly, with lower values of R^2 and large increases in BIC compared with the classic algorithm.

4.2. Grip strength

For grip strength, Table 4 shows that the proportion of variance explained by the classic AL scoring algorithm (including age and sex) was 60.8 %, and this value did not vary substantially across the other measures; +0.06 % for the best fitting model (overall two-tailed deciles) compared with the classic model. Two-tailed decile measures, clinical and z-scores were the only AL scoring algorithms that exhibited statistically significant associations with grip strength independently of age and sex. The strength of these relationships were small, however, and the direction of these relationships differed; negative associations were noted for overall (β = −0.16, 95 % CI: −0.27, −0.06) and sex-specific (β = −0.11, 95 % CI: −0.02, −0.01) two-tailed deciles, whilst positive associations were noted for the clinical measure (β = 0.13, 95 % CI: 0.02, 0.23), as well as the overall (β = 0.05, 95 % CI: 0.01, 0.09) and sex-specific (β = 0.05, 95 % CI: 0.01, 0.09) z-score measures.

4.3. Self-rated health

Table 5 shows that the classic AL measure accounted for 1.1 % of the variance in SRH, compared with 0.02 % for the worst fitting models (overall high-risk deciles, two-tailed high-risk decile measures) and 1.9 % for the best fitting model (high-risk quartiles incorporating medications). The sex-specific high-risk quartile measure, overall weighted measure, the measure incorporating medications, and the clinical measure led to strong improvements in BIC values compared to the classic AL index, whilst the two-tailed measures performed substantially worse.

4.4. Area under the receiver operating characteristic curve (AUC)

Supplementary Table 5 reports the AUC estimates for each of the AL scoring algorithms separately for each of the three binary health outcomes. Results were consistent with what was observed when using the continuous measures. Notably, the two-tailed measures (quartiles and deciles) performed worse than the classic algorithm when predicting walking speed and SRH, whilst all measures predicted poor grip strength to the same degree of accuracy as the classic score (Fig. 1). Although a minority of AL algorithms reached statistical significance with grip strength, no substantial differences in AUC were noted between these and the classic algorithm. In addition, the count of high-risk quartiles measure incorporating medications predicted fair/ poor SRH better than the classic measure.

5. Discussion

The present study assessed the explanatory utility of a multitude of AL scoring algorithms for predicting a number of objective and subjective age-related health outcomes in an older adult population residing in Ireland. In accordance with the findings from a previous study involving a Taiwanese sample (Seplaki et al., 2005), we found that the choice of AL scoring algorithm has a relatively modest impact in terms of variance explanation and classification accuracy in the prediction of a number of important health outcomes. Although the differences between the 14 scoring algorithms were not pronounced in this study, there were, nevertheless, some subtle nuances observed.

AL was strongly associated with walking speed across all 14 scoring algorithms independently of age and sex, supporting prior findings regarding the predictive utility of AL for functional decline (Karlamangla et al., 2002; Read and Grundy, 2014; Singer et al., 2004; Szanton et al., 2005). Although the differences were small when compared to the classic scoring algorithm, the count of high-risk quartiles incorporating
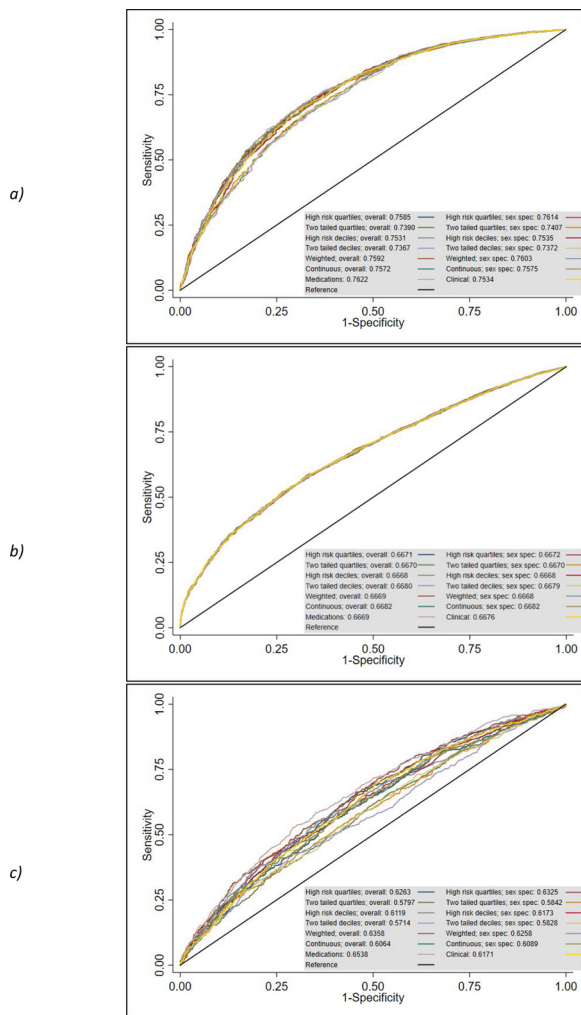
Fig. 1. AUC estimates for different scoring algorithms, predicting a) slow walking speed, b) weak grip strength and c) poor self-rated health.

medications performed marginally better. In contrast, few significant associations were observed between measures of AL and grip strength, and no algorithm fit the data significantly better than the classic method. Although few have investigated this relationship directly, grip strength shows strong age-related decline and is an important component of many frailty indices that have been found to be negatively associated with AL (Gruenewald et al., 2012; Szanton et al., 2005). The null associations of the majority of AL scoring algorithms, and the varied direction of the relationship noted for those algorithms which did statistically predict grip strength is surprising, but not without precedent. Freire et al. (2020) for example, arrived at the rather counter-intuitive conclusion that higher AL burden (classic measure) was associated with higher grip strength in a small sample of older community-dwelling adults in Brazil (n = 256), yet they hypothesised this contradictory result was due to survival bias. Perhaps the predominantly null results arise as there are other potential confounders that we are not controlling for in this analysis, which may help preserve muscle strength at older ages (e.g. height, occupational class etc).

In respect of SRH, results were more variable across scoring algorithms, although differences were small in absolute terms. Similar to walking speed, the algorithm incorporating medication use was the best predictor of SRH in terms of model fit. This pattern of results begs the obvious question as to whether the better predictive performance of a measure incorporating medication when using a SRH measure is simply a reflection of the individual being more aware of their 'health state' by virtue of consciously taking prescribed medications on a regular basis.

5.1. Use of one vs two-tailed risk

Notably, the AL scoring algorithms incorporating risk in both tails of the distribution of each biomarker underperformed relative to the classic one-tailed measure in this study. These findings align with results from the Hawaii Personality and Health Cohort (n = 470), where two-tailed measures of AL were less effective than one-tailed measures in predicting SRH and depressive symptoms. In stark contrast, however, Seplaki et al. (2005) found that two-tailed measures of risk accounted for more variance across self-reported health measures than the classic AL index. This discrepant pattern of results across studies could be explained by differences in the biomarker panels employed to represent AL. Seplaki et al.'s (2005) study included primary mediators of the stress response such as cortisol, in which both high and low values can characterize certain syndromes and diseases (Fries et al., 2005) whereas, similar to the present study, Hampson et al. (2009) did not include any markers of neuroendocrine dysregulation. Moreover, in the context of the current study, it is difficult to see how low scores for HbA1c or cystatin C would increase risk of disease, which reinforces the need to think critically about whether risk on particular biomarkers is best conceptualised as linear or curvilinear when developing AL indices.

5.2. The importance of considering sex differences

Beyond the objective of comparing the predictive utility of various AL algorithms, this study found significant differences across mean biomarker values between male and female individuals. Furthermore, the AL algorithms which were calibrated relative to sex generally performed marginally better in terms of model fit than those derived from the pooled sample with respect to walking speed and SRH. Acknowledging that men and women may react differently to stress, both psychologically and biologically (Stroud et al., 2002); we recommend that future investigations continue to examine the use of sex-specific cut-points, particularly where the distributions on the component biomarkers differ according to sex, or indeed, where there might be large sex differences (Juster et al., 2016), in the outcome measure (e.g. depression).

5.3. The importance of considering medications

The vexed question of whether to include medications in the calculation of the AL measure is a recurring feature in many critiques of the AL literature (Duong et al., 2017; Howard and Sparks, 2016; Rodriquez et al., 2019). This issue has not always been afforded the attention it arguably deserves in many empirical studies, presumably because many studies either do not capture this information, rely on self-report, or perhaps due to operational difficulties in how to treat the data if using continuous measures (e.g. z-scores). We found that including medications led to small improvements in performance relative to the classic measure with respect to walking speed and SRH. It was also the algorithm which was most strongly correlated with age. Nevertheless, we applied only one of a number of potential techniques for including medications in the development of the AL index, and future studies may wish to consider alternative approaches such as including medications as a covariate (Piazza et al., 2018) or adjusting biomarker values upward to adjust for their deflationary effect on the biomarker of interest (Robertson and Watts, 2016).

6. Strengths and limitations

This study is among the first to examine and compare the effects of various AL scoring algorithms in a community-dwelling cohort of older persons. It benefits from having a large sample pulled from a nationally representative cohort, and gold standard measures of objective physiological functioning measured using trained nurses according to

standard operating protocols. It advances the current knowledge base through the inclusion of these objective markers of physical fitness, by considering sex as an important variable in the determination of AL scores, and by accounting for medication use as a potential indicator of physiological wear and tear. Despite these strengths, this study does not have any neuroendocrine markers at baseline which is an important limitation as they are hypothesised to play a central role in the stress response. Furthermore, whilst TILDA is a nationally representative cohort study, the sample employed in this paper is not generalisable to other populations. Similarly, the results from this study are not generalisable to studies employing biomarkers different to those used here. It should also be acknowledged that although significant associations were found between measures of AL and objective health outcomes in this study, the estimated models control only for the effects of age and sex. Instead, this study provides evidence for the construction of a summary score from individual biological components to predict poor health outcomes, without assuming a causal or directional relationship.

7. Conclusion

This study was motivated by the lack of empirical insight into the effects the choice of AL scoring algorithm has on the predictive utility of poor health outcomes. Seeman et al. (2010) claimed that their original method did not represent a gold standard, yet the results of this study suggest that this classic method performs well. We therefore echo the conclusion of others (Berger et al., 2018; Li et al., 2019), that given that this is the technique used in the vast majority of papers, perhaps the notion of harmonising international work around this scoring algorithm should be considered. Nonetheless, the findings of this paper can help allay researchers concerns that the choice of scoring algorithm makes a large difference to the results, but whether the composition of biomarkers employed to reflect physiological dysregulation does remains to be elucidated.

Funding

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.psyneuen.2020. 104789.

References

Allen, A.M., Thomas, M.D., Michaels, E.K., Reeves, A.N., Okoye, U., Price, M.M., Hasson, R.E., Syme, S.L., Chae, D.H., 2019. Racial discrimination, educational attainment, and biological dysregulation among midlife African American women. Psychoneuroendocrinology 99, 225–235. https://doi.org/10.1016/j.psyneuen.2018. 09.001.

Beckie, T.M., 2012. A systematic review of allostatic load, health, and health disparities. Biol. Res. Nurs. 14 (4), 311–346. https://doi.org/10.1177/1099800412455688.

Berger, M., Juster, R.-P., Westphal, S., Amminger, G.P., Bogerts, B., Schiltz, K., Bahn, S., Steiner, J., Sarnyai, Z., 2018. Allostatic load is associated with psychotic symptoms and decreases with antipsychotic treatment in patients with schizophrenia and first-episode psychosis. Psychoneuroendocrinology 90, 35–42. https://doi.org/10.1016/j. psyneuen.2018.02.001.

Bohannon, R.W., 2008. Hand-grip dynamometry predicts future outcomes in aging adults. J Geriatr Phys Ther.

Borrell, L.N., Dallo, F.J., Nguyen, N., 2010. Racial/ethnic disparities in all-cause mortality in U.S. adults: the effect of allostatic load. Public Health Rep 125 (6), 810–816. https://doi.org/10.1177/003335491012500608.

Canney, M., Sexton, D.J., O'Leary, N., Healy, M., Kenny, R.A., Little, M.A., O'Seaghdha, C.M., 2018. Examining the utility of cystatin C as a confirmatory test of chronic kidney disease across the age range in middle-aged and older community-dwelling adults. J. Epidemiol. Community Health 72 (4), 287. https://doi.org/10.1136/jech-2017-209864. LP-293.

Castagné, R., Garès, V., Karimi, M., Chadeau-Hyam, M., Vineis, P., Delpierre, C., Kelly-Irving, M., Alenius, H., Avendano, M., Baltar, V., Bartley, M., Barros, H., Bochud, M., Carmeli, C., Carra, L., Costa, G., Courtin, E., Donkin, A., D'Errico, A., et al., 2018. Allostatic load and subsequent all-cause mortality: which biological markers drive the relationship? Findings from a UK birth cohort. Eur. J. Epidemiol. 33 (5), 441–458. https://doi.org/10.1007/s10654-018-0364-1.

Christensen, D.S., Flensborg-Madsen, T., Garde, E., Hansen, A.M., Pedersen, J.M., Mortensen, E.L., 2018. Parental socioeconomic position and midlife allostatic load: a study of potential mediators. BMC Public Health 18 (1), 1029. https://doi.org/10. 1186/s12889-018-5956-x.

Christensen, D.S., Flensborg-Madsen, T., Garde, E., Hansen, A.M., Mortensen, E.L., 2019. Big five personality traits and allostatic load in midlife. Psychol. Health 1–18. https:// doi.org/10.1080/08870446.2019.1585851.

Cleves, M.A., 2002. From the help desk: comparing areas under receiver operating characteristic curves from two or more probit or logit models. Stata J. 2 (3), 301–313. https://doi.org/10.1177/1536867X0200200307.

Crimmins, E.M., Kim, J.K., Seeman, T.E., 2009. Poverty and biological risk: the earlier "aging" of the poor. J. Gerontol. https://doi.org/10.1093/gerona/gln010.

Daly, M., Sutin, A.R., Robinson, E., 2019. Perceived weight discrimination mediates the prospective association between obesity and physiological dysregulation: evidence from a population-based cohort. Psychol. Sci. https://doi.org/10.1177/ 0956797619849440. 956797619849440.

Duong, M.T., Bingham, B.A., Aldana, P.C., Chung, S.T., Sumner, A.E., 2017. Variation in the calculation of allostatic load score: 21 examples from NHANES. J. Racial Ethn. Health Disparities 4 (3), 455–461. https://doi.org/10.1007/s40615-016-0246-8.

Duru, O.K., Harawa, N.T., Kermah, D., Norris, K.C., 2012. Allostatic load burden and racial disparities in mortality. J. Med. Assoc. 104 (1–2), 89–95. https://doi.org/10. 1016/S0027-9684(15)30120-6.

Freire, A., do, N.F., Barbosa, J.F., de, S., Pereira, D.S., Gomes, Cdos S., Guerra, R.O., 2020. Allostatic load and stress biomarkers in a sample of community-dwelling older adults. Arch. Gerontol. Geriatr. 87, 104006. https://doi.org/10.1016/j.archger.2019. 104006.

Fries, E., Hesse, J., Hellhammer, J., Hellhammer, D.H., 2005. A new view on hypo-cortisolism. Psychoneuroendocrinology 30 (10), 1010–1016. https://doi.org/10. 1016/j.psyneuen.2005.04.006.

Glei, D.A., Goldman, N., Wu, C.H., Weinstein, M., 2013. Does exposure to stressors predict changes in physiological dysregulation? Ann. Behav. Med. https://doi.org/10.1007/ s12160-013-9485-7.

Goldman, N., Weinstein, M., Cornman, J., Singer, B., Seeman, T., Goldman, N., Chang, M.C., 2004. Sex differentials in biological risk factors for chronic disease: estimates from population-based surveys. J. Womens Health (Larchmt) 13 (4), 393–403. https://doi.org/10.1089/154099904323087088.

Gruenewald, T.L., Seeman, T.E., Ryff, C.D., Karlamangla, A.S., Singer, B.H., 2006. Combinations of biomarkers predictive of later life mortality. Proc. Natl. Acad. Sci. U. S. A. 103 (38), 14158–14163. https://doi.org/10.1073/pnas.0606215103.

Gruenewald, T.L., Karlamangla, A.S., Hu, P., Stein-Merkin, S., Crandall, C., Koretz, B., Seeman, T.E., 2012. History of socioeconomic disadvantage and allostatic load in later life. Soc. Sci. Med. 74 (1), 75–83. https://doi.org/10.1016/j.socscimed.2011.09. 037.

Guralnik, J.M., Ferrucci, L., Pieper, C.F., Leveille, S.G., Markides, K.S., Ostir, G.V., Studenski, S., Berkman, L.F., Wallace, R.B., 2000. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J. Gerontol. A Biol. Sci. Med. Sci. 55 (4), 221–231. https://doi.org/10.1093/gerona/55.4.m221.

Gustafsson, P.E., Janlert, U., Theorell, T., Westerlund, H., Hammarstrom, A., 2012. Social and material adversity from adolescence to adulthood and allostatic load in middle-aged women and men: results from the Northern Swedish Cohort. Ann. Behav. Med. 43 (1), 117–128. https://doi.org/10.1007/s12160-011-9309-6.

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. J. Health Psychol. 14 (2), 232–241. https://doi.org/10. 1177/1359105308100207.

Hawkley, L.C., Lavelle, L.A., Berntson, G.G., Cacioppo, J.T., 2011. Mediators of the re-lationship between socioeconomic status and allostatic load in the Chicago Health, Aging, and Social Relations Study (CHASRS). Psychophysiology 48 (8), 1134–1145. https://doi.org/10.1111/j.1469-8986.2011.01185.x.

Howard, J.T., Sparks, P.J., 2016. Does allostatic load calculation method matter? Evaluation of different methods and individual biomarkers functioning by race/ ethnicity and educational level. Am. J. Hum. Biol. 28 (5), 627–635. https://doi.org/ 10.1002/ajhb.22843.

Hwang, A.C., Peng, L.N., Wen, Y.W., Tsai, Y.W., Chang, L.C., Chiou, S.T., Chen, L.K., 2014. Predicting all-cause and cause-specific mortality by static and dynamic mea-surements of allostatic load: a 10-year population-based cohort study in Taiwan. J. Am. Med. Dir. Assoc. 15 (7), 490–496. https://doi.org/10.1016/j.jamda.2014.02. 001.

Johnson, S.C., Cavallaro, F.L., Leon, D.A., 2017. A systematic review of allostatic load in relation to socioeconomic position: poor fidelity and major inconsistencies in bio-markers employed. Soc. Sci. Med. 192, 66–73. https://doi.org/10.1016/j.socscimed. 2017.09.025.

Juster, Robert-Paul, de Torre, Margot Barbosa, Kerr, Philippe, Kheloui, Sarah, Bourdon, Olivier, et al., 2019. Sex Differences and Gender Diversity in Stress Responses and Allostatic Load Among Workers and LGBT People. Curr Psychiatry Rep. https://link. springer.com/content/pdf/10.1007%2Fs11920-019-1104-2.pdf.

Juster, R.P., McEwen, B.S., Lupien, S.J., 2010. Allostatic load biomarkers of chronic stress and impact on health and cognition. Neurosci. Biobehav. Rev. 35 (1), 2–16. https:// doi.org/10.1016/j.neubiorev.2009.10.002.

Juster, R.-P., Pruessner, J.C., Desrochers, A.B., Bourdon, O., Durand, N., Wan, N., Tourjman, V., Kouassi, E., Lesage, A., Lupien, S.J., 2016. Sex and gender roles in relation to mental health and allostatic load. Psychosom. Med. 78 (7), 788–804. https://doi.org/10.1097/PSY.0000000000000351.

Jylha, M., 2009. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Soc. Sci. Med. 69 (3), 307–316. https://doi.org/10.1016/j.socscimed.2009.05.013.

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. J. Clin. Epidemiol. 55 (7), 696–710 internal-pdf://7.245.239.112/Karlamangla-2002-Allostatic load as a predicto.pdf.

Karlamangla, A.S., Singer, B.H., Seeman, T.E., 2006. Reduction in allostatic load in older adults is associated with lower all-cause mortality risk: MacArthur studies of successful aging. Psychosom. Med. 68 (3), 500–507. https://doi.org/10.1097/01.psy.0000221270.93985.82.

LeBron, A.M.W., Schulz, A.J., Mentz, G.B., Israel, B.A., Stokes, C.A., 2019. Social relationships, neighbourhood poverty and cumulative biological risk: findings from a multi-racial US urban community. J. Biosoc. Sci. 1–18. https://doi.org/10.1017/s002193201900004x.

Levine, M.E., Crimmins, E.M., 2014. A comparison of methods for assessing mortality risk. Am. J. Hum. Biol. 26 (6), 768–776. https://doi.org/10.1002/ajhb.22595.

Li, Y., Rosemberg, M.S., Dalton, V.K., Lee, S.J., Seng, J.S., 2019. Exploring the optimal allostatic load scoring method in women of reproductive age. J. Adv. Nurs. https://doi.org/10.1111/jan.14014.

Lipowicz, A., Szklarska, A., Malina, R.M., 2014. Allostatic load and socioeconomic status in Polish adult men. J. Biosoc. Sci. 46 (2), 155–167. https://doi.org/10.1017/s0021932013000345.

Logan, J.G., Barksdale, D.J., Carlson, J., Carlson, B.W., Rowsey, P.J., 2012. Psychological stress and arterial stiffness in Korean Americans. J. Psychosom. Res. 73 (1), 53–58. https://doi.org/10.1016/j.jpsychores.2012.04.008.

Manor, O., Matthews, S., Power, C., 2000. Dichotomous or categorical response? Analysing self-rated health and lifetime social class. Int. J. Epidemiol. 29 (1), 149–157. https://doi.org/10.1093/ije/29.1.149.

Matsumura, K., Noguchi, H., Rolfe, P., Yamakoshi, T., Matsuoka, Y., 2019. Differential effect of two mental stress tasks on arterial stiffness. Japan. Psychol. Res. 61 (4), 249–261. https://doi.org/10.1111/jpr.12235.

McEwen, B.S., Stellar, E., 1993. Stress and the individual. Mechanisms leading to disease. Arch. Intern. Med. 153 (18), 2093–2101.

Merkin, S.S., Basurto-Davila, R., Karlamangla, A., Bird, C.E., Lurie, N., Escarce, J., Seeman, T., 2009. Neighborhoods and cumulative biological risk profiles by race/ethnicity in a national sample of U.S. adults: NHANES III. Ann. Epidemiol. 19 (3), 194–201. https://doi.org/10.1016/j.annepidem.2008.12.006.

Pepe, M.S., Longton, G., Janes, H., 2009. Estimation and comparison of receiver operating characteristic curves. Stata J. https://doi.org/10.1177/1536867x0900900101.

Perez-Zepeda, M.U., Belanger, E., Zunzunegui, M.V., Phillips, S., Ylli, A., Guralnik, J., 2016. Assessing the validity of self-rated health with the short physical performance battery: a cross-sectional analysis of the international mobility in aging study. PLoS One 11 (4), e0153855. https://doi.org/10.1371/journal.pone.0153855.

Petrovic, D., Pivin, E., Ponte, B., Dhayat, N., Pruijm, M., Ehret, G., Ackermann, D., Guessous, I., Younes, S.E., Pechere-Bertschi, A., Vogt, B., Mohaupt, M., Martin, P.Y., Paccaud, F., Burnier, M., Bochud, M., Stringhini, S., 2016. Sociodemographic, behavioral and genetic determinants of allostatic load in a Swiss population-based study. Psychoneuroendocrinology 67, 76–85. https://doi.org/10.1016/j.psyneuen.2016.02.003.

Piazza, J.R., Stawski, R.S., Sheffler, J.L., 2018. Age, daily stress processes, and allostatic load: a longitudinal study. J. Aging Health. https://doi.org/10.1177/0898264318788493. 898264318788493.

Prag, P., Richards, L., 2018. Intergenerational social mobility and allostatic load in Great Britain. J. Epidemiol. Community Health. https://doi.org/10.1136/jech-2017-210171.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociol. Methodol. 25, 111–163. https://doi.org/10.2307/271063.

Read, S., Grundy, E., 2014. Allostatic load and health in the older population of England: a crossed-lagged analysis. Psychosom. Med. 76 (7), 490–496. https://doi.org/10.1097/PSY.0000000000000083.

Robertson, T., Watts, E., 2016. The importance of age, sex and place in understanding socioeconomic inequalities in allostatic load: evidence from the Scottish Health Survey (2008–2011). BMC Public Health 16 (1), 126. https://doi.org/10.1186/s12889-016-2796-4.

Robertson, T., Benzeval, M., Whitley, E., Popham, F., 2015. The role of material, psychosocial and behavioral factors in mediating the association between socioeconomic position and allostatic load (measured by cardiovascular, metabolic and inflammatory markers). Brain Behav. Immun. 45, 41–49. https://doi.org/10.1016/j.bbi.2014.10.005.

Rodriquez, E.J., Kim, E.N., Sumner, A.E., Nápoles, A.M., Pérez-Stable, E.J., 2019. Allostatic load: importance, markers, and score determination in minority and disparity populations. J. Urban Health 96 (1), 3–11. https://doi.org/10.1007/s11524-019-00345-5.

Sallinen, J., Stenholm, S., Rantanen, T., Heliovaara, M., Sainio, P., Koskinen, S., 2010. Hand-grip strength cut points to screen older persons at risk for mobility limitation. J. Am. Geriatr. Soc. 58 (9), 1721–1726. https://doi.org/10.1111/j.1532-5415.2010.03035.x.

Santos-Lozada, A.R., Howard, J.T., 2018. Using allostatic load to validate self-rated health for Racial/Ethnic groups in the United States. Biodemogr. Soc. Biol. 64 (1), 1–14. https://doi.org/10.1080/19485565.2018.1429891.

Scott Long, J., Freese, J., 2014. Regression Models for Categorical Dependent Variables Using Stata, 3rd ed. Stata Press. Stata Press.

Seeman, T., 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. Arch. Intern. Med. 157 (19), 2259–2268.

Seeman, T.E., Singer, B.H., Ryff, C.D., Dienberg Love, G., Levy-Storms, L., 2002. Social relationships, gender, and allostatic load across two age cohorts. Psychosom. Med. 64 (3), 395–406. https://doi.org/10.1097/00006842-200205000-00004.

Seeman, T.E., Crimmins, E., Huang, M.-H., Singer, B., Bucur, A., Gruenewald, T., Berkman, L.F., Reuben, D.B., 2004. Cumulative biological risk and socio-economic differences in mortality: MacArthur Studies of Successful Aging. Soc. Sci. Med. 58 (10), 1985–1997. https://doi.org/10.1016/S0277-9536(03)00402-7.

Seeman, T., Epel, E., Gruenewald, T., Karlamangla, A., McEwen, B.S., 2010. Socio-economic differentials in peripheral biology: cumulative allostatic load. Ann. N. Y. Acad. Sci. 1186, 223–239. https://doi.org/10.1111/j.1749-6632.2009.05341.x.

Seeman, M., Stein Merkin, S., Karlamangla, A., Koretz, B., Seeman, T., 2014. Social status and biological dysregulation: the "status syndrome" and allostatic load. Soc. Sci. Med. 1982 (118), 143–151. https://doi.org/10.1016/j.socscimed.2014.08.002.

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

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.

Singer, B., Ryff, C.D., Seeman, T., 2004. Operationalizing allostatic load. Allostasis, Homeostasis, and the Costs of Physiological Adaptation. Cambridge University Press, pp. 113–149. https://doi.org/10.1017/CBO9781316257081.007.

StataCorp, 2017. Stata Statistical Software: Release 15. StataCorp LLC.

Stoney, C.M., Matthews, K.A., McDonald, R.H., Johnson, C.A., 1988. Sex differences in lipid, lipoprotein, cardiovascular, and neuroendocrine responses to acute stress. Psychophysiology 25 (6), 645–656. https://doi.org/10.1111/j.1469-8986.1988.tb01902.x.

Stroud, L.R., Salovey, P., Epel, E.S., 2002. Sex differences in stress responses: social rejection versus achievement stress. Biol. Psychiatry 52 (4), 318–327. https://doi.org/10.1016/S0006-3223(02)01333-1.

Studenski, S., Perera, S., Wallace, D., Chandler, J.M., Duncan, P.W., Rooney, E., Fox, M., Guralnik, J.M., 2003. Physical performance measures in the clinical setting. J. Am. Geriatr. Soc. 51 (3), 314–322.

Studenski, S., Perera, S., Patel, K., Rosano, C., Faulkner, K., Inzitari, M., Brach, J., Chandler, J., Cawthon, P., Connor, E.B., Nevitt, M., Visser, M., Kritchevsky, S., Badinelli, S., Harris, T., Newman, A.B., Cauley, J., Ferrucci, L., Guralnik, J., 2011. Gait speed and survival in older adults. Jama 305 (1), 50–58. https://doi.org/10.1001/jama.2010.1923.

Szanton, S.L., Gill, J.M., Allen, J.K., 2005. Allostatic load: a mechanism of socioeconomic health disparities? Biol. Res. Nurs. 7 (1), 7–15. https://doi.org/10.1177/1099800405278216.

Upchurch, D.M., Stein, J., Greendale, G.A., Chyu, L., Tseng, C.H., Huang, M.H., Lewis, T.T., Kravitz, H.M., Seeman, T., 2015. A longitudinal investigation of race, socio-economic status, and psychosocial mediators of allostatic load in midlife women: findings from the study of women's health across the nation. Psychosom. Med. 77 (4), 402–412. https://doi.org/10.1097/PSY.0000000000000175.

Vasunilashorn, S., Glei, D.A., Weinstein, M., Goldman, N., 2013. Perceived stress and mortality in a Taiwanese older adult population. Stress 16 (6), 600–606. https://doi.org/10.3109/10253890.2013.823943.

Verma, Rohit, Balhara, Yatan Pal Singh, Gupta, Chandra Shekhar, 2011. Gender differences in stress response: Role of developmental and biological determinants. Ind. Psychiatry J. https://www.ncbi.nlm.nih.gov/pubmed/22969173.

Viccaro, L.J., Perera, S., Studenski, S.A., 2011. Is timed up and go better than gait speed in predicting health, function, and falls in older adults? J. Am. Geriatr. Soc. 59 (5), 887–892. https://doi.org/10.1111/j.1532-5415.2011.03336.x.

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). Soc. Sci. Med. 117, 1–9. https://doi.org/10.1016/j.socscimed.2014.07.019.

von Thiele, U., Lindfors, P., Lundberg, U., 2006. Self-rated recovery from work stress and allostatic load in women. J. Psychosom. Res. 61 (2), 237–242. https://doi.org/10.1016/j.jpsychores.2006.01.015.

Weinstein, M., Goldman, N., Hedley, A., Yu-Hsuan, L., Seeman, T., 2003. Social linkages to biological markers of health among the elderly. J. Biosoc. Sci. 35 (3), 433–453. https://doi.org/10.1017/S0021932003004334.

Whelan, B.J., Savva, G.M., 2013. Design and methodology of the Irish longitudinal study on ageing. J. Am. Geriatr. Soc. 61 (Suppl 2), S265–8. https://doi.org/10.1111/jgs.12199.

Williams, R., 2019. Scalar Measures of Fit: Pseudo R2 and Information Measures (AIC & BIC). https://www3.nd.edu/~rwilliam/stats3/L05.pdf. .

Yang, Y., Kozloski, M., 2011. Sex differences in age trajectories of physiological dysregulation: inflammation, metabolic syndrome, and allostatic load. J. Gerontol. A Biol. Sci. Med. Sci. 66 (5), 493–500. https://doi.org/10.1093/gerona/glr003.