Review article

Measuring allostatic load: Approaches and limitations to algorithm creation

Jason T. Carbone, Jenifer Clift, Nicholas Alexander

School of Social Work, Wayne State University, 5447 Woodward Blvd, Detroit, MI 48202, USA

ARTICLE INFO

Keywords:
Allostatic load
Biological dysregulation
Biomarkers
Scoping review

ABSTRACT

Objective: Allostatic load literature has proliferated over the past three decades, and a growing body of research demonstrates that higher levels of allostatic load are associated with a wide range of negative physical and mental health outcomes. However, there remain significant challenges with operationalization of the concept. A scoping review of the methods employed to create an allostatic load algorithm was conducted and recommendations for future research with an orientation towards advancing clinical application of the theory are discussed.

Methods: A search of seven electronic databases (PubMed, PsycINFO, Social Work Abstracts, Social Service Abstracts, Social Sciences Citation Index (Web of Science), Sociological Abstracts, Scopus) was completed with the search term “allostatic load.” Studies were reviewed, and if they met the inclusion criteria, data was extracted, compiled, and presented in the narrative, table, and figures.

Results: The initial searches yielded 5280 results with the final sample of 395 non-duplicate articles that met the inclusion criteria. More than half (52.5%) of all included publications employed biomarker cutoffs based on the high-risk quartiles of the sample distribution, 11.1% employed the sum of at-risk clinical scores, and the remainder of studies utilized a range of different algorithms.

Conclusion: Allostatic load literature has grown at an exponential rate in recent years, but researchers continue to operationalize the concept via algorithms that may have limited utility moving forward. More nuanced statistical approaches are emerging and should be considered, as should a shift towards an approach that can provide additional clinical utility.

1. Introduction

Allostatic load, which represents the cumulative wear and tear of long-term exposure to stress [47], has become a major focus in the literature since the term was first coined nearly three decades ago. The premise that exposure to stress leads to negative health outcomes is not new. It can be linked to the earliest definitions of stress [61] as well as the concept that stress accumulates in the body over time [62]. Yet the approach taken by McEwen and Stellar [47] was novel. Specifically, the idea that individual biomarkers could be measured in clinical settings, and a score calculated that could be utilized as a predictive tool for negative health outcomes, was a significant contribution to the literature. Since then, research on allostatic load has associated it with a wide range of health outcomes including—but not limited to—cardiovascular disease, diabetes, musculoskeletal disorders, cancer, periodontal disease, mood and anxiety disorder, and post-traumatic stress disorder (see [3,29] for more detailed reviews).

Despite the wide-ranging application of allostatic load theory to better understand health, there remains a lack of consensus related to the operationalization of the theoretical concept [4,48]. One of the more impactful consequences of this heterogeneity is that it limits the ability to make comparisons across studies [4]. The current study presents the findings of a scoping review of the algorithms employed to operationalize allostatic load in order to gain a more holistic picture of the variance in how allostatic load is operationalized. First, a brief overview of the literature is provided for general historical context of allostatic load operationalization and algorithms. Second, the results of the scoping review are presented. Finally, recommendations for future research with an orientation towards practical, clinical applications are discussed.

2. The development of allostatic load theory and research

Allostatic load theory is an evolution of stress-related research that dates to the early 20th century. It is rooted in the work of Selye [61], who coined the term stress, and Cannon [10], who created the word homeostasis. For Cannon [10], homeostasis as about biological systems...
maintaining ‘consistency through stability’ ([68], p. 17), which enabled them to balance the impact of internal and external stimuli in order to sustain life. Selye [63] believed that the homeostatic model did not adequately describe biological responses to extreme stimuli (i.e., stress). Selye created the term heterostasis to represent the new normal that is achieved when organisms adapt to changes in the environment [23]. Yet Selye [62] recognized that when these changes are extreme, they can be maladaptive. He was the first to suggest that stress can accumulate over time and that this can lead to negative biological outcomes—a process he labeled General Adaptation Syndrome (GAS).

Sterling and Eyer [69] published a seminal work wherein they introduced the term allostatic, which was a theoretical leap beyond homeostasis theory. Allostatic proposes that longer-term stability occurs as a result of change and incorporates concepts from evolutionary theory by stating that an organism’s goal is not to remain constant, but to maintain fitness under natural selection and that this process inherently requires the ability to adapt and change [57]. The allostatic model is also a predictive one, allowing organisms to learn from previous experiences in order to maintain fitness [57]. This is a key concept when considering maladaptive allostatic processes and their connections with, for example, mental health disorders such as posttraumatic stress disorder.

McEwen and Stellar [47] extended these earlier lines of work to create allostatic load theory, which integrates the allostatic and chronic stress perspectives into a cause and effect framework [34]. Allostatic load theory proposes a linear progression that begins with exposure to chronic stress and trigger biological changes that, over time, result in negative health outcomes. More specifically, long-term exposure to chronic stress can lead to changes in the management of stress hormones (i.e., cortisol, DHEA-S, epinephrine, and norepinephrine as well as some cytokines). Collectively, these biomarkers are referred to as primary mediators [34,54]. Long-term dysregulation in primary mediators can result in changes at the cellular level, known as primary effects. Over time, this can lead to subclinical changes in what are known as secondary outcomes. Examples include blood pressure, cholesterol, glucose levels, fibrinogen, and albumin. Finally, changes in secondary outcomes lead to the development of tertiary outcomes, which are the disease outcomes and include physical disease (e.g., cardiovascular) and mental health disorders (e.g., depression, anxiety). This final disease stage is known as allostatic overload and occurs when environmental challenges exceed an individual’s abilities to cope [34].

McEwen and Seeman [46] were the first to test allostatic load theory through empirical research with the MacArthur Foundation Study on Successful Aging. This study laid the groundwork for future allostatic load research and set the standard for the biomarkers that would comprise an allostatic load measure. Early studies with this data set (e.g., [36,58-60]) focused on a combination of ten biomarkers (DHEA-S, urinary cortisol, norepinephrine, epinephrine, systolic blood pressure, diastolic blood pressure, waist-to-hip ratio, serum HDL-cholesterol, total cholesterol-to-HDL cholesterol, and HbA1c) to create an allostatic load index (ALI).

The National Health and Nutrition Examination Survey (NHANES) was another early adopter of biomarker collection that contributed to allostatic load research. The third wave of NHANES, known as NHANES III, collected data from 1988 through 1994 and was the first wave of NHANES to include biomarkers [13]. While researchers who utilized the MacArthur Foundation Study data tended to employ the same list of biomarkers, when reviewing NHANES III data, the focus began to change with NHANES. Researchers began to add or delete individual biomarkers from allostatic load scales, resulting in the use of a range of biomarkers from as few as six (e.g., [43]) to as many as 14 (e.g., [1,82]). The biomarkers include some combination of C-reactive protein, plasma fibrinogen, urinary albumin, waist circumference, SBP, DBP, serum triglycerides, HDL cholesterol, fasting glucose, BMI, glycated hemoglobin (HbA1c), serum homocysteine, peak flow, and creatinine clearance. It should be noted that all of these biomarkers are secondary outcomes. It has been noted that the lack of primary mediators (i.e., epinephrine, norepinephrine, dopamine, cortisol) limits the utility of this data set to elucidate the causal pathways that link factors to disease outcomes via an allostatic load model [53]. That is, there are delays in the sequential changes from primary mediator to primary effects to secondary outcome to tertiary outcomes, as described above in the theory. Blending data from multiple points in time (i.e., primary mediators and secondary outcomes) obscures this process and leads to a less precise understanding of how and when biological changes occur. In addition, few researchers have differentiated between primary mediators and secondary outcomes in their algorithm formation (e.g., [64]).

The breadth of biomarkers collected expanded, most notably, with the Midlife in the United States (MIDUS) study. MIDUS is a longitudinal study that began in 1995 to explore the role of psychological, behavioral, and social factors on health as individuals age [6]. From 2005 to 2009, biomarkers were collected from a subsample of the original study population. While MIDUS collected a wide range of biomarker data, most researchers utilize up to 24 biomarkers and group them into seven biological systems (e.g., [67,83]). These biomarkers—and their respective biological systems—include systolic blood pressure, diastolic blood pressure, resting pulse (cardiovascular system); 12-h overnight urine epinephrine, 12-h overnight urine norepinephrine (sensory nervous system); low-frequency heart variability, high-frequency heart variability, the standard deviation of R-R (heartbeat-to-heartbeat) intervals, the root mean square of successive differences (parasympathetic nervous system); 12-h overnight urinary measure of cortisol, serum DHEA-S (HPA-axis); plasma C-reactive protein, fibrinogen, serum IL-6, soluble adhesion molecules E-selectin, intercellular adhesion molecule 1 (inflammatory system); glucose, insulin resistance, HbA1c (glucose metabolism); HDL-cholesterol, LDL-cholesterol, triglycerides, BMI, WHR (lipid metabolism).

Other research studies have collected a wide and varied array of biomarkers, resulting in variability across the literature. Examples of other biomarkers included in allostatic load summary scores include thyroxin [2]; insulin resistance [34,84,85]; transaminases, IGF-1, IL-8, [9]; IL-10 [40,52]; tumor-necrosis factor alpha [26,32,56]; D-dimer, [5,78]; aldosterone [49,77]. This variance in biomarker utilization is symptomatic of the lack of consensus as to which biomarkers should or should not be included [48]. Arguably even more problematic, and the focus of the current scoping review, is the variance in how biomarker values are aggregated into an informative and predictive allostatic load algorithm. The limited research that compares varied approaches to a summary measure has found differing results with modest effects on some—but not all—outcomes [36,64].

3. Creating allostatic load algorithms

There are multiple considerations for developing an allostatic load algorithm. In order to take data from multiple biomarkers and aggregate them into a single allostatic load score, the following decisions about the algorithm structure need to be considered: (1) Should biomarkers be treated as continuous or dichotomous variables? (2) If dichotomized, should a sample distribution or clinical cutoff approach be taken? (3) How should biological systems be taken into consideration when calculating allostatic load? (4) Once a scale is constructed, what score represents high allostatic load?

3.1. Dichotomized versus continuous biomarkers

Researchers must first determine if the values of the biomarker will be treated as a standardized, continuous value (e.g., z-score) or if it will be converted to a dichotomous variable (1 = high level of biomarker dysregulation, 0 = low/no biomarker dysregulation). Allostatic load scales can be constructed based on dichotomized or continuous biomarkers, yet there is not a consensus as to which method is best [3]. Some researchers have noted that both methods present similar results [72,81].
while others suggest that the use of z-scores is preferred as they utilize the full continuum of the data and do not eliminate information [30,80]. Widom et al. [81] utilized and compared both methods. Although they state that the dichotomized variable—based on high-risk quartile—results are statistically significantly correlated with the z-score results, the two share <50% of their variance ($r = 0.69$, $p < 0.001$) (p.64) suggesting that the two approaches are, at least in part, measuring different aspects of biological dysregulation.

3.4. Applying allostatic load scales

If a dichotomous approach is taken, the researcher must assess if individual biomarkers should be treated as being within the normal range or outside of the normal range and, therefore, a marker of allostatic load. Typically, researchers utilize either a cutoff value based on the sample distribution or one based on a clinically determined value. Although there are multiple sample distribution approaches, utilizing the high-risk quartile is the most common. For this approach, the observations in the highest risk quartile of the sample distribution are coded as dysregulated and all others are considered to have a normal value. Some authors consider the observations with either the highest or the lowest 25% of the sample distribution, depending on the clinical role of that specific biomarker, as being in the high-risk quartiles. Some researchers divide the sample into the top 12.5% and lowest 12.5% for certain biomarkers such as cortisol since research suggests that both hypercortisolemia and hypocortisolemia are associated with negative health outcomes (e.g., [5,27]). Using the sample distribution to calculate the high-risk quartile can result in a large amount of variance across studies. Some examples of high-risk quartile cutoffs based on the sample distribution for systolic blood pressure include greater than or equal to 150 mmHg [19], greater than or equal to 130 mmHg [41], greater than or equal to 115 mmHg [31].

An alternative to high-risk quartiles is the use of clinically-established cutoffs for determining high-risk biomarkers. Researchers have noted that results using clinical cutoffs do not vary from those that use distributional approaches such as high-risk quartiles or z-scores (e.g., [12,51,80]). In spite of such statements, evidence in this area is not conclusive. For example, Ahrens et al. [1] found that among reproductive-age women, clinical cutoffs of allostatic load were more highly associated with allostatic load outcomes than quartile cutoffs. There are two drawbacks to this approach. First, clinical cutoffs have not been established for all biomarkers commonly used in allostatic load algorithms [35]. Second, the use of clinical cutoffs is not consistent with allostatic load theory. Theory posits that allostatic load manifests itself as a continuous variable (e.g., sum or average of z-scores, sum of dichotomous biomarkers) or a categorical variable. In categorizing the allostatic load scale, some researchers choose to utilize three categories (low, medium, high) while many dichotomize the indicator. In converting the variable from a continuous to a two or three category variable, researchers must determine an adequate cutoff point or cutoff points. Some select a number above the median score to focus on higher risk individuals (e.g., [76]), while others take a more liberal approach. Theill et al. [71] assessed the level of allostatic load among adolescents by utilizing ten biomarkers. With an allostatic load sample mean of one, they utilized a cutoff of two, which designated 35% of their sample as having high allostatic load. Given that allostatic load is conceptualized as representing pre-clinical dysregulation [44], a more liberal approach (i.e., lower cutoff) can be justified. Yet most cutoff points are not theoretically driven or justified. When researchers note a reason for selecting a specific cutoff, it tends to be based on general convention (e.g., [42]) or cutoffs from previous studies with the same data set (e.g., [15]).

4. The present study

To the authors’ knowledge, there has yet to be a recent, comprehensive review of the literature exploring the different ways in which researchers have constructed an allostatic load algorithm. Juster et al. [34] provide a road map for this work along with both a strong theoretical summary and review of the literature. Yet the extensive growth in allostatic load literature since then necessitates a additional review. There are more limited reviews of this nature, but they focus on allostatic load measures related to a specific stressor (e.g., employment, [44]), specific populations (e.g., adolescents, [79]), or targeted health outcomes (e.g., women’s brain health, [38]). In addition, some researchers have looked at how allostatic load is operationalized across publications with a specific data set (e.g., [20]). This study attempts to take a more holistic view of the science and asks the broad question: what is the general state of the literature as it relates to how researchers operationalize allostatic load? The findings of our scoping review are presented below followed by recommendations for the future trajectory of allostatic load research including how research in this area can be tailored to be more clinically meaningful.

5. Materials and methods

5.1. Protocol and registration

The PRISMA Extension for Scoping Reviews [73] checklist was employed to inform the collection, compilation, and reporting of the present study. The study was pre-registered with the Center for Open Science’s Open Science Framework database (link masked for review).

5.2. Eligibility criteria

For inclusion, studies must have met the following basic criteria: (1) utilized human subjects (i.e., animal studies excluded), (2) explicitly stated and applied allostatic load theory (e.g., cumulative measures of biological dysregulation without any reference to allostatic load were excluded), (3) operationalized allostatic load via a measure that incorporated multiple biomarkers across multiple systems (e.g., studies only utilizing cortisol were excluded). Due to limited translation resources, only studies published in English were included and gray literature were also excluded. Reviews, editorials, and theoretical papers without measured biomarkers were all excluded as ineligible.

6. Information sources and search

Studies included in this review were identified from two searches of the literature. The initial search was completed in July of 2017. Given
that the McEwen and Stellar article that first utilized the term “allostatic load” was published in 1993, the search criteria was limited to articles published between 1993 and the then present date (i.e., July 2017). A subsequent search was conducted in February 2021 to update the previous results and limited the dates of publication to 2017 through February 8, 2021 (the date of the search).

Both searches (2017 and 2021) utilized the same approach and criteria. Seven electronic databases were searched (PubMed, PsycINFO, Social Work Abstracts, Social Service Abstracts, Social Sciences Citation Index (Web of Science), Sociological Abstracts, Scopus) with the term “allostatic load.” While more inclusive search terms could have been utilized (e.g., allostasis, biological dysregulation, cumulative risk), only “allostatic load” was employed as the goal was to focus on how researchers who specifically identify their research within the allostatic load theoretical framework decided to operationalize the concept.

6.1. Data extraction

Each study was reviewed to determine if it met the inclusion criteria. If it did, the following data elements were extracted into a database file: authors, year of publication, article title, list of biomarkers (and their respective biological systems if specifically identified in the study) included, allostatic load algorithm, source of the data, and sample size.

7. Results

7.1. Study selection and characteristics

Fig. 1 is a flow chart of the inclusion/exclusion process and includes the number of studies excluded at each phase. The combined (2017 and 2021) searches yielded 5280 publications. 2319 duplicates were excluded as were 2134 that did not meet inclusion criteria related to the use of human subjects or type of paper (e.g., theoretical papers, editorials, etc.). Of the 827 remaining studies, 432 were assessed as ineligible for not meeting the inclusion criteria with respect to operationalizing allostatic load as a composite score based on multiple biomarkers across multiple biological systems. The final analytic sample of publications that were eligible for inclusion was 395.

7.2. Trend in total number of publications

Fig. 2 displays the total number of publications per year from 1997 (i.e., the first year with publications that met the inclusion criteria) through February of 2021. It also includes a trend line showing the exponential growth in research that utilizes the allostatic load framework to measure biological dysregulation. The growth was slow at first, but beginning in 2013—with 22 publications and the first year to surpass the 20-publication milestone—growth continued at an exponential rate.

7.3. Data sources/studies

The final sample included 395 studies that utilized a total of 425

![Fig. 2. Trend in allostatic load publications (1997-February 2021).](image-url)
### Table 1
Examples of methods for creating composite allostatic load scores by study.

<table>
<thead>
<tr>
<th>Study</th>
<th>Range of sample sizes</th>
<th>Range of biomarkers</th>
<th>Methods creating allostatic load variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Health and Nutrition Examination Survey (NHANES) (United States)</td>
<td>(n = 85-38,000)</td>
<td>Biomarkers</td>
<td>Sum of at-risk clinical scores&lt;sup&gt;1,16&lt;/sup&gt; Sum of at-risk clinical scores (dichotomized)&lt;sup&gt;11,18&lt;/sup&gt; Sum of at-risk clinical scores (trichotomized)&lt;sup&gt;1&lt;/sup&gt; Sum of at-risk clinical &amp; high-risk quartile scores&lt;sup&gt;2,20&lt;/sup&gt; Sum of high-risk quartiles&lt;sup&gt;25,52&lt;/sup&gt; Sum of high-risk quartiles (dichotomized)&lt;sup&gt;41,53,54&lt;/sup&gt; Sum of high-risk quartiles (trichotomized)&lt;sup&gt;55&lt;/sup&gt; Sum of high-risk quartiles (sex-stratified)&lt;sup&gt;56&lt;/sup&gt; Sum of z-scores&lt;sup&gt;44,57&lt;/sup&gt; Mean of z-scores&lt;sup&gt;3&lt;/sup&gt; Other algorithms&lt;sup&gt;58-63&lt;/sup&gt;</td>
</tr>
<tr>
<td>Midlife Development in the United States (MIDUS) (United States)</td>
<td>(n = 76-1255)</td>
<td>Biomarkers</td>
<td>Mean of systems level scores based on sum of at-risk clinical scores&lt;sup&gt;73&lt;/sup&gt; Sum of systems-level based on biomarker high-risk quartile&lt;sup&gt;4-92&lt;/sup&gt; Sum of systems-level based on biomarker high-risk quartile (dichotomized)&lt;sup&gt;135&lt;/sup&gt; Sum of systems-level based on biomarker clinical and high-risk quartile&lt;sup&gt;24,95&lt;/sup&gt; Other algorithms&lt;sup&gt;96&lt;/sup&gt;</td>
</tr>
<tr>
<td>2000 Social Environment and Biomarkers of Aging Study (SEBAS) (Taiwan)</td>
<td>(n = 521-1023)</td>
<td>Biomarkers</td>
<td>Sum of at-risk clinical &amp; high-risk quartile scores&lt;sup&gt;97-98&lt;/sup&gt; Sum of at-risk clinical &amp; high-risk quartile scores (sex-stratified)&lt;sup&gt;65,94&lt;/sup&gt; Sum of high-risk quartiles&lt;sup&gt;100,101&lt;/sup&gt; Sum of high-risk quartile (sex-stratified)&lt;sup&gt;39&lt;/sup&gt; Sum of high-risk deciles&lt;sup&gt;102,103&lt;/sup&gt; Other algorithms&lt;sup&gt;99,104,105&lt;/sup&gt;</td>
</tr>
<tr>
<td>MacArthur Study of Successful Aging (MSSA) (United States)</td>
<td>(n = 171-874)</td>
<td>Biomarkers</td>
<td>Sum of high-risk quartiles&lt;sup&gt;96-110&lt;/sup&gt; Sum of high-risk quartiles (dichotomized)&lt;sup&gt;111&lt;/sup&gt; Sum of weighted biomarker values&lt;sup&gt;112,113&lt;/sup&gt;</td>
</tr>
<tr>
<td>Boston Puerto Rican Health Study (BPRHS) (Boston, Massachusetts)</td>
<td>(n = 787-1387)</td>
<td>Biomarkers</td>
<td>Sum of at-risk clinical score&lt;sup&gt;1,14,115&lt;/sup&gt; Sum of at-risk clinical scores (5 categories)&lt;sup&gt;117&lt;/sup&gt; Sum of at-risk clinical &amp; high-risk quartile scores&lt;sup&gt;118-123&lt;/sup&gt; Sum of high-risk quartiles&lt;sup&gt;124&lt;/sup&gt; Sum of high-risk quartiles (sex-stratified)&lt;sup&gt;125&lt;/sup&gt;</td>
</tr>
<tr>
<td>Jackson Heart Study (Jackson, Mississippi)</td>
<td>(n = 2670-5306)</td>
<td>Biomarkers</td>
<td>Sum of high-risk quartiles&lt;sup&gt;126&lt;/sup&gt; Average of biomarker z-scores&lt;sup&gt;27,128&lt;/sup&gt; Mean of systems-level scores</td>
</tr>
</tbody>
</table>

Table 1 (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Range of sample sizes</th>
<th>Range of biomarkers</th>
<th>Methods creating allostatic load variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>(n = 1263-6123)</td>
<td>Biomarkers</td>
<td>Sum of mean of biomarker z-scores&lt;sup&gt;239&lt;/sup&gt; Sum of systems-level scores based on mean of biomarker z-scores&lt;sup&gt;130,131&lt;/sup&gt;</td>
</tr>
<tr>
<td>Longitudinal Study of Aging (ELSA)</td>
<td>(n = 9-13)</td>
<td>Biomarkers</td>
<td>Sum of at-risk quartiles&lt;sup&gt;132-136&lt;/sup&gt; Sum of high-risk quartiles (sex-stratified)&lt;sup&gt;137,138&lt;/sup&gt; Other algorithms&lt;sup&gt;139&lt;/sup&gt;</td>
</tr>
<tr>
<td>Northern Swedish Cohort</td>
<td>(n = 12)</td>
<td>Biomarkers</td>
<td>Sum of at-risk clinical scores&lt;sup&gt;145-146&lt;/sup&gt;</td>
</tr>
<tr>
<td>Copenhagen Aging and Midlife Biobank Study (CAMB) (Denmark)</td>
<td>(n = 5512)</td>
<td>Biomarkers</td>
<td>Sum of at-risk clinical scores (dichotomized)&lt;sup&gt;148-150&lt;/sup&gt; Sum of at-risk clinical scores (trichotomized)&lt;sup&gt;170&lt;/sup&gt;</td>
</tr>
<tr>
<td>Whitehall II Study (United Kingdom)</td>
<td>(n = 563-7007)</td>
<td>Biomarkers</td>
<td>Sum of at-risk clinical scores (sex-stratified)&lt;sup&gt;171&lt;/sup&gt; Sum of at-risk clinical &amp; high-risk quartile scores&lt;sup&gt;72-179&lt;/sup&gt; Sum of at-risk clinical &amp; high-risk quartile scores (sex-stratified)&lt;sup&gt;169&lt;/sup&gt; Sum of at-risk clinical &amp; high-risk quartile scores (sex-stratified)&lt;sup&gt;174&lt;/sup&gt;</td>
</tr>
<tr>
<td>Other Studies&lt;sup&gt;3&lt;/sup&gt;</td>
<td>(n = 2-7007)</td>
<td>Biomarkers</td>
<td>Sum of at-risk clinical scores (dichotomized)&lt;sup&gt;155-156,211,238,274,354&lt;/sup&gt; Sum of at-risk clinical scores (trichotomized)&lt;sup&gt;155,211,238,274,354&lt;/sup&gt; Sum of high-risk quartiles&lt;sup&gt;155,160,180,314&lt;/sup&gt; Sum of high-risk quartiles (sex-stratified)&lt;sup&gt;155,327-338&lt;/sup&gt; Sum of high-risk quartiles (trichotomized)&lt;sup&gt;174&lt;/sup&gt; Sum of high-risk quartiles (sex-stratified)&lt;sup&gt;155,327-338&lt;/sup&gt; Sum of high-risk quartiles (trichotomized) (sex-stratified)&lt;sup&gt;339-340&lt;/sup&gt; Sum of z-scores&lt;sup&gt;355,264,305,341-353&lt;/sup&gt; Mean of z-scores&lt;sup&gt;3,21,23,27,34,354-361&lt;/sup&gt; Sum of systems-level based on biomarker high-risk quartiles&lt;sup&gt;155,362-366&lt;/sup&gt; Sum of systems-level based on biomarker high-risk quartiles of health controls&lt;sup&gt;97-369&lt;/sup&gt; Sum of systems-level based on biomarker Z-Scores&lt;sup&gt;2,370&lt;/sup&gt; Other algorithms&lt;sup&gt;45,155,160,190,371-399&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Note: Few of the included publications utilized clinical populations. The majority of publications were based on larger, community-based research studies (e.g., NHANES, MIDUS). Of the 134 publications from those labeled as “Other studies,” only 32 were identified as utilizing samples from clinical populations. Articles below that are marked with an asterisk (*) have multiple algorithms for allostatic load scales, therefore they are in the table multiple times.

France (2021),
Kang (2011),
Bruce et al. (2017a),
Akrivos et al. (2020),
Rollings and Evans (2019),
Sibille et al. (2017)*,
Mattei et al. (2010),
Kisler (2013),
Egorov et al. (2017),
Egorov et al. (2020),
274Hough et al. (2020),
374Jack-Roberts et al. (2017),
264Gallo et al. (2019),
374Prior et al. (2018),
375Berger et al. (2018),
376Pietrowski et al. (2019),
379Miskiewicz et al. (2019),
372McMillan et al. (2017),
374Currie et al. (2017),
373Wallace and Harville (2013),
352Mair, Cutchin, & Peek (2011),
353Langelana et al. (2007),
355von Kanel et al. (2003),
354Boneva et al. (2019),
355Prig and Richards (2019),
356Schenk et al. (2018),
357Silva et al. (2018),
358Tampubolon and Maharani (2018),
359Stephan et al. (2016),
360Vie et al. (2014),
361Hawkey et al. (2011),
362Berger et al. (2020),
363Currie et al. (2020),
364Pietrowski et al. (2020),
365Gallo et al. (2019),
366Pietrowski et al. (2019),
367McMillan et al. (2017),
368Currie et al. (2019),
369Egorov et al. (2020),
370Hough et al. (2020),
371Jack-Roberts et al. (2017),
372Merkin et al. (2020),
373Niño et Cai (2020),
374Forrester et al. (2019),
375Moon-Riley et al. (2019),
376Burkard et al. (2019),
377Rogers et al. (2020),
378Nobel et al. (2017),
379Vaccaro et al. (2017),
380Mauz, Jarczok, & Fischer (2016),
381O’Campo et al. (2016),
382Gale et al. (2015),
383Mauz, Jarczok, & Fischer (2015),
384Widom, Horan, & Brzustowicz (2015),
385Lipowicz, Śląska, & Malina (2014),
386Merkin et al. (2014),
387Ivaka et al. (2014),
388Rob- ertson, Popham, & Benzenal (2014),
389Evans & Fuller-Rowell (2013),
390Wallace et al. (2013),
391Orloch, Dacks, & Cicchetti (2011),
392Glover et al. (2010),
393Vaccaro et al. (2009),
394Examples of other studies include: 1958 National Child Development Study (NCDS) (Britain); CAMB (Denmark); Chicago Community Adult Health Study (Chicago, IL, USA); Chicago Health, Aging, and Social Relations Study (CHASRS); Coronary Artery Risk Development in Young Adults Study (CARDIA) (Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA, USA); En- glish Longitudinal Study of Aging (ELSA); Individual Development and Adap- tation (Sweden); Health 2000 Study (Finland); Jackson Heart Study (Jackson, Mississippi, USA); Northern Swedish Cohort. Additional articles were published with data collected outside of an identified and names research study.
allostatic load algorithms (i.e., some studies utilized multiple algorithms). The final sample did not constitute 395 separate studies that collected biomarkers. As displayed in Table 1, the ten research studies with five or more publications represent 39% of all the included allostatic load publications. Furthermore, NHANES and MIDUS, combined, represent nearly one-quarter (24.1%) of all the included allostatic load publications.

7.4. Allostatic load algorithm

A chart displaying the frequency of specific algorithms is displayed in Fig. 3. The most common allostatic load algorithm is the sum of biomarker high-risk quartiles, which was employed in over half the studies (52.5%). The next most common method was the sum of at-risk clinical scores (11.1%), followed by the sum of a combination of clinical and high-risk quartile values (7.3%). Systems level scores based on high-risk quartile were the next most common algorithms (6.8%). Other, less common methods included sum of biomarker \( Z \)-scores (4.2%), mean of biomarker \( Z \)-scores (4.0%) and other systems-level scores based on biomarker at-risk clinical scores (0.2%), \( Z \)-scores (1.2%), or other systems-level calculations (1.2%).

7.5. Specific biomarkers

The frequency of individual biomarkers was tabulated across publications with results displayed in Fig. 4. Heart rate (99.7%), systolic blood pressure (95.2%), diastolic blood pressure (91.1%), HbA1c (79.7%), HDL cholesterol (79.0%), and C-reactive protein (72.7%) were the most commonly employed biomarkers. This was followed by waist-to-hip ratio (64.3%), cortisol (62.3%), body mass index (61.8%) and total cholesterol (56.5%) as the remaining biomarkers that were used at least half of the included studies.

8. Discussion

The sum of at-risk biomarker values was by far the most common approach to creating an allostatic load score. While many researchers utilized clinical values, the use of high-risk quartile based on the sample distribution (as originally utilized by [58]) appears to be the “go to” method for most researchers (52.5% of all studies included in the review). Such an approach is problematic for multiple reasons. As noted by Grunewald et al. [28], this creates an uneven distribution of the impact of specific biological systems. For example, if five biomarkers of immune function are included and only two measures of HPA function are included, immune function will have an outsized impact on the total score. While there may be theoretical justifications for weighting one system more than another, this is rarely—if ever—explicitly stated, let alone justified in publications. A second concern with this approach, as previously noted, is that it is sample-dependent. This makes it difficult to compare allostatic load scores across studies that employ different samples. Finally, the use of high-risk quartile, as opposed to tertiles, deciles, or other method—is rather arbitrary, as neither theory nor evidence to date suggest there is something unique about the most at-risk 25% of the population [17]. Furthermore, allostatic load is meant to represent biological dysregulation in general. Although some researchers split the high-risk quartile into the highest and lowest 12.5% for variables such as cortisol, one could argue that such an approach could be valid across many additional biomarkers.

9. Recommendations for future research

Allostatic load research has evolved over the past three decades. While the results of this review demonstrate the exponential growth in research that utilizes allostatic load theory, there is a need to improve and grow the measurement and statistical methods employed in this
research in order to facilitate a move into the next generation of allostatic load research. Such research should include (1) a more nuanced understanding of the biological dysregulation that underpins the concept of allostatic load and (2) an orientation towards clinical implications of this research.

9.1. Moving measurement forward

In order to formulate a more nuanced understanding of allostatic load, advances in the algorithm used to calculate and apply the concept of allostatic load are sorely needed. Part of this is based on the need to understand the drivers of the allostatic load score. A handful of researchers have moved beyond systems scores and begun to use more advanced data analysis techniques such as latent class analysis to better understand the latent classes or groupings behind a cumulative score. Carbone [11] found that within the MIDUS study, there were three latent classes of biological dysregulation (parasympathetic dysregulation, metabolic and inflammatory dysregulation, and SAM pathway dysregulation), yet only the metabolic and inflammatory dysregulation group and the parasympathetic dysregulation group were found to have a greater risk of depression relative to the baseline (i.e., minimal/no dysregulation) group. Forrester et al. [24] conducted a latent class analysis of a subsample from the Study of Atherosclerosis and found that class membership was associated with health behaviors such as physical activity and alcohol use. Future research should consider approaches such as latent class analysis for a more nuanced understanding of the role of individual biomarkers as well as how they interact.

A limitation of latent class analysis is that it still requires individual biomarkers to be dichotomized. Latent profile analysis is another approach that provides the same benefits as latent class analysis but utilizes continuous data so as not to require the researcher to dichotomize biomarker values. Utilization of this approach in allostatic load research has been limited to date (e.g., [7]). Both latent class and latent profiles analyses presents their own challenges, including the fact that these forms of analysis are inherently exploratory and considerable subjectivity is often employed in selecting the final number of classes [66].

Structural equation modeling with latent variable analyses is another approach that provides the same benefits as latent class analysis but utilizes continuous data so as not to require the researcher to dichotomize biomarker values. Utilization of this approach in allostatic load research has been limited to date (e.g., [39]). Both latent class and latent profiles analyses present their own challenges, including the fact that these forms of analysis are inherently exploratory and considerable subjectivity is often employed in selecting the final number of classes [66].
statistical approach that has had only limited application in allostatic load research. Tampubolon and Maharani [70] found a linear trajectory of allostatic load among individuals in the United States and also found that these trajectories differed by sex. While a handful of other researchers have applied this approach (e.g., [50,75]), there remain opportunities to better explicate the development of biological dysregulation over time, especially as the availability of panel data with numerous data collection time points becomes more readily available.

In addition, algorithms that utilize sex-specific biomarker distributions is an important area for future research. Some scholars have employed this approach (e.g., [33,64]), but research in this area to date is fairly limited and additional work, including additional exploration of what biomarkers are most relevant and what is considered dysregulation by sex, is needed. Such work would have clearer clinical implications and applicability, and align with the National Institutes of Health Reproducibility Guidelines [74].

9.2. Improving clinical applications

Arguably the most important advance in measurement will focus more on change over time and not differences within groups. Currently, the MIDUS study is collecting its second wave of biomarker data from study participants. To date, one of the most significant limitations of allostatic load research relates to the cross-sectional nature of the majority of studies. Few allostatic load studies to date have employed a longitudinal approach, though this is changing as more studies are designed to utilize longitudinal cohorts (e.g., [8,16,18,70]). Indeed, this remains a weakness in allostatic load research. In addition to being able to explore the temporal ordering of allostatic load relative to other key variables related to health outcomes, longitudinal data will allow researchers to better explore the role of different biomarkers—and changes in those biomarkers—over time. Ideally, multiple biomarker samples will be collected over extended periods of time, allowing for causal inferences to be made about the role of specific and groups of biomarkers as their values change. This will allow for expanded theory testing, as it is important to remember that allostatic load is theorized to be a pre-clinical condition. If sub-clinical values can be monitored over time as biological dysregulation shifts and clinically-identified diseases develop, that information can be better utilized to identify individuals that are at a greater risk of developing a given disease. The focus on change over time will likely make the method of creating an allostatic load algorithm less important so long as the same method is employed in longitudinal analyses. This is the true clinical utility of allostatic load: identifying those at risk of, or experiencing low levels of, biological dysregulation and intervening before the development of disease.

It should be noted that to date, allostatic load has been found to be influences by, and associated with, a wide range of clinical and non-clinical factors. Another important area of focus that will require additional research and better integration into the existing literature relates to clinimetric measures. Some criteria to date include the identification of specific sources of stress or distress as well psychosocial manifestations (e.g., sleep disturbances, impairment in social functioning, irritability, restlessness) [21,22]. Clinimetric criteria can provide for a better understanding and linkage between the biological changes—as measured by biomarkers, the upstream environmental factors that induce stress and lead to biological dysregulation, and the downstream disease outcomes. Future research should aim to better integrate these aspects of allostatic load.

10. Conclusion

As allostatic load theory approaches its third decade of study, its use in the literature continues at an exponential rate of growth. While the theory itself is arguably elegant, the implementation in research has still not adequately addressed questions about how the construct should be operationalized. Not least of the operationalization issues relate to the algorithms used to create the allostatic load scores. This scoping review provides a brief overview of the breadth of approaches utilized in the literature, but also reveals that the majority of researchers have employed some version of a summed score if dichotomized biomarkers are based on the high-risk quartile of the sample distribution. As the science around allostatic load theory advances, the use of more nuanced and advanced statistical analysis techniques can aid in both advancing theory and making the results of research more applicable to the clinical setting, while informing the development and improvement of interventions.

Funding to declare

None.

Declaration of Competing Interest

None.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychres.2022.111050.

References*


* Note: This list contains all the references cited in the main text of the manuscript. Due to space constraints, a list of all 395 publications included in the scoping review is provided in Supplement 1 - Scoping Review References.


