Commentary

Neuroendocrine biomarkers, allostatic load, and the challenge of measurement: A commentary on Gersten

Eric B. Loucks[a],*, Robert P. Juster[b], Jens C. Pruessner[b]

[a]*McGill University, Montreal, QC, Canada H4H 1R3*
[b]*Douglas Mental Health Institute, McGill University, Montreal, QC, Canada*

Abstract

In this commentary, we discuss Gersten's findings particularly in relation to the challenge of accurately measuring stress, neuroendocrine markers and allostatic load. Allostatic load is a timely, potentially useful tool to measure the degree in which the body's physiological function is outside of optimal range. As with most biomarkers early on in development, there are sound opportunities to advance methods that will help understand the etiology of allostatic load and allow it to become more accurately measured. We present a biomarker development framework that should aid in furthering measurement of allostatic load, emphasizing the importance of biomarker measurement accuracy, standardization of methods, and relevance to clinically meaningful outcomes.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Allostatic load; Biomarkers; Stress; Measurement

An increasing number of epidemiological studies employ the allostatic load framework. As research on this topic evolves, several challenges have arisen, including (1) is allostatic load affected primarily by psychological stress, or are other factors primarily important, such as sedentary lifestyle, smoking, or diet? (2) how can allostatic load be measured most accurately? (3) to what extent does allostatic load causally influence clinical endpoints such as mortality, cardiovascular disease, cancer, and diabetes?

In the current issue of Social Science & Medicine, Gersten touches on the first two challenges described above (Gersten, 2008). Specifically, his

findings question whether a neuroendocrine subcomponent of allostatic load is reflective of chronic stress. His report using the Social Environment and Biomarkers of Aging Study (SEBAS) (the analytical sample was 880 male and female participants aged 54–90 living in Taiwan) showed that an index of current stress was associated with a panel of four neuroendocrine biomarkers in women and not in men. Interestingly, no association was found in either sex between a measure of chronic stress and the neuroendocrine markers. Current stress was measured using a summary score of six items: family's work situation, financial situation, family members' or children's health, marital situation, familial tension/conflict, and "other" stressors. Measurement of chronic stress was operationalized as the sum of stress duration for the six aforementioned items. Other factors conceptualized as

*Corresponding author. Tel.: +1 514 761 6131;
fax: +1 514 762 3049.
E-mail addresses: ericbarryloucks@gmail.com,
eric.loucks@mcgill.ca (E.B. Loucks).

measures of stress included low education and widowhood. The neuroendocrine biomarker panel, called neuroendocrine allostatic load (NAL), was measured using four markers: cortisol, norepinephrine, and epinephrine measured in urine collected over 12 h (7:00 pm–7:00 am), and dihydroepiandrosterone sulfate (DHEAS) assayed from a blood sample. The NAL score was quantified using different cut-off points in relation to the distribution within the study population (e.g., by quantile or a summed *z*-score), where high levels of cortisol, norepinephrine and epinephrine, as well as low levels of DHEAS were considered to be high risk.

There is a rich literature on biomarker development, used in many fields such as environmental toxicology, cardiovascular disease prediction, and pharmaceutical treatments for HIV/AIDS, among others. Commonalities exist for the phases in which biomarkers are developed. In the years 2000 and 2001, the National Institutes of Health Biomarkers Working Group released publications that provided frameworks on biomarker development (De Gruttola et al., 2001; Downing, 2000; NIH Biomarker Definitions Working Group, 2001). One of the frameworks presented by Xu and Zeger (Xu & Zeger, 2001; Zeger, 2000) considers the usefulness of biomarkers in predicting an outcome (such as mortality, cardiovascular disease, or cancer). It is adapted here to the context of allostatic load, shown in Fig. 1. In the case where a "perfect" biomarker (or surrogate endpoint) is used, the effects of the exposure (e.g. chronic psychological stress) on the outcome should be entirely mediated through the biomarker. This concept is shown in Fig. 1, where only pathways A and B would exist, and not pathways C or D. In reality, this situation does not yet exist for any single biomarker or panel of multiple biomarkers. Some markers come close, such as CD4 cell count and plasma HIV RNA (viral load) in the progression of HIV/AIDS (Mildvan, 2000). However, we must consider the situation (shown in pathway C of Fig. 1) where the exposure variable typically affects the outcome not only through the measured biomarker(s) but also through other pathways. For example, socioeconomic position may influence the incidence of cardiovascular disease not only through one measured biomarker (e.g., blood pressure), but also through other potential mediators such as smoking and obesity. Finally, each biomarker measurement is an approximation of the actual physiological processes taking place in the body (path D in Fig. 1). This point is particularly important in the case of allostatic load, which is in a fairly early phase of development as a biomarker panel. At early stages of biomarker development, there can be reasonable variability in how they are measured between studies. This is currently the case for allostatic load, and has recently been observed in biomarkers used for the metabolic syndrome. With regard to the metabolic syndrome, attempts are currently being made to standardize measurement, with expert consensus definitions proposed by groups such as the National Cholesterol Education Program (Adult Treatment Panel III) subsequently updated by the American Heart Association and
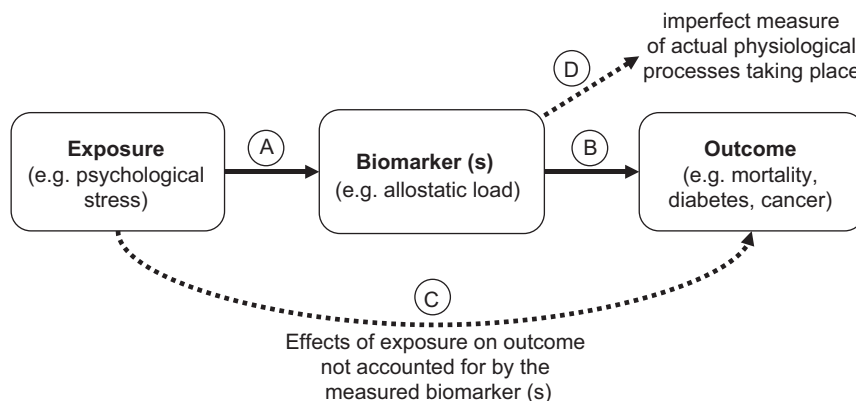

Fig. 1. In biomarker development, perfect biomarkers (i.e. surrogate endpoints) completely mediate the effects of the exposure on the outcome of interest (paths A and B). However, in reality, for most biomarkers, there are unmeasured factors that mediate the effects of the exposure on the outcome (path C). Finally, almost always, measurement of the biological mediators imperfectly represents the true bioactivity of the mediators (path D). This framework is modeled after Xu and Zeger (Xu & Zeger, 2001; Zeger, 2000), and applied to the concept of allostatic load.

National Heart, Lung and Blood Institute (Grundy et al., 2005), the World Health Organization (Alberti & Zimmet, 1998), and the International Diabetes Federation (Alberti, Zimmet, & Shaw, 2005). Acceptance of high-quality, standardized definitions allows for better comparability across studies, thus facilitating higher-quality measures of the biomarker(s) in future studies. As more knowledge becomes available on allostatic load, experts can come closer to a consensus on gold-standard measures in order to help the field develop.

Gersten's article is useful in allowing us to pause and evaluate where in the developmental phase allostatic load is at, and what steps are now required to fully realize its usefulness. In Fig. 1, path A describes the first challenge described above for the field of allostatic load—specifically is allostatic load primarily caused by psychological stress? Gersten's findings suggest that acute stress is related to NAL in women only, and place some question on whether chronic stress is related to NAL. Future studies will be helpful to confirm these results, keeping in mind important methodological factors. For example, it is important to measure the exposure variable (in this case acute and chronic stress) as accurately as possible while keeping in mind the limitations of large epidemiological studies such as this one, where it is difficult to include lengthy, time-consuming questionnaires. In this study, however, there was no information shown on validity or reliability of the stress scales; consequently, the accuracy of the stress scale is uncertain. Future studies that use stress questionnaires with known psychometric properties will provide further information on the relation between stress and neuroendocrine markers of allostatic load. For example, instruments validated in Taiwan have assessed *perceived stress* in postpartum mothers (Chen, Chou, Tseng, & Lee, 1994), and *stressful life events* in hyperthyroid outpatients (Lee et al., 2003); their applicability to older adults is limited and more cross-culturally validated instruments are needed. When considering which stress measures to use, the combination of objective stressors (e.g., natural disasters), subjective stressors (e.g., perceived distress), daily hassles (e.g., limited mobility), and stress duration (e.g., acute, chronic, and anticipated stress) (Brosschot, Pieper, & Thayer, 2005; Cohen, Kessler, & Underwood Gordon, 1995) might be beneficial in delineating the role of stress in the etiology of allostatic load. Some frequently used and validated stress

questionnaires include the Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983), the Social Readjustment Rating Scale (Holmes & Rahe, 1967), the WHO quality of life-BREF (Skevington, Lotfy, & O'Connell, 2004), the Hassles Scale (DeLongis, Folkman, & Lazarus, 1988) and the Trier Inventory of Chronic Stress (Schulz & Schlotz, 1999).

The second challenge currently facing the field of allostatic load described above is "how can allostatic load be measured most accurately?". This brings in the concept shown in Fig. 1, path D. Given the different endocrine and metabolic functions that the allostatic load concept touches upon, a number of guidelines can probably be formulated that should be kept in mind when assessing its biomarkers cortisol, norepinephrine, epinephrine, and DHEAS. With regard to cortisol, its regulation is highly complex and researchers to date struggle to find the best approach to obtain reliable and valid assessments (Haus & Touitou, 1994). Cortisol is the end product of activity of the hypothalamic–pituitary–adrenal (HPA) axis, and is often assessed as a proxy of total HPA axis activity. It can be measured from saliva (only the free, biologically active fraction of cortisol is captured in saliva), blood, and urine. When measured in urine, it is a cumulative measure and the duration of measurement needs to be chosen carefully, as described in more detail below. It is generally accepted that cortisol reaches the highest levels shortly after morning awakening, since cortisol levels are near the peak at the time of awakening (Born, Hansen, Marshall, Mölle, & Fehm, 1999) and another pulsatile secretion is usually associated with awakening (Pruessner et al., 1997). This is followed by declining levels throughout the day as pulsatile secretion decreases in amplitude and frequency. The nadir is usually found around midnight, and cortisol levels start to rise again in the early morning hours. Since individual pulses in response to stress can occur at any time during the day, a single sample in blood or saliva is basically useless for the assessment of baseline regulation of the HPA axis. Repeated measures spread throughout the day can help establish the circadian rhythm of the axis, with five samples usually providing sufficient information to establish the slope and amplitude of the curve (Stone et al., 2001). Since awakening is marked by a single pulsatile secretion, repeated assessment over the course of the first hour after awakening is a more reliable marker of the circadian peak and the

basal reactivity of the axis because it is sensitive to subtle dysregulations of the axis (Pruessner, Hellhammer, & Kirschbaum, 1999; Schulz, Kirschbaum, Pruessner, & Hellhammer, 1998). Norepinephrine and epinephrine, too can be measured in saliva, blood, and urine, but since these variables highly fluctuate, single time-point measures are usually not considered useful in saliva and blood. Since urine measures the cumulative norepinephrine and epinephrine output, it is considered the better agent for these variables when compared against single, one-time measures from plasma (Grassi & Esler, 1999). Cumulative measures like alpha-amylase assessment from saliva have recently been emerging, and show good promise (Nater, Rohleder, Schlotz, Ehlert, & Kirschbaum, 2007). Finally, for DHEAS, it already is a cumulative measure of DHEA; thus a single assessment in blood is considered a reasonable approach (Thomas et al., 1994).

In the data used by Gersten in this article, cortisol, norepinephrine, and epinephrine were measured from overnight urine, which has important implications. Any pulsatile secretion of the targeted variables that occurred during the night will contribute to the overall level established in the morning. Since it is known that HPA axis activity, and norepinephrine and epinephrine are stimulated by awakening, frequent awakenings during the night will likely have influenced those measures. Other state factors that might have influenced this cumulative endocrine measure could include meals, evening activities, number of hours of sleep, etc. Thus, if one is to rely on a single cumulative measure, it is usually recommended to have repeated sampling of this one measure (three nights over a 2-week period, for example), to avoid any state factors largely influencing the variance of the data (Epstein, 1990). Another consideration is the chosen time point and duration of the urine measure: levels were accumulated from 7 pm to 7 am, which corresponds to the nadir of the HPA axis activity, which is generally a quieter period for norepinephrine and epinephrine activity. It is therefore possible that variations in reactivity of these measures were not captured by the measure employed in the SEBAS. Of course, given the financial and feasibility constraints of large epidemiological studies, one has to weigh the practicality of the method against the information that can be derived from it. While it may not have been practical to schedule multiple samplings over the course of 1 day and then repeat this for several days, it might be useful for future studies to consider obtaining the daily urine together with nightly urine in order to assess the full cycle of cortisol and norepinephrine and epinephrine release, and/or to obtain urine on more than 1 day to avoid state factors that may influence the variability of the data.

Another factor to consider is the method with which the assessed neuroendocrine markers are being employed in the statistical analysis. In Gersten's article and the majority of allostatic load articles to date, a cut-off score of "high cortisol" is used for inclusion in allostatic load. However, increasing evidence suggests that low levels of cortisol are reflective of chronic stress states like burnout (Pruessner et al., 1999) which are implicated in stress-related diseases such as fibromyalgia and chronic pain (Crofford, Engleberg, & Demitrack, 1996; Heim, Ehlert, & Hellhammer, 2000; Millea & Holloway, 2000; Scott & Dinan, 1999) and are the most chronic and persevering symptoms in posttraumatic stress disorder and atypical depression (Hull, 2002; McGinn, Asnis, & Rubinson, 1996; Stewart, Quitkin, McGrath, & Klein, 2005; Wessa, Rohleder, Kirschbaum, & Flor, 2006; Yehuda, 1998, 1999). Future studies may also consider the use of both excessively high and low levels of cortisol for inclusion in allostatic load, as recently done by Glover, Stuber, and Poland (2006).

As we continue to consider the optimal measures of allostatic load, we can also investigate the use of DHEAS. Although it was used in many of the original formulations of allostatic load, and was considered to be implicated in successful aging, maintenance of youth, and prevention of disease (Butler, 1997; Khorram, 1996), recent studies have questioned the strength of association between this marker and various disease outcomes (Nair et al., 2006). Consequently, as research moves forward in determining optimal biomarkers for inclusion in allostatic load, the role of DHEAS in the concept of allostatic load may be considered less certain than that of most other biomarkers used in allostatic load.

Allostatic load is a timely, potentially useful tool to measure the degree to which the body's physiological function is outside of its optimal range. The hypothesized ability of allostatic load to predict risk for a variety of diseases is important in an era in which many major challenges to health (including cardiovascular disease, diabetes, and

cancer) share at least some common social and behavioral risk factors, such as socioeconomic position, smoking, obesity, and physical activity. As with most biomarkers early on in development, there are sound opportunities to advance the research that will help understand the etiology of allostatic load and allow it to become more accurately measured. Some of these opportunities include utilizing, and developing where needed, accurate measures of exposure variables (e.g., stress) and biomarkers (e.g., allostatic load) to better delineate paths A and D in Fig. 1. Further information on the causal role of allostatic load in predicting clinical endpoints such as mortality, cardiovascular disease, and cancer will help us understand the health implications of allostatic load (path B in Fig. 1). Gersten's article in this issue of Social Science & Medicine advances the field in providing information that, upon replication in other studies, could lead to de-emphasizing the causal role of chronic stress in the etiology of the neuroendocrine subcomponent of allostatic load.

References

Alberti, K. G., & Zimmet, P. Z. (1998). Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabetic Medicine*, *15*(7), 539–553.

Alberti, K. G., Zimmet, P., & Shaw, J. (2005). The metabolic syndrome—a new worldwide definition. *Lancet*, *366*(9491), 1059–1062.

Born, J., Hansen, K., Marshall, L., Mölle, M., & Fehm, H. L. (1999). Timing the end of the nocturnal sleep. *Nature*, *397*(6714), 29–30.

Brosschot, J. F., Pieper, S., & Thayer, J. F. (2005). Expanding stress theory: Prolonged activation and perseverative cognition. *Psychoneuroendocrinology*, *30*(10), 1043–1049.

Butler, R. N. (1997). Super hype for super hormones. *Geriatrics*, *52*(2), 13–14.

Chen, C. H., Chou, F., Tseng, Y., & Lee, J. N. (1994). The prevalence and predictors of postpartum depression (Chinese). *Nursing Research (China)*, *2*, 263–274.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385–396.

Cohen, S., Kessler, R., & Underwood Gordon, L. (1995). *Measuring stress: A guide for health and social scientists*. New York: Oxford University Press.

Crofford, L. J., Engleberg, N. C., & Demitrack, M. A. (1996). Neurohormonal perturbations in fibromyalgia. *Baillieres Clinical Rheumatology*, *10*(2), 365–378.

De Gruttola, V. G., Clax, P., DeMets, D. L., Downing, G. J., Ellenberg, S. S., Friedman, L., et al. (2001). Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. *Controlled Clinical Trials*, *22*(5), 485–502.

DeLongis, A., Folkman, S., & Lazarus, R. S. (1988). The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology*, *54*(3), 486–495.

Downing, G. (2000). Biomarkers and surrogate endpoints: Clinical research and applications. Proceedings of the NIH-FDA conference 15–16 April 1999, Bethesda, MD, USA. New York: Elsevier.

Epstein, S. (1990). Comment on the effects of aggregation across and within occasions on consistency, specificity, and reliability. *Methodika*, *4*, 95–100.

Gersten, O. (2008). Neuroendocrine biomarkers, social relations, and the cumulative costs of stress in Taiwan. *Social Science and Medicine*, *66*, 507–519.

Glover, D. A., Stuber, M., & Poland, R. E. (2006). Allostatic load in women with and without PTSD symptoms. *Psychiatry*, *69*(3), 191–203.

Grassi, G., & Esler, M. (1999). How to assess sympathetic activity in humans. *Journal of Hypertension*, *17*(6), 719–734.

Grundy, S. M., Cleeman, J. I., Daniels, S. R., Donato, K. A., Eckel, R. H., Franklin, B. A., et al. (2005). Diagnosis and management of the metabolic syndrome. An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Executive Summary. Circulation.

Haus, E., & Touitou, Y. (1994). Principles of clinical chronobiology. In Y. Touitou, & E. Haus (Eds.), *Biologic rhythms in clinical and laboratory medicine* (pp. 6–33). Berlin: Springer Verlag.

Heim, C., Ehlert, U., & Hellhammer, D. H. (2000). The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. *Psychoneuroendocrinology*, *25*(1), 1–35.

Holmes, T. H., & Rahe, R. H. (1967). The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, *11*(2), 213–218.

Hull, A. M. (2002). Neuroimaging findings in post-traumatic stress disorder. Systematic review. *The British Journal of Psychiatry*, *181*, 102–110.

Khorram, O. (1996). DHEA: A hormone with multiple effects. *Current Opinion in Obstetrics Gynecology*, *8*(5), 351–354.

Lee, I. T., Sheu, W. H., Liau, Y. J., Lin, S. Y., Lee, W. J., & Lin, C. C. (2003). Relationship of stressful life events, anxiety and depression to hyperthyroidism in an Asian population. *Hormone Research*, *60*(5), 247–251.

McGinn, L. K., Asnis, G. M., & Rubinson, E. (1996). Biological and clinical validation of atypical depression. *Psychiatry Research*, *60*(2–3), 191–198.

Mildvan, D. (2000). Biomarkers as surrogate endpoints in clinical trials for HIV/AIDS: A model for other diseases? In G. J. Downing (Ed.), *Biomarkers and surrogate endpoints: Clinical research and applications* (pp. 13–17). New York: Elsevier.

Millea, P. J., Holloway, R. L. (2000). Treating fibromyalgia. American Family Physician, *62*(7), 1575–1582, 1587.

Nair, K. S., Rizza, R. A., O'Brien, P., Dhatariya, K., Short, K. R., Nehra, A., et al. (2006). DHEA in elderly women and DHEA or testosterone in elderly men. *New England Journal of Medicine*, *355*(16), 1647–1659.

Nater, U. M., Rohleder, N., Schlotz, W., Ehlert, U., & Kirschbaum, C. (2007). Determinants of the diurnal course of salivary alpha-amylase. *Psychoneuroendocrinology*.

NIH Biomarker Definitions Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*, *69*(3), 89–95.

Pruessner, J. C., Hellhammer, D. H., & Kirschbaum, C. (1999). Burnout, perceived stress, and cortisol responses to awakening. *Psychosomatic Medicine*, *61*(2), 197–204.

Pruessner, J. C., Wolf, O. T., Hellhammer, D. H., Buske-Kirschbaum, A., von Auer, K., Jobst, S., et al. (1997). Free cortisol levels after awakening: A reliable biological marker for the assessment of adrenocortical activity. *Life Sciences*, *61*(26), 2539–2549.

Schulz, P., Kirschbaum, C., Pruessner, J. C., & Hellhammer, D. H. (1998). Increased free cortisol secretion after awakening in chronically stressed individuals due to work overload. *Stress Medicine*, *14*, 91–97.

Schulz, P., Schlotz, W. (1999). Trier Inventar zur Erfassung von chronischem Streß (TICS): Skalenkonstruktion und teststatistische Überprüfung. Trier: Universität Trier.

Scott, L. V., & Dinan, T. G. (1999). The neuroendocrinology of chronic fatigue syndrome: Focus on the hypothalamic–pituitary–adrenal axis. *Functional Neurology*, *14*(1), 3–11.

Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res*, *13*(2), 299–310.

Stewart, J. W., Quitkin, F. M., McGrath, P. J., & Klein, D. F. (2005). Defining the boundaries of atypical depression: Evidence from the HPA axis supports course of illness distinctions. *Journal of Affective Disorders*, *86*(2-3), 161–167.

Stone, A. A., Schwartz, J. E., Smyth, J., Kirschbaum, C., Cohen, S., Hellhammer, D., et al. (2001). Individual differences in the diurnal cycle of salivary free cortisol: A replication of flattened cycles for some individuals. *Psychoneuroendocrinology*, *26*(3), 295–306.

Thomas, G., Frenoy, N., Legrain, S., Sebag Lanoe, R., Baulieu, E. E., & Debuire, B. (1994). Serum dehydroepiandrosterone sulfate levels as an individual marker. *Journal of Clinical Endocrinology & Metabologism*, *79*(5), 1273–1276.

Wessa, M., Rohleder, N., Kirschbaum, C., & Flor, H. (2006). Altered cortisol awakening response in posttraumatic stress disorder. *Psychoneuroendocrinology*, *31*(2), 209–215.

Xu, J., & Zeger, S. L. (2001). The evaluation of multiple surrogate endpoints. *Biometrics*, *57*(1), 81–87.

Yehuda, R. (1998). Psychoneuroendocrinology of post-traumatic stress disorder. *Psychiatric Clinics & North American*, *21*(2), 359–379.

Yehuda, R. (1999). Biological factors associated with susceptibility to posttraumatic stress disorder [see comments]. *Canadian Journal of Psychiatry*, *44*(1), 34–39.

Zeger, S. L. (2000). Biomarkers as surrogates in clinical research. In G. J. Downing (Ed.), *Biomarkers and surrogate endpoints: Clinical research and applications* (pp. 22–26). New York: Elsevier.