Assessment of Chronic Stress: Comparison of Hair Biomarkers and Allostatic Load Indices*

Pia-Maria Wippert, J. Honold, V. Wang
University of Potsdam, Potsdam, Germany

C. Kirschbaum
Technical University of Dresden, Dresden, Germany

Background: The measurement of long-term stress is challenging and depends on the population under study. New methods, such as hair markers or ALIs (allostatic load indices), hold promise as a means of assessing personal stress levels. These methods are relatively new and there is still a lack of knowledge concerning their validity and reliability.

Objective: This preliminary study aims to determine the test-retest reliability of hair hormone levels (progesterone, cortisol, dehydroepiandrosteron) and two different ALIs as well as the comparability of these biomarker sets among themselves and with different psychometric measures.

Methods: Psychometric and physiological data were collected from 13 participants (10 f, 3 m; age = 28.2 ± 2.1 yr.) on two consecutive days with an interval of 24 hours. An 18-biomarker ALI included: HbA1c (glycosylated hemoglobin), fasting glucose, insulin resistance, triglycerides, HDL (high-density lipoprotein), LDL (low-density lipoprotein) waist-hip ratio, BMI (body mass index), BPSYS (systolic blood pressure), BPDIA (diastolic blood pressure), resting pulse, DHEA-S (dehydroepiandrosteron-serum), cortisol, epinephrine, norepinephrine (12 h urine), CRP (C-reactive protein), sICAM-1 (soluble intracellular adhesion molecule-1) and e-Selectin. ALI-HPA included only cortisol and DHEA-S. Hair samples (3 cm) were analyzed using high-performance LC-MS/MS (liquid chromatography-mass spectrometry). Subjective stress was documented using TICS (the Trier Inventory of Chronic Stress).

Descriptive statistics were quantified as means ± SD and correlations with Pearson’s correlation coefficient. For reliability, ICC (inter-class correlation) 2.1 and Bland-Altman analysis (LoA (limits of agreement) 2.00*SD) were calculated.

Results: HCC (Hair cortisol concentrations) and ALIs showed high test-retest reliabilities while other hair hormones (i.e., DHEA, progesterone) were only moderate or good in this respect. ALI and HCC correlated moderately, as did hair DHEA and serum DHEA. Neither HCC and ALI-HPA, nor HCC and 12h-urine cortisol showed significant correlations.

Conclusion: HCC and ALIs are reliable techniques for the assessment of chronic stress. Regarding the comparability of the measurement techniques, further studies and larger samples are needed to better understand their cut-off points and predictive power for health risks.

Keywords: allostatic load, hair cortisol, hair steroids, stress biomarker

Introduction

The measurement of biomarkers during an acute stress situation is a well-established practice in...
bio-behavioural research. Dynamic test situations are used in order to assess the function or feedback sensitivity of neuroendocrine axes. However, identifying the gold standard biomarker (set) to pinpoint physiological responses to chronic stress has proven to be challenging. Chronic stress can be measured by assessing biomarkers in urine, blood or saliva. The interpretation of the results is limited and depends on the population under study, the research question, the specific study design and inter- and intra-individual variability in biomarker levels. More sophisticated methods are required for a better description of physiological states and to determine whether these have prognostic implications in terms of poor physical health of people under chronic stress (Juster, McEwen, & Lupien, 2010). There are currently two measurement concepts that claim to be capable of assessing physiological changes due to long-term exposure to stress: analyses of hair cortisol concentration (HCC (Hair cortisol concentrations) (Kirschbaum, Tietze, Skoluda, & Dettenborn, 2009)) and of allostatic load indices (ALIs (allostatic load indices) (Seeman et al., 2010)). These approaches are based on different theoretical backgrounds.

As cortisol is incorporated in hair as it grows, HCC should be useful in analysing central HPA (hypothalamic-pituitary-adrenal axis) activity over prolonged periods of time retrospectively (Stalder & Kirschbaum, 2012). Different studies have shown that exposure to chronic stress leads to an increase in hair cortisol (Davenport, Tiefenbacher, Lutz, Novak, & Meyer, 2006; Xie et al., 2011).

The ALI is an integrated score of multiple interacting systems (PNS (parasympathetic nervous system) and SNS (sympathetic nervous system), HPA, cardiovascular system, lipid metabolism, glucose metabolism, and inflammatory immune activity) that considers both primary and secondary mediators in adapting to long-term stress exposure (Juster et al., 2010). The allostatics-adaptation process follows theoretically three stages of stress mediation. In the first stage, primary mediators (e.g., adrenalin, norepinephrine, DHEA, cortisol) and anti-inflammatory cytokines (e.g., Interleukin-6) will be activated. In the second, secondary mediators lead to changes in metabolic (e.g., insulin, glucose, HDL (high-density lipoprotein), LDL (low-density lipoprotein), total cholesterol, triglycerides, visceral fat depositing), cardiovascular (e.g., systolic and diastolic blood pressure, heart rate) and immune system (e.g., CRP (C-reactive protein), Fibrinogen) (Goldman et al., 2006; Seeman, McEwan, Rowe, & Singer, 2001; Sun, Wang, Zhang, & Li, 2007) changes. The index may provide more information than just the HPA-response by representing the multisystem biological dysregulation and cumulative physiological cost of allostatic load, which is associated with a greater risk of stress-related diseases (Beckie, 2012).

Despite this promising research, these methods have also been criticised. It has been suggested that HCC may stem from passive diffusion from blood, sebaceous, eccrine or hair follicle secretions (Xie et al., 2011) and that wash-out effects can influence results (Kirschbaum, Tietze, Skoluda, & Dettenborn, 2009). Furthermore, it is unclear if HCC only maps the HPA activity or provides broader information about the adaption to long-term stress. The ALI has been criticized for a lack of test-retest studies and the fact that cut-offs for health risk have yet to be determined at the level of the individual person. To reconcile such methodological issues, but also for practical use, it is worthwhile to understand better the validity and reliability of each measurement method and the comparability between both measurement concepts. Therefore, in the present preliminary study we assessed: (1) the test-retest reliability of HCC, hair DHEA, hair progesterone and ALIs; as well as (2) the comparability of these biomarkers measured in different human biological samples. For a complete picture of the assessment methods, we also acquired information regarding stress levels, measured by psychometric scales.
Material and Methods

Participants

*N* = 13 academic employees of the University of Potsdam (10 females, 3 males; age: *M* = 28.23, *SD* = 2.13; ethnicity: white = 100%) participated in this study on a voluntary basis. This study was approved by the university’s ethics board as a preliminary study for a larger-scale longitudinal study (petition 44/2012-decision 6/34-6.5.2013). In this proof of concept study (validation of methods), individuals were excluded if they had an acute infection, suffered from any major somatic disease (cardiovascular, metabolism, thyroid dysfunction, angiopathy, hemophiliac, or psychological disorders such as depression or anxiety), used medication (glucocorticoids, antibiotics), were pregnant, had short hair (< 3 cm) or suffered severe hair loss.

Design

One day before the investigation, participants were informed about the study and the required diet during the study period. They provided written informed consent before inclusion and received their urine vessel. Participants were asked to only drink water and not to eat for 12 hours prior to the assessment. They were also instructed to abstain from coffee, tea, alcohol, certain foods (e.g., bananas, cheese, nuts, vanilla, and citrus fruits), intensive exercise and any medication. For proofing the diurnal variability (and reliability itself), all measurements were completed on two consecutive days with an interval of 24 hours, following the same order and by the same staff at the University of Potsdam. In addition to biological data (derived from samples of blood, urine, hair, cardiovascular, and heart rate variability measurements), participants provided information on sociodemographic factors, common health conditions, alcohol and nicotine consumption and completed a questionnaire on the first measurement day.

Biochemical Assessments

Two strands of hair (approx. 4-5 mm diameter) were cut each day as closely as possible to the scalp from a posterior vertex position. A minimum of 20 mg of hair for a 3 cm segment was obtained from each participant and stored in aluminium foil until steroid concentrations were determined. Assuming a hair growth rate of 1 cm/month (Wenning, 2000), the 3 cm hair segment most proximal to the scalp may be used as a retrospective measurement of stress (i.e., the amount of HPA-regulated steroid hormones produced) covering the last three months.

12-hour overnight urine samples (7:00 pm to 7:00 am) were collected in tubes containing acetic acid as preservative and refrigerated throughout the overnight collection phase. Until the urine hormone concentrations were assessed, the samples were stored in a freezer (-60°C to -80°C). The total 12-hour hormone excretion was determined by multiplying the urine hormone concentration (ng/ml) by the urine volume (ml).

Morning blood samples were drawn from the arm between 7:00 am and 8:00 am (1-hour after awakening). Blood was collected in plain blood collection tubes or tubes containing EDTA, citrate or sodium fluoride for subsequent analysis by partner laboratories.

All hair hormone levels were assayed at the University of Dresden (CK) by high performance LC-MS/MS (liquid chromatography coupled with tandem mass spectrometry) as described in detail elsewhere (Gao, Stalder, Foley, Rauh, Deng, & Kirschbaum, 2013).

Assays for metabolic biomarkers such as HbA1c (glycosylated hemoglobin), fasting glucose, fasting insulin, triglycerides, HDL as well as LDL were performed at the “Institut für Medizinische Diagnostik Berlin-Potsdam MVZ GbR” (Lab CV: D-ML-13142-01-00). HbA1c was determined with the HPLC Bio-Rad
Variant II (Bio-Rad Laboratories, CA, USA). Plasma glucose was measured via a hexokinase enzymatic reaction using the Roche Cobas Integra 400 Plus System (Roche Diagnostics Ltd., Basel, Switzerland). Fasting insulin was determined by ECLIA (the electrochemiluminescence enzyme immunoassay method) using a Roche Cobas 8,000 Modul E620 (Roche Diagnostics Ltd., Basel, Switzerland). To estimate insulin resistance, HOMA (the homeostatic model assessment) index was calculated by the formula: glucose (mmol/L) × insulin (µU/ml)/22.5 (see Matthews et al., 1985). Triglycerides, HDL and LDL were measured using an Architect ci8200 analyzer (Abbott Laboratories, Illinois, USA) via enzymatic colorimetric assays. The medical staff at the University of Potsdam assessed height (Seca 222 telescopic measuring rod; seca, Hamburg, Germany), weight (Kern MPS scale; Kern & Sohn GmbH, Balingen, Germany), waist and hip circumference (conventional measuring tape). During the physical measurements, participants only wore underwear. Waist-hip ratio was calculated as the ratio of the waist (assessed at the narrowest point between ribs and the iliac crest) to the hip circumference (measured at the maximum diameter of the buttocks). BMI (body mass index) was computed as weight in kg/(height in m)².

Cardiovascular biomarkers such as SBP (systolic blood pressure) and DBP (diastolic blood pressure) were taken three times with an interval of 5 minutes and the 2nd and 3rd measurements were averaged to obtain the final values (BOSO BS 90 Blood pressure instrument, BOSCH + SOHN GmbH u. Co. KG, Jungingen, Germany)¹.

Hypothalamic-pituitary-adrenal hormones (i.e., urine cortisol and serum-DHEA), sympathetic nervous systems hormones (i.e., urine epinephrine and urine norepinephrine) and inflammatory biomarkers (i.e., sICAM-1 (soluble intracellular adhesion molecule-1), IL-6 (interleukin-6), soluble (e-Selectin) were assayed at the University of Potsdam with a panel of enzyme-linked immunosorbent assays: RE52241 for cortisol, RE52181 for serum-DHEA, RE59251 for epinephrine, RE59261 for norepinephrine, BE59011 for sICAM-1 and BE59061 for e-Selectin (IBL International GmbH, Hamburg, Germany). CRP (C-reactive protein) and plasma fibrinogen were analyzed as additional inflammatory biomarkers. CRP levels were determined at the University of Potsdam by a latex immunoturbidimetric assay using the ABX Pentra 400 System (HORIBA ABX SAS, Montpellier, France). Plasma fibrinogen was determined at the “Institut für Medizinische Diagnostik Berlin-Potsdam MVZ GbR” via a Siemens BCS XP coagulation analyzer (Siemens AG, Erlangen, Germany).

Questionnaires

Chronic stress was assessed with the TICS (Trierer Inventory of Chronic Stress) (Schulz, Schlotz, Becker, & (TICS), 2004), where participants rated 57 items on a 5-point Likert scale (from 0 = never to 4 = very often) and indicated how often they experienced described stressful situations within the last three months. The test consists of nine scales: Work Overload, Social Overload, Pressure to Perform, Work Discontent, Excessive Demands at Work, Lack of Social Recognition, Social Tensions, Social Isolation and Chronic Worrying. Internal consistency coefficients for the different scales vary in this sample between 0.70 and 0.94.

Statistics

The 18 biomarkers² were first controlled for outliers and then aggregated into a multi-system ALI

¹ Due to technical problems with the ECG, heart rate variability indicators (SDNN, rMSSD, SDANN, SDNNidx) and resting pulse could not be assessed.
² IL-6 was not detectable in sample: value < detectable threshold 10.
representing six different physiological systems (SNS, HPA axis, cardiovascular system, glucose metabolism, lipid metabolism and inflammatory immune activity). The ALI was calculated as the sum of single biomarkers that fell into the high-risk quartile of the sample\(^3\). System risk values were scored from 0 to 1 (indicating that 0-100% of a participant’s system biomarkers fell within the high-risk range (Gruenewald et al., 2012)). Because the six different physiological systems had varying numbers of biomarker indicators, the aggregate risk scoring methodology used here generated a comparable “scaling” of scores across the various systems. For further comparisons between HCC and ALI, we aggregated urine cortisol and serum-DHEA into a second AL-score that only represents the HPA system (termed ALI-HPA). This was based on different theoretical conclusions in which HCC is described as a presenter of the HPA activity (Stalder & Kirschbaum, 2012; Xie et al., 2011).

Descriptive statistics were quantified as means ± SD and correlations with Pearson’s correlation coefficient using SPSS (SPSS Statistics 21, IBM, USA) and Microsoft Excel. For reliability, inter-class correlation (ICC 2.1) and Bland-Altman analysis (bias ± limits of agreement (LoA 2.00*SD)) were calculated.

### Results

Table 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>(M (SD))</th>
<th>Min. ± Max.</th>
<th>Pearson’s (r)</th>
<th>ICC 95% CI</th>
<th>Precision</th>
<th>LoA (t1-t2, MD ± SD) (t1-t2, 95% CI of difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hair (3 cm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hair cortisol (pg/mg)</td>
<td>13.14 ± 8.90</td>
<td>10.80 ± 7.96</td>
<td>4.75 ± 36.27</td>
<td>0.95**</td>
<td>0.94</td>
<td>2.33 ± 2.94 0.56 ± 4.12</td>
</tr>
<tr>
<td>Hair progesterone (pg/mg)</td>
<td>6.16 ± 3.69</td>
<td>4.55 ± 2.73</td>
<td>0.68 ± 11.81</td>
<td>0.68*</td>
<td>0.65</td>
<td>1.61 ± 2.72 -0.04 ± 3.25</td>
</tr>
<tr>
<td>Hair DHEA (pg/mg)</td>
<td>9.83 ± 5.21</td>
<td>10.46 ± 7.51</td>
<td>3.21 ± 20.88</td>
<td>0.81**</td>
<td>0.75</td>
<td>-0.63 ± 4.53 -3.37 ± 2.11</td>
</tr>
<tr>
<td>ALI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALI total (18)</td>
<td>4.23 ± 2.65</td>
<td>4.00 ± 3.00</td>
<td>1.00 ± 10.00</td>
<td>0.89**</td>
<td>0.88</td>
<td>0.23 ± 1.36 -0.59 ± 1.05</td>
</tr>
<tr>
<td>ALI-HPA (2)</td>
<td>0.54 ± 0.66</td>
<td>0.54 ± 0.52</td>
<td>0.00 ± 2.0</td>
<td>0.79**</td>
<td>0.76</td>
<td>0.01 ± 0.41 -0.25 ± 0.25</td>
</tr>
<tr>
<td>Test-retest values of comparison data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12h-Urine Cortisol rel. (μg/d)</td>
<td>48.20 ± 21.40</td>
<td>54.72 ± 23.13</td>
<td>13.10 ± 75.40</td>
<td>0.75**</td>
<td>0.74</td>
<td>-6.52 ± 15.96 -16.16 ± 3.13</td>
</tr>
<tr>
<td>DHEA-s rel. (μg/ml)</td>
<td>1.78 ± 0.95</td>
<td>1.77 ± 0.95</td>
<td>0.66 ± 3.27</td>
<td>0.99**</td>
<td>0.99</td>
<td>0.01 ± 0.13 -0.07 ± 0.09</td>
</tr>
</tbody>
</table>

**Notes.** LoA: limits of agreement; a) calculated for \(n = 13\); \(^*\) correlation significant at 0.05; ** correlation significant at 0.01 (2-tailed); hair calculated for \(n = 13\), segment 3 cm, extraction with 1.800 µl MeOH, 18 h.

As shown in Table 1, ICC and LoA results pointed to a high stability of HCC extraction and hair-DHEA, while ICC and LoA of hair progesterone analysis seems only moderately reliable. Urine cortisol and serum-DHEA levels correlated highly between day 1 and 2, which suggests good test-retest reliability for the HPA system. ALI and ALI-HPA also showed excellent test-retest reliability between the two consecutive measurement days.

As it was shown in Table 2, ALI and HCC correlated moderately and for one measurement point significantly, whereas ALI and hair-DHEA do not significantly correlate (see Table 2). In contrast the ALI-HPA was not significantly correlated with HCC. Neither HCC and 12h-urine cortisol, nor DHEA levels of hair and serum were significantly correlated.

Psychometric measures showed average stress level for the TICS subscales: social overload (\(M = 6.08, SD\)

---

\(^3\) All participants layed within the laboratory reference values of the biomarkers.
pressure to perform ($M = 14.75, SD = 4.03$), work discontent ($M = 9.17, SD = 4.43$), excessive demands at work ($M = 5.50, SD = 2.71$), lack of social recognition ($M = 4.27, SD = 2.69$), social tensions ($M = 3.83, SD = 3.24$), and social isolation ($M = 5.08, SD = 4.46$). The TICS scales: work overload ($M = 14.83, SD = 7.28$), chronic worrying ($M = 5.42, SD = 4.03$), and the chronic subjective stress scale ($M = 15.46, SD = 7.92$) were slightly increased.

Table 2

<table>
<thead>
<tr>
<th>Measures</th>
<th>Pearson’s $r$ (t1)</th>
<th>Pearson’s $r$ (t2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALI total-Hair Cortisol</td>
<td>0.53 ($p = 0.06$)</td>
<td>0.55* ($p = 0.05$)</td>
</tr>
<tr>
<td>ALI total-Hair DHEA</td>
<td>0.33 ($p = 0.27$)</td>
<td>0.19 ($p = 0.53$)</td>
</tr>
<tr>
<td>ALI HPA-Hair Cortisol</td>
<td>0.20 ($p = 0.51$)</td>
<td>0.24 ($p = 0.43$)</td>
</tr>
<tr>
<td>ALI HPA-Hair DHEA</td>
<td>-0.14 ($p = 0.64$)</td>
<td>-0.14 ($p = 0.65$)</td>
</tr>
<tr>
<td>12h-Urine Cortisol-Hair Cortisol</td>
<td>0.13 ($p = 0.68$)</td>
<td>-0.16 ($p = 0.61$)</td>
</tr>
<tr>
<td>Hair-DHEA-Serum DHEA-s</td>
<td>0.44 ($p = 0.14$)</td>
<td>0.37 ($p = 0.21$)</td>
</tr>
</tbody>
</table>

Notes. Calculated for $n = 13$; * correlation significant at 0.05; ** correlation significant at 0.01 (2-tailed).

**Discussion**

The aim of this test-retest-study was to investigate if different stress measurement methods are reliable in a test-retest situation, and if similar biomarkers measured in different human biological samples are comparable.

We found that HCC, hair-DHEA and ALI showed excellent reliability, whereas hair progesterone and ALI-HPA showed moderate or good test-retest reliability between two consecutive measurement days. Thus, these measures seem to be methodologically useful. The stability of the stress-associated hormones confirms that these methods could be used for assessing chronic stress, as they seem less prone to day-to-day fluctuations than measures derived from other matrices.

The comparison between methods showed that ALI and HCC correlated moderately and in one measurement day significantly, while ALI-HPA correlated only weakly and nonsignificantly with HCC. The low correlation between ALI-HPA and HCC (which should give information about the HPA activity, see Xie et al., 2011) was not expected and might be explained in the measurement of different fluids and time frames (hair should map the last three months). The moderate correlations between ALI and HCC, however, may underline this and suggest that HCC provides broader information about the stress adaptation compared to only momentary urine/serum HPA biomarkers. Further research on comparisons between HCC and different AL-indices (e.g., which give us a information about the primary stress adaption mediators: cortisol, s-DHEA, epinephrine, norepinephrine, IL-6 (Seeman et al., 2001) or further IGF1, Dopamin (Goldman et al., 2006) as well as secondary stress adaption mediators: SBP, DBP, HbA1c, cholesterol, ratio of total cholesterol to HDL, BMI (Goldman et al., 2006)) could clarify these questions and would be helpful in determining cut-off points for ALI and HCC measurements and their different predictive power for health risks. In this case, it would also be important to contrast different steroid extraction protocols for the hair analysis (ELISA or LC-MS/MS) to gain more information about the comparison possibilities between ALI and HCC.

The comparison between ALI and hair-DHEA, as well as between DHEA levels in hair and serum,
showed moderate but non-significant correlations. Still, the DHEA measurements\(^4\) are reliable and make them suitable for usage in future studies analysing this functional antagonism of the actions of glucocorticoids and resilience factor (Wolf & Kirschbaum, 1999).

As it was shown by Sauvé, Koren, Walsh, Tokmakejian, and Van Uum (2007), we could not reproduce the correlation between cortisol levels assayed from hair and 12h-urine samples with 24h-urine samples. This might be explained by the difference of 12h- and 24h-urine samples, the usage of longer hair samples or the usage of an ELISA-oriented steroid extraction protocol by the authors (Sauvé et al., 2007).

Regarding the stress levels of the group, most psychometric scales showed average age- and status-related stress levels. In relation to different German reference groups, three TICS scales namely “work overload”, “chronic worrying” and the “chronic stress screening scale” were increased (WO: \(M = 11.4\); CW: \(M = 4.9\), CSSS: \(M = 14.4, n = 1,087\) (Petrovski, Sören, Albani, & Brähler, 2012; Schulz et al., 2004)) and give the indication of a partially higher work-related demand in our study group. Referring to the biomarkers, it is to resume, that the HCC mean value is low in comparison to the rare reference values of this analysis method LC-MS/MS (\(M = 18\) pg/mg (Raul, Cirimele, Ludes, & Kintz, 2004)). For the Allostatic load indices, we miss reference values for this age group, wherefore further studies are needed and a grading is impossible (Beckie, 2012).

In sum, we have a partially demanded group which is very suitable for a test-retest situation, but of course our study had some limitations: (1) The study sample was small and rather homogeneous introducing a possible sex bias, wherefore we abstained from further comparisons between psychometric scales and biomarkers; (2) It is well known that physiological measures could be influenced from laboratory techniques and calibration assays (Contois, Warnick, & Sniderman, 2011); and (3) Due to missing six biomarkers, we were significantly restricted in the attempt to measure the 24-AL-indicator. Nevertheless, the topic seems to be very important in research and for the understanding of cut-off points of different measurement methods.

The results suggest that HCC and ALIs are reliable methodological approaches for assessing chronic stress biomarkers. Further studies with significantly larger sample sizes and the usage of different steroid extraction protocols (regarding analyses like ELISA or LC-MS/MS and the amount of hair) are required to understand more about the comparability between ALI and HCC, their cut-off points, predictive power for health risks and their potential for clinical interventions.

References

\(^4\) High reliability between DHEA-s sulfat was expected.


