Biological Psychology 129 (2017) 207–230

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Review

Psychosocial functioning and the cortisol awakening response: Meta-analysis, *P*-curve analysis, and evaluation of the evidential value in existing studies

Ian A. Boggero[a,*], Camelia E. Hostinar[b], Eric A. Haak[a], Michael L.M. Murphy[c], Suzanne C. Segerstrom[a]

[a] Department of Psychology, University of Kentucky, 125 Kastle Hall, Lexington, KY 40506, United States
[b] Department of Psychology, University of California, Davis, 103 Young Hall, Davis, CA 95616, United States
[c] Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, United States

ARTICLE INFO

Keywords:
HPA axis
Stress
Depression
Physiology
Endocrinology

ABSTRACT

Cortisol levels rise immediately after awakening and peak approximately 30–45 min thereafter. Psychosocial functioning influences this cortisol awakening response (CAR), but there is considerable heterogeneity in the literature. The current study used *p*-curve and meta-analysis on 709 findings from 212 studies to test the evidential value and estimate effect sizes of four sets of findings: those associating worse psychosocial functioning with higher or lower cortisol increase relative to the waking period (CARi) and to the output of the waking period (AUCw). All four sets of findings demonstrated evidential value. Psychosocial predictors explained 1%-3.6% of variance in CARi and AUCw responses. Based on these effect sizes, cross-sectional studies assessing CAR would need a minimum sample size of 617–783 to detect true effects with 80% power. Depression was linked to higher AUCw and posttraumatic stress to lower AUCw, whereas inconclusive results were obtained for predictor-specific effects on CARi. Suggestions for future CAR research are discussed.

The replication crisis in psychology has raised questions about the best methods to promote confidence in scientific findings. Some have suggested that "meta-analysis provides the best foundation for progress, even for messy applied questions" (Cumming, 2008, p. 292), because aggregating across studies reduces noise. Others have noted that publication bias and questionable research practices in individual studies can skew meta-analytic results (Simmons, Nelson, & Simonsohn, 2011), proposing additional ways to analyze the evidential value of a set of findings (Simonsohn et al., 2014; van Aert, Micherts, & van Assen, 2016). Research on relations between hormones and behavior can be particularly "messy" and vulnerable to inconsistency. The literature on psychosocial functioning and the cortisol awakening response (CAR) has more than doubled in the last six years, but heterogeneity in published findings has partially obscured the nature and direction of these effects. The present study aims to assess the evidential value of the published literature linking psychosocial functioning to the CAR.

Cortisol is a steroid hormone involved in glucose regulation and metabolism that is frequently labeled a stress hormone because its levels change markedly in response to stressors (see Sapolsky, 2004 for a review). Cortisol levels rise steeply immediately after awakening in the morning and peak approximately 30–45 min thereafter; from there, they slowly decrease throughout the day. This early peak in cortisol is known as the CAR (Pruessner et al., 1997). The CAR is responsive to stress perception and anticipation of daily stressors, making it particularly interesting to psychologists (Fries, Dettenborn, & Kirschbaum, 2009).

A meta-analysis of 62 studies revealed considerable heterogeneity in findings linking psychosocial functioning to the CAR (Chida & Steptoe, 2009). For instance, CAR measures were negatively correlated with depression in one study (O'Donnell, Badrick, Kumari, & Steptoe, 2008) but positively correlated in another (Mommersteeg, Heijnen, Verbraak, & van Doornen, 2006). Heterogeneity in the literature linking psychosocial functioning to the CAR is the rule rather than the exception; in fact, in four of seven types of psychosocial predictors, studies showing both positive and negative associations exist (Chida & Steptoe, 2009). Heterogeneity may result from inconsistency in how the CAR is measured (Stalder et al., 2016) and from small sample sizes ($N < 100$) that are characteristic of the literature. With small sample sizes, observed effects may misestimate both the size and the direction of any true underlying effect (Gelman & Carlin, 2014). Misestimations of

* Corresponding author.
E-mail addresses: ian.boggero@uky.edu (I.A. Boggero), cehostinar@ucdavis.edu (C.E. Hostinar), eric.haak@uky.edu (E.A. Haak), michaelmurphy@cmu.edu (M.L.M. Murphy), segerstrom@uky.edu (S.C. Segerstrom).

http://dx.doi.org/10.1016/j.biopsycho.2017.08.058
Received 15 January 2017; Received in revised form 5 July 2017; Accepted 29 August 2017
Available online 14 September 2017

direction are unlikely to be detected because there are no norms for the CAR, making it impossible to tell whether CAR values represent hyporesponsiveness, normal responses, or hyperresponsiveness.

To further complicate matters, different researchers operationalize and measure the CAR differently. Most researchers quantify the dynamic increases that occur during the first waking hour (CARi). Others compute the area under the curve relative to the waking period (AUCw) to quantify total cortisol output, although this is not strictly a measure of CAR because it is influenced by cortisol levels prior to awakening and not entirely by dynamic increases in cortisol that happen post-wakening (Stalder et al., 2016). In studies in which both CARi and AUCw were measured, their relationships to measures of psychosocial functioning were not the same (e.g., Bhattacharyya, Molloy, & Steptoe, 2008; Chida & Steptoe, 2009; Ellenbogen, Hodgins, Walker, Couture, & Adam, 2006; Mommersteeg et al., 2006; Sonnenschein et al., 2007; Whitehead, Perkins-Porras, Strike, Magid, & Steptoe, 2007), suggesting possible measure-dependent effects.

Yet another source of heterogeneity in the literature emerges from specific subytpes of psychosocial predictors having unique associations with the CAR. Chida and Steptoe (2009) described seven different types of psychosocial predictors: job stress, general life stress (non-work-related), depression, anxiety (including neuroticism and negative affect), fatigue/burnout/exhaustion, posttraumatic stress, and positive psychosocial traits. These different predictors may have different relationships with the CAR. For example, AUCw was positively related to general life stress but negatively related to posttraumatic stress (Chida & Steptoe, 2009). Similarly, CARi was positively correlated with job stress and general life stress but negatively correlated with fatigue, burnout, or exhaustion and not reliably associated with positive affect. Meta-analytic findings, therefore, suggest that psychosocial predictors can be related to higher or lower CAR, depending on the nature of the predictor.

Meta-analysis is designed to aggregate across individual findings and extract the commonalities from the idiosyncrasies. The latest meta-analysis on psychosocial functioning and CAR was a step in the right direction (Chida & Steptoe, 2009). However, a potential problem with meta-analysis is the file drawer problem, where studies are "filed away" unpublished if they fail to find statistically significant relationships. When studies are statistically significant, they make their way into the literature. Over time, this practice artificially inflates Type I error and may make it appear that relationships exist when in reality they do not (Rosenthal, 1979). To the degree that a meta-analysis includes a disproportionate number of Type I errors, its results − although superior to individual findings − may nonetheless be biased.

New statistical tools allow researchers to test a set of findings for evidential value in less biased ways. Specifically, *p*-curve analysis is based on the mathematical principle that if the null hypothesis is true, the probability of a *p*-value falling within a range of possible values is the same for all bins of the same size (Simonsohn, Nelson, & Simmons, 2014a; Simonsohn, Nelson, & Simmons, 2014b; Simonsohn et al., 2015). Thus, given that the null hypothesis is true (i.e., studies lack evidential value), the probability of *p* falling between 0.01 − 0.02 is equal to the probability of it falling between 0.02 − 0.03, 0.03 −0.04, or 0.04 − 0.05. In this case, a curve of all the significant *p*-values within a set of findings (i.e., a *p*-curve) will be flat. However, if a set of findings contains evidential value, more significant *p*-values from that set will fall in the 0.01 − 0.02 and 0.02 − 0.03 bins than in the 0.03 − 0.04 and 0.04 − 0.05 bins, and the *p*-curve will be positively skewed. Thus, by assembling the *p*-values for any set of significant findings and calculating the slope of that line, evidential value can be determined (Simonsohn et al., 2014a, 2014b, 2015). Moreover, if one knows all the *p*-values from a set of findings, one can approximate the distribution used to obtain those *p*-values, resulting in an unbiased estimate of effect size (Simonsohn et al., 2014a). This point estimate can then be compared to a meta-analytic point estimate to assess for systematic bias in the literature.

1. The current study

Although the extant literature suggests that psychosocial functioning and the CAR are associated, the nature of these associations remains unclear. In some studies, psychosocial variables were related to higher CAR and in other studies, to lower CAR (Chida & Steptoe, 2009). Moreover, the relationship between psychosocial functioning and the CAR may not be the same for different measures of the CAR, and may not be consistent for all psychosocial predictors. The current study aims to clarify the following questions regarding the relationships between psychosocial functioning and the CAR by combining *p*-curve and meta-analytic techniques: First, do the sets of findings associating psychosocial functioning to higher or lower CAR differ in evidential value? Second, what is the effect size for each of these sets of findings? Third, is there evidence of systematic bias in the extant literature? A secondary aim of the study was to examine predictor-specific effects.

2. Methods

2.1. Data sources and study selection

The current study updated the literature testing relationships between psychosocial functioning and the CAR. It included studies previously reviewed by Chida and Steptoe (2009) and added new findings from the past six years. Methods for updating the literature were based on those described by Chida and Steptoe (2009). Articles were identified using the search terms "cortisol" and "awakening" and "response" simultaneously (thus allowing the terms to be out of order) using the Endnote X7.2 (Thompson Reuters) software to search across the following databases: Medline, PsycINFO, Pubmed, and Web of Science. The search terms were allowed to appear in "Any field" for the Pubmed and PsycINFO databases, or in "Title/Keywords/Abstract" for Web of Science and Medline. Articles that were duplicates across these databases, published in foreign language journals, dissertations, conference abstracts, and letters to the editor were removed at a first pass. Reference sections of the identified articles were searched for other relevant articles. To avoid overlap with the previous meta-analysis, the search dates for the updated literature were restricted to articles published between October 1, 2008 and July 1, 2015. Inclusion criteria required articles to (1) be written in English, (2) be published in a peer-reviewed journal, (3) include at least one measure of psychosocial functioning, and (4) include at least one measure of the CAR. Articles were excluded if the population of interest had medical conditions/disorders known to influence cortisol, including pregnant or postpartum women, chronic pain disorders, circadian rhythm disorders, endocrine disorders, hormonal treatments, diabetes, cardiovascular disease, or other medical conditions. Biological predictors of the CAR (genetics or gene x environment interactions, family history of psychopathology, sleep patterns and disorders, chronotype), as well as intervention studies or effects of laboratory manipulations, were also excluded. Studies including participants with intellectual or developmental disabilities, dementia, or other cognitive impairments were excluded because there was insufficient research to synthesize these studies separately. Lastly, studies focusing on within-subject psychosocial changes were excluded because the current study focused on individual differences instead of intraindividual shifts, and there has recently been a thorough review on within-subject effects on the CAR (Law, Hucklebridge, Thorn, Evans, & Clow, 2013).

Where different articles reported on the same cohort (e.g., from large publicly available datasets), they were treated as multiple findings from the same study. In cases where articles met inclusion criteria but provided insufficient data for analysis, the corresponding authors were contacted. Of 32 authors contacted, 19 responded, and 17 provided additional data. In total, 159 articles met criteria for the updated literature review. These were combined with 54 articles from the Chida and Steptoe (2009) analysis. Thus, the current study included 709