Psychoneuroendocrinology (2013) 38, 941—946

Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

SHORT COMMUNICATION

Transformation techniques for cross-sectional and longitudinal endocrine data: Application to salivary cortisol concentrations

Robert Miller *, **Franziska Plessow**

Department of Psychology, Technische Universität Dresden, Germany

Received 13 June 2012; received in revised form 18 September 2012; accepted 18 September 2012

KEYWORDS
Transformation;
Distribution;
Scaling;
General linear model;
Salivary cortisol;
Secretory episode;
Psychosocial stress

Summary Endocrine time series often lack normality and homoscedasticity most likely due to the non-linear dynamics of their natural determinants and the immanent characteristics of the biochemical analysis tools, respectively. As a consequence, data transformation (e.g., log-transformation) is frequently applied to enable general linear model-based analyses. However, to date, data transformation techniques substantially vary across studies and the question of which is the optimum power transformation remains to be addressed. The present report aims to provide a common solution for the analysis of endocrine time series by systematically comparing different power transformations with regard to their impact on data normality and homoscedasticity. For this, a variety of power transformations of the Box—Cox family were applied to salivary cortisol data of 309 healthy participants sampled in temporal proximity to a psychosocial stressor (the Trier Social Stress Test). Whereas our analyses show that un- as well as log-transformed data are inferior in terms of meeting normality and homoscedasticity, they also provide optimum transformations for both, cross-sectional cortisol samples reflecting the distributional concentration equilibrium and longitudinal cortisol time series comprising systematically altered hormone distributions that result from simultaneously elicited pulsatile change and continuous elimination processes. Considering these dynamics of endocrine oscillations, data transformation prior to testing GLMs seems mandatory to minimize biased results.

1. Introduction

Statistical analyses of endocrine data by utilization of parametric models based on the general linear model (GLM), for example, conventional analyses of variance (ANOVAs), require several assumptions in order to reliably infer on the presence of hypothesized effects (Harwell

* Corresponding author at: Department of Psychology, Technische Universität Dresden, D-01062 Dresden, Germany.
Tel.: +49 351 463 33708; fax: +49 351 463 37274.
E-mail addresses: miller@biopsych.tu-dresden.de (R. Miller), plessow@biopsych.tu-dresden.de (F. Plessow).

0306-4530/$ — see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.psyneuen.2012.09.013

et al., 1992; Sakia, 1992). Among these, homoscedasticity and normality of models' residuals are of particular importance when analyzing endocrine time series. First of all, any manifestation of an endocrine signal is subject to heteroscedasticity, that is, an overproportional increase of variance with growing hormone concentrations, which is likely to be caused by inherent properties of biochemical measurement tools (Miller et al., in press). Second, hormone distributions at specific sampling points have repeatedly been shown to violate normality (e.g., Hanson et al., 2000; Mueller et al., 2011). The inherently non-normal distribution pattern of hormone concentrations across time is assumed to result from the complex non-linear dynamics of secretory pulses and hormone elimination (Brown et al., 2001).

These shortcomings in meeting the assumptions of GLM-based analysis led to the development of sophisticated computational procedures, for example, the deconvolution analysis (Veldhuis, 1997), which are preferable as compared to GLM-based analyses but entail other design complications (e.g., require more extensive and frequent sampling than feasible in psychophysiological studies). Thus, the utilization of basic non-parametric methods would offer a compromise, as they are robust to the presence of non-normality and heteroscedasticity. However, they are less powerful as compared to parametric methods.

In order to circumvent these issues, and consequently to generate endocrine variables that allow application of GLM-based analyses, many researchers have dealt with heteroscedasticity and non-normality in a rather data-driven way by applying variance stabilizing and normalizing transformations (e.g., Hanson et al., 2000; Mueller et al., 2011; Plessow et al., 2012a). The log-transformation, in particular, represents a popular account, but the currently observable inconsistent application of transformations points toward a lack of consensus on what technique is to be considered as appropriate.

Therefore, with the present report, we intend to resolve some ambiguity on the ''right'' transformation technique for endocrine time series by systematically evaluating which transformation most effectively reduces heteroscedasticity and, at the same time, generates endocrine variables that are likely to follow Gaussian (normal) distributions. For this, we applied power transformations of the Box–Cox family (Box and Cox, 1964; Sakia, 1992) to salivary cortisol concentrations being longitudinally sampled before and after induction of psychosocial stress with the Trier Social Stress Test (TSST; Kirschbaum et al., 1993). Such cortisol data are particularly suitable for demonstrating the outlined transformation approach, as they feature the typical characteristics of endocrine time series, that is, pulsatile change and continuous elimination of hormones (Brown et al., 2001). Thus, it enables to check for systematically occurring shifts of hormonal concentrations' distributional properties at baseline (i.e., hypothalamus–pituitary–adrenal [HPA] axis activity prior to stress induction), which result from interindividually varying, randomly occurring pulsatile oscillations, toward systematically altered hormonal distributions being subject to a substantial fraction of simultaneously elicited secretory pulses (to investigate HPA axis reactivity across time).

2. Methods

2.1. Sample

Salivary cortisol data from previously published (Mueller et al., 2011; Plessow et al., 2011, 2012a, 2012b) and unpublished studies were merged. This data set comprised cortisol samples of 309 participants, which were obtained in temporal proximity to the TSST. More specifically, five samples were taken from each participant at $t_0 = -6$ min, $t_1 = +16$ min, $t_2 = +25$ min, $t_3 = +35$ min, and $t_4 = +45$ min relative to TSST onset. Sample t_0 is considered to reflect a random (baseline) cortisol sample within the diurnal cortisol course, as it would be employed in clinical practice for reference. By the time of saliva sampling, all participants were aged between 18 and 65 years, and had a body mass index below 30 kg/m^2. The proportion between male and female participants was approximately balanced. All test sessions were conducted in the afternoon. All participants reported absence of smoking, no intake of HPA axis-altering medication (including hormonal contraceptives for females), and declared their written informed consent prior to testing.

2.2. Biochemical analyses

Salivary cortisol was determined on site at the endocrine laboratory in Dresden, Germany, by utilization of a commercially available chemiluminescence immunoassay (IBL, Hamburg, Germany), which was conducted according to the protocols provided by the manufacturer. Intra- and inter-assay coefficients of variation were below 10%. Operational range was 0.50 to 110.40 nmol/l implying that no measured cortisol concentration was below or above quantification limits.

2.3. Statistical analyses

Box–Cox power transformations[1] being defined as $X' = (X^{\lambda} - 1)/\lambda$ for $\lambda \neq 0$, and $X' = \log_e(X)$ for $\lambda = 0$ (see Sakia, 1992), were applied to all cortisol sampling points by employing a series of λ-values ranging from $\lambda = 1$ (data remain untransformed) to $\lambda = 0$ (log-transformation of data), resulting in transformed variable distributions, which were submitted to further analyses. Considering that most between-subject factors examined in psychophysiological research (e.g. sex) do not account for structural effects on hormone distributions, simple fixed-effects ANOVAs were employed to investigate the overall change of cortisol concentrations across time. In order to acquire powerful overall criteria to check for presence of heteroscedasticity and deviation from normality, we calculated χ^2-statistics from Breusch–Pagan (Breusch and Pagan, 1979) and Doornik–Hansen tests

[1] Speaking of Box–Cox power transformations, the authors are referring to the family of power functions defined by Box and Cox (1964; formula 1), but not to Box–Cox's procedure to obtain the maximum-likelihood estimates of a parameter λ. This family is extremely flexible as it represents a generalized form of the most popular transformations (i.e. inverse-, log- and root-transformations).