Psychoneuroendocrinology (2005) **30**, 698–714

PNEC

www.elsevier.com/locate/psyneuen

Estimating between- and within-individual variation in cortisol levels using multilevel models

Daniel J. Hruschka*, Brandon A. Kohrt, Carol M. Worthman

Department of Anthropology, 1557 Dickey Drive, Emory University, Atlanta, GA 30322, USA

Received 2 June 2004; received in revised form 20 November 2004; accepted 3 March 2005

KEYWORDS
Cortisol;
Mixed effects;
Multilevel;
Hierarchical linear model

Summary Cortisol measures often are used to examine variation in hypothalamic–pituitary–adrenal axis (HPA) activity as well as broader patterns of differential health. However, substantial within-individual variation renders single cortisol measurements unreliable as estimates for probing differences between individuals and groups. A standard practice to clarify between-individual differences involves collecting multiple samples from each participant and then deriving person-specific averages. By ignoring information about variation at between- and within-individual levels, this technique impedes cross-study comparison of results, ignores data useful for future study design, and hinders the analysis of cross-level interactions. This report describes how multilevel approaches can simultaneously model between- and within-individual variation in diurnal cortisol levels without using crude averages. We apply these models to data from children in Nepal ($n=29$, 11–15 samples per child), Mongolia ($n=47$, 8–12 samples per child) and the US ($n=1269$, 1–6 samples per child). Using the Nepal data, we show how an analysis of crude time-adjusted aggregates does not detect an association between aggressive behavior and cortisol levels, while a multilevel analysis does. More importantly, we argue that the 'roadmap' to variation generated by these multilevel models provides meaningful information about the predictive accuracy—not just statistical significance—of relationships between cortisol levels and individual-level variables, such as psychopathology, age, and gender. The 'roadmap' also facilitates comparison between the results from different studies and estimation of the necessary number of cortisol measurements for future investigations.

1. Introduction

As an essential and measurable product of the HPA system, cortisol has become a key variable in studies of human response to the environment (Flinn and England, 1997; Gunnar, 2001; Kirschbaum and Hellhammer, 2000; Pollard, 1995; Worthman, 1999). Such studies have established

Abbreviations: ANOVA, analysis of variance; CV, coefficient of variation; GSMS, Great Smoky Mountain Study; HLM, hierarchical linear model; HPA, hypothalamic–anterior pituitary–adrenal; ICC, intra-class correlation coefficient; LNCORT, natural log of cortisol; ML, maximum likelihood; OLS, ordinary least squares; REML, restricted maximum likelihood.

* Corresponding author. Tel.: +1 404 636 0144; fax: +1 404 727 2860.
E-mail address: dhrusch@sph.emory.edu (D.J. Hruschka).

0306-4530/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.psyneuen.2005.03.002

the existence of stable between-individual differences in several aspects of cortisol secretion, including overall mean or basal levels (Kirschbaum et al., 1990), diurnal trends (Smyth et al., 1997; Stone et al., 2001), and response to stressors (Kirschbaum et al., 1998) or morning awakening (Kudielka and Kirschbaum, 2003; Wust et al., 2000). Between-individual differences in these aspects of cortisol secretion have been associated with a variety of health measures, including post-traumatic stress disorder (Carrion et al., 2002; Glover and Poland, 2002; Yehuda et al., 2000) and other psychiatric conditions (Goodyer et al., 1996, 2001; Harris et al., 2000; Weber et al., 2000), physical illness (Flinn and England, 1997; Heim et al., 2000), and psychosocial functioning (Adam and Gunnar, 2001; Decker, 2000; Kiecolt-Glaser et al., 1997; Koertge et al., 2002; Melamed et al., 1999; Nicolson and Van Diest, 2000; Pruessner et al., 1999; Schulz et al., 1998; Van Eck et al., 1996).

Consequently, comparisons involving cortisol levels assume that individuals differ in their patterns of cortisol secretion and that these differences exhibit some stability over time. However, momentary assessments of cortisol depend on a number of factors, including whether the person recently consumed food or caffeine, the time of day at which the measurement was taken, the person's general basal cortisol level, whether infectious or inflammatory processes are active, or whether the person was currently anticipating a stressful situation (Pollard, 1995). The measurement can also reflect error due to the processes involved in extracting, storing and analyzing samples. These factors can be divided into three rough categories—between-individual differences, within-individual variation, and measurement error.

Such complexities present well-recognized challenges to study design and interpretation of cortisol data. The relative contributions of each of these three sources of variation—individual difference, acute effects, and method error—cannot be parsed by a single measure from each member of a population. Collecting multiple measurements from each individual in a sample allows comparison of the amount of variation that exists *between* individual means relative to the amount of variation observed *within* each individual. If variation in individual means is high relative to within-individual variation then we can be confident that a single measure of cortisol reflects stable between-individual differences in tonic levels (see Fig. 1). If, however, within-individual variation is high relative to between-individual variation, then any single measure of cortisol tells us less about an individual's tonic levels and more about the situation in which the cortisol measurement was taken.[1]

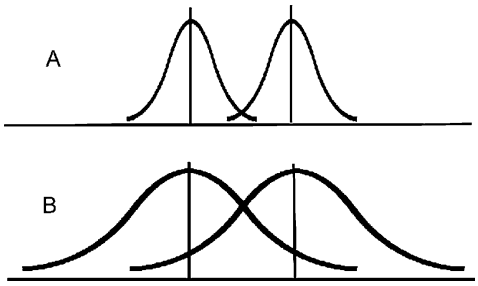

Fig. 1 Within-individual variation and the estimation of mean differences. A and B represent two possible distributions of measurements on two individuals. The difference in means between the two individuals is the same in A and B. In situation B, however, the within-individual variation is so great that the person with the lower mean will often have a higher measurement than the person with the greater mean. This would rarely occur in situation A.

A clear understanding of these sources of variation prevents incorrectly interpreting within-individual variation as a signal of between-individual difference. Conversely, it avoids the erroneous conclusion that *all* variations in cortisol measurements is a result of within-individual variation or measurement error, and that we can say nothing reliable about individual differences.

Researchers aiming to assess between-individual differences in HPA activity have designed studies to control (experimentally and/or statistically) for the known and unknown situational factors that can generate within-individual variation. A standard practice in research conducted in naturalistic settings is to collect repeated observations on each individual to estimate reliable individual cortisol level means or medians (Decker, 2000; Dettling et al., 2000; Fisher et al., 2000; Flinn and England, 1997; Harris et al., 2000; Koertge et al., 2002; McBurnett et al., 2000; Watamura et al., 2002; Weber et al., 2000; Wolf et al., 2002; Wust et al., 2000). *However, simply analyzing aggregates removes information about within-individual variation which can be used (1) to improve interpretations of current results, (2) to inform future study design, and (3) to explore interactions across levels of analysis.*

In this paper, we describe how a class of models, alternatively referred to as multilevel

[1] Within-individual variation described here also includes variation due to measurement error. Most studies that have examined this issue show that measurement error contributes very little (<5%) of the total variance in cortisol measures (Kirschbaum et al., 1990).