

# Handbook of Physiological Research Methods in Health Psychology

## Measurement of Cortisol

Contributors: Linda J. Luecken & Linda G. Gallo

Print Pub. Date: 2008

Online Pub. Date: June 22, 2009

Print ISBN: 9781412926058

Online ISBN: 9781412976244

DOI: 10.4135/9781412976244

Print pages: 37-75

This PDF has been generated from SAGE knowledge. Please note that the pagination of the online version will vary from the pagination of the print book.

Arizona State UniversitySan Diego State University  
10.4135/9781412976244

[p. 37 ↓ ]

## Chapter 3: Measurement of Cortisol

Nancy A. Nicolson

### Introduction to the Hypothalamic-Pituitary-Adrenocortical Axis

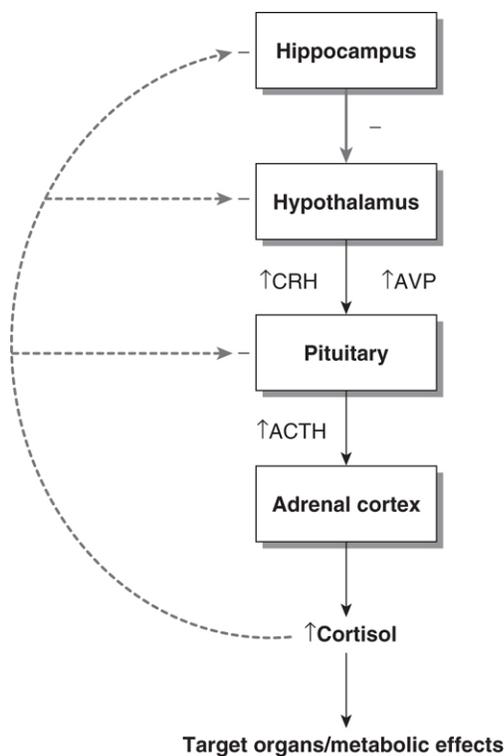
The hypothalamic-pituitary-adrenocortical (HPA) axis and its end product, cortisol, are thought to be important mediators of the relationship between stressful life experiences and health outcomes. The HPA response is a component of the organism's adaptive system for maintaining function under changing environmental circumstances. Over the long term, however, chronic overactivation following repeated stressors can give rise to wear and tear or *allostatic load* (McEwen, 2003). Both maladaptive responses to stress and disturbances in the functioning of the HPA axis have been implicated in a wide variety of syndromes and illnesses, including cardiovascular illness, insulin resistance syndrome and diabetes, cognitive decline during aging, fatigue and pain syndromes, and psychiatric disorders such as depression and posttraumatic stress disorder (PTSD), among others (Charmandari, Tsigos, & Chrousos, 2005).

As the name indicates, the main components of the HPA axis are the hypothalamus, the pituitary gland, and the adrenal cortex (see [Figure 3.1](#)). The hypothalamus releases corticotropin-releasing hormone (CRH, also known as CRF) into the portal blood vessels connecting the hypothalamus to the anterior pituitary. CRH, which works synergistically with arginine vasopressin (AVP) released from the hypothalamus, then triggers the pituitary to secrete adrenocorticotrophic hormone (ACTH) into the bloodstream. After reaching the adrenal cortex, ACTH stimulates the release of glucocorticoids (GCs)—in humans, cortisol. This entire process takes place within a

matter of minutes. The HPA axis is regulated by a complex negative feedback system, with circulating glucocorticoids inhibiting activity at the level of the hippocampus, the hypothalamus, and the pituitary. In general, hippocampal structures exert inhibitory influences on the axis at the level of the hypothalamus, whereas the amygdala plays an activating role (Herman & Cullinan, 1997). Mineralocorticoid (MR) and glucocorticoid (GR) receptors in the brain are thought to play different but complementary roles in regulating normal circadian activity, preparing the organism to respond to external stimuli, and facilitating recovery of disturbed homeostasis after acutely stressful situations (de Kloet, 1991).

[p. 38 ↓ ]

Figure 3.1 Schematic Overview of the Hypothalamic-Pituitary-Adrenocortical (HPA) Axis



NOTE: CRH = corticotropin-releasing hormone, AVP = arginine vasopressin, ACTH = adrenocorticotrophic hormone. Dashed lines indicate negative feedback effects.

Activity of the HPA axis shows a pronounced circadian rhythm, controlled by the primary endogenous pacemaker, the suprachiasmatic nucleus. ACTH and cortisol are secreted in short pulsatile episodes, concentrated in the morning hours in humans, but occurring throughout the day, even in the absence of stressors. In a 24-hour cycle, approximately 15 to 18 ACTH pulses can be identified. In people who have a normal routine of nocturnal sleep and daytime activity, cortisol levels are lowest between 10 p.m. and 4 a.m. After a quiescent period of HPA activity lasting from 2.5 to 6 hours (Linkowski et al., 1985), cortisol levels begin to rise several hours before awakening, with an additional sharp increase in the 30 to 40 minutes following awakening. Thereafter, cortisol levels steadily decrease, except for a moderate rise following lunch. Although cortisol levels decline over the rest of the afternoon and throughout the evening until sleep onset, the slope of the diurnal curve is relatively flat compared to the morning hours.

## A Brief Overview of Research Approaches

Because of its central role in regulating the psychobiological stress response, the HPA axis is one of the most heavily investigated physiological systems in health psychology and psychiatry. Hans Selye's conception of the general adaptation syndrome, in particular, [p. 39 ↓ ] called attention to the importance of the HPA axis in regulating a wide range of bodily functions and their disturbance by acute physiological stressors, such as exposure to toxins (Selye, 1956). A deeper understanding of the effects of psychological stress on the HPA axis, however, began to emerge in the 1960s, when now-classic studies employed new methods to assess endocrine responses to stress in rodents, nonhuman primates, and humans (Levine, 2000; Mason, 1968; Rose, 1984). The widespread involvement of the HPA axis in both healthy adaptation and common disorders, combined with the increasing ease with which its activity can be measured, have led to an enormous growth over the last two decades in research on this system, in settings ranging from the laboratory to the community.

Research approaches include studies of spontaneous hormone secretion throughout the day, pharmacological manipulations to determine how feedback mechanisms are functioning, and studies of reactivity to acute real-life or experimental stressors. Assessment of the HPA axis at multiple levels is not feasible in most studies, because of the invasive procedures involved. Ignoring the vast literature on animal models and clinical research, this review focuses on methods that can be more generally applied by health psychologists studying human subjects in a wide variety of real-life and laboratory settings, without undue inconvenience or risk to the research participants or the need for specialized medical personnel. This means that measures of CRH, ACTH, GC receptor characteristics, or responses of the HPA axis to challenge tests in which CRH, ACTH, or other substances are administered are not covered, despite their utility in psychoneuroendocrine research and clinical studies. Furthermore, this chapter does not discuss the rationale or procedures for measuring dehydroepiandrosterone (DHEA), a steroid hormone produced primarily by the adrenal cortex, although there is evidence that DHEA may counteract some of the effects of elevated glucocorticoids and play a role in stress-related disorders such as depression and chronic fatigue (Goodyer, Park, Netherton, & Herbert, 2001; Khorram, 1996; Wolkowitz, Brizendine, & Reus, 2000).

This chapter focuses specifically on cortisol, the end product of the HPA axis. As a cautionary note, it is important to realize that cortisol is a peripheral measure and secretory patterns can be deviant in the statistical sense without necessarily reflecting dysregulation at a higher level. In some cases, apparent abnormalities may be the result of an adaptive response to environmental demands. On the other hand, cortisol levels can also be perfectly normal when other probes indicate regulatory abnormalities; excessive CRH or ACTH secretion might, for example, be coupled with decreased adrenal sensitivity. The HPA axis is a complex and dynamic system, and cortisol measures can provide only a partial window into how this system is regulated—or dysregulated.

# Investigating Spontaneous Activity of the HPA Axis

## Basal Cortisol Levels

Researchers have long been interested in obtaining overall basal measures of glucocorticoid output, as overactivation of the HPA axis resulting from chronic stress or illness was expected to result in higher levels of circulating cortisol. Because of the inherent novelty of hospital settings as well as the trouble and expense of bringing healthy subjects to the clinic, ambulatory procedures have distinct advantages. Numerous studies have used 24-hour urinary measures or repeated salivary sampling to examine genetic, developmental, and especially environmental influences on HPA activity in healthy adults and children. Others have investigated HPA [p. 40 ↓] abnormalities in stress-related disorders. It is now clear that not only hypercortisolism but also hypocortisolism can occur, for example in PTSD, pain, and fatigue syndromes. The processes by which stress could lead to such divergent outcomes are still poorly understood, but recent reviews have summarized a number of hypotheses (Fries, Hesse, Hellhammer, & Hellhammer, 2005; Gunnar & Vazquez, 2001; Heim, Ehlert, & Hellhammer, 2000; Yehuda, 2002). For example, hypocortisolism could be the long-term effect of adverse early experiences that permanently alter the axis. Down-regulation might even be seen as a protective mechanism, set in motion following long-term hyperactivation to reduce the negative effects of allostatic load. Alternatively, hypocortisolism might represent a preexisting risk factor, of genetic or early developmental origin, which later undermines the individual's ability to respond adaptively to trauma or chronic stressors.

# Circadian Rhythm and Diurnal Patterns of HPA Axis Activity

In addition to overall cortisol levels, the diurnal patterning of hormone secretion can provide important clues to HPA axis dysregulation. Sophisticated chronobiological analyses of circadian rhythms (see, e.g., Posener et al., 2000; Van Cauter, Leproult, & Kupfer, 1996) require more frequent sampling than is feasible in ambulatory settings, not to mention the problem of obtaining nocturnal measures. For this reason, simpler measures of the shape of the diurnal curve are more frequently employed, in particular the steepness of the decline in cortisol levels from morning to evening. Loss of diurnal variation, as reflected in flatter slopes, has been reported in various disorders and at-risk groups (Bower et al., 2005; Sephton, Sapolsky, Kraemer, & Spiegel, 2000).

Even if the diurnal slope is not of direct relevance to the goals of a study, collecting several samples over the course of a day is good practice; differences between groups being compared may be restricted to a certain time of day, which often cannot be predicted on theoretical grounds. For this reason, studies with only a single diurnal sampling time will inevitably raise questions about how results generalize to the rest of the day.

## Cortisol Response to Awakening

In recent years, interest has been growing in the cortisol awakening response (CAR). Cortisol levels rise sharply (50–160% in saliva) during the first 30 to 40 minutes after wakeup, returning to the awakening baseline within 60 to 75 minutes, and declining more gradually thereafter (Clow, Thorn, Evans, & Hucklebridge, 2004; Pruessner et al., 1997; Wüst et al., 2000). The function of the CAR is not yet clear, but general agreement is that this response is a discrete aspect of cortisol's circadian rhythm, with its own regulatory processes (Clow et al., 2004; Schmidt-Reinwald et al., 1999).

The CAR appears to be moderately stable within persons, from day to day and over longer periods of several weeks to months, and it has a clear genetic component (Wüst

et al., 2000). Nevertheless, it can vary in relation to short-term influences, such as the stressfulness of a workday compared to a weekend (Kunz-Ebrecht, Kirschbaum, Marmot, & Steptoe, 2004), or an early-shift compared to a late-shift workday (Williams, Magid, & Steptoe, 2005). In addition, the CAR may be either enhanced or blunted in chronic stress, burnout, depression, and other disorders (e.g., Bhagwagar, Hafizi, & Cowen, 2005; Grossi et al., 2005; Pruessner, Hellhammer, Pruessner, & Lupien, 2003; Stetler & Miller, 2005).

[p. 41 ↓ ]

## Within-Person Variability

One important aspect of spontaneous cortisol secretion that has received relatively little research attention, despite its potential significance as an index of HPA dysregulation, is within-person variability. Greater irregularity in within-day cortisol measures has been observed in affective disorders, even when overall levels are normal (Peeters, Nicolson, & Berkhof, 2004; Posener et al., 2004; Yehuda, Teicher, Trestman, Levengood, & Siever, 1996), and may predict worse clinical outcomes (Goodyer, Tamplin, Herbert, & Altham, 2000). There is some evidence that a subset of individuals lacks a consistent diurnal slope pattern (Smyth et al., 1997), but day-to-day variation in cortisol measures remains largely unexplored. One major obstacle is that investigating within-person variability requires many more samples per person.

## Summary

The degree of detail with which a given study is able to characterize spontaneous cortisol secretory patterns depends on its specific goals, but also on the available budget and logistical considerations. Thus, large epidemiological surveys are often restricted to obtaining only a few samples per subject and may have to choose between the response to awakening and/or a diurnal slope measure (either of which can be estimated with a minimum of two saliva samples; see, e.g., Young & Breslau, 2004), perhaps in combination with a urinary measures if nighttime or total cortisol secretion are of interest. At the other extreme, intensive daily process designs may collect 60 or

more saliva samples per subject in order to estimate not only overall levels and diurnal slopes, but also the association between cortisol at a particular point in time with current mood, symptoms, daily hassles, and uplifts (Smyth et al., 1998; van Eck, Berkhof, Nicolson, & Sulon, 1996). As we don't yet know which measures of spontaneous cortisol secretion are most relevant for understanding disease processes, a conservative approach would be to obtain reliable measures of cortisol basal levels, diurnal slopes, and the CAR in the same protocol (see Sampling Strategy under A Framework for Designing a Study and Interpreting the Results).

The availability of noninvasive sampling methods (described in Measuring Activity of the HPA Axis) has greatly increased the range of research applications. These include cross-cultural field studies (Flinn, 1999; Hruschka, Kohrt, & Worthman, 2005), large-scale longitudinal studies in the community (Rosmalen et al., 2005), intervention studies (Carlson, Speca, Patel, & Goodey, 2004; Gaab et al., 2003), and prediction of disease outcomes (Sephton et al., 2000).

## Sensitivity of the HPA Axis to Negative Feedback

Measuring the response of the HPA axis to synthetic glucocorticoids provides a measure of the strength of negative feedback inhibition. Following an oral dose of 1 mg dexamethasone late in the evening, cortisol levels are normally suppressed the next day; incomplete suppression or early escape from suppression indicates deficits in feedback regulatory mechanisms. The dexamethasone suppression test (DST) was originally developed as a diagnostic tool in major depression, a disorder in which hypercortisolism is often observed (Carroll et al., 1981). A low-dose (0.25–0.5 mg) version of the DST has been used to investigate more subtle deficits in feedback regulation in individuals with chronic stress (Powell et al., 2002) or to determine whether sensitivity of the HPA axis to glucocorticoid negative feedback is heightened in disorders in which hypocortisolism is more frequently observed, such as PTSD (Yehuda et al., 1993) or chronic fatigue syndrome (Gaab et al., 2002).

[p. 42 ↓ ]

# Response of the HPA Axis to Acute Stressors

Early studies of the HPA axis focused on the hormonal response to acute stressors, and this remains one of the primary interests of health psychologists. In humans, the cortisol response to stress can be studied in real life or under more controlled conditions in the laboratory. Compared to the quick but short-lived response of the catecholamines, the cortisol response to acute stress is relatively slow. Within minutes of the onset of a discrete stressful stimulus or event, such as public speaking, cortisol levels begin to rise, superimposed on the diurnal profile of basal HPA activity. After termination of the stressor, cortisol levels gradually return to their prestress baseline; full recovery can take an hour or more, in part reflecting the approximately one-hour half-life of cortisol in blood or saliva.

Basal levels of glucocorticoids act permissively to prepare the individual to respond to a stressful episode. The cortisol response to stress mobilizes energy for coping with the stressor, but also shuts down the initial fight or flight responses of the sympathetic nervous and immune systems to prevent them from overshooting and damaging the organism (Munck, 2000). Glucocorticoid release during stress is thus primarily a protective response. If, however, cortisol levels are delayed in their poststress recovery, or repeated stress exposures result in sensitization instead of habituation of the HPA axis, a chronic hyperactivation of this system can be maladaptive, leading to stress-related disorders (McEwen, 2003).

## What Kinds of Stimuli Activate the HPA Axis?

It is a common misconception, probably going back to the work of Hans Selye (1956), that the HPA axis will respond to all types of stressful experiences and the acute cortisol response can therefore serve as the gold standard for determining whether a particular experience was stressful. Many physiological systems are involved in stress

responses, and each system varies in terms of the types of stressors that activate it, its temporal dynamics, and its relations to other systems (Baum & Grunberg, 1995). For example, aversive stimuli that activate the sympathetic nervous system and adrenal medulla, producing elevations in heart rate, blood pressure, and catecholamines, do not necessarily lead to measurable changes in cortisol.

Certain types of psychosocial stressors do have consistent effects. Reviews of early studies in humans, rodents, and nonhuman primates concluded that situations characterized by novelty, unpredictability, or low perceived control were most likely to activate the HPA axis (Mason, 1968; Rose, 1984). A recent metaanalysis of experimental studies showed that social-evaluative threat during task performance and low control over the situation were the two best predictors of acute cortisol responses in humans (Dickerson & Kemeny, 2004). Although an individual's appraisal of the stressor, coping, and degree of distress are predicted, on the basis of transactional stress theory (Lazarus & Folkman, 1984), to moderate or (in the case of distress) mediate the cortisol response, laboratory studies have shown surprisingly low correlations between individual self-reports of these variables and cortisol measures.

Physical stressors such as intense exercise also activate the HPA axis. The observation that cortisol elevations are often greater during competitive sports than during training at the same level of physical exertion (Cook, Ng, Read, Harris, & Riad-Fahmy, 1987) indicates that physical and psychosocial components of competition have additive effects. Cortisol levels also increase following experimentally induced pain (al'Absi, Petersen, & Wittmers, 2002).

[p. 43 ↓ ]

# Measuring Activity of the HPA Axis

## Salivary Cortisol

### Background

The first assays for salivary steroids were described in 1959, but this method did not gain widespread acceptance until researchers at the Tenovus Institute in Wales developed reliable assays for steroids in small volumes of whole saliva (Riad-Fahmy, Read, Walker, & Griffiths, 1982; Walker, Riad-Fahmy, & Read, 1978). Over the past 20 years, there has been an explosive growth in the number of studies using salivary measures to assess cortisol levels in a wide variety of applications in psychology, psychiatry, endocrinology, and beyond. The advantages of salivary cortisol sampling, compared to traditional procedures for blood sampling, have been summarized in several reviews (Kirschbaum & Hellhammer, 1989, 1994; Vining, McGinley, Maksvytis, & Ho, 1983). In addition to the ease and noninvasive nature of sample collection, the fact that salivary cortisol is “free,” unbound by corticosteroid-binding globulin (CBG) or other carriers, is advantageous, as free cortisol thus represents the biologically active fraction of the hormone (Mendel, 1989).

As noted earlier, salivary cortisol is ideal for assessing acute responses to experimental stressors. In addition, repeated measurement by subjects in their daily environment allows a good estimate of basal levels, diurnal variation, and response to awakening; some naturalistic designs also permit individual estimates of day-to-day variability and stress reactivity.

### Comparison with Blood Measures

Cortisol levels measured in saliva correlate highly with free cortisol in blood. However, because of partial conversion of cortisol to cortisone during passage through the

salivary glands, the absolute level of free cortisol in saliva is 10% to 35% lower than it is in blood (Vining et al., 1983). Correlations with total blood concentrations (bound and free fractions) are also high, but the slope of the regression line becomes steeper at higher cortisol concentrations, after CBG-binding sites in blood are fully occupied. CBG levels can vary both within and between individuals, for example during pregnancy or with oral contraceptive use.

Movement of cortisol from blood to saliva occurs by passive diffusion, so that salivary levels are independent of the flow rate of saliva (Vining et al., 1983). Changes in plasma and salivary cortisol levels are closely synchronized. After injections of cortisol, salivary levels increased within 1 minute (Walker, 1984), and peak concentrations in blood are seen 2 to 3 minutes later in saliva (Kirschbaum & Hellhammer, 2000). Cortisol responses to awakening and to meals appear to be more pronounced in salivary than in plasma measures, and salivary cortisol returns to baseline more slowly after psychosocial stressors (Kirschbaum & Hellhammer, 2000).

## Collection

The popularity of salivary cortisol measures is largely due to the ease of collecting samples from participants in both laboratory and field settings. A number of different techniques for collecting saliva samples have been described; which is most appropriate for a given research question will depend on characteristics of the participants, the setting, and frequency with which samples will be collected, and whether other substances will be measured in the same samples.

Saliva samples are usually obtained from infants and toddlers with pipettes or other devices that aspirate saliva from the floor of the mouth, cotton ropes, swabs, or sponges held by the researcher or parent (Gunnar & Talge, 2007). In older children and adults, [p. 44 ↓] cotton dental rolls—including the widely used salivettes® (Sarstedt, Nümbrecht, Germany)—have convenient features for both research participants and laboratory personnel. Because of aspecific binding to the swabs or interference by other substances such as phytoestrogens that may be present, cotton salivettes or swabs should not be used when other steroids (e.g., DHEA, testosterone, progesterone) or salivary immunoglobulin A (IgA) are also being measured (Shirtcliff, Granger, Schwartz,

& Curran, 2001). In 2007 Sarstedt began production of a new synthetic salivette, designed to eliminate the risk of batch-to-batch variation in the performance of cotton swabs. Swabs are unnecessary if participants can collect saliva by drooling into a tube, either directly or through a straw. Drooling may be less acceptable in studies where repeated samples need to be collected as rapidly and unobtrusively as possible, for example, during participants' daily activities outside the home. In one comparison, subjects collected adequate amounts of saliva in 1 to 2 minutes with cotton salivettes or cellulose-cotton tip "eyespears," whereas passive drooling took from 1 to 15 minutes to produce the same amount (Strazdins et al., 2005). For all collection methods, it is important that the plastic storage tubes and stoppers are made of materials, such as unrecycled polypropylene (IBL, 2006), that do not absorb the hormone. Stoppers also need to fit tightly, because evaporation of saliva will lead to inaccurate cortisol results.

Most cortisol assays require only 20 to 50  $\mu$ l of saliva per tube, and therefore twice these amounts for a duplicate assay. In practice, larger volumes of saliva need to be collected when cotton-based methods are used, because up to 450  $\mu$ l of saliva can remain in the cotton after centrifuging (de Weerth, Graat, Buitelaar, & Thijssen, 2003). Specialized techniques make it possible to extract cortisol from smaller sample volumes, which may be a great advantage in studies of infants (de Weerth et al., 2003). In subgroups with low spontaneous flow rates (e.g., babies and small children, depressed patients, the elderly), saliva flow can be stimulated with powdered drink mix crystals, candies containing citric acid, or lemon juice. Salivettes prepared with citric acid are also meant to stimulate salivary flow. Extreme caution is warranted in using such procedures, however, as they can lower the pH of the resulting saliva sample. Many currently available immunoassays produce false high values when sample pH is lower than 3.5 to 4 (Kirschbaum & Hellhammer, 2000; Schwartz, Granger, Susman, Gunnar, & Laird, 1998; Talge, Donzella, Kryzer, Gierens, & Gunnar, 2005; Vialard-Miguel, Belaidi, Lembeye, & Corcuff, 2005). Chewing on an inert substance (e.g., plain salivette, sugarless chewing gum, parafilm) or just making chewing movements are good alternatives for stimulating salivary flow.

## Instructions to Subjects

Subjects should be trained how to collect saliva samples and given the opportunity to practice under supervision to ensure that they collect adequate volumes of saliva. With salivettes, subjects should be instructed to chew lightly on the swab and to keep it fully inside the mouth until it feels saturated. (This can take 1 to 2 minutes, depending on salivary flow rate.) It is standard practice to ask subjects not to brush their teeth in the 30 minutes before scheduled collection of a salivary sample. Acidic drinks, milk, and use of inhaled steroids (as examples of substances that could interfere with assay performance) should be avoided shortly before taking a saliva sample. If rinsing with water is considered necessary, it should be done at least 10 minutes before saliva collection to avoid diluting the cortisol concentration. Recent food intake and smoking can influence cortisol responses to acute [p. 45 ↓ ] stressors and possibly to morning awakening and should be avoided in the hour before sampling. It is crucial that subjects close the tubes tightly and label them with the exact time the sample was collected! Clear instructions should be given concerning storage (i.e., whether tubes should be kept at room temperature, in the refrigerator, or in the home freezer); samples should not be left exposed to heat or sunlight.

## Storage and Handling

Saliva samples can be stored at room temperature (RT) or in participants' home refrigerator or freezer until they are mailed or delivered to the lab. Estimates of how long cortisol is stable at RT range from 7 days (Groschl, Wagner, Rauh, & Dorr, 2001) to at least 4 weeks (Kirschbaum & Hellhammer, 2000). Centrifuging samples before storage appears to prolong the stable period (Groschl et al., 2001); nevertheless, increasing variance as well as decreasing levels over time indicate that storage at RT for more than 2 to maximally 4 weeks should be avoided (Garde & Hansen, 2005). Salivette samples develop mold and a bad odor after about 4 days at RT; this does not affect the cortisol concentrations, but makes the work of lab technicians unpleasant.

The benefits of refrigeration at 4° to 5°C, compared to RT, are unclear. In one study (Groschl et al., 2001), cortisol levels decreased in samples refrigerated for 11 days or longer; in contrast, Garde and Hansen (2005) found no change in cortisol levels in polyester salivettes refrigerated up to 3 months. Freezing clearly prolongs the stability of salivary cortisol. In samples frozen at either -20° or -80°C, cortisol concentrations remain stable for 9 months (Aardal & Holm, 1995) to 1 year (Garde & Hansen, 2005); freezing for as long as 2 years is probably possible.

In settings where there is no access to refrigerators or freezers, stability of samples can be prolonged by adding preservatives such as sodium azide (Groschl et al., 2001), citric acid (alone or with sodium benzoate), or ethyl and propyl paraben (Nimmagudda, Ramanathan, & Putcha, 1997). Cortisol in samples treated with citric acid and sodium benzoate remained stable for 180 days at RT (Nimmagudda et al., 1997). As noted earlier, preservatives, especially those that lower pH, may invalidate certain assays. Blood spots offer an alternative to saliva when extended storage at RT is necessary (Worthman & Stallings, 1997). (See Blood Spot Measures, below.)

Salivary cortisol levels are relatively insensitive to repeated thawing and refreezing; in recent studies, cortisol levels remained stable in samples undergoing up to three (Groschl et al., 2001) or four (Garde & Hansen, 2005) freeze/thaw cycles prior to assay. In the laboratory, samples collected by passive drool are frozen and thawed at least once before assaying in order to break down mucins that can interfere with pipetting (Vining & McGinley, 1986). Centrifuging helps remove particulate matter that can interfere with immunoassay. In salivettes, clear saliva collects in the bottom of the outer tube after centrifuging.

There is normally no need to transport samples to the laboratory on ice (Clements & Parker, 1998). However, when the time in transit is more than a few days, shipping on dry ice will prevent molding and may be required by some laboratories. (For information on international shipping, see International Air Transport Association IATA regulations; adjustments made in 2005 exempt saliva samples from regulations for hazardous biological substances.)

Prior to assay, saliva samples should be checked for blood contamination, as this can artificially elevate the cortisol concentration. Deficient diet, poor oral hygiene, and

overly strenuous toothbrushing can cause bleeding gums. In a recent study (Kivlighan et al., 2004), subjects first brushed their teeth [p. 46 ↓ ] vigorously and then collected saliva by direct drool. Minor injuries to the oral cavity led to detectable blood leakage in the samples, as assessed by three different methods: trans-ferrin immunoassay, dipsticks for detecting hemoglobin in urine, and visual inspection. A moderate degree of blood contamination (samples visibly pink) had a negligible effect on cortisol levels, but darker saliva samples were more problematic. Visual inspection and discarding of saliva samples that are discolored therefore appears to be adequate to control this source of error under normal circumstances. This is good news, because assay of transferrin—the most accurate method for assessing blood contamination—is relatively expensive, and dipsticks can yield false-positive results (Worthman & Stallings, 1997).

## Types of Assays

Free cortisol in the blood represents only 4% to 5% of total cortisol released; moreover, during passive diffusion into the salivary glands, approximately one-third of the free cortisol is lost through conversion to cortisone. Sensitive assay procedures are therefore necessary. Several methods currently allow reliable measurement of salivary cortisol without the necessity of extraction procedures. These include radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA), fluorescence immunoassay (FIA), and chemiluminescence immunoassay (LIA). The last three are nonradioactive assays in microtiter plate format, which can be run either manually or on automated equipment. Special laboratory equipment is required. Before the late 1990s, salivary cortisol assays were often adaptations of protocols designed for plasma/serum measures. Currently, assay kits developed for salivary determinations have standards suspended in a saliva matrix (in contrast to a serum or buffer matrix). Regardless of the assay used, the following procedures are recommended: (1) assay samples in duplicate and use the mean value in statistical analyses,<sup>1</sup> (2) repeat the assay for samples with duplicate values that differ by more than 20%, and (3) measure all samples from a given subject in the same assay run.

## Choosing a Lab

Over the last several years, commercial labs have proliferated, in some cases offering assay services for salivary cortisol as well as kits for use in the investigator's own lab. Details are available through the laboratories' websites, for example <http://www.salimetrics.com>, <http://www.ibl-hamburg.com>, and <http://www.dslabs.com>. In addition, many hospital and research labs have expertise in salivary assays; some use commercially available kits, while others have developed their own in-house assays. Price is an important consideration; the costs of a duplicate cortisol determination can range from roughly \$6 to \$30, often with a discount for large quantities. Quality should also enter into the choice, as not all assays are equally sensitive or reliable. Fortunately, assay quality and cost are likely to be inversely related, because laboratories with tailored salivary assays also tend to have more experience, higher volumes, and automated procedures.

Sensitivity refers to the minimum concentration of cortisol that can be distinguished from zero. Salivary cortisol assays generally have a lower detection limit of less than .01 µg/dL, which is below the concentration normally observed until late in the evening when the HPA axis becomes quiescent. The reliability of an assay, which is even more crucial for most research questions, is reflected in the intra- and interassay coefficients of variation (CV). The intra-assay CV can be calculated by dividing the standard deviation by the mean and then multiplying this figure by [p. 47 ↓ ] 100, over a representative subsample (high, medium, and low cortisol concentrations) of duplicate measures from the same assay run. The interassay CV is calculated over several assay runs; laboratories should be able to provide this information on request. Both CVs should be under 12% to 15% (most are lower). In general, CVs are higher for cortisol concentrations in the lower part of the range. Even when intra- and interassay coefficients of variation are acceptable, some laboratories may obtain higher or lower absolute cortisol values than others (Hansen, Garde, Christensen, Eller, & Netterstrom, 2003; Kraemer et al., 2006). For this reason, it is inadvisable to switch from one type of assay to another or from one laboratory to another during the same study.

Other indicators of assay performance are range of calibration, range of linearity of the assay (the linear region of the standard curve should cover the range in which

most salivary cortisol values are found), spike recovery, and specificity (the percentage of crossreactivity with other endogenous or exogenous substances, like cortisone or prednisone, that are or may be present in saliva). Laboratories should be able to provide this information for their assays. Reviewers are likely to request more detailed information for in-house assays. Recent implementation of voluntary quality assessment programs for salivary assays (IBL, 2005) will hopefully make it easier for researchers to evaluate and compare laboratories.

## Urinary Cortisol

### Background

Urinary measures of glucocorticoid metabolites (17-hydroxycorticoids) were among the first techniques available for studying activity of the HPA axis in humans, going back to the 1950s. An impressive body of knowledge emerged from the early psychoendocrine studies of 17-HOCs levels (Mason, 1968). New techniques soon allowed researchers to measure cortisol directly, as small amounts are excreted as free cortisol in the urine (UFC). Urine samples collected over 24 hours provide an integrated measure of total free cortisol excretion. Mean UFC values are approximately 20 $\mu$ g/24 h (range 3–43  $\mu$ g/24 h) in healthy adult women (Murphy, 2003).

To reduce participant burden, collection over shorter periods may prove adequate for a specific research question. In addition, urine collection can be scheduled in such a way that more refined analyses are possible. As an example, Jerjes and colleagues were able to investigate diurnal patterns of HPA activity by having subjects collect urine every 3 hours for 15 hours (Jerjes et al., 2006). Another recent study compared women's urinary cortisol levels when they were at home, at work, or asleep (Dettenborn et al., 2005).

A distinct advantage of urinary measures is that they allow assessment of nighttime cortisol levels, which may be crucial in certain disorders in which daytime levels are often normal (anxiety: Abelson & Curtis, 1996; PTSD: Yehuda, 2002). Urinary measures also have some disadvantages, which explain why they are less popular than salivary

cortisol. First, integrated measures are not very informative for research questions concerning acute stress responses. Second, the burden of collecting complete urine samples should not be underestimated, as it can lead to low participation in studies or poor compliance. Finally, transporting large volumes of urine from field to laboratory is cumbersome.

## Collection, Storage, and Handling

At the beginning of the sampling period, subjects void and discard the first urine. All urine produced thereafter is collected in large [p. 48 ↓ ] plastic containers designed for this purpose, or several containers if the study entails separate measurements. Samples can be kept at room temperature, without preservatives, for at least 24 hours without degradation of glucocorticoids (Gouarne, Foury, & Duclos, 2004).

## Assays

Urinary free cortisol represents a small fraction of total cortisol released by the adrenal cortex. Commercially available RIA kits for measuring UFC may yield falsely high values, as results can be influenced by the presence of cortisol metabolites as well as other interfering substances; UFC values obtained with these assays are potentially two to four times higher than the true values established with chromatography (Murphy, 2002). Immunoassays have been reported to show particularly low specificity and poor precision at low cortisol concentrations, leading to widely discrepant results in studies of adrenal suppression (Fink et al., 2002). In choosing a laboratory, it is therefore important to make sure that the assay has been validated and is monitored according to established standards for UFC; details concerning the assay (accuracy, recovery, precision, antibody used, crossreactivity, extraction method) should also be reported in publications. Accurate methods, for example liquid chromatography/tandem mass spectrometry (McCann, Gillingwater, & Keevil, 2005; Turpeinen & Stenman, 2003), are becoming more widely available and affordable. Because the tiny percentage (2–3%) of UFC in relation to total urinary cortisol metabolites may vary due to changes in steroid metabolism, measuring urinary cortisone, the ratio of cortisone to cortisol, or total

cortisol metabolites may provide additional insights into HPA axis function (Gouarne, Groussard, Gratas-Delamarche, Delamarche, & Duclos, 2005; Jerjes et al., 2006). UFC results are often corrected for creatinine levels.

## Blood Spot Measures

### Background

Finger-prick blood spot sampling provides an alternative to salivary measures of cortisol; this technique combines the advantages of traditional blood samples, in terms of the range of substances that can be measured, with greater ease of sample collection and more convenient storage and handling procedures (Wong, Yan, Donald, & McLean, 2004; Worthman & Stallings, 1997). Using devices designed to allow diabetics to monitor their own glucose levels, collection of finger-prick samples in capillary blood is quick and minimally invasive. Because of the tiny amount of blood required, obtaining repeated samples from an individual is feasible. Blood spot cortisol is highly correlated with serum levels. The method also has some disadvantages: not all participants can be trained to collect their own samples, so that research personnel may have to be involved; finger-pricks are not entirely painless, and recruitment of subjects may be more difficult for this reason; subjects may be concerned about the safety of the procedure; and so on.

### Collection, Handling, and Storage

Capillary blood from a finger-prick is dropped, without blotting or smearing, onto specially designed filter paper of the sort widely used in neonatal screening programs. One drop (yielding a blood spot of approximately 50  $\mu$ L of whole blood) is sufficient for cortisol determination. After samples on filter paper are air-dried for several hours, they can be easily stored in plastic bags for transport and even be mailed directly to the lab by ordinary post.

## Assays

Special, highly sensitive assays have been developed to determine cortisol levels in [p. 49 ↓ ] blood spots. These assays are currently performed by an increasing number of laboratories, including commercial laboratories such as Salimetrics and DSL. Additional hormones and other substances can be measured in blood spots, including those not measurable in saliva, such as prolactin and markers of immune function (McDade et al., 2000). Accurate measures of estradiol and progesterone can also be obtained, enabling the researcher reliably to assess the stage of the menstrual cycle (Shirtcliff, Reavis, Overman, & Granger, 2001), for example.

## A Framework for Designing a Study and Interpreting the Results

The first key to designing an effective study with clear results is awareness of the temporal dynamics of HPA axis activity, as these will dictate the sampling strategy. For studies of stress reactivity, the choice of a stressor can have a major impact on the results and their interpretation. Finally, the design should take into consideration the range of possible moderators, mediators, and confounders that might affect the hypothesized relationship between HPA measures and biopsychosocial variables of interest. This review attempts to summarize current recommendations and practice, without claiming that evidence in all cases is so consistent and complete that researchers have reached a consensus.

## Sampling Strategy

As previously described, integrated (in urine, UFC) and momentary (in saliva or blood) measures of cortisol are available. For UFC, the main decision is whether to collect samples over 24 hours or shorter time periods; the choice should be based on theoretical grounds, but subject burden and logistics often play a role. For salivary cortisol and blood spots, the optimal number and timing of samples depends on the

aspects of HPA activity being investigated (e.g., basal levels, diurnal variation, response to awakening, negative feedback inhibition, or response to acute stressors) and the stability of these measures over time.

## Basal Cortisol Levels and Diurnal Variation

Although investigators seem to agree that cortisol should be measured several times a day for a number of days to get reliable estimates of mean basal levels and diurnal slope (Goodyer et al., 2001; Stewart & Seeman, 2000), clear recommendations with supporting data are difficult to find. Hruschka and colleagues (2005) recently presented formulas for determining these figures on the basis of variance estimates from multilevel regression models. Based on data from a number of studies using different sampling protocols, these calculations suggested that—depending on the spacing of the samples in time—as few as four samples taken on one day might be adequate for estimating individual mean levels, but that 14 or more days of sampling with four to five samples per day might be necessary to obtain a reliable estimate of an individual's diurnal slope. In contrast, a study in an older population with very good protocol compliance found that five samples per day for 3 days provided a reliable estimate of daytime slope; moreover, slopes based on as few as two daily time points (wake and 9:00 p.m.) correlated highly with those based on four points, and additional days did not substantially increase reliability (Kraemer et al., 2006). These findings underscore the need for more analyses of existing “daily profile” datasets. To establish an optimal sampling protocol for a specific population, a pilot study with at least [p. 50 ↓ ] 50 participants is desirable (Kramer et al., 2006). Until more empirical results are available, a conservative recommendation would be to collect three to five samples a day for at least 3 days if basal levels are of primary interest and for 6 to 7 days if diurnal variation is a major focus. Increasing the number of subjects can increase statistical power when reliability of the cortisol measures is not optimal.

Because individual differences in sleep patterns may be associated with shifts in the circadian cycle, some researchers have chosen to collect samples at fixed intervals from the habitual time of awakening, instead of at fixed times of day. Another elegant but logistically simpler design is to sample at fixed times of day and then model effects of time since awakening statistically (Cohen et al., 2006). In all cases, efforts should be

made to obtain accurate information concerning the actual sample collection times. The most foolproof method is some form of electronic monitoring, for example, devices that record whenever a participant opens a vial to remove a cotton swab (Broderick, Arnold, Kudielka, & Kirschbaum, 2004; Jacobs et al., 2005; Kudielka, Broderick, & Kirschbaum, 2003) or handheld computers that generate time stamps with which participants must label their tubes (Stetler, Dickerson, & Miller, 2004). Awareness that compliance is being monitored increases the probability that samples will be taken as directed (Kudielka et al., 2003). Prompting participants with an audible or vibrating signal can also help. Finally, instructions to participants should emphasize the importance of accuracy and honesty in reporting actual collection times. In older adults, self-reported collection times were close to automatically recorded times, and test-retest reliability of slope estimates was actually slightly better when based on self-reported times (Kraemer et al., 2006).

## Cortisol Awakening Response (CAR)

The time course of this response has been well characterized. The peak response occurs 30 to 45 minutes after awakening; by 60 minutes, cortisol levels are decreasing and may no longer be reliably distinguishable from the levels at awakening. At least two samples (at awakening and either 30 or 45 minutes later) are needed to characterize the response; more samples (e.g., at 0, 30, 45, and 60 minutes) may increase reliability and allow calculation of AUC measures (see Statistical Analysis). The CAR should preferably be measured on at least 2 days. Given the narrow window of response, accurate timing of samples is crucial. Because participants appear to have difficulty in taking early morning samples as directed (Kudielka et al., 2003), it may be wise to reduce the sample burden to the minimum, emphasizing quality rather than quantity. Some kind of alarm device is useful to remind the participant to collect samples at the appropriate times, and compliance should be monitored electronically (see above) if this is possible. Activity monitors can be helpful in confirming time of awakening, but this is not considered essential for all studies.

Methodological issues relevant to study design have been summarized by Clow and colleagues (2004). Instructions to subjects should be standardized along the following lines:

- Place all materials next to your bed before going to sleep.
- Take the first sample in bed immediately after awakening, with lights on and eyes open.
- Do not go back to sleep; get out of bed (within 15 minutes) before taking another sample.
- The second sample should be taken [n] minutes after awakening (and so on for each sample).
- Do not brush your teeth, smoke, eat, or drink anything except water until you have finished taking the [n] morning samples.
- Remember to record the exact time each sample was taken on the tube, even if this was not the scheduled time.

## Dexamethasone Suppression as a Measure of Negative Feedback Sensitivity

Dexamethasone (DEX) can be safely ingested by participants at home and its effects measured in salivary cortisol (Lindley, Carlson, & Benoit, 2004; Powell et al., 2002). In most studies, cortisol measures on a control day are compared with measures taken at the same times of day following intake of 0.25 to 0.5 mg (low dose) or 1.0 mg (high dose) dexamethasone late the previous evening (at 11 p.m. or an agreed-on bedtime). The original dexamethasone suppression test (DST), developed as a diagnostic test for major depression, was scored as positive if, following administration of 1 mg DEX at 11 p.m. on day 1, cortisol at 4 p.m. on day 2 was above an established cutoff point (Carroll et al., 1981). For research purposes, analyzing the cortisol results as continuous instead of dichotomized measures yields more information. The optimal timing of the samples depends on the DEX dosage, as cortisol will “escape” from suppression earlier with lower doses. Collecting a number of post-DEX saliva samples at intervals of several hours will increase reliability of the results and gives added information about the time course of feedback inhibition.

# The Cortisol Response to Acute Stressors

Cortisol responses to acute stressors can be studied in the laboratory and in real life, where anticipated as well as unanticipated stressors occur. Design issues vary according to the setting. In the laboratory, important decisions include the best time of day to schedule the experiment, how many samples are needed to characterize the stress response, the timing of these samples in relation to the stress task, the nature of the task, how to control for effects of novelty and anticipation, and habituation to repeated stressors. For a detailed overview of many of these design issues, see Dickerson and Kemeny (2004).

*Time of Day.* Although the HPA axis is capable of responding to acute stress at any point in the diurnal cycle (Kudielka, Schommer, Hellhammer, & Kirschbaum, 2004), scheduling experiments in the mid to late afternoon (roughly between 3 and 6 p.m.) has advantages. First, the cortisol response to stress is more readily distinguishable in the afternoon than in the morning from background noise in the form of spontaneous pulsatile episodes and the natural decline in basal levels over the morning hours; the metaanalysis performed by Dickerson and Kemeny (2004) showed moderate effect sizes for cortisol response to stress tasks performed in the afternoon, compared to small effect sizes in the morning. Second, cortisol responses are easier to provoke in the afternoon than in the late evening, when the HPA axis becomes quiescent. Third, effects of potential confounders such as recent awakening and lunch (see below) are easier to exclude. Thus far, study design has been influenced mainly by such practical considerations, and little attention has been paid to theoretically important issues, such as the consequences of differential activation of MR and GR systems by stressors occurring at the trough versus the peak of the diurnal cycle (Dallman, Akana, Bhatnagar, Bell, & Strack, 2000).

*Number and Timing of Samples.* To characterize the cortisol response to an acute stressor, samples are taken at fixed intervals during baseline (30–40 minutes), stress exposure (10–20 minutes), and recovery (40–60 [p. 52 ↓ ] minutes) periods. Participants thus need to remain at the laboratory for a total of 1½ to 2 hours. The stress exposure includes both the preparation period, if there is one, and the actual task performance. Peak cortisol levels are usually observed 20 to 40 minutes after

task onset, depending on the intensity and duration of the task, with a gradual return to baseline levels over the next hour or longer (Dickerson & Kemeny, 2004). Even with an identical task, however, there are marked individual differences in latency to peak response (Gunnar & Talge, 2007). If time and budget allow, an optimal design would include two to three baseline measures, one to two measures shortly after and possibly during the stress task, and two to three measures during the recovery period. Minimalistic designs with one prestress and one poststress measure run the risk of missing the peak response and provide no information about speed of recovery.

*Controlling for Novelty and Anticipation.* Prestress baseline measures are highly sensitive to the novelty of the setting. Previous visits to the lab or an extended acclimation period (30 minutes or more) after arrival can reduce the probability of elevated baseline levels. Anxiety in anticipation of the task remains difficult to control. Because information provided earlier as part of informed consent and pretask instructions can either heighten or reduce anxiety, procedures need to be fully standardized in terms of both content and timing. Obtaining a saliva sample at home, at the same time on another day, is very useful in determining whether lab baseline levels are elevated (Nicolson, Storms, Ponds, & Sulon, 1997). This is important to know, because high baseline cortisol is often associated with a blunted response to stress (Kudielka, Schommer, et al., 2004; Young & Nolen-Hoeksema, 2001). Interestingly, infants and young children tend to show lower cortisol levels at lab arrival than at home (Gunnar & Talge, 2007).

*Type of Stressor.* Health psychology studies most frequently apply psychosocial stress tasks, as these are thought to have the greatest ecological validity. The HPA axis can be activated by physical stressors such as intense exercise or pain, or by pharmacological challenges; individuals' responses to different classes of stressors, however, do not appear to be highly intercorrelated. Among the psychosocial stressors, performance tasks with elements of social-evaluative threat, uncontrollability, or both produce the largest and most consistent increases in cortisol. Less consistent results are found for passive tasks (e.g., watching a film or other emotion induction procedures, noise exposure) and performance tasks without evaluative threat or uncontrollability (Dickerson & Kemeny, 2004).

Only a few stress tasks have been described in sufficient detail that results can be compared across studies and populations. The best known of these is the Trier Social Stress Test (TSST) (Kirschbaum, Pirke, & Hellhammer, 1993). The widespread use of the TSST reflects the fact that it has been extensively studied, can be applied in subjects varying in age and educational status, and induces a cortisol response in the majority of participants. For the TSST and other tasks with a combination of social-evaluative threat and uncontrollability, effect sizes, on average, are large (Dickerson & Kemeny, 2004). Results are sensitive to changes in the procedure, however: when the interval between instructions to subjects about the task and performance was extended from 10 minutes to 1 hour, the cortisol response to the task was obliterated (Young & Nolen-Hoeksema, 2001).

*Interpretation of Results.* Although laboratory stress tasks are probably the single most [p. 53 ↓ ] common approach to investigating the HPA axis in health psychology, results can be difficult to interpret. Failure to detect a statistically significant cortisol response is a rather common occurrence, even when the task appears to be experienced as stressful. Weaknesses in study design are responsible for some of these negative findings. With adequate sample size and the right choice of stress task and timing of measures, however, the majority of subjects are likely to show an increase in cortisol from baseline to posttask. Theoretically, one would expect the magnitude of this response to reflect the individual's experience of the situation, in terms of appraised threat, coping possibilities, and the intensity of emotional distress. Unfortunately, this is rarely the case with laboratory tasks, possibly because self-report instruments are not sensitive to the relevant processes or because other aspects of the situation—for example, its novelty—are more salient. Another discouraging finding is that cortisol responses often show no correlation with personality traits linked to stress reactivity, such as neuroticism (Schommer, Kudielka, Hellhammer, & Kirschbaum, 1999). Furthermore, little is known about the generalizability of lab reactivity measures to real-life situations (for studies involving cortisol, see Houtman & Bakker, 1987; Lundberg, Melin, Fredrikson, Tuomisto, & Frankenhaeuser, 1990; van Eck, Nicolson, Berkhof, & Sulon, 1996). Giving a speech on a particular topic or performing mental arithmetic before an audience may also have different meanings for individuals or groups, depending on variables such as cognitive ability, occupation, and cultural background. The challenge for researchers is to design laboratory stressors that convincingly tap

into the processes of interest in a given population and also reliably activate the HPA axis. Examples include a task involving competition with an antagonistic peer in children (van Goozen et al., 1998) and a standardized lecture in student teachers (Houtman & Bakker, 1987).

Test-retest reliability of acute stress response measures appears to be low, probably because of both the underestimated noise introduced by spontaneous pulsatile activity (Young, Abelson, & Lightman, 2004) and the tendency of the cortisol response to habituate following repeated exposures. Low reliability of cortisol outcome measures means that laboratory stress experiments are particularly vulnerable to Type 2 error. This issue is important in all studies, but especially needs to be taken into account in intervention studies, where stress reactivity is compared pre- and postintervention. More analyses are needed to determine how many cortisol measures per session and how many repeated sessions are necessary to obtain reliable measures of stress reactivity for different subject populations and stressors (Gunnar & Talge, 2007; Hruschka et al., 2005).

*Habituation and Sensitization.* If the cortisol response to experimental stressors is to be considered a valid indicator of what goes on in daily life, it is important to know what happens following repetitive stressful experiences. The acute response is adaptive, but is expected to habituate over repeated exposures as novelty decreases and control increases. Failure to habituate or sensitization to repeated stressors, in contrast, is regarded as maladaptive, contributing in the long run to allostatic load. Habituation versus sensitization of the HPA response has been extensively examined in animal models (e.g., Pitman, Ottenweller, & Natelson, 1990), but comparatively little research has been conducted in humans (exceptions include al'Absi et al., 1997; Epel et al., 2000; Gerra et al., 2001; Gunnar, Hertzgaard, Larson, & Rigatuso, 1991; Kirschbaum, Prüssner et al., 1995; Wüst, Federenko, van Rossum, Koper, & Hellhammer, 2005). Findings indicate [p. 54 ↓ ] rapid habituation from the first to the second and later exposures, but also show marked individual differences, with a subset of individuals failing to habituate after repeated exposures. In addition, correlations between the cortisol response and trait characteristics may increase (Kirschbaum, Prüssner et al., 1995) or decrease (al'Absi & Lovallo, 1993) over successive task performances. These findings suggest that repeated testing (for example, three times in a week, or once

a week for 3 weeks) is much more informative than single exposures in elucidating individual differences in stress reactivity that are relevant to long-term health outcomes.

*Naturalistic Experiments.* Responses of the HPA axis can also be investigated in response to real-life activities that entail some level of challenge or threat. Examples include exams, parachuting, sports competitions, musical performances, and occupational stressors, to name just a few. Tension-reducing activities such as yoga, meditation, and massage, on the other hand, may lower cortisol levels. Like lab experiments, these activities are usually scheduled or at least anticipated in advance, so that baseline, response, and recovery measures can be obtained. Certain activities, like parachute jumping, also lend themselves to studies of the habituation process (Deinzer, Kirschbaum, Gresele, & Hellhammer, 1997; Levine, 1978). Correlations between cortisol responses and subjective distress measures may be higher for real-life than for laboratory stressors (Nicolson, 1992).

HPA axis responses to life events can also be investigated, although prestress baseline measures are rarely available because of the unpredictable nature of individual life events (e.g., rape, sudden death of a family member) and natural or manmade catastrophes (e.g., earthquakes, hurricanes, war). In this case, cortisol levels in exposed individuals are compared with those in an unexposed comparison group and are often examined longitudinally, in relation to symptoms.

*Daily Hassles and Emotions.* Combining repeated self-reports with salivary measures, experience sampling or ecological momentary assessment studies have investigated cortisol reactivity to daily life hassles and uplifts and accompanying emotions. Real-life stressors vary widely in duration, and participants are often unable to report exactly when a stressful situation began or when it ended. Although the timing of cortisol measures in relation to daily hassles is therefore imprecise, multilevel regression analyses can assess associations between the two. The association between daily events and cortisol is probably mediated by changes in negative affect. The finding of higher salivary cortisol in association with daily hassles or negative affects has been replicated in several samples of adults (Hanson, Maas, Meijman, & Godaert, 2000; Peeters, Nicholson, & Berkhof, 2003; Smyth et al., 1998; van Eck, Berkhof et al., 1996) and children (Adam, 2006). In some but not all studies, positive affects were associated with lower cortisol (Adam, 2005; Polk, Cohen, Doyle, Skoner, & Kirschbaum, 2005).

## Moderators and Confounders

Interpretation of cortisol results can be facilitated by considering a number of between- and within-individual factors that can influence HPA axis activity. These include age, gender-related variables, and ethnicity; somatic variables such as illness, medications, and obesity; daily lifestyle variables such as food intake, smoking, and sleep patterns; psychosocial variables related to stress; and genetic differences.

## Age and Gender-Related Variables

Findings concerning the effects of age and gender on the HPA axis vary from study to **[p. 55 ↓ ]** study in magnitude and sometimes even in direction. It remains essential to consider and if necessary to control for the independent effects of these two variables and their possible interactions. In brief summary of the most consistent findings, cortisol levels increase with age, especially in the very old, and changes also occur, sometimes dependent on gender, in circadian amplitude and phase (Van Cauter et al., 1996). Age-related differences in acute stress reactivity have also been reported (Nicolson et al., 1997; Otte et al., 2005).

Gender appears to have a negligible effect on basal cortisol levels or diurnal slopes, but males and females often show different responses to experimental stressors. As summarized in recent reviews (Kajantie & Phillips, 2006; Kudielka & Kirschbaum, 2005), gender differences in cortisol reactivity have been attributed to the influence of female reproductive hormones as well as exogenous estrogens, but also to differences in cognitive, emotional, and behavioral responses to specific stressors (e.g., Kirschbaum, Klauer, Filipp, & Hellhammer, 1995; Stroud, Salovey, & Epel, 2002). Basal levels remain stable throughout the menstrual cycle, but both menstrual phase and oral contraceptive use may influence cortisol reactivity to stressors. Although findings are conflicting, current evidence indicates that salivary cortisol responses to psychosocial stress are blunted during the follicular as compared to the luteal phase of the menstrual cycle, whereas responses of women in the luteal phase are similar to those of men. Oral contraceptive users show responses similar to those observed in the follicular phase

(Kajantie & Phillips, 2006; Kudielka & Kirschbaum, 2005). Pubertal development in girls (Netherton, Goodyer, Tamplin, & Herbert, 2004), pregnancy (de Weerth & Buitelaar, 2005), and menopause (Kajantie & Phillips, 2006; Kudielka, Buske-Kirschbaum, Hellhammer, & Kirschbaum, 2004) have all been reported to moderate either basal cortisol levels or stress reactivity. Female hormonal and reproductive status should therefore be taken into account in study design and analysis.

## Race/Ethnicity

Studies examining ethnic differences in cortisol measures have produced mixed results (Bennett, Merritt, & Wolin, 2004; Cohen et al., 2006; Polk et al., 2005; Reynolds et al., 2006). Given race differences in other physiological measures and their relevance to disparities in health outcomes, additional research is needed.

## Somatic Variables

*Illness.* Participants who are acutely ill, with fever and malaise, should be excluded or rescheduled after full recovery. Chronic disorders such as Type 1 diabetes and other endocrine disorders, epilepsy, autoimmune disorders, and severe psychiatric disorders are often excluded because of their known or suspected direct effects on the HPA axis or effects of associated medications. Adrenal disorders should clearly be excluded, and it is standard practice to exclude other severe or unmanaged chronic disorders. For more prevalent and manageable disorders, exclusion criteria should be considered in light of study objectives and population. In a community sample of older men and women, for example, exclusion of all with hypertension, Type 2 diabetes, asthma, fibromyalgia, osteoarthritis, or a lifetime or family history of psychiatric disorder would leave few participants, and results of such a study would not be generalizable. Self-reports of these conditions are also unreliable, as many of these illnesses remain undiagnosed.

*Medication.* A similar situation applies to confounding effects of medications. These [p. 56 ↓ ] should not be underestimated, and a conservative approach (as in most clinical studies of the HPA axis) would require all subjects to be drug free. However, due to the

range of medications in widespread use and their small or as yet unknown effects of the HPA axis, it is not feasible in health psychology studies to exclude all subjects who take any medication. All medications should therefore be carefully recorded. Some classes of drugs, in particular systemic GCs like prednisone, prednisolone, and hydrocortisone, can have long-term effects on the HPA feedback system, and individuals who have used them in the past 6 months should be excluded. Anticonvulsants such as phenytoin and carbamazepine should also be excluded (Kunzel et al., 2003), as well as pure agonist opioids (Hibel et al., 2006). Use of low-dose GC inhalers, intranasal sprays, and topically applied creams can lead to mild suppression of the HPA axis in some individuals (Masharani et al., 2005), but this is unlikely to be a serious confounder (Hibel et al., 2006). Clinical dosages of zolpidem, a frequently used non-benzodiazepine hypnotic, does not alter cortisol rhythms (Copinschi et al., 1995). Use of antidepressants, low-dose benzodiazepines, non-steroidal anti-inflammatory drugs (NSAIDs), antihypertensives, and even over-the-counter drugs such as acetylsalicylic acid and acetaminophen (Hibel et al., 2006) should be evaluated in light of study goals and controlled for as necessary.

*Body Weight.* A comprehensive review of the literature on cortisol in human obesity (Björntorp & Rosmond, 2000) indicates that cortisol secretion rate is elevated in obesity, but cortisol is removed more rapidly from the circulation; the net effect is normal or lower-than-normal basal levels. However, studies that take the type of obesity into account have found that hypercortisolemia and dysregulation of the HPA axis are more common in central, abdominal obesity than in peripheral obesity. Men and women with central obesity often show elevated cortisol responses to laboratory stressors and to food intake, but there is marked heterogeneity, also in patterns of diurnal salivary cortisol secretion (Rosmond & Björntorp, 2001). Depending on the population and research questions, waist-to-hip ratio (WHR) may be a more informative measure than body mass index (BMI) (Epel et al., 2000; Ljung et al., 2000). BMI is a useful index of abnormally low body weight due to fasting or malnutrition, which has also been associated with HPA axis irregularities.

## Daily Activities and Lifestyle

*Sleep Patterns.* The circadian cycle is sensitive to disturbances and individual differences in the sleep-wake cycle. Studies of diurnal variation or the CAR should certainly control for effects of wake-up time, sleep duration, acute sleep loss (Leproult, Copinschi, Buxton, & Van Cauter, 1997), and disturbances in the sleep-wake pattern, including those due to jet lag and shift work. Related variables include individual differences in morningness-eveningness and seasonal changes in zeitgebers that affect HPA axis activity directly or through changes in sleep and activity patterns (Polk et al., 2005; Touitou et al., 1983). Seasonal effects may be more pronounced in certain disorders (Sher et al., 2005).

*Food Intake.* Controlled experiments have shown that food intake, particularly at lunch, increases cortisol secretion. In studies where salivary cortisol was repeatedly sampled over the day under naturalistic conditions, recent food intake at any time of day was associated with higher cortisol (Peeters et al., 2003; van Eck, Berkhof, et al., 1996). Studies have shown that the magnitude of the response depends on the macronutrient composition of the meal. Protein-rich meals lead to an [p. 57 ↓] increase of 50% to 100% in cortisol concentrations, with levels starting to rise approximately 30 minutes after meal onset, peaking around 60 minutes, and returning to baseline within 2 hours (Gibson et al., 1999; Slag, Ahmed, Gannon, & Nuttall, 1981). Glucose intake enhances cortisol response to acute stressors (Gonzalez-Bono, Rohleder, Hellhammer, Salvador, & Kirschbaum, 2002), which implies that food intake prior to undergoing an experimental stressor should be carefully controlled. Researchers might consider offering participants a standardized snack, with uniform caloric and carbohydrate content, an hour before baseline measures or else ask participants to refrain from eating. Similarly, subjects should not eat before completing all assessments of the CAR. For assessment of diurnal profiles with fixed schedules, researchers should choose sampling times long enough after usual meal times to minimize the effects of food intake (e.g., 11 a.m., 4 p.m., and 9 p.m. instead of 9 a.m., 2 p.m., and 7 p.m.). In studies with more frequent measures over the entire day, participants should be asked to record whether they have eaten in the past hour.

*Caffeine Intake.* Dietary doses of caffeine have been shown to increase cortisol secretion under experimental conditions (Lovallo, Al'Absi, Blick, Whitsett, & Wilson, 1996). In a recent study, acute cortisol responses to cumulative caffeine administration during a single day were reduced but not eliminated when subjects had consumed caffeine on the preceding 5 days, compared to a placebo condition (Lovallo et al., 2005). This suggests that even regular coffee drinkers may display some degree of HPA axis activation, especially in the afternoon. A single cup of coffee or tea, on the other hand, is unlikely to trigger an acute cortisol response (Quinlan, Lane, & Aspinall, 1997). Taken together, these results suggest that habitual caffeine consumption may be a relevant trait variable to assess, especially if one suspects that groups being compared might differ in percentage of coffee drinkers. Caffeine intake in the past hour is not likely to be a serious confounder in experimental studies.

*Smoking.* A number of studies in different age groups have reported higher cortisol levels in habitual smokers than in nonsmokers. For example, teenagers who smoked more than 10 cigarettes a day had higher basal cortisol levels than nonsmokers or light smokers; this effect was especially pronounced in girls (Canals, Colomina, Domingo, & Domenech, 1997). In college student smokers, serum cortisol levels were higher than in nonsmokers (Gilbert, Stunkard, Jensen, Detwiler, & Martinko, 1996). Habitual smoking was similarly associated with higher serum cortisol levels in postmenopausal women; this affect was not attributable to acute effects of smoking cigarettes during the test day (Baron, Comi, Cryns, Brinck Johnsen, & Mercer, 1995). In middle-aged men and women, salivary cortisol levels were higher throughout the day in smokers than in nonsmokers, and smokers' cortisol responses to awakening were also greater (Steptoe & Ussher, 2006). Smokers in the process of quitting show an acute reduction in cortisol levels in the early weeks, with levels gradually returning to somewhat under the preabstinence baseline after 4 to 6 weeks (Frederick et al., 1998; Steptoe & Ussher, 2006). Because smoking status can be an important confounder and may also partially mediate effects of other variables of interest (for examples, see Cohen et al., 2006; Olf et al., 2006), it should always be assessed, including number of cigarettes or other sources of nicotine per day and recent cessation or reduction.

Smoking may have greater trait than state influences on cortisol. One-day abstinence compared to ad libitum smoking had no effect on cortisol measures in habitual smokers (al'Absi, Amunrud, & Wittmers, 2002). **[p. 58 ↓ ]** On the other hand, several studies

have reported that recent smoking can cause transient cortisol elevations (Baron et al., 1995; Kirschbaum, Wust, & Strasburger, 1992) and attenuates the cortisol response to acute psychosocial stressors (Kirschbaum, Strasburger, & Langkrar, 1993; Rohleder & Kirschbaum, 2006; Tsuda, Steptoe, West, Fieldman, & Kirschbaum, 1996). Subjects should therefore be asked to refrain from smoking for at least an hour prior to cortisol measurements in the laboratory or scheduled sampling in real life. With random sampling, participants should record whether they have smoked in the last hour so that possible effects can be controlled for in the statistical analysis.

*Alcohol Intake.* Acute alcohol intake is associated with an increase in cortisol in light drinkers; this response is attenuated in heavy drinkers (King, Houle, de Wit, Holdstock, & Schuster, 2002). In a study of non-alcohol-dependent binge drinkers, moderate consumption of white wine (4 glasses over roughly 2 hours) reduced the cortisol response to food intake (Kokavec & Crowe, 2001). Moderate alcohol consumption does not appear to significantly influence basal cortisol levels, but alcohol dependence can alter both basal activity and reactivity of the HPA axis. Compared to alcohol-dependent individuals who were currently abstinent, those who were recently intoxicated displayed elevated cortisol, which increased further during withdrawal (Adinoff, Ruether, Krebaum, Iranmanesh, & Williams, 2003). Depending on study goals and population, these findings suggest that subjects should be screened for alcohol or other substance dependence. This is often an exclusion criterion.

*Physical Activity.* Both recent physical exertion and habitual athletic training can influence HPA activity. Experiments have shown that an hour of high-intensity exercise (at 70% VO<sub>2</sub> peak)

leads to pronounced increases in cortisol levels, whereas lower intensities or shorter durations had no significant effects (Jacks, Sowash, Anning, McGloughlin, & Andres, 2002). Moderate increases in physical activity are thus unlikely to have a measurable effect on cortisol. Postural changes (e.g., from supine to standing) do not affect cortisol measures (Hucklebridge, Mellins, Evans, & Clow, 2002).

## Psychosocial Variables

Past and current exposure to life events and chronic stressors are known to moderate basal cortisol levels and stress reactivity, as do individual traits like neuroticism and habitual coping styles. Depending on the research question, it may be useful to assess childhood adversity, recent life events, chronic stress (low socioeconomic status, difficulties with work, family, or other life domains), neuroticism, trait positive affect, coping styles, and current symptoms of depression, anxiety, or fatigue. Psychological state at the time samples are collected is also important to assess, including separate measures of positive and negative affect, and recent as well as anticipated stressors.

## Genetic Polymorphisms

There is growing evidence of individual differences in genetic vulnerability to stress, as well as gene-environment interactions (Caspi et al., 2003). Polymorphisms in genes involved in HPA axis regulation (DeRijk, Schaaf, & de Kloet, 2002; Wüst, DeRijk, et al., 2005) or in other stress-sensitive systems may predispose individuals to show different patterns of cortisol response to environmental stressors. DNA can be obtained noninvasively from buccal swabs, salivettes, or mouth rinses (Etter, Neidhart, Bertrand, Malafosse, & Bertrand, 2005).

[p. 59 ↓ ]

# Statistical Analysis of Cortisol Data

## Preparing the Data for Analysis

### Checking the Cortisol Distributions

Before statistical analysis, cortisol data should be screened to eliminate outliers, in particular those values exceeding the normal physiological range. The highest unstimulated salivary cortisol levels observed in healthy subjects are approximately 45 to 50 nmol/L, most likely to occur in the early morning. Cortisol outliers can also be defined statistically, for example, as values greater than 4 standard deviations above the mean. The researcher should also define criteria for excluding individuals who have a relatively high percentage of suspect measures, even if some samples are in the normal range.

Even after excluding physiologically abnormal data points, cortisol values display skewed distributions, especially in the early morning and evening hours. Data are usually logarithmically transformed prior to analysis, to avoid violating assumptions of common statistical procedures. Transformation, when necessary, should be done on the measure (e.g., area under the curve, or AUC) actually being entered into the analysis.

### Checking Compliance with Timing of Samples

If saliva collection is unsupervised, it may be necessary to exclude samples not taken close enough to the intended collection time. This is essential for accurate measurement of the cortisol response to awakening, as the peak response occurs within a narrow time window (30–45 minutes after awakening); participants who collect

saliva samples too late will thus falsely appear to have a blunted response. Later in the day, when cortisol levels are fairly stable over periods of a few hours, deviations from planned collection times are less likely to have a serious impact on the results; here, it is appropriate to exclude samples outside preestablished windows of 30 to 60 minutes on either side of the scheduled collection time. Some researchers define even larger windows of acceptability (Cohen et al., 2006). In multilevel regression approaches, the diurnal curve can be more accurately modeled using actual collection times (self-reported or electronically monitored) (Ranjit et al., 2005), and in this case, cortisol measures taken later than scheduled need not be excluded. Intensive, semi-random sampling schedules appear to enhance participant compliance and improve the reliability of cortisol results obtained over the day, even when participants believe that their compliance is not being monitored (Jacobs et al., 2005).

## Statistical Methods

# Total Cortisol Concentration and Diurnal Variation

Urinary cortisol provides an integrated measure of total output over the collection interval. Momentary assessments of cortisol in saliva or blood, however, need to be aggregated or statistically modeled in order to test effects of other variables on daily cortisol levels or slopes. Individual summary measures of basal levels or diurnal slope are also needed when cortisol is an independent variable in the analysis—for example, in longitudinal studies where cortisol measures are examined as predictors of health outcomes (Sephton et al., 2000). There are several options. For total levels, the simplest approach is to calculate the sum or average of two or more samples on a given day. With three or more daily samples, calculating an AUC for total levels takes into account that time intervals between samples may not be equal. Because the preceding measures will be heavily influenced by morning cortisol (when normative levels are high), [p. 60 ↓] other researchers have standardized cortisol values relative to the sample mean at each time of day before calculating a daily average cortisol measure

(DAC; Gunnar, Morison, Chisholm, & Schuder, 2001). An underlying assumption of this approach is that the biological “meaning” of a given cortisol level may vary according to the time of day, as studies in animal models suggest (Dallman et al., 2000). Advantages and disadvantages of aggregated measures have been discussed (Hruschka et al., 2005; Kraemer et al., 2006; Rosmalen et al., 2005). The most important drawback is that these measures ignore and obscure diurnal variation, which is fundamental to understanding the nature of HPA axis dysregulation. The current consensus that cortisol levels and slopes represent different constructs argues against combining them in the same measure (Kraemer et al., 2006).

The simplest measure of diurnal change entails calculating the difference between the morning and the evening values. When cortisol is the dependent variable, analysis of variance for repeated measures also allows differentiation between overall basal level and diurnal variation. A practical problem is that missing data at one of the sampling times results in exclusion of the entire day. When multiple days are sampled, it may be possible to reduce the percentage of missing data in the analysis by first aggregating data at each time of day, but this approach is far from elegant. In short, these models do not allow accurate modeling of diurnal variation in cortisol secretion.

In comparison to these traditional approaches, multilevel regression (hierarchical linear modeling) offers more accurate and statistically powerful approaches to analyzing cortisol data. The multilevel model is a variant of multiple linear regression, appropriate for data sets with a hierarchical structure (Raudenbush & Bryk, 2002). These methods offer many advantages for analyzing cortisol data, when the hormone measures are nested within days, and days are often nested within participants. First, they allow estimation of effects of independent variables on overall cortisol level, slope, and other characteristics of diurnal activity, as well as individual estimates of these parameters. Second, because discrete or continuous explanatory variables can enter the model at any level of the hierarchy, trait and state influences on cortisol can be teased apart. Models can thus be extended to test associations between cortisol and such time-varying covariates as emotional state or recent food intake, as well as individual characteristics like age or gender. Third, the multilevel model makes maximum use of the data, as it can deal flexibly with missing data and does not require fixed time intervals between successive measures. This is particularly useful in studies outside the laboratory, in which missing data are inevitable when participants are asleep, forget, or

are unable to comply with the sampling schedule. Moreover, the problem of subjects' failing to collect samples on schedule is reduced, as actual collection times can be accommodated in the analysis (Ranjit et al., 2005). Fourth, multilevel models explicitly take into account the dependencies among repeated cortisol measures taken within days and within individuals, thus allowing valid inferences. Finally, statistical power is increased compared to analyses of aggregated data.

Hruschka and colleagues (2005) and Ranjit and colleagues (2005) provide more detailed rationales for using multilevel methods in the analysis of cortisol data, as well as examples. Initially applied in experience sampling (ecological momentary assessment) studies with semi-random sampling intervals (Smyth et al., 1998; van Eck, Berkhof et al., 1996), multilevel modeling is also ideal for analysis of diurnal profile data collected at fixed time points. Despite its many advantages, multilevel modeling has the drawback of requiring more statistical expertise than some more familiar methods. Accurately [p. 61 ↓ ] modeling the diurnal cortisol curve, for example, is more complex than it might first appear. Moreover, although recent versions of standard statistical software packages (SAS, Stata, SPSS) offer more extensive procedures for mixed models, consultation with a statistician remains advisable for all but the simplest analyses.

## Cortisol Response to Awakening (CAR)

The CAR is usually operationalized as the absolute change in cortisol levels from awakening to either a fixed time point (e.g., 30 minutes) or the peak value of repeated measurements over the first hour. Alternatively, an AUC can be calculated, either as the response from waking baseline or as the total area relative to zero (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). Results should be presented in such a way that it is clear whether a larger CAR is due to relatively low cortisol at awakening. Although cortisol generally shows a pronounced increase following awakening, declining levels (negative CARs) are also observed and are not necessarily artifactual (Wüst et al., 2000). Given the current lack of consensus concerning the best way to characterize the CAR and the extent to which nonresponse is due to confounders, researchers are urged to provide as much information as possible (Clow et al., 2004). The CAR should be considered a discrete part of the circadian cycle, and samples

taken to assess the response should be excluded from analyses of basal levels and diurnal slopes.

## Cortisol Response to Acute Stress

For assessing cortisol responses to experimental stressors, traditional statistical approaches such as repeated measures ANOVA/ANCOVA are often used. The literature varies considerably in how constructs such as baseline, peak response, total response, and recovery are operationalized. An AUC can be computed, using a trapezoidal integration, as a measure of total response. The AUC is usually considered to be the area above the baseline level, but should also be allowed to take negative values (Grice & Jackson, 2004; Gunnar & Talge, 2007; Pruessner et al., 2003). Stress response measures should be corrected for baseline levels, as higher baseline is often associated with an attenuated response. A recent study used growth curve analysis (a form of multilevel modeling) to characterize changes in salivary cortisol levels during the TSST over three baseline samples, two stress response samples, and four recovery samples for each subject (Taylor et al., 2006). Estimates for the cortisol intercept and slopes of baseline, reactivity, and recovery measures were all significant and varied from person to person (random effects); to address the study's hypotheses, effects of between-subject predictors were tested in an extension of this model (in this case, oxytocin level at baseline was a significant predictor of cortisol intercept and baseline slope, whereas the presence of an audience was associated with steeper reactivity and recovery slopes). Given the general advantages of multilevel modeling mentioned earlier, plus the specific advantages of being able to model separate and theoretically important aspects of cortisol variability in the laboratory, this new statistical approach is likely to become standard practice.

## How to Report Findings

Laboratories often report cortisol concentrations in metric units (either  $\mu\text{g/dL}$  or  $\text{ng/dL}$ ), whereas research journals may prefer molar units of measurement. For salivary measures, for example, the conversion from metric ( $\text{ng/dL}$ ) to molar units ( $\text{nmol/L}$ )

simply entails dividing by 36.2. Similarly, urinary free cortisol can be expressed in micrograms/24 [p. 62 ↓ ] hours or nanomoles/24 hours (or per unit time for shorter periods); micrograms/24 hours divided by 3.62 transforms the value to nanomoles/24 hours.

The methods section of papers should briefly mention the following: exclusion criteria (for subjects or samples) intended to reduce confounding of HPA axis measures, frequency and timing of measures, whether compliance was electronically monitored, sample collection method, storage conditions, type of assay performed (including name and location of manufacturer if a commercial kit was used), and assay performance characteristics (inter- and intraassay coefficients of variation and lower detection limit; for in-house assays, reviewers will expect more details, e.g., spike recovery). For laboratory studies, a detailed description of the stress task is essential, including the time of day when experiments were performed.

The results section should include information about compliance with ambulatory collection procedures and times of awakening or meals, if these could have influenced the cortisol results. Presentation of statistical results should include information concerning effect sizes and observed power, so that the meaning of nonsignificant results can be judged. Providing intraclass correlations (ICCs) for mixed models enables readers to evaluate the strength of the results (what percentage of the variance in cortisol is explained by within- or between-person variables, after controlling for time of day effects) and to estimate sample sizes needed for future studies (see Hruschka et al., 2005).

## Conclusion

The availability of reliable and noninvasive methods for measuring cortisol, in combination with an appreciation of the pivotal role of stress in psychiatric and psychosomatic disorders, has spurred many researchers to consider incorporating cortisol measures into their investigations. The goals of this chapter were to describe basic features of the HPA axis and to alert the reader to possibilities and constraints at each stage in the process from study design, sample collection, assays, and statistical analysis through interpretation and presentation of results. Attention to these details can

help ensure that new studies will continue to extend our understanding of how the HPA axis contributes to human health and disease.

## Specialized Reviews on Related Topics

- Clow, A., Thorn, L., Evans, P., & Hucklebridge, F. (2004). The awakening cortisol response: Methodological issues and significance. *Stress, 7*(1), 29–37.
- de Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: From adaptation to disease. *National Review of Neurosciences, 6*, 463–475.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychology Bulletin, 130*(3), 355–391.
- Goodyer, I. M., Park, R. J., Netherton, C. M., & Herbert, J. (2001). Possible role of cortisol and dehydroepiandrosterone in human development and psychopathology. *British Journal of Psychiatry, 179*, 243–249.
- Gunnar, M. R., & Talge, N. M. (2007). Neuroendocrine measures in developmental research. In L. A. Schmidt & S. J. Segalowitz (Eds.), *Developmental psychophysiology* (pp. 343–366). Cambridge, UK: Cambridge University Press.
- 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.
- Kirschbaum, C, & Hellhammer, D. H. (2000). Salivary cortisol. In G. Fink (Ed.), *Encyclopedia of stress* (Vol. 3, pp. 379–383). New York: Academic Press.

10.4135/9781412976244.n3