Running head: HORMONE ASSAYS

Hormone assays

Oliver C. Schultheiss

Friedrich-Alexander University, Erlangen, Germany

Anja Schiepe

Technische Universität München, Munich, Germany

Maika Rawolle

Technische Universität München, Munich, Germany

Appeared as: Schultheiss, O. C., Schiepe, A., & Rawolle, M. (2012). Hormone assays. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.), *Handbook of Research Methods in Psychology* (Vol. 1: Foundations, planning, measures, and psychometrics, pp. 489-500). Washington DC: American Psychological Association.

Hormone assays

Hormones can be assayed from blood, urine, and saliva. Because measuring hormones in saliva samples is the easiest and least stressful method for research participants, it is often the method of choice for psychologists. In this chapter we will give an overview about the measurement of hormones in saliva samples. We will first describe hormones that can be assessed in saliva and their general effects. In the next section, we will discuss how hormones can be to assessed and how to deal with saliva samples. In closing, we will make suggestions about how to analyze data obtained with hormone assays and how to report hormone data in research journals.

1. A primer on concepts and measurement issues in behavioral endocrinology

Hormones are messenger molecules that are released by specialized neurons in the brain and by glands in the body into the blood stream or the interstitial fluid and that carry a signal to other parts of the brain or body. Which specific responses they trigger in target organs depends on the receptors involved and the function of the organ. Thus, one hormone can drive several different physiological and psychological functions through its effects on several target organs.

Generally, two broad classes of hormonal effects on physiology and behavior must be distinguished: *organizational effects* are lasting influences that hormones exert on the organism, thus changing its shape and functional properties various ways. Organizational hormone effects often occur during development or when significant hormonal changes take place, such as puberty. For instance, the development of the female and male body morphology is largely under hormonal control during fetal development, and deviations from typical gendered body morphology are frequently the result of deviations in hormone production, enzymatic conversion, or receptor action. In contrast to organizational effects, *activational effects* are those that

hormones exert temporarily, without producing lasting changes in the brain or the body. For instance, due to peaking estradiol levels around ovulation, women become more sensitive to sexual stimuli, as indicated by an enhanced pupillary response. This effect vanishes again after ovulation, when estradiol levels decrease (Laeng & Falkenberg, 2007).

The relationship between hormones and behavior is *bidirectional*. Hormones can have a facilitating effect on behavior, such as when high levels of testosterone increase aggressive responses in a game setting (Pope, Kouri, & Hudson, 2000). Such hormone → behavior effects can be most conclusively demonstrated through experimental manipulation of hormone levels, a method that uses hormone assays only to verify that circulating hormone levels were indeed affected by the experimental manipulation. Conversely, the situational outcome of a person's behavior as well as the stimuli and events impinging on the person can influence current hormone levels, such as when watching romantic movies leads to an increase in viewers' progesterone levels (Schultheiss, Wirth, & Stanton, 2004).

Because hormones have far-reaching and broad effects on physiology and behavior, their release is tightly controlled and monitored, primarily through negative feedback loops. For instance, circulating levels of the steroid hormone cortisol are monitored by the brain. If levels fall below a critical threshold, the hypothalamus releases corticotropine releasing hormone (CRH), which in turn triggers the release of the adrenocorticotrope hormone (ACTH). ACTH travels from the brain to the cortex of the adrenals, small glands that sit on top of the kidneys, where it stimulates the release of cortisol. If rising levels of cortisol exceed a certain threshold, CRH release, and thus the subsequent release of ACTH and cortisol, are suppressed until cortisol levels fall below the critical threshold again due to metabolic clearance. As a

consequence of this negative-feedback-loop mechanism, many hormones are released in repeated bursts occurring every 30 to 120 min. Hormones can also influence the release of other hormones. The quick (i.e., within minutes) testosterone increase in response to dominance challenges observed in male mammals (Mazur, 1985) is a good example. These rapid changes are the result of the stimulating effects of epinephrine and norepinephrine (NE), which are released within seconds after the onset of a situational challenge, on the testes (which produce testosterone in males). This effect is independent of the hypothalamic-pituitary-gonadal feedback mechanism normally involved in testosterone release (Sapolsky, 1987).

Like other physiological measures, such as heart rate or blood pressure, hormone levels are *multiply determined*, and in order to tease out the effects of interests (i.e., relationships between hormones and behavior), it is almost always necessary to control for, or hold constant, other influences on hormone levels. Chief among those influences are the strong circadian variations observed in many endocrine systems (hormones like testosterone, estradiol, and cortisol start out high in the morning and then decline through the course of the day), menstrual cycle changes in hormone levels (e.g., progesterone is low in the first half of the cycle and rises in the second), age and reproductive status (e.g., many hormonal systems operate differently in prepubertal children than in adults in their reproductive years), and the use of medications that alter hormone levels or endocrine responses (e.g., oral contraceptives). The variance generated by these factors can easily drown out whatever between-subjects differences one hopes to observe in an experiment if it is not taken into account through proper study design (e.g., test participants only in the afternoon) or measurement of potential confounds (e.g. recording of information

related to cycle stage, age, time of day, or use of medications; see Schultheiss & Stanton, 2009, for a screening questionnaire covering the most important variables).

The second issue of concern to a behavioral scientist who wants to use endocrine measures is how easy or difficult it is to assess a particular hormone. This in turn depends primarily on the biochemical properties of the hormone. *Peptide hormones* (i.e., short protein molecules, composed of a small number of amino acids), such as insulin, arginine-vasopressin (AVP), ACTH, NE, or oxytocin (OXY), are, by molecular standards, large structures and therefore do not easily pass through cell membranes. As a consequence, they can only be measured in the medium or body compartment into which they have been released or actively transported. Also, peptide hormone concentrations measured in the body may not accurately reflect peptide hormone levels in the brain, because they are released by different hypothalamic neuron populations. Moreover, peptide hormones break down easily and special precautions are necessary to stabilize their molecular structure after sampling. The other major class of hormones besides those that are peptides are *steroid hormones*, which are synthesized in the body from cholesterol. In contrast to peptide hormones, steroid hormones are highly stable, and in their free, bioactive form (i.e., not bound to larger proteins) can pass through cell membranes, leading to roughly similar levels of the free fraction of a hormone across body compartments. This means that, for instance, cortisol levels measured in saliva are similar to (free) cortisol levels measured in blood or cortisol levels in the brain. For this reason, and because saliva sampling is much easier and free of stress for research participants than the collection of blood samples or spinal fluid samples (to get at hormone levels within the CNS), salivary hormone assessment has become the method of choice among behavioral endocrinologists and psychologists working with human populations (Dabbs, 1992;

Hofman, 2001; Kirschbaum & Hellhammer, 1994). Table 1 provides an overview of hormones that can be assessed in saliva, their psychological correlates and effects, and references that discuss the validity of saliva assays for each hormone.

2. How can hormones be assessed?

Assays are procedures for determining the presence and amount of a substance in a biological sample. For the assessment of hormones, a variety of assays are available that utilize the capacity of antibodies produced by an organism's immune system to bind in a very precise way to a specific substance – hence the term immunoassay. For instance, antibodies can be raised in animals against human cortisol and these antibodies are then used in immunoassays for the quantification of cortisol. One of the oldest and still most precise immunoassays is the radioimmunoassay (RIA). "Radio" signifies that in RIAs, a fixed quantity of hormone molecules with radioactive labels (typically radioiodine; 125-I), also called tracer, is added to the assay and these molecules compete with molecules from samples collected from research participants for antibody binding sites. After a fixed incubation time, all excess tracer and sample are discarded and only the antibody-bound molecules (both with and without radioactive label) are retained. The more signal from the radioactive substance is detected in a test tube, the more tracer-labeled hormone and the less natural sample hormone is present in the tube. Conversely, the less signal is detected, the more natural hormone is present. Adding known amounts of unlabeled hormone to the RIA enables the researcher to construct a standard curve – essentially a regression formula that allows estimating the amount of hormone present in a given sample from the strength of the signal detected in the tube (see Schultheiss & Stanton, 2009, for more details on this procedure).

While RIAs are still widely considered the most valid and direct way of assessing hormones, they are increasingly being replaced by enzymatic immunoassays (EIAs) because of the hassles of licensing and properly running a radioisotope laboratory. EIAs operate according to the same principles as RIAs, except that the tracer signal is not based on radioactive decays but on enzymatic reactions leading, for instance, to differences in sample coloration or luminescence that can be quantified. Among the drawbacks of EIAs, in comparison to RIAs, are the complexity of the assay protocols and the relatively lower accuracy and sensitivity (cf. Raff, Homar & Burns, 2002).

A hormone assay, regardless of whether it is a RIA or an EIA, has to meet a number of criteria to be deemed valid and reliable. Assay validity is assessed through specificity, sensitivity, and accuracy; assay reliability is assessed through the precision of the measurements. Because it is important to understand these concepts, regardless of whether one reads assay quality information in a published paper, receives this type of information along with the data from an assay service, or is conducting one's own assays, we will briefly discuss each of them in the following (see O'Fegan, 2000, for a thorough discussion of hormone assay validation).

Specificity is defined as the ability of an assay to maximize measurement of the targeted analyte and minimize measurement of other analytes. Specificity is often established by measuring the degree to which an assay produces measurements different from zero for non-targeted analytes (e.g., in the case of a cortisol assay, measurements greater zero for progesterone, aldosterone, pregnenolone, and other related steroid hormones). Cross-reactivity with such non-target analytes is estimated by dividing the measured, apparent concentration of the analyte by the amount added

(e.g., 1000 ug/dL aldosterone added, 0.4 ug/dL measured: (0.4/1000)*100 = 0.04% cross-reactivity).

Sensitivity is defined as the lowest dose of an analyte that can be distinguished from a sample containing no analyte. It is often pragmatically derived by calculating the lower limit of detection (LLD), which is defined as signal obtained from a sample with zero analyte (B_0), minus three times the standard deviation of the signal at B_0. Values outside of the $B_0 – 3$ x SD range are considered valid non-zero measurements.

Accuracy is defined as the ability of the assay to measure the true concentrations of the analyte in the samples being tested. Accuracy is measured by including control samples with known amounts of analyte in the assay and then comparing the amount of analyte estimated by the assay (e.g., 46 pg/mL) with the actual amount added (e.g., 50 pg/mL testosterone). The result is expressed as the percentage of the actual amount that is recovered by the assay (e.g., accuracy = (46/50)*100 = 92%). Recovery coefficients between 90% and 110% reflect good accuracy.

Precision is defined as the degree of agreement between test results repeatedly and independently obtained under stable conditions. Precision is typically estimated by the coefficient of variation (CV%), which is calculated as the mean of replicate measurements of a given sample, divided by the standard deviation of the measurements, multiplied by 100. The *intra-assay CV%* is calculated as the average of the CV%s of all duplicate samples in a given assay or set of assays; the *inter-assay CV%* is calculated from the between-assay mean and SD of a control sample (e.g., a saliva pool) included in all assays. Intra- and inter-assays CV%s less than 10% are considered good.

Information about sensitivity, accuracy, and precision should be routinely obtained in hormone assays and reported in published research. Omission of these measurement credentials of an assay makes it difficult to judge the validity and reliability of the assay method used and may call into question the results obtained with it. Measures of specificity are not routinely included in hormone assays or research reports in behavioral endocrinology, but specificity should at least be carefully examined when a new assay is adopted.

3. Collecting, processing, and assaying saliva samples

High-quality hormone assessment starts with the careful collection and processing of the samples that will later be assayed for hormone concentrations. In keeping with our focus on salivary hormone assessment, we will restrict our review of the necessary steps to the collection, processing, and analysis of saliva samples.

The goal of the saliva collection phase is to collect high-quality samples (i.e., samples free of contaminants) in an exactly identified sequence and with a sufficient amount of saliva to allow the measurement of all targeted hormones later on. Because hormone release is influenced by distal factors (e.g., age, genes) and by more proximal factors such as circadian changes and pulsatile secretion patterns on the order of hours, timing of the sampling process is critical for making sense of the data later and capturing the effects of interest.

If salivary hormones are assessed with the goal of using the hormone measurements as individual difference variables (e.g., to address a question such as "Is salivary testosterone correlated with having a committed relationship to a partner?"; cf. Burnham et al., 2003), samples should be collected at the same time of day from all participants and participants need to indicate the number of hours since waking up as a key control variable which can be used as a covariate in later analyses.

If a researcher is interested in using hormone levels as a dependent variable in a study (e.g., to address a question such as "Does arousal of affiliation motivation lead to increases in salivary progesterone?"; cf. Brown et al., 2009; Schultheiss et al., 2004), at least two samples are needed to address the research question meaningfully: one baseline sample before the critical intervention or experimental procedure and one sample taken after the procedure.

The baseline sample should be collected immediately before the intervention takes place to account in later statistical analyses for as much hormone level variance as possible that is not attributable to the intervention itself. Placement of the post-intervention sample depends on the dynamics of salivary hormone changes and the properties of the intervention itself. Hormones released from their glands take seconds to minutes to spread out via the blood stream through the body and reach their target organs. It takes another 5 to 10 min for these hormones to then cross from the blood stream to the salivary glands (Riad-Fahmy et al., 1987). Thus, it can take as much as 15 to 20 min from the onset of a critical event that changes the amount of hormone released from a gland to the effect of these changes to show up in saliva as a robust signal that makes it across the statistical threshold. A meta-analysis on the effects of social-evaluative threat on salivary cortisol release found that the highest effect sizes can be detected about 20 to 30 min after the intervention (Dickerson & Kemeny, 2004). Similarly, research on the effects of winning or losing a dominance contest on testosterone changes in men found replicable evidence for a testosterone response maximum 15 to 30 min after the end of the contest (for summaries, see Schultheiss, 2007; Stanton & Schultheiss, 2009). In comparison, samples taken immediately after the end of an intervention or more than 45 min later are less likely to yield detectable

effects of an experimental manipulation (Dickerson & Kemeny, 2004; Schultheiss et al., 2005; Schultheiss & Rohde, 2002; Wirth & Schultheiss, 2006).

Thus, if only one post-intervention sample can be assayed in addition to a baseline sample, it should be placed about 20 to 30 min after the end of the experimental procedure (or, if the intervention has a long duration [> 10 min], after its presumably most impactful component). If more than one post-intervention sample can be taken, they should cover this "sweet spot" of situationally induced changes in steroid hormone levels, but, due to the time it takes for the salivary glands to fill up again and for hormones to pass into saliva, with intervals no less than 10 min between individual samples (e.g., taking samples at 15 - 20 and 30 - 35 min post-intervention). Note that collecting samples is not expensive: it usually costs only cents in terms of expenses for collection tubes and storage. Therefore, more than one post-intervention sample can be collected without much additional cost and the decision to assay only one or several of them can be deferred to a later date.

Over the years, several methods for collecting saliva have been introduced and evaluated. Some methods have aimed at stimulating saliva flow and speeding up the collection process (e.g., through the use of chewing gum or citric acid); others have attempted to combine this with a reduction of the embarrassment of letting spit drool out of one's mouth (e.g., through collecting saliva with the use of dental rolls that participants chew on). However, many of these methods alter the levels of the measured hormones and can add unsystematic error to the assessment of salivary hormones (see Dabbs, 1991; Shirtcliff, Granger, Schwartz, & Curran, 2001). As the overview of saliva collection methods provided in Table 2 indicates, only very few methods can be recommended across the board for the assessment of salivary hormones. The method that is least likely to produce interference through the

collection process is having participants spit directly into 50 mL centrifugation tubes, perhaps with the aid of a plastic straw through which saliva can flow into the tube. The only drawback of this method is that it can take some participants a long time to collect a sufficient quantity of saliva. Perhaps the only viable alternative to this is the use of sugarless chewing gum (Dabbs, 1991, found Trident Original Flavor sugarless gum to produce no interference) to stimulate saliva flow. However, this method has only been validated for use with salivary testosterone assessment and its validity for the assessment of other hormones remains to be tested.

The amount of saliva to be collected for each sample depends on the number and type of assays to be performed on them later. For instance, if only one hormone will be assessed, the net amount needed for an assay is < 1 mL; if more hormones will be measured, the amount collected needs to be increased accordingly. We recommend obtaining information about how much saliva will be needed for each hormone assessment, adding up the volumes and adding one 1 mL to account for sample attrition during saliva processing to calculate the target sample volume. Thus, if testosterone and cortisol will be assessed via RIA, 2 x 400 uL will be needed for each hormone, and a total volume of 2.6 mL (4 x 400 uL plus 1 mL) should be collected at each sampling point. In our experience, it is easy to collect as much as 5 mL within 5 min, which easily accommodates the assessment of 3 or 4 hormonal parameters per sample. To ensure that participants collect a sufficient amount of saliva, we routinely mark the collection tubes at the targeted volume and instruct participants to fill the tube to the mark (see Schultheiss & Stanton, 2009, for more details on sample collection instructions).

To avoid contaminants like blood or residues from a meal, participants should be instructed to refrain from eating and brushing their teeth for at least one hour

before coming to the lab. Research indicates that even though teeth-brushing can lead to blood leaking into saliva and thus to altered salivary hormone levels, this effect is transitory and no longer detectable 1 hour after brushing for all hormones tested so far (cortisol, DHEA, testosterone, progesterone, estradiol; see Kivlighan et al., 2004, 2005). Upon arrival, participants are asked to rinse their mouths with water. Collection of the first sample should start no earlier than 5 min after rinsing to avoid dilution of saliva samples with water.

More generally, because research participants may not always comply with instructions or fail to inform the experimenter beforehand of conditions and medications that may alter hormone levels and endocrine functions (see Granger, Hibel, Fortunato, & Kapelewski, 2009), we recommend including a screening questionnaire in behavioral endocrinology studies that covers the most important factors that can influence circulating hormone levels (including menstrual cycle stage and duration). Schultheiss and Stanton (2009) include in their appendix such a questionnaire. We recommend adding two questions to this questionnaire, namely "How many hours did you sleep last night?" and "At what time did you wake up this morning?", which will allow researchers to calculate the time that has elapsed between getting up and the start of the data collection session for each participant and to control for the effect of sleep duration on the functional capacity of the endocrine system. Finally, date and start time of each session should routinely be noted for each participant.

After a data collection session has ended, all samples should be sealed and frozen immediately. Salivary hormone levels can undergo significant changes if samples are not frozen for extended periods of time, presumably because of bacterial activity in the samples (Granger et al., 2004; Whembolua et al., 2006). If samples are

stored only for a couple of weeks before assaying, a regular -20 °C chest freezer will suffice to preserve salivary steroid concentrations. If, however, other analytes (alpha amylase, oxytocin) are targeted or samples have to be stored for extended periods of time (> 6 months), a -80 °C freezer is necessary to keep salivary hormone levels stable (Dabbs, 1991; Granger et al., 2004).

The goal of the second phase – saliva processing – is to make the saliva samples amenable to precise pipetting in the actual assay. To achieve this goal, all samples are first thawed and frozen three times after all data collection has been completed for the study. This procedure helps break down the long molecule chains (e.g., mucins) that make saliva sticky and viscous and turn it into a more watery, and thus precisely pipettable, fluid. The break-down of molecular chains can be enhanced by speeding up freezing and thawing through the use of dry ice and a warm water bath – the stronger shearing forces associated with the fast temperature differential induced by the use of these aids facilitates the degradation of the molecule chains. After the third thaw, samples are spun for 10 min at 1000 g in a refrigerated centrifuge to push all coarse content to the bottom of the tube (this process is similar to the separation of serum and plasma in blood samples). After centrifugation, the supernatant (i.e., the watery part of the sample that stays on top after centrifugation) of each sample is transferred to aliquot tube(s) (e.g., 5 mL, 2 mL, or 1.5 mL tubes). Care must be taken to avoid stirring up and transferring the coarse, sticky contents of saliva from the bottom of the tube during transfer. For this reason, we recommend centrifuging and aspirating only small batches of tubes (≤ 12) at a time, because coarse and watery components of saliva tend to mingle again after long waits between centrifugation and sample transfer to aliquots, particularly if samples are not refrigerated during and after

centrifugation. After aliquoting, samples can either be assayed right away or refrozen for later assaying.

The goal of the assaying the samples is to provide a specific, sensitive, accurate, and reliable measurement of their hormone content. There are several ways to get saliva samples assayed for hormone content. Ideally, one's own university already has an endocrinology lab whose services one can use or which at least provides one with an opportunity to run one's own assays. Such a lab can often be found in biochemistry or anthropology departments or in medical schools and university hospitals. For the novice, we recommend teaming up with another researcher who already has experience with the assessment of hormones and who can provide help and know-how with (salivary) hormone assays.

For a researcher who plans to include salivary hormone measures in her or his research on a regular basis, setting up a dedicated salivary hormone laboratory may be another option. The bare bones of such a lab include sufficient bench space (approx. 8 to 10 m) and storage space, a lab freezer (-20 °C: approx. $2,000; -80 °C: approx. $8,000), a fridge (approx. $500), a refrigerated centrifuge (approx. $7,000), a water bath (approx. $500), a vacuum pump for the aspiration of fluids (approx. $300) and an assortment of pipets covering the range from 10 uL to 1 mL (approx $1,000 for 3 pipets). Depending on whether the lab uses only enzymatic immunoassays or also works with radioimmunoassay, the purchase of a plate reader in the former case (approx. $5,000) and a gamma counter in the latter ($15,000 to $50,000, depending on whether it is used or new and how many samples can be counted simultaneously) will be required. Finally, another $1,000 should be dedicated to the purchase of lab glassware, tubing, stoppers, etc. Thus, the price tag for a salivary hormone lab can run anywhere between $17,000 and $70,000, depending on the type of equipment

purchased and not factoring in the cost of providing room, bench space, water supply, and air conditioning.

A third option is to have saliva sample analysis done by commercial assay labs that specialize in salivary hormone measurement. We strongly recommend that researchers not simply trust the claims these labs are making, but actually test their validity before and after sending off the samples. Of course, a thorough understanding of the quality parameters of good endocrine measurement as outlined in section 2 is essential for this. One simple way to pick a good assay service is to compare the claims of the assay provider with the published literature. Good assay services offer assays that cover the range of hormone concentrations typically observed in salivary hormones and also report validity data (i.e., specificity, sensitivity, accuracy, and precision) for this range. We also recommend including one's own accuracy checks calibrated for the hormone concentrations expected in the study sample (e.g., cortisol accuracy checks at 1.5 and 3.5 ng/mL, corresponding to low and high salivary levels of this hormone). The investment into a set of commercially available calibrator samples (e.g., Lyphochek from BioRad, Hercules, CA), a pipet, and a couple of tubes and pipet tips is comparatively small (less than $700) and pays off in the form of an independent verification of the quality of the outsourced assays. Finally, customers of commercial assay services should expect to receive a complete set of data that includes not only the mean hormone level and CV% for each sample, but also the values for each individual measurement (for verification of the intra-assay CV%), the values for standard pools used across assays (for verification of inter-assay CV%s) and the complete data on the standard curve, including the zero-concentration calibrator, which can be used to verify the service's claims about the sensitivity (LLD) of the assay.

Hormone assays come at a cost in terms of collection and processing materials, reagents, and work hours invested, and researchers need to factor in this cost when planning their studies and applying for funding. In our experience, the following rule of thumb works reasonably well when calculating the cost of a hormone assay, regardless of whether one conducts it in one's own or a colleague's lab or whether one sends off the samples to a commercial lab: For each sample and each hormone assessed in duplicate (i.e., when the same sample is assayed twice to get an estimate not only of the mean concentration, but also the measurement error in CV%), a cost of approximately $10 should be expected. Thus, if a researcher wants to collect 3 samples each from 60 participants and would like to have them assayed for cortisol and testosterone, she should plan a budget of $3,600 for the hormone assays (3 samples x 60 participants x 2 hormones x $10). Shipping and handling fees for mailing samples to an external lab need to be added to this estimate.

4. Data analysis and research report writing

Once the raw data from the assays have been collected, they need to be processed to arrive at estimates of the actual sample hormone concentrations. Schultheiss and Stanton (2009) provide a guided tour through the steps of data processing and an excellent in-depth treatment of the ins and outs of assay data processing can be found in Nix and Wild (2000). In the following, we will concentrate on data analytic strategies and presentation of the findings in a research report.

In general, the same rules and best practices for analyzing and reporting of other kinds of data also apply to hormone measures. Thus, hormone data distributions should be examined for skew and, if necessary, transformed to bring them closer to a normal distribution (this is frequently necessary for salivary cortisol data and may be required for other hormones in some cases, too); this should be reported. If outliers

are present in the hormone data (e.g., elevated estradiol due to ovulation, high progesterone levels sometimes observed in women in the luteal phase or in the early stages of pregnancy, or extreme levels of cortisol sometimes observed in individuals with undiagnosed endocrine disorders) and they cannot be accommodated through standard data transformations, analyses should be run and reported with and without the outliers. If the findings hold up to scrutiny either way, nothing is lost by pointing this out; if they emerge only in one or the other case, this needs to be considered in the discussion section and perhaps even before writing the paper.

Almost all hormone measures are influenced by one or several factors such as gender (e.g., men have 4 to 6 times higher levels of testosterone than women), time of day (e.g., steroid hormone levels are higher in the morning than in the evening), menstrual cycle stage (progesterone is higher in the second half than in the first half of the cycle), or use of medications (e.g., oral contraceptives lead to reduced salivary levels of estradiol, progesterone, and testosterone). The influence of such factors needs to be controlled for in data analyses, particularly when salivary hormone levels have been assessed only once, often as a marker of a personality or behavioral disposition (see Sellers, Mehl, & Josephs, 2007). In this case, an ANCOVA approach, in which the effect of the hormone of interest on the criterion measure is tested after such outside influences have been held constant, is suitable. Keep in mind, however, that the ANCOVA approach is only valid if the controlled-for variables exert only main effects on the criterion and do not significantly interact with the hormone in question in its effect on the criterion.

If more than one sample has been collected in pre-/post-intervention designs, the effects of extraneous factors such as time of day, gender, etc are already represented in the baseline hormone measure. If the baseline is being held constant in

a (repeated-measures) ANCOVA design while testing the effects of an experimental manipulation on subsequently assessed hormone levels, controlling for extraneous factors usually does not account for additional significant portions of variance in the post-intervention hormone measure(s), because their effect is already represented in the variance of the baseline hormone measure (see Schultheiss et al., 2005, for an example). Here, too, the researcher needs to make sure that the effects of the extraneous factors contained in the baseline measure is strictly additive and that no interaction effects with the experimental manipulation are present. This should not be presumed a priori but needs to be verified empirically. For instance, Stanton and Schultheiss (2007) tested the effects of implicit power motivation and winning or losing a dominance contest on changes in women's salivary estradiol levels. Because estradiol release is suppressed by oral contraceptives, simply controlling for baseline salivary estradiol (or, in addition, for oral contraceptive use) and then testing for effects of the predictors of interest on estradiol changes would have obscured the fact that the predicted estradiol changes could only be observed in women who did not take oral contraceptives: in this group of participants, salivary estradiol increased in power-motivated winners and decreased in power-motivated losers. This effect did not emerge in women who were taking oral contraceptives, and simply controlling for oral contraceptive use (instead of running analyses separately for women on and off oral contraceptives) would have amounted to throwing apples and oranges into one data-analytical basket.

Note that although we used the term "ANCOVA design" in the preceding paragraphs, we did this as a conceptual shorthand for regression-analytic designs in which hormone levels as predictor or baseline variable are entered as a quantitative predictor. Median-splits of hormone measures to make the design conform more to

the "design cells" approach of classical ANOVA is not recommended due to the loss of test power associated with it (Cohen & Cohen, 1983).

Similar to the reporting of other findings, reporting the results of hormone assays involves two steps. First, the method of assessment and its quality should be reported in the methods section. Second, the actual findings are reported in the results section. Description of the method should include the exact type and make of the assay, a short summary of the sample processing and sample assay protocol, and also state the main quality control parameters of the assay, that is, measures of validity (accuracy, lower limit of detection (LLD), analytical range) and reliability (intra- and interassay CV). Assay quality parameters provided by the manufacturers of commercially available assays should *not* be reported, as these typically represent best-case scenarios that are included with the assay to promote sales and that may have little to do with the quality of the actually conducted assays one's data are based on. Reporting of findings should include descriptive data on the hormone levels observed in the sample and their relationship to major influences on endocrine function, such as gender, menstrual cycle stage, use of oral contraceptives, and time of day when samples were collected. In behavioral endocrinology journals like *Hormones and Behavior* and *Psychoneuroendocrinology*, presentation of actual hormone data and their associations with the key variables of interest in a given study typically takes the form of line or bar graphs with error bars in the case of repeated-measures designs or scatter plots with fitted regression lines in the case of correlations. Researchers who want to publish their findings in psychology journals should follow these standards, as they conform to the guidelines of the APA task force on statistical inference (Wilkinson & Task Force on Statistical Inference, 1999) and allow the

reader to evaluate the suitability of the statistical procedures used and inferences made, given the shape of the data.

5. Conclusion

To summarize, hormones are messenger molecules that are released in the brain and by glands in the body to carry a signal to other parts of the brain or body and thereby exert broad effects on physiology and behavior. For this reason, exploring the role of hormones to identify endocrine indicators of specific psychological phenomena and processes (e.g., stress, motivational and emotional states) has become a fruitful field of inquiry for psychologists. Hormone measures bear the advantage that they simultaneously meet personality psychologists' need for rank-order stability and social psychologists' need for measures being sensitive to the social stimuli impinging on the person. Potential drawbacks of the use of hormone measures, however, might be that (1) a particular hormone does not map onto a particular psychological construct in a one-to-one fashion, and (2) that there are numerous complex interactions of hormonal systems with each other, the immune system, various brain systems, and peripheral organs. We believe, however, that a good understanding of the basic endocrine literature and careful, open-minded analysis of the findings one obtains from research employing hormone assays helps to master this complexity and can suggest exciting new lines of inquiry and discovery. This is work well invested, because linking psychology with endocrinology paves the way for more fruitful interdisciplinary work in the behavioral and brain sciences.

References

Brown, S. L., Fredrickson, B. L., Wirth, M. M., Poulin, M. J., Meier, E. A., Heaphy, E. D., et al. (2009). Social closeness increases salivary progesterone in humans. *Hormones and Behavior, 56*, 108-111.

Burnham, T. C., Chapman, J. F., Gray, P. B., McIntyre, M. H., Lipson, S. F., & Ellison, P. T. (2003). Men in committed, romantic relationships have lower testosterone. *Hormones and Behavior, 44*(2), 119-122.

Carter, C. S., Pournajafi-Nazarloo, H., Kramer, K. M., Ziegler, T. E., White-Traut, R., Bello, D., et al. (2007). Oxytocin: behavioral associations and potential as a salivary biomarker. *Annals of the New York Academy of Sciences, 1098*, 312-322.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2 ed.). Hillsdale NJ: Erlbaum.

Dabbs, J. M. (1991). Salivary testosterone measurements: Collecting, storing, and mailing saliva samples. *Physiology and Behavior, 49*, 815-817.

Dabbs, J. M. (1992). Testosterone measurements in social and clinical psychology. *Journal of Social and Clinical Psychology, 11*, 302-321.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355-391.

Granger, D. A., Hibel, L. C., Fortunato, C. K., & Kapelewski, C. H. (2009). Medication effects on salivary cortisol: tactics and strategy to minimize impact in behavioral and developmental science. *Psychoneuroendocrinology, 34*(10), 1437-1448.

Granger, D. A., Shirtcliff, E. A., Booth, A., Kivlighan, K. T., & Schwartz, E. B. (2004). The "trouble" with salivary testosterone. *Psychoneuroendocrinology, 29*(10), 1229-1240.

Hofman, L. F. (2001). Human saliva as a diagnostic specimen. *Journal of Nutrition, 131*(5), 1621S-1625S.

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: recent developments and applications. *Psychoneuroendocrinology, 19*(4), 313-333.

Kivlighan, K. T., Granger, D. A., & Schwartz, E. B. (2005). Blood contamination and the measurement of salivary progesterone and estradiol. *Hormones and Behavior, 47*(3), 367-370.

Kivlighan, K. T., Granger, D. A., Schwartz, E. B., Nelson, V., Curran, M., & Shirtcliff, E. A. (2004). Quantifying blood leakage into the oral mucosa and its effects on the measurement of cortisol, dehydroepiandrosterone, and testosterone in saliva. *Hormones and Behavior, 46*, 39-46.

Kudielka, B. M., Hellhammer, D. H., & Wüst, S. (2009). Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology, 34*(1), 2-18.

Laeng, B., & Falkenberg, L. (2007). Women's pupillary responses to sexually significant others during the hormonal cycle. *Hormones and Behavior, 52*(4), 520-530.

Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces, 64*, 377-402.

Nater, U. M., & Rohleder, N. (2009). Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: current state of research. *Psychoneuroendocrinology, 34*(4), 486-496.

Nix, B., & Wild, D. (2000). Data processing. In J. P. Gosling (Ed.), *Immunoassays: A practical approach* (pp. 239-261). Oxford: Oxford University Press.

O'Fegan, P. (2000). Validation. In J. P. Gosling (Ed.), *Immunoassays: A practical approach* (pp. 211-238). Oxford: Oxford University Press.

Pope, H. G., Jr., Kouri, E. M., & Hudson, J. I. (2000). Effects of supraphysiologic doses of testosterone on mood and aggression in normal men: a randomized controlled trial. *Archives of General Psychiatry, 57*(2), 133-140; discussion 155-136.

Raff, H., Homar, P. J., & Burns, E. A. (2002). Comparison of two methods for measuring salivary cortisol. *Clinical Chemistry, 48*(1), 207-208.

Riad-Fahmy, D., Read, G. F., Walker, R. F., Walker, S. M., & Griffiths, K. (1987). Determination of ovarian steroid hormone levels in saliva. An overview. *Journal of Reproductive Medicine, 32*(4), 254-272.

Sapolsky, R. M. (1987). Stress, social status, and reproductive physiology in free-living baboons. In D. Crews (Ed.), *Psychobiology and reproductive behavior: An evolutionary perspective* (pp. 291-322). Englewood Cliffs, NJ: Prentice-Hall.

Schultheiss, O. C. (2007). A biobehavioral model of implicit power motivation arousal, reward and frustration. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behavior* (pp. 176-196). New York: Guilford.

Schultheiss, O. C., & Rohde, W. (2002). Implicit power motivation predicts men's testosterone changes and implicit learning in a contest situation. *Hormones and Behavior, 41*, 195-202.

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in the neurobiology of social and personality psychology* (pp. 17-44). New York, NY: Guilford.

Schultheiss, O. C., Wirth, M. M., & Stanton, S. J. (2004). Effects of affiliation and power motivation arousal on salivary progesterone and testosterone. *Hormones and Behavior, 46*(5), 592-599.

Schultheiss, O. C., Wirth, M. M., Torges, C. M., Pang, J. S., Villacorta, M. A., & Welsh, K. M. (2005). Effects of implicit power motivation on men's and women's implicit learning and testosterone changes after social victory or defeat. *Journal of Personality and Social Psychology, 88*(1), 174–188.

Sellers, J. G., Mehl, M. R., & Josephs, R. A. (2007). Hormones and personality: Testosterone as a marker of individual differences. *Journal of Research in Personality, 41*, 126-138.

Shirtcliff, E. A., Granger, D. A., Schwartz, E., & Curran, M. J. (2001). Use of salivary biomarkers in biobehavioral research: cotton-based sample collection methods can interfere with salivary immunoassay results. *Psychoneuroendocrinology, 26*(2), 165-173.

Stanton, S. J., & Schultheiss, O. C. (2007). Basal and dynamic relationships between implicit power motivation and estradiol in women. *Hormones and Behavior, 52*(5), 571-580.

Stanton, S. J., & Schultheiss, O. C. (2009). The hormonal correlates of implicit power motivation. *Journal of Research in Personality, 43*, 942-949.

Whembolua, G. L., Granger, D. A., Singer, S., Kivlighan, K. T., & Marguin, J. A. (2006). Bacteria in the oral mucosa and its effects on the measurement of cortisol, dehydroepiandrosterone, and testosterone in saliva. *Hormones and Behavior, 49*(4), 478-483.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604.

Wirth, M. M., & Schultheiss, O. C. (2006). Effects of affiliation arousal (hope of closeness) and affiliation stress (fear of rejection) on progesterone and cortisol. *Hormones and Behavior, 50*, 786-795.

Wolkowitz, O. M., & Reus, V. I. (2003). Dehydroepiandrosterone in psychoneuroendocrinology. In O. M. Wolkowitz & A. J. Rothschild (Eds.), *Psychoneuroendocrinology: The scientific basis of clinical practice* (pp. 205-242). Washington, DC: American Psychiatric Publishing, Inc.

Table 1

Hormones and endocrine markers that can be assayed in saliva

Hormone	*Psychological functions*	*Reviews*
Cortisol	Indicates activation of the hypothalamus-pituitary-adrenal axis during stress; affects cognitive processes (memory, executive functions)	Kirschbaum & Hellhammer (1994); Kudielka, Hellhammer, & Wüst (2009)
Alpha-amylase	Marker of sympathetic nervous system activation during stress, release is stimulated by norepinephrine	Nater & Rohleder (2009)
Dehydroepi-androsterone (DHEA)	Associated with high psychological and physical well-being, memory function	Wolkowitz & Reus (2003)
Testosterone	Facilitates aggressive and non-aggressive forms of social dominance; enhances libido in men and women; supports male sexual behavior; influences cognition (e.g., mental rotation)	Dabbs (1992)
Estradiol	Enhances libido; involved in social dominance and sexual behavior; has profound effects on cognitive processes (e.g., verbal ability, memory); potentiates sensitization to psychostimulants	Riad-Fahmy, Read, Walker, Walker, & Griffiths (1987)
Progesterone	Decreases libido (particularly in men); anxiolytic; associated with affiliation	Riad-Fahmy, Read, Walker, Walker, &

	motivation	Griffiths (1987)
Oxytocin	Facilitates pair bonding and parental behavior; increases trust; counteracts the stress response	Carter et al (2007)

Table 2

Overview and evaluation of saliva collection methods

Collection method	*Recommendation*
Passive drooling into plastic collection tube (can be aided by plastic straw)	Recommended, produces no interference; potential drawback: some participants may take several minutes to collect a sample (Dabbs, 1991; Shirtcliff et al., 2001)
Sugarless chewing gum	Recommended, produces little interference and speeds up sample collection; potential drawback: slight transient rise (< 1 min after onset of chewing) in salivary testosterone levels, but can be avoided by asking participants to start collecting saliva only 2 min after onset of chewing (Dabbs, 1991; Granger et al, 2004); effects on other salivary hormones are unknown
Sugared chewing gum	Not recommended, produces elevated salivary testosterone levels (Dabbs, 1991); effects on others salivary hormones are unknown
Cotton rolls (Salivette)	Not recommended, produces increased measurements (relative to passive drooling into tube) for DHEA, estradiol, testosterone, and progesterone, though not for cortisol (Shirtcliff et al, 2001)
Polyester rolls	Not recommended, produces increased measurements (relative to passive drooling into tube) for testosterone (Granger et al., 2004); effects on others salivary hormones are unknown
Citric acid (crystals or	Not recommended, alters pH value of samples, which

powder)	may later interfere with pH-critical immunoassays, and produces increased measurements (relative to passive drooling into tube) for testosterone (Granger et al., 2004); effects on other salivary hormones are unknown