CHAPTER 25

SALIVARY HORMONE ASSAYS

Linda Becker, Nicolas Rohleder, and Oliver C. Schultheiss

Final accepted manuscript version, to appear as:

Becker, L., Rohleder, N., & Schultheiss, O. C. (in press). Salivary hormone assays. In H. Cooper, A. Panter, D. Rindskopf, K. Sher, M. Coutanche, & L. McMullen (Eds.), *APA Handbook of Research Methods in Psychology* (2 ed.). American Psychological Association.

Hormones can be assayed from blood, urine, spinal fluid, hair, and saliva samples, among other sources. Because measuring hormones in saliva is the easiest and least stressful method for participants, it is often the method of choice for psychologists. In this chapter, which represents a revised and updated version of an earlier chapter by Schultheiss et al., 2012[1], we give an overview about the assessment of hormones from saliva samples. We first describe which hormones can be assessed in saliva and typical research questions that are addressed in biobehavioral research. Furthermore, we provide an overview of quality criteria that should be considered when choosing an analysis method. Next, we discuss how hormones can be assessed and how to deal with saliva samples. Additionally, we provide a screening questionnaire with which the most important confounding variables can be assessed. In closing, we make suggestions about how to analyze data obtained with hormone assays, how to report hormone data in research journals, and how to implement open science practices in behavioral endocrinological research.

HORMONES IN BIOBEHAVIORAL RESEARCH

Many questions in biobehavioral research focus on the interrelationships between hormones, brain

[1] We thank Maika Rawolle and Anja Schiepe-Tiska for contributing to the first version of this chapter.

functioning, human experience, and behavior. Hormones are messenger molecules that are released by specialized neurons in the brain and by glands in the body. Free circulating hormones in the blood stream or in the interstitial fluid interact through positive and negative feedback loops with other parts of the body and with the brain, where they can affect neural processing. This can result in behavioral changes that are related with changes in emotional and cognitive processes. Therefore, the assessment of hormones is particularly interesting for researchers from different domains such as clinical, motivational, health, or cognitive psychology and neuroscience.

Generally, two broad classes of hormonal effects on physiology and behavior can be distinguished. *Organizational effects* are lasting effects that hormones exert on the organism, thus changing its shape and functional properties in various ways. Organizational hormone effects often occur during development or when significant hormonal changes take place, such as during 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. Although direct endocrine assessment (e.g., through amniocentesis or umbilical cord blood analysis for pre- and perinatal effects as well as repeated blood or saliva assessment during puberty) provides some limited insights in the role of organizational effects of hormones on human development, such effects are currently more typically assayed from bone-growth patterns that integrate hormonal action over longer stretches of time, albeit with less endocrine specificity (e.g., Köllner et al., 2019; Manning et al., 2014). In contrast to organizational effects, *activational effects* are those effects that hormones exert temporarily, without producing lasting changes in the brain or the body.

The relationship between hormones and behavior is *bidirectional*. Hormones can have a

facilitating or deteriorating effect on cognition and behavior, such as when high levels of testosterone increase aggressive responses in a game setting (Geniole et al., 2019), or high levels of cortisol decrease memory performance (Becker & Rohleder, 2019; Het et al., 2005). Such hormone-behavior effects can be most conclusively demonstrated through experimental manipulation of hormone levels (e.g., through stress tasks which induce the release of stress hormones such as cortisol; Becker et al., 2019; Kirschbaum et al., 1993). An alternative to an experimental manipulation with the intention to alter hormone levels is the direct application of hormones via oral drugs or nasal sprays and the investigation of their effects on psychological (e.g., behavioral, emotional, or cognitive) processes, such as decreased memory performance after hydrocortisone administration (e.g, Wolf, 2019). However, while oral hormone administration is a widely used and accepted method in experinmental studies (e.g., Geniole et al., 2019; Wolf, 2009), other methods such as the nasal application of oxytocin have been questioned and the presence of results supporting this method have been attributed to publication bias (Lane et al., 2016). 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 watching experimentally manipulated winning or losing leads to changes in individuals' gonadal steroid levels (Oxford et al., 2017; Vongas et al., 2020). These are all examples of studies, in which acute effects of hormones on behavior or vice versa were investigated in experimental laboratory settings. Another frequent research topic is the investigation of the hormone systems themselves and their circadian secretion patterns. Secretion of most hormones follows a circadian rhythm (e.g., for cortisol: an increase after wake-up and a decline throughout the day until bedtime). Several measures can be obtained from the diurnal course such as, for cortisol, the awakening response, the diurnal slope, or the area under the daytime curve as a measure for average cortisol exposure (Adam & Kumari,

2009). Altered diurnal secretion patterns have been found in psychopathology (e.g., flatter diurnal cortisol and testosterone slopes have been found in depression; Adam et al., 2017).

One issue of concern to biobehavioral researchers who want to use endocrine measures is how easy or difficult it is to assess a particular hormone. This depends primarily on the biochemical properties of the hormone. One class of hormones are *steroid hormones*, which are synthesized from cholesterol. The most important steroid hormones are glucocorticoids (e.g., the stress hormone cortisol), mineralocorticoids (e.g., aldosterone), and sex hormones (e.g., testosterone, estradiol, and progesterone). Steroid hormones are highly stable, and in their free bioactive form (i.e., not bound to larger proteins) they can pass through cell membranes, leading to roughly similar levels of the free fraction of a hormone across body fluid compartments. More importantly, they can also pass the blood-brain barrier and thus affect neural processing directly. Therefore, concentrations of steroid hormone levels in saliva are usually proportional to hormone levels in the blood. However, as most steroids are bound to transport proteins in blood, only the free fraction (about 10% for example of cortisol) can be measured in saliva. For this reason, and because saliva sampling is much easier and relatively stress-free for research participants than the collection of blood samples, salivary hormone assessment has become the method of choice among psychoneuroendocrinologists working with human populations (Schultheiss & Stanton, 2009; Strahler et al., 2017). Another class of hormones are *peptide hormones*, which are short protein molecules, composed of a small number of amino acids. Important peptide hormoens are insulin, arginine-vasopressin, adrenocorticotrope hormone (ACTH), norepinephrine (NE), and oxytocin. In contrast to steroid hormones, peptide hormones are not liposoluble and are large structures by molecular standards and therefore do not easily pass through cell membranes. As a consequence, peptide hormones can be measured only in the medium or body compartment into which they have

been released or actively transported. Importantly, in many cases peptide hormone concentrations measured in the body do not accurately reflect peptide hormone concentrations in the brain because they are released by different hypothalamic neuron populations. Furthermore, peptide hormones break down easily, and special precautions are necessary to stabilize their molecular structure after sampling. Some ways of measuring peptide hormone levels indirectly through the assessment of enzymes (that are related with hormones) from saliva have been suggested. One example is the assessment of salivary alpha-amylase (sAA) as an indicator for sympathetic nervous system (SNS) activity, which is associated with the fight-or-flight stress response. sAA and blood NE responses to stimulation (which is also released when the SNS is activated) are highly correlated (Nater & Rohleder, 2009). However, although sAA is an established and suitable marker for stress reactivity, you have to be methodologically well informed when measuring sAA and have to be aware that sAA mainly reflects SNS reactivity and that it is not necessaryly a suitable measure for absolute SNS activity (Bosch et al., 2011; Rohleder & Nater, 2009). Another class of substances that are not considered hormones, but that can also be obtained from saliva and which are also of interest for psychologists, are *markers of inflammation* (e.g., interleukines, C-reactive protein, tumor necrosis factor-alpha, immunoglobulin-A). However, the validity and reliability of using such salivary markers in psychological research still needs further investigation before it can actually be recommended (Engeland et al., 2019; Slavish et al., 2015). Table 25.1 provides an overview of relevant hormones and hormone-related substances that can be assessed in saliva, their psychological correlates and effects, and references that discuss the validity of saliva assays and their main area of application for each marker.

TABLE 25.1
Hormones and Other Markers That Can Be Assayed in Saliva

Marker	Psychological functions	References

Alpha-amylase	Marker of sympathetic nervous system activation; related with blood norepinephrine stress responses	Bosch et al., 2011; Nater & Rohleder, 2009; Rohleder & Nater, 2009
Cortisol	Indicates activation of the hypothalamus–pituitary–adrenal axis (e.g., during stress); associated with cognitive processes (memory, executive functions)	Hellhammer et al., 2009; Smyth et al., 2013
C-reactive protein	Inflammatory marker; associated with the risk for cardio-vascular disease	Christodoulides et al., 2005; Slavish et al., 2015
Dehydroepiandrosterone	Biomarker of aging and development; associated with mood and well-being and cognitive processes (e.g., memory and attention)	Kamin & Kertes, 2017; Wolkowitz & Reus, 2003
Estradiol	Enhances libido; involved in social dominance and sexual behavior; associated with cognitive processes (e.g., verbal ability, memory) and emotional processes	Riad-Fahmy et al., 1987; Toffoletto et al., 2014
Interleukines (e.g., IL-1, IL-6, IL-10)	Inflammatory markers; associated with the acute stress response and long-term health	Slavish et al., 2015
Oxytocin	Supports prosocial behavior and social cognition (e.g., social memory and social decision making); associated with pair bonding and parental behavior	Lane et al., 2016; Martins et al., 2020; McCullough et al., 2013
Progesterone	Decreases libido (particularly in men); anxiolytic; associated with affiliation motivation; associated with cognitive and emotional processes	Riad-Fahmy et al., 1987; Toffoletto et al., 2014
Testosterone	Associated with aggressive behavior and social dominance; enhances libido; supports sexual behavior; associated with cognitive performance (e.g., memory, executive functions, spatial performance)	Carré et al., 2017; van Anders et al., 2014

QUALITY CRITERIA

Hormone assays have to meet a number of quality criteria. Assays have to demonstrate validity,

which is assessed through accuracy, sensitivity, and specificity. They also have to demonstrate reliability, which is assessed through precision. We will briefly discuss each of these concepts in the following (for a thorough discussion of hormone assay validation, see O'Fegan, 2000).

Accuracy is the ability of the assay to measure the true analyte concentration in a sample. Accuracy is determined by including control samples with known analyte concentrations in the assay and comparing the amount of analyte estimated by the assay with the actual amount added. The result is expressed as the percentage of the actual amount that is recovered by the assay. Recovery coefficients between 90% and 110% reflect good accuracy.

Sensitivity is the lowest concentration of an analyte that can be distinguished from a sample containing no analyte. Sensitivity is often derived by calculating the lower limit of detection (*LLD*), which refers to a signal obtained from a sample with zero analyte (B0) minus three times the standard deviation (*SD*). Values outside this range are considered as measurements that can be validly differentiated from zero. Only values that are at least 3 *SD* units above B0 are interpreted as non-zero concentrations.

Specificity refers to the ability of an assay to maximize detection of the targeted analyte and minimize detection of other analytes. Specificity is often established by measuring the degree to which an assay produces measurements different from zero for nontargeted analytes (e.g., in the case of a cortisol assay, measurements greater than zero for related steroid hormones such as aldosterone). Cross-reactivity with nontargeted analytes is estimated by dividing the measured, apparent concentration of the target analyte by the amount of non-target added, multiplied by 100.

Precision is the degree of agreement between test results repeatedly and independently obtained under stable conditions. Precision is typically reported in terms of 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* refers to the average of the *CV*s of all duplicate samples (i.e., when the sample is assayed twice). 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 interassay *CV*s < 10% are considered as indicating good precision.

HOW CAN HORMONES BE ASSESSED?

Assays are procedures for determining the presence and amount of a substance in a biological sample such as hormone concentrations in saliva samples. For the assessment of hormones, a variety of assays are available. In many laboratories, *immunoassays* are used. The mechanism behind this procedure is the binding of antibodies to antigens. Antibodies are produced by an organism's immune system to bind in a precise way to specific substances, the antigens. For instance, an immunoassay for cortisol contains antibodies that specifically bind to cortisol. From the number of bound molecules, the concentration can be quantified. One of the oldest and still most precise immunoassays is the *radioimmunoassay* (RIA). In RIAs, a fixed quantity of hormone molecules with radioactive labels (also called *tracer*; e.g., radioiodine [125-I]) is added to the assay. The tracer molecules compete with the hormones contained in samples for antibody binding. 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 bound radioactivity is then measured. The more signal from the radioactive substance is detected, the more tracer-labeled hormone and the less natural sample hormone is present. Conversely, the less signal is detected, the more natural hormone is present. Quantification is carried out using a standard curve that is generated using samples with a known concentration of the analyte. From this, a regression

equation can be derived, i.e. a 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). Although RIAs are widely considered the most valid and direct way of assessing hormones by means of immunoassays, they are increasingly being replaced, because of the drawbacks of running a radioisotope laboratory. Common non-radioactive alternatives are *enzymatic immunoassays* (EIAs) *and enzyme-linked immunosorbent assays* (ELISAs; Lequin, 2005). Both EIAs and ELISAs operate according to the same principles as RIAs, except that the tracer signal is based on enzyme-labeled antigens and antibodies, leading, for instance, to differences in sample coloration or luminescence that can be quantified according to the same principles as described above for RIA. The drawbacks of EIAs/ELISAs, in comparison to RIAs, are the complexity of the assay protocols and the relatively lower accuracy and sensitivity (Schultheiss et al., 2019; Welker et al., 2016). An alternative method, which is often considered as the "gold standard" for salivary hormone assays, is (ultra-) high pressure *liquid chromatography* (LC), combined with *mass spectrometry* (MS). LC-MS systems include four main components, a chromatograph, an ionizer, a mass analyzer, and a detector. In the LC component of the system, the analytes are separated by their chemical and physical properties according to chromatographical properties (i.e., with respect to their stationary and a mobile phases). These separated analytes then enter the MS interface where they become ionized. The ionized molecules enter the mass analyzer, the part of the system in which they are analyzed with respect to their mass and charge. The molecules are accelerated in a vacuum through an electromagnetic field toward a detector. Only ions with a specific, predefined mass-to-charge ratio can pass the detector, enabling measurement with high specificity (see Schultheiss et al., 2019 for a detailed description of this method). The use of LC–MS systems has become popular in a growing number of

laboratories, because this technique demonstrates high levels of specificity and is argued to be free of many of the limitations of immunoassays such as cross-reactivity (Welker et al., 2016). In several studies, the assessment of different hormones from saliva samples with RIA, EIA/ELISA, or LC-MS has been compared. For example, for salivary testosterone, EIA/ELISA assays fail to capture the typical difference between male and female samples revealed by LC-MS, and correlations between LC-MS and EIA/ELISA measurements were disappointingly low, particularly in the female (low) range of concentrations (Prasad et al., 2019; Welker et al., 2016). RIA assays appear to to provide more valid measurements vis-à-vis LC-MS (Schultheiss et al., 2019).

DESIGNING THE STUDY AND SCREENING PARTICIPANTS

High-quality hormone assessment starts with the careful planning and study designing. Hormone levels are associated with many factors of which psychological variables are particularly important in biobehavioral research. To tease out the effects of interests (i.e., relationships between hormones and psychological variables), it is almost always necessary to control for or hold constant other influences on hormone levels. The most important factors are age and gender, which should always be recorded and controlled for in psychoneuroendocrinological research. Furthermore, as outlined above, for most hormones, strong circadian fluctuations can be found (Liening et al., 2010). Therefore, we recommend to control for time of day and if possible not to collect samples in the morning when the variations are strongest and hormone release is closest to its physiological maximum. Furthermore, it is important to control for menstrual-cycle changes in hormonal levels for female participants. For instance, progesterone levels are low in the first half of the cycle and rise in the second half (e.g., Gangestad et al., 2016; Hampson, 2020). Several

procedures to assess menstrual-cycle stage have been suggested. At least, the day of the beginning of the last cycle as well as the cycle length should be assessed. However, because of strong intra-personal variations (Hampson, 2020), the start of the next cycle after saliva collection should be verified by the participant and cycle stage should be determined by counting backward from the beginning of the subsequent cycle (Gangestad et al., 2016). Moreover, the use of hormonal contraceptives can alter hormone levels and, therefore, it is recommended to exclude women who take hormonal contraceptives if possible or otherwise to control for this factor. Further important factors that should be considered are the reproductive status (e.g., many hormonal systems operate differently in prepubertal children than in adults in their reproductive years), and the use of further medications (other than hormonal contraceptives) that alter hormone levels or endocrine responses (e.g., glucocorticoid-containing medication). We strongly recommend to include a screening questionnaire in behavioral endocrinology studies that covers the most important factors that can influence circulating hormone levels. We provide an example in Table 25.2. If children are included, stages of development (e.g., Tanner stage) should be additionally assessed. Finally, date and start time of each session should routinely be noted for each participant to control for diurnal variations in hormone levels and seasonality (e.g., androgens are highest in autumn; Stanton et al., 2011).

On the day of sample collection, participants should be instructed to refrain from eating at least two hours before saliva collection to avoid contaminants like blood or residues from meals ending up in the samples. For the same reasons, teeth should not be brushed at least 2 hours before sample collection. Furthermore, on arrival, participants should 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. Moreover, participants should refrain from excessive physical activity at least two

hours prior to the first saliva sample.

TABLE 25.2
Screening Questionnaire For Use In Studies With Hormone Assessments

Variable	Question
Age	Please enter your age (in years).
Gender	Please enter your gender (male, female, other (please specify)).
Weight	Please enter your weight (in customary unit (e.g., pounds)).
Height	Please enter your height (in customary unit (e.g., inches)).
Menstrual cycle (women only)	What was the date on which your last menstrual period started? What is the average duration of your menstrual cycle (in days)? How much does your menstrual-cycle length typically vary (0 to 1 day, 2 to 3 days, 4 days or more) (By "menstrual cycle" we mean the time from the start of one menstrual period to the start of the next.)
Menopause (women only)	Have you entered menopause? If so, at what age approximately?
Oral infections, oral lacerations, oral bleeding	Have you experienced any oral bleeding, oral lacerations, or oral infections over the past day?
Physical activity	Did you engage in any intense physical activity during the last two hours (e.g., sports, cycling, running). If yes, please specify.
Caffeine consumption	How many hours ago has it been since you consumed caffeine (contained in coffee, tea, soda, chocolate)?
Food consumption	Did you consume any food or beverages (except water) during the last two hours? If yes, please specify.
Smoking	Do you smoke? If yes, when did you have your last cigarette (in hours)? How many cigarettes do you smoke a day?
Alcohol consumption	How many hours ago has it been since you consumed an alcoholic beverage?
Drugs	Do you use any recreational drugs (e.g., marijuana, Ecstasy, speed, cocaine)? If yes, please specify (which type, how often, time of last consumption).
Anabolic steroids	Do you take anabolic steroids?
Hormonal contraceptives (women only)	Do you currently take hormonal contraceptives (i.e., the "pill" or a patch)? If yes, please specify (name). Do you take the pill continously? If applicable, in which intake phase are you currently (intake or pill break)?
Medication	Are you currently on any kind of medication(s) (prescription or non-prescription)? If yes, please provide the name of the prescription(s).
Endocrine disorders	Do you have a diagnosed endocrine disorder? If yes, please name the disorder.
Further disorders	Do you have any further diagnosed disorders (psychiatric or physical)? If yes, please name the disorder.
Relationship status	Are your currently involved in a steady relationship?

Handedness	Please indicate the hand (left or right) you typically use in activities such as writing, brushing your teeth, holding a glass, etc.
Sleep	What time did you wake up today? How many hours did you sleep last night?

The number of saliva samples that should be collected in a study strongly depends on the research question and the study design. If a researcher is interested in using hormone levels as a dependent variable (e.g., to assess the question "Does performing a neuropsychological test lead to an increase in cortisol levels"; Becker, Schade, & Rohleder, 2020), at least two samples are needed to address the research question meaningfully: one baseline sample and at least one sample that is taken after the procedure at the time point of the expected peak in the hormonal level. However, if the goal is to measure a precise time course, more samples may be required (e.g., to assess the question "What is the precise time course of sAA and cortisol levels during and after a resistance training?"; Becker, Semmlinger, & Rohleder, 2020). The baseline sample should be collected immediately before the intervention. Placement of the postintervention samples depends on the dynamics of salivary hormone changes and the properties of the intervention (e.g., the type of the stressor). For testosterone and cortisol, the peak of the highest concentration can be expected approximately 20 to 30 min after the end of the intervention (e.g., the stressor; Goodman et al., 2017). Markers that are associated with the activation of the SNS such as sAA typically peak immediately after the end of the intervention (e.g., the stressor). Samples taken more than 45 min after the end of the intervention are less likely to yield detectable effects of an experimental manipulation but are usually used to assess recovery of the hormonal response. In many studies, more than one post-intervention sample is collected. We recommend to collect more than one postintervention sample if the exact time course is not known in advance. However, it should be kept in mind that hormone assays are associated with costs that should be calculated in advance. In our experience, the following rule of thumb works reasonably well when calculating the cost of

a hormone assay: For each sample and each hormone assessed in duplicate costs of approximately $15 can be expected. Thus, if a researcher wants to collect three samples each from 100 participants and would like to have them assayed for cortisol and testosterone, he or she should plan a budget of $9,000 for the hormone assays (3 samples × 100 participants × 2 hormones × $15). These costs should be considered while planning and designing study and applying for funding.

COLLECTING, PROCESSING, AND ASSAYING SALIVA SAMPLES

A further key point for high-quality hormone assessment is the collection and processing of the samples that will later be assayed for hormone concentrations. The goal of the saliva-collection stage is to collect high-quality samples (i.e., samples free of contaminants). 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, citric acid, or Parafilm®); 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 cotton or polyester rolls that participants chew on). As the overview of saliva collection methods provided in Table 25.3 indicates, very few methods can be recommended 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 collection tubes (e.g., 50 mL centrifugation tubes), perhaps with the aid of a plastic straw through which saliva can flow into the tube (i.e., passive drooling). The only drawback of this method is that it can take some participants a long time to collect a sufficient quantity of saliva. Possibly the only viable alternative to the passive drool method is the use of Parafilm® as a stimulant, which has been proven suitable for the assessment of cortisol, testosterone, progesterone, and estradiol (Dlugash & Schultheiss, 2021). However, this method has been validated only for use of RIA, and its validity for the assessment with EIA/ELISA or LC-

MS remains to be tested. For a long time (and also in the previous version of this chapter) the use of sugarless chewing gum was recommended; new findings suggest that this recommendation is no longer tenable (Schultheiss, 2013; van Anders, 2010). Other stimulants (e.g., citric acid, cotton, or polyester rolls) are also not recommended, because they fail to provide consistently interference-free measurments for more than one target hormone.

However, the choice of collection device also depends on the hormones that are intended to be measured. In particular for cortisol and sAA, the use of polyester or cotton rolls does not pose any of the above problems (Rohleder & Nater, 2009). Especially, for sAA, polyester rolls lead to a uniform stimulation of saliva flow, and therefore the impact of flow rate on amylase concentrations is reduced (Rohleder et al., 2006). Therefore, the use of polyester rolls is recommended for sAA collection.

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 EIA/ELISA, lower volumes are needed than for RIA. For EIA/ELISA, a total volume of 1 mL is sufficient for the analysis of several analytes. For RIA, a sample volume of 1 mL is sufficient if one hormone will be measured only, and the amount collected needs to be increased if more hormones need to be assayed. We recommend obtaining information about how much saliva will be needed for each hormone assessment, adding up the volumes and adding 1 mL to account for sample attrition during saliva processing to calculate the target sample volume for testosterone assessment via RIA. For cortisol and sAA assessment via EIA/ELISA lower volumes are sufficient (e.g., 200 µL to assess both in duplicates). In our experience, it is easy to collect as much as 5 mL within 5 min, which easily accommodates the assessment of three or four hormonal parameters per sample in RIAs. To ensure that participants collect a sufficient amount of saliva, we recommend to mark the collection tubes at the targeted

volume and instruct participants to fill the tube to the mark (see Schultheiss & Stanton, 2009, for further details).

TABLE 25.3
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 (Schultheiss, 2013)
Parafilm®	Recommended, has been validated for the assessment of cortisol, testosterone, progesterone, and estradiol for RIA (Dlugash & Schultheiss, 2021)
Sugared or sugarless chewing gum	Not recommended, increases testosterone, estradiol (van Anders, 2010), and progesterone levels (Schultheiss, 2013)
Cotton or polyester rolls (salivettes)	Recommended if only one specific hormone is to be measured for that the device has been optimized (e.g., cortisol). Not recommended if other hormones than the one it was optimized for should be assessed or different hormones should be assayed from one sample
Citric acid (crystals or powder)	Not recommended, alters pH value of samples, which may later interfere with pH-critical immunoassays (Dlugash & Schultheiss, 2021)

After saliva sample-collection, all samples should be sealed and frozen immediately. We recommend storage at −80 °C if storage for several weeks or months is intended. For shorter times, storage in regular −20 °C chest freezers will suffice to preserve salivary hormone concentrations (Toone et al., 2013). When using polyester rolls for the assessment of sAA and cortisol, it is not necessary to proceed as strictly, because these molecules remain stable at room temperature for several weeks within this collection device.

The next step is preparation of the samples for the actual assay. Several procedures have been suggested. We recommend the following: First, all samples should be thawed and frozen three times. This procedure helps to break down the long molecule chains that make saliva sticky and viscous and turn the chains into a more watery, and thus more precisely pipettable, fluid. If necessary, this step can be speeded up by freezing and thawing through the use of dry ice and a

warm water bath. After the third freezeing-thawing cycle, the samples should be centrifuged for 10 min at 1,000 – 2,000 g in a refrigerated centrifuge to push all coarse content to the bottom of the tube (note that centrifugation may add bias (e.g., lower testosterone levels after centrifugation; Durdiaková et al., 2013) to the assessment of some hormones). 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) or deepwell plates (e.g., 5-mL, 2-mL, or 1.5-mL tubes or 96 x 1-mL deepwell plates). Care must be taken to avoid stirring up and transferring the detritus at the bottom of the tube during transfer. For this reason, we recommend centrifuging and aspirating only small batches of tubes (≤ 12) at a time or to use a pipetting robot that can pipette faster and with less error than humans. After aliquoting, samples can be assayed or refrozen for later assaying. Note that coarse and watery components of saliva tend to mingle again after long waits between centrifugation and sample transfer to aliquots. Therefore, repeating the centrifugation procedure may become necessary.

There are several ways to get saliva samples assayed for hormone concentration. If the researcher's university does not have an endocrinology lab whose services can be used or that at least provides researchers with an opportunity to run their assays, the set-up of a dedicated salivary hormone laboratory may be an option. The bare bones of such a lab include sufficient bench and storage space. Depending on the equipment and the assay method, the price tag for a sufficiently equipped laboratory starts at about $50,000. Furthermore, running costs for e.g. technical staff should be included in the cost calculation. A further option is to have saliva sample analysis conducted 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. A thorough understanding of the quality

parameters (i.e., specificity, sensitivity, accuracy, and precision, which we described in the section "Quality Criteria") of good endocrine measurement is essential for this testing. A simple way to select 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 quality data. 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 (see Schultheiss & Stanton, 2009, for further recommendations).

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 & Stanton, 2009, provide a guided tour through the steps of data processing, and Nix & Wild, 2000 provide an excellent in-depth treatment of the ins and outs of assay data processing. We next concentrate on data analytic strategies and presentation of the findings.

In general, the same rules and best practices for analyzing and reporting other kinds of data also apply to hormone measures. Most important is transparency; that is, all analysis steps (including reasons for and number of exclusions of participants) should be reported. Hormone data distributions should be examined for skewness and, if necessary, transformed to bring them closer to a normal distribution. For cortisol and sAA data, applying a natural-logarithm transformation is suitable in most cases. If outliers are present in the data and they cannot be accommodated through

standard data transformations, analyses should be run and reported with and without them.

As discussed above, almost all hormone measures are influenced by factors such as age, gender, and time of day (see Table 25.2). The influence of such factors needs to be controlled for in the data analyses, if it was not controlled in the study itself (e.g., by testing only female participants of the same age and menstrual-cycle stage). If these potentially confounding factors should be considered in the statistical analysis, regression analysis in general and analysis of covariance (ANCOVA) as a specific instantiation of this analytical approach can be used. With the use of an ANCOVA, the effect of the hormone of interest on the criterion measure is tested after such outside influencing factors have been held constant. Keep in mind, however, that the ANCOVA approach is valid only if the covariates exert only main effects on the criterion and do not significantly interact with the hormone in question (see Stanton & Schultheiss. 2009 for a discussion of how to deal with covariates for pre- and postintervention designs).

Reporting the results of hormone assays involves describing the method of assessment and its quality in the methods section and reporting the actual findings in the results section. The method description should include the exact type and make of the assay; a short summary of the sample processing and sample assay protocol, and a statement of the main quality control parameters of the assay (see Table 25.4 for two examples).

TABLE 25.4

Two examples for description of the methods in an article for two scenarios (adapted from Oxford et al. (2017) and Pretscher et al. (2021): 1) Assessment of multiple hormones by means of RIA, 2) cortisol assessment using EIA/ELISA

Step	Examples
Saliva collection	1) For each of the salivary samples, participants collected up to 7.5 mL unstimulated saliva in a sterile polypropylene vial. 2) Saliva was collected by means of polyester rolls ([product name], [company name]). Participants were instructed to keep the polyester roll in their mouth for at least one minute and to move it back and forth, but not to chew on it.
Storage	1) Participants sealed the vials immediately after each collection.

	The experimenter placed the vials in frozen storage (-20 °C) after each experimental session. 2) Saliva samples were stored in the participants fridge or freezer for approximately 1−2 weeks until they were given back to the experimenters and stored at -30 °C.
Sample processing in the laboratory	1) Samples were freed from mucopolysaccharides and other residuals by three freeze–thaw cycles, followed by centrifugation for 10 min at 3,000 rpm. All standards and controls were diluted with water (resulting analytic ranges: 5–400 pg/mL for testosterone; 0.51–19.50 pg/mL for estradiol; 5–400 pg/mL for progesterone; 0.5–25 ng/mL for cortisol). 2) For analysis, the polyester rolls were thawed at room temperature and were centrifuged at 2000 g and 20 °C for ten minutes immediately before analysis.
Laboratory analysis	1) Salivary hormones were determined by solid-phase 125-I radioimmunoassays ([assay name], [company name]), using assay protocols previously validated for use with saliva [references]. Lower limits of detection (B0—3xSD) were 3.02 pg/mL for testosterone, 0.60 pg/mL for estradiol, 1.04 pg/mL for progesterone, and 0.06 ng/ml for cortisol. Low and high control samples were included in each assay ([product name], [company name]; 20 and 100 pg/mL for testosterone; 1.23 and 3 pg/mL for estradiol; 27 and 101 ng/mL for progesterone; 1.5 and 3.5 ng/ml for cortisol). Average recovery values for low and high controls were 107% and 94% for testosterone; 128% and 131% for estradiol; 54% and 72% for progesterone; 89% and 88% for cortisol. Median intra-assay CVs were 10.94% for testosterone; 21.86% for estradiol; 8.91% for progesterone; 4.59% for cortisol. 2) Salivary cortisol concentrations were determined in duplicate using enzyme-linked immunosorbent assays (ELISA, ([assay name], [company name]). Intra- and inter-assay coefficients of variation were below 10%.
Statistical analysis	1) Visual histogram inspection and Shapiro-Wilks tests indicated that hormone values deviated from normal distributions. We therefore subjected progesterone, estradiol, and cortisol to log transformations and testosterone to a square-root transformation, after adding a constant of 1, and used these transformed scores in all inferential statistical procedures. Descriptive statistics are given for untransformed values. 2) The total diurnal cortisol output was calculated as area under the curve with respect to ground (AUCg). Because of positive skewness, AUCgs were transformed by means of the natural logarithm prior to further statistical analysis.

Reporting of findings should include descriptive data about the hormone levels observed in the sample and their relationship to major confounding variables. Researchers who want to publish

their findings in psychology and psychoendocrinology journals should follow the guidelines of the American Psychological Association Task Force on Statistical Inference (Wilkinson & APA Task Force on Statistical Inference, 1999) and allow readers to evaluate the suitability of the statistical procedures used and inferences made, given the shape of the data (e.g. by using 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).

In recent years, a debate about reproducibility in science has started and is changing the way we are doing research (Baker, 2016). This debate has also included findings and practices in behavioral endocrinology. Schultheiss & Mehta, 2019 give a comprehensive overview about this development in social neuroendocrinology and provide recommendations for conducting transparent and reproducible science. We agree with these recommendations and will emphasize the main points: 1) Preregister your study and hypotheses, 2) run power analyses, 3) make your raw data and analysis scripts available, and 4) use well-validated methods. Furthermore, we recommend running replication studies of your own and other researcher's work.

CONCLUSION

To summarize, the assessment of hormones from saliva samples is of broad interest in many fields in biobehavioral research. The areas of application are diverse and saliva sample collection is straightforward. However, the measured hormone concentration depends on the data collection procedure and the sensitivity and specificity of the analysis method. Furthermore, the study should be well planned and confounding factors should be identified and controlled. The rules of good scientific practice should be adhered to throughout the research process and open-science practices should be used. In this chapter, we have provided a comprehensive overview of the most important

steps, which is a good starting point. Nevertheless, we recommend newcomers to seek support from experienced researchers. We believe that linking psychology with endocrinology paves the way for fruitful interdisciplinary work in the behavioral, cognitive, and health sciences.

References

Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*, *34*(10), 1423–1436.

Adam, E. K., Quinn, M. E., Tavernier, R., McQuillan, M. T., Dahlke, K. A., & Gilbert, K. E. (2017). Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. *Psychoneuroendocrinology*, *83*, 25–41.

Baker, M. (2016). Reproducibility crisis. *Nature*, *533*(26), 353–366.

Becker, L., & Rohleder, N. (2019). Time course of the physiological stress response to an acute stressor and its associations with the primacy and recency effect of the serial position curve. *PLoS One*, *14*(5), e0213883.

Becker, L., Schade, U., & Rohleder, N. (2019). Evaluation of the socially evaluated cold-pressor group test (SECPT-G) in the general population. *PeerJ*, *7*, e7521.

Becker, L., Schade, U., & Rohleder, N. (2020). Activation of the hypothalamic-pituitary adrenal axis in response to a verbal fluency task and associations with task performance. *PLoS One*, *15*(4), e0227721.

Becker, L., Semmlinger, L., & Rohleder, N. (2020). Resistance training as an acute stressor in healthy young men: associations with heart rate variability, alpha-amylase, and cortisol levels. *Stress*, 1–13.

Bosch, J. A., Veerman, E. C. I., Geus, E. J. de, & Proctor, G. B. (2011). α-Amylase as a reliable

and convenient measure of sympathetic activity: don't start salivating just yet! *Psychoneuroendocrinology*, *36*(4), 449–453.

Carré, J. M., Ruddick, E. L., Moreau, B. J. P., & Bird, B. M. (2017). Testosterone and human aggression. *The Wiley Handbook of Violence and Aggression*, 1–14.

Christodoulides, N., Mohanty, S., Miller, C. S., Langub, M. C., Floriano, P. N., Dharshan, P., Ali, M. F., Bernard, B., Romanovicz, D., & Anslyn, E. (2005). Application of microchip assay system for the measurement of C-reactive protein in human saliva. *Lab on a Chip*, *5*(3), 261–269.

Dlugash, G., & Schultheiss, O. C. (2021). Suitability of saliva stimulants for valid assessment of steroid hormones via radioimmunoassay. *Psychoneuroendocrinology*, *127*, 105175.

Durdiaková, J., Fábryová, H., Koborová, I., Ostatníková, D., & Celec, P. (2013). The effects of saliva collection, handling and storage on salivary testosterone measurement. *Steroids*, *78*(14), 1325–1331.

Engeland, C. G., Bosch, J. A., & Rohleder, N. (2019). Salivary biomarkers in psychoneuroimmunology. *Current Opinion in Behavioral Sciences*, *28*, 58–65.

Gangestad, S. W., Haselton, M. G., Welling, L. L. M., Gildersleeve, K., Pillsworth, E. G., Burriss, R. P., Larson, C. M., & Puts, D. A. (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evolution and Human Behavior*, *37*(2), 85–96.

Geniole, S. N., Procyshyn, T. L., Marley, N., Ortiz, T. L., Bird, B. M., Marcellus, A. L., Welker, K. M., Bonin, P. L., Goldfarb, B., & Watson, N. V. (2019). Using a psychopharmacogenetic approach to identify the pathways through which—and the people for whom—testosterone promotes aggression. *Psychological Science*, *30*(4), 481–494.

Goodman, W. K., Janson, J., & Wolf, J. M. (2017). Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology, 80*, 26–35.

Hampson, E. (2020). A brief guide to the menstrual cycle and oral contraceptive use for researchers in behavioral endocrinology. *Hormones and Behavior, 119*, 104655.

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, *34*(2), 163–171.

Het, S., Ramlow, G., & Wolf, O. T. (2005). A meta-analytic review of the effects of acute cortisol administration on human memory. *Psychoneuroendocrinology*, *30*(8), 771–784.

Kamin, H. S., & Kertes, D. A. (2017). Cortisol and DHEA in development and psychopathology. *Hormones and Behavior*, *89*, 69–85.

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1-2), 76–81.

Köllner, M. G., Janson, K. T., Bleck, K. (2019). The social biopsychology of implicit motive development. *Routledge International Handbook of Social Neuroendocrinology*, 568–585.

Lane, A., Luminet, O., Nave, G., & Mikolajczak, M. (2016). Is there a publication bias in behavioural intranasal oxytocin research on humans? Opening the file drawer of one laboratory. *Journal of Neuroendocrinology*, *28*(4).

Lequin, R. M. (2005). Enzyme immunoassay (EIA)/enzyme-linked immunosorbent assay (ELISA). *Clinical Chemistry*, *51*(12), 2415–2418.

Liening, S. H., Stanton, S. J., Saini, E. K., & Schultheiss, O. C. (2010). Salivary testosterone, cortisol, and progesterone: two-week stability, interhormone correlations, and effects of time

of day, menstrual cycle, and oral contraceptive use on steroid hormone levels. *Physiology & Behavior*, *99*(1), 8–16.

Manning, J., Kilduff, L., Cook, C., Crewther, B., & Fink, B. (2014). Digit ratio (2D: 4D): a biomarker for prenatal sex steroids and adult sex steroids in challenge situations. *Frontiers in Endocrinology*, *5*, 9.

Martins, D., Gabay, A. S., Mehta, M., & Paloyelis, Y. (2020). Salivary and plasmatic oxytocin are not reliable trait markers of the physiology of the oxytocin system in humans. *Elife*, *9*, e62456.

McCullough, M. E., Churchland, P. S., & Mendez, A. J. (2013). Problems with measuring peripheral oxytocin: can the data on oxytocin and human behavior be trusted? *Neuroscience & Biobehavioral Reviews*, *37*(8), 1485–1492.

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.), *The Practical Approach series. Immunoassays: A practical approach* (pp. 239–261). Oxford University Press.

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

Oxford, J. K., Tiedtke, J. M., Ossmann, A., Özbe, D., & Schultheiss, O. C. (2017). Endocrine and aggressive responses to competition are moderated by contest outcome, gender, individual versus team competition, and implicit motives. *PLoS One*, *12*(7), e0181610.

Prasad, S., Lassetter, B., Welker, K. M., & Mehta, P. H. (2019). Unstable correspondence between salivary testosterone measured with enzyme immunoassays and tandem mass

spectrometry. *Psychoneuroendocrinology*, *109*, 104373.

Pretscher, A., Kauzner, S., Rohleder, N., & Becker, L. (2021). Associations between social burden, perceived stress, and diurnal cortisol profiles in older adults: implications for cognitive aging. *European Journal of Ageing*, 1-16.

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. *J Reprod Med*, *32*(4), 254–272.

Rohleder, N., & Nater, U. M. (2009). Determinants of salivary α-amylase in humans and methodological considerations. *Psychoneuroendocrinology*, *34*(4), 469–485.

Rohleder, N., Wolf, J. M., Maldonado, E. F., & Kirschbaum, C. (2006). The psychosocial stress-induced increase in salivary alpha-amylase is independent of saliva flow rate. *Psychophysiology*, *43*(6), 645–652.

Schultheiss, O. C., Dlugash, G., & Mehta, P. H. (2019). Hormone measurement in social neuroendocrinology: A comparison of immunoassay and mass spectrometry methods. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge international handbook of social neuroendocrinology* (pp. 26–40). Routledge.

Schultheiss, O. C., & Mehta, P. H. (2019). Reproducibility in social neuroendocrinology: Past, present, and future. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge international handbook of social neuroendocrinology* (pp. 41–64). Routledge.

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in social neuroscience* (pp. 17–44). Guilford.

Schultheiss, O. C. (2013). Effects of sugarless chewing gum as a stimulant on progesterone, cortisol, and testosterone concentrations assessed in saliva. *International Journal of*

Psychophysiology, *87*(1), 111–114.

Schultheiss, O. C., Schiepe-Tiska, 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.

Slavish, D. C., Graham-Engeland, J. E., Smyth, J. M., & Engeland, C. G. (2015). Salivary markers of inflammation in response to acute stress. *Brain, Behavior, and Immunity*, *44*, 253–269.

Smyth, N., Hucklebridge, F., Thorn, L., Evans, P., & Clow, A. (2013). Salivary cortisol as a biomarker in social science research. *Social and Personality Psychology Compass*, *7*(9), 605–625.

Stanton, S. J., O'Dhaniel, A., & Huettel, S. A. (2011). Seasonal variation of salivary testosterone in men, normally cycling women, and women using hormonal contraceptives. *Physiology & Behavior*, *104*(5), 804–808.

Strahler, J., Skoluda, N., Kappert, M. B., & Nater, U. M. (2017). Simultaneous measurement of salivary cortisol and alpha-amylase: application and recommendations. *Neuroscience & Biobehavioral Reviews*, *83*, 657–677.

Toffoletto, S., Lanzenberger, R., Gingnell, M., Sundström-Poromaa, I., & Comasco, E. (2014). Emotional and cognitive functional imaging of estrogen and progesterone effects in the female human brain: a systematic review. *Psychoneuroendocrinology*, *50*, 28–52.

Toone, R. J., Peacock, O. J., Smith, A. A., Thompson, D., Drawer, S., Cook, C., & Stokes, K. A. (2013). Measurement of steroid hormones in saliva: Effects of sample storage condition. *Scandinavian Journal of Clinical and Laboratory Investigation*, *73*(8), 615–621.

van Anders, S. M. (2010). Chewing gum has large effects on salivary testosterone, estradiol, and secretory immunoglobulin A assays in women and men. *Psychoneuroendocrinology*, *35*(2), 305–309.

van Anders, S. M., Goldey, K. L., & Bell, S. N. (2014). Measurement of testosterone in human sexuality research: Methodological considerations. *Archives of Sexual Behavior*, *43*(2), 231–250.

Vongas, J. G., Al Hajj, R., & Fiset, J. (2020). Leader emergence and affective empathy: A dynamic test of the dual-hormone hypothesis. *PLoS One*, *15*(12), e0244548.

Welker, K. M., Lassetter, B., Brandes, C. M., Prasad, S., Koop, D. R., & Mehta, P. H. (2016). A comparison of salivary testosterone measurement using immunoassays and tandem mass spectrometry. *Psychoneuroendocrinology*, *71*, 180–188.

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

Wolf, O. T. (2019). The impact of psychosocial stress on cognition. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge international handbook of social neuroendocrinology* (pp. 441–453). Routledge. Wolf, O. T. (2009). Stress and memory in humans: twelve years of progress? *Brain Research*, *1293*, 142–154.

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). American Psychiatric Publishing, Inc.