
Abstract
Since the initial reporting of salivary hormone measurements in marathon runners in the early 1980s, the practice of utilizing salivary testosterone (T) and cortisol (C) to reflect acute and rhythmic changes to their systemic counterparts has gained considerable momentum. However, substantial variability exists between studies with respect to methodological protocols, laboratory techniques, and interpretation of study findings. These differences can directly influence the salivary hormone values, thus hampering interpretation, limiting cross-study comparison, and constraining the generalizability of individual study findings. This article examines the current body of literature before proposing a sequence of practical guidelines to minimize sample variability in salivary hormone research. The guidelines are grouped into 3 major categories that limit comparison between studies: A) study design, B) sample acquisition and biological variation, and C) technical and analytical error. To achieve this, the present article critically appraises research employing salivary T and C measurements, identifies potential sources of error before proposing appropriate methodological considerations for researchers and practitioners wishing to obtain T and C measurement from saliva.

Glossary of Terms
- C: Cortisol (general term whereby the binding status of the hormone is undefined).
- T: Testosterone (general term whereby the binding status of the hormone is undefined).
- Sal-C: Salivary cortisol.
- Sal-T: Salivary testosterone.
- Bio-C: Bioavailable cortisol in serum. i.e., cortisol bound loosely to albumin or unbound.
- Bio-T: Bioavailable testosterone in serum. i.e., testosterone bound loosely to albumin or unbound.
- Free-C: Free cortisol in serum. i.e., unbound cortisol.
- Free-T: Free testosterone in serum. i.e., unbound testosterone.
- TC: Total cortisol. i.e., serum cortisol concentration which encompasses serum bioavailable cortisol, serum free cortisol, and serum cortisol which is bound to corticosteroid binding globulin.
- TT: Total testosterone. i.e., serum testosterone concentration which encompasses serum bioavailable testosterone, serum free testosterone, and serum testosterone which is bound to sex hormone binding globulin.

Introduction
Salivary testosterone (sal-T) and cortisol (sal-C) measurements are widely used in clinical, psychobiological, and exercise sciences as a surrogate for blood sampling. Sal-C was initially detected in the early 1960s [14,62] whilst sal-T was first reported in 1976 by Landman and colleagues [93]. In clinical settings salivary measurements are rarely used, except for the most extreme cases, such as use of sal-C for initial screening of Cushing’s syndrome [114] whilst sal-T is not used at all, having recently been ruled out of diagnosing hypogonadism [53,78]. In contrast, sports medicine and psychobiology research have widely adopted both sal-C and sal-T [11,29,35,71–73] promoting salivary measurement from novel surrogate to pervasive biomarker largely due to the widely held assumption that they mirror their respective systemic (blood) concentrations. This has resulted in the exponential growth in popularity of sal-T and sal-C in these fields, depicted by Fig. 1, which represents search results using salivary search terms in PubMed, performed in all fields. Whilst these data depict all returned results, this still
underestimates the total number of studies due to authors frequently omitting the term ‘salivary’ and describing the measure as simply ‘free testosterone’, or merely ‘testosterone’, or ‘cortisol’ in their manuscript title ([29, 30, 92] ○ Table 1). Despite increased prevalence in applied research, there are multiple methodological considerations that can hinder sound interpretation of resulting sal-T and sal-C data [74, 75]. Recently, the smallest observable change required before changes in sal-T and sal-C that can be considered biologically significant was determined (termed the ‘critical difference’ [79]) as 90% and 148% respectively, when determined by enzyme-linked immunosorbent assay (ELISA). Moreover, Valero-Politi and Fuentes-Arderiu [145] reported the critical difference of sal-T measured at monthly intervals to be 78%, using radioimmunoassay (RIA). In both cases, these investigations were conducted in controlled laboratory environments, while conversely, investigations conducted in naturalistic settings undoubtedly exhibit a greater degree of uncontrolled variability (higher critical difference). Such studies are at greater risk of falling victim to classic ‘type 1’ error. In addition, a recent meta-analysis described a profound effect of study design and sampling regimen on sal-T and sal-C reactivity in response to exercise [69]. Further to the observed heterogeneity in research design, exemplars of limitations with salivary hormone interpretation in psychobiology [74], behavior [75], and endocrinology [60] have been previously acknowledged. However, drawing attention to the limitations of otherwise well-designed studies, has marginal impact to inform progress within these fields of research.

With the ultimate aim of informing progress in salivary hormone research, there has been a recent call for standardized method for measurement of sal-T and sal-C and the use of standardized control conditions, [69]. Furthermore, whilst guidelines for detection of T and C in serum are well described [103, 124, 125, 150], there are currently a lack of guidelines for detection of T and C in saliva. As such, we propose that, in the spirit of good scientific practice, it is necessary to outline a series of procedures for researchers and practitioners to adhere to when conducting salivary hormone research. These guidelines can be grouped into 3 major categories; A) Study design, B) Sample acquisition and biological variation, and C) Technical and analytical error. Each of these aspects (A, B and C) coalesce to inform comparison between studies. Therefore, the aim of this article is to provide a critical appraisal of research using salivary measurements within the field of exercise science and sports medicine, identify possible sources of error, and propose appropriate methodological procedures for researchers wishing to...
Table 1  Sources of error in sal-C and sal-T measurement and methodological and technical guidelines recommended to limit error rates. Where evidence exists that disregarding the guideline will influence sal-T and sal-C, a checked box is present. Where evidence exists that disregarding the guideline may influence sal-T or sal-C, based purely on serum data, ‘evidence in TT’, or ‘evidence in TC’ is checked.

<table>
<thead>
<tr>
<th>Source of Systematic Error</th>
<th>Guideline</th>
<th>Evidence in sal-T</th>
<th>Evidence in sal-C</th>
<th>Evidence in TT</th>
<th>Evidence in TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 Randomised Controlled Trial (RCT)/Cohort Study</td>
<td>Is the study suitable for an RCT design? If not proceed to “collection timing”.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>A2 Collection Timing</td>
<td>Baseline samples collected on a control day to abrogate the anticipatory effect of impending exercise/physical activity bout or challenge. Increasing the number of baseline samples (i.e., on 2 separate rest days) may improve reliability but would require additional financing.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B1 Chronological Variation</td>
<td>Collected sal-T and sal-C at the same time of day on each occasion (including control days), during repeated measures studies. Considered circadian and seasonal pattern of sal-T and sal-C release for interventions over numerous hours, and months respectively.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B2 Chronobiological Age</td>
<td>Age-matched participants for between-subject investigation. Control for age during longitudinal studies. Used narrow age ranges or counterbalance studies of broad age ranges. The stage of each genders’ lifespan should be taken into account (E.g., andropause, menopause).</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B3 Sex</td>
<td>Sal-T and sal-C not compared between sexes.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B4 Sex; Male</td>
<td>Screened for androgenic anabolic steroid use (minimum verbally) and used as exclusion criteria. Used hypogonadism as exclusion criteria.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B5 Sex; Female</td>
<td>Reordered oral contraceptive pill use and stage of menstrual cycle. Used entirely pre/post-menopausal. Screened for polycystic ovary syndrome (minimum verbally) and used as exclusion criteria.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B6 Long Haul Travel</td>
<td>Participants who have travelled across multiple time zones within previous 7 days or spent time in hypoxic hypobaric conditions were excluded.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B7 Dietary Intake</td>
<td>Provided standardized meals or acknowledge increased biological variation. Instructed participants to avoid alcohol and caffeine for 24 h before sample collection.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B8 Postural Changes</td>
<td>Where possible, ensure participants maintained their posture for the duration of sampling. Subjects are seated or supine.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B9 Physical Activity</td>
<td>Minimized subjects’ exercise 24 h prior to sampling. If not possible, recorded programme variables to ensure comparability between sampling points.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B10 Smoking</td>
<td>Used as exclusion criteria. Smoking considered as significant confounder, particularly in cross-sectional comparison.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B11 Sexual Activity</td>
<td>Abstinence observed for 24 h prior to saliva collection (verbally screened).</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B12 Illness</td>
<td>Used as exclusion criteria.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B13 Psychological Influence</td>
<td>Familiarization with sampling was implemented. Erotic or aggressive stimulus avoided. Where psychological stress is a primary outcome measure, administered a stress questionnaire. E.g., PSQ [155].</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B14 Geographical Differences</td>
<td>Used a homogenous group of participants. In applied settings, practitioners must be aware that ethnic differences in sal-C and sal-T exist. Reported ethnic demographic in methods section.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C1 Collection Method</td>
<td>Avoided use of flavored beverage crystals, gum, and lemon juice to stimulate salivary flow. Dental roll, cotton swab and sugar free gum also avoided. Unstimulated saliv was collected mid-flow.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C2 Blood Contamination</td>
<td>Participants instructed to avoid teeth brushing, hot drinks, and gum shields prior to collection.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C3 Storage</td>
<td>Samples frozen quickly and stored at ~80 °C if storing longer than 12 months, ~20 °C for durations below 12 months. Room temperature for storage up to 6 h.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C4 Analysis</td>
<td>Ensured samples were analyzed in the same laboratory, by the same researcher, using the same assay kit. Reported coefficient of variation in all cases and not make direct comparisons between techniques.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Reporting of Data

| R1 Study Title | If the main outcome of the study regards sal-T or sal-C changes, then includes reference to “salivary” in the title, in other cases use “salivary” as a search term (e.g., keywords). | ✓                 | ✓                 | ✓              | ✓              |
| R2 Use of Standardized Reporting | If reporting relative changes in sal-T and/or sal-C then also include absolute values (mean ± standard deviation) at each time point. | ✓                 | ✓                 | ✓              | ✓              |

Sources of Systematic Error: This includes all major sources of potential imprecision (see A; B; C below) that contrive to increase measurement error in studies that employ salivary hormone measurement

A: Considerations for Study Design: These factors should inform study design. Unlike B and C, they are not always practicable
B: Sources of Biological Error: These factors should inform sample collection and study protocols. Sources of pre-sampling biological error should inform part of participant/athlete information sheet
C: Sources of Analytical Error: These factors should inform sample measurement
Table 2  Conversion between commonly used units for salivary testosterone within physiologically observed ranges. The molecular mass of testosterone is 362.46 g·mol⁻¹ and therefore 1 mol = 362.46 g. As such, 2 nmol·L⁻¹ = 724.92 ng·L⁻¹ = 72.5 ng·dl⁻¹.

<table>
<thead>
<tr>
<th>Conventional Units</th>
<th>SI Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>362.46 pg·ml⁻¹</td>
<td>1 nmol·L⁻¹</td>
</tr>
<tr>
<td>72.5 pg·dl⁻¹</td>
<td>1 nmol·L⁻¹</td>
</tr>
<tr>
<td>108.7 ug·dl⁻¹</td>
<td>3 nmol·L⁻¹</td>
</tr>
<tr>
<td>145.0 ug·dl⁻¹</td>
<td>4 nmol·L⁻¹</td>
</tr>
<tr>
<td>181.2 ug·dl⁻¹</td>
<td>5 nmol·L⁻¹</td>
</tr>
<tr>
<td>236.5 ug·dl⁻¹</td>
<td>10 nmol·L⁻¹</td>
</tr>
<tr>
<td>543.7 ug·dl⁻¹</td>
<td>15 nmol·L⁻¹</td>
</tr>
<tr>
<td>725.0 ug·dl⁻¹</td>
<td>20 nmol·L⁻¹</td>
</tr>
</tbody>
</table>

Table 3  Conversion between commonly used units for salivary cortisol within physiologically observed ranges. The molecular mass of cortisol is 288.42 g·mol⁻¹ and therefore 1 mol = 288.42 g. As such, 2 pmol·L⁻¹ = 576.84 pg·L⁻¹ = 0.57 pg·ml⁻¹.

<table>
<thead>
<tr>
<th>Conventional Units</th>
<th>SI Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>576.84 pg·L⁻¹</td>
<td>2 pmol·L⁻¹</td>
</tr>
<tr>
<td>288.42 pg·L⁻¹</td>
<td>1 pmol·L⁻¹</td>
</tr>
<tr>
<td>144.21 pg·L⁻¹</td>
<td>0.5 pmol·L⁻¹</td>
</tr>
<tr>
<td>72.10 pg·L⁻¹</td>
<td>0.25 pmol·L⁻¹</td>
</tr>
<tr>
<td>36.05 pg·L⁻¹</td>
<td>0.1 pmol·L⁻¹</td>
</tr>
<tr>
<td>18.02 pg·L⁻¹</td>
<td>0.05 pmol·L⁻¹</td>
</tr>
<tr>
<td>9.01 pg·L⁻¹</td>
<td>0.025 pmol·L⁻¹</td>
</tr>
<tr>
<td>4.50 pg·L⁻¹</td>
<td>0.01 pmol·L⁻¹</td>
</tr>
</tbody>
</table>

Obtain T and C measurement from saliva samples. Due to the significant heterogeneity of study design in studies that examine behavioral or psychological stimuli, this article is restricted to studies involving physiological outcome measures. However, the proposed guidelines can be applied across all studies using single or serial salivary hormone measurements.

A: Study Design

A recent meta-analysis concerning exercise-induced sal-T and sal-C reactivity to acute exercise identified that effect sizes were independent upon study design [69]. For example, pooled analysis of power exercise interventions that collected their baseline sample immediately prior to exercise observed a relative decrease in sal-T following a power-based exercise intervention [48, 105], whereas several other studies that collected a sample immediately prior to exercise observed an increase in sal-T [141, 152]. Without comparison to a control condition, it is difficult to discern the true effect of an intervention on sal-T and sal-C. Therefore, we recommend utilizing an RCT design where possible. Furthermore, if RCT designs are not feasible, and where participant data is to be compared with their own baseline data, the timing of salivary sampling becomes of paramount importance.

Study title

While the majority of studies correctly report the use of salivary hormone methodology within titles, this is not always the case [29, 30, 92]. It has been suggested that sal-T and sal-C ostensibly measure the free fraction of the hormone [118] which is considered to diffuse passively across the salivary glands [122]. Khan-Dawood [86] reported free testosterone (free-T) to account for 78% of sal-T, compared with approximately 4% of TT in adult plasma, although the variability around this value is yet to be determined. Furthermore, ‘free testosterone’ is traditionally interpreted as the unbound component of serum concentrations. However, despite these issues a number of authors describe sal-T as ‘free testosterone’ [36, 38] within study titles. Such descriptions are both confusing and premature, at least until studies using gold-standard ultrafiltration measures of serum T confirm the T:sal-T ratio and the reliability of sal-T to reflect serum derived free-T [53]. Therefore, in the interest of good practice, we believe it necessary to state ‘salivary’ in the title of research articles for clarity and indexing purposes, and throughout manuscripts for transparency.

Study design

A recent meta-analysis of randomized control trials (RCTs) and uncontrolled trials (UCTs) resulted in different magnitudes of sal-T and sal-C response to exercise [69]. It was suggested that the anticipatory rise (or fall) prior to exercise was the underlying mechanism for this outcome. To further elucidate this point, two investigations that collected baseline samples on a rest day observed a relative decrease in sal-T following a power-based exercise intervention [48, 105], whereas several other studies that collected a sample immediately prior to exercise observed an increase in sal-T [141, 152]. Without comparison to a control condition, it is difficult to discern the true effect of an intervention on sal-T and sal-C. Therefore, we recommend utilizing an RCT design where possible. Furthermore, if RCT designs are not feasible, and where participant data is to be compared with their own baseline data, the timing of salivary sampling becomes of paramount importance.

Timing of pre- and post-intervention samples

Since the first account of salivary hormone collection around the time of exercise [32], an anticipatory effect of the forthcoming activity has been observed. Cook and colleagues [32] reported increased sal-C and sal-T prior to a marathon run when compared to rest days. Therefore, we suggest that baseline samples are collected on a control day (without the anticipatory effect of intervention), as well as immediately pre-intervention, to elucidate the true effect of: A) the anticipatory effect on sal-T and sal-C compared to a control day, B) the intervention effect on sal-T and sal-C compared to pre-intervention, and C) the combined effect of anticipation and intervention on sal-T and sal-C. With regards to post-intervention sample timing, Crewther et al. [34] observed a negative sal-T response immediately after completion of an exercise intervention, which when reassessed 15 min later, resulted in a positive change from baseline. We therefore propose post-intervention samples are taken between 15–30 min post-intervention as sal-T has been shown to peak 15 min post-exercise intervention and return to baseline values 30 min post-exercise [153].

Data reporting

There are a number of issues in data reporting which, if standardized, should improve the field of salivary hormone research. For example, reporting raw mean data in combination with either standard deviation (SD), standard error of the mean (SEM) or confidence intervals (CI) rather than percentage change from baseline would add value and clarity to the data. Although reporting delta changes may be useful to contextualize findings, it may underestimate the absolute change in individuals with
currently, a mix between pmol · L$^{-1}$ following intervention experiences an absolute increase of 100 pmol · L$^{-1}$ and relative change of 33% yet an individual whose sal-T rises from 200 to 280 pmol · L$^{-1}$ following intervention experiences an absolute increase of 80 pmol · L$^{-1}$ but a relative change of 40%. We therefore suggest reporting mean ± SD/SE as a standard minimum. Relative changes may be reported, to further contextualise findings in relation to a minimum meaningful change from baseline [91]. Regarding the minimum meaningful change, it may be useful to refer to the critical difference of sal-T and sal-C to determine whether changes observed are biologically meaningful [79, 145]. This can be conducted for interventions within [79], or between days [145]. Furthermore, authors should adhere to the Système International d’Unités (SI Units) when reporting concentrations of sal-T and sal-C for continuity (Table 2–3). Currently, a mix between pmol · L$^{-1}$ and pg · ml$^{-1}$ is used for sal-T whilst a mix between nmol · L$^{-1}$ and ug·dl$^{-1}$ is used for sal-C.

Strength of evidence and recommendations

There is considerable evidence that identifies study design and timing of pre- and post-intervention sample to adversely influence study outcomes from sal-T and sal-C research. Furthermore, discrepancies in data reporting and terminology compound inconsistencies within the literature.

B: Sample Acquisition and Biological Variation

Variability of sal-T and sal-C have been previously reported [70, 79, 137, 145]. Recent evidence has described wide variation in the measurement of sal-T and sal-C within a highly controlled laboratory environment [79]. Therefore, in research outside of a laboratory, variability of sal-T and sal-C will likely be greater. In this section we discuss causes of some of the major sources of biological variation and how these may be controlled in an applied research environment.

Chronobiological variation

Circadian rhythmicity has been established in serum C [129] and T [97, 156], as well as in sal-C [111, 112] and sal-T [70, 90, 139]. As a result of these diurnal variations, it is imperative to employ similar sampling times during repeated measures research design. Furthermore, in prospective cohort studies where a baseline measure will be used as control data (i.e., without a control group), it may be difficult to determine true sal-T and sal-C response to exercise if the exercise intervention is protracted. For example, if examining sal-T and sal-C response to a marathon run, recreational athletes would take > 3 h to complete the course. If using a baseline-as-control design, it would prove difficult to determine whether hormone alterations observed were a result of running, or simply a product of chronological advancement. Therefore, we suggest the use of dual baseline sampling for protracted interventions, i.e., Samples should be collected pre-intervention, and time-matched post-intervention on a control day to eliminate the effect of diurnal variation.

Chronological age

Long haul travel

Long haul travel across multiple time zones causes a depression in sal-C the day following travel [18]. Whilst there is no information from sal-T data, data from serum indicate that after a simulated long haul flight, TT circadian rhythm was not altered [33]. However, simulation of hypoxia and hypobaria was 12000 ft., whereas commercial airplanes fly at over 30000 ft. In addition, as a result of circadian disruption due to crossing multiple time zones, long haul travel may influence serum TC and TT release. Therefore, recent long haul travel should be used as exclusion criteria where possible when obtaining sal-T and sal-C measurements.

Dietary intake

There are a wide variety of dietary factors that have been shown to influence sal-T and sal-C. Consuming an evening meal causes an increase in sal-T compared to fasting [143], whilst 3 weeks of daytime fasting, reduces sal-C [146]. High alcohol consumption
[110], low zinc intake [2], and a low carbohydrate diet [4] all decrease TT. Volek and colleagues [148] have demonstrated that the relative contribution to energy intake from protein, fat, saturated fatty acids, monounsaturated fatty acids, the polyunsaturated fat-to-saturated fat ratio, and the protein-to-carbohydrate ratio influenced TT but not TC concentrations, using a 17-day dietary recall. Heikkonen and coworkers [80] reported that alcohol consumption depressed TT, but increased TC concentrations. Controlling for diet may further reduce biological error in sal-T and sal-C research. This can be achieved by providing standardized meals. Dietary recall may be more cost-effective, if the known problems with dietary recall in children [107] and adults [54] can be reconciled.

Postural changes
Changes in plasma volume with movement between supine and standing positions are now well established though sometimes overlooked. Movement from an initial standing position to a supine position has demonstrated significant haemodilution [64,140], although a method for plasma volume correction has been outlined by Dill and Costill [44]. Failing to standardize body position or failing to correct for any changes in plasma volume increases error and has important interpretative implications [52]. In cases where plasma hormones are measured, correcting for plasma volume changes becomes an important consideration. Sal-C has been shown to increase following 20 min of standing, compared to sitting and lying [81]. Hucklebridge et al. [84] proposed the orthostatic challenge, adjusting from sitting to standing, stimulated the hypothalamic-pituitary-adrenal (HPA) axis as a result of hypotension, increasing sal-C. However, these same authors [84] observed no alteration in the sal-C awakening response when participants remained supine for 45 min or standing immediately upon wakening. Whilst no articles exist to our knowledge, concerning postural change and sal-T, hydration status [100], and salivary flow rate [6] did not influence sal-T concentrations. However, this area requires further confirmatory data as few studies exist addressing postural change and salivary hormone concentrations, particularly sal-T. As there are few data to confirm the influence postural change has on salivary hormones, we propose that repeated measures designs utilize the same posture for sampling at all sampling points, and we advise samples are collected seated or supine.

Acute and habitual physical activity
Exercise has been shown to increase levels of sal-T and sal-C acutely and chronically [19,31,38,120,147], however there is some controversy surrounding chronic elevations in TT [99]; as some reports suggest depressed TT levels as a result of chronic endurance training [8]. The adaptation of the HPA axis to training is mediated by decreased sensitivity to TC [45,95] and altered tissue sensitivity to glucocorticoids [46,47]. However, during a rest day, endurance athletes’ serum TC concentrations have been reported as normal [47]. As the T:C ratio has been implicated as a marker of overtraining and training status [3,102], prior physical activity may influence findings of sal-T and sal-C investigations when examining individuals with a high training volume. It is common practice to exclude physical activity for 24 h prior to study commencement, which is preferential. However, in an applied setting, athletes may not have 24 h without training. In these instances, duration, and programme variables of last training session should be matched for repeated-measures designs.

Smoking
Basally, smokers demonstrate altered sal-T and sal-C compared to controls [9,106,155]. Furthermore, English and colleagues [49] reported elevated levels of total, free, and bioavailable T in smokers compared with non-smokers. Cigarette smoking is further associated with acutely elevated TC levels. However, the results of TC comparison in smokers and non-smokers have been inconsistent, and the significance of TC responses in smoking cessation is unclear [136,144]. Therefore, for precaution, we advise smoking status to be used as exclusion criteria when not considered the independent variable.

Sexual activity
Mohammed [41] reported that, in both males and females, sal-T increased during an evening when there was intercourse and decreased when there was not. Moreover, sal-T has been shown to increase during a visit to a sex club in men [50], but no increase in sal-T or sal-C was observed following sexual thoughts [58]. Sexual activity influences TT concentrations in males and females [119]. Dabbs and Hamilton [65] demonstrated TC response to sexual arousal was highly individualized in females and Exton et al. [51] reported no significant influence of sexual arousal of TC levels in males. We are unaware of any studies to date that report the influence of sexual activity on sal-C or sal-T response to intervention. As such, abstinence should be requested 24 hr prior to salivary collection, similar to alcohol, and physical activity.

Illness
Shattuck and colleagues [130] reported reduced sal-T in response to immune activation. Critical illness is often accompanied by hypercortisolemia, which has been attributed to stress-induced activation of the HPA axis, observed in both sera and saliva [115]. However, low corticotropin levels have been reported in critically ill patients, which may be due to reduced TC metabolism [15]. In men, serum TT decreases during sepsis, burns, myocardial infarction, and surgery [133–135]. Spratt and colleagues [134] reported patients admitted to critical care units displayed decreased TT that varied according to severity of illness. We are unaware of any study to date that evaluates the influence of low-grade illness (i.e., cold or flu) on sal-C or sal-T. This is likely due to difficulty recruiting and opportunity of sampling yet this may be an area for further exploration. Illness should be controlled for during sal-T and sal-C research to the best of researchers’ abilities (i.e., pre-participation questionnaire or verbal screening).

Psychological influence
Cortisol is the primary stress hormone and therefore is elevated in physiologically and psychologically stressful situations [126]. An increasing number of psychophysiology studies use sal-T and sal-C to reflect differences or changes in mood state [20,113,158], though there are a number of contradictory overlaps and methodological issues [74,75]. Bernhardt and colleagues [12] investigated the sal-T response of football fans watching their teams win or lose. The study found that elation the winning team’s fans felt was accompanied by a rise in post-game sal-T levels compared to pre-game. The despondency felt by the losing team’s fans was accompanied by a decrease in sal-T. Increased sal-T in response to, or anticipation of, competition and sports performance has previously been observed [127]. This phenomenon has also been observed in the non-physical competition of chess [16,68], reporting increased pre-competition sal-T in eventual...
winners. Territorial aggression of home teams is also well documented [5] linked to agonistic animalistic behavior defending a home territory [113].

Chronic stress can decrease TT [1, 104]. It has been suggested that the suppression of steroidogenesis in the testes is due to reduced synthesis of testicular androgens, caused by the inhibitory effect of high adrenocorticotrophic hormone (ACTH) levels that accompany chronic stress [24]. Conversely, acute stress can increase TT [123], which may be a consequence of increased sensitivity to LH [24] which is supported by documented sympathetic stimulation by catecholamine release in males and females [26, 27]. The use of saliva sampling as opposed to venous blood draws may reduce stress in some instances. However, familiarization with passive drool is still advised to avoid apprehension. Fingerprick sampling and capillary tubes may be a viable alternative due to the relatively small volume of serum required for analysis and therefore less apprehension than during venous blood collection. Capillary TC accurately reflects samples from venous blood [55], yet further investigation is required as to whether this is a viable technique for testosterone measurement.

An often cited relationship exists between T and aggression, which is further demonstrated by the reported increased aggression or ‘roid-rage’ following consumption of supraphysiological dose T in anabolic androgenic steroid users (AAS). However, in individuals within normal physiological ranges of TT, only a weak positive relationship between TT and aggression exists in humans [5]. We propose that aggressive or stressful stimuli should be used as inclusion/exclusion criteria, preceding saliva collection. Where this may not be possible, practitioners may wish to control for stress by administering a stress questionnaire (e.g., perceived stress questionnaire (PSQ)) [89].

Ethnographical differences

Martin et al. [101] described racial differences in diurnal sal-C rhythms whereby African Americans exhibited dampened morning-to-evening sal-C slopes than Caucasians. Christiansen [25] observed that sal-T from men in !Kung San was lower in morning-to-evening sal-C slopes than Caucasians. Christiansen [116] described significant heritability in sal-T measures (−0.42 and −0.47 for at-home and in-lab values respectively). Serum TT levels are higher in adult male Bangladeshi migrants to the United Kingdom compared to residents of Bangladesh [108], and in native Anyamara men in urban vs. rural Bolivia [10], as well as Chinese men living in Pennsylvania vs. Beijing [128]. However, this may be as a consequence of dietary intake. Winter and colleagues [154] have reported higher TT concentrations in African-American men compared to Caucasian men, whereas Litman and coworkers [98] observed no such difference. Therefore, in between-group investigations, it may be useful to control for ethnic differences or at least acknowledge the influence ethnic or geographical location may exert.

Strength of evidence and recommendations

In summary, a number of our recommendations are based on serum data, which we have extrapolated to the determination of salivary hormones. This in itself is problematic due to some authors reporting weak, or no relationships between salivary and serum values [61, 76, 78, 87, 131]. There is however, evidence that chronological variation, chronological age, sex, dietary intake, acute and habitual physical activity, smoking, illness, mood state and psychology, and geographical differences influence sal-T and sal-C. Conversely, whilst there is evidence of long haul travel influencing sal-C, currently there is no investigation to our knowledge concerning the influence of travel on sal-T. Whilst TT was reportedly unaltered following hypoxic hypobaria, conditions investigated did not closely match those of long haul flights and thus we cannot declare for certain whether sal-T is influenced by long haul travel. Sal-C has been shown to vary following postural change. However, there are no data addressing postural changes and sal-T. Extrapolating from serum data, Hoffman and colleagues [82] observed no change in TT following reduced plasma volume and therefore sal-T may remain unchanged. However, this is purely speculation until further confirmatory data.

C: Technical and Analytical Error

As previously described [79], the critical difference includes both biological error, and technical and analytical error. As with the biological error discussed above, steps should be taken to minimize the technical and analytical error when assessing samples. Lazarou and colleagues [94] reported that in a study examining reference values for hypogonadism in 25 laboratories, there were 17 different threshold values for TT. Indeed, the threshold for hypogonadism diagnosis varied by 350% (130 ng·dl⁻¹ to 450 ng·dl⁻¹). Whilst there is some discrepancy concerning clinical thresholds in serum testosterone, the innate variability of sal-T means it is unsuitable for clinical diagnosis [53, 78]. To exemplify this point, Jensen et al. [85] compared analysis of sal-T and sal-C samples amongst 4 and 3 laboratories respectively. These authors reported recovery of spiked material for testosterone and cortisol was 80–94% and 83–100% respectively. Moreover, substantial differences existed between laboratories, as a result of traceability, clean-up procedures, or issues with calibration Crewther and colleagues [37] reported power trained males had mean basal sal-T concentrations of 107 pg·ml⁻¹ (371 pmol·L⁻¹) whereas Ghigarielli et al. [57] reported basal levels of as high as 180 pg·ml⁻¹ (624 pmol·L⁻¹). In a multicentre study [40], mean sal-T concentrations from the same 100 males ranged from 240 ± 95 to 410 ± 191 pmol·L⁻¹ (69–118 pg·ml⁻¹) suggesting considerable heterogeneity between laboratories. Fiers et al. [53] recently provided evidence for T binding to salivary proteins, therefore limiting the agreement with free-T in serum, determined by equilibrium dialysis. These taken together, suggest large variability in sal-T, so precautions are needed in order to reduce sampling error. In this section, we detail some of the major analytical variation and provide advice for researchers using sal-T and sal-C.

Collection method

Flavored beverage crystals and lemon juice have been used to stimulate flow of saliva. Flavored crystals may cause an increase in measured sal-C concentrations whereas lemon juice may compromise sal-C determination as a result of decreased pH [59]. With regards to sample acquisition, Granger et al. [61] observed chewing sugar free dental gum resulted in increased sal-T levels after 1 min (mean 168 pmol·L⁻¹, SEM = 58 pmol·L⁻¹ compared to 138 pmol·L⁻¹, SEM = 53 pmol·L⁻¹) but thereafter no significant differences were observed. Granger et al. [61] investigated the effects of different sample collection techniques on the measurement of sal-T. In this study, sal-T collected using cotton dental roll, cotton swab and sugar free gum were com-
pared to un-stimulated saliva collection. Compared to un-stimulated saliva (mean = 10.1 pg·ml⁻¹ [34.9 pmol·L⁻¹]), sal-T levels were 2-fold higher (mean = 20.7 pg·ml⁻¹ [71.4 pmol·L⁻¹]) using dental cotton roll, and almost 3-fold higher (28.3 pg·ml⁻¹ [98.1 pmol·L⁻¹]) after using a cotton swab. Moreover, Shircliff and coworkers [131] reported compromised assay results when using cotton materials to absorb saliva. It was shown that the cotton interference effect was of sufficient magnitude to attenuate the association between serum and saliva levels. Therefore, unstimulated, mid-flow saliva is recommended for analysis. It may also be advised to wash out the mouth with distilled water approximately 10 min before sample collection to clear the oral cavity of debris.

Blood contamination

Hormones in saliva are present at far lower concentrations than in circulation. As a result, blood leakage into the oral mucosa can compromise the validity of salivary hormone analysis [61, 88]. Blood contamination may be caused by teeth brushing, drinking hot fluid, or the use of swabs for sample collection [96]. Blood can leak into saliva as a result of micro-injuries such as burns, cuts or abrasions or gum disease/ poor oral hygiene. In sport, both the abrasion to the gums from the use of gum shields and facial injuries can lead to the presence of blood in saliva. Kivilähtinen et al. [88] investigated the effect of blood leakage on sal-T and sal-C concentrations. Saliva samples were taken before, immediately after, and every 15 min for 1 h following vigorous tooth brushing and the same protocol without tooth brushing was performed by a control group. There was a significant increase in sal-T at 15, 30 and 45 min post micro-injury compared to control. paradoxically, there was no difference in sal-C levels between the micro-injury and the control condition. This suggests that sal-C is unaffected by blood contamination in saliva although given the appreciable concentration gradient between serum and saliva C, this requires further investigation. Whilst few studies report the acceptable transferrin level in saliva although given the appreciable concentration gradient between serum and saliva C, this requires further investigation. Whilst few studies report the acceptable transferrin level in saliva, this could be used as exclusion criteria [39, 88]. We propose that investigators ensure avoidance of hot food and drinks, and teeth brushing prior to sample collection. Furthermore, we suggest that following gum shield use, samples are visually inspected for blood contamination and then leakage confirmed by transferring analysis.

Storage

Steroid hormones such as C have generally been considered to be stable in saliva even when stored at room temperature for a number of days [23, 56]. However, more recent studies have raised concern with respect to sample treatment and storage prior to analysis [61, 117, 142]. Concentrations of sal-C were found to decrease by 9.2% per month in samples stored at room temperature [56] compared to baseline. Toone et al. [142] observed that following 7 days of storage at 4 °C, sal-T decreased by 26±15% whereas sal-C remained unchained compared to baseline. With regard to storing saliva samples, 2 freezing options are generally used. Saliva can be stored for a year in a domestic freezer (−20 °C), and possibly several years in a laboratory-based freezer (−80 °C [66]). After collection, saliva should be frozen as soon as possible [117] to precipitate mucus; however if unavailable, saliva can be stored at room temperature for up to 6 h [66]. Storing saliva at >−5 °C will not freeze the samples and is not generally recommended due to bacterial growth which may degrade salivary components and interfere with antibody binding [61]. Granger et al. [61] analyzed saliva samples stored at 4 °C which were assayed for sal-T on a weekly basis for 4 consecutive weeks. Contradictory to the data of Toone et al. [142], measured sal-T levels increased by 20.6% after one week and 330.8% after 4 weeks compared to baseline. Therefore, we recommend studies freeze saliva samples as quickly as possible in a freezer as cold as possible. −80 °C is preferred but −20 °C is suitable as long as samples are analyzed within 12 months.

Analysis

Numerous techniques have been used to measure sal-C and sal-T concentrations, including enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and liquid chromatography–tandem mass spectrometry (LC–MS/MS). Analytical sensitivity differs between these methods, as discussed elsewhere [53, 124]. Despite the ease and speed of analysis compared to LC–MS/MS, results from ELISA and RIA are generally considered less accurate, possibly due to the lack of consistent reference standards. More recently, handheld detection devices have come to market offering salivary hormone determination [28]. However, whilst there are conference proceedings suggesting acceptable reliability and validity of these devices, no full text articles exist to date confirming suitability of sal-T and sal-C values using this method.

Variability of serum thresholds varies between laboratories [94], and possibly as a result, reference ranges for sal-T have not been agreed to date. As such, sal-T and sal-C assays are limited to interpretation by identifying ‘outliers’ within a group/team and/or comparison with previously determined measure. In an attempt to minimize error through analytical procedure, we suggest LC–MS/MS be used where possible. When this is not viable, due to the innate laborious methodology, it is imperative all analysis is conducted in the same laboratory, by the same investigator, in duplicate or triplicate. We also advise that direct comparisons are made only with findings from the same methodological technique. Coefficient of variation needs to be reported in all cases.

Strength of evidence and recommendations

In summary, there is evidence that sal-T and sal-C is directly influenced by collection method, blood contamination, storage duration and temperature, and analysis method. Whilst saliva collection has obvious advantages over blood, accuracy, variability, and salivary protein binding may limit interpretation of results. As capillary sampling has been validated for analysis of TC, this may be an option for TT determination, however this currently requires validation.

Conclusion

Salivary analysis of C and T in sports people will soon enter its fourth decade as a research tool. The benefits of this method are well documented and primarily focus on ease of sample and subsequent analysis. However, there is a general lack of consistency between studies that use salivary measures of T and C as surrogates for their systemic counterparts. These include issues related to A) study design, B) sample acquisition and biological variation, and C) technical and analytical error. We would
encourage researchers to consider each of A, B, and C, as outlined above in the formulation of study design, participant information sheets, and in their reporting of study outcomes.

Affiliations
1Department of Medical and Sport Sciences, University of Cumbria, Lancaster, United Kingdom of Great Britain and Northern Ireland
2Institute of Clinical Exercise and Health Science, University of the West of Scotland, Hamilton, United Kingdom of Great Britain and Northern Ireland
3English Institute of Sport, English Institute of Sport, Marlow, United Kingdom of Great Britain and Northern Ireland
4Faculty of Health, Federation University, Victoria, Australia

References

12 Bernhardt PA, Dabbis JM Jr, Fielden JA, Ludder CD. Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. Psychol Behav 1998; 65: 59–62
29 Cook CJ, Crewther BT, Kidduff LP. Are free testosterone and cortisol concentrations associated with training motivation in elite male athletes? Psychol Sport Exerc 2013; 14: 882–885
30 Cook CJ, Crewther BT, Smith AA. Comparison of baseline free testosterone and cortisol concentrations between elite and non-elite female athletes. Am J Hum Biol 2012; 24: 856–858
36 Crewther BT, Cook CJ. Effects of different post-match recovery interventions on subsequent athlete hormonal state and game performance. Physiol Behav 2012; 106: 471–475
41 Dabbis JM Jr, Mohammed S. Male and female salivary testosterone concentrations before and after sexual activity. Physiol Behav 1992; 51: 195–197
42 Di Luigi I, Romanelli F, Sgro P, Lenzi A. Andrological aspects of physical exercise and sport medicine. Endocrine 2012; 42: 278–284
