Suitability of saliva stimulants for valid assessment of steroid hormones via radioimmunoassay

In press, *Psychoneuroendocrinology*

Gelena Dlugash*, Oliver C. Schultheiss

Chair for Experimental Psychology, Motivation, and Affective Neuroscience

Department of Psychology

Friedrich-Alexander University,

Nägelsbachstrasse 49b, 91052, Erlangen, Germany

*correspondence to G.D.: Email: gelena.dlugash@fau.de

Phone: +49 (0)9131/85-20871

Abstract

Salivary steroid measurement is a popular way to assess endocrine hormone levels, but efficient sample collection can be challenging because the use of stimulants can interfere with valid measurement. The aim of this study was therefore to identify a stimulant that can be used in assessment of the steroid hormones cortisol (C), testosterone (T), progesterone (P) and estradiol (E2) without impairing their quantification by radioimmunoassay. Study 1 and 2 explored the suitability of potential stimulants in comparison to unstimulated saliva collection. Study 3 tested stimulants under standardized conditions in water. Across all three studies, Parafilm® wax foil performed best and was therefore tested once more and validated as a saliva stimulant in Study 4. No significant differences between unstimulated saliva and Parafilm®-stimulated saliva could be found for any of the four hormones assayed. Therefore, Parafilm® appears to be a suitable saliva flow stimulant for assaying the salivary steroid hormones C, T, P and E2 by radioimmunoassay.

Keywords: saliva stimulant, steroid hormones, radioimmunoassay

Introduction

The measurement of steroid hormones in saliva offers undeniable advantages over alternative biofluids such as blood or urine. Saliva collection is simple, cost-effective and non-invasive and therefore stress-free for research participants (Kaczor-Urbanowicz et al., 2017). Moreover, salivary steroids represent only the unbound, bioactive fraction of steroids circulating in blood. These properties have made salivary hormone measures successful in various disciplines like psychology, endocrinology, or pediatrics (Kaczor-Urbanowicz et al., 2017). Depending on the analyzed hormone, the number of hormones analyzed simultaneously, and the assay method, however, large sample volumes are often required. For instance, to assay the steroid hormones cortisol (C), testosterone (T), progesterone (P) and estradiol (E2) in duplicate by radioimmunoassay, 4-5 ml of saliva have to be collected. Such large amounts are required as commercially available radioimmunoassay kits are designed for assaying steroids in blood where much higher effective analyte concentrations are present. (Enzyme-based steroid immunoassay methods often require much smaller samples, but in some cases their validity appears to be less robust than that of steroid radioimmunoassay (Schultheiss, Dlugash, & Mehta, 2019). The 20- to 50-fold lower concentration of steroids in saliva in comparison to blood can therefore only be assessed reliably by using larger sample volumes.

Saliva sampling therefore becomes an issue. Obtaining such large volumes requires a long time for participants to collect a sample, and this can be difficult to realize under laboratory or field conditions. One previously employed solution was the use of chewing gum to increase saliva flow rate and thus shorten the time needed to collect a sample. Although chewing gum containing sugar yields inflated salivary steroid hormone measurements by cross-reacting with specific antibodies of an assay (Lipson & Ellison, 1989), in the early 1990s sugarless chewing gum has been shown to produce almost identical salivary testosterone levels in comparison to unstimulated saliva (Dabbs, 1991). Additionally, use of chewing gum was considered to be safe because higher flow rates did not substantially alter the concentration of steroid hormones passing into saliva via passive diffusion (e.g. C, T, P or E2) (Buttler et al., 2018; Vining, McGinley, & Symons, 1983), although there were hints that flow rate affects the concentration of other biomarkers like DHEAS (Justino, Teixeira, Peixoto, Jaramillo, & Espindola, 2017). Nevertheless, for a long time sugar free chewing gum became the

stimulant of choice for psychoendocrinology.

Later, van Anders (2010) reported that sugarless chewing gum can inflate E2 and T measurements by 50% to 150%. Moreover, Schultheiss (2013) observed that sugarless chewing gum could double P and attenuate T and C levels by around 30%. These studies made many laboratories refrain from the further use of chewing gum as a saliva flow stimulant and switch back to unstimulated sampling. However, this also brought back the issue of time-consuming saliva collection.

Therefore, the goal of the present research is to find a stimulant that speeds up collection of saliva samples through the stimulation of saliva flow while also allowing the valid simultaneous assessment of multiple hormones (C, T, P and E2) without inducing measurement bias. Further requirements were that the stimulant must be non-toxic, show no interactions with the analytes, and be cheap and easy to obtain. In total, we conducted four studies. The first two studies were explorative, testing an array of potential stimulants under real conditions with human participants. In Study 3, stimulants were investigated in water under standardized laboratory conditions. Study 4 verified the equivalence of salivary steroid hormone concentrations in saliva samples obtained with or without stimulant use.

Study 1: Exploratory examination of saliva-flow stimulants

In Study 1 we tested a variety of materials as saliva-flow stimulants and compared their effects on steroid levels to those obtained with unstimulated saliva collection. We selected four possible stimulants that we expected to be inert towards steroid or antibody interactions. They included: a) a piece of paraffin wax foil (hereinafter referred to as Parafilm®), b) chewing gum base without added sugar or flavors (gum base), c) latex from a pacifier (latex), and d) chewing gum created in-house from gluten protein, glycerol, and water (gluten gum). We aimed to determine which of the four selected stimulant materials were suitable for saliva stimulation without biasing hormone measurements.

Method

Ethics statement. This study and all following were approved by the Institutional Review Board of Friedrich-Alexander University.

Saliva collection and hormone assays. Samples were collected between 2 and 4 pm to minimize effects of circadian variation (Vining, McGinley, Maksvytis, & Ho, 1983). Nine women (22-28 years old, mean age: 24.11 years) were instructed to refrain from drinking or eating, chewing gum or brushing teeth for 30 min prior to the experiment. Before collecting each sample, they rinsed their mouths with a sip of water, followed by a 5-min pause. Next, they chewed on a stimulant until they had collected 5 ml of saliva in a 50 ml polypropylene tube. Additionally, for control purposes, participants collected an unstimulated saliva sample. Thus, each participant provided 5 samples in total. Stimulant sequence was permutated within-subjects. Subsequently, all samples were frozen at - 20 °C.

In this study and all remaining studies, sample preparation, assessment and statistical analysis were conducted as follows. To break down mucopolysaccharides and thus allow subsequent precipitation of proteins, samples were thawed three times at room temperature and refrozen at - 20 °C. After the last thawing they were centrifuged for 20 min at 4 °C and 2000 rpm. The watery supernatant was transferred into a new tube and the residue was discarded. Samples were frozen at - 20 °C until assay.

We conducted radioimmunoassay (RIA)[1] using the following kits: solid phase CortiCote Cortisol (MP Biomedicals LLC, Orangeburg), ImmunoChem™ Double Antibody Testosterone (MP Biomedicals LLC, Solon), solid phase Progesterone (Beckman Coulter, Brea) and Ultra-Sensitive Estradiol (Beckman Coulter, Brea). The manufacturers declare the kits to be suitable for hormone assessment in blood and urine. Therefore, our laboratory adjusted and validated assay procedures for

[1] RIA is an established analytical method that is widely used in clinical as well as in research laboratories and shows stronger validity and higher convergence with LC-MS as compared to ELISA. (See Schultheiss, Dlugash & Mehta, 2019, for review).

use with saliva samples (Campbell, Schultheiss, & McClelland, 1999; Schultheiss, Dargel, & Rohde, 2003; Wirth & Schultheiss, 2006). Standards were diluted in water to yield the following calibration curve ranges: For C, 0.5 – 30 ng/ml, for T, 5 – 500 pg/ml, for E2, 0.625 – 20 pg/ml, and for P, 2.5 – 171 pg/ml.

For E, one assay was run that included samples from all four studies, whereas for C, T, and P two or more assays were run to cover samples from this and subsequent studies. In the following, we therefore report quality control checks for all four studies. Calibration curves were linear, with R^2s of 1 for C, and 0.99 T, E2, and P. Lower limit of detection ($B_0 - 3 \times SD$) was 0.02 ng/ml for C, 1.14 pg/ml for T, 0.47 pg/ml for E2, and 2.75 pg/ml for P. Recovery was calculated for the low, middle, and high range of calibration curve. For C, corresponding concentrations were 1, 2 and 4 ng/ml; for T, 5, 26 and 67 pg/ml; for E2, 2.3, 8.9 and 16.9 pg/ml, and for P, 5.5, 27.5 and 105 pg/ml. Recovery coefficients were 96.96%, 107.51% and 103.89% for C, 40.45%, 98.91% and 109.18% for T, 138.97%, 117.14% and 93.04% for E2, and 78.40%, 91.33% and 112.90% for P. Intra-assay coefficients of variation were 9.49% for C, 44.29% for T, 17.47% for E2, and 12.24% for P.

Statistical analysis. Statistical analysis was performed with SYSTAT 13 (Version 13.00.05). Values of samples with non-detectable hormone concentrations were set to zero for the respective hormone(s). Because data was not normally distributed, but samples were dependent, we used for all comparisons Wilcoxon signed-rank tests, a non-parametric test that is robust with regard to small sample sizes and deviations from normal distributions. Because our aim was to err on the side of caution when it comes to identifying a stimulant whose effects on salivary hormone concentrations do not differ from those obtained with unstimulated collection, we did not use Bonferroni adjustment and therefore retained maximum sensitivity for significant differences. We also calculated Spearman's correlations to evaluate the correspondence between stimulated and unstimulated saliva samples.

Open Science. All data files and analysis scripts related to reported results are available at https://osf.io/x9yqr/.

Results

The aim of Study 1 was to evaluate four different materials with regard to their suitability as saliva stimulant. In so doing, we focused not only on ease of handling but also on the equivalence of hormone concentrations in stimulated and unstimulated saliva samples. Table 1 displays median and range information for each hormone across saliva-collection conditions (see also Fig.1A). For gluten gum, one participant did not provide a sample due to celiac disease. Unless otherwise noted, in this and all following studies all comparisons were calculated with reference to unstimulated saliva.

For C, no significant differences could be found with Parafilm®, $z = 1.48$, $p = .14$, $n = 9$, gum base, $z = 0.41$, $p = .68$, $n = 9$, latex, $z = 1.36$, $p = .17$, $n = 9$, or gluten gum, $z = 0.14$, $p = .89$, $n = 8$. Therefore, for measuring C all stimulants appeared to be suitable. For T, due to the low levels of this hormone in women, several participants had non-detectable concentrations (= zero values), bringing the median to zero. Although the value range of Parafilm®, $z = -1.83$, $p = .07$, $n = 9$, suggested a trend-level decrease of T, the value range was comparable to the unstimulated-sample measurements. Notably, samples of the same participants showed zero values when collected without stimulation or with Parafilm®, suggesting that Parafilm® is comparable to unstimulated samples. Samples that were collected with other stimulants showed values above zero-level for several of these participants, suggesting that these stimulants emit some of their ingredients into saliva. Median levels for gum base, $z = 1.48$, $p = .14$ $n = 9$, latex, $z = 1.46$, $p = .14$, $n = 9$, and gluten gum, $z = 1.75$, $p = .08$, $n = 8$, were lower in comparison to unstimulated samples, while value ranges for these stimulants were two to four times higher. For P, samples stimulated by Parafilm®, $z = 0.56$, $p = .58$, $n = 8$ and latex, $z = -0.70$, $p = .48$, $n = 8$ were indistinguishable from unstimulated samples. (For P assessment, one participant was an outlier across several collection stimulants and her data were therefore omitted from analyses of P.[2]) In contrast, gum base samples were elevated by about 60%, $z = 2.52$, $p = .01$,

[2] The participant showed the following salivary concentrations [pg/ml]: 53.81; 22.76; 20.28; 6.73; 5.10. All samples were obtained within ca. 90 min. Such a strong decrease in P can´t be explained by a natural hormone decline. Therefore, we suspect that the subject used a chewing gum right before the test, which, as already known, can interfere with RIA assay.

n = 8, and gluten samples also showed a significant increase compared to unstimulated samples, z = 2.37, p = .02, n = 7. E2 levels did not differ significantly for Parafilm®, z = -1.40, p = 0.16, n = 9 and gluten gum, z = 1.26, p = .21, n = 8. But we observed significantly increased E2 concentrations for gum base, z = 1.95, p = .05, n = 9, and for latex z = 2.07, p = .04, n = 9.

For C, Parafilm® and gum base correlated significantly with unstimulated samples, while latex and gluten gum revealed only a weak correspondence (see also Table 2). For T, correlations of all stimulants were strong and highly significant. For P, strong correlations were found for latex and guten gum, but not for Parafilm® and gum base. For E2, correlations of all stimulants were weak and sometimes even negative, probably owing to the very low concentrations of analyte. For more detailed information, we also provide Passing-Bablok regressions and Bland Altman Plot analysis in the supplement.

Table 1

Median (range) of hormone concentrations for cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml] under various stimulated and unstimulated saliva collection conditions in Study 1.

	Unstimulated	Parafilm®	Gum base	Latex	Gluten gum
Cortisol	0.80 (0.36 - 1.79)	0.83 (0.43 - 1.97)	0.90 (0.39 - 2.09)	0.80 (0.33 - 2.55)	0.76 (0.20 - 2.22)
Testosterone	0.00 (0.00 - 2.92)	**0.00† (0.00 - 1.47)**	0.00 (0.00 - 5.16)	0.00 (0.00 - 12.01)	**0.18† (0.00 - 5.16)**
Progesterone	4.00 (2.07 - 12.55)	4.64 (0.76 - 12.74)	**6.43*** (4.13 - 13.76)	3.86 (1.77 - 9.91)	**12.71*** (9.18 - 23.68)
Estradiol	0.40 (0.00 - 1.10)	0.26 (0.00 - 1.12)	**0.70*** (0.38 - 1.34)	**0.98*** (0.40 - 1.91)	0.70 (0.25 - 22.39)

Note. All differences according to Wilcoxon signed-rank test in comparison to unstimulated sample.

** $p < 0.05$; † $p < .10$*

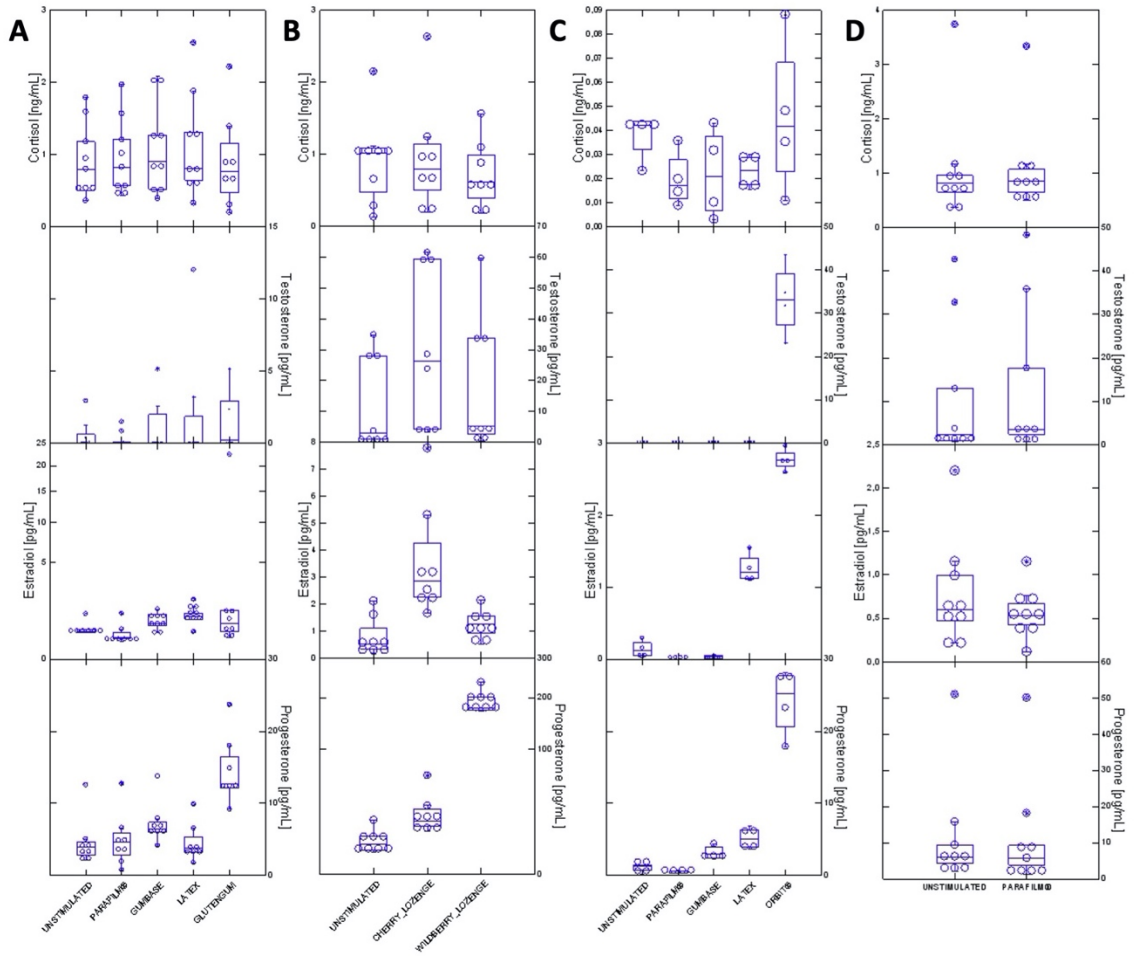
Table 2

Spearman´s correlation coefficients for cortisol, testosterone, progesterone and estradiol under various stimulated saliva collection conditions in Studies 1, 2, and 4.

	Parafilm® Study 1		Gumbase Study 1		Latex Study 1		Glutengum Study 1		Cherry lozenge Study 2		Wildberry lozenge Study 2		Parafilm® Study 4		Parafilm® Study 1 + 4	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Cortisol	.75	.02	.67	.05	.48	.19	.33	.42	.74	.04	.74	.04	.87	.003	.80	.0001
Testosterone	.92	.001	.82	.01	.90	001	.87	.005	.71	.05	.69	.06	.50	.17	.87	<.0001
Progesterone	.52	.18	.62	.10	.81	.01	.82	.02	.76	.03	.00	1.00	.98	<.0001	.85	<.0001
Estradiol	- .03	.93	- .20	.60	-.38	.32	-.37	.37	.11	.78	.43	.29	.70	.04	.54	.02

Figure 1:

Value distributions of cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml] under various stimulated and unstimulated saliva collection conditions in Studies 1 (**A**) *and 2 (**B**), for water-only samples and samples obtained by incubation with stimulants in Study 3 (**C**) and for unstimulated samples and samples obtained with Parafilm® in Study 4 (**D**).*



Discussion

Overall, our findings suggest that every tested material was suitable for assaying one or several of the four hormones but not for others. However, we think it is important to consider to what extent distribution patterns of individual values between unstimulated and stimulated samples are equivalent, as clear deviations can be seen in some cases even though median differences did not become significant. As Figure 1A illustrates, in terms of equivalence of median and range values, Parafilm® performed best for all hormones.

C in saliva is present in nanogram range and may therefore be more robust to possible interference from simulants. In contrast, T, P, and E2 are in the lower picogram range. For these latter three steroids we found some stimulants to increase hormone values significantly. Thus, some stimulants may not interfere with the assessment of high-concentration hormones (C) but may do so for hormones at lower concentrations.

Nevertheless, detected alterations in values could have been caused not only by the stimulant itself, but also by its serial position in the collection protocol. Every participant had to collect five samples in a short time, and this may have affected saliva flow rate or its composition. Further, our 5 min collection intervals may have been too short to clear the mouth completely from previous stimulants, resulting in contamination of the next sample.

It was not possible to achieve a balanced ratio of both sexes, and results are therefore strongly influenced by women's samples. Nonetheless, this may have worked to our advantage, because women tend to show significantly lower concentrations of T and stimulant-induced T increases would therefore be more visible and also would have a larger impact on female than on male samples. Viewed from this angle, our findings strongly suggest that Parafilm® shows no sign of increasing T measurements even in our female-biased sample.

Spearman´s correlations suggested that all tested stimulants were suitable for assaying T, but only some were also suitable for assaying C, P and E2. Nonetheless, correlation analyses should be treated with caution and viewed only as one approach to understanding each stimulant's suitability. Correlations were based on small sample sizes of 8 to 9 samples and are therefore particularly susceptible to outlier and leverage effects. For E2, the natural concentration range of 0 to 1 pg/ml is very restricted, low, and associated with comparatively high measurement error. This may explain the weak correlation coefficients. Negative coefficients resulted from samples with concentrations below 0 pg/ml that were set for analysis purposes to 0 pg/ml (see also statistical analysis), but in the comparison condition had a relatively high concentration of 1 pg/ml. This caused correlation coefficients to turn negative.

Additionally, gluten gum required more preparation time as it cannot be purchased ready-made and has to be manufactured in-house. Finally, gluten gum cannot be used by individuals

suffering celiac disease, which makes it unsuitable for wide application. Therefore, we dropped gluten chewing gum as a stimulant from further investigations.

Study 2: Lozenges as a stimulant?

In Study 2, we explored the suitability of two different types of lozenges as stimulants. In particular, saliva samples were stimulated with cherry flavored lozenge (cherry lozenge) and wildberry flavored lozenges (wildberry lozenge) and were compared to unstimulated saliva.

Method

The experiment was conducted between 2 and 4 pm. Eight participants (5 women, 3 men, 21-37 years old, mean age: 28.6 years) each collected two stimulated samples and one unstimulated saliva sample. Saliva collection, steroid assays, and statistical analyses followed the procedures described in Study 1.

Results

Table 3 displays median and ranges per stimulant for each hormone (see Figure 1B for an illustration of the value distributions). For C, Wilcoxon signed-rank tests revealed a trend towards reduced concentrations for wildberry lozenge, $z = -1.82$, $p = .07$, $n = 8$. For cherry lozenge, C was also lower, although this effect was not significant, $z = 0.28$, $p = .78$, $n = 8$. For T, wildberry lozenge increased concentrations by 40%, $z = 2.10$, $p = .04$, $n = 8$, and cherry lozenge increased concentrations more than 900%, $z = 2.52$, $p = .01$, $n = 8$. For P, cherry increased by over 300%, $z = 2.52$, $p = .01$, $n = 8$, wildberry lozenge samples showed an increase of 3100%, $z = 2.52$, $p = .01$, $n = 8$. For E2, concentrations increased significantly in samples collected using wildberry lozenges, $z = 2.24$, $p = .03$, $n = 8$, and cherry lozenges, $z = 2.52$, $p = .01$, $n = 8$.

As shown in Table 2, samples collected with cherry lozenges revealed a strong and highly significant correlation with unstimulated saliva for C, T and P, but not for E2. Wildberry-lozenge-stimulated samples showed a strong correlation only for C, while correlations for T and E2 were moderate. For P, no substantial correlation could be found in this case.

Table 3

Median (range) of hormone concentrations for cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml] under stimulated and unstimulated saliva collection conditions in Study 2.

	Unstimulated	Wildberry lozenge	Cherry lozenge
Cortisol	1.02 (0.13 - 2.15)	**0.61 † (0.18 - 1.56)**	0.80 (0.19 - 2.63)
Testosterone	2.82 (0.00 – 34.90)	**5.13* (0.43 – 59.64)**	**26.19* (3.20 – 61.58)**
Progesterone	5.80 (3.18 – 18.87)	**178.27* (171.66 -236.64)**	**18.27* (11.86 – 62.70)**
Estradiol	0.52 (0.17 – 2.12)	**1.25* (0.52 – 2.14)**	**2.86* (1.66 – 7.78)**

Note. All differences according to Wilcoxon rank sum test in comparison to unstimulated sample.

** p< .05; † p < .10*

Discussion

The aim of Study 2 was to test different lozenges as an alternative to the chewing substances tested in Study 1. Informal observations suggested that saliva collection durations were lower even in comparison to chewing stimulants. Sampling with the lozenges lasted merely up to 1 min, and some participants were able to collect more than 5 ml saliva in about 45 sec. However, this raises questions about whether hormones can still diffuse quickly enough from the blood stream into saliva to equilibrate concentrations in this medium relative to blood. Findings of reduced C after stimulated saliva collection would be consistent with the possibility of lagging equilibration through an increased flow rate. However, this issue is complicated by the observation of substantially increased gonadal steroid concentrations and heterogenous distributions after stimulation with lozenges.

Correlation analyses revealed both stimulants to be suitable for C. But in total, cherry lozenge showed to be a better fit for 3 of the 4 assayed hormones. Neither lozenge was suitable for assaying E2. Nevertheless, as a result of a small sample size, single samples can cause significant differences of correlations.

Because we deemed the reductions in C and the increases in T, P, and E2 unacceptable for the valid measurement of salivary hormones, we did not include lozenges in further investigations.

Study 3: Laboratory test of stimulants

For a better understanding of potential stimulant-assay interactions, further testing was conducted under standardized laboratory conditions. The aim of this experiment was to discover whether those materials that we deemed as potentially suitable in Study 1 emit any ingredients into water as a medium and therefore simulate an increased analyte level by creating cross-reactions with the assays' antibodies.

Method

The three materials we tested were Parafilm®, gum base, and latex. Additionally, we also included sugar-free chewing gum (Orbit sugar free®, hereinafter: Orbit®), because Schultheiss (2013) has already reported strong interferences of this stimulant with antibodies in RIA assays. Therefore, it served as a positive control and a comparison tool. Assays were carried out as described for Study 1. Chewing gum was prepared by cutting a stripe of Orbit® into four pieces for each sample.

Each stimulant was added with 10 ml of deionized water into a 50 ml polypropylene tube and was stamped 15 times to mimic masticatory movements. After vigorous vortexing, the samples created in this manner were incubated for 24 h at room temperature. The next day, tubes were vigorously vortexed again and centrifuged for 20 min at 4 °C and 2000 rpm. Subsequently, the supernatant was transferred into a new tube and frozen at -80 °C until assay. 10 ml of deionized water was used as baseline and underwent the same procedure, excluding a stimulant. Every sample was assayed four times. No steroids were added to any sample.

Results

Table 4 summarizes median and range of assayed hormone concentrations in deionized water (see also Figure 1C). All comparisons were calculated with water-only samples as reference. For C, Parafilm®, z = -1.83, p = .07, and gum base, z = - 1.83, p = .07, showed a trend to lower C values, albeit in comparison to water-only samples for which the assay produced some non-zero measurements. No significant differences were detectable for latex, z = -1.46, p = .14, or Orbit®, z = 0.73, p = .47. For T, Orbit® yielded a median concentration of 33 pg/ml, z = 1.83, p = .07, while

all other stimulants produced non-detectable concentrations and were therefore set to zero. P did not significantly differ between water-only and Parafilm® samples, z = -1.46, p = .14, but increased by 200% to 400% with gum base, z = 1.83, p = .07, and latex, z = 1.83, p = .07. Samples incubated with Orbit® produced up to 1900% higher P values, z = 1.83, p = .07. For E2, in samples incubated with Parafilm® no hormone was detectable, z = - 1.83, p = .07. Gum base samples and water-based samples did not significantly differ, z = -1.46, p = .14. However, latex, z = 1.83, p = .07, and Orbit®, z = 1.83, p = .07, produced elevated levels.

Table 4

Median (range) of hormone concentrations for cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml] for water-only samples and samples incubated with particular stimulants in Study 3.

	Water only	Parafilm®	Gum base	Latex	Orbit®
Cortisol	0.04 (0.02 – 0.04)	**0.02† (0.01 -0.04)**	**0.02† (0.00 – 0.04)**	0.02 (0.02 – 0.03)	0.04 (0.01 -0.09)
Testosterone	0.00 (0.00 – 0.00)	0.00 (0.00 – 0.00)	0.00 (0.00 – 0.00)	0.00 (0.00 – 0.00)	**33.18† (23.07-43.42)**
Progesterone	1.34 (0.23 – 1.46)	0.39 (0.32 – 0.73)	**3.00† (2.30 – 4.43)**	**5.08† (3.61 -6.87)**	**25.20† (17.90 – 28.16)**
Estradiol	0.12 (0.03 -0.29)	**0.00† (0.00 – 0.00)**	0.03 (0.00 – 0.06)	**1.20† (1.10 – 1.54)**	**2.76† (2.59 – 2.96)**

Note. All differences according to Wilcoxon rank sum test in comparison to unstimulated sample.

** $p < .05$; † $p < .10$*

Discussion

Stimulants were tested in deionized water to investigate their suitability under standardized conditions. Because no interference due to real sample components was possible, in this study hormone concentration changes can only occur additively; that is, by a stimulant partly emitting molecules into the sample. If these molecules are capable of cross-reacting with assay antibodies, the assay will erroneously detect a non-zero concentration of a hormone. Such an effect was clearly in evidence with Orbit® for T, P and E2[3]. Thus, we replicated and extended similar findings reported by

[3] This effect should be considered when evaluating prior studies using sugar-free chewing gum as saliva stimulant. As long as pre- and post-experimental samples are both collected under the same

Schultheiss (2013) for P and T in water. However, our findings are also consistent with van Anders´ (2010) observation that chewing gum increases measured T levels in saliva. None of the stimulants used in saliva Studies 1 and 2 were assumed to interact with the assays. Because van Anders (2010) has shown that sugarless chewing gum increases hormone values, we did not use this stimulant in the studies with real saliva sample collection, because we wanted to prevent a carry-over effect to samples collected afterwards. Therefore, we used Orbit® in water-only experiments exclusively. Gum base as well as latex also cross-reacted with antibodies in P and E2 assays, although to a smaller extent than Orbit®. Even though for C differences between water-only and stimulant-based samples were detectable for Parafilm® and gum base, these effects were close to the detection limit and quantification at this level is noisy. Further, our study design did not allow the detection of potential subtractive effects of stimulants, as no hormones were added to water in the first place.

Summarizing our results across all studies so far, then, hormone value distributions in real samples (Studies 1 and 2) and in water samples (Study 3) suggest Parafilm® to be the best fitting material for saliva flow stimulation.

Study 4: Verification of Parafilm® as suitable saliva stimulant

conditions using sugar-free chewing gum, effects of stimulant-induced changes in hormone values might be not severe, because researchers are typically more interested in *relative* hormone changes than *absolute* levels. However, Schultheiss (2013) has shown that sugar-free chewing gum does not elevate hormone concentrations by a constant, proportional value. Rather, the stimulant-induced bias appears to vary across individuals and thus can affect a study´s results non-systematically. Therefore, results of prior studies that were conducted using sugar-free chewing gum should be considered with caution. However, this does not apply to studies examining C. C is present in saliva at much higher concentrations than T, P or E2. Chewing gum ingredients that cause cross-reactions in picogram range may play no significant role in the nanogram range of salivary C and thus have little effect on reported findings.

The previous studies suggested that Parafilm® is the most suitable material for enhancing saliva flow rate. It showed little evidence of interactions with radioimmunoassay as gauged by alterations in hormone concentrations relative to control samples. Therefore, the last study was conducted to verify its applicability as a suitable stimulant by a direct comparison with unstimulated saliva samples under typical sampling conditions with human research participants.

Method

The study took place between 2 and 4 pm. Nine subjects (6 women, 3 men, 21-38 years old, mean age: 25.9 years) were instructed to provide one unstimulated saliva sample by spitting in a 50 ml polypropylene tube and one sample stimulated by chewing Parafilm®. Order of sample collection was systematically varied to counterbalance circadian decreases in steroids. Before each sampling, participants rinsed their mouths with water and waited for 5 minutes before collecting saliva. After the collection, samples were immediately frozen at -20°C. Sample preparation, assay and statistical analysis were conducted as described previously in Study 1.

Saliva flow rate of both methods was assayed by collecting 5 ml of saliva with or without Parafilm® and measuring the required time.

Results

For C, the median (range) was 0.81 (0.37-3.74) ng/ml for unstimulated samples and 0.85 (0.51-3.34) ng/ml for Parafilm® samples. For T, medians (ranges) were 2.20 (0.75 – 42.69) pg/ml and 3.53 (0.60 – 48.29) pg/ml for unstimulated and Parafilm® samples, respectively. For P, the median (range) was 6.12 (2.07 – 50.93) pg/ml for unstimulated and 5.79 (1.33 - 50.02) pg/ml for Parafilm® samples. For E2, medians (ranges) for unstimulated and Parafilm® samples were 0.60 (0.22 – 2.20) pg/ml and 0.53 (0.12 – 1.15) pg/ml, respectively. The distribution patterns of single values were very similar, too (see Figure 1D). All differences in concentrations between the two sampling methods were not statistically significant for any of the four hormones (for C, $z = 0.18$, $p = .86$; for T, $z = 1.36$, $p = .17$; for P, $z = - 0.65$, $p = .51$; and for E2, $z = - 0.89$, $p = .37$).

Parafilm® samples were strongly correlated with unstimulated saliva concentrations for C, P and E2. For T, Spearman´s correlation was moderate. Additionally, we calculated Spearman´s correlations by combining Parafilm® samples of Study 1 and Study 4. We thus achieved a larger

sample size, yielding more reliable correlations. When analyzing the combined sample, correlations were highly significant for C, T and P. Nevertheless, for E2, the combined sample again showed a moderate correlation.

The time required to collect 5 ml of saliva dropped from 11.00 min (± 6.31 min) without stimulant to 4.40 min (± 2.18 min) with Parafilm® as stimulant (z = - 2.02, p = .04), representing a reduction of more than half of the time needed to collect a sample.

Discussion

The aim of Study 4 was to subject Parafilm® to a final test as a suitable saliva stimulant for measuring hormones by RIA. In contrast to Study 1, this time participants collected only two samples. Therefore, the collection of stimulated and unstimulated saliva was conducted within the considerably shorter period of 20-30 min and interference from circadian changes was minimal. Additionally, potential cross-contamination by other preceding stimulants could be excluded in Study 4.

When assaying C, P or E2, hormone values were almost identical for both sampling methods. Despite a similar distribution pattern, mean T increased slightly, but nonsignificantly when saliva flow was stimulated by Parafilm®. In contrast, in Studies 1 and 3 no such elevation was detectable, suggesting that the slight increase observed in the present study is likely a chance finding.

Overall, in comparison to all tested stimulants Parafilm® showed the strongest correlation for all four hormones. For C, T and P, we obtained strong and highly significant Spearman correlations of > r = .80. Although Parafilm® had only a moderate correlation of r = 0.54 for E2, this was the strongest correlation found among all tested stimulants. As mentioned in Study 1, correlations of E2 need to be interpreted with caution. The small natural E2 concentration range of 0 to 1 pg/ml observed in our samples results, particularly in relation to measurement error, in restricted variance and thus limits the potential size of correlation coefficients. More generally, Spearman ´s correlation coefficients should not be interpreted in isolation from other statistical analyses reported in this paper.

Overall, consistent with the findings emerging from the previous studies, Study 4 confirmed the overall suitability of Parafilm® as an appropriate saliva stimulant for the assessment of C, T, P, and E2.

Conclusions

The aim of this series of studies was to find a stimulant to facilitate and speed up saliva sample collection for measuring the steroids C, T, P and E2 by radioimmunoassay. Under laboratory conditions as well as in typical saliva collection settings, Parafilm® met requirements by producing near identical hormone concentrations comparable to unstimulated saliva samples. It enhanced the speed of sampling by 250 % and can be conveniently applied and stored as well. Due to its tastelessness and odor neutrality it elicits no unpleasant feeling in the mouth, and informal queries suggested that participants reported to prefer collecting samples with Parafilm® rather than without. Further, it is declared by the manufacturer as non-hazardous, is approved as a food contact material, and has already been used as saliva stimulant in other research (Kaufman & Lamster, 2002).

It is important to note that in this study Parafilm® was verified only for the specific hormones and assay kits we mentioned above. We recommend that researchers who want to use Parafilm® as a saliva stimulant in their own studies should first ensure that this method of saliva collection does not introduce bias for the specific hormones and assay methods they use in their laboratories. Furthermore, one should keep in mind that when reagents of kits change (e.g. antibodies), hitherto undocumented biases may occur and diminish the suitability of Parafilm® as a saliva stimulant. Therefore, it is important to conduct checks periodically and especially once it becomes known that a manufacturer changed assay kit ingredients.

Declarations of interest: none

References

Buttler, R. M., Bagci, E., Brand, H. S., Heijer, M. D., Blankenstein, M. A., & Heijboer, A. C. (2018). Testosterone, androstenedione, cortisol and cortisone levels in human unstimulated, stimulated and parotid saliva. *Steroids, 138*, 26-34. doi:10.1016/j.steroids.2018.05.013

Campbell, K. L., Schultheiss, O. C., & McClelland, D. C. (1999). A necessary adjustment of protocol for use of DPC coat-a-count total testosterone assay with saliva. *Clin Biochem, 32*(1), 83-85. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10074898

Dabbs, J. M., Jr. (1991). Salivary testosterone measurements: collecting, storing, and mailing saliva samples. *Physiol Behav, 49*(4), 815-817. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1881989

Justino, A. B., Teixeira, R. R., Peixoto, L. G., Jaramillo, O. L. B., & Espindola, F. S. (2017). Effect of saliva collection methods and oral hygiene on salivary biomarkers. *Scand J Clin Lab Invest, 77*(6), 415-422. doi:10.1080/00365513.2017.1334261

Kaczor-Urbanowicz, K. E., Martin Carreras-Presas, C., Aro, K., Tu, M., Garcia-Godoy, F., & Wong, D. T. (2017). Saliva diagnostics - Current views and directions. *Exp Biol Med (Maywood), 242*(5), 459-472. doi:10.1177/1535370216681550

Kaufman, E., & Lamster, I. B. (2002). The diagnostic applications of saliva--a review. *Crit Rev Oral Biol Med, 13*(2), 197-212. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12097361

Lipson, S. F., & Ellison, P. T. (1989). Development of protocols for the application of salivary steroid analysis to field conditions. *Am J Hum Biol, 1*(3), 249-255. doi:10.1002/ajhb.1310010304

Schultheiss, O. C. (2013). Effects of sugarless chewing gum as a stimulant on progesterone, cortisol, and testosterone concentrations assessed in saliva. *Int J Psychophysiol, 87*(1), 111-114. doi:10.1016/j.ijpsycho.2012.11.012

Schultheiss, O. C., Dargel, A., & Rohde, W. (2003). Implicit motives and gonadal steroid hormones: effects of menstrual cycle phase, oral contraceptive use, and relationship status. *Horm Behav, 43*(2), 293-301. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12694639

Schultheiss, O. C., Dlugash, G., & Mehta, P. H. (2019). *Hormone measurement in social neuroendocrinology: A comparison of immunoassay and mass spectrometry methods*: Abingdon: Routledge Press.

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. doi:10.1016/j.psyneuen.2009.06.009

Vining, R. F., McGinley, R. A., Maksvytis, J. J., & Ho, K. Y. (1983). Salivary cortisol: a better measure of adrenal cortical function than serum cortisol. *Ann Clin Biochem, 20 (Pt 6)*, 329-335. doi:10.1177/000456328302000601

Vining, R. F., McGinley, R. A., & Symons, R. G. (1983). Hormones in saliva: mode of entry and consequent implications for clinical interpretation. *Clin Chem, 29*(10), 1752-1756. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/6225566

Wirth, M. M., & Schultheiss, O. C. (2006). Effects of affiliation arousal (hope of closeness) and affiliation stress (fear of rejection) on progesterone and cortisol. *Horm Behav, 50*(5), 786-795. doi:10.1016/j.yhbeh.2006.08.003

Supplemental materials for:

Suitability of saliva stimulants for valid assessment of steroid hormones via radioimmunoassay

Gelena Dlugash, Oliver C. Schultheiss

Chair for Experimental Psychology, Motivation, and Affective Neuroscience

Department of Psychology

Friedrich-Alexander University,

Nägelsbachstrasse 49b, 91052, Erlangen, Germany

Study 1: Exploratory examination of saliva-flow stimulants

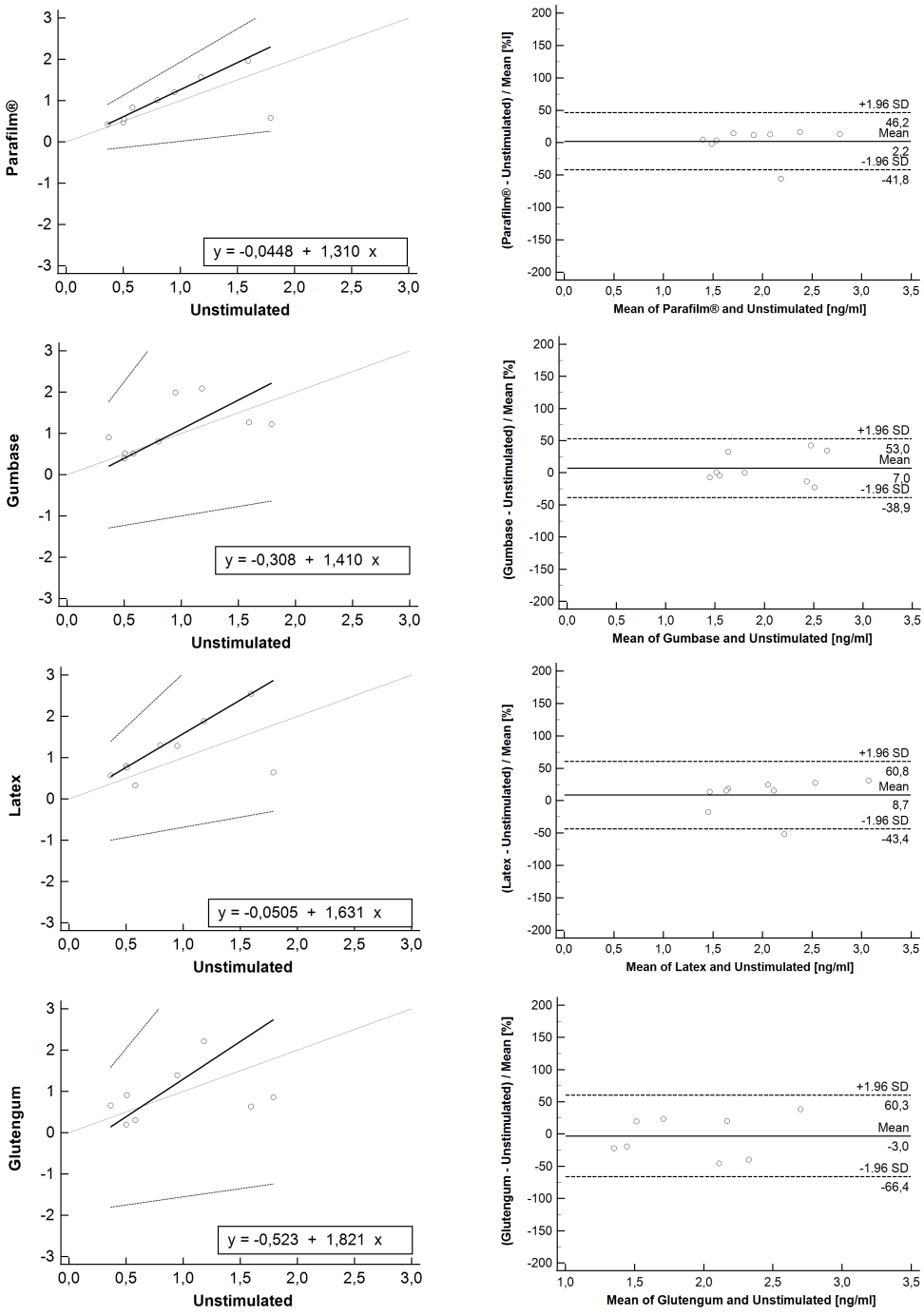


Figure S1: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **cortisol (C)** concentrations for stimulated versus unstimulated saliva samples.

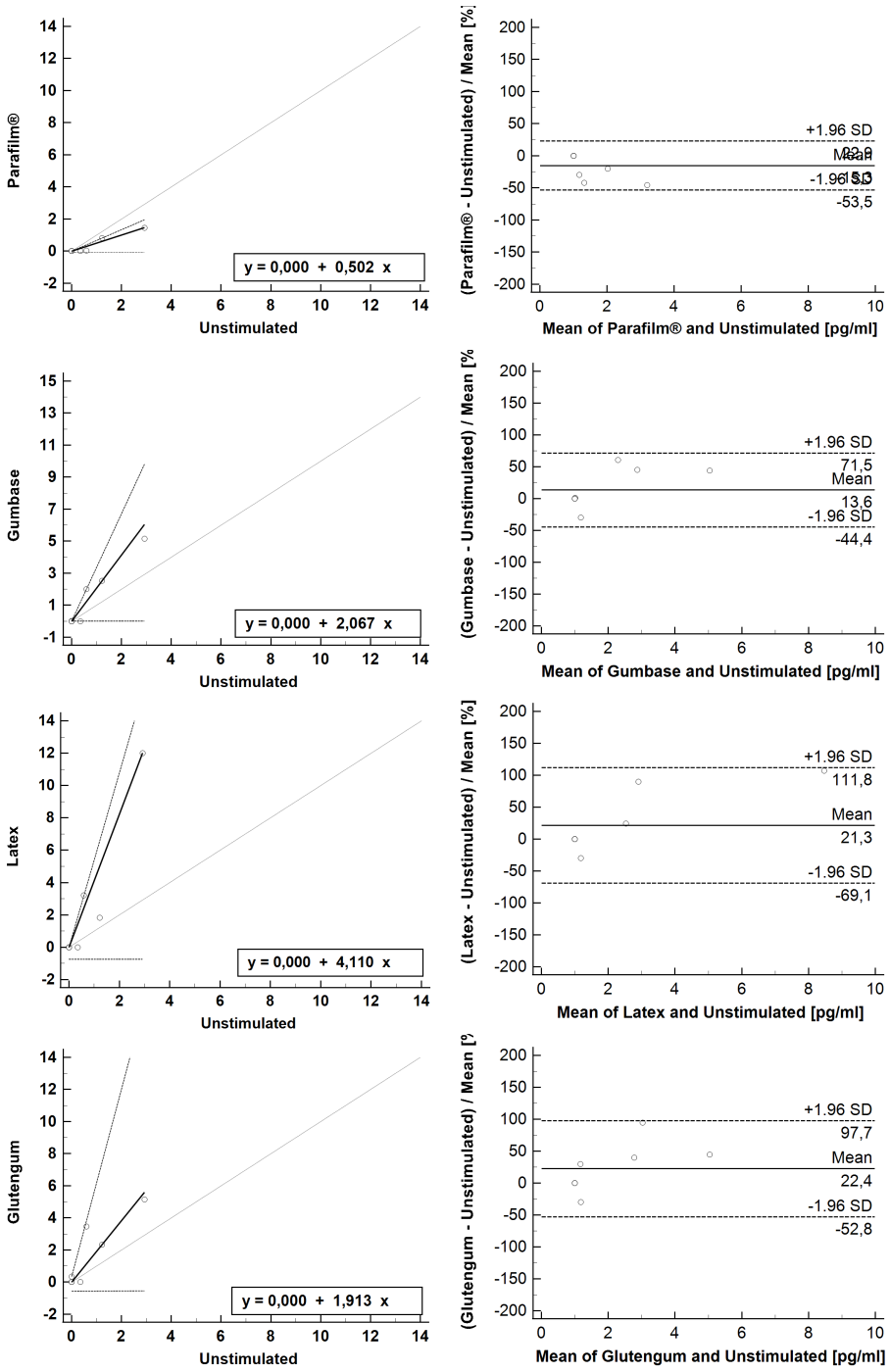


Figure S2: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **testosterone (T)** concentrations for stimulated versus unstimulated saliva samples.

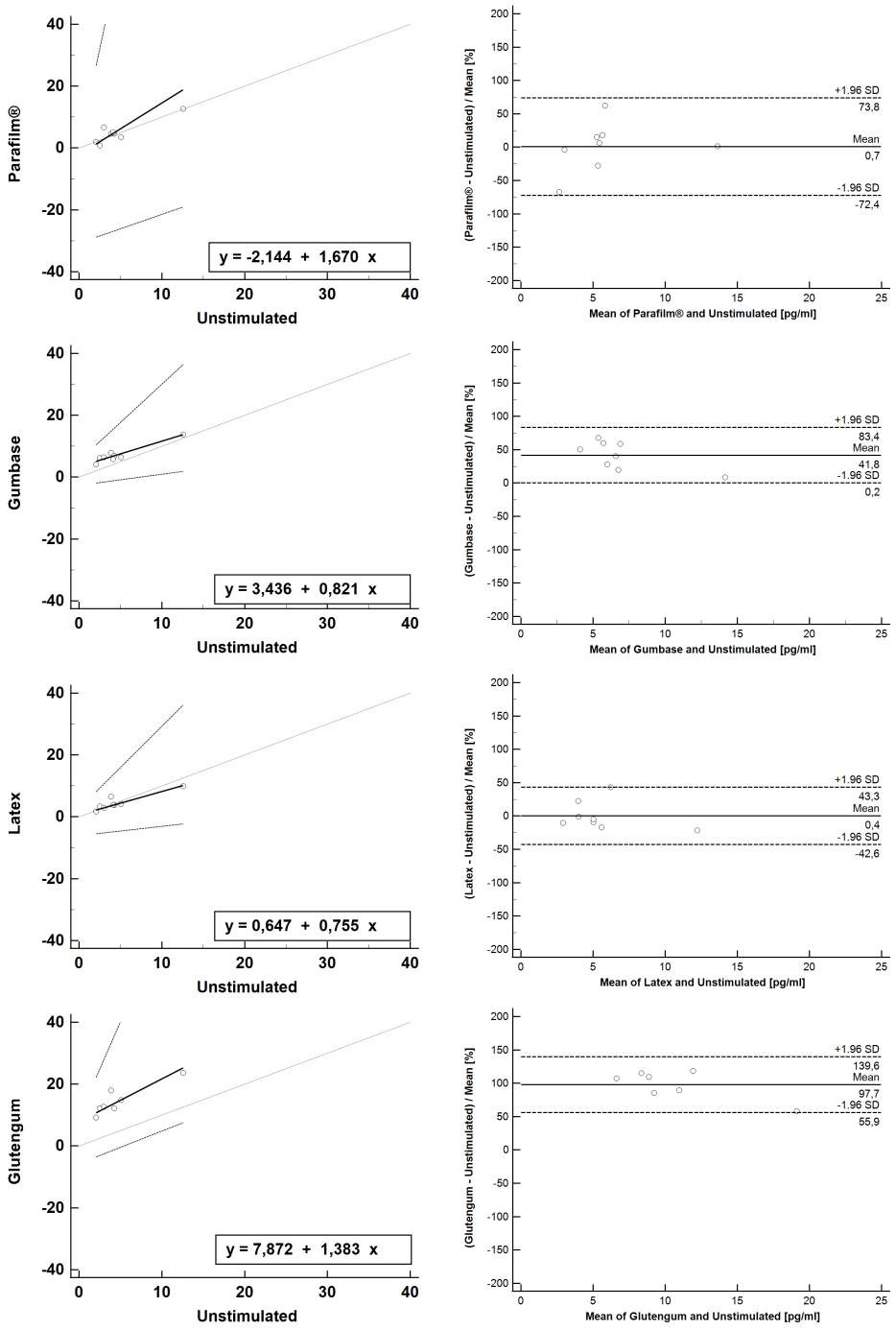


Figure S3: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **progesterone (P)** concentrations for stimulated versus unstimulated saliva samples.

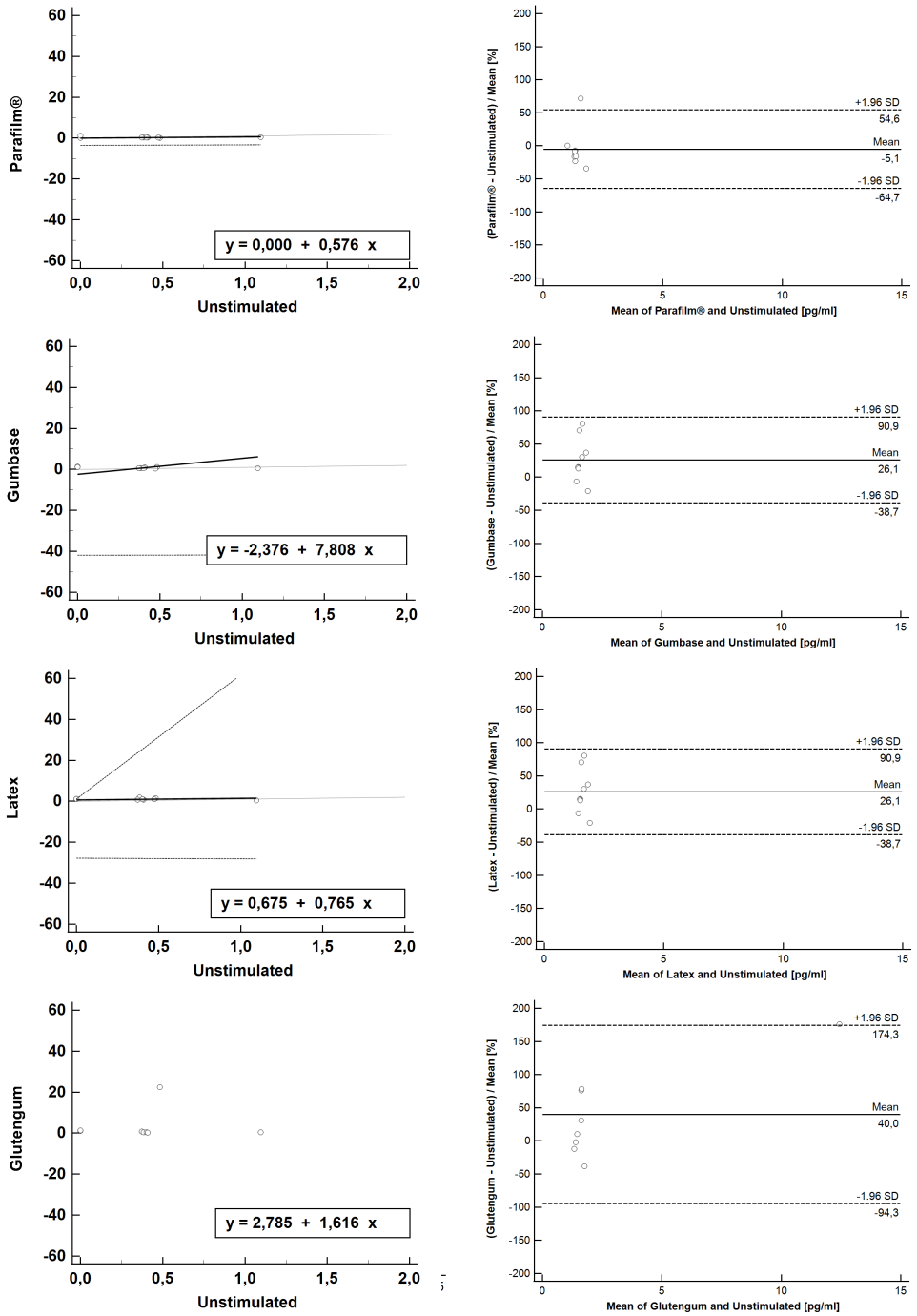


Figure S4: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **estradiol (E2)** concentrations for stimulated versus unstimulated saliva samples.

Table S1

Slopes and intercepts with 95 % confidence intervals (CI) for corresponding Passing-Bablok regressions showed in Figures S1 to S4 for cortisol, testosterone, progesterone and estradiol under various stimulated collection conditions in comparison to unstimulated saliva collection.

	Parafilm®		Gumbase		Latex		Gluten gum	
	Intercept (95% CI)	Slope (95% CI)	Intercept (95% CI)	Slope (95% CI)	Intercept (95% CI)	Slope (95% CI)	Intercept (95% CI)	Slope (95% CI)
Cortisol	-0.29 to 0.32	0.31 to 1.61	-1.46 to 0.46	0.46 to 3.58	-1.18 to 0.45	0.49 to 2.59	-1.95 to 0.36	0.39 to 3.35
Testosterone	-0.06 to 0.00	0.00 to 0.67	0.00 to 0.01	0.00 to 3.35	-0.74 to 0.00	0.00 to 5.39	-0.57 to 0.35	0.01 to 5.83
Progesterone	-30.75 to 0.98	0.93 to 12.44	-0.20 to 5.32	0.34 to 2.10	-4.43 to 2.71	0.29 to 2.67	-5.69 to 9.44	1.07 to 6.12
Estradiol	-3.56 to 0.17	0.24 to 89.88E+306	-41.88 to 1.09	---	-27.79 to 1.13	-0.29 to 60.51	---	---

Passing-Bablok regressions (Passing & Bablok, 1983) were conducted for a deeper understanding of correspondence between stimulated and unstimulated saliva collection methods. Thereby, the identity line represents the best fit of regression with an intercept of 0 and a slope of 1. For C, confidence intervals for slope and intercept of all collection methods contained the identity line, suggesting good comparability of the stimulated collection methods relative to unstimulated saliva sampling. The inclusion of identity line in confidence intervals was also found for the hormones T, P and E2 in almost all stimulants. However, for T when using Parafilm®, and for P when using gluten gum, the inclusion criterion was not met. For E2, when using gluten gum., 95% confidence intervals of intercept and slope could not be calculated.

Table S2

Mean differences (fixed bias) with 95 % confidence intervals (CI) and Pearson´s.correlations of value difference and mean hormone values (proportional bias) for corresponding Bland-Altman plots shown in Figures S1 to S4 for cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml] under various stimulated collection conditions in comparison to unstimulated saliva collection.

	mean difference	95% CI for mean difference	*r*	*p*
		Parafilm®		
Cortisol	0.04	-0.34 to 0.42	.56	.12
Testosterone	-0.31	-0.68 to 0.06	-.94	.0002
Progesterone	5.02	1.98 to 8.05	.19	.66
Estradiol	-0.07	-0.43 to 0.30	-.01	.99
		Gum base		
Cortisol	0.16	-0.26 to 0.58	.24	0.54
Testosterone	0.51	-0.18 to 1.20	.95	.0001
Progesterone	2.51	1.59 to 3.43	-.43	.28
Estradiol	0.42	-0.00 to 0.84	.00	.99
		Latex		
Cortisol	0.21	-0.26 to 0.67	.38	.31
Testosterone	1.33	-1.01 to 3.67	.99	< .0001
Progesterone	-0.13	-1.40 to 1.15	-.53	.18
Estradiol	-0.3332	-0.82 to 0.16	.30	.43
		Gluten gum		
Cortisol	-0.03	-0.62 to 0.55	.19	.65
Testosterone	0.78	-0.22 to 1.77	.85	.01
Progesterone	9.92	7.81 to 12.02	.57	.18
Estradiol	3.03	-3.38 to 9.43	1.00	<.0001

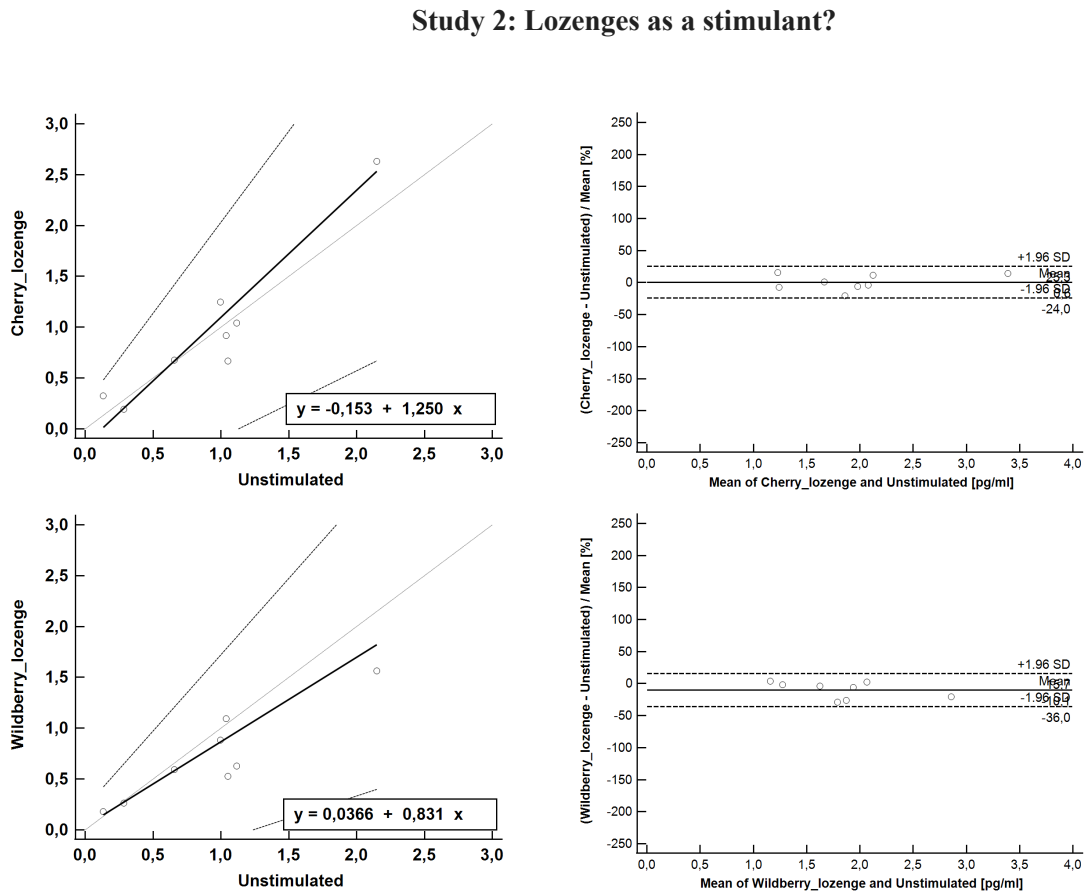


Figure S5: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **cortisol** concentrations for stimulated versus unstimulated saliva samples.

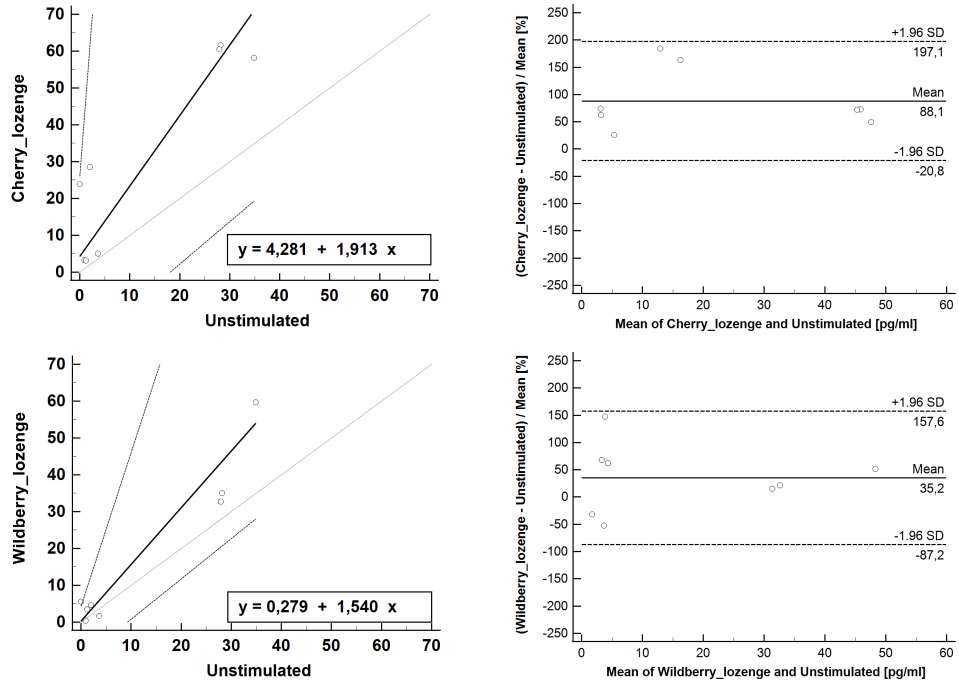


Figure S6: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **testosterone** concentrations for stimulated versus unstimulated saliva samples.

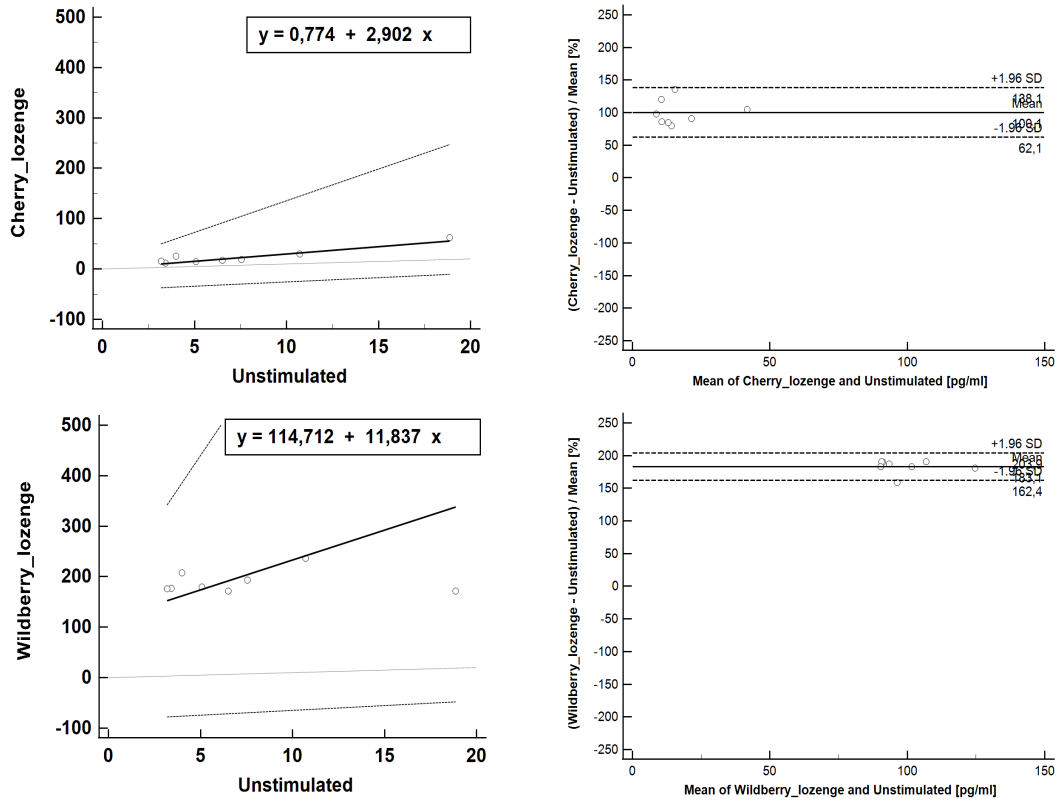


Figure S7: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **progesterone** concentrations for stimulated versus unstimulated saliva samples.

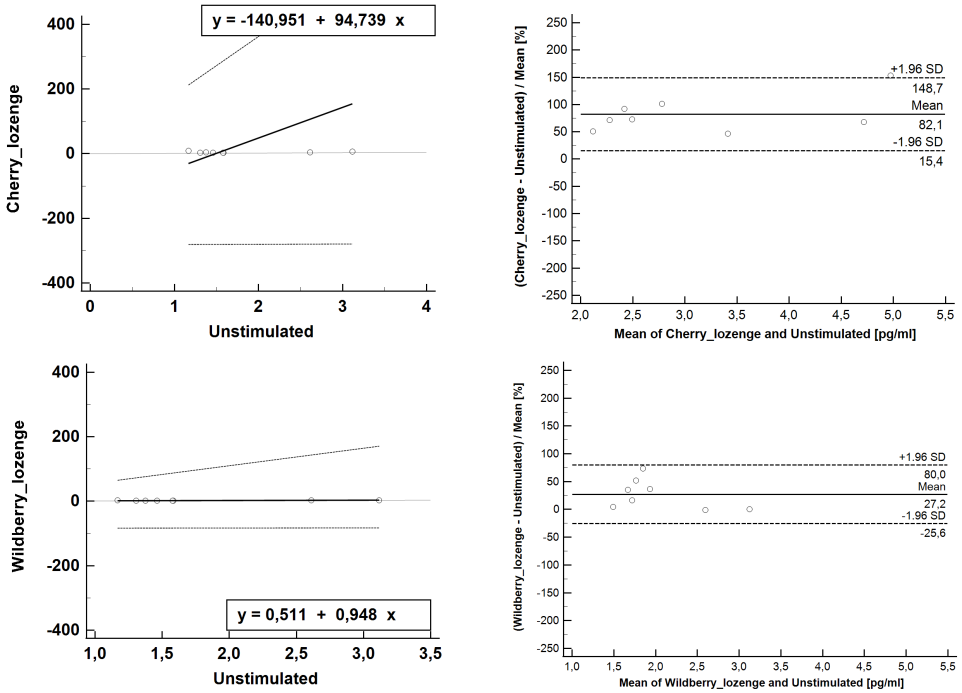


Figure S8: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **estradiol (E2)** concentrations for stimulated versus unstimulated saliva samples.

Table S3

Slopes and intercepts with 95 % confidence intervals (CI) corresponding Passing-Bablok regression showed in Figures S5 to S8 *for cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml] under various stimulated collection conditions in comparison to unstimulated saliva collection.*

	Cherry Lozenge		Wildberry Lozenge	
	Intercept (95% CI)	Slope (95% CI)	Intercept (95% CI)	Slope (95% CI)
Cortisol	-1.22 to 0.24	0.63 to 1.79	-1.05 to 0.14	0.48 to 1.50
Testosterone	-20.79 to 26.03	1.15 to 17.44	-10.16 to 4.29	1.09 to 4.17
Progesterone	-42.62 to 9.83	1.70 to 12.58	-83.80 to 170.09	1.91 to 53.99
Estradiol	-281.72 to 2.18	0.78 to 180.15	-83.58 to 1.84	0.29 to 54.25

Passing-Bablok regressions were conducted for comparison of cherry lozenge and wildberry lozenge stimulated samples to unstimulated saliva. For C and E2, confidence intervals of cherry lozenge as well as of wild berry lozenge met the inclusion criterion for slope and intercept. Nevertheless, for T and P, 95% confidence intervals of both methods did not contain the identity line and are therefore not acceptable.

Table S4

Mean differences (fixed bias) with 95 % confidence intervals (CI) and correlations of value difference and mean hormone values (proportional bias) for corresponding Bland-Altman plots showed in Figures S5 to S8 for cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml]under various stimulated collection conditions in comparison to unstimulated saliva collection.

	mean difference	95% CI for mean difference	*r*	*p*
Cherry lozenge				
Cortisol	0.03	-0.19 to 0.26	1.00	<.0001
Testosterone	18.20	6.52 to 29.88	.79	.02
Progesterone	17.15	7.36 to 26.95	.97	.0001
Estradiol	2.75	1.01 to 4.49	.79	.02
Wild berry lozenge				
Cortisol	-0.21	-0.44 to 0.01	-.62	.10
Testosterone	5.57	-1.37 to 12.52	.83	.01
Progesterone	181.89	162.51 to 201.27	.90	.002
Estradiol	0.49	0.074 to 0.90	-.44	.28

Study 4: Verification of Parafilm® as suitable saliva stimulant

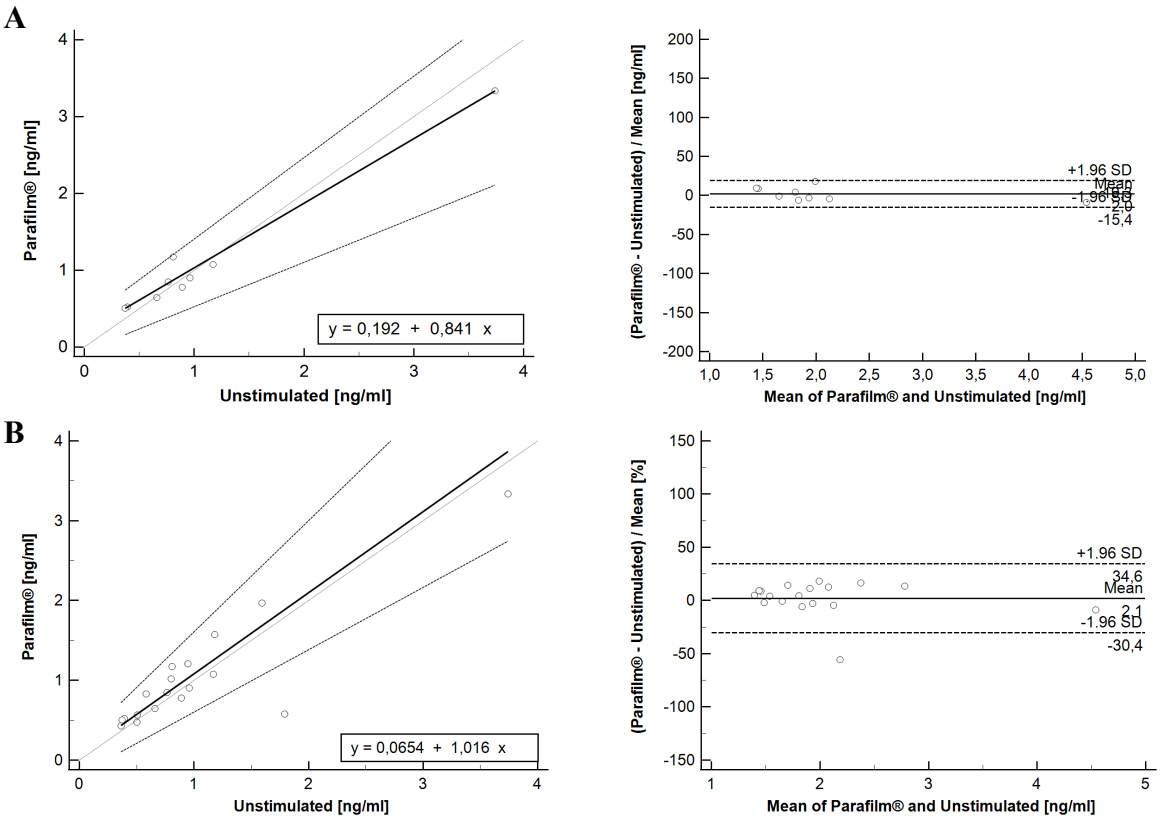


Figure S9: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **cortisol** concentrations for stimulated versus unstimulated saliva samples from Study 4 (A) and samples from Study 1 and 4 combined (B).

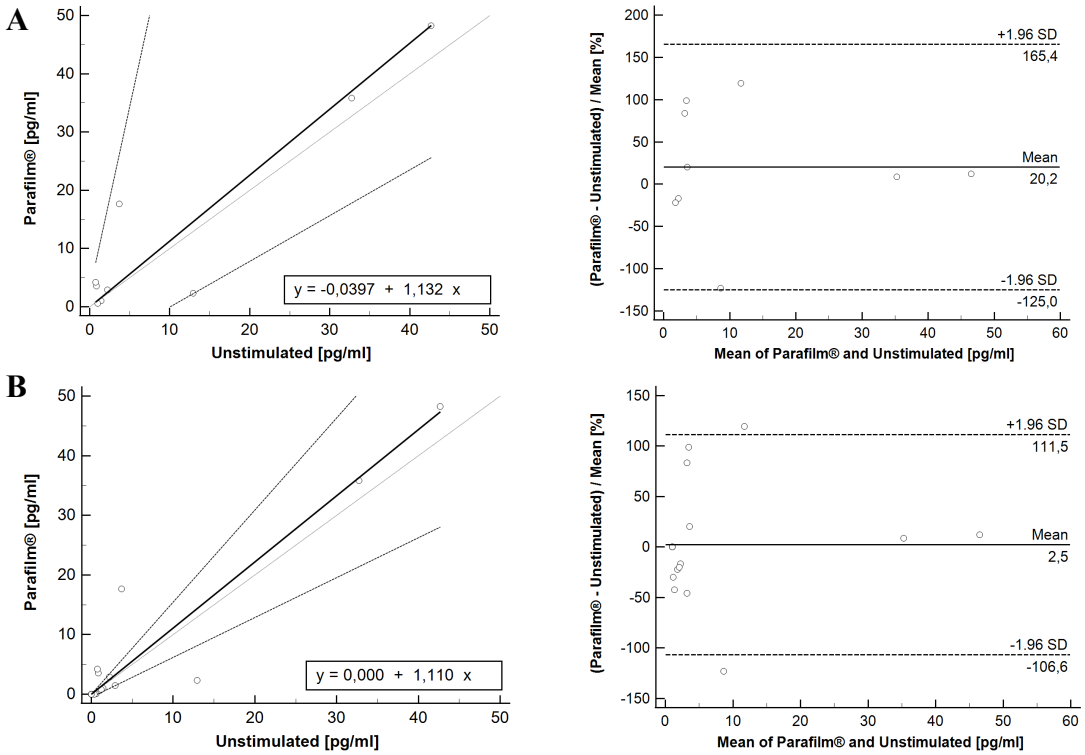


Figure S10: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **testosterone** concentrations for stimulated versus unstimulated saliva samples from Study 4 (A) and samples from Study 1 and 4 combined (B).

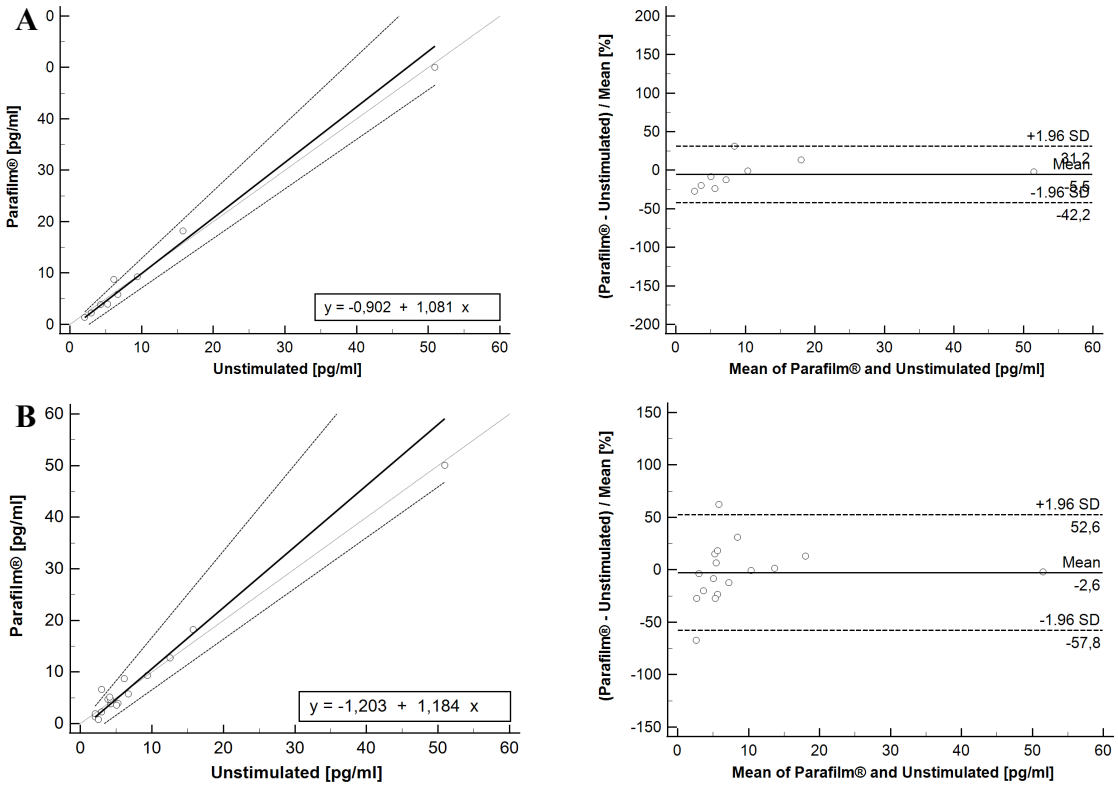


Figure S11: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **progesterone** concentrations for stimulated versus unstimulated saliva samples from Study 4 (A) and samples from Study 1 and 4 combined (B).

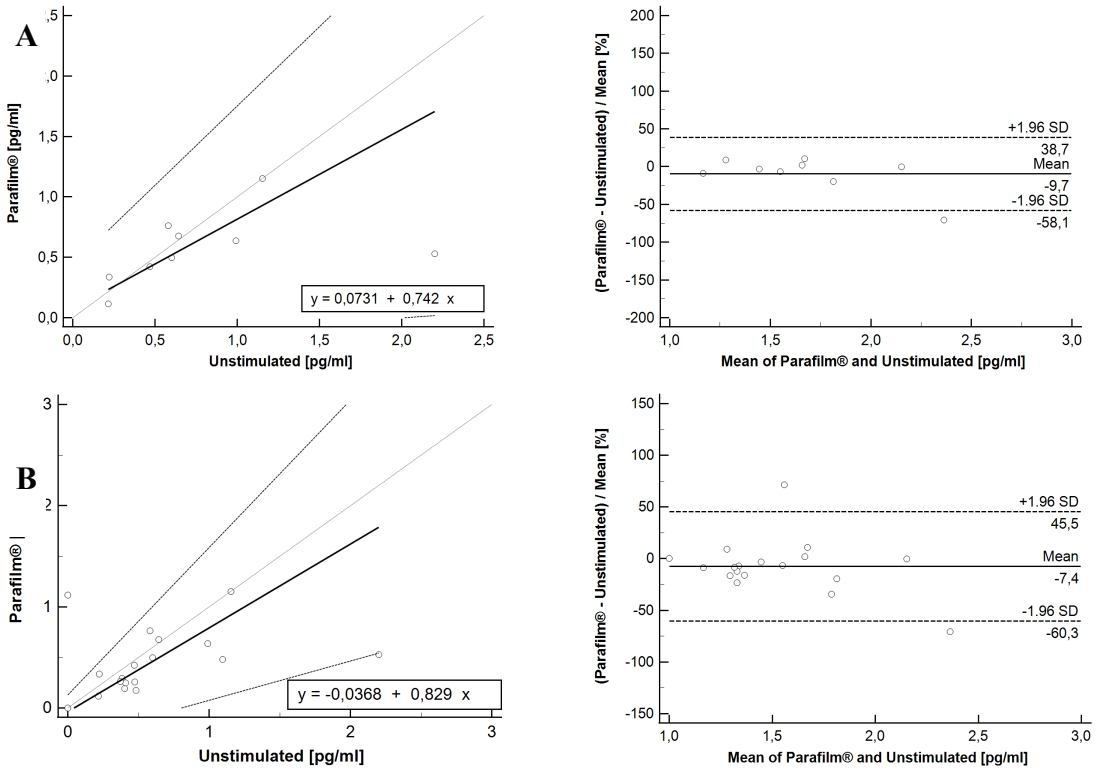


Figure S12: Passing-Bablok regression (____ regression line; ----- 95% C.I. of regression line; ..... x=y identity line) and Bland-Altman plots comparing **estradiol (E2)** concentrations for stimulated versus unstimulated saliva samples from Study 4 (A) and samples from Study 1 and 4 combined (B).

Table S5

Slopes and intercepts with 95 % confidence intervals (CI) corresponding Passing-Bablok regression showed in Figures S9 to S12 for cortisol, testosterone, progesterone and estradiol. Parameters were calculated for samples of Study 4 only and additionally for Study 1 and 4 combined.

	Parafilm® Study 4		Parafilm® Study 1+4	
	Intercept (95% CI)	Slope (95% CI)	Intercept (95% CI)	Slope (95% CI)
Cortisol	-0.16 to 0.27	0.60 to 1.56	-0.18 to 0.22	0.78 to 1.39
Testosterone	-5.67 to 2.68	0.15 to 5.88	-0.51 to 0.00	0.67 to 1.54
Progesterone	-2.94 to -0.04	0.98 to 1.35	-3.26 to 0.19	0.97 to 1.84
Estradiol	-0.26 to 0.52	0.06 to 1.46	-0.31 to 0.13	0.39 to 1.46

Passing-Bablok regression showed good correspondence between unstimulated samples and samples stimulated by Parafilm®. For all four hormones, 95% confidence intervals contained the slope of 1 and intercept of 0 and were therefore acceptable. When calculating regression with samples of Study 1 and 4 combined, we were able to achieve the same acceptable ranges as well. Therefore, Passing-Bablok regression and Bland-Altman Plot analysis confirm Parafilm® to be a suitable saliva collection stimulant.

Table S6

Mean differences (fixed bias) with 95 % confidence intervals (CI) and correlations of value difference and mean hormone values (proportional bias) for corresponding Bland-Altman plots showed in Figures S9 to S12 for cortisol [ng/ml], testosterone [pg/ml], progesterone [pg/ml] and estradiol [pg/ml]. Parameters were calculated for samples of Study 4 only and additionally for Study 1 and 4 combined.

	mean difference	95% CI for mean difference	*r*	*p*
		Study 4		
Cortisol	0.00	-0.16 to 0.17	-.73	.03
Testosterone	2.01	-2.94 to 6.96	.27	.48
Progesterone	-0.01	-1.13 to 1.12	-.01	.97
Estradiol	-0.22	-0.65 to 0.22	-.66	.05
		Study 1+ 4 combined		
Cortisol	0.02	-0.16 to 0.21	-.24	.34
Testosterone	0.85	-1.43 to 3.13	.35	.16
Progesterone	0.15	-0.63 to 0.94	-.02	.95
Estradiol	-0.14	-0.40 to 0.11	-.46	.05

Because our data contains samples that were assayed as 0 pg/ml for both methods, it was not possible to divide by the mean of zero. Therefore, Bland-Altman Plot (Bland & Altman, 1999)was created by adding a constant of +1 to all values of both methods.

References:

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res, 8*(2), 135-160. doi:10.1177/096228029900800204

Passing, H., & Bablok. (1983). A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem, 21*(11), 709-720. doi:10.1515/cclm.1983.21.11.709