Diagnosis of secondary adrenal insufficiency in patients with hypothalamic–pituitary disease: comparison between serum and salivary cortisol during the high-dose short synacthen test

Timo Deutschbein, Nicole Unger, Klaus Mann and Stephan Petersenn
Department of Endocrinology and Division of Laboratory Research, Medical Center, University of Duisburg-Essen, Hufelandstraße 55, 45122 Essen, Germany

Abstract

Objective: Accurate assessment of adrenal function is essential in patients with hypothalamic–pituitary–adrenal (HPA) disease. The measurement of salivary cortisol (SaC) instead of serum cortisol (SeC) offers several advantages, such as the determination of the free hormone. We evaluated the diagnostic value of SeC and SaC both unstimulated and during a high-dose short synacthen test (HDT) in comparison to the insulin tolerance test (ITT).


Methods: Fifty-five patients with HP A impairment and 21 healthy controls were enrolled. Samples were collected in the early morning and over 120 min during the HDT. Receiver operating characteristic analysis revealed individual thresholds for four HDT periods (0–30, 0–60, 0–90, and 0–120 min).

Results: The ITT identified 30 subjects as adrenal insufficient. With respect to the four HDT periods, sensitivity and specificity were 67–79% and 71–88% for SeC, compared with 63–72% and 72–86% for SaC. If upper and lower thresholds (with specificities > 95%) were applied, patients were diagnosed in 40–45% by SeC and in 25–31% by SaC. The combination of basal cortisol and HDT allowed a diagnosis in 47–49% (SeC) and in 42–45% (SaC) respectively.

Conclusion: We suggest the determination of basal SeC or SaC as first-line test. In comparison to the ITT, the HDT has only limited value in screening for alterations of the HPA axis. If the HDT is performed, sampling may be limited to 30 min post-synacthen, using either SeC or SaC. Due to the ease of collection and the independence of binding proteins, SaC may be preferable.

Introduction

An accurate assessment of adrenal function is essential in patients with suspected or proven disease of the hypothalamic–pituitary–adrenal (HPA) axis. Up to now, most dynamic endocrine tests are based on the analysis of serum samples. Serum cortisol (SeC), however, represents the protein bound rather than the free hormone (1). Therefore, its concentration depends on transport proteins like albumin and cortisol-binding globulin (CBG), which are often adversely affected by certain diseases, drugs, and clinical conditions (2, 3). Moreover, venipunctures are potentially stressful and might lead to an artificial increase of cortisol. By contrast, saliva samples are easy to collect, and their non-invasive obtainment is more pleasant in serial sample collections than the use of i.v. cannulas (4). Besides, salivary cortisol (SaC) is not influenced by binding proteins, reflecting the bioactive free hormone (5, 6).

Dynamic testing is often performed with the insulin tolerance test (ITT), which is widely regarded as the gold standard for dynamic stimulation of the HPA axis (7). However, this test is potentially unpleasant, associated with serious complications and sometimes even life threatening (8). In addition, its practicability is limited by numerous contraindications, like cardio- and cerebrovascular diseases, severe metabolic decompensation or former epileptic seizures (9). Thus, it is a costly and labor-intensive test with respect to the degree of medical supervision required, and its use is controversially discussed in children, the elderly and seriously ill patients. As a consequence, alternative means have been sought, for instance testing with metyrapone, CRH or ACTH. Even though the stimulation with metyrapone has been described as simple and reliable (10), this test is not only uncomfortable, but also contraindicated in patients with severe adrenal insufficiency. In addition, the determination of 11-deoxycortisol is difficult and more expensive than the measurement of cortisol. Due to its low sensitivity found in a previous study, we also do not recommend the CRH test as a first-line procedure (11).

The stimulation with exogenous ACTH (synacthen) has been described as a reliable screening method in patients with suspected adrenal impairment due to hypothalamic
or pituitary dysfunction (12). This is because sustained ACTH deprivation causes an adrenal atrophy, resulting in a reduced secretion of cortisol after the administration of synacthen. However, a too short interval between surgery and hormonal evaluation may result in false-negative tests, as both the duration and the degree of ACTH deficiency determine the severity of adrenal impairment (13).

The conventional high-dose short synacthen test (HDT) is carried out with 250 μg synacthen. Several benefits entailed its widespread use, such as the cost-effectiveness, the simplicity of testing procedures, as well as the absence of severe side effects or contraindications. A number of newer studies discussed the low-dose short synacthen test (LDT) with only 1 μg synacthen as a more physiological stimulus, thereby leading to less misinterpretation (14–17). However, an unequivocal proposal has not been made yet, and some authors still recommend that the ITT is the most accurate test (18–21).

The aim of our present study was to evaluate the diagnostic validity of serum and SaC in patients with suspected secondary adrenal insufficiency. For this, samples were collected without stimulation in the early morning, and during an i.v. HDT. Results were compared with the gold standard ITT, and reference values were established for the HDT.

Subjects and methods

Control group

Twenty-one healthy subjects constituted the control group (pertinent data are given in Table 1). All were free of endocrine disorders and none of them received medication (including systemic corticosteroids and hormonal contraceptives). The study protocol was approved by the local ethics committee and an informed consent was obtained from all participants. Control subjects were tested by HDT only.

Patients

Fifty-five patients were investigated because of suspected or proven disease of the HPA axis (pertinent data are given in Table 1). Eight subjects had a present neoplasia (4 prolactinomas, 2 somatotropic adenomas, 1 non-functioning adenoma, 1 meningioma), and 40 subjects (24 non-functioning adenomas, 8 somatotropic adenomas, 3 craniopharyngiomas, 3 prolactinomas, 2 meningiomas) were tested at least 3 months after surgical treatment (median interval: 4.8 months). Seven patients suffered from pituitary hormone impairment without detectable sellar tumors (3 secondary hypogonadisms, 3 congenital hormone insufficiencies, 1 Sheehan syndrome). At the time of hormonal evaluation, female patients were neither on contraceptives nor estrogen. Patients on chronic corticosteroid replacement therapy (generally 10–15 mg hydrocortisone per day) received their last dosage at 1400 h the day before testing, resulting in a drug restriction period of at least 18 h. No subject had to be excluded because of contraindications to insulin-induced hypoglycemia. Dynamic tests were carried out randomly, but the ITT was performed prior to the HDT in 73% of patients. The minimum and maximum intervals between both tests were 1 day and 17 days, respectively, with a median interval of 2 days.

Collection of samples

After catheterization of a superficial cubital vein and a recovery period of 15 min to avoid stress-induced bias, blood samples were directly obtained into serum tubes (Monovetten, Sarstedt, Germany), whereas saliva was collected by chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h, whereas dynamic tests were performed between 0800 and 1100 h. During HDT, serum and saliva samples were taken at 0, 30, 60, 90, and 120 min after i.v. application of 250 μg synthetic ACTH (Synacthen, Novartis). HDT evaluation involving special attention to individual cortisol peak levels focused on four sampling periods: 0–30, 0–60, 0–90, and 0–120 min post-synacthen, respectively. The ITT served as reference test, using a peak cortisol cut-off point of 500 nmol/l to distinguish between adrenal insufficient (AI) and adrenal sufficient (AS) patients (22). After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycemia were required as evidence of sufficient stress (23). Samples for blood glucose and SeC were taken at 0, 15, 30, 45, 60, 90, and 120 min. Other pituitary axes were evaluated

| Table 1 Clinical characteristics in patients and healthy controls (C). |
|--------------------|------|------|-------|-------|
| Subjects (n)       | AI   | AS   | AI + AS | C     |
| Sex                |      |      |        |       |
| M: 14, F: 16       | 30   | 25   | 55     | 21    |
| M: 12, F: 13       | 65   | 45   | 110    |       |
| Age (years)        | 45.4±2.8 | 46.4±3.2 | 45.9±2.1 | 33.8±2.4 |
| BMI (kg/m²)        | 28.2±0.7 | 29.2±1.3 | 28.6±0.7 | 23.9±0.8 |
| Additional axis deficiencies (n) | 2.4±0.2 | 1.6±0.2 | 2.0±0.2 |       |
| Median postoperative interval (months) | 5.0 | 4.5 | 4.8 |       |

Using a peak serum cortisol cut point of 500 nmol/l during the insulin tolerance test (ITT), patients were classified as adrenal insufficient (AI) or adrenal sufficient (AS). Values are means ± S.E.M.
by baseline hormone levels and dynamic endocrine tests, as required. Serum and salivary samples were measured directly or stored at \(-20^\circ\text{C}\).

**Measurement of samples**

SeC levels were determined by competitive immunoassay, using commercial kits (Advia Centaur, Bayer). The analytical sensitivity of the assay was 5.5 nmol/l. Intra-assay variations as coefficient of variation for various cortisol values were 3.7% (107.1 nmol/l), 3.1% (155.3 nmol/l), 2.9% (391.0 nmol/l), 3.8% (759.6 nmol/l), and 3.0% (1025.0 nmol/l). Inter-assay variations for the cortisol concentrations mentioned above were 5.5, 3.8, 3.1, 1.9, and 4.0%. Saliva samples were measured by a modification of the ‘GammaCoat’ RIA for cortisol (DiaSorin, Stillwater, MN, USA), decreasing the sample volume from 200 to 100 \(\mu\text{l}\). The intra-assay and inter-assay coefficients of variation were 5.4% and 15.9%, respectively.

**Statistical analysis**

GraphPad Prism 5.0 (GraphPad Software Inc., La Jolla, CA, USA) was used for statistical analysis. Results are expressed as means \(\pm\) S.E.M. The standard deviation (S.D.) was used to determine lower normal cortisol cut-offs (LNCC) for the HDT in a healthy control group (mean \(\pm\) 2S.D.). The diagnostic accuracies of both the HDT and the ITT were compared by receiver operating characteristic (ROC) analyses and its corresponding areas under the curve (AUC). A multiple stepwise linear regression analysis was carried out to determine the influence of clinical parameters like age, body mass index (BMI), and sex on basal cortisol levels. For further statistical analysis, Mann–Whitney and Kruskal–Wallis tests were performed where appropriate. A \(P<0.05\) was considered significant.

**Results**

**Comparison of cortisol levels in serum and saliva during the HDT**

**Control group** The application of synacthen resulted in a strong stimulation of both serum and SaC levels (Fig. 1, Table 2). For SeC, a mean peak time of 76 \(\pm\) 8 min was calculated, compared with a mean peak time of 110 \(\pm\) 3 min for SaC (\(P<0.005\)). Individual peak SaC values were exclusively observed within the second half of the HDT: at 90 min in 7 (33%), and at 120 min in 14 controls (67%). There was a weak, but significant, correlation between individual serum and SaC levels at each sampling time point during the HDT (\(r=0.46, P<0.0001\)). The calculated LNCC are presented in Table 2.

![Figure 1](https://www.eje-online.org)
Table 2  Mean peak cortisol (MPC) values and lower normal cortisol cut-offs (LNCC) that were individually calculated for serum and salivary cortisol as well as each period during the high-dose short synacthen test (HDT).

<table>
<thead>
<tr>
<th>Sample</th>
<th>HDT-period</th>
<th>MPC (nmol/l)</th>
<th>LNCC (nmol/l)</th>
<th>Corresponding diagnosis between ITT and LNCC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AI</td>
<td>AS</td>
<td>C</td>
<td>AI + AS</td>
</tr>
<tr>
<td>Serum</td>
<td>0–30'</td>
<td>325±35</td>
<td>559±28</td>
<td>684±38</td>
</tr>
<tr>
<td></td>
<td>0–60'</td>
<td>373±37</td>
<td>623±34</td>
<td>729±31</td>
</tr>
<tr>
<td></td>
<td>0–90'</td>
<td>402±39</td>
<td>688±48</td>
<td>753±33</td>
</tr>
<tr>
<td></td>
<td>0–120'</td>
<td>422±41</td>
<td>730±47</td>
<td>788±31</td>
</tr>
<tr>
<td>Saliva</td>
<td>0–30'</td>
<td>25.4±4.0</td>
<td>43.1±2.8</td>
<td>60.6±4.6</td>
</tr>
<tr>
<td></td>
<td>0–60'</td>
<td>34.2±5.3</td>
<td>59.0±5.1</td>
<td>86.0±5.9</td>
</tr>
<tr>
<td></td>
<td>0–90'</td>
<td>42.0±6.7</td>
<td>68.5±5.2</td>
<td>110.1±8.7</td>
</tr>
<tr>
<td></td>
<td>0–120'</td>
<td>45.1±8.3</td>
<td>76.5±6.3</td>
<td>119.0±9.9</td>
</tr>
</tbody>
</table>

MPC levels are shown for adrenal insufficient (AI) as well as adrenal sufficient (AS) patients, and healthy controls (C). Values are means ± S.E.M. The LNCC were calculated in healthy controls, subtracting 2 SDS (SD) from corresponding mean cortisol levels. Finally, the number of subjects who were correctly identified by these LNCC is given separately for all patients (AI + AS) as well as healthy controls.

Comparison of cortisol levels during ITT and HDT

In order to balance between high sensitivity and high specificity, individual ROC analyses were performed for each of the four HDT periods. That way, optimal cut-offs with the best ratio of sensitivity and specificity, upper cut-offs with high specificity (>95%) for AS, and lower cut-offs with high specificity (>95%) for AI were calculated (Table 3). For SeC, sensitivities for the optimal cut-offs ranged from 67 to 79%, while specificity varied from 71 to 88% respectively. By contrast, sensitivities from 63 to 72% were observed for SaC, compared with specificities from 72 to 86%. Similar AUC values did not indicate a relevant advantage of any HDT period. However, higher AUC values for SeC (0.82–0.83) suggested slightly better accuracy in comparison with SaC (0.75–0.77). If both upper and lower cut-offs were applied at the same time, either 40–45% or 25–31% of patients were diagnosed by the cortisol peaks in serum or saliva respectively (Table 3). The comparison between the ITT and the HDT period 0–60 min is shown in Fig. 3.

Influence of the time lag between surgery and testing procedures

Patients with former surgical intervention were further analyzed. Group A consisted of 23 patients (12 × AI, 11 × AS) who were tested 3–6 months after operation, whereas group B consisted of 17 patients (11 × AI, 6 × AS), all of whom had a postoperative interval of at least 6 months. If upper and lower cut-offs were applied, SeC levels allowed a diagnosis in 35–43% in group A and in 53–59% in group B, depending on the HDT period assessed. The corresponding values for SaC were 17–26% for group A and 47% for group B.
Evaluation of a diagnostic algorithm: combination of basal cortisol and peak cortisol levels during the HDT

Patients were not only assessed by dynamic testing procedures, but also by measurement of basal cortisol levels. Regarding SeC, ROC analysis revealed an optimal cut-off of 260 nmol/l (Sens. 73%, Spec. 72%, AUC 0.78, \(P < 0.0005\)), an upper cut-off of 382 nmol/l, and a lower cut-off of 103 nmol/l respectively. By applying the upper as well as the lower cut-off, 18 of 55 patients (33%) were diagnosed by their basal SeC levels, leaving 37 subjects for an additional evaluation. Of these subjects then investigated by upper and lower serum cut-offs defined for each HDT period, an additional eight subjects were diagnosed during HDT period 0–60 min, and nine subjects during each of the other HDT periods (0–30, 0–90, and 0–120 min) respectively. Concerning basal SaC, an optimal cut-off of 7.6 nmol/l (Sens. 53%, Spec. 83%, AUC 0.74, \(P < 0.0005\)), an upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/l were calculated. The use of both the upper and lower thresholds allowed a diagnosis in 19 of 55 patients (35%). If the remaining 36 subjects were then investigated by the HDT, the combination of upper and lower cut-offs diagnosed 3 (HDT periods 0–60 and 0–120 min), 4 (HDT period 0–30 min), and 5 additional patients (HDT period 0–90 min). In conclusion, the combination of basal cortisol and HDT allowed a diagnosis in our patients as follows: in 47–49% by SeC and in 42–45% by SaC respectively.

Side effects

During ITT, seven patients needed an additional injection of insulin due to absent hypoglycemia after the first dose. Apart from that, no clinical intervention was necessary, even though all patients showed variable hypoglycemic symptoms. As to be expected for an endogenous-like hormone, the administration of synacthen did not cause serious complications, and mild flushing was the only side effect observed.

Table 3 Receiver operating characteristics (ROC) analysis of serum and salivary cortisol levels during the high-dose short synacthen test (HDT).

For each of the four HDT periods, an optimal cut-off with the best ratio between sensitivity and specificity, an upper cut-off with high specificity for adrenal sufficiency, and a lower cut-off with high specificity for adrenal insufficiency are given. Every cut-off is listed with its corresponding sensitivity and specificity. Moreover, additional information is given for the optimal cut-offs: a) areas under the curve (AUC); b) stars that indicate statistical significance (\(* P < 0.005\), \(\dagger P < 0.001\)). Finally, the number of patients who were diagnosed by upper and lower cut-offs is mentioned.

Discussion

In comparison to common SeC analyses, the measurement of SaC offers several potential advantages. Nevertheless, comparatively little is known about its practicability after ACTH stimulation. Some studies

Figure 3 Correlation of individual peak cortisol levels during the ITT and the high-dose short synacthen test (HDT 0–60). Results are shown for (A) serum as well as (B) salivary cortisol peaks during the HDT.
investigated SaC levels in healthy controls after different amounts of synacthen (24–27). Others assessed subjects who were on chronic glucocorticoid therapy because of diseases like asthma or chronic fatigue (28, 29). Up to now, however, only two groups collected saliva samples in patients with hypothalamic–pituitary illness after stimulation with ACTH (30, 31).

In the first study, Contreras et al. (30) tested various doses of synacthen, including the conventional dose of 250 μg. Drugs were administered intramuscularly, and normal responses were defined as the lowest cortisol levels observed 30 min post-synacthen in 21 healthy subjects. That way, thresholds of 552 and 20 nmol/l were established for serum and SaC. By contrast, we calculated LNCC values as the mean-2S.D. during an i.v. HDT, obtaining similar results for saliva (18 nmol/l), but somewhat lower values for serum (332 nmol/l). When Contreras et al. used their cut-offs in 10 patients with proven secondary adrenal insufficiency (including four patients with pituitary disease) they identified all of them correctly.

In the second study, Marcus-Perlman et al. (31) evaluated SaC levels during an i.v. LDT. When 14 healthy subjects were compared with 10 patients with both pituitary disease and adrenal insufficiency, a good separation was found. With respect to the latter group, altered HPA axes were verified by pathological serum responses during the LDT. Of note, the authors also found that SeC levels were altered by elevated estrogen responses during the LDT. Therefore, the use of LNCC with its lower levels seems to bear a risk of insufficient glucocorticoid substitution in patients who require medical treatment.

In addition to calculating optimal cut-offs (with its compromise between sensitivity and specificity), we determined two alternative thresholds for the HDT: one with high specificity for AS (upper cut-off), the other with high specificity for AI (lower cut-off). If the latter cut-offs were used, a highly specific diagnosis was established in a large number of patients, thereby reducing the necessity of additional testing by more complicated procedures.

Arguably, the ITT was used as our gold standard, defining an insufficient adrenal response by peak SeC levels below 500 nmol/l. Tsatsoulis et al. (34) detected some patients with only mild adrenal impairment, who were misclassified by this cut-off. Because of its life-threatening potency, such false-negative results are quite critical. Therefore, the necessity of higher cut-offs has been discussed, mainly with the intention to raise sensitivity (7, 35). By contrast, a cut-off of 500 nmol/l has been accepted in most of the studies published to date (20).

Calculation of LNCCs represents an alternative to reference tests and their controversial cut-offs. A meta-analysis published by Dorin et al. (20) summarized 11 studies involving 340 healthy subjects. Serum LNCC (mean-2S.D.) ranged from 390 to 620 nmol/l at 30 min and from 500 to 725 nmol/l at 60 min. Similarly, both the serum and salivary peak cortisol levels extended over a wide range in our healthy controls. We could not establish any significant predictors like age, BMI, or sex suggesting a high variability of endogenous responses to synacthen. Dorin et al. (20) mentioned several other potential factors that may also have caused some variability between our individuals, like differences in HPA set points, CBG levels (at least for SeC), body composition, stress levels, the time of testing, and assay performance across runs. Our LNCC values allowed corresponding diagnoses between the HDT and the ITT in about two third of patients. Nevertheless, it is currently unknown whether the remaining one third of patients is correctly identified by LNCC or by ITT. Of note, opposite classifications were mainly found in patients considered to be AI by ITT. Therefore, the use of LNCC with its lower levels seems to bear a risk of insufficient glucocorticoid substitution in patients who require medical treatment.

In comparison to the ITT, the HDT demonstrated rather low accuracy, although our cut-offs for the HDT are in good agreement with well-established thresholds, ranging from 500 to 550 nmol/l (20). Several groups have reported similar low sensitivities and/or specificities when they compared the HDT with the ITT (16, 18, 36). Nevertheless, other authors reported high accuracy of the HDT with sensitivities as well as specificities above 90% (14, 37, 38). Despite the widespread use of the HDT, the reason for this discrepancy is currently unknown. All of the studies mentioned above included patients with various pituitary diseases in a sufficient number. However, while some studies performed testing already 6 weeks after surgery (18), others required a postoperative interval of at least 3 months (36) as in our current study. Here, we clearly demonstrated that adrenal atrophy requires sufficiently prolonged ACTH deficiency. If specific ROC cut-offs were used for HDT, patients with a shorter postoperative interval to testing were less likely to be diagnosed than patients with a longer postoperative interval. However, although the latter patients had better results, the HDT outcome was still unsatisfactory. Some studies reported a better performance of the LDT, as its dose of synacthen already exceeds the amount required to elicit an adrenal response (17, 30, 39).
In the vast majority of our patients and healthy controls, individual cortisol peaks occurred in the second half of the HDT. These peaks would have been missed during a regular HDT, because sampling is usually carried out for no longer than 60 min (20). However, the outcome of our test was not influenced by its duration, as the percentage of true assessments was almost identical for each of the four HDT periods. Therefore, serum as well as saliva sampling may be limited to 30 min post-synacthen. Of note, cortisol peaks occurred later in saliva than in serum. Similar time lags between serum and SaC peaks have been described not only for the HDT, but also for other dynamic testing procedures (5, 24, 40). Aardal-Eriksson et al. (24) regarded these lags as a result of an increased cellular uptake of free cortisol during the initial stress response.

Basal serum as well as SaC levels already allowed a correct diagnosis in about one third of patients. With regard to basal SeC, our lower cut-off 103 nmol/l corresponds well to previously described thresholds that were ranging from 80 to 110 nmol/l (9, 11, 39–41). By contrast, the upper cut-off is less standardized. While our present value of 382 nmol/l is in good accordance to Erturk et al. (41) and Hägg et al. (42) respectively, Watts et al. (43) as well as ourselves (11) published lower values, while Jones et al. (9) suggested an upper cut-off of even 500 nmol/l. This demonstrates that investigation of larger cohorts may eventually result in expanded grey zones in which further analysis is required. If both the upper and lower cut-offs were used, the determination of basal cortisol levels followed by a HDT allowed a diagnosis in almost half of the patients. Such an approach may be useful in patients with contraindications to undergoing insulin-induced hypoglycemia.

In conclusion, the HDT had only low sensitivity in comparison to the ITT, regardless of whether serum or SaC was analyzed. Thus, it is our opinion that the HDT cannot be recommended as a general screening test for adrenal insufficiency. As a consequence, we suggest the determination of basal cortisol levels in either serum or saliva as first-line test. By using both an upper cut-off with high specificity for AS and a lower cut-off with high specificity for AI, such an approach obviates dynamic testing in about one third of cases. However, if the determination of basal cortisol levels does not allow unequivocal diagnosis of either AS or AI, we propose the ITT as test of choice unless there are specific contraindications to hypoglycemia. Whenever testing with synacthen is performed, serum or saliva samples should be taken within 30 min post-synacthen, because prolonged sampling periods during the HDT did not improve its outcome. Likewise to basal cortisol, the use of highly specific upper and lower cut-offs seems to be more favorable. Due to the ease of collection and the independence of binding proteins, SaC appears to be preferable.

Declaration of interest
The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

Acknowledgements
We wish to thank Helmut Tourné for skillful technical support.

References


