Evaluation of the effectiveness of midnight serum cortisol in the diagnostic procedures for Cushing’s syndrome

Giuseppe Reimondo, Barbara Allasino, Silvia Bovio, Piero Paccotti, Alberto Angeli and Massimo Terzolo
Dipartimento di Scienze Cliniche e Biologiche, Medicina Interna I, A.S.O. San Luigi, Università di Torino, Orbassano 10043, Italy
(Correspondence should be addressed to G Reimondo; Email: g.reimondo@virgilio.it)

Abstract

Objective: It is presently unclear whether the accuracy of midnight serum cortisol (F24) in the diagnosis of Cushing’s syndrome (CS) may be replicated under usual conditions of clinical care. The aim of the present study was to assess retrospectively the effectiveness of F24 for confirming the diagnosis in a consecutive series of 106 patients, in 78 of whom a definitive diagnosis of CS was made.

Design and methods: We have compared the results of F24, urinary free cortisol (UFC) and the overnight 1 mg dexamethasone suppression test (DST) with the definitive clinical diagnosis. Receiver operating characteristic (ROC) analysis has been performed to define the best cutoff values, the sensitivity (Se) and the specificity (Sp) of the tests.

Results: The best cutoff value for F24 was 8.3 μg/dl (Se 91.8%; Sp 96.4%). The best cutoff value for the DST was 4.0 μg/dl (Se 89.2%; Sp 90.9%). The best cutoff value for UFC was 238 μg/24 h (Se 73.2%; Sp 96.3%). The area under the curve of F24 was significantly greater than that of UFC, both in the overall series (P = 0.004) and in the subgroup of patients with mild CS (P = 0.02). The differences were analyzed by means of the two-tailed student’s t-test. With the thresholds generated by the ROC analysis, UFC would have failed to achieve the correct diagnosis in a significantly higher percentage of cases than F24 (20.4% vs 7.9%; P = 0.01). The difference was analyzed by means of the chi-squared test with Yates correction.

Conclusions: The present results show that F24 has excellent effectiveness in the diagnostic procedures for CS in stressed conditions (patients studied in a hospital ward in a nonsleeping state). The test appears to be accurate also for patients with mild hypercortisolism.

Introduction

The diagnosis of Cushing’s syndrome (CS) remains a challenge for the clinical endocrinologist since many patients present with a nonspecific clinical phenotype characterized by recent weight gain, elevated blood pressure and hyperglycemia. These features are components of the metabolic syndrome and are increasingly common in the general population (1). Therefore, the imposition of screening tests that maximize sensitivity could result in a large number of patients found to be positive and submitted to further investigation, only to yield a very low number of true-positive diagnoses. Different approaches have been developed over the years, but all of them have demonstrated some limitations, and none of the proposed screening tests can detect all cases of CS (2). CS is a rare disease and the contradictory experiences reported even by referral centers have been explained by the limited number of patients studied, various laboratory techniques employed, different ways of performing the screening tests and the varying types of control subjects (such as healthy subjects, obese individuals and patients with pseudo-CS). This may explain why a wide array of cutoff values have been proposed for each test. Moreover, most published studies have assessed the performance of the screening tests in a research setting (3, 4), whereas it has not been extensively studied how the tests work in clinical practice. Assessment of urinary free cortisol (UFC) excretion in a 24-h period, which has been considered the reference standard for the diagnosis of CS, has some limitations related to either the adequacy of urine collection or the method of measurement (2, 5, 6).

The 1 mg overnight dexamethasone suppression test (DST) represents a mainstay of the biochemical confirmation of CS, although individual variations in the metabolism of the drug or pharmacologic interference may modify the results (7). The classical criterion of adequate cortisol suppression at 5 μg/dl has been recently reduced to less than 1.8 μg/dl. This value enhances sensitivity but has a limited specificity (8, 9), and it may not represent the definitive statement in this controversial field. A successive study performed in normal conditions of clinical care has shown cortisol suppression to be less than 2 μg/dl in 8% of the patients.
with proven CS (10). In recent years, measurement of cortisol concentration in saliva has also been introduced. It seems to be a promising first-line screening test, is easily accepted by patients with suspected CS and has been found to have high diagnostic sensitivity and specificity (11–14). However, validated commercial assays must become generally available to replicate these findings in larger series before the test can be considered a convincing alternative to the classical first-line screening tests (2). Alteration of the circadian cortisol rhythm is common in most patients with CS, and elevated late-night cortisol serum level has been reported to be the earliest and most sensitive marker of the disease (8). In clinical practice, it is extremely difficult to recognize patients with mild signs and symptoms, and midnight serum cortisol has been demonstrated to be very effective in this particular subset of patients (14, 15). Nevertheless, the published studies report a wide range of cutoff values, depending on the patient state at bedtime (sleeping or nonsleeping), the period since patient admission, the characteristics of the control group and whether the evaluation has been performed in a research setting or in normal conditions of clinical care.

Thus, the present study aimed to establish the diagnostic performance of midnight serum cortisol in patients with CS, evaluated under normal conditions of clinical care.

**Subjects and methods**

We performed a retrospective analysis on the effectiveness of midnight serum cortisol used to diagnose CS in a consecutive series of 106 patients admitted to our institute, a university referral center for endocrine diseases, for suspected CS from 1990 through 2003 (Table 1). Most of these subjects were referred by the outpatient service of our and other centers for diseases, for suspected CS from 1990 through 2003 (Table 1). Eighty patients met the diagnostic criteria and underwent further workup. The results of the midnight serum cortisol were not available, or turned out to be in the normal range, in 13 patients, who nevertheless underwent the same workup because of highly suggestive clinical presentation and biochemical data. Further endocrine evaluation included adrenocorticotropic hormone (ACTH) measurement in baseline conditions and after corticotropin-releasing hormone (CRH) stimulation, overnight 8 mg DST, radiologic imaging (pituitary magnetic resonance imaging (MRI) or adrenal computed tomography (CT) depending on ACTH values) and inferior petrosal sinus sampling in conjunction with CRH stimulation in patients with ACTH-dependent hypercortisolism with negative MRI and/or

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Confirmed Cushing’s syndrome</th>
<th>Excluded Cushing’s syndrome</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>48.5 ± 17.2</td>
<td>39.1 ± 16.6</td>
<td>0.01</td>
</tr>
<tr>
<td>Gender, F/M (%F)</td>
<td>55/23 (70.5)</td>
<td>21/7 (75)</td>
<td>NS</td>
</tr>
<tr>
<td>Body-mass index (kg/m²)</td>
<td>27.7 ± 5.2</td>
<td>33.5 ± 5.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Truncal obesity (%)</td>
<td>84.0</td>
<td>92.0</td>
<td>NS</td>
</tr>
<tr>
<td>Fasting glucose (mg/dl)</td>
<td>117.6 ± 56.2</td>
<td>108.9 ± 38.5</td>
<td>NS</td>
</tr>
<tr>
<td>Diabetes or impaired glucose tolerance (%)</td>
<td>37.6</td>
<td>24.0</td>
<td>NS</td>
</tr>
<tr>
<td>Systolic pressure (mmHg)</td>
<td>144.7 ± 17.0</td>
<td>130.6 ± 15.8</td>
<td>0.0004</td>
</tr>
<tr>
<td>Diastolic pressure (mmHg)</td>
<td>88.8 ± 10.9</td>
<td>81.6 ± 8.1</td>
<td>0.003</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>64.1</td>
<td>48.0</td>
<td>NS</td>
</tr>
</tbody>
</table>

Truncal obesity was defined as waist > 88 cm in women and > 102 cm in men. Hypertension was defined as systolic blood pressure greater than 140 mm Hg, or diastolic blood pressure greater than 90 mm Hg, or on antihypertensive treatment (28). Diabetes mellitus was defined if a subject was on insulin or hypoglycemic agents or when the subject’s plasma glucose was greater than 126 mg/dl. Impaired glucose tolerance was defined when the 2-h postload glucose was ≥ 140 and < 200 mg/dl (29).
inconclusive biochemical studies) (18, 19). A definitive diagnosis of CS was made in 78 patients (23 men and 55 women, median age 51 years, range 14–80 years), whereas hypercortisolism was no longer confirmed in two patients. They had slightly high UFC levels (238 and 277 μg/24 h), and the cortisol values after DST were 5.4 and 7.7 μg/dl respectively. Midnight serum cortisol was 5.7 and 6.2 μg/dl. We performed a combined dexamethasone-CRH test (20), and we observed in both patients adequate basal cortisol suppression (<1 μg/dl), without response after CRH. These two patients were re-evaluated after 6 months, and hypercortisolism was excluded. Forty-four patients had pituitary-dependent Cushing’s disease (CD), 18 patients had ACTH-independent CS (ACS), 15 patients had ectopic ACTH-dependent CS (EAS) and one patient had suspected occult EAS. For this last patient, the source of ACTH secretion was not found after extensive imaging evaluation. The diagnosis of CD was confirmed by the finding of a pituitary adenoma with positive ACTH staining at pathologic examination in 38 patients, whereas in the remaining cases the diagnosis was confirmed by the occurrence of postoperative adrenal insufficiency. In all cases of ACS and EAS, the diagnosis was histologically confirmed. Thirty-four patients were diagnosed as having mild CS by two clinicians (G R and M T), who derived the diagnosis from clinical history, physical examination and routine laboratory data, without relying on endocrine data, according to the recently validated CS severity index (CSI) (21). The patients were 8 men and 26 women with a median age of 57 years (range 14–80); etiology of mild CS was 23 CD, 8 ACS and 3 EAS (Table 2). The diagnosis of CS was excluded in 28 patients (7 men and 21 women, median age 37.5 years, range 15–70 years). Seventeen patients qualified for the metabolic syndrome according to the ATP III criteria (22), and 11/28 had polycystic ovary syndrome. They underwent clinical follow-up and repeat screening after 12 months, and in none of them was there progression of the clinical signs that prompted either suspicion of hypercortisolism or endocrine tests. Furthermore, several clinical features of these patients were ameliorated by specific therapy. The institutional review board approved the study, and all patients provided written, informed consent. The DST was performed as follows: 1 mg dexamethasone was administered orally at 2300 h, and blood samples for cortisol determination were obtained at 0800 h on the day before and the morning after dexamethasone. The blood drawing for midnight cortisol was obtained in a nonsleeping state after at least 5 h of fasting. After a careful explanation of the procedure to the patients, urine was collected for 24 h, starting at 0800 h in at least two separate occasions; urinary creatinine excretion and diuresis were measured to assess completeness of urinary collection.

Hormones were measured in-house with commercially available reagents, as previously reported (17, 18). All samples for an individual subject were batched and run in duplicate in the same assay. Intra-and interassay coefficients of variation for all hormone variables were less than 10% and 15% respectively. The results of each test were compared with the definitive diagnosis. Univariate curves of the receiver operating characteristics (ROC) were calculated to define the best cutoff values with relevant sensitivity and specificity for each test. The point on a ROC curve closest to a sensitivity of 100% and a specificity of 0% provides the best cutoff value or measure of test efficiency. The area under each ROC curve was calculated, and a comparison between the areas was performed. Values for the area can be between 0 and 1; a value of 0.5 means that the diagnostic test is no better than chance. Rates and proportions were calculated for categoric data, and means and standard deviations for continuous data; 95% confidence intervals have always been provided. Normality of data was assessed by the Kolmogorov–Smirnov test. For continuous variables, differences were analyzed by the two-tailed Student’s t-test when data were normally distributed and by the Mann-Whitney U test for

Table 2 Characteristics of the patients with mild or florid Cushing’s syndrome. Data are expressed as mean±s.d.

<table>
<thead>
<tr>
<th></th>
<th>Mild Cushing’s syndrome (n = 34)</th>
<th>Florid Cushing’s syndrome (n = 44)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>54.5±16.6</td>
<td>42.6±16.2</td>
<td>0.002</td>
</tr>
<tr>
<td>CD (n, %)</td>
<td>23 (67.6)</td>
<td>23 (52.3)</td>
<td>NS</td>
</tr>
<tr>
<td>ACS (n, %)</td>
<td>8 (23.5)</td>
<td>9 (20.5)</td>
<td>NS</td>
</tr>
<tr>
<td>EAS (n, %)</td>
<td>3 (8.8)</td>
<td>12 (27.7)</td>
<td>0.05</td>
</tr>
<tr>
<td>Gender, F/M (%F)</td>
<td>26/8 (76.4)</td>
<td>29/15 (65.9)</td>
<td>NS</td>
</tr>
<tr>
<td>Body-mass index (kg/m²)</td>
<td>28.7±4.8</td>
<td>26.6±4.7</td>
<td>NS</td>
</tr>
<tr>
<td>Truncal obesity (%)</td>
<td>85.3</td>
<td>86.3</td>
<td>NS</td>
</tr>
<tr>
<td>Fasting glucose (mg/dl)</td>
<td>114.8±54.2</td>
<td>121.3±57.4</td>
<td>NS</td>
</tr>
<tr>
<td>Diabetes or impaired glucose tolerance (%)</td>
<td>38.2</td>
<td>36.3</td>
<td>NS</td>
</tr>
<tr>
<td>Systolic pressure (mmHg)</td>
<td>143.5±15.2</td>
<td>145.2±18.1</td>
<td>NS</td>
</tr>
<tr>
<td>Diastolic pressure (mmHg)</td>
<td>87.8±8.6</td>
<td>88.7±11.9</td>
<td>NS</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>76.4</td>
<td>54.5</td>
<td>NS</td>
</tr>
</tbody>
</table>

CD: Cushing’s disease; ACS: ACTH-independent Cushing’s syndrome; EAS: ectopic ACTH-dependent Cushing’s syndrome.
nonparametric data. For categoric variables, differences were analyzed by the chi-square test with the Yates correction. Bonferroni’s adjustment for multiple comparisons was performed when appropriate. Missing data were dealt with by excluding patients from particular analyses if their files did not contain data on the required variables. When data were expressed as percent values, these referred to valid cases. Levels of statistical significance were set at $P < 0.05$. All analyses were performed by the Statistica software package (Microsoft, Tulsa, OK, USA) with the exception of the ROC curves, which were constructed by the MedCalc package (Mariakerke, Belgium).

**Results**

The operating characteristics of F24, DST and UFC in the diagnosis of CS are shown in Fig. 1. For F24, the point on the ROC curve closest to 1 corresponded to a value of cortisol of 8.3 μg/dl (sensitivity 91.8% (83.0–96.9), specificity 96.4% (81.6–99.4)). For the DST, the point on the ROC curve closest to 1 corresponded to a value of cortisol of 4 μg/dl (sensitivity 89.2% (79.1–95.5); specificity 90.9% (70.8–98.6)). For UFC, the point on the ROC curve closest to 1 corresponded to a value of cortisol urinary excretion of 238 μg/24 h (sensitivity 73.2% (61.4–83.1); specificity 96.3% (81.0–99.4)). The area under the curve of each test was significantly different from that obtained by chance ($P < 0.0001$ for each test). A 100% sensitivity was obtained with a F24 cutoff value of 3.7 μg/dl (specificity 67.9% (47.7–84.1)). The DST reached 100% sensitivity with a cutoff value of 1.5 μg/dl (specificity 68.2% (45.1–86.1)). UFC yielded 100% sensitivity with a cutoff value of 38 μg/24 h (specificity 18.5% (6.4–38.1)). The comparison between the areas under the ROC curves did not demonstrate a significant difference between either F24 and DST or UFC and DST respectively, whereas the area under the ROC curve of F24 was significantly greater than that of UFC ($0.98 (0.93–1.00) \text{ vs } 0.89 (0.81–0.95), P = 0.004$) (Fig. 2). In addition, in the subgroup of patients with mild CS, the F24 area was significantly greater than that of UFC ($0.97 (0.88–1.00) \text{ vs } 0.85 (0.73–0.93), P = 0.02$) (Fig. 2). We also evaluated the effectiveness of each test with the thresholds generated by the ROC analysis. UFC would have failed to achieve the correct diagnosis in a significantly higher number of cases than F24 (20.4% vs 7.9%, $P = 0.01$). However, no single test is able to recognize all the patients with CS in the present series.

**Discussion**

The present data demonstrate that F24 is a reliable test for diagnosis of CS. The test can recognize also patients who would have been missed by UFC or had equivocal

![Figure 1 ROC curve of F24, DST and UFC in the whole series.](https://example.com/figure1.png)
results. F24 shows greater diagnostic accuracy than UFC, and it is possibly better than the DST. Interestingly, a formal comparison between F24 and the DST has never been reported before. Strengths of the present investigation include the evaluation of a rather large series of patients with CS referred to a single center, the assessment of the diagnostic performance of the tests in conditions of clinical practice, and the establishment of the cutoff points for each test by ROC analysis. However, we should acknowledge the limitations of the retrospective nature of our study and the possibility of a selection bias since our patients were collected in a referral center by experienced endocrinologists. The referral pattern of both the patients who did qualify for CS and those who did not is identical; thus, we do not believe that this may affect significantly the outcome of the present study. However, the results of this study may not be readily conveyed to the patients seen by primary care physicians or general endocrinologists. In the recent past, some studies have provided solid demonstrations of the efficacy of F24, but the definition of the cutoff points and relative sensitivity and specificity values (4, 14, 15) is still a matter of debate. The discrepancy of the literature could be explained by various methodological issues, as reported by Findling et al. (10). Major points of concern are whether blood sampling was performed in a sleeping or awake state, the diverse types of tests that served as terms of comparison for F24, and the different control groups (healthy subjects, obese individuals and patients with pseudo-CS). Moreover, few studies have assessed the distinction between efficacy, which refers to results obtained under scientifically ideal conditions of study, and effectiveness, which refers to results achieved in normal conditions of clinical care (10, 19). In fact, the literature on F24 has produced a variety of threshold levels, ranging from 1.8 to 12.0 µg/dl, to distinguish patients with CS from controls. Newell-Price et al. (4), who initially proposed this screening test, compared 150 patients with CS with 20 healthy subjects. The blood sample was taken in a sleeping state and the desired outcome was to achieve 100% sensitivity. The aim was reached with a cutoff value of 1.8 µg/dl, but the specificity was not tested. Other authors who assessed patients in a nonsleeping state found higher threshold values that displayed both good sensitivity (90.2–96%) and high specificity (96.5–100%) (14, 15). We evaluated the diagnostic accuracy of F24 in inpatients sampled in a nonsleeping state because the test should prove useful in this setting to be recommended in clinical practice. Obviously, these conditions are stressful and yield higher F24 values in patients without CS; thus, doubts about test specificity should be raised. However, we observed fairly good sensitivity and specificity with a cutoff value of 8.3 µg/dl, which is consistent with that found by Papanicolaou et al. (15). They observed that an F24 value of 7.2 µg/dl obtained in a nonsleeping state had 97% sensitivity to distinguish between CS and pseudo-CS in the largest series up to now reported (15). An important result of our study is that a satisfactory diagnostic accuracy should be obtained also in patients with mild CS. F24 appears superior to UFC in the diagnostic approach to a patient with suspected CS. The high UFC threshold value (238 µg/24 h) found in the present series reflects the limits of UFC determination by RIA. In this respect, it is confirmed that UFC may miss patients with mild CS even if better sensitivity should be expected with HPLC (2). Concerning the DST, we have confirmed the results recently reported by Findling et al. (10), who found that 14 of 80 patients with CD (18%) suppressed serum cortisol to less than 5 µg/dl, and six (8%) showed post-dexamethasone cortisol lower than 2 µg/dl. In our study, we did indeed obtain 100% sensitivity with a threshold value of 1.5 µg/dl, but the specificity was only 68.2% at that level. The best cutoff value found with ROC analysis is 4.0 µg/dl.
which is remarkably greater than that proposed by the recent consensus statement (2). This reflects the particular composition of our series, including a high share of ACS and EAS caused by malignancies; it is known that these variants of CS are usually associated with higher post-dexamethasone cortisol levels than CD (2, 5, 23). To summarize, the present study demonstrates the effectiveness of F24 for the diagnosis of CS even if patients are studied in a hospital ward. Although the best cutoff value is expectedly higher than that found in a research setting, the test displayed good sensitivity also for mild CS. The test is significantly superior to UFC, which fails to diagnose a higher number of patients and offers no significantly better accuracy that the DST. However, no single test could recognize all the patients with CS in the present series. Since the initial diagnosis of CS may be extremely difficult in patients with mild or fluctuating cortisol overproduction, the evaluation of the combined results of both the DST and F24 could be recommended when UFC is normal. In our experience, this approach had only three false-negative results (4.8%) and no false-positive results. We acknowledge that midnight serum cortisol is burdened by the cost of inpatient admission, and some authors have recently reported that midnight salivary cortisol determination may be a feasible alternative for outpatients (11–14, 24–26). At present, however, only a limited number of laboratories can perform this analysis, and the different assays available are not comparable and provide variable diagnostic cutoffs. However, we believe that our results in measurement of F24 may be more readily applicable because they were obtained in normal conditions of clinical practice and in a modern series including many patients with mild clinical phenotype without the classic stigmata of florid hypercortisolism (27).

References


Received 2 May 2005
Accepted 2 September 2005