Technical details influence the diagnostic accuracy of the 1 μg ACTH stimulation test

Matthew Wade1, Smita Baid1, Karim Calis2, Hershel Raff3, Ninet Sinaii4 and Lynnette Nieman1

1Program in Reproductive and Adult Endocrinology, National Institute of Child Health and Human Development and 2Pharmacy Department, National Institutes of Health, Clinical Center, Building 10, CRC, 1 East, Room 1-3140, 10 Center Drive, MSC 1109, 53215 Bethesda, Maryland, USA, 3Endocrine Research Laboratory, Medical College of Wisconsin, Aurora St Luke’s Medical Center, Milwaukee, Wisconsin, USA and 4Biostatistics and Clinical Epidemiology Service, National Institutes of Health, Clinical Center, Bethesda, Maryland, USA

(Correspondence should be addressed to L Nieman; Email: niemanl@nih.gov)

Abstract

Objective: To examine the factors causing inadequate cortisol responses to the 1 μg ACTH stimulation test.

Design: Random test assignment (by age and gender) at 0800 or 1600 h.

Methods: We recruited 20 healthy adults to each of the three age groups (<40 years, 40–55 years, and >55 years; half females in each group). ACTH stimulation tests were performed in an outpatient clinic at the NIH Clinical Research Center. Plasma cortisol was measured just before, and 30 and 60 min after the administration of 1 μg ACTH (1–24). The ACTH concentration in diluted and administered solutions was measured.

Results: Twenty-five volunteers (19 at 1600 h) had a subnormal cortisol response (peak cortisol 10.4–17.5 μg/dl), using a criterion <18 μg/dl (497 nmol/l), for a specificity of 58% (confidence interval (CI) 45–71%). Afternoon testing had a significant impact on failure rates (odds ratio 6.98, CI 2.17–22.43), while gender and age did not. The stock solution contained 1 μg ACTH, but after administration through tubing it contained only 0.5–0.8 μg.

Conclusions: The high rate of abnormal results, especially in the afternoon, and loss of ACTH through tubing suggest that morning testing and minimal tubing should be adopted to avoid an inappropriate diagnosis of adrenal insufficiency. Earlier time points and standardized protocols would facilitate comparison of studies.

Introduction

Cosyntropin (ACTH 1–24, ACTH) stimulation is commonly used to diagnose adrenal insufficiency. Initially, the test was performed with a 250 μg dose (1). However, in response to concerns that patients with mild or recent secondary adrenal insufficiency may respond normally to this supraphysiologic dose, investigators developed a 1 μg ‘low dose’ test (2–4).

Based on a report that 3% of adults are prescribed high-dose corticosteroids, the population’s potential risk of secondary adrenal insufficiency is significant (5). However, there is no consensus about which diagnostic test is optimal (6–9). Two meta-analyses have agreed that the 1 μg test has better sensitivity, but only one meta-analysis found a worse specificity (79 vs 94%) for central adrenal insufficiency (7, 10).

Since a falsely positive test may lead to unnecessary life-long glucocorticoid replacement therapy, with its attendant risks, it is important to enhance the specificity. Few studies have examined the effects of time or method of administration of the agent or the influence of age or gender on the test responses.

Therefore, to clarify the optimal protocol for the 1 μg ACTH test, we evaluated factors associated with a falsely abnormal response in 60 healthy volunteers aged 22–71 years.

Methods

Volunteers and tests

We recruited healthy adults from three age groups (<40 years, 40–55 years, and >55 years, evenly split by gender) using community flyers. Exclusion criteria included uncontrolled illnesses, abnormal cell blood count or electrolytes, pregnancy, lactation, recent use of imidazole or glucocorticoid medications, or the presence of signs or symptoms of adrenal insufficiency (unintentional weight loss, nausea, fatigue, or joint pain). Well-controlled chronic illnesses (e.g. hypertension) were allowed. Subjects received $100 (USD) upon study completion. The NICHD IRB approved the study (ClinicalTrials.gov identifier NCT00156767). Subjects provided written informed consent.
An outpatient visit at the NIH Clinical Center ensured an eligibility based on medical history and measurement of blood count and chemistries and, in women, β-human chorionic gonadotropin.

Within each age group, volunteers were evenly allocated to outpatient testing at 0800 or 1600 h. At least 30 min after the insertion of an i.v. line, 1 µg ACTH 1–24 (Cortrosyn, Amphastar Pharmaceuticals, Rancho Cucamonga, CA, USA) was given, followed by 10 ml saline. Serum cortisol was measured just before the ACTH administration, and 30 and 60 min later.

**Preparation of ACTH**

In US, ACTH is available only in vials containing 250 µg sterile lyophilized powder. A 1 µg dose was prepared just before administration as follows: 2.5 ml of sterile 0.9% saline was injected into the vial, yielding a 100 µg/ml solution, and 0.1 ml was injected into a vial containing 9.9 ml of 0.9% saline, yielding a 1 µg/ml solution for administration.

**I.v. tubing**

The routine i.v. tubing used for endocrine testing in our clinic included a 2.5 cm catheter with attached polyurethane tubing and hub (ProtectIV Plus, Medex, Inc., Carlsbad, CA, USA) connected to a 20.3 cm tubing with an injection port (Baxter Healthcare Corp., Deerfield, IL, USA).

**Assays**

The NIH Department of Laboratory Medicine measured cortisol using a chemiluminescent competitive immunoassay (Siemens, Los Angeles, CA, USA). The inter-assay and intra-assay coefficients of variation (CV) are 11.1 and 7.4% respectively. The functional detectable value is 28 nmol/l. Corticosteroid-binding globulin (CBG) was measured by RIA (Quest Diagnostics Nichols Institute). The inter-assay and intra-assay CV are 22.4 and 5.8% respectively. The least detectable value is 0.1 mg/l (2 nmol/l).

**In vitro study**

To investigate a higher failure rate than anticipated failure rate (see Results), we evaluated if ACTH was diluted correctly, absorbed by the i.v. tubing, or degraded by delayed administration. This was done by measuring ACTH levels in the following samples: (a) 1 µg/ml stock solution prepared as described, (b) 1 ml aliquot of the stock solution and the saline flush after pushing through the i.v. tubing; samples (c) and (d) repeated these steps after leaving the stock solution syringe at room temperature for 60 min.

An alternative dilution technique was evaluated whereby 250 µg ACTH was added to 250 ml of 0.9% saline. Samples a–d were collected for each dilution method, flash frozen, and stored at −80 °C. A single sample for the two dilutions at steps a–d was assayed in duplicate for ACTH (below), and the results were averaged.

**ACTH RIA**

ACTH levels were measured by RIA with reagents purchased from MP Biomedicals (Orangeburg, NY, USA). The rabbit anti-porcine ACTH antibody has 100% cross-reactivity with ACTH (1–24) verified using synthetic, HPLC-purified ACTH 1–24 (Phoenix Pharmaceuticals, Belmont, CA, USA). The tracer was synthetic human ACTH–125I. Samples were assayed in duplicate at dilutions of 1:5000 and 1:10 000 using zero ACTH standard as a diluent to give an assayed ACTH result of 22–300 pg/ml (standard curve range: 11–1050 pg/ml). Intra-assay and inter-assay CV are 4.1–6.8 and 3.9–10.7% respectively.

**Data analysis**

Clinical characteristics were summarized using descriptive statistics.

A published criterion for a normal cortisol response to ACTH (497 nmol/l (18 µg/dl) or greater at 30 min) was chosen to minimize falsely abnormal results, and it was used to classify the results (11–13).

Results are presented as mean (s.d). Logistic regression analysis determined the influence of time of test, gender, and age on the ‘pass’ or ‘fail’ rate. Odds ratios (ORs) and their 95% confidence interval (CI) (14) were determined from logistic regression analyses, and specificity and Fisher’s exact 95% CIs were computed.

All statistical analyses were two-tailed, with significance defined as a P value ≤0.05. Data were analyzed using SAS system software, release 9.1 (SAS Institute, Inc., Cary, NC, USA).

**Results**

**Volunteers**

Characteristics of the subjects are presented in Table 1. Four women (13%) were taking oral contraceptives (n = 3) or estrogen-containing hormone replacement therapy (n = 1); 12 were postmenopausal. No other subjects were receiving medication known to affect cortisol or CBG levels. Some had well-controlled chronic illnesses, including hypertension (n = 4), hyperlipidemia (n = 6), diabetes (n = 2), depression/anxiety (n = 7), migraine headache (n = 2), sleep apnea (n = 1), gastroesophageal reflux disease (GERD) (n = 3), seasonal allergies (n = 6), osteopenia (n = 3), and osteoarthritis (n = 1).

All subjects completed the test without adverse events.
Cortisol responses

Surprisingly, in this group of healthy volunteers without signs or symptoms of adrenal insufficiency, 25/60 (42%, CI 29–55%) had cortisol responses <497 nmol/l (range 287–483 nmol/l; 10.4–17.5 µg/dl) (Fig. 1).

The time of testing influenced the failure rate: 19/30 (63%, CI 44–80%) volunteers tested at 1600 h and 6/30 (20%, CI 8–39%) tested at 0800 h had abnormal responses (Fig. 1). Afternoon testing had a significant impact on failure rates (OR 6.98, CI 2.17–22.43), while gender and age did not.

The absolute cortisol increment (30-min value minus 0-min value) was not affected by the time of the test (AM: 10.6 ± 3.5 µg/dl (292 ± 97 nmol/l) versus PM: 10.2 ± 3.7 µg/dl (281 ± 102 nmol/l, P = 0.74)), but was marginally significant by pass/fail status (pass: 11.1 ± 4.0 µg/dl (306 ± 110 nmol/l) versus fail: 9.4 ± 2.5 µg/dl (259 ± 69 nmol/l, P = 0.049; Fig. 2)). However, the absolute change in cortisol did not predict a pass or fail outcome for individual patients (Fig. 2). Those who failed the test had lower baseline values but similar incremental results (Fig. 2). The absolute serum cortisol concentration at 30 minutes was significantly higher with morning testing (AM: 20.5 ± 4.0 µg/dl vs PM: 16.9 ± 3.5 µg/dl, P = 0.0006).

CBG results

Basal CBG levels were higher in women taking estrogen (50.0 ± 25.5 mg/dl, n = 2) than in women not taking estrogen (34.7 ± 13.0 mg/dl, n = 17), or men (26.3 ± 5.7, n = 22) (P = 0.003). As a result, total cortisol values were significantly greater in the three women taking oral contraceptives than in the other women who were tested in the morning (20.9 ± 5.9 vs 8.5 ± 3.2 µg/dl, P = 0.002). Basal CBG values were slightly higher in subjects who passed the ACTH stimulation test than in those who failed, but this did not reach statistical significance (32.0 ± 9.8 vs 29.4 ± 14.1 mg/dl, P = 0.0524).

In vitro ACTH levels

The stock solutions contained at least 1 µg/ml ACTH (range 1.1–1.6 µg/ml). Values were similar in samples kept at room temperature for 60 min (standard dilution method: 1.6 µg/ml, alternative: 1.15 µg/ml) or processed immediately (standard: 1.25 µg/ml, alternative: 1.1 µg/ml). Values in samples pushed through the i.v. tubing were lower (immediately collected: standard 0.047 µg/ml, alternative 0.074 µg/ml; kept at room temperature: standard 0.070 µg/ml, alternative 0.082 µg/ml). Thus, among samples pushed through the i.v. tubing, 21.6–58.6% of ACTH was not recovered, corresponding to the administration of 0.5–0.8 µg of ACTH.

Discussion

This study of healthy volunteers revealed a surprisingly high rate of abnormal responses to the 1 µg ACTH test when a cortisol value of <18 µg/dl (497 nmol/l) was considered abnormal. This 58% specificity (CI 45–71%) is less than that reported by most others (mean 79%, CI 74–84%) (10). Further analysis attributed these
abnormal results to afternoon testing and loss of cosyntropin in the i.v. tubing, suggesting a need for the standardization of the testing protocol. Age and gender did not influence outcome.

In this study, afternoon testing was associated with a sevenfold increased likelihood of failing the 1 µg test. However, time of day did not explain all the abnormal responses since 6/30 (20%) failed morning testing. The incremental (delta) cortisol response was similar at both times of the day, but both the 0- and 30-min values were lower in the afternoon.

Our findings of abnormal afternoon responses to 1 µg ACTH differ from those of Park et al., who reported lower peak cortisol values in the afternoon than in the morning, but a normal response in all eight volunteers (12). The larger size of the current study may underlie these differences. Another study found no influence of afternoon testing in ten subjects (3). Perhaps, as a result of this variability, there is no published consensus on the optimal time for testing. For example, one report suggests that the test may be performed at any time of the day (15).

The overall cortisol increment above baseline was similar at each time of the day, but values for individual subjects did not predict the pass or fail outcome. As a result, although the incremental cortisol response is not influenced by the time of testing, these data do not support the use of this value for diagnostic purposes.

What might explain the 20% failure rate with morning testing? The cortisol dose–response curve is extremely steep when ACTH is delivered as a s.c. dose of 2.5–10 µg (16). Thus, possible loss of ACTH during dilution or administration might sharply reduce the cortisol response. As reported earlier, we found that short-term exposure to room temperature did not degrade ACTH (3). Although the lack of a commercial 1 µg preparation might lead to dilution errors, in this study loss of ACTH due to adherence to plastic tubing appeared to be more significant. These findings corroborate the results of Murphy who showed losses up to 70% when ACTH was passed through a plastic 30 cm scalp vein set (17), but they contrast with the report of Wood, who found that storage in a plastic tube did not decrease the delivered dose (3). The importance of minimizing or eliminating catheter length was emphasized in one recent review (18), but was not mentioned in others (7, 19). Relatively few papers explicitly cite catheter length; of those that do, both short (12) and long (14) catheters were used. We did not investigate the possibilities that the type of plastic and the length of tubing independently influence this effect. Taken together, these data suggest that the shortest possible length of tubing, or direct venous injection, should be used for ACTH administration.

Most previous studies used a 30-min endpoint based on data showing similar cortisol values at 20 and 30 min (20) or similar 30-min values after the administration of 1 and 250 µg ACTH (21), and a recent meta-analysis preferred this time point (7). However, the 30-min time point may not be optimal as up to 20% of the healthy volunteers fail the test at this time point (6, 22–24). Other studies demonstrated that the peak cortisol response in healthy subjects occurs 20–30 min after the administration of low-dose ACTH (1 µg or 500 ng/1.73 m²) (6, 12), and another study recommends the measurement of cortisol at both times (15). In another study, no responses were abnormal when both 20- and 30-min values were used. Notably, in patients with adrenal insufficiency, the times of peak total cortisol response tended to occur later than in the healthy volunteers, further supporting the use of earlier measurement intervals (12).

Differences among cortisol assays and different cut-off points might lead to either a ‘fail’ or a ‘pass’, and might underlie the variability in published diagnostic accuracy. A recent meta-analysis found the 1 µg test superior to the 250 µg test only after converting plasma cortisol levels to serum values (7). Additionally, cortisol threshold criteria have been proposed for various time points of the 1 µg (500–600 nmol/l; 18–21.9 µg/dl) and 250 µg (500–725 nmol/l; 18–26.3 µg/dl) tests (15, 20). It is not clear whether this variability results from differing assays, study populations, or ‘gold standard’ criteria for disease. Taken together, this variability suggests that it is important to validate diagnostic threshold criteria at each center.

As has been reported, higher CBG levels are associated with higher cortisol measurements, and oral estrogens can significantly increase CBG levels (25). Thus, estrogen administration is another confounding factor that might cause a falsely normal test response.

In conclusion, we identified two potential reasons for abnormal cortisol stimulation after the administration of 1 µg ACTH in healthy volunteers: afternoon testing and long plastic tubing. Previous publications suggest that the use of a single 30-min endpoint may also lead to false results. Presumably, these factors would influence falsely abnormal results in patients suspected of having adrenal insufficiency, especially those with long indwelling lines. We suggest that morning testing with 20- and 30-min time points, minimal i.v. tubing, and standardized dilution methods be adopted routinely, and that the potential influence of CBG be considered. If these requirements cannot be met, it may be prudent to consider alternative ‘gold standard’ tests for the diagnosis, such as metyrapone stimulation and insulin stimulation, in patients suspected of having a recent or mild secondary adrenal insufficiency (26). Alternatively, the standard 250 µg test works well in the majority of the patients, even in those with ACTH deficiency (27).

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 work was supported in part by the intramural program of the National Institute of Child Health and Human Development, National Institutes of Health.

Acknowledgements
We thank Nghi Huynh for his technical assistance and Bob Wesley for his helpful statistical suggestions.

References

Received 22 September 2009
Accepted 30 September 2009

www.eje-online.org