

## Corticotropin Tests for Hypothalamic-Pituitary-Adrenal Insufficiency: A Metaanalysis

Rasa Kazlauskaitė, Arthur T. Evans, Carmen V. Villabona, Tariq A. M. Abdu, Bruno Ambrosi, A. Brew Atkinson, Cheung Hei Choi, Richard N. Clayton, C. Hamish Courtney, E. Nazli Gonc, Mohamad Maghnie, Susan R. Rose, Steven G. Soule, and Karen Tordjman; Consortium for Evaluation of Corticotropin Test in Hypothalamic-Pituitary Adrenal Insufficiency

John H. Stroger Jr. Hospital of Cook County and Rush Medical College (R.K., A.T.E., C.V.V.), Chicago, Illinois 60612; Walsall Manor Hospital National Health Service Trust (T.A.M.A.), Walsall, WS2 9PS United Kingdom; University of Milan and Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Policlinico San Donato (B.A.), San Donato Milanese 20097, Italy; Regional Centre for Endocrinology and Diabetes (A.B.A., C.H.C.), Royal Victoria Hospital, Belfast BT12 6BA, United Kingdom; Queen Elizabeth Hospital (C.H.C.), Hong Kong Special Administrative Region; Keele University and North Staffordshire Hospital (R.N.C.), Stoke-on-Trent ST4 7QB, United Kingdom; Hacettepe University (E.N.G.), 06100 Ankara, Turkey; IRCCS G. Gaslini (M.M.), University of Genova, 5-16147 Genova, Italy; Cincinnati Children's Hospital Medical Center and University of Cincinnati (S.R.R.), Cincinnati, Ohio 45229; Christchurch Hospital (S.G.S.), Christchurch, New Zealand; and Institute of Endocrinology, Metabolism, and Hypertension (K.T.), Tel Aviv Sourasky Medical Center, Tel Aviv 64239, Israel

**Context:** The diagnostic value of tests for detecting hypothalamic-pituitary adrenal insufficiency (HPAI) is controversial.

**Objective:** Our objective was to compare standard-dose and low-dose corticotropin tests for diagnosing HPAI.

**Data Sources:** We searched the PubMed database from 1966–2006 for studies reporting diagnostic value of standard-dose or low-dose corticotropin tests, with patient-level data obtained from original investigators.

**Study Selection:** Eligible studies had more than 10 patients. All subjects were evaluated because of suspicion for chronic HPAI, and patient-level data were available. We excluded studies with no accepted reference standard for HPAI (insulin hypoglycemia or metyrapone test) if test subjects were in the intensive care unit or if only normal healthy subjects were used as controls.

**Data Extraction:** We constructed receiver operator characteristic (ROC) curves using patient-level data from each study and then merged results to create summary ROC curves, adjusting for study size and cortisol assay method. Diagnostic value of tests was measured by calculating area under the ROC curve (AUC) and likelihood ratios.

**Data Synthesis:** Patient-level data from 13 of 23 studies (57%; 679 subjects) were included in the metaanalysis. The AUC were as follows: low-dose corticotropin test, 0.92 (95% confidence interval 0.89–0.94), and standard-dose corticotropin test, 0.79 (95% confidence interval 0.74–0.84). Among patients with paired data (seven studies, 254 subjects), diagnostic value of low-dose corticotropin test was superior to standard-dose test (AUC 0.94 and 0.85, respectively;  $P < 0.001$ ).

**Conclusions:** Low-dose corticotropin test was superior to standard-dose test for diagnosing chronic HPAI, although it has technical limitations. (*J Clin Endocrinol Metab* 93: 4245–4253, 2008)

Chronic hypothalamic-pituitary-adrenal (HPA) insufficiency (HPAI) is characterized by impaired ACTH secretion as a result of disease or injury to the pituitary or hypothalamus or from prolonged exogenous glucocorticoid administration that exceeds physiological doses.

The reference tests for establishing the integrity of the HPA axis require assessing the response to either a strong stimulus (e.g. insulin-induced hypoglycemia) or an interruption of the negative feedback from cortisol (overnight metyrapone test). However, these reference tests have major drawbacks. The in-

0021-972X/08/\$15.00/0

Printed in U.S.A.

Copyright © 2008 by The Endocrine Society

doi: 10.1210/jc.2008-0710 Received March 31, 2008. Accepted August 4, 2008.

First Published Online August 12, 2008

Abbreviations: AUC, Area under the curve; CI, confidence interval; HPA, hypothalamic-pituitary-adrenal; HPAI, hypothalamic-pituitary adrenal insufficiency; LDCT, low-dose corticotropin stimulation test; LR, likelihood ratio; ROC, receiver-operator characteristic; SDCT, standard-dose corticotropin stimulation test.

sulin tolerance test is contraindicated in the elderly and those with a history of seizures or cardiovascular disease and requires continuous physician supervision to monitor for serious adrenergic or neurological symptoms (1). The overnight metyrapone test carries a risk of adrenal crisis, and errors can occur from other drugs affecting metyrapone clearance (2). Thus, there is a great clinical demand for alternative tests that are quicker, cheaper, and safer.

The rationale for using the corticotropin analog stimulation test is the assumption that in chronic endogenous corticotropin deficiency, acute responsiveness of the adrenal zona fasciculata is diminished and fails to mount an adequate cortisol response (3). We examined the published literature for evidence on two corticotropin analog stimulation tests using either a standard (250  $\mu$ g) or low (1  $\mu$ g) dose of the corticotropin analog. The primary objective of our quantitative metaanalysis was to compare the performance of the standard-dose corticotropin stimulation test (SDCT) and low-dose corticotropin stimulation test (LDCT) in diagnosing HPAI (defined by results of the insulin tolerance test or overnight metyrapone test). The second objective was to determine how best to incorporate the tests in clinical practice, especially in relation to the basal cortisol level.

## Materials and Methods

### Data collection

We searched the PubMed (www.PubMed.gov) database from 1966–2006 for articles with keywords adrenal insufficiency and diagnosis and limited the search to human studies published in the English language. We selected studies with at least 10 subjects with suspected HPAI and required that the disease be verified with either the insulin tolerance test or metyrapone test. We then tried to contact the principal investigator of each relevant study to request their patient-level data on the following variables: results of the integrated HPA axis reference test, baseline cortisol value, and cortisol values after the SDCT and LDCT.

### Population characteristics

To be eligible for inclusion, adult and pediatric subjects had to be suspected of HPAI from disease or injury to the pituitary or hypothalamus or from prolonged exogenous glucocorticoid administration in supraphysiological doses. Patients had to be affected by hypothalamic-pituitary disease for at least 4 wk to exclude acute hypothalamic or pituitary disorders. A normal sleep-wake cycle was required (or assumed, if no information). We did not include data from studies done in critical care settings.

In actual clinical practice, testing for HPAI is performed when there is some suspicion for HPAI. We have therefore investigated the performance of the tests in this at-risk population to avoid the problem of spectrum bias (4), which occurs when tests are evaluated among patients who are different from the ones who will be tested in practice. Thus, we excluded from analysis those subjects who were described as normal healthy control subjects (those recruited as healthy volunteers and not due to suspected pituitary disease based on signs/symptoms or imaging).

The most common reason for testing (43%) was because of a pituitary macroadenoma (before or after surgical or radiation treatment). The second most common reason was treatment of other intracranial tumors (23%); 14% of patients were tested due to pituitary disease other than pituitary macroadenoma; 13% were tested due to other pituitary hormone deficiencies (GH or panhypopituitarism). Prolonged supraphysiological glucocorticoid treatment prompted HPAI evaluation in 7% of total patients, but in only eight patients were there paired data on LDCT and SDCT (5–7). The only study (5) that tested patients solely due to suspected glucocorticoid-induced HPAI was analyzed separately [see Fig. 1 in Kane *et al.* (6)], but this study did not have paired SDCT and LDCT data.

### Reference test

The diagnosis of HPAI was based on an abnormal response to one of the two reference standards for evaluating the integrity of the HPA axis: insulin tolerance test or overnight metyrapone test. We relied on individual study investigators to correctly dichotomize the reference test results into HPAI present or absent.

### Cortisol assay

Cortisol assays are not standardized and vary across hospitals and studies (8–10). The cortisol assay methods used in various studies included RIA or commercially available immunometric methods. To adjust for cortisol assay variability, first we excluded five studies that used the fluorescence polarization assay, which is less specific than more modern methods (8) and has never been used to evaluate LDCT. Second, because plasma cortisol is known to be consistently higher than serum cortisol levels, we converted plasma values (used in two studies) (6, 11) to their expected serum value by multiplying by the published correction factor of 0.877 (12).

We also investigated whether the results are sensitive to assay type by reassessing the SDCT and LDCT comparisons using the same thresholds for all studies (weighted mean cortisol thresholds) rather than individual study cortisol thresholds (see statistical analysis below).

### SDCT

One of the two available synthetic corticotropin analogs, cosyntropin (Cortrosyn, Amphastar Pharmaceuticals, Inc.) or tetracosactrin (Synacthen, Novartis Pharma, Switzerland), was administered iv at a dose of 250  $\mu$ g, and serum cortisol levels were obtained at baseline and at least once after injection (most commonly at 30 or 60 min).

A dose of 250  $\mu$ g (0.25 mg) of cosyntropin or tetracosactrin is equivalent to 25 USP units of corticotropin, indicating the equivalence of Synacthen and Cortrosyn formulations. For brevity, we use the term corticotropin for both analogs, while acknowledging that Synacthen and Cortrosyn are synthetic corticotropin analogs, different from the native ACTH molecule.

### LDCT

The low-dose test was performed in the morning with patients fasting. One of the two synthetic corticotropin analogs (cosyntropin or tetracosactrin) was administered intravenously in a 1- $\mu$ g dose, after being prepared using the method of Dickstein and colleagues (13). Serum cortisol was measured at baseline and at 30 min after injection (except for one study, which measured it at 20 min). We excluded studies that used a different dose or different protocol.

### Basal cortisol

All studies measured serum cortisol between 0800 and 1000 h after an overnight fast (basal cortisol). The database from one study (11) did not provide basal cortisol.

### Statistical analysis

We conducted data analysis using Stata statistical software, version 10.0 (StataCorp LP, College Station, TX).

To compare performance of LDCT and SDCT (and basal cortisol), we used receiver-operator-characteristic (ROC) curve analysis. From each study's data, we calculated the area under the ROC curve (AUC) with 95% confidence intervals (CI) (Fig. 1). The same methods were used to compare test performance at different time points for measuring stimulated cortisol after corticotropin injection (usually, 30 or 60 min).

For each study, dependent on data availability, we conducted paired analyses of the AUC for the three test pairs, SDCT and LDCT, basal and SDCT, and basal and LDCT, to evaluate their relative performance in predicting results of the integrated HPA axis testing (reference standard). The AUC is independent of measurement units; thus, overall comparison of SDCT and LDCT is not influenced by type of cortisol assay.

To combine data across studies, we categorized cortisol responses into three intervals (high, indeterminate, and low likelihood of HPAI),

using standard criteria applied consistently to all studies. Two thresholds were defined for each test: the threshold below which cortisol values had high likelihood of HPAI [likelihood ratio (LR)  $> 9$ ; rule-in threshold] and the threshold above which cortisol values had a low likelihood of HPAI (LR  $< 0.15$ ; rule-out threshold). Cortisol values between these thresholds (LR = 0.15–9) defined the interval with indeterminate likelihood of HPAI.

LR were calculated as the ratio of two probabilities: the probability of the test result among patients with HPAI divided by the probability of the same test result among patients without HPAI.

The use of categorized cortisol responses recalibrates individual study results to common metric, therefore permitting paired comparisons of the ROC areas for SDCT, LDCT, and basal cortisol, stratified by study and cortisol assay.

### Comparison of corticotropin stimulation tests using summary cortisol thresholds

We combined results across studies by calculating a weighted average of the cortisol values defining the rule-in threshold and the rule-out threshold for each test (Table 1). The weights were based on study sample size. Using these new summary thresholds, regardless of cortisol assay method, we recategorized the data from all studies and combined results in summary ROC curves. This analysis was performed to determine whether the results of SDCT and LDCT performance comparisons were sensitive to cortisol assay type (Fig. 2B).

### Corticotropin-stimulated cortisol in subjects with indeterminate basal cortisol

To address the research goal of defining an optimal testing strategy, we tested the sequential testing strategy of first measuring basal cortisol and then measuring a stimulated cortisol in only those subjects with an intermediate basal value. We used paired ROC curve analyses, adjusted for study size and cortisol assay (as described above), to compare LDCT and SDCT in subjects with indeterminate basal cortisol results (LR = 0.15–9 for HPAI).

### Optimal testing strategy

Based on results of metaanalyses, we derived an optimal testing strategy algorithm (Fig. 3). The basal cortisol and LDCT summary thresholds described in Fig. 3 were based on the mean cortisol values weighted by study sample size (last row of Table 1).

We used summary thresholds for the two sequential tests to define five subgroups: 1) low basal cortisol, 2) high basal cortisol, 3) indeterminate basal cortisol and low stimulated cortisol, 4) indeterminate cortisol and indeterminate stimulated cortisol, and 5) indeterminate basal cortisol and high stimulated cortisol. We calculated the expected probability of HPAI, with 95% CI, within each of the five subgroups.

## Results

We were able to analyze patient-level data from 13 published studies, with seven studies having paired data comparing results of the LDCT and SDCT in the same patients (Table 1). The prevalence of HPAI in the study samples ranged from 18–58%, with a mean of 33% (Table 1).

Cortisol testing methods varied from individual RIA kits (6, 7, 11, 14–17) to immunometric test kits from various manufacturers (5, 18–22). The lack of a standard cortisol assay method (8–10) explains some of the variability in diagnostic cortisol thresholds reported across studies.

### SDCT

After standard-dose corticotropin stimulation, there was variability across studies in the optimal timing for measuring cortisol response; however, in no study was there a statistically significant difference in diagnostic discrimination at 30 or 60 min or at peak response. In six studies, there was a trend favoring 60-min measurements; in six studies, the peak cortisol appeared best; and in four studies, 30-min testing appeared best. In our analyses of standard-dose testing, we used 30-min serum cortisol values and used peak cortisol in the studies where 30-min values were not available (7, 15, 17).

Combining results from 10 studies of standard-dose stimulation (346 subjects), 30-min cortisol values of less than 16  $\mu\text{g/dl}$  (440 nmol/liter) were highly predictive of HPAI. Values greater than 30  $\mu\text{g/dl}$  (833 nmol/liter) best predicted a normal reference test result (ruling out HPAI). Intermediate values, 16–30  $\mu\text{g/dl}$ , were diagnostically indeterminate. The AUC for these categorized test results was 0.82 (95% CI = 0.78–0.86).

In paired analyses of the standard-dose stimulation test and the basal cortisol test (nine studies, 302 subjects), diagnostic discrimination was similar; AUC was 0.79 for the SDCT and 0.80 for the basal cortisol test ( $P = 0.45$ ).

### LDCT

After low-dose corticotropin stimulation, 30-min cortisol measurements had superior test characteristics in three studies (18, 19, 21), whereas two studies found 20-min values to be best (although not statistically different from 30-min values) (20, 22). In our analyses of LDCT, we used 30-min cortisol values (5, 11, 14, 16, 18–22) or, if not available, the 20-min (17) or peak values (7).

A metaanalysis of the 11 studies (589 subjects) using the LDCT found that values less than 16  $\mu\text{g/dl}$  (440 nmol/liter) best predicted HPAI, whereas values greater than 22  $\mu\text{g/dl}$  (600 nmol/liter) best predicted a normal reference test result. The AUC for these diagnostic thresholds was 0.94 (95% CI = 0.90–0.94).

In paired analyses (10 studies, 545 subjects), the LDCT was superior to the basal cortisol test in overall diagnostic discrimination; AUC was 0.80 for basal cortisol and 0.92 for LDCT ( $P = 0.01$ ).

### Comparison of LDCT and SDCT

In the seven studies with paired 30-min cortisol data for both tests (254 subjects), the LDCT had a larger AUC compared with the SDCT, 0.94 vs. 0.85 ( $P < 0.001$ ), after adjusting for type of cortisol assay (Fig. 2C). Excluding the eight patients with glucocorticoid-induced HPAI (7) did not change the results.

Among subjects with indeterminate basal cortisol values (5–13  $\mu\text{g/dl}$ , or 138–365 nmol/liter) (paired data from six studies), the LDCT was more discriminating than the SDCT in diagnosing HPAI; AUC was 0.94 for low-dose and 0.75 for standard-dose test ( $P < 0.001$ ).

### Basal cortisol test

In a metaanalysis of 12 studies (635 subjects), a basal cortisol less than 5  $\mu\text{g/dl}$  (138 nmol/liter) best predicted HPAI, whereas values greater than 13  $\mu\text{g/dl}$  (365 nmol/liter) best predicted normal HPA axis testing. The AUC was 0.79 (95% CI = 0.75–0.82).

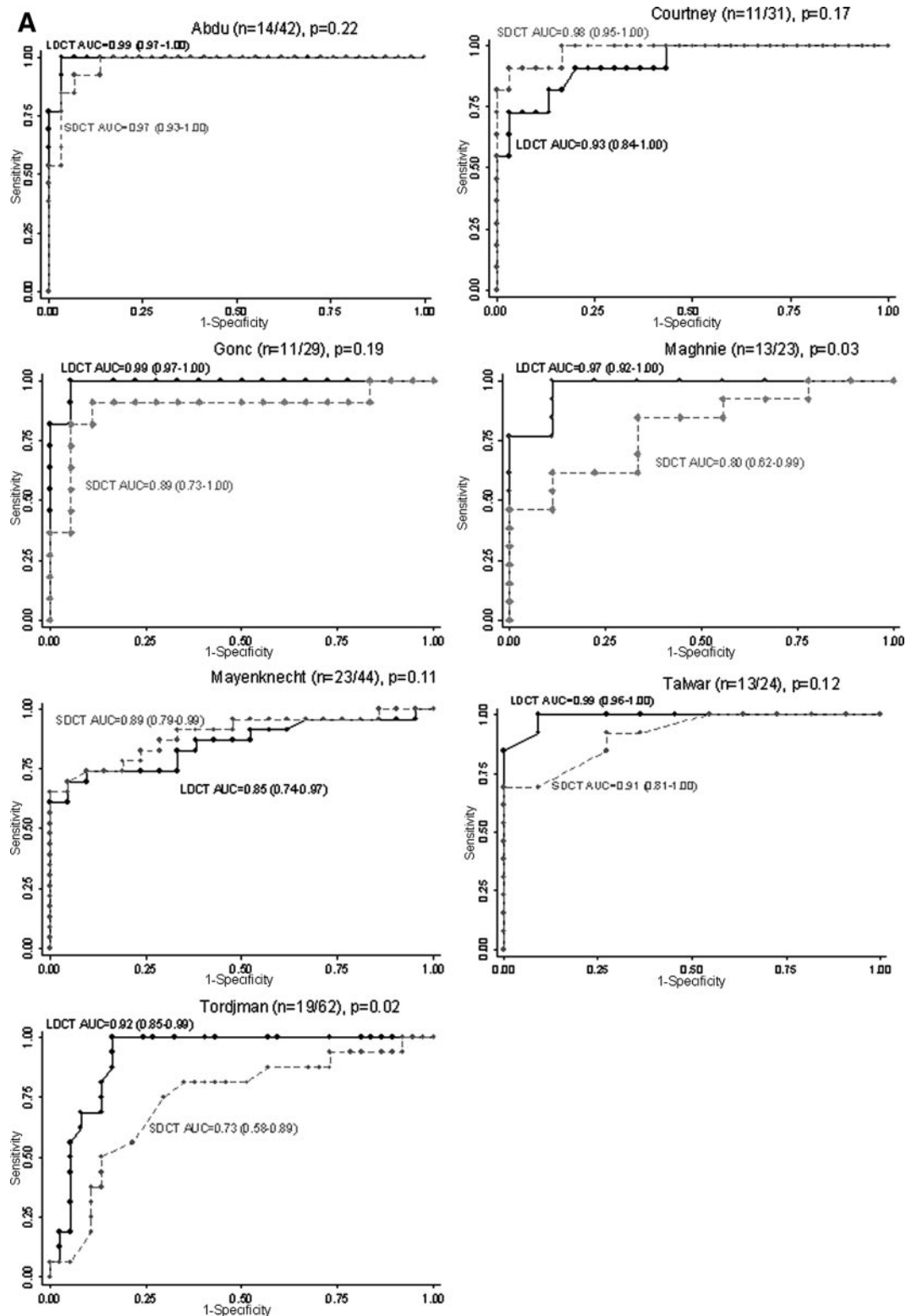


FIG. 1. ROC curves (A, paired data; B, unpaired data) for LDCT and SDCT (30-min cortisol levels) in diagnosing HPA insufficiency.

## Discussion

This metaanalysis demonstrates that the LDCT is superior to the SDCT in diagnosing HPA insufficiency. Because all study subjects were ambulatory and presumably had normal sleep-wake

cycles, these findings may not generalize to hospital settings or patients with acute illnesses.

Our results differ from the metaanalysis of Dorin and colleagues (23), who reported similar operating characteristics for the LDCT and SDCT. There are three possible reasons for the

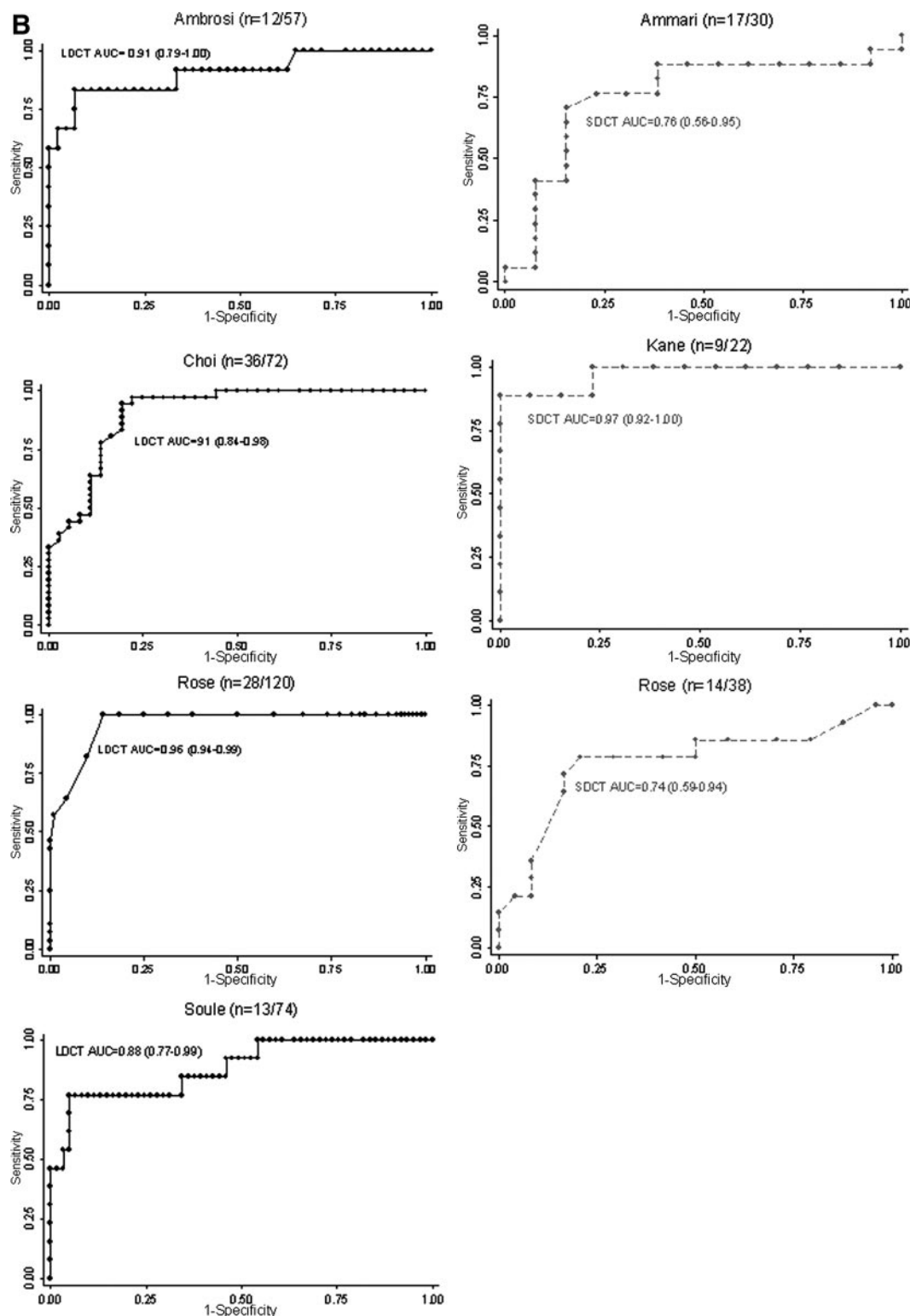


FIG. 1. (continued).

discrepant results. First, the metaanalyses differed in the studies that were included. All studies included in our metaanalysis, except three (5, 16, 17), were also included in the metaanalysis by Dorin and colleagues. The three exceptions were published after Dorin's literature search. Dorin also included 12 studies that we decided not to include for the following reasons: five

studies (published before 1990) tested cortisol levels using the fluorescence polarization assay, which is less cortisol specific than more modern assay methods (8) and has never been used to evaluate LDCT; two studies evaluated subjects in the early post-operative period; and five studies (3 with paired comparisons of LDCT and SDCT) were considered eligible for inclusion in our



**TABLE 1.** Study characteristics

| Study                      | Basal cortisol (early morning) |                           |                              | SDCT (30 min/peak) |                           |                              | LDCT (20–30 min) |                           |                              |
|----------------------------|--------------------------------|---------------------------|------------------------------|--------------------|---------------------------|------------------------------|------------------|---------------------------|------------------------------|
|                            | n HPAI/n total                 | HPAI <sup>a</sup> (μg/dl) | No HPAI <sup>b</sup> (μg/dl) | n HPAI/n total     | HPAI <sup>a</sup> (μg/dl) | No HPAI <sup>b</sup> (μg/dl) | n HPAI/n total   | HPAI <sup>a</sup> (μg/dl) | No HPAI <sup>b</sup> (μg/dl) |
| Studies with paired data   |                                |                           |                              |                    |                           |                              |                  |                           |                              |
| Abdu (22)                  | 13/42                          | ≤3.2                      | ≥12.2                        | 13/42              | ≤8.6                      | ≥19.1                        | 13/42            | ≤12.7                     | ≥16.8                        |
| Courtney (16)              | 11/41                          | ≤3.0                      | ≥11.4                        | 11/41              | ≤17.1                     | ≥22.4                        | 11/41            | ≤7.1                      | ≥22.1                        |
| Gonc (18)                  | 11/29                          | ≤5.7                      | ≥9.0                         | 11/29              | ≤15.6                     | ≥36.1                        | 11/29            | ≤15.8                     | ≥18.6                        |
| Maghnie (19)               | 14/24                          | ≤9.1                      | ≥15.0                        | 13/23              | ≤19.9                     | ≥33.0                        | 14/24            | ≤19.6                     | ≥22.7                        |
| Mayenknecht (11)           | NA                             | NA                        | NA                           | 23/44              | ≤16.0                     | ≥32.0                        | 23/44            | ≤15.8                     | ≥28.0                        |
| Talwar (7) <sup>d</sup>    | 13/24                          | ≤10.0                     | ≥13.5                        | 13/24              | ≤22.5                     | ≥31.7                        | 13/24            | ≤17.9                     | ≥21.0                        |
| Tordjman (21)              | 19/62                          | ≤4.8                      | ≥17.9                        | 16/53              | ≤12.8                     | ≥38.5                        | 19/62            | ≤13.9                     | ≥17.9                        |
| Studies with unpaired data |                                |                           |                              |                    |                           |                              |                  |                           |                              |
| Ambrosi (14)               | 12/57                          | ≤3.6                      | ≥9.3                         | NA                 | NA                        | NA                           | 12/57            | ≤16.6                     | ≥26.8                        |
| Ammari (15)                | 17/30                          | ≤7.4                      | ≥14.7                        | 17/30              | ≤17.0                     | ≥33.9                        | NA               | NA                        | NA                           |
| Choi (5) <sup>d</sup>      | 36/72                          | ≤4.8                      | ≥13.3                        | NA                 | NA                        | NA                           | 36/72            | ≤15.3                     | ≥18.9                        |
| Kane (6) <sup>d</sup>      | 9/22                           | ≤4.4                      | ≥8.6                         | 9/22               | ≤12.2                     | ≥15.7                        | NA               | NA                        | NA                           |
| Rose (17) <sup>c</sup>     | 42/158                         | ≤3.2                      | ≥12.5                        | 14/38              | ≤16.0                     | ≥39.0                        | 28/120           | ≤17.5                     | ≥20.5                        |
| Soule (20)                 | 13/74                          | ≤3.5                      | ≥18.0                        | NA                 | NA                        | NA                           | 13/74            | ≤17.0                     | ≥24.5                        |
| Mean (95% CI)              | 33%                            | ≤5                        | ≥13                          | 40%                | ≤16                       | ≥30                          | 33%              | ≤16                       | ≥22                          |
|                            | 210/635                        | (4.7–5.3)                 | (12.9–13.6)                  | 140/346            | (15.2–16.4)               | (29.9–32.3)                  | 193/586          | (15.2–16.0)               | (20.9–21.9)                  |

To convert serum cortisol in micrograms per deciliter to nanomoles per liter, multiply by 27.56. Mean values were weighted by the size of the studies. NA, No data available.

<sup>a</sup> Threshold below which cortisol values had a LR > 9 (rule-in threshold).

<sup>b</sup> Threshold above which cortisol values had a LR < 0.15 (rule-out threshold).

<sup>c</sup> SDCT and LDCT were performed on different subsets of patients (no paired data).

<sup>d</sup> Included patients with glucocorticoid-induced HPAI.

metaanalysis, but we were unable to obtain patient-level data. Among the three eligible studies with paired comparisons that were not included, one did not recruit consecutive patients (24); the second (25) reported superiority of SDCT, but no data were provided regarding statistical significance (their study also included subjects with glucocorticoid-induced HPAI); and the third (26) demonstrated results favoring LDCT, although the results were not statistically significant.

A second reason is that the metaanalyses differed in the study subjects that were included. Whereas the Dorin metaanalysis used summary data reported in the published articles, we obtained patient-level data from study investigators and could therefore apply patient-level eligibility criteria. For example, subjects were not eligible in our metaanalysis if they were healthy control subjects without any suspicion of HPAI (to reduce the risk of spectrum bias) (4) or if they had pituitary surgery within 4 wk of testing.

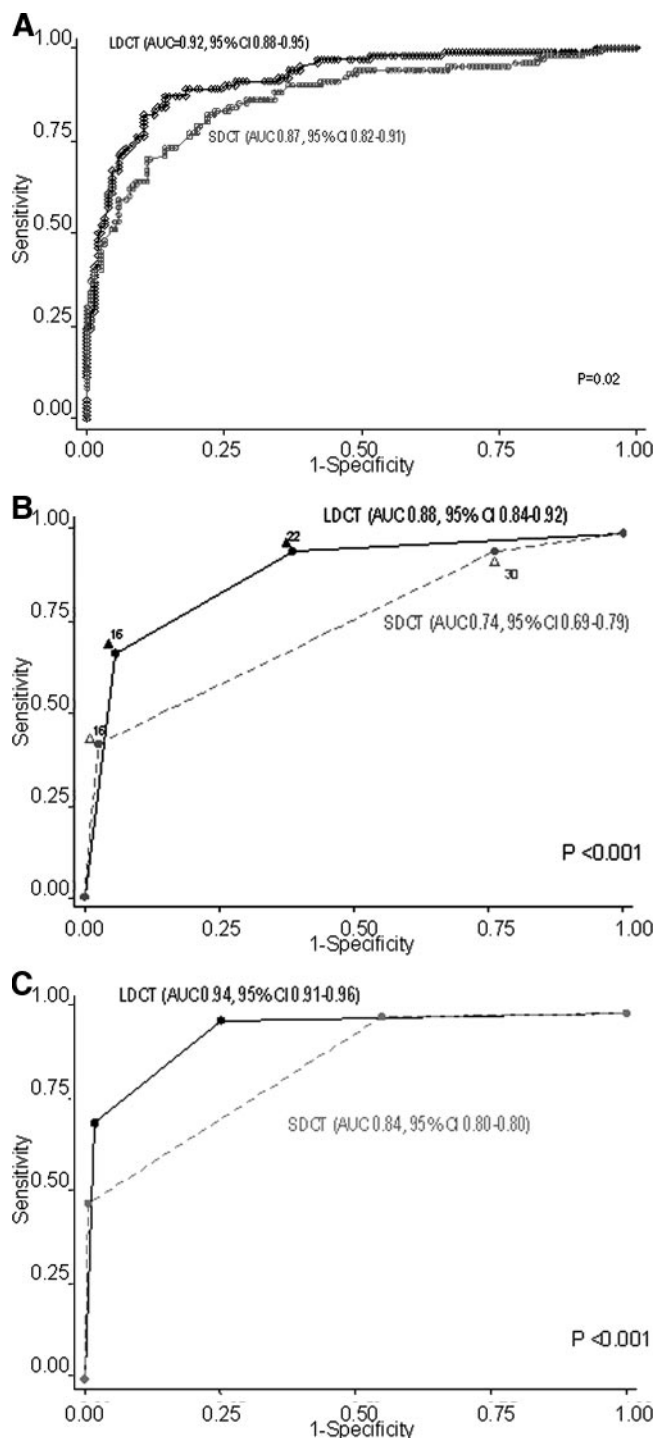
A third reason for the discrepancy is the difference in analytic methods. Because we had access to patient-level data, we were able to adjust for cortisol assay and sample size and also able to reanalyze data to define two (rule-in and rule-out) cortisol thresholds per test, instead of relying on the single cortisol threshold available in published reports.

In the unadjusted analysis (Fig. 2A), the low-dose and standard-dose tests perform similarly; the AUC is slightly better for the LDCT, but the difference is probably clinically unimportant. However, when the analysis is adjusted for study size and cortisol assay, the superiority of the LDCT is more dramatic (Fig. 2C) and likely to be clinically relevant. Using summary cortisol thresh-

olds, regardless of cortisol assay method, we have found that the results of SDCT and LDCT performance comparisons were not sensitive to cortisol assay type (Fig. 2B). Nevertheless, there may be clinical settings where SDCT is more appropriate to diagnose HPAI, especially if the quality of the low-dose testing protocol cannot be assured.

Based on our findings, we suggest a three-step approach for evaluating patients with possible HPAI (Fig. 3). The first step is measuring a morning basal cortisol. If the results are not convincingly normal or abnormal (basal cortisol level falls in indeterminate range of 5–13 μg/dl), then the second step is performing a LDCT. If this test is indeterminate and there are no contraindications to integrated HPA axis testing, we suggest the third step of an insulin hypoglycemia test or metyrapone test. This three-step approach will accurately diagnose the majority of patients, but because it is not perfect, there will still be an important role for clinical judgment, especially regarding use of glucocorticoid supplementation during extreme stress. For convenience, in appropriate clinical circumstances, the first and second steps (basal and LDCT or SDCT) can be done at the same clinical visit to reduce the number of visits for testing.

Gleeson and colleagues (27) published a retrospective study of 10 yr of observational data on patients undergoing SDCT. The clinical diagnosis of HPAI was ascertained by an unblinded assessment of the clinical course depicted in medical records. Although the authors report a 97% negative predictive value for SDCT, the data do not allow calculations of sensitivity, specificity, or accuracy. The impressive negative predictive value might be partly explained by a low prevalence of newly diag-



**FIG. 2.** ROC curves for LDCT and SDCT (30-min cortisol levels) in diagnosing HPA insufficiency. A, Unadjusted for study size and cortisol assay; B, after categorizing cortisol values, but without other adjustments (arrows represent mean cortisol values across the studies; in micrograms per deciliter (multiply by 27.57 to convert to nanomoles per liter); C, after categorizing cortisol values and adjusting for cortisol assay.

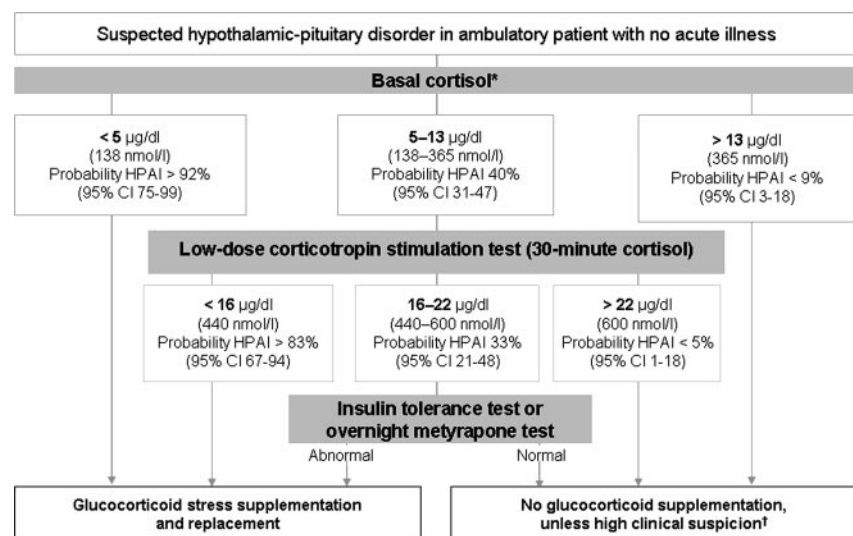
nosed HPAI (19%), which is lower than in most studies included in our metaanalysis. The lower prevalence might be a result of reference test bias, because the clinical diagnosis of milder HPAI might be easily missed without the aid of integrated HPA axis testing. In addition, there is a significant risk of selection bias because 38% of patients were excluded from analysis.

Agha and colleagues (28) reported that only 3–8% of patients with an intermediate response to SDCT (30 min cortisol, 18.5–23  $\mu\text{g}/\text{dl}$  or 510–635 nmol/liter) developed signs of adrenal insufficiency over an average follow-up of 4.2 yr. The implication is that a clearly negative standard-dose test result ( $>23 \mu\text{g}/\text{dl}$ , or  $>635 \text{ nmol}/\text{liter}$ ) should perform even better than the intermediate results and effectively rule out HPAI. However, no data are provided regarding patients with either positive or negative standard-dose test results, and there is no assessment of the overall prevalence of adrenal insufficiency in the entire test population. In fact, for certain cortisol assay methods (6), a 30-min cortisol value in the range studied by Agha (18.5–23  $\mu\text{g}/\text{dl}$ , or 510–635 nmol/liter) would be classified as a negative test, with a low likelihood of HPAI (Table 1). Unfortunately, the data from Agha's study do not allow the calculation of test characteristics for positive or negative SDCT results.

Due to the lack of cortisol assay standardization and other reasons for measurement variability, the error in measuring cortisol can be up to 6  $\mu\text{g}/\text{dl}$  (165 nmol/liter); thus, caution is advised when making clinical decisions based on cortisol values close to threshold values. In addition to high variability in the cortisol thresholds, another concern is that a test with a low likelihood of HPAI does not exclude the possibility of future HPAI, especially if there is progression of hypothalamic-pituitary disease or radiation therapy. Therefore, longitudinal assessments may be necessary.

The LDCT has not been validated in patients with acute illnesses, abnormal sleep-wake cycles, or acute hypothalamic-pituitary disorders (e.g. within 1 month of pituitary surgery). In addition, all studies of the LDCT that were included in this meta-analysis were conducted in the morning with the patients fasting. Afternoon cortisol values tend to be lower by 1–1.5  $\mu\text{g}/\text{dl}$  (28–58 nmol/liter) (13, 29), and the effect of eating or drinking is uncertain. We also have no information on how the LDCT would perform among patients with low serum protein levels, because circulating cortisol is highly protein bound.

There are several technical details to performing a low-dose test that must be rigorously addressed to avoid false-positive test results (falsely low 30-min stimulated cortisol value). Currently, there are two acceptable corticotropin analogs that can be used, cosyntropin (Cortrosyn, Amphastar Pharmaceuticals) or tetracosactrin (Synacthen, Novartis Pharma), supplied in vials containing 250  $\mu\text{g}$  powder. Preparing the 1- $\mu\text{g}$  dose requires a several-step process of first reconstituting with 250 ml normal saline and then using a 1-ml aliquot (1  $\mu\text{g}$ ) for iv injection. There are additional steps for minimizing adherence of the medication to plastic tubing (30). In addition, the timing of cortisol sampling after low-dose corticotropin administration is very important (we recommend collecting the blood sample 20–30 min after corticotropin analog administration), because later sampling may result in false-positive results (31). Thus, low-dose testing should be performed only by personnel knowledgeable of the multiple steps required for preparation and administration. If the quality of administering a 1- $\mu\text{g}$  dose is suspect, then we recommend using the standard dose of 250  $\mu\text{g}$  (reconstituted with 1 ml sterile diluent) and measuring serum cortisol 30 min after iv injection. A result less than 16  $\mu\text{g}/\text{dl}$  (440 nmol/liter), which is the



\* A normal DHEA-S blood level makes HPAI highly unlikely (32) and therefore supports the finding of a normal basal cortisol level.

† This algorithm is valid only for chronic HPAI (longer than 4–6 weeks duration). The corticotropin analog stimulation tests have no proven validity in suspected acute or subacute HPAI (less than 4–6 weeks). Insulin tolerance test or overnight metyrapone test can be used to make a diagnosis of acute or subacute HPAI. In addition, even when testing with a corticotropin analog indicates good adrenal responsiveness, the HPA integrity may need reassessment over time to assess for progression of hypothalamic-pituitary dysfunction.

FIG. 3. Optimal testing strategy for evaluating patients with possible HPA insufficiency.

same threshold used for LDCT, strongly suggests HPAI. However, with standard-dose testing, the 30-min cortisol value must be greater than 30 µg/dl (833 nmol/liter) to be reasonably confident in ruling out HPAI.

Research has suggested (32, 33) that dehydroepiandrosterone sulfate (DHEA-S) blood levels might also help with assessing the HPA axis, particularly when the results of the LDCT are close to either of the two threshold values.

In establishing the diagnosis of HPAI, the accepted reference standard is an abnormal insulin tolerance test or metyrapone test. Both tests, however, can be unreliable. The average intra-subject variability in peak cortisol response to insulin-induced hypoglycemia is 8–12% (34), but in men with hypopituitarism, it can vary by 42% (35). Healthy control subjects have been known to fail this test. Neither of these reference tests has been validated by assessing predictive accuracy, that is, the ability to predict adrenal crisis.

A limitation of this metaanalysis is that we could not include data of four studies that had published paired results of LDCT and SDCT, because we were unable to obtain the patient-level data (24–26, 32). One study favored LDCT (26), and another study (25) apparently favored SDCT; however, the latter study had an incomplete ROC to estimate the magnitude of the difference. The third study (32) found no difference in performance of the two tests. The fourth study (24) used nonconsecutive patients for testing and is therefore not pertinent to our metaanalysis.

In conclusion, the performance of the LDCT was superior to the standard-dose test for evaluating HPA insufficiency. However, LDCT must be used by personnel knowledgeable of the multiple steps required for preparation and administration. We describe a three-step testing strategy that proposes use of basal cortisol and the low-dose stimulation test before proceeding to one of the reference tests for HPA insufficiency.

## Acknowledgments

Address all correspondence and requests for reprints to: Rasa Kazlauskaitė, M.D., M.Sc., 1900 West Polk Street, Room 806, Chicago, Illinois 60612. E-mail: rasa\_kazlauskaitė@rush.edu.

Disclosure Summary: The authors have nothing to disclose.

## References

- Grinspoon SK, Biller BM 1994 Laboratory assessment of adrenal insufficiency. *J Clin Endocrinol Metab* 79:923–931
- Hartband PI, Van Herle AJ, Sorger L, Cope D 1988 Assessment of hypothalamic-pituitary-adrenal (HPA) axis dysfunction: comparison of ACTH stimulation, insulin-hypoglycemia and metyrapone. *J Endocrinol Invest* 11:769–776
- Wood JB, Frankland AW, James VH, Landon J 1965 A rapid test of adrenocortical function. *Lancet* 191:243–245
- Ransohoff DF, Feinstein AR 1978 Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 299:926–930
- Choi CH, Tiu SC, Shek CC, Choi KL, Chan FK, Kong PS 2002 Use of the low-dose corticotropin stimulation test for the diagnosis of secondary adrenocortical insufficiency. *Hong Kong Med J* 8:427–434
- Kane KF, Emery P, Sheppard MC, Stewart PM 1995 Assessing the hypothalamic-pituitary-adrenal axis in patients on long-term glucocorticoid therapy: the short synacthen versus the insulin tolerance test. *QJM* 88:263–267
- Talwar V, Lodha S, Dash RJ 1998 Assessing the hypothalamic-pituitary-adrenal axis using physiological doses of adrenocorticotrophic hormone. *QJM* 91:285–290
- Nye EJ, Grice JE, Hockings GI, Strakosch CR, Crosbie GV, Walters MM, Jackson RV 1999 Comparison of adrenocorticotropin (ACTH) stimulation tests and insulin hypoglycemia in normal humans: low dose, standard high dose, and 8-hour ACTH-(1–24) infusion tests. *J Clin Endocrinol Metab* 84:3648–3655
- Odagiri E, Naruse M, Terasaki K, Yamaguchi N, Jibiki K, Takagi S, Tanabe M, Takano K 2004 The diagnostic standard of preclinical Cushing's syndrome: evaluation of the dexamethasone suppression test using various cortisol kits. *Endocr J* 51:295–302
- Clark PM, Neylon I, Raggatt PR, Sheppard MC, Stewart PM 1998 Defining the normal cortisol response to the short Synacthen test: implications for the investigation of hypothalamic-pituitary disorders. *Clin Endocrinol (Oxf)* 49:287–292
- Mayenknecht J, Diederich S, Bahr V, Plockinger U, Oelkers W 1998 Comparison of low and high dose corticotropin stimulation tests in patients with pituitary disease. *J Clin Endocrinol Metab* 83:1558–1562
- Tuchelt H, Dekker K, Bahr V, Oelkers W 2000 Dose-response relationship between plasma ACTH and serum cortisol in the insulin-hypoglycemia test in 25 healthy subjects and 109 patients with pituitary disease. *Clin Endocrinol (Oxf)* 53:301–307
- Dickstein G, Shechner C, Nicholson WE, Rosner I, Shen-Orr Z, Adawi F, Lahav M 1991 Adrenocorticotropin stimulation test: effects of basal cortisol level, time of day, and suggested new sensitive low dose test. *J Clin Endocrinol Metab* 72:773–778
- Ambrosi B, Barbetta L, Re T, Passini E, Faglia G 1998 The one microgram adrenocorticotropin test in the assessment of hypothalamic-pituitary-adrenal function. *Eur J Endocrinol* 139:575–579
- Ammari F, Issa BG, Millward E, Scanlon MF 1996 A comparison between short ACTH and insulin stress tests for assessing hypothalamic-pituitary-adrenal function. *Clin Endocrinol (Oxf)* 44:473–476
- Courtney CH, McAllister AS, Bell PM, McCance DR, Leslie H, Sheridan B, Atkinson AB 2004 Low- and standard-dose corticotropin and insulin hypoglycemia testing in the assessment of hypothalamic-pituitary-adrenal function after pituitary surgery. *J Clin Endocrinol Metab* 89:1712–1717



17. Rose SR, Lustig RH, Burstein S, Pitukcheewanont P, Broome DC, Burghen GA 1999 Diagnosis of ACTH deficiency. Comparison of overnight metyrapone test to either low-dose or high-dose ACTH test. *Horm Res* 52:73–79
18. Gonc EN, Kandemir N, Kinik ST 2003 Significance of low-dose and standard-dose ACTH tests compared to overnight metyrapone test in the diagnosis of adrenal insufficiency in childhood. *Horm Res* 60:191–197
19. Maghnie M, Uga E, Temporini F, Di Iorgi N, Secco A, Tinelli C, Papalia A, Casini MR, Loche S 2005 Evaluation of adrenal function in patients with growth hormone deficiency and hypothalamic-pituitary disorders: comparison between insulin-induced hypoglycemia, low-dose ACTH, standard ACTH and CRH stimulation tests. *Eur J Endocrinol* 152:735–741
20. Soule S, Van Zyl Smit C, Parolis G, Attenborough S, Peter D, Kinvig S, Kinvig T, Coetzer E 2000 The low dose ACTH stimulation test is less sensitive than the overnight metyrapone test for the diagnosis of secondary hypoadrenalism. *Clin Endocrinol (Oxf)* 53:221–227
21. Tordjman K, Jaffe A, Trostanetsky Y, Greenman Y, Limor R, Stern N 2000 Low-dose (1 microgram) adrenocorticotrophin (ACTH) stimulation as a screening test for impaired hypothalamo-pituitary-adrenal axis function: sensitivity, specificity and accuracy in comparison with the high-dose (250 microgram) test. *Clin Endocrinol (Oxf)* 52:633–640
22. Abdu TA, Elhadd TA, Neary R, Clayton RN 1999 Comparison of the low dose short synacthen test (1 µg), the conventional dose short synacthen test (250 µg), and the insulin tolerance test for assessment of the hypothalamo-pituitary-adrenal axis in patients with pituitary disease. *J Clin Endocrinol Metab* 84:838–843
23. Dorin RI, Qualls CR, Crapo LM 2003 Diagnosis of adrenal insufficiency. *Ann Intern Med* 139:194–204
24. Weintrob N, Sprecher E, Josefsberg Z, Weininger C, Aurbach-Klipper Y, Lazard D, Karp M, Pertzelan A 1998 Standard and low-dose short adrenocorticotropin test compared with insulin-induced hypoglycemia for assessment of the hypothalamic-pituitary-adrenal axis in children with idiopathic multiple pituitary hormone deficiencies. *J Clin Endocrinol Metab* 83:88–92
25. Suliman AM, Smith TP, Labib M, Fiad TM, McKenna TJ 2002 The low-dose ACTH test does not provide a useful assessment of the hypothalamic-pituitary-adrenal axis in secondary adrenal insufficiency. *Clin Endocrinol (Oxf)* 56:533–539
26. Rasmuson S, Olsson T, Hagg E 1996 A low dose ACTH test to assess the function of the hypothalamic-pituitary-adrenal axis. *Clin Endocrinol (Oxf)* 44:151–156
27. Gleeson HK, Walker BR, Seckl JR, Padfield PL 2003 Ten years on: safety of short synacthen tests in assessing adrenocorticotropin deficiency in clinical practice. *J Clin Endocrinol Metab* 88:2106–2111
28. Agha A, Tomlinson JW, Clark PM, Holder G, Stewart PM 2006 The long-term predictive accuracy of the short synacthen (corticotropin) stimulation test for assessment of the hypothalamic-pituitary-adrenal axis. *J Clin Endocrinol Metab* 91:43–47
29. Park YJ, Park KS, Kim JH, Shin CS, Kim SY, Lee HK 1999 Reproducibility of the cortisol response to stimulation with the low dose (1 microg) of ACTH. *Clin Endocrinol (Oxf)* 51:153–158
30. Murphy H, Livesey J, Espiner EA, Donald RA 1998 The low dose ACTH test: a further word of caution. *J Clin Endocrinol Metab* 83:712–713
31. Oelkers W 1998 The role of high- and low-dose corticotropin tests in the diagnosis of secondary adrenal insufficiency. *Eur J Endocrinol* 139:567–570
32. Nasrallah MP, Arafah BM 2003 The value of dehydroepiandrosterone sulfate measurements in the assessment of adrenal function. *J Clin Endocrinol Metab* 88:5293–5298
33. Fischli S, Jenni S, Allemann S, Zwahlen M, Diem P, Christ ER, Stettler C 2008 Dehydroepiandrosterone sulfate in the assessment of the hypothalamic-pituitary-adrenal axis. *J Clin Endocrinol Metab* 93:539–542
34. Vestergaard P, Hoeck HC, Jakobsen PE, Laurberg P 1997 Reproducibility of growth hormone and cortisol responses to the insulin tolerance test and the short ACTH test in normal adults. *Horm Metab Res* 29:106–110
35. Pfeifer M, Kanc K, Verhovec R, Kocijancic A 2001 Reproducibility of the insulin tolerance test (ITT) for assessment of growth hormone and cortisol secretion in normal and hypopituitary adult men. *Clin Endocrinol (Oxf)* 54:17–22