Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses


CRD summary
This well-conducted review concluded that commonly used tests to diagnose Cushing's syndrome appeared to be highly accurate in referred patients, but their performance in usual clinical practice remained unclear. These conclusions are likely to be reliable.

Authors' objectives
To summarise the evidence for the accuracy of common tests for diagnosing Cushing's syndrome.

Searching
MEDLINE, EMBASE, Web of Science, and Scopus were searched for articles from 1975 to September 2007. Search terms were not reported, but the search strategy was available from the authors, on request. Experts from a relevant society were contacted and neither language nor publication restrictions were applied.

Study selection
Diagnostic cohort studies were eligible for inclusion if they evaluated the accuracy of urinary free cortisol, serum and salivary midnight or bedtime cortisol, 1mg overnight dexamethasone suppression test (DST), or the two-day 2mg DST, compared with a reference standard for Cushing's syndrome, in individuals in whom there was true diagnostic uncertainty. The eligible reference standards were pathological diagnosis, response to therapy targeting Cushing's syndrome, or clinical follow-up. Studies had to report data on sensitivity and specificity or likelihood ratios.

Most studies enrolled patients with a suspicion of Cushing's syndrome from referral centres and had a high prevalence of Cushing's syndrome; some studies enrolled patients without a suspicion of Cushing's syndrome. The prevalence of Cushing's syndrome ranged from 0.1% to 91%. Reference standards used in the included studies were pathology alone or in combination with clinical information. Where reported, the ages ranged from five to 92 years.

Two reviewers independently assessed studies for inclusion and disagreements were resolved through consensus or arbitration.

Assessment of study quality
Two reviewers independently assessed study quality using Quality Assessment of Diagnostic Accuracy Studies (QUADAS). Where more than one test was evaluated in a single study, the relevant QUADAS items were applied to each test within the study.

Data extraction
Two reviewers independently extracted data. Thresholds in the primary studies were used to estimate accuracy. Where results were reported for more than one threshold, the threshold that offered the best test performance was selected. Authors were contacted and asked to verify the extracted data and to provide missing data.

Methods of synthesis
Random-effects models were used to estimate summary sensitivity, specificity, positive and negative likelihood ratios, and diagnostic odds ratios together with 95% confidence intervals (CIs) stratified according to the index test. Analyses were repeated using the bivariate model. Heterogeneity was assessed using the I² statistic. Summary receiver operating characteristic curves were estimated. The following possible sources of heterogeneity were investigated, using a test for interaction: severity of Cushing's syndrome, selection bias, types of patients, threshold rationale, and test characteristics.

Results of the review
Twenty-five studies were included in the review (n=8,631). Study quality was generally good. All but one of the
studies enrolled an appropriate spectrum of patients and all but one clearly reported the selection criteria. Verification bias was a potential problem in nine studies.

**Urinary free cortisol** (14 studies): The summary positive likelihood ratio was 10.6 (95% CI 5.5 to 20.5) and summary negative likelihood ratio was 0.16 (95% CI 0.08 to 0.33). There was moderate heterogeneity in the diagnostic odds ratio ($I^2=44\%$) and values for the likelihood ratio were not reported.

**Midnight serum cortisol** (six studies): The summary positive likelihood ratio was 9.5 (95% CI 1.7 to 54.1) and summary negative likelihood ratio was 0.09 (95% CI 0.03 to 0.28). There was considerable heterogeneity in the diagnostic odds ratio ($I^2=78\%$). In studies in which thresholds were assay driven, the test had poorer performance (two studies) than those in which thresholds were outcome driven (four studies): positive likelihood ratio 1.8 compared with 26.6 and negative likelihood ratio 0.47 compared with 0.05.

**Midnight salivary cortisol** (four studies): The summary positive likelihood ratio was 8.8 (95% CI 3.5 to 21.8) and summary negative likelihood ratio was 0.07 (95% CI 0.00 to 1.20). There was moderate heterogeneity in the diagnostic odds ratio ($I^2=50\%$).

**1mg overnight DST** (14 studies): The summary positive likelihood ratio was 11.6 (95% CI 5.8 to 23.1) and summary negative likelihood ratio was 0.09 (95% CI 0.05 to 0.14). There was little heterogeneity in the diagnostic odds ratio ($I^2=11\%$). Studies in which the prevalence of Cushing's syndrome was under 50% showed better performance (11 studies) than those in which the prevalence was over 50% (three studies): positive likelihood ratio 16.4 compared with 2.8 and negative likelihood ratio 0.06 compared with 0.11.

**Two-day 2mg DST** (eight studies): The summary positive likelihood ratio was 7.3 (95% CI 3.6 to 15.2) and summary negative likelihood ratio was 0.18 (95% CI 0.06 to 0.52). There was no heterogeneity in the diagnostic odds ratio ($I^2=0\%$).

**Urinary free cortisol plus 1mg overnight DST** (three studies): The summary positive likelihood ratio was 15.4 (95% CI 0.7 to 358.0) and summary negative likelihood ratio was 0.11 (95% CI 0.01 to 1.57). There was considerable heterogeneity in the diagnostic odds ratio ($I^2=90\%$).

All subgroups analyses, with the exception of those detailed above, failed to show significant associations.

**Authors' conclusions**
Commonly used tests to diagnose Cushing's syndrome appeared to be highly accurate in samples of patients referred with suspected Cushing's syndrome. Their performance in usual clinical practice remained unclear.

**CRD commentary**
This review addressed a focused question, supported by clearly defined inclusion criteria. The search was comprehensive and included some attempts to locate unpublished studies. Appropriate steps were taken to minimise bias at all stages of the review process. Study quality was assessed, using appropriate criteria, and the results were clearly presented in tables and considered in the analysis. Methods used to pool the studies were appropriate and heterogeneity was investigated.

This was a well conducted review and the authors' conclusions are likely to be reliable.

**Implications of the review for practice and research**
**Practice:** The authors stated that the implications for practice were summarised as a proposed algorithm in an accompanying Endocrine Society practice guideline (see Other Publications of Related Interest).

**Research:** The authors stated that their proposed algorithm needed evaluation in prospective studies. Future studies should report findings using likelihood ratios for the test result ranges rather than forcing a single threshold on the data, and they should use diagnostic categories that include indeterminate results.
**Funding**
The Endocrine Society.

**Bibliographic details**

**PubMedID**
18334594

**DOI**
10.1210/jc.2008-0139

**Original Paper URL**
http://jcem.endojournals.org/cgi/content/abstract/93/5/1553

**Other publications of related interest**

**Indexing Status**
Subject indexing assigned by NLM

**MeSH**
Cushing Syndrome /diagnosis; Dexamethasone; Humans; Hydrocortisone /urine

**AccessionNumber**
12008107209

**Date bibliographic record published**
18/11/2009

**Date abstract record published**
07/04/2010

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.