Accuracy of Diagnostic Tests for Cushing's Syndrome: A Systematic Review and Metaanalyses

Mohamed B. Elamin, M. Hassan Murad, Rebecca Mullan, Dana Erickson, Katherine Harris, Sarah Nadeem, Robert Ennis, Patricia J. Erwin, and Victor M. Montori

Knowledge and Encounter Research Unit (M.B.E., M.H.M., R.M., P.J.E., V.M.M.), Division of Preventive Medicine (M.H.M.), Division of Endocrinology, Diabetes, Metabolism, Nutrition (D.E., S.N., V.M.M.), Department of Medicine (M.H.M., D.E., K.H., R.E., V.M.M.), Mayo Clinic, and Mayo Clinic Libraries (P.J.E.), Mayo Clinic, Rochester, Minnesota 55905

Context: The diagnosis of Cushing's syndrome (CS) requires the use of tests of unregulated hypercortisolism that have unclear accuracy.

Objective: Our objective was to summarize evidence on the accuracy of common tests for diagnosing CS.

Data Sources: We searched electronic databases (MEDLINE, EMBASE, Web of Science, Scopus, and citation search for key articles) from 1975 through September 2007 and sought additional references from experts.

Study Selection: Eligible studies reported on the accuracy of urinary free cortisol (UFC), dexamethasone suppression test (DST), and midnight cortisol assays *vs.* reference standard in patients suspected of CS.

Data Extraction: Reviewers working in duplicate and independently extracted study characteristics and quality and data to estimate the likelihood ratio (LR) and the 95% confidence interval (CI) for each result.

Data Synthesis: We found 27 eligible studies, with a high prevalence [794 (9.2%) of 8631 patients had CS] and severity of CS. The tests had similar accuracy: UFC (n = 14 studies; LR+ 10.6, CI 5.5–20.5; LR− 0.16, CI 0.08–0.33), salivary midnight cortisol (n = 4; LR+ 8.8, CI 3.5–21.8; LR− 0.07, CI 0–1.2), and the 1-mg overnight DST (n = 14; LR+ 16.4, CI 9.3–28.8; LR− 0.06, CI 0.03–0.14). Combined testing strategies (*e.g.* a positive result in both UFC and 1-mg overnight DST) had similar diagnostic accuracy (n = 3; LR+ 15.4, CI 0.7–358; LR− 0.11, CI 0.007–1.57).

Conclusions: Commonly used tests to diagnose CS appear highly accurate in referral practices with samples enriched with patients with CS. Their performance in usual clinical practice remains unclear. **(*J Clin Endocrinol Metab* 93: 1553–1562, 2008)**

Cushing's syndrome (CS) results from the excessive exposure of the body to glucocorticoids, either from endogenous or, more commonly, exogenous sources. Severe CS is rare and requires urgent attention due to the natural history of this condition, which is associated with important morbidity and mortality (1). Because of the overt presentation of severe CS, clinicians familiar with this condition can often make a firm diagnosis on clinical and biochemical grounds (2).

The aging population and the obesity epidemic are making some features of CS, such as central obesity, hypertension, hyperglycemia, and bone fragility, common. Therefore, detecting patients with CS, particularly those with milder forms, requires

0021-972X/08/$15.00/0

Printed in U.S.A.

doi: 10.1210/jc.2008-0139 Received January 22, 2008. Accepted March 3, 2008.

First Published Online March 11, 2008

See article p. 1526

Abbreviations: CS, Cushing's syndrome; DST, dexamethasone suppression test; ROC, receiver operator characteristic; UFC, urinary free cortisol.

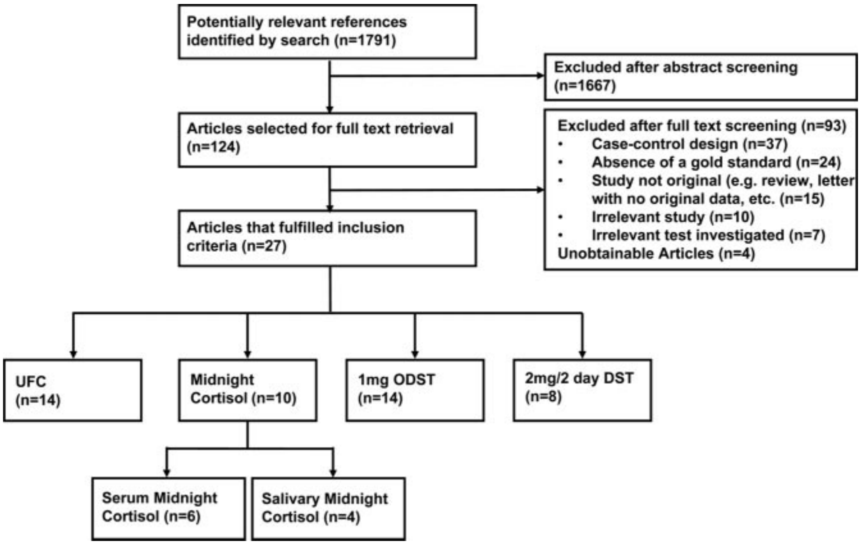
Potentially relevant references identified by search (n=1791)
Excluded after abstract screening (n=1667)
Articles selected for full text retrieval (n=124)
Excluded after full text screening (n=93)
• Case-control design (n=37)
• Absence of a gold standard (n=24)
• Study not original (e.g. review, letter with no original data, etc. (n=15)
• Irrelevant study (n=10)
• Irrelevant test investigated (n=7)
Unobtainable Articles (n=4)
Articles that fulfilled inclusion criteria (n=27)
UFC (n=14)
Midnight Cortisol (n=10)
1mg ODST (n=14)
2mg/2 day DST (n=8)
Serum Midnight Cortisol (n=6)
Salivary Midnight Cortisol (n=4)

FIG. 1. Results of the systematic review with flow of studies for eligibility into the review and into each metaanalysis. ODST, Overnight DST.

accurate tests that are able to discriminate patients with and without hypercortisolism (3–5).

To summarize the available evidence of diagnostic accuracy of tests of abnormal cortisol overproduction, The Endocrine Society Cushing's Syndrome Task Force commissioned us to conduct a systematic review of diagnostic accuracy of diagnostic tests for CS.

Materials and Methods

The protocol of this review, approved by the Task Force, adheres to current methodological guidelines on the conduct of systematic reviews of diagnostic accuracy (6).

Eligibility criteria

We included cross-sectional and longitudinal studies that enrolled participants with true diagnostic uncertainty. Therefore, the diagnosis of CS could not be a criterion for enrollment in these studies, so-called phase II and III diagnostic studies (7). These studies may have included individuals selected because they had physical findings or comorbid conditions suggestive of CS.

Tests of interest were urinary free cortisol (UFC), serum and salivary midnight/bedtime cortisol, 1-mg overnight dexamethasone suppression test (DST) or the 2-d 2 mg DST. Eligible studies had a reference standard for diagnosing CS. Eligible reference standards included a pathological diagnosis, response to therapy targeting CS, or clinical follow-up (*i.e.* consensus among treating clinicians about a diagnosis of CS). Eligible studies measured the accuracy of test results with results expressed as 1) both sensitivity and specificity or 2) likelihood ratio. We included studies regardless of their publication status, language, or size.

Study identification

An expert reference librarian (P.J.E.) designed and conducted the electronic search strategy with input from study investigators with expertise in conducting systematic reviews. To identify eligible studies, we searched electronic databases (MEDLINE, EMBASE, Web of Science, Scopus, and citation search for key articles) from 1975 through September 2007. The detailed search strategy is available upon request. We also sought references from experts from The Endocrine Society Cushing's Syndrome Task Force.

Reviewers working independently and in duplicate reviewed all abstracts and titles and, upon retrieval of potentially eligible studies, the full text publications for eligibility with adequate chance-adjusted inter-reviewer agreement (κ statistic = 0.6; 95% confidence interval 0.4–0.7). Disagreements were resolved by consensus or arbitration.

Quality assessment

Reviewers working independently and in duplicate analyzed the eligible articles to assess the reported quality of the methods. We followed the tool for quality assessment of studies of diagnostic accuracy included in systematic reviews (QUADAS) (8).

Data extraction

Reviewers working independently and in pairs used a standardized form to extract a full description of study participants, including judgments about the extent of diagnostic uncertainty, the presence of comorbid conditions as eligibility criteria (not as characteristics of the sample), the tests and the procedures followed to conduct them, the cutoff or range definitions of diagnostic tests, whether these cutoffs were derived from previous research or determined by study authors, and the nature and characteristics of the reference standard used. To extract data to estimate diagnostic accuracy measures, we used the cutoffs authors chose to use in the primary studies. If more than one cutoff was reported or if the results were reported at the individual patient level, then we chose to use cutoffs that offered the best test performance.

Author contact

We sent letters to the corresponding authors (or any other author with contact address listed on the main manuscript) of each of the eligible studies by electronic mail (regular mail if we could not obtain an active e-mail). We asked these authors to verify the data we extracted and to complete missing data we could not identify in the published record. In case of no response, we repeated the request 2 wk later.

Statistical analysis

We used Meta-DiSc Software for Meta-analysis for Screening and Diagnostic tests version 1.4 (9). Using random effects metaanalyses, we pooled the sensitivities, specificities, likelihood ratios, and diagnostic odds ratio and estimated the 95% confidence intervals for the outcomes. Because the pooled sensitivity and the pooled specificity are interrelated, we focused our analyses on estimating and pooling likelihood ratios and diagnostic odds ratios. The diagnostic odds ratio of a test describes the ratio of the odds of a positive test result in patients with disease compared with patients without disease (10) and can be calculated as the ratio of the likelihood ratios for a positive and a negative test. It has the advantage of being a single indicator of test performance that provides a global meaning of agreement between a test and a reference standard and allows for pooling across studies when the main source of inconsistency is the threshold to consider a test positive [*i.e.* when there is a common receiver operator characteristic (ROC) curve across all studies].

Summary ROC curves allow readers to visually inspect the consistency of results across studies (answering the question of whether there is a single ROC curve across all these studies) and the accuracy of the test, as judged by the area under the summary ROC curve, in discriminating between patients with and without CS. In contrast to ROC curves in which individual data points represent different test cutoffs, in summary ROC curves, each point represents a study (11). We assessed the inconsistency among studies using the I^2 statistic, which represents the proportion of variability across studies that is not due to chance. I^2 values of

TABLE 1. Baseline characteristics of included studies

Author, year (Ref.)	Cohort characteristics	Mean age, yr (range)	No. cohort (% women)	No. CS (% CS, % CD, % adrenal origin, % ectopic, % other)	No. indeterminate (%)	No. lost to follow-up (%)	CS definition	Follow-up length (months)	Test studied
Eddy, 1973 (20)	Suspicion referral[a]	(21–64)	39 (74.4)	24 (61.5, 45.8, 41.6, 12.5, 0)	0 (0)	0 (0)	Path and clinical	>12	UFC (immunoassay, single assay-driven cutoff), ODST (assay-driven cutoff), LDDST (assay-driven cutoff)
Barbarino, 1979 (16)	Suspicion referral[a]	NR	23 (NR)	10 (43.5, 40, 50, 0, 10)	0 (0)	0 (0)	Path and clinical	NR	ODST (assay-driven cutoff)
Ashcraft, 1982 (15)[b]	Suspected Cushing, but not specified if referral or not	28 (5–60)	21 (49)	NR (NR, NR, NR, 13.3, NR)	NR	NR	Path and clinical	NR	LDDST
Meikle, 1982 (28)	Suspected Cushing, but not specified if referral or not	(15–70)	175 (NR)	17 (9.7, 70.6, 29.4, 0, 0)	0 (0)	1 (0.6)	Path and clinical	Several	ODST (assay-driven cutoff)
Kreze, 1983 (24)	Suspicion referral[a]	NR	35 (65.7)	5 (14.3, 40, 0, 0, 60)	0 (0)	0 (0)	Path and clinical	12–36	UFC (immunoassay, single assay-driven cutoff), LDDST (assay-driven cutoff)
Vidal Trecan, 1983 (40)	Suspicion referral	11.9 (9–72)	130 (88)	26 (20, 77, 23, 0, 0)	0 (0)	NR	Path	>12	UFC (Immunoassay, single assay driven cutoff)
Dunlap, 1985 (19)	Suspicion-nonreferral	NR	43 (74.4)	18 (41.8, 55.5, 27.7, 16.7, 0)	0 (0)	0 (0)	Path	>12	UFC (immunoassay, single assay-driven cutoff), LDDST (assay-driven cutoff)
Cronin, 1990 (18)	No suspicion	30 (14–46)	100(80)	4 (4, 4, 0, 0, 0)	0 (0)	0 (0)	Path	NR	ODST (outcome-driven cutoff[c] with assay sensitivity of 5 nmol/liter)
Yanovski, 1993 (41)	Suspicion referral	38.5	58 (71)	39 (67.3, 89.7, 5.1,5.1, 0)	0 (0)	0 (0)	Path and clinical	NR	LDDST (outcome-driven cutoff with assay sensitivity of 5.5–22 nmol/liter)
Leibowitz, 1996 (25)	No suspicion[a]	53.5 (21–78)	90 (64)	3 (3.33, 66.67, 33.33,0, 0)	0 (0)	0 (0)	Path	NR	ODST (assay-driven cutoff with assay sensitivity of 28 nmol/liter)
Papanicolaou, 1998 (32)	Suspicion referral[d]	35 (5–77)	263 (75)	240 (91.3, 83, 6, 11, 0)	0 (0)	59 (18.3)	Path and clinical	21	UFC (immunoassay, single outcome-driven cutoff), MSerC (outcome-driven cutoff)
Raff, 1998 (35)	Suspicion referral[d]	44	78 (NR)	39 (50, 76.9, 12.8, 10.2, 0)	2 (2.6)	0 (0)	Path and clinical	NR	MSalC (assay driven cutoff with assay sensitivity of 0.4 nmol/liter)
Ness-Abramof, 2002 (29)	No suspicion[a]	42.9 (26–69)	86 (85)	5 (6, 60, 20, 0, 20)	0 (0)	0 (0)	Path and clinical	NR	UFC (immunoassay, single assay-driven cutoff), ODST (assay-driven cutoff with assay sensitivity of 5.8 nmol/liter), LDDST (assay-driven cutoff with assay sensitivity of 5.8 nmol/liter)
Papanicolaou, 2002 (31)	Suspicion referral[d]	NR	156 (NR)	122(78.21, 80.33, 9.84, 9.84, 0)	4 (2.6)	0 (0)	Path and clinical	NR	UFC (immunoassay, single outcome-driven cutoff), MSerC (outcome-driven cutoff), MSalC (outcome-driven cutoff)
Catargi, 2003 (17)	No suspicion[a]	58.6 (22–84)	200 (75.5)	11 (5.5, 27.3, 72.7, 0, 0)	3 (1.5)	NR	Path	NR	ODST (assay-driven cutoff)
Omura, 2004 (30)	No suspicion	NR	1020 (NR)	11 (1.1, 45.5, 54.5, 0, 0)	0 (0)	NR	Path	NR	ODST (assay-driven cutoff)
Holleman, 2005 (23)	Suspicion referral	40 (17–76)	144 (78)	17 (12, 47, 29, 24, 0)	0 (0)	10 (6.9)	Path and clinical	41.2	UFC (liquid chromatography, ROC/multiple outcome-driven cutoffs), ODST (assay-driven cutoff with assay sensitivity of 50 nmol/liter)
Liu, 2005 (26)[b]	No suspicion	61.8	141 (0)	0 (0, 0, 0, 0, 0)	0 (0)	1 (0.7)	Path and clinical	NR	UFC, MSalC, ODST, LDDST
Reimondo, 2005 (36)	Suspicion referral	NR	106 (71.7)	78 (73.6, 56.4, 23.1, 19.2, 1)	0 (0)	0 (0)	Path and clinical	12	UFC (ROC/multiple outcome-driven cutoffs), MSerC (outcome-driven cutoff), ODST (outcome-driven),
Viardot, 2005 (39)	Suspicion referral	48.8 (18–68)	26 (69.23)	12 (46.2, 41.67, 25, 33.3, 0)	0 (0)	0 (0)	Path and clinical	>6	UFC (RIA, ROC/multiple outcome-driven cutoffs), MSalC (outcome-driven cutoff with assay sensitivity of 0.8 nmol/liter), ODST (outcome driven cutoff)
Martin, 2006 (27)	Suspicion referral[d]	44 (17–77)	36 (61)	12 (33.3, 66.6, 33.3, 0, 0)	0 (0)	0 (0)	Path and clinical	12	LDDST (assay-driven cutoff with assay sensitivity of 15 nmol/liter)
Erickson, 2007 (21)	Suspicion referral	45	51 (72.5)	21 (41, 100, 0, 0, 0)	0 (0)	15 (27.7)	Path and clinical	11.5–13.5	UFC (liquid chromatography, ROC/multiple outcome-driven cutoffs)
Friedman, 2007 (22)	Suspicion referral	36.1	87 (96)	24 (27.6, 100, 0, 0, 0)	0 (0)	35 (40.2)	Path and clinical	>12	UFC (liquid chromatography, single assay-driven cutoff), MSerC (assay-driven cutoff), MSalC (assay-driven cutoff)
Giraldi, 2007 (33)	Suspicion referral	41.7 (13–92)	4126(76.3)	22 (0.5, 90.9, 9.1, 0, 0)	0 (0)	0 (0)	Path and clinical	29	UFC (Immunoassay, ROC/ multiple outcome driven cutoffs), MSerC (outcome driven cutoff), ODST (outcome driven cutoff), UFC+ODST
Giraldi, 2007 (34)	Suspicion referral[d]	36.6 (12–65)	55 (83.6)	32 (58.2, 91, 9, 0, 0)	0 (0)	0 (0)	Path and clinical	37	UFC (immunoassay, single assay-driven cutoff), MSerC (assay-driven cutoff with assay sensitivity of 13.5 nmol/liter), ODST (assay-driven cutoff with assay sensitivity of 13.5 nmol/liter), LDDST (assay-driven cutoff with assay sensitivity of 13.5 nmol/liter)
Reimondo, 2007 (37)	No suspicion	61 (30–87)	99 (37)	1 (1, 100, 0, 0, 0)	0 (0)	1 (1)	Path and clinical	NR	ODST (assay-driven cutoff)
Reinehr, 2007 (38)	No suspicion	11.9 (5–16)	1405 (52)	1 (0.07, 0, 100, 0, 0)	0 (0)	0 (0)	Path	NR	LDDST (assay-driven cutoff)

For cohort characteristics, suspicion referral indicates clinicians referred patients for further evaluation for CS, and suspicion nonreferral indicates clinicians suspected CS because of history (of diabetes or hypertension), physical examination (central obesity, easy bruising, striae, cervical or supraclavicular fat pad), or laboratory findings. For CS definition pathological finding refers to a pituitary tumor or other tumor that stained for ACTH or cortisol, and clinical indicates clinical and laboratory follow-up leading to overt syndrome (postoperative adrenal crisis or adrenal insufficiency, need for steroid replacement, follow-up confirmation of Cushing through symptoms, signs, or tests) or rule out of the condition. CD, Cushing disease; LDDST, 2-d 2-mg DST; MSalC, midnight salivary cortisol; MSerC, midnight serum cortisol; ODST, 1-mg overnight DST; Path, pathological finding.

[a] Milder CS cases with mean cortisol levels less than the median value across the studies.

[b] Excluded from analysis.

[c] Outcome-driven cutoff refers to investigators setting cutoffs maximizing sensitivities, specificities, or both.

[d] More severe CS cases with mean cortisol levels more than the median value across studies.

TABLE 2. General quality assessment of studies of diagnostic accuracy included in systematic reviews (QUADAS)

	Author, year (Ref.)									
	Eddy, 1973 (20)	**Barbarino, 1979 (16)**	**Ashcraft, 1982 (15)**	**Meikle, 1982 (28)**	**Kreze, 1983 (24)**	**Vidal Trecan, 1983 (40)**	**Dunlap, 1985 (19)**	**Cronin, 1990 (18)**	**Yanovski, 1993 (41)**	**Leibowitz, 1996 (25)**
General QUADAS										
1. Was the spectrum of patients representative of the patients who will receive the test in practice?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
2. Were selection criteria clearly established?	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
3. Is the reference standard likely to correctly classify the target condition?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
5. Did the whole sample or a random selection of the sample receive verification of Cushing syndrome using a reference standard of diagnosis?	Y	N	U	Y	N	Y	Y	N	Y	N
6. Was the execution of the reference standard described in sufficient detail to permit its replication?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
7. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
8. Were withdrawals from the study explained?	Y̲	Y̲	U	N	Y̲	Y̲	Y̲	Y̲	Y̲	Y̲
UFC test-specific QUADAS										
1. Did patients receive the same reference standard regardless of test result?	Y	NR	NR	NR	N	Y	N	NR	NR	N
2. Was the reference standard independent of the index test?	Y	NR	NR	NR	Y	Y	Y	NR	NR	Y
3. Was the execution of the test described in sufficient detail to permit replication of the test?	Y	NR	NR	NR	Y	Y	Y	NR	NR	Y
4. Were the index test results interpreted without the knowledge of the results of the reference standard?	Y	NR	NR	NR	Y	Y	Y	NR	NR	Y
5. Were the reference standard results interpreted without knowledge of the results of the index test?	Y	NR	NR	NR	U	N	U	NR	NR	N
6. Were uninterpretable/intermediate test results reported?	U	NR	NR	NR	N	N̲	Y	NR	NR	Y
Midnight serum cortisol test-specific QUADAS										
1. Did patients receive the same reference standard regardless of test result?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
2. Was the reference standard independent of the index test?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
3. Was the execution of the test described in sufficient detail to permit replication of the test?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
4. Were the index test results interpreted without the knowledge of the results of the reference standard?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
5. Were the reference standard results interpreted without knowledge of the results of the index test?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
6. Were uninterpretable/intermediate test results reported?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Midnight salivary cortisol test-specific QUADAS										
1. Did patients receive the same reference standard regardless of test result?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
2. Was the reference standard independent of the index test?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
3. Was the execution of the test described in sufficient detail to permit replication of the test?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
4. Were the index test results interpreted without the knowledge of the results of the reference standard?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
5. Were the reference standard results interpreted without knowledge of the results of the index test?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
6. Were uninterpretable/intermediate test results reported?	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
1-mg overnight DST-specific QUADAS										
1. Did patients receive the same reference standard regardless of test result?	Y	N	NR	N	NR	NR	NR	N	NR	N
2. Was the reference standard independent of the index test?	Y	Y	NR	Y	NR	NR	NR	N	NR	N
3. Was the execution of the test described in sufficient detail to permit replication of the test?	Y	Y	NR	Y	NR	NR	NR	Y	NR	Y
4. Were the index test results interpreted without the knowledge of the results of the reference standard?	Y	U	NR	Y	NR	NR	NR	Y	NR	Y
5. Were the reference standard results interpreted without knowledge of the results of the index test?	Y	U	NR	U	NR	NR	NR	N	NR	N
6. Were uninterpretable/intermediate test results reported?	Y	Y	NR	Y	NR	NR	NR	Y	NR	Y
2-d 2-mg DST-specific QUADAS										
1. Did patients receive the same reference standard regardless of test result?	Y	NR	U	NR	N	NR	N	NR	N	NR
2. Was the reference standard independent of the index test?	Y	NR	N	NR	Y	NR	Y	NR	Y	NR
3. Was the execution of the test described in sufficient detail to permit replication of the test?	Y	NR	Y	NR	N	NR	Y	NR	Y	NR
4. Were the index test results interpreted without the knowledge of the results of the reference standard?	Y	NR	U	NR	Y	NR	Y	NR	Y	NR
5. Were the reference standard results interpreted without knowledge of the results of the index test?	Y	NR	N	NR	U	NR	U	NR	N	NR
6. Were uninterpretable/intermediate test results reported?	Y	NR	N	NR	Y	NR	Y	NR	N	NR

(*Table continues*)

N, No; N̲, there were no uninterpretable or indeterminate results; NR, test was not reported; U, unclear; Y, yes; Y̲, yes, there were no withdrawals.

TABLE 2. (*Continued*)

Author, year (Ref.)																
Papanicolaou, 1998 (32)	Raff, 1998 (35)	Ness-Abramof, 2002 (29)	Papanicolaou, 2002 (31)	Catargi, 2003 (17)	Omura, 2004 (30)	Holleman, 2005 (23)	Liu, 2005 (26)	Reimondo, 2005 (36)	Viardot, 2005 (39)	Martin, 2006 (27)	Erickson, 2007 (21)	Friedman, 2007 (22)	Giraldi, 2007 (33)	Giraldi, 2007 (34)	Reimondo, 2007 (37)	Reinehr, 2007 (38)
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Y	U	Y	U	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Y	Y	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	Y	Y	N	N
Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	U
Y	NR	N	N	N	NR	Y	N	N	Y	N	Y	N	N	N	NR	NR
N	NR	N	Y	N	NR	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR
Y	NR	Y	Y	Y	NR	Y	Y	Y	Y	Y	Y	Y	Y	Y	NR	NR
Y	NR	Y	N	Y	NR	Y	Y	U	Y	Y	U	Y	U	Y	NR	NR
N	NR	N	N	N	NR	Y	U	U	N	U	U	U	N	N	NR	NR
U	NR	U	Y	N	NR	N	Y	Y	N	U	U	N	U	Y	NR	NR
Y	NR	NR	Y	N	NR	NR	NR	N	NR	NR	NR	N	N	N	NR	NR
Y	NR	NR	Y	N	NR	NR	NR	Y	NR	NR	NR	Y	Y	Y	NR	NR
Y	NR	NR	Y	Y	NR	NR	NR	Y	NR	NR	NR	Y	N	Y	NR	NR
Y	NR	NR	N	Y	NR	NR	NR	U	NR	NR	NR	Y	U	Y	NR	NR
N	NR	NR	N	N	NR	NR	NR	U	NR	NR	NR	U	N	N	NR	NR
N	NR	NR	Y	Y	NR	NR	NR	Y	NR	NR	NR	N	U	Y	NR	NR
NR	Y	NR	Y	NR	NR	NR	N	NR	Y	NR	NR	N	NR	NR	NR	NR
NR	Y	NR	Y	NR	NR	NR	Y	NR	Y	NR	NR	Y	NR	NR	NR	NR
NR	Y	NR	Y	NR	NR	NR	Y	NR	Y	NR	NR	Y	NR	NR	NR	NR
NR	N	NR	N	NR	NR	NR	U	NR	Y	NR	NR	Y	NR	NR	NR	NR
NR	N	NR	N	NR	NR	NR	U	NR	Y	NR	NR	U	NR	NR	NR	NR
NR	Y	NR	Y	NR	NR	NR	N	NR	N	NR	NR	N	NR	NR	NR	NR
NR	NR	Y	NR	N	N	Y	N	N	Y	NR	NR	NR	N	N	N	NR
NR	NR	Y	NR	Y	Y	Y	Y	Y	Y	NR	NR	NR	Y	Y	N	NR
NR	NR	Y	NR	Y	Y	Y	Y	Y	Y	NR	NR	NR	N	Y	Y	NR
NR	NR	Y	NR	Y	N	Y	Y	U	Y	NR	NR	NR	U	Y	N	NR
NR	NR	U	NR	N	N	Y	U	U	Y	NR	NR	NR	N	N	N	NR
NR	NR	N	NR	Y	Y	N	Y	Y	N	NR	NR	NR	U	Y	N	NR
NR	NR	N	NR	NR	NR	NR	N	NR	NR	N	NR	NR	NR	N	NR	N
NR	NR	N	NR	NR	NR	NR	Y	NR	NR	Y	NR	NR	NR	Y	NR	N
NR	NR	Y	NR	NR	NR	NR	Y	NR	NR	Y	NR	NR	NR	Y	NR	N
NR	NR	Y	NR	NR	NR	NR	Y	NR	NR	Y	NR	NR	NR	Y	NR	Y
NR	NR	Y	NR	NR	NR	NR	U	NR	NR	U	NR	NR	NR	N	NR	U
NR	NR	Y	NR	NR	NR	NR	Y	NR	NR	Y	NR	NR	NR	Y	NR	N

TABLE 3. Summary of pooled results

Diagnostic test	LR positive test (95% CI)	LR negative test (95% CI)	Diagnostic OR (95% CI)[a]	I^2 (%)
Individual tests (n = no. of studies)				
UFC (n = 14)				
Pooled results	10.6 (5.5–20.5)	0.16 (0.08–0.33)	95.4 (37.8–240.3)	44
Midnight serum cortisol (n = 6)				
Pooled results	9.5 (1.7–54.1)	0.09 (0.03–0.28)	122.1 (15.3–974.6)	78
Assay driven (n = 2)	1.8 (0.5–6.9)	0.47 (0.23–0.96)	6.47 (1.6–26.6)	0
Outcome driven (n = 4)	26.6 (0.9–768.5)	0.05 (0.03–0.08)	581.11 (155.7–2169.5)	0[a]
Midnight salivary cortisol (n = 4)				
Pooled results	8.8 (3.5–21.8)	0.07 (0.00–1.20)	165.4 (26.9–1015.0)	50
1-mg overnight DST (n = 14)				
Pooled results	11.6 (5.8–23.1)	0.09 (0.05–0.14)	146.6 (67.8–316.9)	11
<50% had CS (n = 11)	16.4 (9.3–28.8)	0.06 (0.03–0.14)	328.7 (125.9–857.9)	0
>50% had CS (n = 3)	2.8 (1.3–6.3)	0.11 (0.06–0.19)	48.1 (16.9–136.3)	0[b]
2-day 2 mg DST (n = 8)				
Pooled results	7.3 (3.6–15.2)	0.18 (0.06–0.52)	51.6 (20.0–133.3)	0
Test combinations[c]				
UFC + 1-mg overnight DST (n = 3)				
Pooled results	15.4 (0.7–358.0)	0.11 (0.007–1.57)	149.4 (1.3–16811.5)	90
UFC + Midnight serum cortisol (n = 1)				
Pooled results	73.0 (29.1–183.2)	0.02 (0.001–0.34)	3315 (173–63513)	NA
UFC + 1-mg overnight DST + midnight serum cortisol (n = 1)				
Pooled results	174.1 (11.0–2764.2)	0.02 (0.001–0.34)	7965 (153.8–412492)	NA

CI, Confidence interval; LR, Likelihood ratio; NA, incalculable for less than three studies; OR, odds ratio.

[a] Subgroup-interaction test, $P = 0.000005$.

[b] Subgroup-interaction test, $P = 0.0008$.

[c] Judged positive when all included tests were positive.

25, 50, and 75% indicate low, moderate, and high heterogeneity, respectively (12).

Subgroup analyses

A priori hypotheses to explain potential heterogeneity among studies included severity of CS, selection bias (*i.e.* samples of consecutive patients with high prevalence of CS), type of patients (referred because of clinician's suspicion of CS *vs.* no CS suspicion), cutoff rationale (driven by outcomes in the same sample, *e.g.* chosen to maximize specificity, or by the upper limit of the assay), and tests characteristics (sensitivity of the assay, use of liquid chromatography *vs.* RIA). We tested these hypotheses using a test for interaction considering $P < 0.05$ as significant (13), because we did not have enough studies to conduct meta-regression (14).

Results

Study identification

Initial search of the literature yielded 1791 publications, of which 124 were potentially relevant to this review based on titles and abstracts (Fig. 1). After full text review, we found 27 eligible studies (15–41). We excluded one study from analyses because there were no CS cases in the sample (26) and excluded another study because we could not obtain essential data from the author (15).

We contacted all of the corresponding authors (another author in two studies) by electronic mail or, in five instances, by regular mail of which 70% were successfully contacted. Ninety percent of the authors successfully contacted either contributed missing data (where these data had been collected but not reported in the format we needed for analyses) or confirmed study characteristics, quality assessments, and data as collected.

Study characteristics

Table 1 summarizes the baseline characteristics of eligible studies. Fourteen studies assessed the diagnostic accuracy of UFC, six midnight serum cortisol, four midnight salivary cortisol, 14 the 1-mg overnight DST, and eight the 2-d 2 mg DST. Of 8631 patients enrolled in these studies, 794 (9.2%) had CS.

Study quality

Table 2 summarizes the methodological quality of the 27 included studies. Almost all studies enrolled patients with apparent diagnostic uncertainty of spectrum similar to the population in whom clinicians would use the tests in clinical practice (42). However, there is a strikingly broad range in the prevalence of CS across these studies, suggesting some degree of selection or referral bias. Their selection criteria were clearly described, and all received a reference standard that either diagnosed or excluded CS.

Metaanalyses

The appendix (published as supplemental data on The Endocrine Society's Journals Online web site at http://jcem.endojournals.org) includes tables with the test accuracy data from each included study (supplemental Tables 1–6). Table 3 shows pooled likelihood ratios for test results considered positive and negative. Table 3 also reports the diagnostic odds ratio and its associated inconsistency statistic (I^2). Where the subgroup analyses revealed a significant interaction, we report the effect in each of the subgroups in addition to the pooled estimates, because the latter may have less validity. [Supplemental Figs. 1–5 (published as supplemental data on The Endocrine Society's Journals Online web site at http://jcem.endojournals.org) show summary ROC curves for the tests of interest].

Although comparisons across tests require comparisons across studies that may have involved patients with a different spectrum of disease, the diagnostic odds ratio column in Table 3 can help readers identify tests with better discriminating power. Pooled sensitivities and specificities cannot be interpreted directly (because they were pooled independently, yet they are closely related) as if they were coming from a single study, and thus, we do not report them here. Instead, readers should focus on the likelihood ratio results; tests with a high likelihood ratio for a positive test indicate tests that can help rule in CS, and tests with a very low likelihood ratio for a negative test indicate tests that can help rule out CS. Figure 2 summarizes the likelihood ratio results in a Fagan nomogram (43). Clinicians can use this nomogram to estimate the posttest probability of CS using the pretest probability of CS and the pooled estimate and 95% confidence intervals for the likelihood ratios of the tests evaluated.

Subgroup analyses

Except where noted in Table 3, all other subgroup analyses explored were not associated with significant test accuracy-subgroup interactions (see supplemental Tables 7–11).

Sensitivity analyses

Most patients included in the metaanalysis were enrolled in a single study (33). A sensitivity analysis, in which we removed this study, revealed similar pooled accuracy results (data not shown).

Zwinderman and Bossuyt (44) have proposed the use of bivariate random-effects metaanalysis to analyze the sensitivities and specificities together from which one could derive pooled likelihood ratios, rather than pooling the likelihood ratios directly; in this data set, however, the bivariate approach yields results consistent with those presented here (data not shown).

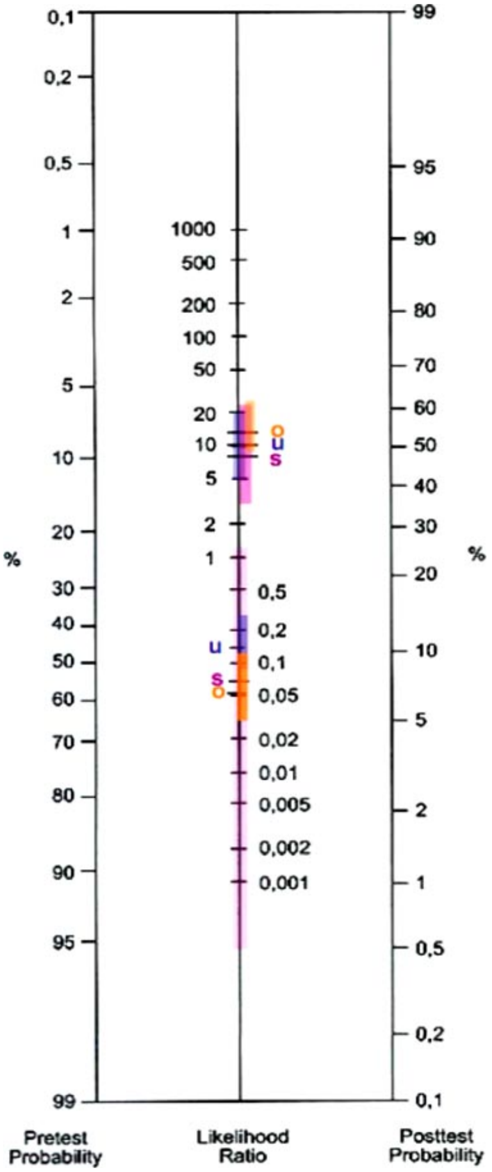

FIG. 2. Fagan nomogram summarizing the likelihood ratios of selected tests. Use a straight edge to link pretest probability of CS with the posttest probability by crossing the likelihood ratio line at a point that describes the results obtained. The *colored shadows* represent the 95% confidence interval around the likelihood ratios for each of the tests. o, Overnight DST; s, midnight salivary cortisol; u, UFC. Adapted from Fagan (43).

Discussion

Summary of findings

We conducted a systematic review and metaanalyses of studies that enrolled patients with diagnostic uncertainty and conducted a test for hypercortisolism and a satisfactory reference standard test. This review offers 1) a survey comprised of mostly small studies with high prevalence of CS from referral centers, 2) pooled test characteristics that represent the best estimates of test accuracy for each of the tests assessed and their combinations, and 3) inconsistent results across studies that are not explained by the choice of test thresholds but likely represent differences in the spectrum of patients with and without CS, in the characteristics of the tests used, and in the definitions of CS. These inconsistencies remain unexplained given the limitations in our ability to explore these differences with few studies.

In all, we found that the UFC and the overnight DST have the most evidence supporting their use for the detection of CS, with limited evidence supporting the use of salivary and serum midnight cortisol tests. Limited evidence also supports the use of these tests in combination to both identify and exclude patients with CS. In two instances in which the inconsistency across studies was important, we were able to identify potential explanations. For the midnight serum cortisol test compared with assay-driven thresholds, outcome-driven thresholds overestimated test accuracy (*i.e.* test interpretation was fitted to the data in the studies). For the 1-mg overnight DST, studies in which the prevalence of CS was greater than 50% (the median across studies) reported more modest test characteristics, especially more false-positive test results. This paradoxical result may be due to chance, to a lower cortisol threshold for positivity, or to patients without CS who had other syndromes associated with impaired cortisol suppression.

Limitations and strengths

The key limitations of this review refer to the relative paucity of evidence of test accuracy for the evaluated tests and to the methodological quality of the included studies. In particular, the prevalence and severity of CS varies importantly across studies despite the authors' representation of their populations as consecutive samples of patients referred without clear diagnosis. It is also striking that these studies rarely report indeterminate cases, given how often there is residual diagnostic uncertainty even among patients evaluated in centers of excellence. Finally, the report of a single cutoff in many of these studies precludes the estimation of likelihood ratios for ranges of test results. The arbitrary choice of test threshold and the dichotomy of the test results into positive and negative may contribute to a dichotomous view of diagnosis in which patients either have or do not have CS rather than a Bayesian approach in which additional test results modify the probability that a given patient has CS.

Incomplete searching, arbitrary study selection, poor quality of the primary studies, misguided analyses, and results that cannot be applied in practice represent potential limitations of systematic reviews. The extent to which publication bias affects studies of test accuracy is unknown, and the performance of tests of publication bias in the context of heterogeneous results is problematic (45); the accuracy of the indexing of such studies in the electronic databases is also unclear (46). Yet, our overlapping search strategies and extensive input from clinical experts should have minimized the chances that we missed studies that could substantially change the inferences drawn from this study.

Our review has the strengths of systematic reviews that summarize the totality of the available evidence following a protocol-driven procedure with explicit eligibility criteria, reproducible judgments about study quality and selection, and focused analyses (47). We also provide in the appendix the data from each of the studies to facilitate readers' secondary analyses. Given our focus on samples of patients in whom there was diagnostic uncertainty (phase II and III diagnostic studies) (7), we may have successfully ameliorated the overestimation of test accuracy that results from so-called phase I diagnostic accuracy studies in which investigators evaluate the accuracy of the test in distinguishing patients with clear confirmed disease and individuals who are clearly free of disease. We were forced to use a single cutoff when many were reported from a given study with the subsequent loss of information and gain in simplicity and transparency. Yet, our analyses take into account inconsistencies associated with the choice of threshold (*i.e.* using the diagnostic odds ratio).

Because of our study selection criteria, this review's results do not apply to patients with adrenal incidentaloma or to patients with suspected intermittent or so-called cyclical CS. Because of the high prevalence of CS in the included studies, the applicability of this study to general practice settings or to general endocrine practices is unclear.

With these limitations and strengths, clinicians seeking to apply these results in their practice can use a Fagan nomogram to update their estimates of the probability their patients have CS (Fig. 2). Given the close biological relationship between the tests assessed here, it may be unwise to use this procedure to estimate the posttest probability when several of these tests are performed in series.

Implications for practice and research

The accompanying Endocrine Society practice guideline on the diagnosis of CS contains the practical implications of the results of this review. The Task Force recommends a particular algorithm that seeks to balance diagnostic accuracy with practical and logistical considerations.

Our systematic review has uncovered several research gaps in this area. From the laboratory perspective, laboratory and test manufacturers should seek and maintain standards for measuring cortisol in urine, serum, and saliva. Variability today introduces variability in the literature and in clinical practice and impairs clinicians' ability to apply published cutoffs and results to their practice.

From the diagnostic accuracy perspective, prospective studies of the proposed algorithm may uncover further advantages and disadvantages of the proposed approach, including the downstream consequences of patient misclassification. Further work to evaluate the accuracy of testing algorithms in consecutive patients in whom clinical features suggest CS should 1) yield more accurate estimates of the diagnostic power of test results, 2) report findings using likelihood ratios for test result ranges rather than forcing a single cutoff on the data, and 3) use diagnostic categories that include those who clearly have and do not have CS and those with indeterminate results (48). Given the low incidence of CS and the increasing incidence of conditions with similar features (truncal obesity, bone loss, hyperglycemia, and hypertension), rigorous research is likely to yield more conservative estimates of test performance than those summarized here.

For stronger recommendations in the future, guideline panels will require evidence that patients are better off in important ways when they receive a diagnosis when the disease is subtle and mild rather than when it is florid and severe. The paucity of both patients and resources mandates collaboration across centers of excellence (*i.e.* endocrinologists with an interest in CS working in academic medical centers) tightly integrated with their referral sources (*i.e.* primary care and internal medicine clinicians) to generate this much-needed research evidence.

Conclusions

Commonly used tests to diagnose CS appear highly accurate, particularly when used in combination, in referral practices with samples enriched with patients with CS. Their performance in usual clinical practice remains unclear.

Acknowledgments

We are grateful to the authors of primary studies who responded to our requests for data confirmation and missing data (F. Cavagnini, R. L. Eddy, D. Erickson, T. Friedman, W. E. Grizzle, F. Holleman, G. Leibowitz, H. Liu, K. Meeran, A. W. Meikle, R. Ness-Abramof, H. Raff, G. Reimondo, T. Reinehr, A. Viardot, and G. Vidal Trecan). We are also grateful to the members of The Endocrine Society Task Force on Cushing's Syndrome for their expert input into the conduct and interpretation of our review.

Address all correspondence and requests for reprints to: Victor M. Montori, M.D., M.Sc., Mayo Clinic, W18A, 200 First Street SW, Rochester, Minnesota 55905. E-mail: montori.victor@mayo.edu.

This work was supported by a contract from The Endocrine Society.

Disclosure Statement: M.B.E., M.H.M., R.M., D.E., K.H., S.N., R.E., P.J.E., and V.M.M. have nothing to declare.

References

1. **Lindholm J, Juul S, Jorgensen JOL, Astrup J, Bjerre P, Feldt-Rasmussen U, Hagen C, Jorgensen J, Kosteljanetz M, Kristensen LO, Laurberg P, Schmidt K, Weeke J** 2001 Incidence and late prognosis of Cushing's syndrome: a population-based study. J Clin Endocrinol Metab 86:117–123
2. **Arnaldi G, Angeli A, Atkinson AB, Bertagna X, Cavagnini F, Chrousos GP, Fava GA, Findling JW, Gaillard RC, Grossman AB, Kola B, Lacroix A, Mancini T, Mantero F, Newell-Price J, Nieman LK, Sonino N, Vance ML, Giustina A, Boscaro M** 2003 Diagnosis and complications of Cushing's syndrome: a consensus statement. J Clin Endocrinol Metab 88:5593–5602
3. **Newell-Price J, Bertagna X, Grossman AB, Nieman LK** 2006 Cushing's syndrome. Lancet 367:1605–1617
4. **Nieman LK, Ilias I** 2005 Evaluation and treatment of Cushing's syndrome. Am J Med 118:1340–1346
5. **Raff H, Findling JW** 2003 A physiologic approach to diagnosis of the Cushing syndrome. Ann Intern Med 138:980–991
6. **Deville WL, Buntinx F, Bouter LM, Montori VM, de Vet HC, van der Windt DA, Bezemer PD** 2002 Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2:9
7. **Sackett DL, Haynes RB** 2002 The architecture of diagnostic research. BMJ 324:539–541
8. **Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J** 2003 The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3:25
9. **Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A** 2006 Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 6:31
10. **Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM** 2003 The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 56:1129–1135
11. **Deeks JJ** 2001 Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. BMJ 323:157–162
12. **Higgins JP, Thompson SG, Deeks JJ, Altman DG** 2003 Measuring inconsistency in meta-analyses. BMJ 327:557–560
13. **Altman DG, Bland JM** 2003 Interaction revisited: the difference between two estimates. BMJ 326:219
14. **Lijmer JG, Bossuyt PM, Heisterkamp SH** 2002 Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 21:1525–1537
15. **Ashcraft MW, Van Herle AJ, Vener SL, Geffner DL** 1982 Serum cortisol levels in Cushing's syndrome after low- and high-dose dexamethasone suppression. Ann Intern Med 97:21–26
16. **Barbarino A, de Marinis L, Liberale I, Menini E** 1979 Evaluation of steroid laboratory tests and adrenal gland imaging with radiocholesterol in the aetiological diagnosis of Cushing's syndrome. Clin Endocrinol (Oxf) 10:107–121
17. **Catargi B, Rigalleau V, Poussin A, Ronci-Chaix N, Bex V, Vergnot V, Gin H, Roger P, Tabarin A** 2003 Occult Cushing's syndrome in type-2 diabetes. J Clin Endocrinol Metab 88:5808–5813
18. **Cronin C, Igoe D, Duffy MJ, Cunningham SK, McKenna TJ** 1990 The overnight dexamethasone test is a worthwhile screening procedure. Clin Endocrinol (Oxf) 33:27–33
19. **Dunlap NE, Grizzle WE, Siegel AL** 1985 Cushing's syndrome. Screening methods in hospitalized patients. Arch Pathol Lab Med 109:222–229
20. **Eddy RL, Jones AL, Gilliland PF, Ibarra JD, Jr., Thompson JQ, MacMurry Jr JF** 1973 Cushing's syndrome: a prospective study of diagnostic methods. Am J Med 55:621–630
21. **Erickson D, Natt N, Nippoldt T, Young Jr WF, Carpenter PC, Petterson T, Christianson T** 2007 Dexamethasone-suppressed corticotropin-releasing hormone stimulation test for diagnosis of mild hypercortisolism. J Clin Endocrinol Metab 92:2972–2976
22. **Friedman TC, Zuckerbraun E, Lee ML, Kabil MS, Shahinian H** 2007 Dynamic pituitary MRI has high sensitivity and specificity for the diagnosis of mild Cushing's syndrome and should be part of the initial workup. Horm Metab Res 39:451–456
23. **Holleman F, Endert E, Prummel MF, van Vessem-Timmermans M, Wiersinga WM, Fliers E** 2005 Evaluation of endocrine tests. B: screening for hypercortisolism. Neth J Med 63:348–353
24. **Kreze A, Veleminsky J, Spirova E** 1983 A follow-up of the "low dose suppressible" hypercortisolism. Endocrinol Exp 17:119–123
25. **Leibowitz G, Tsur A, Chayen SD, Salameh M, Raz I, Cerasi E, Gross DJ** 1996 Pre-clinical Cushing's syndrome: an unexpected frequent cause of poor glycaemic control in obese diabetic patients. Clin Endocrinol (Oxf) 44:717–722
26. **Liu H, Bravata DM, Cabaccan J, Raff H, Ryzen E** 2005 Elevated late-night salivary cortisol levels in elderly male type 2 diabetic veterans. Clin Endocrinol (Oxf) 63:642–649
27. **Martin NM, Dhillo WS, Banerjee A, Abdulali A, Jayasena CN, Donaldson M, Todd JF, Meeran K** 2006 Comparison of the dexamethasone-suppressed corticotropin-releasing hormone test and low-dose dexamethasone suppression test in the diagnosis of Cushing's syndrome. J Clin Endocrinol Metab 91:2582–2586
28. **Meikle AW** 1982 Dexamethasone suppression tests: usefulness of simultaneous measurement of plasma cortisol and dexamethasone. Clin Endocrinol (Oxf) 16:401–408
29. **Ness-Abramof R, Nabriski D, Apovian CM, Niven M, Weiss E, Shapiro MS, Shenkman L** 2002 Overnight dexamethasone suppression test: a reliable screen for Cushing's syndrome in the obese. Obes Res 10:1217–1221
30. **Omura M, Saito J, Yamaguchi K, Kakuta Y, Nishikawa T** 2004 Prospective study on the prevalence of secondary hypertension among hypertensive patients visiting a general outpatient clinic in Japan. Hypertens Res 27:193–202
31. **Papanicolaou DA, Mullen N, Kyrou I, Nieman LK** 2002 Nighttime salivary cortisol: a useful test for the diagnosis of Cushing's syndrome. J Clin Endocrinol Metab 87:4515–4521
32. **Papanicolaou DA, Yanovski JA, Cutler GB, Jr., Chrousos GP, Nieman LK** 1998 A single midnight serum cortisol measurement distinguishes Cushing's syndrome from pseudo-Cushing states. J Clin Endocrinol Metab 83:1163–1167
33. **Pecori Giraldi F, Ambrogio AG, De Martin M, Fatti LM, Scacchi M, Cavagnini F** 2007 Specificity of first-line tests for the diagnosis of Cushing's syndrome: assessment in a large series. J Clin Endocrinol Metab 92:4123–4129
34. **Pecori Giraldi F, Pivonello R, Ambrogio AG, De Martino MC, De Martin M, Scacchi M, Colao A, Toja PM, Lombardi G, Cavagnini F** 2007 The dexamethasone-suppressed corticotropin-releasing hormone stimulation test and the desmopressin test to distinguish Cushing's syndrome from pseudo-Cushing's states. Clin Endocrinol (Oxf) 66:251–257
35. **Raff H, Raff JL, Findling JW** 1998 Late-night salivary cortisol as a screening test for Cushing's syndrome. J Clin Endocrinol Metab 83:2681–2686
36. **Reimondo G, Allasino B, Bovio S, Paccotti P, Angeli A, Terzolo M** 2005 Evaluation of the effectiveness of midnight serum cortisol in the diagnostic procedures for Cushing's syndrome. Eur J Endocrinol 153:803–809
37. **Reimondo G, Pia A, Allasino B, Tassone F, Bovio S, Borretta G, Angeli A, Terzolo M** 2007 Screening of Cushing's syndrome in adult patients with newly diagnosed diabetes mellitus. Clin Endocrinol (Oxf) 67:225–229
38. **Reinehr T, Hinney A, de Sousa G, Austrup F, Hebebrand J, Andler W** 2007 Definable somatic disorders in overweight children and adolescents. J Pediatr 150:618–622, 622.e1–e5
39. **Viardot A, Huber P, Puder JJ, Zulewski H, Keller U, Muller B** 2005 Reproducibility of nighttime salivary cortisol and its use in the diagnosis of hypercortisolism compared with urinary free cortisol and overnight dexamethasone suppression test. J Clin Endocrinol Metab 90:5730–5736
40. **Vidal Trecan G, Laudat MH, Thomopoulos P, Luton JP, Bricaire H** 1983 Urinary free corticoids: an evaluation of their usefulness in the diagnosis of Cushing's syndrome. Acta Endocrinol (Copenh) 103:110–115
41. **Yanovski JA, Cutler GB, Jr., Chrousos GP, Nieman LK** 1993 Corticotropin-releasing hormone stimulation following low-dose dexamethasone administration. A new test to distinguish Cushing's syndrome from pseudo-Cushing's states. JAMA 269:2232–2238
42. **Montori VM, Wyer P, Newman TB, Keitz S, Guyatt G** 2005 Tips for learners of evidence-based medicine. 5. The effect of spectrum of disease on the performance of diagnostic tests. CMAJ 173:385–390
43. **Fagan TJ** 1975 Letter: nomogram for Bayes theorem. N Engl J Med 293:257
44. **Zwinderman AH, Bossuyt PM** 2007 We should not pool diagnostic likelihood ratios in systematic reviews. Stat Med 27:687–697
45. **Deeks JJ, Macaskill P, Irwig L** 2005 The performance of tests of publication

bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58:882–893

46. **Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM** 2006 Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J Clin Epidemiol 59:234–240
47. **Montori VM, Guyatt GH** 2003 Summarizing studies of diagnostic test performance. Clin Chem 49:1783–1784
48. **Montori VM, Guyatt GH** 2003 Evidence-based medicine and the diagnostic process. In: Price C, Christenson R, eds. Evidence-based laboratory medicine. Washington, DC: AACC Press; 1–19