



# Development of diagnostic algorithm for Cushing's syndrome: a tertiary centre experience

A. Efthymiadis<sup>1</sup> · H. Loo<sup>1</sup> · B. Shine<sup>2</sup> · T. James<sup>2</sup> · B. Keevil<sup>3</sup> · J. W. Tomlinson<sup>1</sup> · A. Pal<sup>1</sup> · R. Pofi<sup>1</sup>

Received: 10 January 2024 / Accepted: 27 February 2024  
© The Author(s) 2024

## Abstract

**Purpose** No consensus exists as the gold standard for Cushing's Syndrome (CS) screening. This study aimed to evaluate the diagnostic accuracy and utility of late-night salivary cortisol (LNSC) and cortisone (LNSE), overnight dexamethasone suppression test (ODST), and urinary free cortisol (UFC) in developing a screening algorithm for CS.

**Methods** A retrospective, single-centre analysis on 93 adult patients referred to the Oxford Centre for Diabetes, Endocrinology, and Metabolism for CS evaluation (2017–2022). Data were analysed using binomial logistic regression and area under the receiver-operating curve (AUROC).

**Results** Fifty-three patients were diagnosed with CS. LNSC (sensitivity 87.5%, specificity 64.9%, AUC 0.76), LNSE (sensitivity 72.4%, specificity 85.7%, AUC 0.79), and ODST (sensitivity 94.7%, specificity 52.1%; AUC 0.74) demonstrated comparable effectiveness for CS diagnosis. Their combined application increased diagnostic accuracy (AUC 0.91). UFC was not statistically significant. Pre-test clinical symptom inclusion improved screening test performance (AUC LNSC: 0.83; LNSE: 0.84; ODST: 0.82). For CD diagnosis, LNSE + LNSC (AUC 0.95) outperformed ODST. Combining these with ACTH levels < 12.6 pmol/L perfectly distinguished MACS (AUC 1.00). ODST (AUC 0.76) exhibited superior performance (sensitivity 100.0%, specificity 52.2%) in MACS detection.

**Conclusions** LNSC, LNSE, and ODST are robust tools for CS screening, with their combined use offering the highest diagnostic precision. LNSE, especially when used with LNSC, is highly effective for CD diagnosis, exceeding ODST accuracy. ODST is preferable for MACS identification. Integrating ACTH levels markedly improves differentiation between CD and MACS. Conversely, UFC shows limited diagnostic utility.

**Keywords** Late night salivary cortisol · Late night salivary cortisone · Overnight dexamethasone test · 24 h urinary free cortisol · Cushing's Syndrome

## Introduction

Endogenous Cushing Syndrome (CS) refers to pathologic hypercortisolism and is associated with significant morbidity, mortality, and reduction in quality of life [1–4]. The

diagnosis of CS can often be challenging [5] as it is typically characterized by the presence of multiple symptoms (such as hypertension, diabetes, weight gain, or osteoporosis) which are very common in the general population. Clinical pre-test likelihood of CS should be evaluated [6] and screening conducted in those with either signs or symptoms of low discriminatory value (e.g., hypertension and osteoporosis) occurring at an unusually early age, or patients with clinical features of higher specificity for CS (e.g., easy bruising, facial plethora, proximal myopathy, and Striae rubrae).

Several diagnostic testing strategies have been proposed, and at least two clinical scores have been recently developed to identify patients deserving screening for CS [7, 8]; however, their value in everyday clinical practice remains uncertain, since validation studies for these scores are missing. To date, there is no consensus as to the gold standard screening test for

✉ R. Pofi  
riccardo.pofi@ocdem.ox.ac.uk

<sup>1</sup> Oxford Centre for Diabetes, Endocrinology and Metabolism, NIHR Oxford Biomedical Research Centre, University of Oxford, Churchill Hospital, Oxford, UK

<sup>2</sup> Department of Clinical Biochemistry, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

<sup>3</sup> Department of Clinical Biochemistry, Manchester University Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK

the diagnosis of CS, and the presence of at least two abnormal tests with high diagnostic accuracy is needed [9, 10].

The decision on which screening test to choose is influenced by a combination of factors, including the index of clinical suspicion for hypercortisolism, the suspected underlying diagnosis, and, importantly, local availability of specific tests. The latest consensus recommended that if CS is suspected any combination of overnight dexamethasone suppression test (ODST), urinary free cortisol (UFC), and late-night salivary cortisol (LNSC) tests can be helpful [9]. For patients with adrenal incidentalomas being evaluated for hypercortisolism, ODST is recommended as a first test, with consideration of additional UFC and/or LNSC measurements [9, 11]. The role of Late-Night Salivary Cortisone (LNSE) as a CS screening is still unclear. Moreover, the cost-effectiveness of various hypercortisolism screening strategies remains a subject of ongoing debate. This is particularly relevant given the increasing frequency of hypercortisolism screenings in low-risk populations (e.g., obese patients) and their use in the diagnostic workup for incidentally discovered adrenal lesions. The challenge lies in identifying a screening strategy that exhibits high diagnostic accuracy and acceptable costs but is simultaneously acceptable from the patient's perspective.

The aims of this study were: (a) to compare the ability of LNSC, LNSE, ONDT, and UFC as screening test for CS; (b) to evaluate the diagnostic performance of each test in distinguishing Cushing's disease (CD) or mild autonomous cortisol secretion (MACS) from patients without CS; (c) to suggest a screening algorithm.

## Materials and methods

### Patient selection

We retrospectively reviewed all consecutive adult patients referred at the Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM) to be evaluated for hypercortisolism and who had LNSC measured, from January 2017 to November 2022. Reasons to screen for hypercortisolism included an incidental diagnosis of adrenal adenoma, the presence of phenotypic characteristics of CS (i.e., dorsal fat pad, central adiposity, facial plethora, easy bruising, purple striae, hirsutism, and proximal myopathy) as well as combinations of hypertension, insulin-resistant diabetes, oligomenorrhea, osteoporosis, and mood disorders.

### Screening tests

We collected data for the following screening tests: LNSC, LNSE, ODST, UFC, and Low-Dose Dexamethasone Suppression Test (LDDST).

A cut off of 50 nmol/L (1.8 µg/dL) for 9 AM cortisol levels after dexamethasone (DEX) challenge was used to

interpret ODST (1 mg of DEX between 23:00–24:00 and subsequent 9am cortisol) and LDDT (0.5 mg of DEX 6-hourly for 48 h with cortisol checked at time 0' and 48 h). At least two 24-h urine samples were requested to the patients screened through UFC and values less than 135 nmol/L were considered normal. Total urine volume and creatinine clearance were measured to ensure adequate urine collection. Similarly, all the patients screened with LNSC and LNSE had at least two evaluations. Cut-off values of < 1.7 nmol/L for LNSC, < 18.0 nmol/L for LNSE were considered normal based on assay-specific reference ranges [12, 13]. We also collected data on ACTH levels performed as baseline sample for LDDST. In our institution, the latter is used either in lieu of ODST or where results of ODST are equivocal. The diagnosis of CD was confirmed through inferior petrosal sinus sampling and histology. Patients with adrenal lesions and a ODST above the 50 nmol/L cut-off were diagnosed as mild autonomous cortisol secretion (MACS) as per current guidelines [14] by experienced endocrinologists. Subjects in whom CS was ruled out based on the outcomes of the screening tests (ODST, UFC, LNSC, and LNSE) will be referred to as "*controls*" throughout the manuscript. Patients with proven CS will be referred to as "*cases*".

### Data collection

Electronic patient records were reviewed. Patient demographics and anthropometric measurements (i.e., age, sex, weight, and body mass index), relevant comorbidities (i.e., obesity, hypertension, insulin-resistant diabetes, osteoporosis, obstructive sleep apnoea, mood disorders, and hypokalaemia), relevant medications (i.e., steroids, hormonal replacement therapies), clinical signs, and symptoms suggestive of CS (dorsal fat pad, facial plethora, proximal myopathy, purple striae, hyperpigmentation, and fatigue) were collected.

### Laboratory analysis

As described previously, saliva cortisol/cortisone [12, 13] were measured by electrospray positive ion mode liquid chromatography tandem mass spectrometry. The lower limit of quantification was 0.46 nmol/L for salivary cortisol and 0.42 nmol/L for salivary cortisone. Between-batch imprecision for cortisol showed coefficient variations of 13.4% to 2.7% across a range of concentrations from 4.2 to 118 nmol/L. Between-batch imprecision for salivary cortisone showed coefficient variations of 8.6% to 2.3% across a range of concentrations from 5.0 to 130.9 nmol/L. Recovery was 93% and 96% for cortisol and cortisone, respectively. 20 Alpha and 20 beta dihydrocortisone showed baseline separation with cortisone and

did not interfere in the assay. Serum cortisol was measured through the Abbott Architect i2000 (Abbott Laboratories, Maidenhead, UK) which had between-batch coefficient of variation of 4.1% at 118 nmol/L, 2.8% at 427 nmol/L and 2.8% at 967 nmol/L. UFC was determined following dichloromethane solvent extraction and liquid chromatography tandem mass spectrometry. Between-batch precision was 3.2% at 144 nmol/L and 2.2% at 692 nmol/L. The lower limit of quantification was 16 nmol/L, and the upper limit of linearity was 1600 nmol/L. ACTH measurements were undertaken using an Immulite 2000 analyser (Siemens Healthineers, Frimley, UK), a solid-phase, two-site sequential chemiluminescent immunometric assay (CLIA). Method imprecision, expressed as CV% was 8.2% at 26.8 pmol/L and 4.9% at 375.7 pmol/L.

### Statistical analysis

Sensitivity, specificity, likelihood ratios, as well as predictive values were computed for each biochemical test (LNSC, LNSE, ODST, UFC, and LDDST). The  $\chi^2$  test was used to test statistical significance associations between binomial variables. A  $p < 0.05$  was considered indicative of a statistically significant difference. A binomial logistic regression (LR) was performed to ascertain the ability of each test in diagnosing CS, calculate odds ratios (OR) and extract predictive values to compute ROC curve analysis. Comparison between the AUCs of each test was made for patients having both tests (De Long et al. [15]) as well as with independent comparisons. A stepwise backward LR was run to assess the utility of clinical symptoms to predict the diagnosis of CS. Only the symptoms retaining statistically significant association at univariate analysis were included into the model as independent variables (hypertension, dorsal fat pad, facial plethora, striae rubrae, myopathy, and easy bruising). Independent meaningful correlation was defined by a  $p$  value  $< 0.05$  with the outcome of interest. The extracted predictive value from LR analysis on clinical symptoms was then used to refine the diagnostic accuracy of biochemical tests by combining the relative AUCs. The analysis was then stratified according to disease aetiology (either CD or MACS). ROC curve analysis was used to confirm optimal cut-offs stratified for disease aetiology using the raw data. To evaluate the ability of ACTH levels to predict CD vs MACS, the baseline ACTH levels of the LDDST were included into the binomial logistic regression model. Statistical analyses were performed using SPSS (version 29, Chicago, IL, USA), illustrations were made with GraphPad Prism 8.0 software package (GraphPad Software, Inc. La Jolla, CA, USA) and ROC curve comparisons were conducted using MedCalc software Ltd (Ostend, Belgium).

### Ethics

The study was approved by the Trust audit team and conducted and registered as a local audit of practice (reference number 8352).

### Results

A total of 93 subjects were included in the analysis. Fifty-three patients were diagnosed with CS (79% females, mean age  $56 \pm 16$  years, and mean body mass index  $31.7 \pm 7.2$  kg/m<sup>2</sup>), whereas the diagnosis was excluded in 40 subjects (mean age was  $51 \pm 18$  years, 72% females, and mean body mass index  $34.8 \pm 9.3$  kg/m<sup>2</sup>). Among patients with CS, 24 (47%) and 27 (53%) were diagnosed with MACS and CD, respectively. Two patients were diagnosed with ectopic disease and were excluded from the analysis.

### The performance of biochemical screening tests

Sensitivity and specificity of LNSC in distinguishing *cases* and *controls* were 87.5% (95% CI 73.2–95.8) and 71.1% (95% CI 73.2–95.8), respectively, and the NLR was 0.19, with good diagnostic accuracy (AUC 0.762, 95% CI 0.650–0.873,  $p < 0.001$ ) (Fig. 1a, Table 1). LNSE showed higher specificity (96.1%; 95% CI 80.4–99.9) but lower sensitivity (55.8%; 95% CI 37.9–72.8), a NLR of 0.46, with comparable diagnostic accuracy (AUC 0.791, 95% CI 0.660–0.921,  $p < 0.001$ ) (Fig. 1b, Table 1). The ODST also demonstrated similar performance (AUC 0.740, 95% CI 0.59–0.88;  $p = 0.02$ ), with sensitivity and specificity of 94.7% (95% CI 82.2–99.3) and 52.1% (95% CI 30.6–73.2) respectively, and a NLR of 0.10 (Fig. 1c, Table 1). On the contrary, UFC was not able to reach statistical significance in distinguishing *cases* from *controls* (AUC 0.59, 95% CI 0.43–0.76;  $p = 0.26$ ) (Fig. 1d, Table 1), showing low sensitivity (60.0%, 95% CI 43.3–75.1) and specificity (58%, 95% CI 32.9–81.6).

Comparison of AUC for LNSC, LNSE, and ODST demonstrated a negligible, non-significant, difference between areas ( $p > 0.05$  for all comparisons). However, considering that LNSC and LNSE can be performed at the same time, we combined the two tests' prediction and found a better performance AUC of 0.85 (95% CI 0.71–0.94,  $p < 0.001$ ) (Fig. 1f) in distinguishing *cases* and *controls* compared to the LNSC alone (delta AUC 0.09, 95% CI 0.01–0.17,  $p = 0.02$ ). Adding ODST prediction to the model further increased diagnostic accuracy (AUC of 0.91, 95% CI 0.77–0.98,  $p < 0.001$ ; delta AUC 0.08, 95% CI 0.01–0.17,  $p = 0.06$ ) with 1.7 nmol/L, 15.2 nmol/L and 50 nmol/L being the best cut-offs for LNSC, LNSE, and ODST, respectively. Interestingly, including the UFC to each test or each combination modelling,

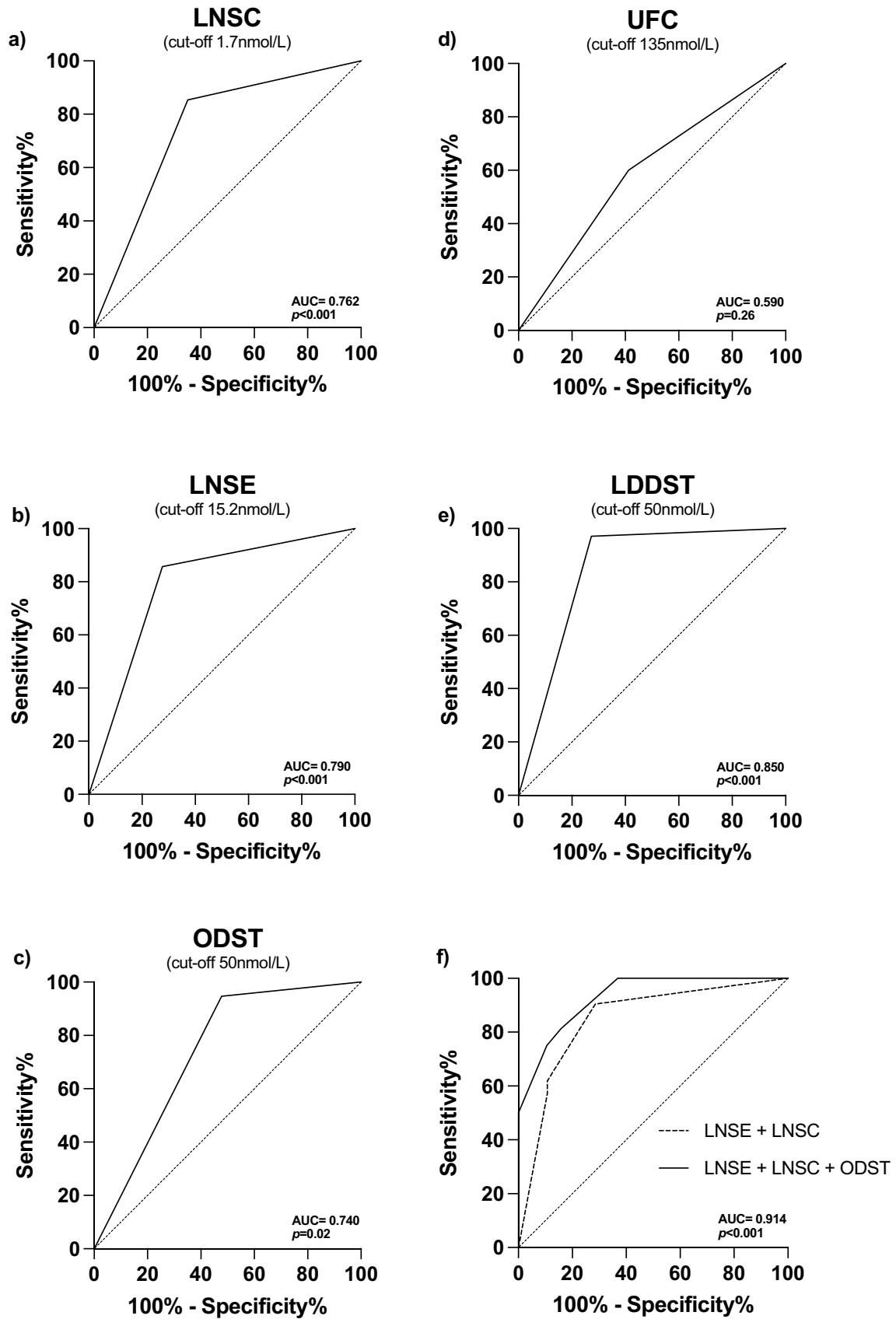


Fig. 1 ROC curves of commonly used screening test for the diagnosis of Cushing's syndrome

**Table 1** The diagnostic accuracy of screening tests stratified for disease aetiology

Test	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Positive likeli- hood ratio	Negative likeli- hood ratio	Positive predic- tive value	Negative predictive value
<b>Cases vs controls</b>							
LNSC	0.76 (0.65–0.87) <i>p</i> < 0.001	87.5% (73.2–95.8)	71.1% (73.2–95.8)	2.49	0.19	76.1%	84.3%
LNSE	0.79 (0.66–0.92) <i>p</i> < 0.001	55.8% (37.9–72.8)	96.1% (80.4–99.9)	14.53	0.46	95.0%	62.5%
ODST	0.74 (0.59–0.88) <i>p</i> = 0.020	94.7% (82.2–99.3)	52.1% (30.6–73.2)	1.98	0.10	76.6%	85.7%
LDDST	0.85 (0.69–1.0) <i>p</i> < 0.001	97.4% (86.5–99.9)	72.7% (39.0–94.0)	3.57	0.04	92.7%	88.9%
UFC	0.59 (0.43–0.76) <i>p</i> = 0.260	60.0% (43.3–75.1)	58.0% (32.9–81.6)	1.46	0.68	77.4%	38.5%
<b>CD vs controls</b>							
LNSC	0.82 (0.72–0.93) <i>p</i> < 0.001	100% (80.5–100)	64.9% (47.5–79.8)	2.85	0.00	56.7%	100%
LNSE	0.93 (0.83–1.00) <i>p</i> < 0.001	86.7% (59.5–98.3)	96.1% (80.4–99.9)	22.53	0.14	92.9%	92.6%
ODST	0.71 (0.55–0.87) <i>p</i> = 0.03	88.9% (65.3–96.6)	52.2% (30.6–73.2)	1.86	0.21	59.3%	85.7%
LDDST	0.86 (0.70–1.0) <i>p</i> = 0.001	100% (84.6–100)	72.7% (39.0–94.0)	3.67	0.00	88.0%	100%
UFC	0.75 (0.58–0.91) <i>p</i> = 0.008	90.9% (70.8–98.9)	58.8% (32.9–81.6)	2.21	0.15	74.1%	83.3%
<b>MACS vs controls</b>							
LNSC	0.71 (0.57–0.85) <i>p</i> = 0.007	77.2% (54.6–92.2)	64.8% (47.5–79.8)	2.20	0.3	56.7%	82.8%
LNSE	0.66 (0.47–0.85) <i>P</i> = 0.102	27.8% (9.7–53.5)	96.1% (80.4–99.9)	7.22	0.75	83.3%	65.8%
ODST	0.76 (0.66–1.00) <i>p</i> = 0.004	100% (82.4–100.0)	52.2% (30.6–73.2)	2.09	0.00	63.3%	100%
LDDST	0.83 (0.66–1.00) <i>p</i> = 0.004	93.8% (69.8–99.8)	72.7% (39.0–94.0)	3.44	0.09	83.3%	88.9%
UFC	0.62 (0.43–0.80) <i>p</i> = 0.242	17.7% (3.8–43.4)	58.8% (32.9–81.6)	0.43	1.4	30.0%	41.7%

LNSC late-night salivary cortisol, LNSE late-night salivary cortisone, ODST overnight dexamethasone test, LDDST low-dose (2 mg) dexamethasone test, UFC 24 h urinary free cortisol. CD Cushing's disease, MACS mild autonomous cortisol secretion

results in worsening of the diagnostic performance (data not shown). As expected [16], (being used as second screening test in our centre) LDDST had the highest sensitivity

(97.4%, 95% CI 86.5–99.9) and specificity (72.7%, 95% CI 39.0–94.0) and diagnostic accuracy (AUC 0.85, 95% CI 0.69–1.0; *p* < 0.001), with the lowest NLR 0.04 (Table 1,

Fig. 1d). However, there were no differences compared to the combined performance of LNSC + LNSE + ODST ( $p=0.49$ ).

**Table 2** Backward linear regression modelling of clinical important symptoms in 93 subjects screened for Cushing's syndrome

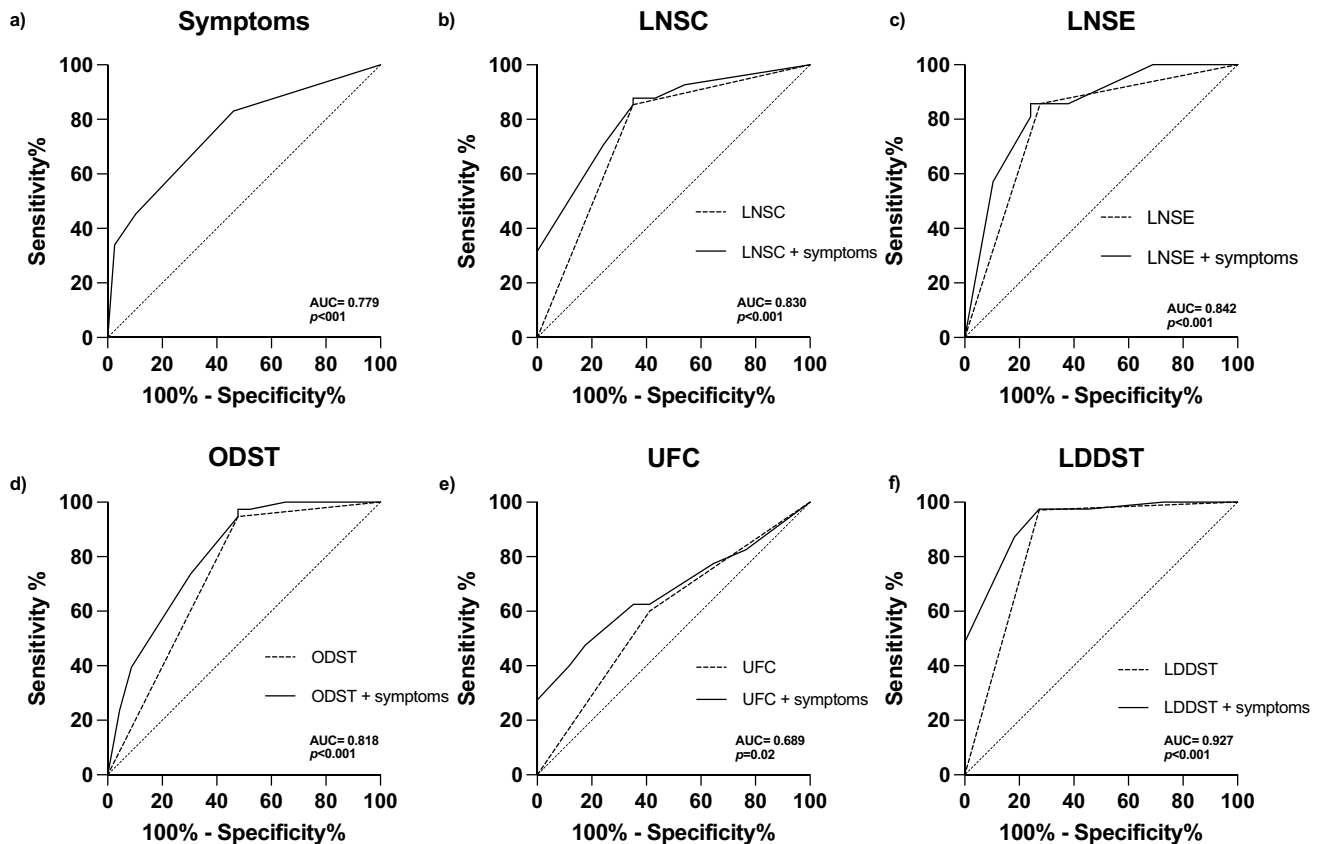
	<i>B</i>	<i>p</i>	Exp (B)	95% confidence interval	
				Lower	Upper
$\chi^2=29.72, p<0.001$ Step 1					
Hypertension	1.170	<b>0.023</b>	3.223	1.178	8.819
Striae rubrae	0.643	0.519	1.902	0.270	13.404
Facial plethora	1.758	0.069	5.801	0.874	38.483
Dorsal fat pad	0.783	0.354	2.188	0.418	11.449
Easy bruising	0.850	0.236	2.339	0.574	9.540
Myopathy	-1.038	0.274	0.354	0.055	2.272
Osteopenia	2.078	0.097	7.989	0.685	93.115
$\chi^2=25.81, p<0.001$ Step 5					
Hypertension	1.281	<b>0.005</b>	3.600	1.372	9.446
Facial plethora	1.875	<b>0.003</b>	6.518	1.902	22.332
Osteopenia	1.694	0.091	5.443	0.591	50.139

Significant *p*-values are highlighted in bold

## The importance of pre-test probability

Univariate analysis showed hypertension ( $\chi^2=10.2, p=0.001$ ), dorsal fat pad ( $\chi^2=11.9, p<0.001$ ), facial plethora ( $\chi^2=13.0, p<0.001$ ), striae rubrae ( $\chi^2=6.2, p=0.013$ ), myopathy ( $\chi^2=5.3, p=0.021$ ), and easy bruising ( $\chi^2=6.9, p=0.008$ ) as clinical important signs and symptoms associated with the diagnosis of CS. A backward multiple LR analysis was conducted to assess the contribution of each symptom in diagnosing CS ( $\chi^2=26.08, p<0.001$ ): hypertension (OR 3.6, 95% CI 1.53–10.21,  $p=0.005$ ) and facial plethora (OR 6.5, 95% CI 2.10–23.82,  $p=0.002$ ) where those with strongest prediction for CS (Table 2). ROC curve analysis confirmed good diagnostic performance of pre-test clinical symptoms (AUC 0.779, 95% CI 0.686–0.872,  $p<0.001$ ) (Fig. 2a). Patients presenting with both hypertension and facial plethora had 19 times higher chance of being diagnosed with CS (OR 19.54, 95% CI 2.47–154.15;  $p<0.001$ ).

Combining AUC of the pre-test clinical symptoms with those of the screening tests significantly improved diagnostic performance of LNSC (AUC 0.83, 95% CI 0.74–0.92,  $p<0.001$ ), LNSE (AUC 0.84, 95% CI 0.73–0.95,  $p<0.001$ ), ODST (0.82, 95% CI 0.71–0.93,  $p<0.001$ ), and LDDST (AUC 0.93, 95% CI 0.84–1.0,  $p<0.001$ ). Despite significant



**Fig. 2** Pre-test probability performance alone and in combination with screening tests for Cushing's syndrome



improvement in diagnostic performance of UFC with symptoms (AUC 0.689, 95% CI 0.552–0.826,  $p=0.025$ ), the results still did not reach good diagnostic accuracy (i.e.,  $AUC < 0.7$ ) (Fig. 2b–f).

### Subgroup analysis according to disease aetiology: CD

LNSE had the best sensitivity (100%, 95% CI 80.5–100.0) and NLR (0.00) compared to LNESC (86.7%, 95% CI 59.5–98.3; NLR 0.14) but lower specificity (LNESC 64.9%, 95% CI 47.5–79.8; LNESC 96.1%, 95% CI 80.4–99.9) in distinguishing between patients with CD and those who had hypercortisolism excluded ( $AUC_{LNESC}$  0.82, 95% CI 0.72–0.93,  $p < 0.001$ ;  $AUC_{LNESC}$  0.93, 95% CI 0.83–1.00,  $p < 0.001$ ). However, optimal cut-off derived by ROC analysis revealed lower threshold for LNESC of 15.2 nmol/L as best distinguishing between CD and controls, with significantly higher sensitivity (100%, 95% CI 78.2–100) and specificity (78.1%, 95% CI 60.0–90.7).

Interestingly, in this context, UFC reached statistical significance (AUC 0.75, 95% CI 0.58–0.91,  $p=0.008$ ) in distinguishing CD and controls with good sensitivity (90.9%, 95% CI 70.8–98.9), low specificity (58.8%, 95% CI 32.9–81.6), and acceptable NLR (0.15). Nevertheless, ODST still retained good sensitivity (88.9%, 95% CI 65.3–96.6) albeit with low specificity (52.2%, 95% CI 30.6–73.2) and NLR (0.21) in diagnosing CD (AUC 0.71, 95% CI 0.55–0.87,  $p=0.03$ ).

As for the whole cohort, LDDST had high sensitivity (100.0%, 95% CI 84.6–100.0) and specificity (72.7%, 95% CI 39.0–94.0), NLR (0.00) with good diagnostic performance (AUC of 0.86, 95% CI 0.70–1.0,  $p=0.001$ ). Independent AUCs comparisons showed LNESC to perform better than ODST and UFC in CD. All other comparisons did not show significant differences. As for the whole cohort, combining LNESC and LNESC result in a significant increase in diagnostic accuracy (AUC of 0.95, 95% CI 0.81–1.0,  $p < 0.001$ ), with 1.7 nmol/L and 15.2 nmol/L being the best cut-offs for LNESC and LNESC, respectively.

A summary of ROC analyses for the screening tests in distinguish patients with CD from controls is reported in *supplementary Fig. 1*.

### Subgroup analysis according to disease aetiology: MACS

The ODST was the best test in distinguishing between patients with MACS and those who had hypercortisolism excluded (AUC of 0.76, 95% CI 0.66–1.00,  $p=0.004$ ) with a sensitivity of 100.0% (95% CI 82.4–100.0), specificity of 52.2% (95% CI 30.6–73.2), and NLR of 0.00. LNESC was not as good as in CD when used to rule out MACS diagnosis:

the sensitivity was 77.2% (95% CI 54.6–92.2), the specificity 64.8% (95% CI 47.5–79.8), and NLR 0.3 (AUC 0.711, 95% CI 0.574–0.848,  $p=0.007$ ). Interestingly, LDDST confirmed high diagnostic performance AUC of 0.83 (95% CI 0.66–1.00,  $p=0.004$ ) also in this context, albeit with lower sensitivity of 93.8% (95% CI 69.8–99.8), higher specificity of 72.7% (95% CI 39.0–94.0), and NLR of 0.09 when compared to ODST. As per the whole cohort, UFC alone did not distinguish MACS and controls (AUC 0.62, 95% CI 0.43–0.80,  $p=0.242$ ), and the combination with any other test worsens relative diagnostic performance (data not shown). Interestingly, LNESC also did not reach statistical significance (AUC of 0.66, 95% CI 0.47–0.85,  $p=0.102$ ). There were no differences between the AUCs of the tests at independent comparisons.

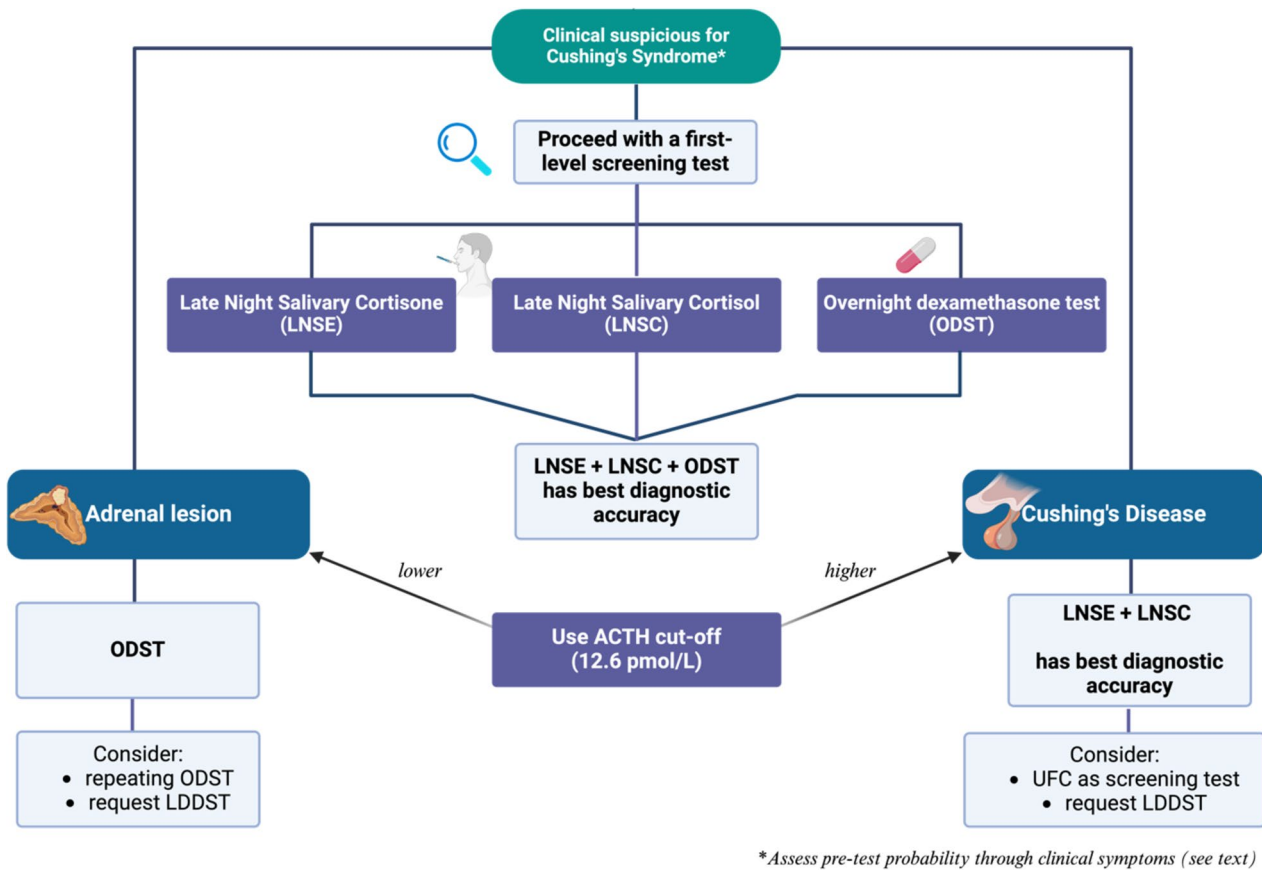
A summary of the diagnostic accuracy of the screening tests in distinguish patients with MACS from controls is reported in *supplementary Fig. 2*.

### The significance of ACTH levels in distinguishing CD from MACS

As expected, ACTH levels (median  $ACTH_{MACS}$  8.6 pmol/L, min–max 5.0–25.7 vs  $ACTH_{CD}$  57.2 pmol/L, min–max 13.6–273;  $p < 0.001$ ) were lower in patients with MACS as compared with those with CD. Baseline ACTH levels of LDDST were used as dependent variable in a logistic regression model to evaluate its ability in predicting the diagnosis of CD against MACS. The model was significant ( $\chi^2 = 33.78$ ,  $p < 0.001$ ), and ACTH levels were able to localize the disease ( $B = 0.258$ ,  $p = 0.017$ ). An ROC curve analysis (AUC 0.98, 95% CI 0.87–1.00,  $p < 0.001$ ) showed  $ACTH > 12.6$  pmol/L as cut-off distinguishing CD from MACS with 100% sensitivity, 86.7% specificity, and negative likelihood ratio (NLR) of 0.00. Interestingly, combining AUC of the ACTH levels with those of the LNESC + LNESC (as the one with the highest performance in diagnosing CD) improved overall diagnostic performance to an AUC of 1.00 (95% CI 1.00–1.00,  $p < 0.001$ ).

## Discussion

The diagnosis of CS is one of the most challenging in endocrinology. This is the first study comparing the diagnostic performance of five screening test used in the diagnosis of CS with the aim of describing the best combination of tests to be used as a screening strategy and according to disease aetiology. We show that LNESC, LNESC, and ODST individually offer comparable performance in screening for hypercortisolism. However, their combination significantly increases the overall diagnostic performance. In our study, UFC, which is widely used as a first-line screening



**Fig. 3** In-house proposed algorithm for Cushing's syndrome diagnosis

test in many countries due to its historical application and availability, showed limited diagnostic utility in detecting pathological hypercortisolism when considering all cases collectively. However, the subgroup analysis reveals that whilst UFC demonstrates moderate diagnostic ability in CD, it fails to effectively identify MACS. This outcome aligns with expectations, given that MACS is characterized by subclinical or mild cortisol excess, typically resulting in lower plasma and, consequently, free urinary cortisol levels. This is further reflected by the absence of overt Cushingoid features in MACS patients. Notably, adding the pre-test probability (assessed by clinical symptoms) to biochemical evaluation demonstrated superior diagnostic performance than each separate test alone, and matched the performance of LDDST, a finding not replicated with UFC. Of note, we found that test performance was dependent on aetiology of hypercortisolaemia; in CD, LNSE was most discriminatory whereas for investigation of adrenal incidentaloma and eventual diagnosis of MACS, ODST was the 'best' test. When LNSE and ODST are used in combination, the overall diagnostic performance increases significantly. Finally, ACTH levels performed well in confirming CD versus MACS.

Our study emphasises the importance of clinical context when deciding upon if, and how one should screen for Cushing's syndrome. The existence of clinical features of cortisol excess were clearly associated with diagnosis confirmation (we find hypertension and facial plethora to be independent predictors) and if pre-test probability is moderate to high, the use of LNSC, LNSE, and ODST represents a robust screening approach with 1.7 nmol/L, 15.2 nmol/L, and 50 nmol/L being the best cut-offs for each test, respectively. This challenges the need for the more cumbersome and time-consuming LDDST and UFC. Building upon previous smaller studies that reported increased diagnostic accuracy through the combination of LNSC and LNSE [17], our findings advocate for including ODST to enhance diagnostic precision further. In line with other studies and current guidelines [14, 18], we find that when investigating incidental adrenal lesions, ODST was the superior screening tool. Notably, the specificity of ODST can be further improved by measuring dexamethasone levels. It has been observed that about 6% of patients who do not demonstrate cortisol suppression during ODST actually have suboptimal dexamethasone levels [19]. This underscores the potential benefit of incorporating dexamethasone level measurement in ODST protocols.



Regrettably, data on dexamethasone levels were not available in our cohort.

Whilst adrenocorticotrophic hormone (ACTH) levels are known to help disease localisation, the optimal ACTH cut-off for distinguishing between adrenal and pituitary CS is still undefined. We found 12.6 pmol/L as the cut-off distinguishing the two aetiologies with high sensitivity and specificity. Previous studies have proposed similar but not identical cut-offs [20, 21], albeit with lower sensitivity and specificity. This variability across studies is likely due to the use of different assays, which hampers the interpretation and comparison of results [22]. Although additional focused research is required to validate our findings, in our cohort, combining ACTH with LNSC and LNSE gave excellent diagnostic accuracy in distinguishing CD from MACS.

Our findings are consistent with the current literature. Updated guidelines recommend the use of at least two screening tests when investigating CS and there is intermediate to high pre-test probability [9], because the diagnostic accuracy increases significantly [23]. A recent systematic review and meta-analysis of 139 studies including 14,140 patients showed that sensitivity and specificity of LNSC, ODST, and UFC for diagnosing CD were close to 90% and, using meta-regression, ODST and UFC were reported as the best and the worst screening tests respectively, albeit the CIs overlapped significantly [24]. We confirmed high sensitivity for all the three tests in diagnosing CD but we found lower specificity, probably due to the lower number of subjects included in the analysis. The fact that our hospital is a tertiary referral centre where subjects have been filtered by primary and secondary care physicians before CS testing could also be contributing to this discrepancy.

Current guidelines emphasise the specificity of LNSC in diagnosing CD. However, as highlighted in previous research [25, 26], LNSC demonstrates limited diagnostic accuracy for adrenal incidentalomas, where the ODST remains the preferred screening tool [14, 27]. Whilst early studies predominantly focused on comparing LNSC with standard screening tests, the diagnostic accuracy of LNSE had been less well described. Recently, other studies reported similar diagnostic accuracy for LNSE in the diagnosis of CD compared with LNSC using LCMS/MS [28, 29] using different cut-offs (13.5–19.9 nmol/L) [30–32]. In our study, using 18 nmol/L as assay-specific cut-off we found sensitivity of 86% and specificity of 96.1% for LNSE. Screening tests should prioritize high sensitivity over specificity [33, 34] and, when using 15.2 nmol/L as optimal cut-off computed by ROC analysis, the sensitivity of LNSE rose to 100%, with minimal effects on specificity.

Our findings indicate that LNSE is the most predictive test for CD, surpassing LNSC in this regard. Supporting our observation, a study focused on CD patients revealed notable fluctuations in LNSC levels over time [31], which

might reduce its reliability as a diagnostic tool in certain cases. In contrast, LNSE is derived from the rapid and efficient conversion of free serum cortisol in the salivary glands by the enzyme 11- $\beta$ -hydroxysteroid dehydrogenase type 2. This process remains effective even when serum cortisol levels are low. Consequently, LNSE generally shows a more consistent correlation with serum cortisol levels compared to LNSC, which can be undetectable at lower concentrations [35, 36]. Yet, from a practical perspective, LNSC and LNSE can be routinely measured simultaneously, and we have shown that measuring both together led to better diagnostic sensitivity than using either test alone in CD.

Nevertheless, like LNSC, LNSE also showed reduced diagnostic accuracy in cases of adrenal hypercortisolism. This discrepancy might stem from the overall lower circulating glucocorticoid levels in MACS compared with full blown CD as well as from their different pattern of fluctuations throughout the day [37], and other researchers showed how peaks in LNSE measurements do not consistently align with elevations detected by other tests [17]. Albeit salivary cortisone is present at a higher concentration than cortisol (cortisone/cortisol ratio 4:1), a retrospective analysis on 173 patients undergoing ODST, LNSC, and LNSE demonstrated that post-ODST cortisone (but not LNSE) was the most significant independent predictor for abnormal ODST, as evidenced in both univariate and multivariate analyses [38]. Whilst our findings align with other studies that confirm LNSC diagnostic accuracy [39–42] and suggest LNSE as a potential tool to enhance the current screening strategy for CS, further research is still needed to clarify the role of LNSE in diagnosing adrenal Cushing's.

We acknowledge that this study has some limitations and potential biases. First, as a retrospective analysis conducted at a single tertiary referral centre, the findings may not be generalizable to all patient populations. Variations in patient demographics, comorbid conditions, and healthcare settings could influence the performance of the screening tests. Additionally, the retrospective nature of the study introduces the potential for selection bias, as the sample may not represent all patients typically evaluated for CS. Finally, dexamethasone serum levels during ODST were not available in our analysis. Future research should aim to include multicentre, prospective studies to validate these findings across diverse clinical settings and reduce the influence of selection biases.

## Conclusions

Current guidelines for CS screening remain ambiguous as to the optimal screening approach tailored on patient's characteristics. Our study is novel in its combination of various screening tests to identify a method with the highest diagnostic precision, rather than assessing each test in

isolation. We underscore the superiority of LNESC, LNSE, and ODST over UFC in CS screening and show high diagnostic accuracy comparable to the more time-consuming and labour intensive LDDST. We also stress the value of incorporating clinical probability into the screening, which heighten diagnostic accuracy. We propose a simple screening algorithm based on performance of screening tests in our centre (Fig. 3) which now uses LNSE/C in combination with ODST for screening moderate/high clinical pre-test probability patients, and ODST for adrenal incidentalomas. Our analysis contributes to the current literature examining performance of screening tests in Cushing's by analysing a sizeable cohort from a single centre, where previously data, particularly on LNESC/E measurement are lacking. This will hopefully inform the development of more precise clinical guidelines in the investigation of CS.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40618-024-02354-x>.

**Author contributions** RP and AP contributed to the study conception and design. AE contributed to material preparation and data collection. RP performed statistical analysis. The first draft of the manuscript was written by AE and RP. All the other authors reviewed, edited, and commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** The authors did not receive support from any organization for the submitted work.

**Data availability** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** The other authors have no relevant financial or non-financial interests to disclose.

**Ethical approval** This retrospective study involving human participants was in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Research involving human participants and/or animals** The study was approved by the Trust audit team and conducted and registered as a local audit of practice (reference number 8352).

**Informed consent** For this type of study, no informed consent is required.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Pivonello R, Isidori AM, De Martino MC, Newell-Price J, Biller BM, Colao A (2016) Complications of Cushing's syndrome: state of the art. *Lancet Diabetes Endocrinol* 4:611–629
2. De Alcubierre D, Ferrari D, Mauro G, Isidori AM, Tomlinson JW, Pofi R (2023) Glucocorticoids and cognitive function: a walk-through in endogenous and exogenous alterations. *J Endocrinol Invest* 46:1961–1982
3. Pofi R, Caratti G, Ray DW, Tomlinson JW (2023) Treating the side effects of exogenous glucocorticoids; can we separate the good from the bad? *Endocr Rev* 44:975–1011
4. Pivonello R, Scaroni C, Polistena B, Migliore A, Giustina A (2023) Unmet needs on the current medical management of Cushing's syndrome: results from a Delphi panel of Italian endocrinologists. *J Endocrinol Invest* 46:1923–1934
5. Newell-Price J, Grossman A (1999) Diagnosis and management of Cushing's syndrome. *Lancet* 353:2087–2088
6. Elamin MB, Murad MH, Mullan R, Erickson D, Harris K, Nadeem S, Ennis R, Erwin PJ, Montori VM (2008) Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses. *J Clin Endocrinol Metab* 93:1553–1562
7. Leon-Justel A, Madrazo-Atutxa A, Alvarez-Rios AI, Infantes-Fontan R, Garcia-Arnes JA, Lillo-Munoz JA, Aulinas A, Urgell-Rull E, Boronat M, Sanchez-de-Abajo A, Fajardo-Montanana C, Ortuno-Alonso M, Salinas-Vert I, Granada ML, Cano DA, Leal-Cerro A, Spanish CSG (2016) A probabilistic model for Cushing's syndrome screening in at-risk populations: a prospective multicenter study. *J Clin Endocrinol Metab* 101:3747–3754
8. Parasiliti-Caprino M, Bioletto F, Frigerio T, D'Angelo V, Ceccato F, Ferrau F, Ferrigno R, Minnetti M, Scaroni C, Cannavo S, Pivonello R, Isidori A, Broglio F, Giordano R, Spinello M, Grotoli S, Arvat E (2021) A new clinical model to estimate the pre-test probability of Cushing's syndrome: the Cushing score. *Front Endocrinol (Lausanne)* 12:747549
9. Flieseriu M, Auchus R, Bancos I, Ben-Shlomo A, Bertherat J, Biermasz NR, Boguszewski CL, Bronstein MD, Buchfelder M, Carmichael JD, Casanueva FF, Castinetti F, Chanson P, Findling J, Gadelha M, Geer EB, Giustina A, Grossman A, Gurnell M, Ho K, Ioachimescu AG, Kaiser UB, Karavitaki N, Katznelson L, Kelly DF, Lacroix A, McCormack A, Melmed S, Molitch M, Mortini P, Newell-Price J, Nieman L, Pereira AM, Petersenn S, Pivonello R, Raff H, Reincke M, Salvatori R, Scaroni C, Shimon I, Stratakis CA, Swearingen B, Tabarin A, Takahashi Y, Theodoropoulou M, Tsagarakis S, Valassi E, Varlamov EV, Vila G, Wass J, Webb SM, Zatelli MC, Biller BMK (2021) Consensus on diagnosis and management of Cushing's disease: a guideline update. *Lancet Diabetes Endocrinol* 9:847–875
10. Ceccato F, Barbot M, Scaroni C, Boscaro M (2021) Frequently asked questions and answers (if any) in patients with adrenal incidentaloma. *J Endocrinol Invest* 44:2749–2763
11. Nieman LK, Biller BM, Findling JW, Newell-Price J, Savage MO, Stewart PM, Montori VM (2008) The diagnosis of Cushing's syndrome: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 93:1526–1540
12. Jones RL, Owen LJ, Adaway JE, Keevil BG (2012) Simultaneous analysis of cortisol and cortisone in saliva using XLC-MS/MS for fully automated online solid phase extraction. *J Chromatogr B Analyt Technol Biomed Life Sci* 881–882:42–48

13. Owen LJ, Adaway JE, Davies S, Neale S, El-Farhan N, Ducroq D, Evans C, Rees DA, MacKenzie F, Keevil BG (2013) Development of a rapid assay for the analysis of serum cortisol and its implementation into a routine service laboratory. *Ann Clin Biochem* 50:345–352
14. Fassnacht M, Tsagarakis S, Terzolo M, Tabarin A, Sahdev A, Newell-Price J, Pelsma I, Marina L, Lorenz K, Bancos I, Arlt W, Dekkers OM (2023) European Society of Endocrinology clinical practice guidelines on the management of adrenal incidentalomas, in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol* 189:G1–G42
15. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845
16. Vilar L, Freitas MC, Naves LA, Canadas V, Albuquerque JL, Botelho CA, Egito CS, Arruda MJ, Silva LM, Arahata CM, Agra R, Lima LH, Azevedo M, Casulari LA (2008) The role of non-invasive dynamic tests in the diagnosis of Cushing's syndrome. *J Endocrinol Invest* 31:1008–1013
17. Garrahy A, Forde H, O'Kelly P, McGurran K, Zia-Ul-Hussain HM, Noctor E, Tormey WP, Smith D, Denny MC, Bell M, Javadpour M, Agha A (2021) The diagnostic utility of late night salivary cortisol (LNSF) and cortisone (LNSE) in Cushing's syndrome. *Ir J Med Sci* 190:615–623
18. Reimondo G, Allasino B, Bovio S, Saba L, Ardito A, Angeli A, Terzolo M (2011) Pros and cons of dexamethasone suppression test for screening of subclinical Cushing's syndrome in patients with adrenal incidentalomas. *J Endocrinol Invest* 34:e1–5
19. Ceccato F, Artusi C, Barbot M, Lizzul L, Pinelli S, Costantini G, Niero S, Antonelli G, Plebani M, Scaroni C (2020) Dexamethasone measurement during low-dose suppression test for suspected hypercortisolism: threshold development with and validation. *J Endocrinol Invest* 43:1105–1113
20. Gunes M, Celik O, Kadioglu P (2013) Reliability of the diagnostic tests for Cushing's syndrome performed in a tertiary referral center. *Pituitary* 16:139–145
21. Hong AR, Kim JH, Hong ES, Kim IK, Park KS, Ahn CH, Kim SW, Shin CS, Kim SY (2015) Limited diagnostic utility of plasma adrenocorticotropic hormone for differentiation between adrenal Cushing syndrome and Cushing disease. *Endocrinol Metab (Seoul)* 30:297–304
22. Pecori Giraldi F, Saccani A, Cavagnini F (2011) E. study group on the hypothalamo-pituitary-adrenal axis of the Italian Society of, assessment of ACTH assay variability: a multicenter study. *Eur J Endocrinol* 164:505–512
23. Araujo-Castro M, Garcia Cano A, Jimenez Mendiguchia L, Escobar-Morreale HF, Valderrabano P (2021) Diagnostic accuracy of the different hormonal tests used for the diagnosis of autonomous cortisol secretion. *Sci Rep* 11:20539
24. Galm BP, Qiao N, Klibanski A, Biller BMK, Tritos NA (2020) Accuracy of laboratory tests for the diagnosis of Cushing syndrome. *J Clin Endocrinol Metab* 105:2081–2094
25. Masserini B, Morelli V, Bergamaschi S, Ermetici F, Eller-Vainicher C, Barbieri AM, Maffini MA, Scillitani A, Ambrosi B, Beck-Peccoz P, Chiodini I (2009) The limited role of midnight salivary cortisol levels in the diagnosis of subclinical hypercortisolism in patients with adrenal incidentaloma. *Eur J Endocrinol* 160:87–92
26. Ceccato F, Barbot M, Zilio M, Ferasin S, Occhi G, Daniele A, Mazzocut S, Iacobone M, Betterle C, Mantero F, Scaroni C (2013) Performance of salivary cortisol in the diagnosis of Cushing's syndrome, adrenal incidentaloma, and adrenal insufficiency. *Eur J Endocrinol* 169:31–36
27. Vieira-Correa M, Giorgi RB, Oliveira KC, Hayashi LF, Costa-Barbosa FA, Kater CE (2019) Saliva versus serum cortisol to identify subclinical hypercortisolism in adrenal incidentalomas: simplicity versus accuracy. *J Endocrinol Invest* 42:1435–1442
28. Antonelli G, Ceccato F, Artusi C, Marinova M, Plebani M (2015) Salivary cortisol and cortisone by LC-MS/MS: validation, reference intervals and diagnostic accuracy in Cushing's syndrome. *Clin Chim Acta* 451:247–251
29. Meszaros K, Karvaly G, Marta Z, Magda B, Toke J, Szucs N, Toth M, Racz K, Patocs A (2018) Diagnostic performance of a newly developed salivary cortisol and cortisone measurement using an LC-MS/MS method with simple and rapid sample preparation. *J Endocrinol Invest* 41:315–323
30. Mohamed RS, Abuelgasim B, Barker S, Prabhudev H, Martin NM, Meeran K, Williams EL, Darch S, Matthew W, Tan T, Wernig F (2022) Late-night salivary cortisol and cortisone should be the initial screening test for Cushing's syndrome. *Endocr Connect*. <https://doi.org/10.1530/EC-22-0050>
31. Sandouk Z, Johnston P, Bunch D, Wang S, Bena J, Hamrahian A, Kennedy L (2018) Variability of late-night salivary cortisol in Cushing disease: a prospective study. *J Clin Endocrinol Metab* 103:983–990
32. Ponzetto F, Settanni F, Parasiliti-Caprino M, Rumbolo F, Nonnato A, Ricciardo M, Amante E, Priolo G, Vitali S, Anfossi L, Arvat E, Ghigo E, Giordano R, Mengozzi G (2020) Reference ranges of late-night salivary cortisol and cortisone measured by LC-MS/MS and accuracy for the diagnosis of Cushing's syndrome. *J Endocrinol Invest* 43:1797–1806
33. Petersenn S (2021) Biochemical diagnosis of Cushing's disease: screening and confirmatory testing. *Best Pract Res Clin Endocrinol Metab* 35:101519
34. Pofi R, Gunatilake S, Macgregor V, Shine B, Joseph R, Grossman AB, Isidori AM, Cudlip S, Jafar-Mohammadi B, Tomlinson JW, Pal A (2019) Recovery of the hypothalamo-pituitary-adrenal axis after transsphenoidal adenectomy for non-ACTH-secreting macroadenomas. *J Clin Endocrinol Metab* 104:5316–5324
35. Blair J, Adaway J, Keevil B, Ross R (2017) Salivary cortisol and cortisone in the clinical setting. *Curr Opin Endocrinol Diabetes Obes* 24:161–168
36. Harrison RF, Debono M, Whitaker MJ, Keevil BG, Newell-Price J, Ross RJ (2019) Salivary cortisone to estimate cortisol exposure and sampling frequency required based on serum cortisol measurements. *J Clin Endocrinol Metab* 104:765–772
37. Cunningham JM, Buxton OM, Weiss RE (2002) Circadian variation in Cushing's disease and pseudo-Cushing states by analysis of F and ACTH pulsatility. *J Endocrinol Invest* 25:791–799
38. Issa BG, Hanna FWF, Fryer AA, Ensah G, Ebere I, Marshall D, Keevil B (2023) The utility of salivary cortisone in the overnight dexamethasone suppression test in adrenal incidentalomas. *J Clin Endocrinol Metab* 108:e937–e943
39. Raff H (2009) Utility of salivary cortisol measurements in Cushing's syndrome and adrenal insufficiency. *J Clin Endocrinol Metab* 94:3647–3655
40. Elias PC, Martinez EZ, Barone BF, Mermejo LM, Castro M, Moreira AC (2014) Late-night salivary cortisol has a better performance than urinary free cortisol in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab* 99:2045–2051
41. Carroll T, Raff H, Findling JW (2009) Late-night salivary cortisol for the diagnosis of Cushing syndrome: a meta-analysis. *Endocr Pract* 15:335–342
42. Ceccato F, Marcelli G, Martino M, Concettoni C, Brugia M, Trementino L, Michetti G, Arnaldi G (2019) The diagnostic accuracy of increased late night salivary cortisol for Cushing's syndrome: a real-life prospective study. *J Endocrinol Invest* 42:327–335

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.