National Institute for Health and Care Excellence

Final

Adrenal insufficiency: identification and management

Evidence review D: Diagnostic tests and diagnostic thresholds for referral

NICE guideline NG243

Evidence reviews underpinning recommendations 1.2.5 to 1.2.13 and recommendation for research 1 in the NICE guideline

August 2024

Final

This evidence review was developed by NICE

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-6465-9

Contents

1. Diagnostic Tests 6
1.1. Review question 6
1.1.1. Introduction 6
1.1.2. Summary of the protocol 6
1.1.3. Methods and process 7
1.1.4. Diagnostic evidence 7
1.1.5. Summary of studies included in the diagnostic evidence. 8
1.1.6. Summary of the diagnostic evidence 21
1.1.7. Analyses by index test 21
1.1.8. Analyses by study (multiple index tests in single population) 25
1.1.9. Economic evidence 29
1.1.10. Economic model 29
1.1.11. Unit costs 29
2. Diagnostic thresholds for referral 30
2.1. Review question 30
2.1.1. Introduction 30
2.1.2. Summary of the protocol 30
2.1.3. Methods and process 31
2.1.4. Diagnostic evidence 31
2.1.5. Summary of studies included in the diagnostic evidence. 32
2.1.6. Summary of the diagnostic evidence 37
2.1.7. Analyses by index test 37
2.1.8. Additional data not suitable for GRADE analysis 41
2.1.9. Summary of thresholds 41
2.1.10. Economic evidence 42
2.1.11. Economic model 42
2.1.12. Unit costs 42
3. The committee's discussion and interpretation of the evidence 43
3.1.1. The outcomes that matter most 43
3.1.2. The quality of the evidence 43
3.1.3. Benefits and harms 44
3.1.4. Cost-effectiveness and resource use 47
3.1.5. Recommendations supported by this evidence review. 48
References 49
Appendices 51
Appendix A Review protocols 51
Appendix B Literature search strategies 68

Appendix C Diagnostic evidence study selection 80
Appendix D Diagnostic evidence..... 81
Appendix E Forest plots..... 126
Appendix F Economic evidence study selection 131
Appendix G Economic evidence tables 133
Appendix H Health economic model 133
Appendix I Excluded studies..... 134
Appendix J Recommendations for research – full details 156

1. Diagnostic Tests

1.1. Review question

What initial investigations should be done by the non-specialist for people with suspected adrenal insufficiency?

1.1.1. Introduction

Healthcare providers in a wide range of primary and secondary care settings commonly wonder whether a patient has adrenal insufficiency. The definitive (Short Synacthen) test is complicated and costly to perform, requires specialist interpretation, and is not generally readily available to non-endocrine specialists. In practice, random serum cortisol levels are commonly measured by non-specialists as an adrenal insufficiency screening test, and the patient is referred for further testing if the level is considered abnormal. However, this approach can result in unnecessary referrals. Cortisol secretion is pulsatile and follows a circadian rhythm, and secretion is maximal in the early morning, with levels declining during the evening/night. Cortisol level is also affected by many factors including stress, exogenous steroid, and other medications such as oral oestrogen preparations. Unless these factors are taken into consideration, random cortisol testing is rarely helpful. There are new methods of checking adrenal function -testing salivary cortisol and cortisone.

This review seeks to determine which initial tests should be performed by non-specialists to screen for adrenal insufficiency, and how these tests should be interpreted to determine which patients should be referred for specialist testing and those in whom adrenal insufficiency can be safely excluded.

1.1.2. Summary of the protocol

For full details see the review protocol in Appendix A.

Table 1: PICO characteristics of review question

Population	People with suspected adrenal insufficiency Note: Could be well or unwell. Could be in primary or in hospital not under the care of endocrinology. Exclusion: Critically ill patients
Target condition	Adrenal Insufficiency
Index test	• Serum cortisol (8-9 am) • Salivary cortisol • Random cortisol • Electrolytes • Blood glucose • Combination of tests may be included. Exclude: Specialist test for example those that determine the type or cause of AI such as: • Plasma ACTH test • Corticotropin releasing hormone stimulation test.

	• DHEAs
Reference standards	• Short Synacthen Test (standard and low dose) Or • Insulin tolerance test (insulin hypoglycaemia test) Or Clinical diagnosis by a specialist (a specialist will take into account the full clinical picture, including signs, symptoms, risk factors and test results)
Statistical measures	Diagnostic accuracy data • Sensitivity (prioritised) [fewer false negatives i.e. very few people with the condition will be missed] • Specificity
Study design	• Cross-sectional (single gate) studies • Systematic reviews of diagnostic accuracy studies

1.1.3. Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in Appendix A and the methods document.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.1.4. Diagnostic evidence

1.1.4.1. Included studies.

A search was conducted for cross-sectional (single gate) studies reporting the diagnostic accuracy of initial tests, that can be performed by non-specialists, to identify people who have adrenal insufficiency. Diagnosis of adrenal insufficiency was confirmed by the reference standard which could be a short Synacthen test (standard and low dose), insulin tolerance test or clinical diagnosis by a specialist.

Eleven diagnostic studies with prospective data collection were included in the review;[1, 3-7, 9, 11, 13, 17, 18] these are summarised in Table 2 below. A variety of index tests and thresholds were used and evidence from these studies is summarised in the clinical evidence summary below in Table 3 and references in References .

Evidence for the following index tests were identified; basal serum cortisol (8 studies), basal salivary cortisol (7 studies) and basal salivary cortisone (2 studies). No relevant diagnostic test accuracy studies of random cortisol, electrolytes or blood glucose were identified.

The reference standards used to diagnose adrenal insufficiency varied between the studies. six studies used the low-dose or standard dose short Synacthen test (SST) and five studies used the insulin tolerance test (ITT).

Studies took place in a range of countries from around the world including Germany (3 studies), Hong Kong (2 studies), the UK (2 studies), Singapore, Korea, India and the Netherlands (1 study). The majority of studies were conducted in adult populations and one study (Agwu 1999[1]) looked at children from 2-19 years old with suspected adrenal insufficiency.

The assessment of the evidence quality was conducted with emphasis on test sensitivity and specificity as this was identified by the committee as the primary measure in guiding decision-making. The committee also agreed that sensitivity is more important than

specificity, as avoiding false negatives would be the main aim in assessing for this condition. The committee set clinical decision thresholds as sensitivity/specificity 0.9 and 0.70 above which a test would be recommended and 0.6 and 0.5 below which a test is of no clinical use.

See also the study selection flow chart in Appendix C, sensitivity and specificity forest plots in Appendix E and study evidence tables in Appendix D.

1.1.4.2. Excluded studies.

See the excluded studies list in Appendix E.

1.1.5. Summary of studies included in the diagnostic evidence.

Table 2: Summary of studies included in the evidence review.

Study	Population	Target condition	Index test	Reference standard	Comments
Agwu 1999[1]	32 patients who presented with symptoms suggestive of adrenal insufficiency. **Age (years):** 62.8±1.7 **Ratio male:female:** 2:19 **Underlying diagnoses:** 14 had been irradiated for brain tumours; 2 had total body irradiation for relapsed non-Hodgkin's lymphoma and acute lymphoblastic leukaemia; 13 had other endocrinopathies; 2 had growth failure; and 1 had histiocytosis with prolonged dexamethasone treatment.	Adrenal Insufficiency	08:00am Serum cortisol. Measured via indwelling IV catheter before the patient got out of bed. **Threshold**: 500 nmol/l)	Low dose Synacthen test (LDST) was performed at 14:00. A bolus injection of 500 ng/1.73 m^2 of ACTH was given intravenously and sampling took place at 0, 10, 15, 20, 25, 30, 35, 40, and 45 minutes. The next day, a standard short Synacthen test (SSST) was performed at 09:00: 250 µg/1.73 m2 of Synacthen was given as an intravenous bolus. The 30-minute peak was studied. A normal response to the Synacthen test was defined as a peak serum cortisol of ≥500 nmol/Land/or incremental concentration of ≥200 nmol/l. **Assays**: Cortisol was measured using a direct coated tube assay (Euro DPC Ltd, Llanberies, Gwynedd, UK).	Country: UK

Study	Population	Target condition	Index test	Reference standard	Comments
				This assay has a lower limit of detection of 6 nmol/l. The within assay coefficients of variation were 5.7% and 2.6% at serum concentrations of 28 nmol/Land 552 nmol/l, respectively. The between assay coefficients of variation were 9.1% and 6.8% at serum concentrations of 95 nmol/Land 459 nmol/l, respectively.	
Choi 2002[3]	72 patients with clinically suspected secondary adrenocortical insufficiency **Age**: mean; 46 years; range (28-74 years) **Ratio male:female:** 30:42 **Underlying diagnoses**: reasons for suspicion of dysfunction of the hypothalamic-pituitary-adrenocortical axis were: nasopharyngeal carcinoma with radiotherapy to the pituitary region (n=22), iatrogenic Cushing's syndrome (n=20), non-functioning pituitary macroadenoma (n=14), empty sella syndrome (n=8),	Secondary adrenal insufficiency	Morning fasting serum cortisol concentration (0900am) **Threshold:** to maximise sensitivity or specificity	Insulin tolerance test. Actrapid insulin (NovoNordisk, Bagsvaerd, Denmark) 0.1 U/kg was given intravenously after a baseline blood sample was taken for glucose and cortisol assay. When hypoglycaemia developed a blood sample was taken for glucose and cortisol assay, and 50% dextrose solution 40 mL was given to the patient intravenously, followed by oral food. Blood sampling for cortisol measurement was done at intervals of 15 minutes during the first 60 minutes, and thereafter at 30-minute intervals. **Threshold:** peak cortisol response of ≥550 nmol/L for adrenal sufficiency	Country: Hong Kong

Study	Population	Target condition	Index test	Reference standard	Comments
	acromegaly (n=5), Sheehan's syndrome (n=1), Cooley's anaemia with secondary haemochromatosis (n=1), and prolactin-secreting pituitary macroadenoma (n=1).			**Assays**: Cortisol was assayed by the chemiluminescence method (Bayer-Centaur, New York, US). The coefficient of variation for the assay was less than 5%.	
Debono 2023[5]	220 patients at high risk for adrenal insufficiency. All patients referred for an ACTH stimulation test to assess for a new diagnosis of adrenal insufficiency or for recovery from a previous diagnosis of adrenal insufficiency. **Age (SD):** 55.1(15.8) years **Gender** (male to female ratio): 106:102 **Reason for referral**: Patients who were dependent on any type of glucocorticoid, who had been on an oral glucocorticoid prednisolone-equivalent dose of 5 mg/d for 4 weeks, and who were referred for adrenal testing only after they had been weaned down to prednisolone 5 mg/d or equivalent or converted to physiologic	Adrenal Insufficiency	Morning fasting salivary cortisone (performed by participants at home) **Thresholds:** 7 nmol/Land 17 nmol/l Morning fasting salivary cortisol (performed by participants at home) **Thresholds:** 1 nmol/Land 5 nmol/l Baseline serum cortisol (performed in clinic approx. 2 hours later) **Thresholds:** 152 nmol/Land 310 nmol/l	**Reference standard** Standard dose ACTH stimulation test At the endocrine clinic. An ACTH stimulation test was performed with intravenous injection of 250 mg of Synacthen (Atnahs Pharma UK Limited), followed by a serum cortisol level blood draw at 30 minutes. **Threshold:** peak cortisol level of ≥430 nmol/Lfor adrenal sufficiency **Assays**: Serum cortisol was analysed by immunoassay (Elecsys Cortisol II assay; Roche) and interpreted immediately at Sheffield Teaching Hospitals NHS Foundation Trust. An extra serum cortisol sample was stored at -80C and, together with the salivary sample, was then analysed and interpreted by liquid chromatography–tandem mass spectrometry (LC-MS/MS) as a batch	Country: UK **Setting**: NHS hospital for ACTH and home-based salivary cortisone test

Study	Population	Target condition	Index test	Reference standard	Comments
	doses of hydrocortisone 25 mg/d. Patients receiving any intermediate- or long-acting intramuscular or intra-articular glucocorticoid injections were recruited at least 3 months after their last injection. Patients with pituitary disease, such as tumours, inflammatory disease, or those with a history of cranial radiotherapy, were considered eligible for inclusion.			at the end of the study in a different laboratory in Manchester University NHS Foundation Trust.	
De Lange 1993[4]	58 patients with pituitary adenoma or hypothalamic-pituitary disease **Age:** range 17-73 years **Ratio male:female:** 33:25 **Underlying diagnoses:** 45 had pituitary adenoma (30 before treatment; 10 after hypophysectomy and 5 after X-ray therapy). 13 had proven or suspected HPA disease.	Adrenal Insufficiency	Morning fasting plasma cortisol concentration (0900-1000) **Thresholds:** 160 and 260 nmol/L **Assay**: Plasma cortisol was measured by RIA.	Insulin tolerance test. Performed after overnight fast. Soluble insulin 0.15 U/kg was given intravenously. Blood sampling for glucose and cortisol measurement was taken before and at 0, 30, 45, 60 and 90 minutes after injection. **Threshold:** rise in plasma cortisol above 500 nmol/L for adrenal sufficiency **Assay**: Blood glucose was measured by a hexokinase method.	Country: Netherlands
Deutschbein 2009[7]	77 patients with suspected or	Adrenal Insufficiency	Basal salivary cortisol. Saliva was collected by	Insulin tolerance test. After administration of	Country: Germany

Study	Population	Target condition	Index test	Reference standard	Comments
	proven disease of the HPA axis **Age** mean (SD): 44.2 (1.8) years **Ratio male:female:** 41:36 **Underlying diagnoses:** 65 patients had sellar masses, of whom 11 were not operated; all 54 surgically treated patients were tested at least 3 months after surgery (median postoperative interval: 6.3 months; range: 3 – 68 months). The remaining 12 patients showed an impaired secretion of various hormones but did not have any detectable tumours within the HPA.		chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h. **Threshold:** Basal salivary cortisol - an optimal cut-off of 7.6 nmol/l, an upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. Basal serum cortisol: after catheterisation of a superficial cubital vein and a recovery period of 15 min to avoid stress-induced bias, blood samples were directly obtained into serum tubes (Monovetten, Sarstedt, Germany). During High dose SST, serum and saliva samples were taken at 0, 30, 60, 90, and 120 min after i.v. application of 250 mg synthetic ACTH (Synacthen, Novartis). **Threshold:** ROC analysis revealed an optimal cut-off of	insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycaemia were required as evidence of sufficient stress. Samples for blood glucose and serum cortisol were taken at 0, 15, 30, 45, 60, 90, and 120 min. Threshold: Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L. **Assays:** Serum cortisol was measured by a competitive immunoassay (Advia Centaur, Bayer, Germany). The lower detection limit of this assay was 5.5 nmol/l, and intra- and inter-assay coefficients of variation were less than 3.8 and 5.5%, respectively.	

Study	Population	Target condition	Index test	Reference standard	Comments
			260 nmol/l, an upper cut-off of 382 nmol/l, and a lower cut-off of 103 nmol/Lrespectively. Assay: Salivary cortisol was determined using a modification of a commercial radioimmunoassay (RIA) (GammaCoat, DiaSorin, USA), decreasing the sample volume from 200 to 100 µ l. The intra- and inter-assay coefficients of variation were 5.4 and 15.9 %, respectively.		
Deutschbein 2009a[6]	55 patients with suspected or proven disease of the HPA axis **Age** mean (SD): 45.9 (2.1) years **Ratio male:female:** 26:29 **Underlying diagnoses**: Eight subjects had a present neoplasia (4 prolactinomas, 2 somatotropic adenomas, 1 non-functioning adenoma, 1 meningioma), and 40 subjects (24 non-functioning adenomas, 8 somatotropic adenomas, 3 craniopharyngiomas, 3	Adrenal Insufficiency	Basal salivary cortisol. Saliva was collected by chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h. **Threshold:** Basal salivary cortisol - an optimal cut-off of 7.6 nmol/l, an upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. Basal serum cortisol: after catheterisation of a superficial cubital vein and	Insulin tolerance test. After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycaemia were required as evidence of sufficient stress. Samples for blood glucose and serum cortisol were taken at 0, 15, 30, 45, 60, 90, and 120 min. Threshold: Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L. **Assays:** Serum cortisol was measured by a competitive immunoassay (Advia Centaur,	Country: Germany

Study	Population	Target condition	Index test	Reference standard	Comments
	prolactinomas, 2 meningiomas) were tested at least 3 months after surgical treatment (median interval: 4.8 months). Seven patients suffered from pituitary hormone impairment without detectable sellar tumors (3 secondary hypogonadisms, 3 congenital hormone insufficiencies, 1 Sheehan syndrome).		a recovery period of 15 min to avoid stress-induced bias, blood samples were directly obtained into serum tubes (Monovetten, Sarstedt, Germany). During High dose SST, serum and saliva samples were taken at 0, 30, 60, 90, and 120 min after i.v. application of 250 mg synthetic ACTH (Synacthen, Novartis). **Threshold:** ROC analysis revealed an optimal cut-off of 260 nmol/l, an upper cut-off of 382 nmol/l, and a lower cut-off of 103 nmol/Lrespectively. Assay: Salivary cortisol was determined using a modification of a commercial radioimmunoassay (RIA) (GammaCoat, DiaSorin, USA), decreasing the sample volume from 200 to 100 µ l. The intra- and inter-assay coefficients of variation were 5.4 and 15.9 %, respectively.	Bayer, Germany). The lower detection limit of this assay was 5.5 nmol/l, and intra- and inter-assay coefficients of variation were less than 3.8 and 5.5%, respectively.	
George 2020[9]	67 patients with suspected	Adrenal Insufficiency	Basal salivary cortisol.	ACTH short Synacthen test (SST).	Country: India

Study	Population	Target condition	Index test	Reference standard	Comments
	adrenal insufficiency **Age** mean (SD): AI present= 40.69 (18.3), AI absent = 32.99 (15.8) **Ratio male:female:** 29:38 **Reason for referral**: patients with clinical suspicion of AI (anorexia, nausea, weight loss, unexplained hyponatraemia, progressive hyperpigmentation, history of pituitary macroadenoma or surgery)		All tests were initiated between 0800 and 0900 hours. Saliva was collected in plain plastic containers (CML Biotech, Kochi) using passive drooling method until a sufficient volume (approximately 2 ml) was obtained. Salivary samples were collected at 0 (Basal), 60 and 120 minutes for measurement of salivary cortisol. **Thresholds:** 14.1 nmol/L - optimal cut-off values for diagnosis of AI based on SST with a diagnostic cut-off value of <500nmol/L. 3.0 nmol/L cut-off values for diagnosis of AI based on SST with a diagnostic cut-off value of 400.1 nmol/L.	The SST was performed by administering 250 µg of Synacthen intramuscularly, and blood sampling was done at 0, 30 and 60 minutes. Threshold: Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L(18 µg/dl). **Assay:** Serum and salivary cortisol were measured by electrochemilumine scence method (ECLIA) using Cobas e 411 analysers with commercially available Elecsys Cortisol II (second generation, monoclonal antibody) kits which have showed a good correlation with gas chromatography mass spectrometry (GC-MS). Untreated centrifuged saliva was used for the assay. The measuring range was 0.054-63.4 µg/dl. Intra-assay coefficient of variation (CV) at high and low concentrations were 2.5% and 6.1%, respectively. The functional sensitivity (lower limit of quantification) of the assay was 3 nmol/L.	
Kim 2020[11]	120 subjects who underwent the Short Synacthen	Adrenal Insufficiency	Basal salivary cortisol. Samples were	ACTH. Stimulated serum cortisol levels were	Country: Korea

Study	Population	Target condition	Index test	Reference standard	Comments
	test because they were suspected of having primary or secondary Adrenal insufficiency. **Age** mean (range): AI= 58 (42-67), non AI= 58 (43-67) **Ratio male:female:** 50:70 **Underlying diagnoses**: The subjects had a history of pituitary disease (n=67), adrenal disease (n=10), or suspected secondary AI with iatrogenic Cushing syndrome (n=43). Of these subjects, 31 patients underwent pituitary surgery and only one patient took a steroid replacement postoperatively.		collected between 8:00 AM and 10:00 AM and the SST was performed subsequently. Saliva was collected by chewing an oral cotton swab (Salivette, Sarstedt, Germany) for 2 to 5 minutes and the samples were frozen at −20°C until analysis. **Threshold:** Optimal baseline salivary cortisol cut point of 3.2 nmol/L **Assay:** Salivary cortisol was analysed using an enzyme immunoassay kit (EIA, Salimetrics Inc., State College, PA, USA) [24]. The intra-assay CV was 3.2% to 6.3%, and the inter-assay CV was 5.7% to 6.8%. The expected morning ranges of salivary cortisol derived using this kit were 3.1 to 22.4 nmol/L and 4.1 to 20.4 nmol/L in adult men and women aged 51 to 70, respectively.	measured at 30 and 60 minutes after intravenous administration of 250 µg of synthetic adrenocorticotropic hormone (ACTH1-24) (Synacthen, Novartis, Basel, Switzerland). Stimulated serum blood samples were immediately centrifuged at 4°C for 15 minutes and the resulting serum was stored at −20°C until use. **Threshold:** The AI group was defined by a level of stimulated serum cortisol less than 496.8 nmol/L. **Assay**: Serum total cortisol was measured using a Packard Cobra Gamma Counter analyser with commercial radioimmunoassay (RIA) kits (CIS Bio International, Saclay, France; inter-assay coefficient of variation [CV], 4.7%; intra-assay CV, 4.2%).	
Mak 2017[13]	171 patients suspected of having AI from the clinical context.	Adrenal Insufficiency	Basal salivary cortisol. A saliva sample for steroid measurement	Post low dose short Synacthen test peak serum cortisol. All subjects underwent the low-	Country: Hong Kong

Study	Population	Target condition	Index test	Reference standard	Comments
	Age mean (SD): AI group= 56.9 (14.2), non AI group= 56.6 (12.4) **Ratio male:female:** 82:89 **Reasons for investigation**: nonspecific clinical features (n = 55) such as unexplained dizziness and fatigue, electrolyte disturbance (n = 18) such as hyponatremia, diseases or interventions involving the sellar and suprasellar regions (n = 124), previous administration of exogenous glucocorticoids (n = 32), and post adrenalectomy (n = 4)		was collected by placing the cotton tubes (Salivette) in the mouth, chewing for 2 to 3 minutes, and then carefully putting the Salivette into a plastic container without touching it with hands. Simultaneous saliva and serum samples were collected at baseline. **Threshold:** 1.7 nmol/L Basal salivary cortisone. As above. **Threshold:** 12.5 nmol/L Basal serum cortisol. Simultaneous saliva and serum samples were collected at baseline. **Threshold:** 170 nmol/L **Assay**: Salivary cortisol and cortisone were assayed with LC-MS/MS using the Waters Xevo TQ MS system (Waters, Milford, MA). The assay CV for salivary cortisol was ~5% to 7% across all ranges; that for salivary cortisone was	dose SST (LDSST) in the morning. A 1-mg bolus of Synacthen was then injected intravenously. At 30 and 60 minutes after the injection, 2 more pairs of saliva and serum samples were collected. The higher value of the tested parameters, at either 30 or 60 minutes, was regarded as the "peak" value. **Threshold:** The value corresponding to the mean 2 SDs of the peak serum cortisol in this healthy cohort (376 nmol/L) was used as the gold standard **Assay:** Serum total cortisol was assayed with the competitive chemiluminescent microparticle immunoassay using the Abbott Architect i2000SR system (Abbott Laboratories, Abbott Park, IL). The assay coefficient of variation (CV) was 4.0% to 6.2% at low levels and 3.3% to 4.3% at high levels. Serum CBG was measured using a commercial human CBG enzyme-linked immunosorbent assay kit (BioVendor–Laboratorni Medicina, Brno, Czech Republic).	

Study	Population	Target condition	Index test	Reference standard	Comments
			~7% to 10% across all ranges. The lower limit of detection was 0.5 nmol/L for both salivary cortisol and cortisone.		
Schmidt 2003[17]	Fifty-four patients were evaluated because of suspected disease of the HPA axis **Age** mean (SD): 46.6 5 ± 2.5 years **Ratio male:female:** 19:22 **Underlying diagnoses:** Thirty-five had a history of tumours in the pituitary area (14 nonfunctioning adenomas, seven somatotropic adenomas, four prolactinomas, one case of neurosarcoidosis, two cases of hypophysitis, six craniopharyngeomas, and one chordoma), one patient had diabetes insipidus, and in five cases disease was suspected because of clinical symptoms.	Adrenal Insufficiency	Basal cortisol. Unstimulated serum cortisol values between 0800 and 0900h were taken for comparison with the peak cortisol response to the dynamic test. **Threshold:** Twelve morning cortisol values of the total of 20 healthy volunteers were available. The mean basal cortisol value was 439.3 =/- 24.9 nmol/liter (326.0–600.0 nmol/liter). The lower limit of a normal basal cortisol was calculated as 267 nmol/liter (mean 2 sd). ROC analysis suggested an optimal baseline cortisol cut point of 285 nmol/liter.	Insulin tolerance test. Patients underwent an ITT between 0900 and 1030 h by injection of 0.15 IU/kg regular insulin (Actrapid, Novo Nordisk, Mainz, Germany) to achieve blood glucose levels less than 40 mg/dl and until symptoms of hypoglycaemia developed. Blood samples were taken at 0, 15, 30, 45, 60, 90, and 120 min. **Threshold:** Patients were defined as adrenal insufficient or sufficient based on their cortisol peak response to hypoglycaemia of less than 500 nmol/liter or more than 500 nmol/liter. **Assays:** Serum cortisol levels (nanomoles per liter) were assayed at each time point by competitive immunoassay (ADVIA Centaur System, Bayer, Fernwald, Germany). The lower detection limit was assessed to be 5.5 nmol/liter (0.20 g/dl). Intraassay variations as	Country: Germany

Study	Population	Target condition	Index test	Reference standard	Comments
				coefficient of variation for various cortisol values were 3.69% (107.05 nmol/L), 3.09% (155.33 nmol/L), 2.89% (390.95 nmol/L), 3.82% (759.55 nmol/L), and 2.98% (1024.97 nmol/L). Interassay variations for the above-mentioned cortisol concentrations were 5.45, 3.83, 3.07, 1.86, and 3.99%.	
Tan 2023[18]	42 subjects who were planned for adrenocorticotrophic hormone (ACTH) stimulation test (AST) for the evaluation of suspected AI. **Age** mean (SD): AI group = 62.2 (14.6), non AI group = 51.1 (16.4) years **Ratio male:female:** 26:16 **Underlying diagnoses:** Thirty-five had a history of tumors in the pituitary area (14 nonfunctioning adenomas, seven somatotropic adenomas, four prolactinomas, one case of neurosarcoidosis, two cases of hypophysitis, six craniopharyngeomas, and one	Adrenal Insufficiency	Salivary cortisol. Conducted in the outpatient endocrine unit between 08:00 and 10:00 by specialised nursing staff. An intravenous cannula was inserted followed by the simultaneous collection of baseline plasma and saliva samples (0-min sample. Saliva specimens were collected using the SARSTEDT Salivette®. Subjects were instructed to place the SARSTEDT Salivette® into their mouth for 2 min to obtain at least 1.5 mL of saliva per specimen. **Threshold**: cut off 2.7 nmol/L **Assay:** Salivary cortisol	Serum cortisol after AST. Conducted in the outpatient endocrine unit between 08:00 and 10:00am. An intravenous cannula was inserted. Blood was collected in an EDTA tube and analysed upon receipt by the laboratory. About 250 µg of ACTH (Synacthen®, Novartis) was then injected intravenously, followed by the simultaneous collection of serum and salivary cortisol samples at 30 and 60 min. **Threshold:** AI was diagnosed if peak serum cortisol levels failed to reach 500 nmol/L **Assay:** Serum cortisol was measured using the Beckman Coulter UniCel DxI 800 Access	Country: Singapore

Study	Population	Target condition	Index test	Reference standard	Comments
	chordoma), one patient had diabetes insipidus, and in five cases disease was suspected because of clinical symptoms.		measurement was performed at the Mayo Medical Laboratories. Salivary cortisol was extracted from the specimen using online turbulent flow high-performance liquid chromatography and analysed by liquid chromatography tandem mass spectrometry (LC-MS/MS). The detection limit was 0.11 nmol/L. The intra-assay coefficient of variation was 7.2% at 3.0 nmol/L, and the inter-assay coefficient of variation was 5.8% at 1.4 nmol/L.	Immunoassay Systems. The detection limit was 11 nmol/L. The assay exhibited a total imprecision of <12% at approximately 138 nmol/L and <10% for higher concentrations of cortisol.	

See Appendix D for full evidence tables.

1.1.6. Summary of the diagnostic evidence

The assessment of the evidence quality was conducted with emphasis on test sensitivity and specificity as this was identified by the committee as the primary pair of measures in guiding decision-making, with sensitivity being prioritised. The committee set clinical decision thresholds as sensitivity/specificity =0.9/0.7 above which a test would be recommended and 0.6/0.5 below which a test is of no clinical use.

1.1.7. Analyses by index test

Table 3: Clinical evidence summary: diagnostic test accuracy for morning serum cortisol

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol (threshold ≤112 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	72	Very serious[1]	Not serious	Not serious	Unclear[2]	Not reported	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	Sensitivity 1.00 (no false negatives)	VERY LOW
Morning serum cortisol (threshold 152 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Serious[6]	Sensitivity = 0.65 (0.54-0.75)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity = 0.95 (0.89-0.98	HIGH
Morning serum cortisol (threshold <160 nmol/L nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	58	Not serious	Not serious	Serious[3]	Unclear[2]	Sensitivity = 0.64	LOW
		Not serious	Not serious	Serious[3]	Unclear[2]	Specificity = 0.91	LOW
Morning serum cortisol (threshold 170 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	171	Serious[4]	Not serious	Not serious	Not Serious	Sensitivity=0.76 (0.63-0.86)	MODERATE
		Serious[4]	Not serious	Not serious	Not serious	Specificity=0.88 (0.81-0.94)	MODERATE
Morning serum cortisol (threshold 260 nmol/l) to detect AI in people with suspected secondary adrenal insufficiency							
1 cross-sectional study	55	Very serious[5]	Not serious	Not serious	Serious[6]	Sensitivity=0.73 (0.54-0.88)	VERY LOW
		Very serious[5]	Not serious	Not serious	Serious[7]	Specificity=0.72 (0.51-0.88)	VERY LOW
Morning serum cortisol (threshold <260 nmol/L nmol/L) to detect AI in people with suspected							
1 cross-sectional study	58	Not serious	Not serious	Serious[3]	Unclear[2]	Sensitivity = 0.96	LOW
		Not serious	Not serious	Serious[3]	Unclear[2]	Specificity = 0.64	LOW

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol (threshold 283 nmol/l) to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
1 cross-sectional study	77	Very serious[5]	Not serious	Not serious	Serious[6]	Sensitivity=0.73 (0.57-0.86)	VERY LOW
		Very serious[5]	Not serious	Not serious	Serious[7]	Specificity=0.69 (0.52-0.84)	VERY LOW
Morning serum cortisol (threshold 285 nmol/L) to detect AI in people with suspected hypothalamic-pituitary-adrenal disease							
1 cross-sectional study	41	Serious[4]	Not serious	Not serious	Serious[8]	Sensitivity=1.00 (0.83-1)	LOW
		Serious[4]	Not serious	Not serious	Very serious[9]	Specificity=0.62 (0.38-0.82)	VERY LOW
Morning serum cortisol (threshold 310 nmol/L nmol/L) to detect AI in people with suspected AI							
1 cross - sectional study	208	Not serious	Not serious	Not serious	Serious[8]	Sensitivity=0.96 (0.89-0.99)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity=0.41 (0.32-0.50)	HIGH
Morning serum cortisol (threshold ≥420 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	72	Very serious[1]	Not serious	Not serious	Unclear[2]	Sensitivity 1.00 (no false negatives)	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	Not reported	VERY LOW
Morning serum cortisol (threshold 500 nmol/L) to detect AI in children with suspected AI							
1 cross-sectional study	32	Serious[4]	Not serious	Not serious	Serious[8]	Sensitivity=1.00 (0.72-1)	LOW
		Serious[4]	Not serious	Not serious	Not serious	Specificity=0.33 (0.15-0.57)	MODERATE

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and inappropriate time interval between the index test and reference standard.
[2] No confidence interval reported and primary data not available so unable to calculate imprecision. Downgraded by 2 increments.
[3] The evidence was downgraded by 1 increment due to serious indirectness (serious intervention indirectness due to concerns over the conduction of the index test. Basal plasma cortisol was measured on 2 consecutive days and the mean of these two cortisol measures is reported).
[4] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 1 increment due to high risk of bias arising from unclear reporting on patient selection.
[5] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and patient flow.
[6] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).
[7] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high Specificity' (70%).
[8] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high sensitivity' (90%).
[9] Downgraded by 2 increments as the confidence interval crossed two decision thresholds (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low specificity' (60%) and high specificity (70%).

Table 4: Clinical evidence summary: diagnostic test accuracy for morning salivary cortisol

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning salivary cortisol **(threshold 1 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.52 (0.41-0.62)	HIGH
		Not serious	Not serious	Not serious	Not serious	Specificity=0.97 (0.91-0.99)	HIGH
Morning salivary cortisol **(threshold 1.7 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	171	Serious[1]	Not serious	Not serious	Serious[2]	Sensitivity=0.63 (0.49-0.75)	LOW
		Serious[1]	Not serious	Not serious	Not serious	Specificity=0.96 (0.90-0.99)	MODERATE
Morning salivary cortisol **(threshold 2.7 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	42	Very serious[3]	Not serious	Not serious	Serious[2]	Sensitivity=0.52 (0.30-0.74)	VERY LOW
		Very serious[3]	Not serious	Not serious	Serious[4]	Specificity=0.86 (0.64-0.97)	VERY LOW
Morning salivary cortisol **(threshold 3.0 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	67	Serious[1]	Not serious	Not serious	Serious[2]	Sensitivity=0.74 (0.57-0.88)	LOW
		Serious[1]	Not serious	Not serious	Serious[4]	Specificity=0.84 (0.67-0.95)	LOW
Morning salivary cortisol **(threshold 3.2 nmol/L)** to detect AI in people with suspected primary or secondary AI							
1 cross-sectional study	120	Serious[1]	Not serious	Not serious	Serious[5]	Sensitivity=0.85 (0.69-0.95)	LOW
		Serious[1]	Not serious	Not serious	Serious[4]	Specificity=0.73 (0.63-0.82)	LOW
Morning salivary cortisol **(threshold 5.0 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Serious[5]	Sensitivity=0.95 (0.88-0.98)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity=0.61 (0.51-0.70)	HIGH
Morning salivary cortisol **(threshold 7.6 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency							
1 cross-sectional study	55	Very serious[6]	Not serious	Not serious	Serious[2]	Sensitivity=0.53 (0.34-0.72)	VERY LOW
			Very serious[6]	Not serious	Serious[4]	Specificity=0.84 (0.64-0.95)	VERY LOW
Morning salivary cortisol **(threshold 14.1 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	67	Serious[1]	Not serious	Not serious	Serious[5]	Sensitivity=0.94 (0.79-0.99)	LOW
		Serious[1]	Not serious	Not serious	Not serious	Specificity=0.14 (0.05-0.30)	MODERATE

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning salivary cortisol **(threshold 15.1 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
1 cross-sectional study	77	Very serious[6]	Not serious	Not serious	Serious[5]	Sensitivity=0.85 (0.71-0.94)	VERY LOW
		Very serious[6]	Not serious	Not serious	Very serious[7]	Specificity=0.56 (0.38-0.72)	VERY LOW

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 1 increment due to high risk of bias arising from unclear reporting on patient selection.
[2] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).
[3] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and patient flow.
[4] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high Specificity' (70%).
[5]Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high sensitivity' (90%).
[6] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and patient flow.
[7] Downgraded by 2 increments as the confidence interval crossed two decision thresholds (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low specificity' (60%) and high specificity (70%).

Table 5: Clinical evidence summary: diagnostic test accuracy for morning salivary cortisone

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning salivary cortisone (threshold 7 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.74 (0.63-0.82)	HIGH
		Not serious	Not serious	Not serious	Not serious	Specificity=0.97 (0.91-0.99)	HIGH
Morning salivary cortisone (threshold 12.5 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	171	Serious[1]	Not serious	Not serious	Not serious	Sensitivity=0.73 (0.60-0.84)	MODERATE
		Serious[1]	Not serious	Not serious	Not serious	Specificity=0.88 (0.81-0.94)	MODERATE
Morning salivary cortisone (threshold 17 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.97 (0.91-0.99)	HIGH
		Not serious	Not serious	Not serious	Serious[2]	Specificity=0.56 (0.46-0.65)	MODERATE

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 1 increment due to high risk of bias arising from unclear reporting on patient selection.

[2] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'Low Specificity' (50%).

1.1.8. Analyses by study (multiple index tests in single population)

***Note**: these data are duplicated in the analyses by index test above.*

Table 6: Clinical evidence summary: diagnostic test accuracy for morning serum and salivary cortisol

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol **(threshold 283 nmol/L)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
1 cross-sectional study	77	Very serious[1]	Not serious	Not serious	Serious[2]	Sensitivity=0.73 (0.57-0.86)	LOW
		Very serious[1]	Not serious	Not serious	Serious[3]	Specificity=0.69 (0.52-0.84)	LOW
Morning salivary cortisol **(threshold 15.1 nmol/L)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
1 cross-sectional study	77	Very serious[1]	Not serious	Not serious	Serious[4]	Sensitivity=0.85 (0.71-0.94)	VERY LOW
		Very serious[1]	Not serious	Not serious	Very Serious[5]	Specificity=0.56 (0.38-0.72)	VERY LOW

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and patient flow.
[2] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).
[3] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high Specificity' (70%).
[4] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high sensitivity' (90%).
[5] Downgraded by 2 increments as the confidence interval crossed two decision thresholds (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low specificity' (50%) and high specificity (70%).

Table 7: Clinical evidence summary: diagnostic test accuracy for morning serum and salivary cortisol

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol **(threshold 260 nmol/l**) to detect AI in people with suspected secondary adrenal insufficiency							
	55	Very serious[1]	Not serious	Not serious	Serious[2]	Sensitivity=0.73 (0.54-0.88)	VERY LOW

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
1 cross-sectional study		Very serious[1]	Not serious	Not serious	Serious[3]	Specificity=0.72 (0.51-0.88)	VERY LOW
Morning salivary cortisol **(threshold 7.6 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency							
1 cross-sectional study	55	Very serious[1]	Not serious	Not serious	Serious[4]	Sensitivity=0.53 (0.34-0.72)	VERY LOW
		Very serious[1]	Not serious	Not serious	Serious[5]	Specificity=0.84 (0.64-0.95)	VERY LOW

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and patient flow.
[2] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).
[3] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high Specificity' (70%).
[4] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).
[5] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high Specificity' (70%).

Table 8: Clinical evidence summary: diagnostic test accuracy for morning serum cortisol, morning salivary cortisol and morning salivary cortisone

Lower thresholds

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol (threshold 152 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Serious[6]	Sensitivity = 0.65 (0.54-0.75)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity = 0.95 (0.89-0.98)	HIGH
Morning salivary cortisol (threshold 1nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.52 (0.41-0.62)	HIGH
		Not serious	Not serious	Not serious	Not serious	Specificity=0.97 (0.91-0.99)	HIGH
Morning salivary cortisone (threshold 7 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.74 (0.63-0.82)	HIGH
		Not serious	Not serious	Not serious	Not serious	Specificity=0.97 (0.91-0.99)	HIGH

[1]Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).

Higher thresholds

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol (**threshold 310 nmol/L**) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Serious[1]	Sensitivity=0.96 (0.89-0.99)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity=0.41 (0.32-0.50)	MODERATE
Morning salivary cortisol (threshold 5.0 nmol/L) to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Serious[1]	Sensitivity=0.95 (0.88-0.98)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity=0.61 (0.51-0.70)	HIGH
Morning salivary cortisone **(threshold 17 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.97 (0.91-0.99)	HIGH
		Not serious	Not serious	Not serious	Serious[2]	Specificity=0.56 (0.46-0.65)	MODERATE

[1] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high sensitivity' (90%).

[2] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'Low Specificity' (50%).

Table 9: Clinical evidence summary: diagnostic test accuracy for morning serum cortisol, morning salivary cortisol and morning salivary cortisone

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol (**threshold 170 nmol/L**) to detect AI in people with suspected AI							
1 cross-sectional study	171	Serious[1]	Not serious	Not serious	Not Serious	Sensitivity=0.76 (0.63-0.86)	MODERATE
		Serious[1]	Not serious	Not serious	Not serious	Specificity=0.88 (0.81-0.94)	MODERATE
Morning salivary cortisol **(threshold 1.7 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	171	Serious[1]	Not serious	Not serious	Serious[2]	Sensitivity=0.63 (0.49-0.75)	VERY LOW
		Serious[1]	Not serious	Not serious	Not serious	Specificity=0.96 (0.90-0.99)	VERY LOW
Morning salivary cortisone **(threshold 12.5 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	171	Serious[1]	Not serious	Not serious	No serious	Sensitivity=0.73 (0.60-0.84)	VERY LOW
		Serious[1]	Not serious	Not serious	Not serious	Specificity=0.88 (0.81-0.94)	VERY LOW

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 1 increment due to high risk of bias arising from unclear reporting on patient selection.
[2] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).

1.1.9. Economic evidence

1.1.9.1. Included studies.

No health economic studies were included.

1.1.9.2. Excluded studies.

No relevant health economic studies were excluded due to assessment of limited applicability or methodological limitations.

See also the health economic study selection flow chart in Appendix B.

1.1.10. Economic model

This area was not prioritised for new cost-effectiveness analysis.

1.1.11. Unit costs

Relevant unit costs are provided below to aid the consideration of cost-effectiveness.

Table 10: Unit costs of tests and referral

Resource	Unit costs
Primary care	
Serum cortisol test (community / hospital outpatient) (a)	£5.88 - £6.25
Salivary cortisol test (b)	£28.17
Advice and guidance (c)	£12.08 - £24.17
Secondary care	
Referral to secondary care, adult / paediatric (d)	£293.42 / £417.78
Short Synacthen test (e)	£398.36
Insulin tolerance test (f)	£469.60

Sources:

(a) Consisting of a blood test taken either in the community (ten minutes of a Band 3 health care assistant time £4.33 [calculated based on reported wage from PSSRU 2022[10] and proportional Salary oncosts, Overheads and Capital used by PSSRU for Community-based social care professionals, £26 per hour]) or in an outpatient hospital setting by a phlebotomist (£4.70 [Phlebotomy DAPS08, NHS reference costs 21/22[16]]). Blood test analysis: £1.55 (Clinical biochemistry DAPS04, NHS reference cost 21/22).[16]

(b) Salivary sample taken at home using Salivette, cost of Salivette (£184.95 for 500 units).[12]. Postage and packaging (£3.50) and laboratory analysis approximately (£14.30). Additional posting may be required when the local hospital needs to send to another laboratory (medical courier service £10). These estimates are based on committee opinion and communication with Debono et al.

(c) Cost based on 5 to 10 minutes of consultant endocrinology time required (based on committee opinion). Unit cost from PSSRU 2022.[10]

(d) Weighted average (consultant led, non-consultant led and multiprofessional) cost of first endocrinology face to face appointment. NHS reference costs 2021/2022).[16]

(e) Cost includes: daycase £398.36 (KA08C Other Endocrine Disorders with CC Score 0-1, NHS reference costs 2021/2022) this is assumed to include cost of Synacthen (Synacthen 250µg ampoule: £38, BNF, November 2023).[2]

(f) Cost excluding VAT. Costed as part of NICE Medtech innovation briefing [MIB320] on 'Macimorelin for diagnosing growth hormone deficiency', published March 2023[15]

2. Diagnostic thresholds for referral

2.1. Review question

When should people with suspected adrenal insufficiency be referred to specialists for further investigation?

2.1.1. Introduction

Adrenal insufficiency is a diagnosis that must not be missed because it is inevitably fatal if untreated. With simple daily, oral glucocorticoid medication patients make an excellent recovery and are able to live close to normal lives. The clinical features of adrenal insufficiency are frequently non-specific so clinicians must consider it, as otherwise undiagnosed patients can be missed with disastrous consequences. There is new literature looking at different thresholds to safely diagnose adrenal insufficiency, and also data from newer salivary cortisol and cortisone tests.

This review looks at the investigation process for adrenal sufficiency. It assesses the diagnostic accuracy of different screening tests and also provides information to inform the choice of threshold for test results to inform referral to a specialist for further investigation.

2.1.2. Summary of the protocol

For full details see the review protocol in Appendix A.

Table 11: PICO characteristics of review question

Population	**Inclusion:** • People with suspected adrenal insufficiency. • Patients who are well or unwell. • Patients in primary care or in hospital not under the care of endocrinology. **Exclusion:** • Critically ill patients
Target condition	Adrenal Insufficiency
Index tests	**Include:** • Serum cortisol (8-9 am) • Salivary cortisol • Salivary cortisone • Combined ACTH and cortisol **Exclude:** Specialist tests, including those that determine the type or cause of AI such as: • Plasma ACTH test • Corticotropin releasing hormone stimulation test. • DHEAs
Reference standards	• Short Synacthen Test (standard and low dose) Or • Insulin tolerance test (insulin hypoglycaemia test) Or • Clinical diagnosis by a specialist

Statistical measures	Diagnostic accuracy data • Sensitivity (prioritised) • Specificity
Study design	• Cross-sectional (single gate) studies • Systematic reviews of diagnostic accuracy studies • If no or insufficient diagnostic accuracy studies are identified, prospective cohort studies may be included

2.1.3. Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in Appendix A and the methods document.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

2.1.4. Diagnostic evidence

2.1.4.1. Included studies

A search was conducted for cross-sectional (single gate) studies reporting the diagnostic accuracy of initial tests, that can be performed by non-specialists, to identify people who have adrenal insufficiency. Diagnosis of adrenal insufficiency was confirmed by the reference standard which could be a short Synacthen test (standard and low dose), insulin tolerance test or clinical diagnosis by a specialist.

Five diagnostic studies[3-7] with prospective data were included in the review. These studies assessed the diagnostic accuracy of different index tests and provided information to inform the choice of threshold for referral to a specialist. These studies are summarised in Table 2 below. A variety of index tests and thresholds were used and evidence from these studies is summarised in the clinical evidence summary below in Table 3 and references in References.

Evidence for the following index tests were identified; basal serum or plasma cortisol (5 studies), basal salivary cortisol (3 studies) and basal salivary cortisone (1 study). All evidence was reported in adult populations with suspected adrenal insufficiency and no studies were identified in children.

The reference standard used to diagnose adrenal insufficiency was the insulin tolerance test in 4 studies and the ACTH stimulation test in 1 study.

Of the included studies, 2 were conducted in Germany,1 in the UK, 1 in Hong Kong, and 1 in the Netherlands.

The assessment of the evidence quality was conducted with emphasis on test sensitivity and specificity as this was identified by the committee as the primary measure in guiding decision-making. The committee also agreed that sensitivity is more important than specificity, as avoiding false negatives would be the main aim in assessing for this condition in a non-specialist setting before any further tests were conducted. The committee set clinical decision thresholds as sensitivity/specificity 0.9 and 0.70 above which a test would be recommended and 0.6 and 0.5 below which a test is of no clinical use.

See also the study selection flow chart in Appendix C, and study evidence tables in Appendix D.

Excluded studies.

See the excluded studies list in Appendix E.

2.1.5. Summary of studies included in the diagnostic evidence.

Table 12: Summary of studies included in the evidence review.

Study	Population	Target condition	Index test	Reference standard	Comments
Choi 2002[3]	72 patients with clinically suspected secondary adrenocortical insufficiency **Age**: mean; 46 years; range (28-74 years) **Ratio male: female:** 30:42 **Underlying diagnoses**: reasons for suspicion of dysfunction of the hypothalamic-pituitary-adrenocortical axis were: nasopharyngeal carcinoma with radiotherapy to the pituitary region (n=22), iatrogenic Cushing's syndrome (n=20), non-functioning pituitary macroadenoma (n=14), empty sella syndrome (n=8), acromegaly (n=5), Sheehan's syndrome (n=1),	Secondary adrenal insufficiency	Morning fasting serum cortisol concentration (0900am) **Threshold:** to maximise sensitivity or specificity	Insulin tolerance test. Actrapid insulin (NovoNordisk, Bagsvaerd, Denmark) 0.1 U/kg was given intravenously after a baseline blood sample was taken for glucose and cortisol assay. When hypoglycaemia developed a blood sample was taken for glucose and cortisol assay, and 50% dextrose solution 40 mL was given to the patient intravenously, followed by oral food. Blood sampling for cortisol measurement was done at intervals of 15 minutes during the first 60 minutes, and thereafter at 30-minute intervals. **Threshold:** peak cortisol response of ≥550 nmol/L for adrenal sufficiency **Assays**: Cortisol was assayed by the chemiluminescence method (Bayer-Centaur, New York, US).	Country: Hong Kong

Study	Population	Target condition	Index test	Reference standard	Comments
	Cooley's anaemia with secondary haemochromatosis (n=1), and prolactin-secreting pituitary macroadenoma (n=1).			The coefficient of variation for the assay was less than 5%.	
Debono 2023[5]	220 patients at high risk for adrenal insufficiency. All patients referred for an ACTH stimulation test to assess for a new diagnosis of adrenal insufficiency or for recovery from a previous diagnosis of adrenal insufficiency. **Age:** 55.1–15.8 years **Gender** (male to female ratio): 106:102 **Reason for referral**: Patients who were dependent on any type of glucocorticoid, who had been on an oral glucocorticoid prednisolone-equivalent dose of 5 mg/d for 4 weeks, and who were referred for adrenal testing only after they had been weaned down to prednisolone 5 mg/d or equivalent or converted to physiologic doses of hydrocortisone 25 mg/d.	Adrenal Insufficiency	Morning fasting salivary cortisone (performed by participants at home) **Thresholds:** 7 nmol/Land 17 nmol/L Morning fasting salivary cortisol (performed by participants at home) **Thresholds:** 1 nmol/Land 5 nmol/L Baseline serum cortisol (performed in clinic approx. 2 hours later) **Thresholds:** 152 nmol/Land 310 nmol/l	Reference standard. Standard dose ACTH stimulation test At the endocrine clinic. An ACTH stimulation test was performed with intravenous injection of 250 mg of Synacthen (Atnahs Pharma UK Limited), followed by a serum cortisol level blood draw at 30 minutes. **Threshold:** peak cortisol level of ≥430 nmol/Lfor adrenal sufficiency **Assays**: Serum cortisol was analysed by immunoassay (Elecsys Cortisol II assay; Roche) and interpreted immediately at Sheffield Teaching Hospitals NHS Foundation Trust. An extra serum cortisol sample was stored at -80C and, together with the salivary sample, was then analysed and interpreted by liquid chromatography–tandem mass spectrometry (LC-MS/MS) as a batch at the end of the study in a different laboratory in Manchester University NHS Foundation Trust.	Country: UK Setting: NHS hospital for ACTH and home based salivary cortisone test

Study	Population	Target condition	Index test	Reference standard	Comments
	Patients receiving any intermediate- or long-acting intramuscular or intra-articular glucocorticoid injections were recruited at least 3 months after their last injection. Patients with pituitary disease, such as tumours, inflammatory disease, or those with a history of cranial radiotherapy, were considered eligible for inclusion.				
De Lange 1993[4]	58 patients with pituitary adenoma or hypothalamic-pituitary disease **Age:** range 17-73 years **Ratio male: female:** 33:25 **Underlying diagnoses:** 45 had pituitary adenoma (30 before treatment; 10 after hypophysectomy and 5 after X-ray therapy). 13 had proven or suspected HPA disease.	Adrenal Insufficiency	Morning fasting plasma cortisol concentration (0900-1000). **Thresholds:** 160 and 260 nmol/L **Assay**: Plasma cortisol was measured by RIA.	Insulin tolerance test. Performed after overnight fast. Soluble insulin 0.15 U/kg was given intravenously. Blood sampling for glucose and cortisol measurement was taken before and at 0, 30, 45, 60 and 90 minutes after injection. **Threshold:** rise in plasma cortisol above 500 nmol/L for adrenal sufficiency. **Assay**: Blood glucose was measured by a hexokinase method.	Country: Netherlands
Deutschbein 2009[7]	77 patients with suspected or proven disease of the HPA axis	Adrenal Insufficiency	Basal salivary cortisol. Saliva was collected by chewing a specific cotton swab (Salivette,	Insulin tolerance test. After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of	Country: Germany

Study	Population	Target condition	Index test	Reference standard	Comments
	Age mean (SD): 44.2 (1.8) years **Ratio male: female:** 41:36 **Underlying diagnoses:** 65 patients had sellar masses, of whom 11 were not operated; all 54 surgically treated patients were tested at least 3 months after surgery (median postoperative interval: 6.3 months; range: 3 – 68 months). The remaining 12 patients showed an impaired secretion of various hormones but did not have any detectable tumours within the HPA.		Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h. **Threshold:** Basal salivary cortisol - upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. The remaining subjects were investigated by the standard dose SST. Basal serum cortisol: after catheterisation of a superficial cubital vein and a recovery period of 15 min to avoid stress-induced bias, blood samples were directly obtained into serum tubes (Monovetten, Sarstedt, Germany). During SST, serum and saliva samples were taken at 0, 30, 60, 90, and 120 min after i.v. application of 250 mg synthetic ACTH (Synacthen, Novartis). **Threshold:**	hypoglycaemia were required as evidence of sufficient stress. Samples for blood glucose and serum cortisol were taken at 0, 15, 30, 45, 60, 90, and 120 min. Threshold: Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L. **Assays:** Serum cortisol was measured by a competitive immunoassay (Advia Centaur, Bayer, Germany). The lower detection limit of this assay was 5.5 nmol/l, and intra- and inter-assay coefficients of variation were less than 3.8 and 5.5%, respectively. Salivary cortisol was determined using a modification of a commercial radioimmunoassay (RIA) (GammaCoat, DiaSorin, USA), decreasing the sample volume from 200 to 100 µ l. The intra- and inter-assay coefficients of variation were 5.4 and 15.9 %, respectively.	Patient cohort may overlap with Deutschbein 2009a

Study	Population	Target condition	Index test	Reference standard	Comments
			upper cut-off 382 nmol/l, and lower cut-off 103 nmol/l.		
Deutschbein 2009a[6]	55 patients with suspected or proven disease of the HPA axis **Age** mean (SD): 45.9 (2.1) years **Ratio male: female:** 26:29 **Underlying diagnoses**: Eight subjects had a present neoplasia (4 prolactinomas, 2 somatotropic adenomas, 1 non-functioning adenoma, 1 meningioma), and 40 subjects (24 non-functioning adenomas, 8 somatotropic adenomas, 3 craniopharyngiomas, 3 prolactinomas, 2 meningiomas) were tested at least 3 months after surgical treatment (median interval: 4.8 months). Seven patients suffered from pituitary hormone impairment without detectable sellar tumors (3 secondary hypogonadisms, 3 congenital hormone insufficiencies, 1 Sheehan syndrome).	Adrenal Insufficiency	Basal salivary cortisol. Saliva was collected by chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h. **Threshold:** Basal salivary cortisol - upper cut-off of 17.5 nmol/l, lower cut-off of 5.0 nmol/l. Basal serum cortisol: after catheterisation of a superficial cubital vein and a recovery period of 15 min to avoid stress-induced bias, blood samples were directly obtained into serum tubes (Monovetten, Sarstedt, Germany). During SST, serum and saliva samples were taken at 0, 30, 60, 90, and 120 min after i.v. application of 250 mg	Insulin tolerance test. After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycaemia were required as evidence of sufficient stress. Samples for blood glucose and serum cortisol were taken at 0, 15, 30, 45, 60, 90, and 120 min. **Threshold:** Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L. **Assays:** Serum cortisol was measured by a competitive immunoassay (Advia Centaur, Bayer, Germany). The lower detection limit of this assay was 5.5 nmol/l, and intra- and inter-assay coefficients of variation were less than 3.8 and 5.5%, respectively. Salivary cortisol was determined using a modification of a commercial radioimmunoassay (RIA) (GammaCoat, DiaSorin, USA), decreasing the sample volume from 200 to 100 µ l. The intra- and inter-assay coefficients of variation were 5.4 and 15.9 %, respectively.	Country: Germany Patient cohort may overlap with Deutschbein 2009

Study	Population	Target condition	Index test	Reference standard	Comments
			synthetic ACTH (Synacthen, Novartis). **Threshold:** upper cut-off 382 nmol/l, and lower cut-off 103 nmol/l.		

See Appendix D for full evidence tables.

2.1.6. Summary of the diagnostic evidence

The assessment of the evidence quality was conducted with emphasis on test sensitivity and specificity as this was identified by the committee as the primary pair of measures in guiding decision-making, with sensitivity being prioritised. The committee set clinical decision thresholds as sensitivity/specificity =0.9/0.7 above which a test would be recommended and 0.6/0.5 below which a test is of no clinical use.

2.1.7. Analyses by index test

Table 13: Clinical evidence summary: diagnostic test accuracy for morning serum cortisol

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning serum cortisol **(threshold <103 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
2 cross-sectional studies	132	Very serious[1]	Not serious	Not serious	Unclear[2]	NR	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	Specificity ≥95%	VERY LOW
Morning serum cortisol **(threshold ≤112 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	72	Very serious[3]	Not serious	Not serious	Unclear[2]	Not reported	VERY LOW
		Very serious[3]	Not serious	Not serious	Unclear[2]	Sensitivity 1.00 (no false negatives)	VERY LOW
Morning serum cortisol **(threshold 152 nmol/L)** to detect AI in people with suspected AI							
	208	Not serious	Not serious	Not serious	Serious[5]	Sensitivity = 0.65 (0.54-0.75)	MODERATE

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
1 diagnostic study		Not serious	Not serious	Not serious	Not serious	Specificity = 0.95 (0.89-0.98	HIGH
Morning serum cortisol **(threshold <160 nmol/L nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	58	Not serious	Not serious	Serious[4]	Unclear[2]	Sensitivity = 0.64	LOW
		Not serious	Not serious	Serious[4]	Unclear[2]	Specificity = 0.91	LOW
Morning serum cortisol **(threshold <260 nmol/L nmol/L)** to detect AI in people with suspected							
1 cross-sectional study	58	Not serious	Not serious	Serious[4]	Unclear[2]	Sensitivity = 0.96	LOW
		Not serious	Not serious	Serious[4]	Unclear[2]	Specificity = 0.64	LOW
Morning serum cortisol **(threshold 310 nmol/L nmol/L)** to detect AI in people with suspected AI							
1 diagnostic study	208	Not serious	Not serious	Not serious	Serious[6]	Sensitivity=0.96 (0.89-0.99)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity=0.41 (0.32-0.50)	HIGH
Morning serum cortisol **(threshold <382 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
1 cross-sectional study	55	Very serious[1]	Not serious	Not serious	Unclear[2]	Sensitivity≥95%	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	NR	VERY LOW
Morning serum cortisol **(threshold <420 nmol/L)** to detect AI in people with suspected AI							
1 cross-sectional study	72	Very serious[3]	Not serious	Not serious	Unclear[2]	Sensitivity 1.00 (no false negatives)	VERY LOW
		Very serious[3]	Not serious	Not serious	Unclear[2]	Not reported	VERY LOW
Morning serum cortisol **(threshold <470 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
1 cross-sectional study	77	Very serious[1]	Not serious	Not serious	Unclear[2]	Sensitivity≥95%	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	NR	VERY LOW

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and patient flow.
[2] No confidence interval reported and primary data not available so unable to calculate imprecision. Downgraded by 2 increments.

[3] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and inappropriate time interval between the index test and reference standard

[4] The evidence was downgraded by 1 increment due to serious indirectness (serious population indirectness due to concerns over the conduction of the index test. Basal plasma cortisol was measured on 2 consecutive days and the mean of these two cortisol measures is reported).

[5] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'low sensitivity' (60%).

[6] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high sensitivity' (90%)

Table 14: Clinical evidence summary: diagnostic test accuracy for morning salivary cortisol

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning salivary cortisol **(threshold 1 nmol/L)** to detect AI in people with suspected A							
1 diagnostic study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.52 (0.41-0.62)	HIGH
		Not serious	Not serious	Not serious	Not serious	Specificity=0.97 (0.91-0.99)	HIGH
Morning salivary cortisol **(threshold <5.0 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
2 cross-sectional study	132	Very serious[1]	Not serious	Not serious	Unclear[2]	NR	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	Specificity ≥95%	VERY LOW
Morning salivary cortisol **(threshold 5.0 nmol/L)** to detect AI in people with suspected AI							
1 diagnostic study	208	Not serious	Not serious	Not serious	Serious[3]	Sensitivity=0.95 (0.88-0.98)	MODERATE
		Not serious	Not serious	Not serious	Not serious	Specificity=0.61 (0.51-0.70)	HIGH
Morning salivary cortisol **(threshold <17.5 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							
1 cross-sectional study	55	Very serious[1]	Not serious	Not serious	Unclear[2]	Sensitivity≥95%	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	NR	VERY LOW
Morning salivary cortisol **(threshold <21.1 nmol/l)** to detect AI in people with suspected secondary adrenal insufficiency in patients with hypothalamic-pituitary illness							

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
1 cross-sectional study	77	Very serious[1]	Not serious	Not serious	Unclear[2]	Sensitivity≥95%	VERY LOW
		Very serious[1]	Not serious	Not serious	Unclear[2]	NR	VERY LOW

[1] Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 2 increments due to very high risk of bias arising from unclear reporting on patient selection and patient flow.
[2] No confidence interval reported and primary data not available so unable to calculate imprecision. Downgraded by 2 increments.
[3] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'high sensitivity' (90%).

Table 15: Clinical evidence summary: diagnostic test accuracy for morning salivary cortisone

Studies	N	Risk of bias	Inconsistency	Indirectness	Imprecision	Effect size (95%CI)	Quality
Morning salivary cortisone (threshold 7 nmol/L) to detect AI in people with suspected AI							
1 diagnostic study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.74 (0.63-0.82)	HIGH
		Not serious	Not serious	Not serious	Not serious	Specificity=0.97 (0.91-0.99)	HIGH
Morning salivary cortisone (threshold 17 nmol/L) to detect AI in people with suspected AI							
1 diagnostic study	208	Not serious	Not serious	Not serious	Not serious	Sensitivity=0.97 (0.91-0.99)	HIGH
		Not serious	Not serious	Not serious	Serious[1]	Specificity=0.56 (0.46-0.65)	MODERATE

[1] Downgraded by 1 increment as the confidence interval crossed one decision threshold (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity. Confidence interval crossed the decision threshold corresponding to 'Low Specificity' (50%).

2.1.8. Additional data not suitable for GRADE analysis

Two studies provided further information on the proportion of people with a diagnosis matching the reference test that could not be analysed or assessed in GRADE. This included one study that examined a 2-step diagnostic pathway, with basal serum or salivary cortisol initially, and then a short Synacthen test from those in whom uncertainty regarding the diagnosis remained. These results are summarised in **Table 16**. However, note that it is possible these studies were based on overlapping cohorts of patients but there was insufficient information to determine whether this was the case. A scatter plot of individual serum cortisol peak values during ITT plotted against matched basal serum and salivary cortisol levels is presented in Figure 9.

Table 16: Proportion of patients correctly diagnosed by index tests using upper and lower thresholds with ≥95% sensitivity and specificity, respectively.

Study	Index test	Reference standard	Proportion with correct diagnosis on the index test(s)
Serum cortisol			
Deutschbein 2009	Basal serum cortisol upper cutoff of 470 nmol/l, and a lower cutoff of 103 nmol	ITT (threshold: peak serum cortisol ≤500 nmol/L)	23% (77% required further tests)
Deutschbein 2009a	Basal serum cortisol: upper cutoff of 382 nmol/l, and lower cutoff of 103 nmol followed by short Synacthen test (90 minutes) in those with uncertain results (upper limit 686 nmol/Land lower limit 455 nmol/l)		27% (73% required further tests) 49% after additional test
Salivary cortisol			
Deutschbein 2009	Basal salivary cortisol upper cutoff of 21.1 nmol/l, and a lower cutoff of 5.0 nmol	ITT (threshold: peak serum cortisol ≤500 nmol/L)	33% (67% required further tests)
Deutschbein 2009a	Basal salivary cortisol upper cutoff of 17.5 nmol/l, and a lower cutoff of 5.0 nmol followed by short Synacthen test (90 minutes) in those with uncertain results (upper limit 135.5 nmol/Land lower limit 35.5 nmol/l)		35% after first test; 45% after additional test

2.1.9. Summary of thresholds

The evidence in the above sections is summarised in **Table 17** to highlight the thresholds reported for maximum sensitivity and specificity.

Table 17: Thresholds reported for maximum sensitivity and specificity.

Population	N	Reference standard	Threshold for maximum sensitivity	Threshold for maximum specificity
Index test: morning serum cortisol				
HPA illness and suspected AI	77	ITT	470 nmol/l	103 nmol/l
Suspected AI	72	ITT	420 nmol/l	112 nmol/l

Population	N	Reference standard	Threshold for maximum sensitivity	Threshold for maximum specificity
HPA illness and suspected AI	55	ITT	382 nmol/l	103 nmol/l
Suspected AI	58	ITT	260 nmol/l	160 nmol/l
Suspected AI	208	ACTH	310 nmol/l	152 nmol/l
Index test: morning salivary cortisol				
HPA illness and suspected AI	77	ITT	21.1 nmol/l	5.0 nmol/l
HPA illness and suspected AI	55	ITT	17.5 nmol/l	5.0 nmol/l
Suspected AI	208	ACTH	5 nmol/l	1.0 nmol/l
Index test: morning salivary cortisone				
Suspected AI	208	ACTH	17 nmol/l	7 nmol/l

2.1.10. Economic evidence

2.1.10.1. Included studies.

No health economic studies were included.

2.1.10.2. Excluded studies.

No relevant health economic studies were excluded due to assessment of limited applicability or methodological limitations.

See also the health economic study selection flow chart in Appendix B.

2.1.11. Economic model

This area was not prioritised for new cost-effectiveness analysis.

2.1.12. Unit costs

Relevant unit costs are provided above to aid consideration of cost effectiveness. See Table 10.

2.1.12.1. Economic

- No relevant economic evaluations were identified.

3. The committee's discussion and interpretation of the evidence

The committee discussion of the review on diagnostic thresholds for referral 2.4 is included here in the discussion of the review on initial investigations by non-specialists 2.3.

3.1.1. The outcomes that matter most

Diagnostic accuracy

The committee considered the diagnostic measures of sensitivity and specificity of the index tests; serum cortisol, salivary cortisol and salivary cortisone for detecting adrenal insufficiency. Clinical decision thresholds were set by the committee as sensitivity/specificity=0.9 and 0.7 above which a test would be recommended and 0.6 and 0.5 below which a test is of no clinical use.

The committee were interested in establishing whether any of the index tests had greater diagnostic accuracy than others in detecting adrenal insufficiency. Therefore, sensitivity was considered the most important measure, as adrenal insufficiency is a potentially critical condition and so the consequences of missing a diagnosis could have serious implications.

3.1.2. The quality of the evidence

Eleven cross-sectional studies or (single gate) diagnostic accuracy studies with prospective data were included in this review.

- Evidence for the following index tests were identified:
    - basal serum cortisol (8 studies)
    - basal salivary cortisol (7 studies)
    - basal salivary cortisone (2 studies)

Ten of these studies were in adults and one study, looking at basal serum cortisol, was conducted in children. All studies provided sensitivity and specificity data for the index tests at various cut points. Four studies reported accuracy data for two or three of the index tests which allowed direct comparison, and these were presented separately to the committee to aid their decision-making.

Diagnostic accuracy analyses were carried out at specific cut-off points for each index test in order to diagnose adrenal insufficiency compared to a relevant reference standard. The reference standards used varied between the studies. Six studies used the low-dose or standard-dose short Synacthen test, and five studies used the insulin tolerance test. Results were presented as sensitivity and specificity data along with a 2x2 table where this was possible to calculate.

The evidence available was limited and only eleven studies with small sample sizes of <100 participants were identified. Due to the heterogeneity between the studies, in terms of the different reference standards, including populations and assays employed, meta-analysis of the data was not possible, consequently, all outcomes are based on single studies.

The quality of the evidence ranged from very low to high quality but was most commonly low quality. Evidence was generally downgraded due to the risk of bias and imprecision around the effect estimate. Risk of bias, as determined by the QUADAS-2, was often rated high or very high due to concerns with patient selection, the timing between the index test and reference standard or concerns arising from patient flow due to no details on any missing

data. A number of outcomes were also downgraded due to imprecision. This arose from the confidence interval crossing one or two decision thresholds (thresholds set at 60% and 90% for sensitivity; 50% and 70% for specificity), or due to the study not reporting a confidence interval or sufficient primary data, meaning calculation of imprecision was not possible. In outcomes where the confidence intervals were not reported, it was downgraded twice for imprecision, as this was deemed the most suitable way to capture the uncertainty and poor reporting. One outcome was downgraded for indirectness which was due to the index test being carried out on two consecutive days and taking the mean result of both tests. The committee agreed this would not happen in clinical practice, so the outcome was included but downgraded.

No relevant diagnostic test accuracy studies of random cortisol, electrolytes or blood glucose were identified.

3.1.3. Benefits and harms

The evidence base for index tests (that can be performed by non-specialists) to diagnose adrenal insufficiency was limited. Consequently, the committee decided to use their expertise to inform the recommendations and supplement the available evidence. The sensitivities and specificities of the index tests varied widely within and between studies, depending on the different cut-offs and reference standards used.

Serum cortisol

Evidence for 8-9 am serum cortisol came from 8 studies and sensitivity and specificity data was available for 14 outcomes looking at different thresholds. Thresholds ranged from 112 nmol/Lto 500 nmol/l. Some studies reported lower cut-offs to optimise the specificity or higher cut-offs to maximise sensitivity, while some studies reported optimal thresholds based on area under the curve analyses.

No paired sensitivity and specificity data reached the decision thresholds (sensitivity/specificity=0.9/0.7) for being a clinically useful diagnostic test. However, the committee agreed that as this test would be used as an initial first-line test rather than a stand-alone diagnostic test, then reaching these values was not strictly necessary and avoiding false negative results would be the priority. The committee considered two outcomes with cut-offs of 260 nmol/Land 285nmol/Lthat reported the best-paired values for diagnostic accuracy and reached the decision threshold for sensitivity but just fell short of the threshold for specificity, however, these were based on very small studies of low quality so they did not take these into account in their decision making.

Salivary cortisol

Evidence for morning salivary cortisol came from 7 studies and sensitivity and specificity data was available for 11 outcomes looking at different thresholds. Thresholds ranged from 1 nmol/Lto 21 nmol/Lwith some studies reporting cut-offs to optimise sensitivity or specificity, although the majority reported optimal thresholds. The committee highlighted the large range in reported cut-offs across the outcomes and agreed this was due to differences in the assays employed, with more up-to-date assays enabling lower cut points to be used.

No paired sensitivity and specificity data reached the thresholds for being a clinically useful diagnostic test, however, a recent UK-based study (De Bono 2023[5]) reported a cut-off of 5.0 nmol/Lwith good, paired values for diagnostic accuracy and reached the threshold for sensitivity but fell just below for specificity.

The committee noted that the studies examining salivary cortisol were more up-to-date than those for serum cortisol and used newer assays with greater accuracy. They agreed that the

use of salivary cortisol instead of serum as a first line test is an emerging field and further research is required.

Salivary cortisone

Evidence for salivary cortisone was very limited and only came from 2 studies reporting 3 different cut offs (7 nmol/l, 12.5 nmol/Land 17nmol/l). While these outcomes reported paired diagnostic accuracy values similar to those reported for salivary and serum cortisol, the committee acknowledged that the evidence base was very limited and again further research is necessary.

First line testing

The committee considered the diagnostic accuracy data presented from the 3 index tests and concluded that the evidence base for serum cortisol and salivary were fairly similar with sensitivity and specificity data varying largely between the studies, and dependent on the cut-offs and assays used.

Due to the limited and low-quality evidence, the committee used their consensus opinion and clinical experience to help form recommendations. They agreed that serum cortisol should be used as a first line test for people suspected of having adrenal insufficiency. This could be performed in primary care or in hospital by non-specialists and should be based on signs or symptoms or associated risk factors of adrenal insufficiency which are identified in review F.

The committee agreed to make a strong 'offer' recommendation and to advise clinicians to use 9am serum cortisol tests over the other index tests, as they are already widely used in current practice, readily available, relatively inexpensive and would ultimately lead to cost savings if referrals for unnecessary short Synacthen tests can be avoided. The committee also identified specific cut points of when to refer to a specialist for further testing which are identified in review D and discussed below.

The committee specified that people taking oral oestrogen should stop taking it, 6 weeks prior to having their serum cortisol measured. They explained that oral oestrogen increases the circulating cortisol-binding globin levels which results in an increased total cortisol concentration. This leads to falsely elevated cortisol levels and inaccurate test readings. The committee suggested that if oral oestrogen is being used for hormone replacement therapy then it could be switched to a transdermal preparation which has less effect on the circulating cortisol-binding globin levels. Alternatively, if oestrogen is being used for contraceptive purposes, then other methods of contraception should be considered to avoid unplanned pregnancies.

The committee highlighted that an 8-9 am serum cortisol test must be carried out within this time window in order to measure peak levels. They explained that random cortisol tests carried out later in the day (i.e., 10 am) would not be of any clinical use as levels may be too low and could lead to unnecessary referrals. The committee agreed that random cortisol tests should not be carried out in primary care due to the reasons specified above.

The committee also wished to highlight that shift workers may have variation in diurnal rhythm and this should be taken into consideration when interpreting the results.

Despite not identifying any evidence for the measurement of electrolytes, the committee agreed that everyone who is being tested for serum cortisol should also have their electrolyte levels checked. They reasoned that these tests are quick and easy to perform alongside serum cortisol tests and would provide additional information to help diagnose adrenal insufficiency or indicate when a more urgent referral for emergency care is required (i.e., if the reading is below 125 mmol/l).

The committee discussed the use of salivary cortisol and cortisone as alternative first-line tests and acknowledged the growing evidence base in the field. They noted the benefits of patients being able to carry out these tests by themselves at home and avoiding the necessity for blood tests. They also noted that these tests are being performed with newer assays with improved diagnostic accuracy. However, the committee concluded that as these tests are not routinely used in the UK there are not enough facilities available for them to be used effectively in a non-specialist setting. The committee explained that currently there are only two testing centres in the UK with machines that enable the analysis of mass spectrometry assays. Therefore, tests would need to be sent away to these centres and results may take up to five weeks to process, which is not practical or clinically appropriate in the majority of scenarios. The committee therefore agreed that a research recommendation, looking at the diagnostic accuracy, cut-offs for referral and cost effectiveness of salivary cortisol and cortisone tests for detecting adrenal insufficiency would be useful in order to establish efficacy and encourage implementation in the UK. See Appendix J.

Thresholds for test results

The committee considered the cut-offs reported in the evidence in order to determine when to refer for further second-line testing. Cut-offs reported in the evidence for serum cortisol ranged from 112 nmol/Lto 500 nmol/l. While two cut points of 260 nmol/Land 285 nmol/Lproduced the highest paired sensitivity and specificity values, there was no clear trend throughout the evidence to support any particular cut-offs for 9 am serum cortisol tests.

The committee discussed the difficulties in setting cut-off points, as these vary greatly depending on the assay used and only have clinical use if specific to a particular assay. However, the committee agreed that it would be useful for non-specialists to have some guidance on at what point to refer onwards (providing it is highlighted that the cut-offs are only for use with modern immunoassay assays and that local guidelines may need to be followed if alternative assays are used).

When interpreting 8am to 9am cortisol results it is important to take into account clinical context- patients with a symptoms of adrenal insufficiency together with hyponatremia may warrant discussion with endocrinology especially if cortisol is between 150-200nmol/L. They may have developed acute adrenal insufficiency related to other treatments such as check point inhibitors.

The committee decided to set a cut off for onward referral at ≤150 nmol/Lor for repeat testing if the value is between 150 – 300 nmol/l. This was partly based on the evidence which suggested high sensitivity (>95%) at 265 and 280 nmol/Land on consensus opinion. The committee also highlighted that the commonly used Elecsys Cortisol II assay suggests a cut-off of 300 nmol/l.

The committee discussed the option of setting high cut-offs to maximise sensitivity and ensure no one with the condition is missed. However, they deliberated that high cut-offs would increase the risk of false positive results and lead to increased numbers of referrals and unnecessary tests. Therefore, the committee agreed on the cut-off of 150 nmol/l, and anyone with a reading below this figure should be referred to a specialist for a nurse led short Synacthen test. They agreed that a referral to a consultant would not routinely be necessary.

The committee also defined a range of values from 150 to 300 nmol/Lin which the results of the 9 am serum cortisol may not be an adequate indicator of adrenal insufficiency on its own. They explained that when these readings are combined with further evidence in the form of clinical presentation or with results from electrolyte readings, they may be more indicative of adrenal insufficiency. The committee deliberated over the upper threshold of this 'grey zone', weighing up the risk of false negative results against the impact of offering more people additional testing. They therefore decided that people should be retested if the number was

between 150 and 300 nmol/l/ and if the reading remains in this 'grey zone' they should then be referred to endocrinology if serum cortisol remains at that level.

The committee considered setting a lower threshold, below which immediate treatment should be provided. They agreed that if the cortisol level is below 150 nmol/L then immediate management of adrenal insufficiency should be initiated (as outlined in section 1.3 of the guideline document) and if the person is acutely unwell emergency management (as outlined in section 1.7 of the guideline document) should be initiated. A referral to endocrinology should also be made. For people with readings above the upper threshold of 300 nmol/l, adrenal insufficiency is very unlikely, and an alternative diagnosis should be considered if their symptoms persist.

The committee concluded that 8-9 am serum cortisol testing should be used as a first-line diagnostic test to screen for people 1 year and over who present with signs and symptoms indicative of adrenal insufficiency. Due to the variation in diurnal rhythm, for babies under 1 year, serum cortisol levels can be measured at any time of day but paediatric or paediatric endocrinology advice for interpretation of results should be sought. As these tests are only a first screen for adrenal insufficiency the committee decided to set the threshold for further investigation or repeat testing fairly high at <300nmol/Lin order to avoid any false negative results and missed diagnoses. Correct utilisation of these tests should hopefully reduce the need to perform more costly and time-consuming short Synacthen tests in people presenting with symptoms of adrenal insufficiency.

The committee agreed that these thresholds could be used for children and adults with suspected adrenal insufficiency including those withdrawing from glucocorticoids (i.e. population in evidence review C).

3.1.4. Cost-effectiveness and resource use

No health economic studies were identified for this review. Unit costs were presented and discussed with the committee to aid the committee's consideration of cost-effectiveness.

Serum cortisol tests were estimated to be £6 per test. The costs include the cost of a blood test taken either in the community at a General Practitioner's surgery by a healthcare assistant (ten minutes of a Band 3 health care assistant time £4.33) or in an outpatient hospital setting by a phlebotomist (£4.70) and the cost of laboratory analysis (£1.55 for clinical biochemistry). The committee recommended 9am serum cortisol tests over the other index tests as they are readily available, relatively inexpensive and would ultimately lead to cost savings if referrals for unnecessary short Synacthen tests can be avoided.

The committee specified that an 8-9 am serum cortisol test must be carried out within this time window to measure peak levels. They explained that tests carried out later out in the day (i.e., 10am) would not be of any clinical use as levels may be too low and could lead to unnecessary referrals. This is also true of random cortisol testing which is occurring in current practice. The committee also noted that secondary care will often ask for 9 am cortisol to be done before accepting a referral. The 'do not do random cortisol testing' recommendation may result in cost savings as this will avoid unnecessary duplication of tests or inappropriate referrals.

Providing serum cortisol testing as a first test is considered current practice, although the 8-9 am timing is not well adhered to. The recommendations made are unlikely to result in a significant increase in serum cortisol tests but are likely to reduce unnecessary secondary care referrals and testing, which in turn could result in resource savings.

The committee decided to set a cut-off for onward referral at ≤200 nmol/Lor for repeat testing if the value is between 201 – 300 nmol/l. This was partly based on the evidence which suggested high sensitivity (>95%) at 265 and 280 nmol/Land on consensus opinion. The committee deliberated that higher cut-offs would increase the risk of false positive results and

lead to increased cost to the NHS associated with unnecessary referrals and short Synacthen testing. In terms of retesting, this may be a change in current practice, however, the committee considered that repeating a serum cortisol test was less costly than unnecessary referrals based on a single test with numbers between 201 and 300 nmol/l, which may be happening in current practice.

Salivary cortisol testing includes the cost of the 'Salivette' to collect the saliva and then the laboratory and postage costs. The former was estimated to be £0.37 per unit, the latter £27.80. The committee noted that salivary cortisol testing is not widely used and there are only two testing centres in the UK with machines that enable the analysis of mass spectrometry assays. It was noted that there could be significant set-up costs if it were to be recommended due to the current lack of infrastructure available for analysing these samples nationally. The committee noted that reliance on the available laboratories may result in delays in results of up to 5 weeks which would be impractical and possibly clinically inappropriate. Given the health economic uncertainty and the lack of use in current practice, the committee agreed that a research recommendation looking at the diagnostic accuracy and cuts-offs for referral (including cost-effectiveness) for salivary cortisol tests would be useful.

The unit cost of referral to secondary endocrinology care was also presented, this was £293 for adults and £418 for paediatrics (a weighted average of consultant-led, non-consultant led and multi-professional cost of first endocrinology face-to-face appointment). In addition, the unit cost of endocrinology 'advice and guidance' was estimated to be between £12 and £24 depending on the duration (5 to 10 minutes). The unit cost of a short Synacthen test was presented as a day case cost (£398). The cost of Synacthen would be included in the day case cost as it is not listed as a high-cost drug. This is a broad HRG unit cost which will cover many interventions and so may not provide an exact estimate of the cost of this test. In addition, the committee noted that the unit cost of Synacthen in hospitals currently is higher than that listed in the drug tariff / BNF (£45 versus £38). Insulin tolerance tests are likely to be bundled in the same day case cost as a short Synacthen test, but the committee noted it is likely to be more expensive as it is a more resource-intensive test. The cost of an insulin tolerance test was calculated as part of the NICE Medtech innovation briefing [MIB320] on 'Macimorelin for diagnosing growth hormone deficiency' at £470.

3.1.5. Recommendations supported by this evidence review.

This evidence review supports recommendations1.2.5 – 1.2.13 and the recommendation for research on the clinical and cost-effectiveness of salivary cortisone or cortisol to identify people with adrenal insufficiency.

References

1. Agwu JC, Spoudeas H, Hindmarsh PC, Pringle PJ, Brook CG. Tests of adrenal insufficiency. Archives of Disease in Childhood. 1999; 80(4):330-333

2. BMJ Group and the Royal Pharmaceutical Society of Great Britain. British National Formulary. 2023. Available from: https://bnf.nice.org.uk/ Last accessed: 05/11/2023.

3. Choi CH, Tiu SC, Shek CC, Choi KL, Chan FKW, Kong PS. Use of the low-dose corticotropin stimulation test for the diagnosis of secondary adrenocortical insufficiency. Hong Kong medical journal = Xianggang yi xue za zhi. 2002; 8(6):427-434

4. de Lange WE, Sluiter WJ. The assessment of the functional capacity of the hypothalamic-pituitary-adrenal axis by measurement of basal plasma and urinary cortisol in comparison with insulin-induced hypoglycaemia and metyrapone tests. The Netherlands journal of medicine. 1993; 43(12):64-68

5. Debono M, Caunt S, Elder C, Fearnside J, Lewis J, Keevil B et al. Home waking salivary cortisone to screen for adrenal insufficiency. NEJM Evidence. 2023; 2(2)

6. Deutschbein T, Unger N, Mann K, Petersenn S. Diagnosis of secondary adrenal insufficiency in patients with hypothalamic-pituitary disease: comparison between serum and salivary cortisol during the high-dose short Synacthen test. European Journal of Endocrinology. 2009; 160(1):9-16

7. Deutschbein T, Unger N, Mann K, Petersenn S. Diagnosis of secondary adrenal insufficiency: unstimulated early morning cortisol in saliva and serum in comparison with the insulin tolerance test. Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme. 2009; 41(11):834-839

8. Falorni A, Laureti S. Adrenal autoimmunity and correlation with adrenal dysfunction. Endocrinologist. 2000; 10(3):145-154

9. George GS, Jabbar PK, Jayakumari C, John M, Mini M, Thekkumkara Surendran Nair A et al. Long-acting porcine ACTH stimulated salivary cortisol in the diagnosis of adrenal insufficiency. Clinical Endocrinology. 2020; 93(6):652-660

10. Jones K, Weatherly H, Birch S, Castelli A, Chalkley M, Dargan A et al. Unit costs of health and social care 2022 manual. Canterbury. Personal Social Services Research Unit (University of Kent) & Centre for Health Economics (University of York) K, 2023. Available from: https://www.pssru.ac.uk/unitcostsreport/

11. Kim YJ, Kim JH, Hong AR, Park KS, Kim SW, Shin CS et al. Stimulated salivary cortisol as a noninvasive diagnostic tool for adrenal insufficiency. Endocrinology and metabolism (Seoul, Korea). 2020; 35(3):628-635

12. Lab Unlimited UK. Sarstedt Salivette for cortisol determination 51.1534.500. 2023. Available from: https://www.labunlimited.co.uk/s/ALL/4AJ-4667984/Sarstedt-Salivette%C2%AE-for-cortisol-determination-51.1534.500 Last accessed: 09/11/2023.

13. Mak IYF, Au Yeung BYT, Ng YW, Choi CH, Iu HYP, Shek CC et al. Salivary cortisol and cortisone after low-dose corticotropin stimulation in the diagnosis of adrenal insufficiency. Journal of the Endocrine Society. 2017; 1(2):96-108

14. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London. National Institute for Health and Care Excellence, 2014. Available from: https://www.nice.org.uk/process/pmg20/chapter/introduction

15. National Institute for Health and Care Excellence. Macimorelin for diagnosing growth hormone deficiency. Medtech innovation briefing [MIB320]. London. 2023. Available from: https://www.nice.org.uk/advice/mib320

16. NHS England. National Cost Collection 2021-22. 2023. Available from: https://www.england.nhs.uk/publication/2021-22-national-cost-collection-data-publication/ Last accessed: 27/11/2023.

17. Schmidt IL, Lahner H, Mann K, Petersenn S. Diagnosis of adrenal insufficiency: Evaluation of the corticotropin-releasing hormone test and Basal serum cortisol in comparison to the insulin tolerance test in patients with hypothalamic-pituitary-adrenal disease. The Journal of clinical endocrinology and metabolism. 2003; 88(9):4193-4198

18. Tan SYT, Tan HC, Zhu L, Loh LM, Lim DST, Swee DS et al. Expanding the use of salivary cortisol as a non-invasive outpatient test in the dynamic evaluation of suspected adrenal insufficiency. Endocrine connections. 2023; 12(4):e230004

Appendices

Appendix A Review protocols

A.1 Review protocol for [add key area, for example, unplanned hospital admission]

ID	Field	Content
1.	Review title	Initial investigations by non-specialists
2.	Review question	2.3 What initial investigations should be done by the non-specialist for people with suspected adrenal insufficiency?
3.	Objective	To determine which initial tests, that can be done by non-specialists, are most accurate in identifying people who may have adrenal insufficiency.
4.	Searches	The following databases (from inception) will be searched: • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE • Epistemonikos Searches will be restricted by: • English language studies • Human studies The searches may be re-run 6 weeks before the final committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review. Medline search strategy to be quality assured using the PRESS evidence-based checklist (see methods chapter for full details).

5.	Condition or domain being studied	Adrenal Insufficiency
6.	Population	Inclusion: People with suspected adrenal insufficiency Note: Could be well or unwell. Could be in primary or in hospital not under the care of endocrinology. Exclusion: Critically ill patients
7.	Test	• Serum cortisol (8-9 am) • Salivary cortisol • Random cortisol • Electrolytes • Blood glucose • Combination of tests may be included. Exclude: Specialist test for example those that determine the type or cause of AI such as: • plasma ACTH test • Corticotropin releasing hormone stimulation test. • DHEAs **Note for reviewers:** The short Synacthen test, also known as the cosyntropin, tetracosactide or ACTH stimulation test, is a ‘dynamic' test where a fragment of ACTH (1-24) is given to the patient (IV or IM) and the serum cortisol is measured after 30 or 60 minutes. So, this maximally stimulates the adrenal glands, and you measure how much cortisol is produced. It’s a way to get rid of the normal pulsatile ultradian rhythm and get a standard measure of adrenal function. The standard dose of Synacthen/cosyntropin/tetracosactide is 250mcg, but there is some literature on ‘low-dose’ Synacthen tests, using 1mcg.

		A plasma ACTH test is a direct measurement of the circulating ACTH level normally reading out in pmol/L or pg/dL, whereas the Synacthen test readout is cortisol measured in nmol/L or ng/dL.
8.	Reference standard	Short Synacthen Test (standard and low dose) Or Insulin tolerance test (insulin hypoglycaemia test) Or Clinical diagnosis by a specialist (a specialist will take into account the full clinical picture, including signs, symptoms, risk factors and test results)
9.	Types of study to be included	• Cross-sectional (single gate) studies • Systematic reviews of diagnostic accuracy studies • If no or insufficient diagnostic accuracy studies are identified, prospective cohort studies may be included
10.	Other exclusion criteria	Non comparative cohort studies Before and after studies Case control studies Retrospective cohort studies will only be included if no cross-sectional or prospective cohort studies are identified. Non-English language studies. Conference abstracts will be excluded because they are unlikely to contain enough information to assess whether the population matches the review question in terms of previous medication use, or enough detail on outcome definitions, or on the methodology to assess the risk of bias of the study.
11.	Context	
12.	Primary outcomes (critical outcomes)	Diagnostic accuracy data • Sensitivity (prioritised) *[fewer false negatives i.e., very few people with the condition will be missed]* • Specificity The GC has prioritised sensitivity and specificity as the most important outcomes for their interpretation of the evidence.

		The following thresholds will be used for imprecision for DTA measures and for deciding on the usefulness of the tests in detecting adrenal insufficiency: Sensitivity • Upper 0.9 • Lower 0.6 Specificity • Upper 0.7 • Lower 0.5 Likelihood ratios or other measures such as C statistic or area under ROC curve will only be reported if they are the only measures available, sensitivity and specificity are not reported and cannot be calculated from raw data. Should this be the case, cut-offs for summarising the performance of diagnostic tests or prediction models will be agreed with the guideline committee before the analysis of the evidence is conducted and the protocol will be updated accordingly.
13.	Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. All references identified by the searches and from other sources will be uploaded into EPPI reviewer and de-duplicated. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. The full text of potentially eligible studies will be retrieved and will be assessed in line with the criteria outlined above. A standardised form will be used to extract data from studies (see Developing NICE guidelines: the manual section 6.4). 10% of all evidence reviews are quality assured by a senior research fellow. This includes checking: • papers were included /excluded appropriately. • a sample of the data extractions. • correct methods are used to synthesise data. • a sample of the risk of bias assessments. Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary.

		Study investigators may be contacted for missing data where time and resources allow.	
14.	Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual. These may include: • Systematic reviews: Risk of Bias in Systematic Reviews (ROBIS) • Nonrandomised study, including cohort studies: Cochrane ROBINS-I • Cross sectional study: JBI checklist for cross sectional study • Check list for diagnostic test accuracy studies: QUADAS-2	
15.	Strategy for data synthesis	Diagnostic meta-analysis using Cochrane Review Manager (RevMan5) will be conducted where appropriate. Heterogeneity between the studies in effect measures will be assessed using the I^2 statistic and visually inspected. An I^2 value greater than 50% will be considered indicative of substantial heterogeneity. Sensitivity analyses will be conducted based on pre-specified subgroups using stratified meta-analysis to explore the heterogeneity in effect estimates. If this does not explain the heterogeneity, the results will be presented pooled using random effects. Where meta-analysis is not possible, data will be presented, and quality assessed individually per outcome. GRADEpro will be used to assess the quality of evidence for each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency, and imprecision) will be appraised for each outcome. Publication bias will be considered with the guideline committee, and if suspected will be tested for when there are more than 5 studies for that outcome. The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/	
16.	Analysis of sub-groups	Subgroups that will be investigated if heterogeneity is present: None identified	
17.	Type and method of review	☐	Intervention
		☒	Diagnostic

		☐	Prognostic	
		☐	Qualitative	
		☐	Epidemiologic	
		☐	Service Delivery	
		☐	Other (please specify)	
18.	Language	English		
19.	Country	England		
20.	Anticipated or actual start date			
21.	Anticipated completion date			
22.	Stage of review at time of this submission	Review stage	Started	Completed
		Preliminary searches	☑	☑
		Piloting of the study selection process	☑	☑
		Formal screening of search results against eligibility criteria	☑	☑
		Data extraction	☑	☑
		Risk of bias (quality) assessment	☑	☑
		Data analysis	☑	☑
23.	Named contact	5a. Named contact Guideline Development Team NGC 5b Named contact e-mail.		

		Hypoadrenalism@nice.org.uk 5e Organisational affiliation of the review National Institute for Health and Care Excellence (NICE)
24.	Review team members	From NICE: Sharon Swain [Guideline lead] Saoussen Ftouh [Senior systematic reviewer] Alexandra Bannon [Health economist] Stephen Deed [Information specialist]
25.	Funding sources/sponsor	Development of this systematic review is being funded by NICE.
26.	Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
27.	Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10237.
28.	Other registration details	
29.	Reference/URL for published protocol	-
30.	Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication

		• publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.	
31.	Keywords	-	
32.	Details of existing review of same topic by same authors	-	
33.	Current review status	☐	Ongoing
		☒	Completed but not published
		☐	Completed and published
		☐	Completed, published and being updated
		☐	Discontinued
34.	Additional information	-	
35.	Details of final publication	www.nice.org.uk	

A.2 Review protocol for initial investigations by non-specialists – thresholds for referral

ID	Field	Content
1.	Review title	When to refer for specialist investigation
2.	Review question	2.4 When should people with suspected adrenal insufficiency be referred to specialists for further investigation?
3.	Objective	To determine when a person with suspected AI should be referred to a specialist based on specific cut-offs for cortisol tests.

4.	Searches	The following databases (from inception) will be searched: • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE • Epistemonikos Searches will be restricted by: • English language studies • Human studies The searches may be re-run 6 weeks before the final committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review. Medline search strategy to be quality assured using the PRESS evidence-based checklist (see methods chapter for full details).
5.	Condition or domain being studied	Adrenal Insufficiency
6.	Population	Inclusion: People suspected of having adrenal insufficiency. Exclusion: None identified
7.	Test	Diagnostic accuracy based on cut-off: • Cortisol Tests –8- 9 am • Salivary cortisol • Salivary cortisone • Short Synacthen test

		• ACTH and cortisol Note assay specific cut-offs and which assays being used – exclude if they don't' state the assay.
8.	Reference standard	Short Synacthen Test (standard and low dose) Or Insulin tolerance test (insulin hypoglycaemia test) Or Clinical diagnosis by a specialist (a specialist will take into account the full clinical picture, including signs, symptoms, risk factors and test results)
9.	Types of study to be included	• Cross sectional (single gate) diagnostic studies • Systematic reviews of diagnostic accuracy studies • If no or insufficient diagnostic accuracy studies are identified, prospective cohort studies may be included
10.	Other exclusion criteria	Non comparative cohort studies Before and after studies Case control studies Retrospective cohort studies will only be included if no cross-sectional or prospective cohort studies are identified. Non-English language studies. Conference abstracts will be excluded because they are unlikely to contain enough information to assess whether the population matches the review question in terms of previous medication use, or enough detail on outcome definitions, or on the methodology to assess the risk of bias of the study.
11.	Context	-
12.	Primary outcomes (critical outcomes)	Diagnostic accuracy data • Sensitivity (prioritised) *[fewer false negatives i.e. very few people with the condition will be missed]*

		• Specificity The GC has prioritised sensitivity and specificity as the most important outcomes for their interpretation of the evidence. The following thresholds will be used for imprecision for DTA measures and for deciding on the usefulness of the tests in detecting adrenal insufficiency: Sensitivity • Upper 0.9 • Lower 0.6 Specificity • Upper 0.7 • Lower 0.5 Likelihood ratios or other measures such as C statistic or area under ROC curve will only be reported if they are the only measures available, sensitivity and specificity are not reported and cannot be calculated from raw data. Should this be the case, cut-offs for summarising the performance of diagnostic tests or prediction models will be agreed with the guideline committee before the analysis of the evidence is conducted and the protocol will be updated accordingly.
13.	Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. All references identified by the searches and from other sources will be uploaded into EPPI reviewer and de-duplicated. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. The full text of potentially eligible studies will be retrieved and will be assessed in line with the criteria outlined above. A standardised form will be used to extract data from studies (see Developing NICE guidelines: the manual section 6.4). 10% of all evidence reviews are quality assured by a senior research fellow. This includes checking: • papers were included /excluded appropriately. • a sample of the data extractions • correct methods are used to synthesise data. • a sample of the risk of bias assessments

		Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary. Study investigators may be contacted for missing data where time and resources allow.	
14.	Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual. These may include: • Systematic reviews: Risk of Bias in Systematic Reviews (ROBIS) • Nonrandomised study, including cohort studies: Cochrane ROBINS-I • Cross sectional study: JBI checklist for cross sectional study • Check list for diagnostic test accuracy studies: QUADAS-2	
15.	Strategy for data synthesis	Diagnostic meta-analysis using Cochrane Review Manager (RevMan5) will be conducted where appropriate. Heterogeneity between the studies in effect measures will be assessed using the I^2 statistic and visually inspected. An I^2 value greater than 50% will be considered indicative of substantial heterogeneity. Sensitivity analyses will be conducted based on pre-specified subgroups using stratified meta-analysis to explore the heterogeneity in effect estimates. If this does not explain the heterogeneity, the results will be presented pooled using random effects. Where meta-analysis is not possible, data will be presented, and quality assessed individually per outcome. GRADEpro will be used to assess the quality of evidence for each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency, and imprecision) will be appraised for each outcome. Publication bias will be considered with the guideline committee, and if suspected will be tested for when there are more than 5 studies for that outcome. The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/	
16.	Analysis of sub-groups	[Subgroups that will be investigated if heterogeneity is present: None identified	
17.		☐	Intervention

	Type and method of review	☒	Diagnostic	
		☐	Prognostic	
		☐	Qualitative	
		☐	Epidemiologic	
		☐	Service Delivery	
		☐	Other (please specify)	
18.	Language	English		
19.	Country	England		
20.	Anticipated or actual start date			
21.	Anticipated completion date			
22.	Stage of review at time of this submission	Review stage	Started	Completed
		Preliminary searches	☑	☑
		Piloting of the study selection process	☑	☑
		Formal screening of search results against eligibility criteria	☑	☑
		Data extraction	☑	☑
		Risk of bias (quality) assessment	☑	☑
		Data analysis	☑	☑
23.	Named contact	5a. Named contact Guideline Development Team NGC		

		5b Named contact e-mail Hypoadrenalism@nice.org.uk 5e Organisational affiliation of the review National Institute for Health and Care Excellence (NICE)
24.	Review team members	Sharon Swain [Guideline lead] Saoussen Ftouh [Senior systematic reviewer] [Systematic reviewer] Alexandra Bonnon [Health economist] Stephen Deed [Information specialist]
25.	Funding sources/sponsor	Development of this systematic review is being funded by NICE.
26.	Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
27.	Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10237.
28.	Other registration details	-
29.	Reference/URL for published protocol	-
30.	Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as:

		• notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.	
31.	Keywords	-	
32.	Details of existing review of same topic by same authors	-	
33.	Current review status	☐	Ongoing
		☒	Completed but not published
		☐	Completed and published
		☐	Completed, published and being updated
		☐	Discontinued
34.	Additional information		
35.	Details of final publication	www.nice.org.uk	

A.3 Health economic review protocol

Table 18: Health economic review protocol

Review question	All questions – health economic evidence
Objectives	To identify health economic studies relevant to any of the review questions.
Search criteria	• Populations, interventions, and comparators must be as specified in the clinical review protocol above. • Studies must be of a relevant health economic study design (cost–utility analysis, cost-effectiveness analysis, cost–benefit analysis, cost–consequences analysis, comparative cost analysis). • Studies must not be a letter, editorial or commentary, or a review of health economic evaluations. (Recent reviews will be ordered although not reviewed. The bibliographies will be checked for relevant studies, which will then be ordered.) • Unpublished reports will not be considered unless submitted as part of a call for evidence. • Studies must be in English.
Search strategy	A health economic study search will be undertaken using population-specific terms and a health economic study filter – see appendix B below.
Review strategy	Studies not meeting any of the search criteria above will be excluded. Studies published before 2007, abstract-only studies and studies from non-OECD countries or the USA will also be excluded. Each remaining study will be assessed for applicability and methodological limitations using the NICE economic evaluation checklist which can be found in appendix H of Developing NICE guidelines: the manual (2014).[14] **Inclusion and exclusion criteria** • If a study is rated as both 'Directly applicable' and with 'Minor limitations', then it will be included in the guideline. A health economic evidence table will be completed, and it will be included in the health economic evidence profile. • If a study is rated as either 'Not applicable' or with 'Very serious limitations', then it will usually be excluded from the guideline. If it is excluded, then a health economic evidence table will not be completed, and it will not be included in the health economic evidence profile. • If a study is rated as 'Partially applicable', with 'Potentially serious limitations' or both then there is discretion over whether it should be included. **Where there is discretion** The health economist will make a decision based on the relative applicability and quality of the available evidence for that question, in discussion with the guideline committee if required. The ultimate aim is to include health economic studies that are helpful for decision-making in the context of the guideline and the current NHS setting. If several studies are considered of sufficiently high applicability and methodological quality that they could all be included, then the health economist, in discussion with the committee if required, may decide to include only the most applicable studies and to selectively exclude the remaining studies. All studies excluded on the basis of applicability or methodological limitations will be listed with explanation in the excluded health economic studies appendix below. The health economist will be guided by the following hierarchies. *Setting:*

	• UK NHS (most applicable). • OECD countries with predominantly public health insurance systems (for example, France, Germany, Sweden). • OECD countries with predominantly private health insurance systems (for example, Switzerland). • Studies set in non-OECD countries or in the USA will be excluded before being assessed for applicability and methodological limitations. *Health economic study type:* • Cost–utility analysis (most applicable). • Other type of full economic evaluation (cost–benefit analysis, cost-effectiveness analysis, cost–consequences analysis). • Comparative cost analysis. • Non-comparative cost analyses including cost-of-illness studies will be excluded before being assessed for applicability and methodological limitations. *Year of analysis:* • The more recent the study, the more applicable it will be. • Studies published in 2007 or later but that depend on unit costs and resource data entirely or predominantly from before 2007 will be rated as 'Not applicable'. • Studies published before 2007 be excluded before being assessed for applicability and methodological limitations. *Quality and relevance of effectiveness data used in the health economic analysis:* • The more closely the clinical effectiveness data used in the health economic analysis match with the outcomes of the studies included in the clinical review the more useful the analysis will be for decision-making in the guideline.

Appendix B Literature search strategies

The literature searches for this review are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual.[14]

For more information, please see the Methodology review published as part of the accompanying documents for this guideline.

B.1 Clinical search literature search strategy

Searches were constructed using a PICO framework where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies as these concepts may not be indexed or described in the title or abstract and are therefore difficult to retrieve. Search filters were applied to the search where appropriate.

Table 19: Database parameters, filters and limits applied.

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 26 September 2023	Systematic review studies Observational studies Diagnostic studies Exclusions (animal studies, letters, comments, editorials, case studies/reports) English language
Embase (OVID)	1974 – 26 September 2023	Systematic review studies Observational studies Diagnostic studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
The Cochrane Library (Wiley)	Cochrane Database of Systematic Reviews to Issue 9 of 12, 26 September 2023	Exclusions (clinical trials, conference abstracts)
Epistemonikos (The Epistemonikos Foundation)	Inception to 26 September 2023	Systematic review Exclusions (Cochrane reviews)

Medline (Ovid) search terms

1.	exp Adrenal Insufficiency/
2.	Adrenal Hyperplasia, Congenital/
3.	(addison* disease or addisonian*).ti,ab,kf.
4.	((adrenal* or adrenocort* or adreno cort*) adj3 (insufficien* or inadequa* or deficien* or suppress* or hypofunction* or disorder* or underactiv* or dysfunction* or abnormal* or

	problem* or crisis or crises or dysgenesis or destruction or destroy* or hyperplasia or hypoplasia or failure* or fails or failed or fatigue or inhibit* or damage* or disruption*)).ti,ab,kf.
5.	((cortisol or aldosterone or adrenocorticotrop* or adreno corticotrop* or ACTH or corticotropi* releas* or corticotrophi* releas* or corticoliberin or CRH) adj3 (insufficien* or inadequa* or deficien* or suppress* or reduc* or decreas* or descend* or diminish* or lack* or less or lessen* or low or lower* or limited)).ti,ab,kf.
6.	(hypoadrenal* or hypo adrenal* or hypoadrenocorticism or hypo adrenocorticism or adrenoleukodystrophy or adreno leukodystrophy or adrenomyeloneuropathy or adreno myeloneuropathy or hypoaldosteronism or hypo aldosteronism).ti,ab,kf.
7.	((adrenogenital or adreno genital) adj (syndrome or disorder*)).ti,ab,kf.
8.	((haemorrhag* or hemorrhag* or bleed*) adj3 adrenal*).ti,ab,kf.
9.	(Bronze Schilder* Disease or Melanodermic Leukodystrophy or Schilder-Addison* Complex or Siemerling-Creutzfeldt* Disease).ti,ab,kf.
10.	((Allgrove or 3A or TripleA or AAA) adj syndrome).ti,ab,kf.
11.	(CAH or X-ALD).ti,ab.
12.	(Waterhouse-Friderichsen* syndrome or antiphospholipid syndrome).ti,ab,kf.
13.	Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy.ti,ab,kf.
14.	or/1-13
15.	letter/
16.	editorial/
17.	news/
18.	exp historical article/
19.	Anecdotes as Topic/
20.	comment/
21.	case reports/
22.	(letter or comment*).ti.
23.	or/15-22
24.	randomized controlled trial/ or random*.ti,ab.
25.	23 not 24
26.	animals/ not humans/
27.	exp Animals, Laboratory/
28.	exp Animal Experimentation/
29.	exp Models, Animal/
30.	exp Rodentia/
31.	(rat or rats or mouse or mice or rodent*).ti.
32.	or/25-31
33.	14 not 32
34.	limit 33 to English language
35.	Hydrocortisone/
36.	((cortisol or cortisone) adj3 (assay* or test* or level* or measure* or serum or blood or plasma or saliva* or random or 8am or "8 am" or "8 a.m" or 9am or "9 am" or "9 a.m" or "8-9 am" or "8-9 a.m" or morning or awaken*)).ti,ab,kf.
37.	Electrolytes/
38.	electrolyte*.ti,ab,kf.
39.	Blood Glucose/
40.	((blood or serum or plasma) adj3 (sugar or glucose)).ti,ab,kf.
41.	(insulin adj4 test*).ti,ab,kf.
42.	(Synacthen or tetracosactide).ti,ab,kf.
43.	exp Adrenocorticotropic Hormone/

44.	((adrenocorticotrop* or "adreno corticotrop*" or corticotrop* or ACTH or cosyntrop*) adj3 (assay* or test* or level* or measure* or stimulation)).ti,ab,kf.
45.	Dehydroepiandrosterone Sulfate/
46.	((DHEA* or "dehydroepiandrosterone sulfate" or "dehydroepiandrosterone sulphate" or "prasterone sulfate" or "prasterone sulphate") adj3 (assay* or test* or level* or measure*)).ti,ab,kf.
47.	or/35-46
48.	34 and 47
49.	exp "sensitivity and specificity"/
50.	(sensitivity or specificity).ti,ab.
51.	((pre test or pretest or post test) adj probability).ti,ab.
52.	(predictive value* or PPV or NPV).ti,ab.
53.	likelihood ratio*.ti,ab.
54.	likelihood function/
55.	((area under adj4 curve) or AUC).ti,ab.
56.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
57.	gold standard.ab.
58.	exp Diagnostic errors/
59.	(false positiv* or false negativ*).ti,ab.
60.	Diagnosis, Differential/
61.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness or precision or validat* or validity or differential or error*)).ti,ab.
62.	or/49-61
63.	Epidemiologic studies/
64.	Observational study/
65.	exp Cohort studies/
66.	(cohort adj (study or studies or analys* or data)).ti,ab.
67.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
68.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
69.	Controlled Before-After Studies/
70.	Historically Controlled Study/
71.	Interrupted Time Series Analysis/
72.	(before adj2 after adj2 (study or studies or data)).ti,ab.
73.	exp case control study/
74.	case control*.ti,ab.
75.	Cross-sectional studies/
76.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
77.	or/63-76
78.	Meta-Analysis/
79.	Meta-Analysis as Topic/
80.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
81.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
82.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
83.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
84.	(search* adj4 literature).ab.

85.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
86.	cochrane.jw.
87.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
88.	or/78-87
89.	48 and (62 or 77 or 88)

Embase (Ovid) search terms

1.	exp Adrenal cortex insufficiency/
2.	Congenital adrenal hyperplasia/
3.	(addison* disease or addisonian*).ti,ab,kf.
4.	((adrenal* or adrenocort* or adreno cort*) adj3 (insufficien* or inadequa* or deficien* or suppress* or hypofunction* or disorder* or underactiv* or dysfunction* or abnormal* or problem* or crisis or crises or dysgenesis or destruction or destroy* or hyperplasia or hypoplasia or failure* or fails or failed or fatigue or inhibit* or damage* or disruption*)).ti,ab,kf.
5.	((cortisol or aldosterone or adrenocorticotrop* or adreno corticotrop* or ACTH or corticotropi* releas* or corticotrophi* releas* or corticoliberin or CRH) adj3 (insufficien* or inadequa* or deficien* or suppress* or reduc* or decreas* or descend* or diminish* or lack* or less or lessen* or low or lower* or limited)).ti,ab,kf.
6.	(hypoadrenal* or hypo adrenal* or hypoadrenocorticism or hypo adrenocorticism or adrenoleukodystrophy or adreno leukodystrophy or adrenomyeloneuropathy or adreno myeloneuropathy or hypoaldosteronism or hypo aldosteronism).ti,ab,kf.
7.	((adrenogenital or adreno genital) adj (syndrome or disorder*)).ti,ab,kf.
8.	((haemorrhag* or hemorrhag* or bleed*) adj3 adrenal*).ti,ab,kf.
9.	(Bronze Schilder* Disease or Melanodermic Leukodystrophy or Schilder-Addison* Complex or Siemerling-Creutzfeldt* Disease).ti,ab,kf.
10.	((Allgrove or 3A or TripleA or AAA) adj syndrome).ti,ab,kf.
11.	(CAH or X-ALD).ti,ab.
12.	(Waterhouse-Friderichsen* syndrome or antiphospholipid syndrome).ti,ab,kf.
13.	Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy.ti,ab,kf.
14.	or/1-13
15.	letter.pt. or letter/
16.	note.pt.
17.	editorial.pt.
18.	case report/ or case study/
19.	(letter or comment*).ti.
20.	(conference abstract* or conference review or conference paper or conference proceeding).db,pt,su.
21.	or/15-20
22.	randomized controlled trial/ or random*.ti,ab.
23.	21 not 22
24.	animal/ not human/
25.	nonhuman/
26.	exp Animal Experiment/
27.	exp Experimental Animal/
28.	animal model/
29.	exp Rodent/
30.	(rat or rats or mouse or mice or rodent*).ti.
31.	or/23-30

32.	14 not 31
33.	limit 32 to English language
34.	hydrocortisone/
35.	((cortisol or cortisone) adj3 (assay* or test* or level* or measure* or serum or blood or plasma or saliva* or random or 8am or "8 am" or "8 a.m" or 9am or "9 am" or "9 a.m" or "8-9 am" or "8-9 a.m" or morning or awaken*)).ti,ab,kf.
36.	electrolyte/
37.	electrolyte*.ti,ab,kf.
38.	glucose blood level/
39.	((blood or serum or plasma) adj3 (sugar or glucose)).ti,ab,kf.
40.	(insulin adj4 test*).ti,ab,kf.
41.	tetracosactide/
42.	(Synacthen or tetracosactide).ti,ab,kf.
43.	corticotropin/
44.	((adrenocorticotrop* or "adreno corticotrop*" or corticotrop* or ACTH or cosyntrop*) adj3 (assay* or test* or level* or measure* or stimulation)).ti,ab,kf.
45.	prasterone sulfate/
46.	((DHEA* or "dehydroepiandrosterone sulfate" or "dehydroepiandrosterone sulphate" or "prasterone sulfate" or "prasterone sulphate") adj3 (assay* or test* or level* or measure*)).ti,ab,kf.
47.	or/34-46
48.	33 and 47
49.	exp "sensitivity and specificity"/
50.	(sensitivity or specificity).ti,ab.
51.	((pre test or pretest or post test) adj probability).ti,ab.
52.	(predictive value* or PPV or NPV).ti,ab.
53.	likelihood ratio*.ti,ab.
54.	((area under adj4 curve) or AUC).ti,ab.
55.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
56.	diagnostic accuracy/
57.	diagnostic test accuracy study/
58.	gold standard.ab.
59.	exp diagnostic error/
60.	(false positiv* or false negativ*).ti,ab.
61.	differential diagnosis/
62.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness or precision or validat* or validity or differential or error*)).ti,ab.
63.	or/49-62
64.	Clinical study/
65.	Observational study/
66.	Family study/
67.	Longitudinal study/
68.	Retrospective study/
69.	Prospective study/
70.	Cohort analysis/
71.	Follow-up/
72.	cohort*.ti,ab.
73.	71 and 72

74.	(cohort adj (study or studies or analys* or data)).ti,ab.
75.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
76.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
77.	(before adj2 after adj2 (study or studies or data)).ti,ab.
78.	exp case control study/
79.	case control*.ti,ab.
80.	cross-sectional study/
81.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
82.	or/64-70,73-81
83.	Systematic Review/
84.	Meta-Analysis/
85.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
86.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
87.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
88.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
89.	(search* adj4 literature).ab.
90.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
91.	cochrane.jw.
92.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
93.	or/83-92
94.	48 and (63 or 82 or 93)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Adrenal Insufficiency] explode all trees
#2.	MeSH descriptor: [Adrenal Hyperplasia, Congenital] this term only
#3.	((addison* NEXT disease) or addisonian*):ti,ab,kw
#4.	((adrenal* or adrenocort* or adreno-cort*) near/3 (insufficien* or inadequa* or deficien* or suppress* or hypofunction* or disorder* or underactiv* or dysfunction* or abnormal* or problem* or crisis or crises or dysgenesis or destruction or destroy* or hyperplasia or hypoplasia or failure* or fails or failed or fatigue or inhibit* or damage* or disruption*)):ti,ab,kw
#5.	((cortisol or aldosterone or adrenocorticotrop* or adreno-corticotrop* or ACTH or (corticotropi* NEXT releas*) or (corticotrophi* NEXT releas*) or corticoliberin or CRH) near/3 (insufficien* or inadequa* or deficien* or suppress* or reduc* or decreas* or descend* or diminish* or lack* or less or lessen* or low or lower* or limited)):ti,ab,kw
#6.	(hypoadrenal* or hypo-adrenal* or hypoadrenocorticism or "hypo adrenocorticism" or adrenoleukodystrophy or "adreno leukodystrophy" or adrenomyeloneuropathy or "adreno myeloneuropathy" or hypoaldosteronism or "hypo aldosteronism"):ti,ab,kw
#7.	((adrenogenital or "adreno genital") near/1 (syndrome or disorder*)):ti,ab,kw
#8.	((haemorrhag* or hemorrhag* or bleed*) near/3 adrenal*):ti,ab,kw
#9.	((Bronze NEXT Schilder*) or "Melanodermic Leukodystrophy" or (Schilder NEXT Addison*) or (Siemerling NEXT Creutzfeldt*)):ti,ab,kw
#10.	((Allgrove or 3A or TripleA or AAA) near/1 syndrome):ti,ab,kw
#11.	(CAH or "X-ALD"):ti,ab
#12.	((Waterhouse NEXT Friderichsen*) or "antiphospholipid syndrome"):ti,ab,kw
#13.	"Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy":ti,ab,kw

#14.	(or #1-#13)
#15.	MeSH descriptor: [Hydrocortisone] this term only
#16.	((cortisol or cortisone) near/3 (assay* or test* or level* or measure* or serum or blood or plasma or saliva* or random or 8am or "8 am" or "8 a.m" or 9am or "9 am" or "9 a.m" or "8-9 am" or "8-9 a.m" or morning or awaken*)):ti,ab,kw
#17.	MeSH descriptor: [Electrolytes] this term only
#18.	electrolyte*:ti,ab,kw
#19.	MeSH descriptor: [Blood Glucose] this term only
#20.	((blood or serum or plasma) near/3 (sugar or glucose)):ti,ab,kw
#21.	(insulin near/4 test*):ti,ab,kw
#22.	(Synacthen or tetracosactide):ti,ab,kw
#23.	MeSH descriptor: [Adrenocorticotropic Hormone] explode all trees
#24.	((adrenocorticotrop* or adreno-corticotrop* or corticotrop* or ACTH or cosyntrop*) near/3 (assay* or test* or level* or measure* or stimulation)):ti,ab,kw
#25.	MeSH descriptor: [Dehydroepiandrosterone Sulfate] this term only
#26.	((DHEA* or "dehydroepiandrosterone sulfate" or "dehydroepiandrosterone sulphate" or "prasterone sulfate" or "prasterone sulphate") near/3 (assay* or test* or level* or measure*)):ti,ab,kw
#27.	(or #15-#26)
#28.	#14 and #27
#29.	conference:pt or (clinicaltrials or trialsearch):so
#30.	#28 not #29

Epistemonikos search terms

1.	(title:("adrenal insufficiency" OR "adrenal inadequacy" OR "adrenal deficiency" OR "adrenal suppression" OR "adrenal hypofunction" OR "adrenal disorder" OR "adrenal underactivity" OR "adrenal dysfunction" OR "adrenal crisis" OR "adrenal crises" OR "adrenal hypoplasia" OR "adrenal congenital hyperplasia" OR "addison disease" OR "addisons disease" OR "addison's disease" OR addisonian* OR hypoadrenal* OR "hypo adrenalism" OR hypoadrenocorticism OR "hypo adrenocorticism" OR adrenoleukodystrophy OR "adreno leukodystrophy" OR adrenomyeloneuropathy OR "adreno myeloneuropathy" OR hypoaldosteronism OR "hypo aldosteronism") OR abstract:("adrenal insufficiency" OR "adrenal inadequacy" OR "adrenal deficiency" OR "adrenal suppression" OR "adrenal hypofunction" OR "adrenal disorder" OR "adrenal underactivity" OR "adrenal dysfunction" OR "adrenal crisis" OR "adrenal crises" OR "adrenal hypoplasia" OR "adrenal congenital hyperplasia" OR "addison disease" OR "addisons disease" OR "addison's disease" OR addisonian* OR hypoadrenal* OR "hypo adrenalism" OR hypoadrenocorticism OR "hypo adrenocorticism" OR adrenoleukodystrophy OR "adreno leukodystrophy" OR adrenomyeloneuropathy OR "adreno myeloneuropathy" OR hypoaldosteronism OR "hypo aldosteronism")) AND (title:("cortisol test" OR "serum cortisol" OR "salivary cortisol" OR "salivary cortisone" OR "random cortisol" OR "8am cortisol" OR "8 am cortisol" OR "8 a.m cortisol" OR "9am cortisol" OR "9 am cortisol" OR "9 a.m cortisol" OR "8-9 am cortisol" OR "8-9 a.m cortisol" OR "morning cortisol" OR "awakening cortisol" OR electrolyte* OR "blood sugar" OR "blood glucose" OR "insulin tolerance test" OR "insulin-induced hypoglycemia test" OR Synacthen OR tetracosactide OR "adrenocorticotropic hormone test" OR "adrenocorticotropic hormone stimulation" OR "ACTH test" OR "ACTH stimulation" OR "cosyntropin test" OR "cosyntropin stimulation" OR DHEA* OR "dehydroepiandrosterone sulfate" OR "dehydroepiandrosterone sulphate" OR "prasterone sulfate" OR "prasterone sulphate") OR abstract:("cortisol test" OR "serum cortisol" OR "salivary cortisol" OR "salivary cortisone" OR "random cortisol" OR "8am cortisol" OR "8 am cortisol" OR "8 a.m cortisol" OR "9am cortisol" OR "9 am cortisol" OR "9 a.m cortisol" OR "8-9 am cortisol" OR "8-9 a.m cortisol" OR "morning cortisol" OR "awakening cortisol" OR electrolyte* OR "blood sugar" OR "blood glucose" OR "insulin tolerance test" OR "insulin-induced hypoglycemia test" OR Synacthen OR tetracosactide OR "adrenocorticotropic hormone test" OR "adrenocorticotropic

	hormone stimulation" OR "ACTH test" OR "ACTH stimulation" OR "cosyntropin test" OR "cosyntropin stimulation" OR DHEA* OR "dehydroepiandrosterone sulfate" OR "dehydroepiandrosterone sulphate" OR "prasterone sulfate" OR "prasterone sulphate"))

- Health Economics literature search strategy

Health economic evidence was identified by conducting searches using terms for a broad Adrenal Insufficiency population. The following databases were searched: NHS Economic Evaluation Database (NHS EED - this ceased to be updated after 31st March 2015), Health Technology Assessment database (HTA - this ceased to be updated from 31st March 2018) and The International Network of Agencies for Health Technology Assessment (INAHTA). Searches for recent evidence were run on Medline and Embase from 2014 onwards.

Table 2: Database parameters, filters and limits applied.

Database	Dates searched	Search filters and limits applied
Medline (OVID)	1 January 2014 – 26 September 2023	Health economics studies Exclusions (animal studies, letters, comments, editorials, case studies/reports) English language
Embase (OVID)	1 January 2014 – 26 September 2023	Health economics studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
NHS Economic Evaluation Database (NHS EED) (Centre for Research and Dissemination - CRD)	Inception –31st March 2015	
Health Technology Assessment Database (HTA) (Centre for Research and Dissemination – CRD)	Inception – 31st March 2018	
The International Network of Agencies for Health Technology Assessment (INAHTA)	Inception - 26 September 2023	English language

Medline (Ovid) search terms

1.	exp Adrenal Insufficiency/
2.	Adrenal Hyperplasia, Congenital/
3.	(addison* disease or addisonian*).ti,ab,kf.

4.	((adrenal* or adrenocort* or adreno cort*) adj3 (insufficien* or inadequa* or deficien* or suppress* or hypofunction* or disorder* or underactiv* or dysfunction* or abnormal* or problem* or crisis or crises or dysgenesis or destruction or destroy* or hyperplasia or hypoplasia or failure* or fails or failed or fatigue or inhibit* or damage* or disruption*)).ti,ab,kf.
5.	((cortisol or aldosterone or adrenocorticotrop* or adreno corticotrop* or ACTH or corticotropi* releas* or corticotrophi* releas* or corticoliberin or CRH) adj3 (insufficien* or inadequa* or deficien* or suppress* or reduc* or decreas* or descend* or diminish* or lack* or less or lessen* or low or lower* or limited)).ti,ab,kf.
6.	(hypoadrenal* or hypo adrenal* or hypoadrenocorticism or hypo adrenocorticism or adrenoleukodystrophy or adreno leukodystrophy or adrenomyeloneuropathy or adreno myeloneuropathy or hypoaldosteronism or hypo aldosteronism).ti,ab,kf.
7.	((adrenogenital or adreno genital) adj (syndrome or disorder*)).ti,ab,kf.
8.	((haemorrhag* or hemorrhag* or bleed*) adj3 adrenal*).ti,ab,kf.
9.	(Bronze Schilder* Disease or Melanodermic Leukodystrophy or Schilder-Addison* Complex or Siemerling-Creutzfeldt* Disease).ti,ab,kf.
10.	((Allgrove or 3A or TripleA or AAA) adj syndrome).ti,ab,kf.
11.	(CAH or X-ALD).ti,ab.
12.	(Waterhouse-Friderichsen* syndrome or antiphospholipid syndrome).ti,ab,kf.
13.	Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy.ti,ab,kf.
14.	or/1-13
15.	letter/
16.	editorial/
17.	news/
18.	exp historical article/
19.	Anecdotes as Topic/
20.	comment/
21.	case reports/
22.	(letter or comment*).ti.
23.	or/15-22
24.	randomized controlled trial/ or random*.ti,ab.
25.	23 not 24
26.	animals/ not humans/
27.	exp Animals, Laboratory/
28.	exp Animal Experimentation/
29.	exp Models, Animal/
30.	exp Rodentia/
31.	(rat or rats or mouse or mice or rodent*).ti.
32.	or/25-31
33.	14 not 32
34.	limit 33 to English language
35.	Economics/
36.	Value of life/
37.	exp "Costs and Cost Analysis"/
38.	exp Economics, Hospital/
39.	exp Economics, Medical/

40.	Economics, Nursing/
41.	Economics, Pharmaceutical/
42.	exp "Fees and Charges"/
43.	exp Budgets/
44.	budget*.ti,ab.
45.	cost*.ti.
46.	(economic* or pharmaco?economic*).ti.
47.	(price* or pricing*).ti,ab.
48.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
49.	(financ* or fee or fees).ti,ab.
50.	(value adj2 (money or monetary)).ti,ab.
51.	or/35-50
52.	34 and 51
53.	limit 52 to yr="2014 -Current"

Embase (Ovid) search terms

1.	exp Adrenal cortex insufficiency/
2.	Congenital adrenal hyperplasia/
3.	(addison* disease or addisonian*).ti,ab,kf.
4.	((adrenal* or adrenocort* or adreno cort*) adj3 (insufficien* or inadequa* or deficien* or suppress* or hypofunction* or disorder* or underactiv* or dysfunction* or abnormal* or problem* or crisis or crises or dysgenesis or destruction or destroy* or hyperplasia or hypoplasia or failure* or fails or failed or fatigue or inhibit* or damage* or disruption*)).ti,ab,kf.
5.	((cortisol or aldosterone or adrenocorticotrop* or adreno corticotrop* or ACTH or corticotropi* releas* or corticotrophi* releas* or corticoliberin or CRH) adj3 (insufficien* or inadequa* or deficien* or suppress* or reduc* or decreas* or descend* or diminish* or lack* or less or lessen* or low or lower* or limited)).ti,ab,kf.
6.	(hypoadrenal* or hypo adrenal* or hypoadrenocorticism or hypo adrenocorticism or adrenoleukodystrophy or adreno leukodystrophy or adrenomyeloneuropathy or adreno myeloneuropathy or hypoaldosteronism or hypo aldosteronism).ti,ab,kf.
7.	((adrenogenital or adreno genital) adj (syndrome or disorder*)).ti,ab,kf.
8.	((haemorrhag* or hemorrhag* or bleed*) adj3 adrenal*).ti,ab,kf.
9.	(Bronze Schilder* Disease or Melanodermic Leukodystrophy or Schilder-Addison* Complex or Siemerling-Creutzfeldt* Disease).ti,ab,kf.
10.	((Allgrove or 3A or TripleA or AAA) adj syndrome).ti,ab,kf.
11.	(CAH or X-ALD).ti,ab.
12.	(Waterhouse-Friderichsen* syndrome or antiphospholipid syndrome).ti,ab,kf.
13.	Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy.ti,ab,kf.
14.	or/1-13
15.	letter.pt. or letter/
16.	note.pt.
17.	editorial.pt.
18.	case report/ or case study/
19.	(letter or comment*).ti.
20.	(conference abstract* or conference review or conference paper or conference proceeding).db,pt,su.

21.	or/15-20
22.	randomized controlled trial/ or random*.ti,ab.
23.	21 not 22
24.	animal/ not human/
25.	nonhuman/
26.	exp Animal Experiment/
27.	exp Experimental Animal/
28.	animal model/
29.	exp Rodent/
30.	(rat or rats or mouse or mice or rodent*).ti.
31.	or/23-30
32.	14 not 31
33.	limit 32 to English language
34.	health economics/
35.	exp economic evaluation/
36.	exp health care cost/
37.	exp fee/
38.	budget/
39.	funding/
40.	budget*.ti,ab.
41.	cost*.ti.
42.	(economic* or pharmaco?economic*).ti.
43.	(price* or pricing*).ti,ab.
44.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
45.	(financ* or fee or fees).ti,ab.
46.	(value adj2 (money or monetary)).ti,ab.
47.	or/34-46
48.	33 and 47
49.	limit 48 to yr="2014 -Current"

NHS EED and HTA (CRD) search terms

#1.	MeSH DESCRIPTOR Adrenal Insufficiency EXPLODE ALL TREES
#2.	MeSH DESCRIPTOR Adrenal Hyperplasia, Congenital EXPLODE ALL TREES
#3.	(addison* disease or addisonian)
#4.	(adrenal*) AND (insufficien* or inadequa* or deficien* or hypofunction* or disorder* or underactiv* or dysfunction* or abnormal* or problem* or crisis or crises or dysgenesis or destruction or destroy* or hyperplasia or hypoplasia or failure* or fails or failed)
#5.	(cortisol or aldosterone or adrenocortical or adrenocorticotropi* or ACTH or corticotropi* releas* or corticotrophi* releas* or corticoliberin or CRH) AND (insufficien* or inadequac* or deficien* or reduc* or decreas* or descend* or diminish* or lack* or less or lessen* or low or lower* or produc* or limited)
#6.	(hypoadrenalism or hypoadrenocorticism or adrenoleukodystrophy or adrenomyeloneuropathy or hypoaldosteronism)
#7.	((Bronze Schilder* Disease or Melanodermic Leukodystrophy or Schilder-Addison* Complex or Siemerling-Creutzfeldt* Disease))
#8.	(Allgrove or 3A or TripleA or AAA) AND (syndrome)
#9.	(X-ALD)
#10.	((Waterhouse-Friderichsen* syndrome or antiphospholipid syndrome))

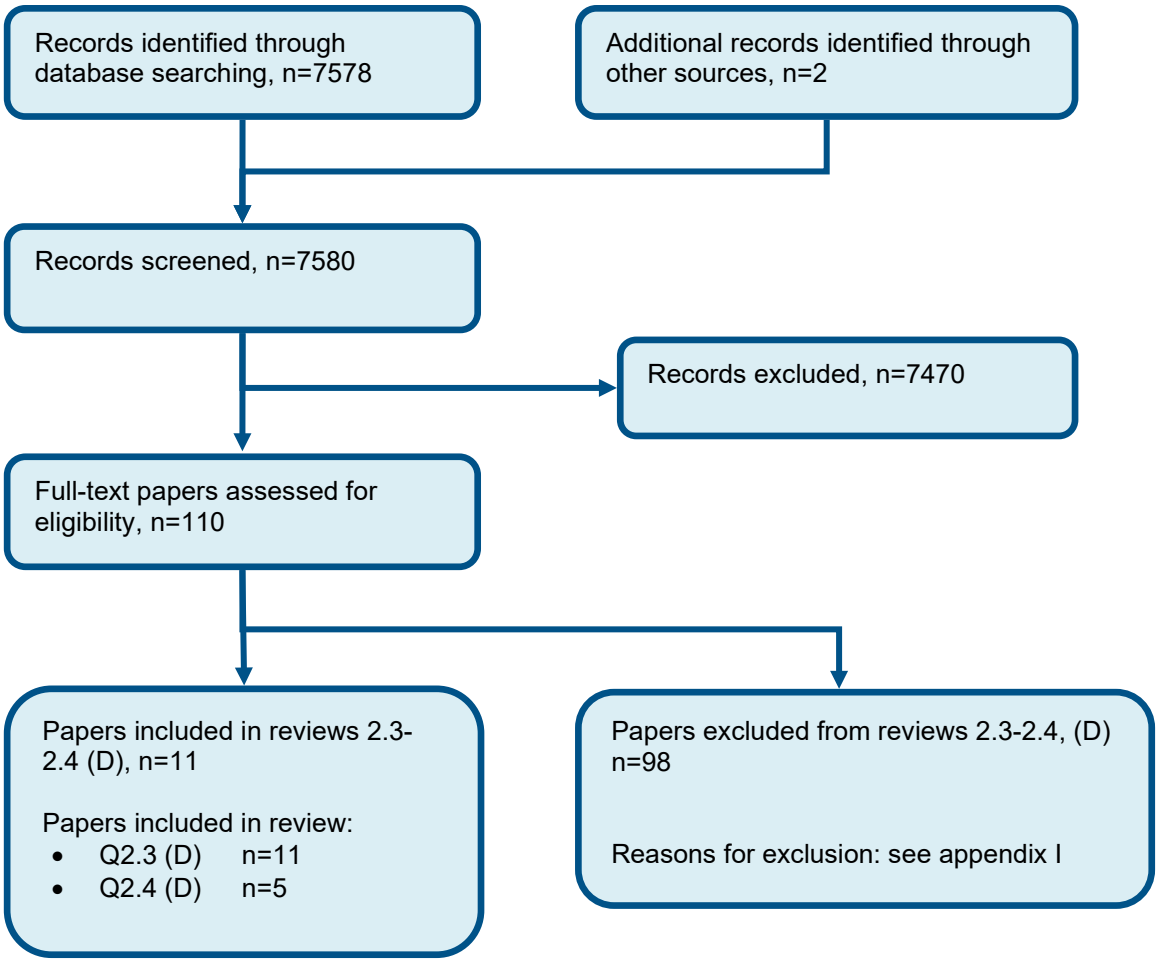
#11.	((Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy))
#12.	(adrenogenital or adreno genital) AND (syndrome)
#13.	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12

INAHTA search terms

1.	(("Adrenal Insufficiency"[mhe]) OR (hypoadrenalism) OR (addison*) OR (adrenal insufficiency) OR (adrenal crisis))

Appendix C Diagnostic evidence study selection

Figure 1: Flow chart of clinical study selection for the review of diagnostic tests and thresholds for referral



Appendix D Diagnostic evidence

D.1 Diagnostic Tests

Reference	**Agwu 1999**[1]
Study type	Cross sectional diagnostic accuracy study
Study methodology	**Data source**: Hospital assessments (unclear if inpatients or outpatients) **Recruitment**: Children presenting with symptoms suggestive of AI were admitted (recruitment period not specified).
Number of patients	n = 32
Patient characteristics	**Age** (range): 2-19 years **Gender** (male to female ratio): 14:18 **Ethnicity**: not reported **Setting**: In hospital **Country**: UK **Inclusion criteria**: symptoms suggestive of AI **Exclusion criteria**: not stated. **Underlying diagnoses**: 14 had been irradiated for brain tumours; 2 had total body irradiation for relapsed non-Hodgkin's lymphoma and acute lymphoblastic leukaemia; 13 had other endocrinopathies; 2 had growth failure; and 1 had histiocytosis with prolonged dexamethasone treatment.
Target condition(s)	Adrenal insufficiency

Reference	**Agwu 1999[1]**
Index test(s) and reference standard	**Index test** 08:00-09:00 hour serum cortisol concentration (measured via indwelling IV catheter before the patient got out of bed) **Threshold:** A normal cortisol level was defined as >500 nmol/l. **Reference standard** Low dose Synacthen test (LDST) was performed at 14:00 after a 2-hour rest after lunch. A bolus injection of 500 ng/1.73 m^2 of ACTH was given intravenously and sampling took place at 0, 10, 15, 20, 25, 30, 35, 40, and 45 minutes. The next day, a standard short Synacthen test (SSST) was performed at 09:00: 250 µg/1.73 m^2 of Synacthen was given as an intravenous bolus with sampling at 0, 30, and 60 minutes. The 30-minute peak was studied. **Thresholds:** A normal response to the Synacthen test was defined as a peak serum cortisol of ≥500 nmol/Land/or incremental concentration of ≥200 nmol/l. **Assays**: Cortisol was measured using a direct coated tube assay (Euro DPC Ltd, Llanberies, Gwynedd, UK). This assay has a lower limit of detection of 6 nmol/l. The within assay coefficients of variation were 5.7% and 2.6% at serum concentrations of 28 nmol/Land 552 nmol/l, respectively. The between assay coefficients of variation were 9.1% and 6.8% at serum concentrations of 95 nmol/Land 459 nmol/l, respectively. **Time:** between measurement of index test and reference standard: 6 hours for LDST and 24 hours for SSST
2×2 table	

	Reference standard + (LDST increment <200nmol/l)	**Reference standard −** (LDST increment <200nmol/l)	**Total**
Index test + (0800 serum cortisol <500 nmol/l)	11	14	25
Index test – (0800 serum cortisol <500 nmol/l)	0	7	7
Total	11	21	32

Statistical measures	**Index text (0800 serum cortisol – threshold 500 nmol/l)** Sensitivity: 100% Specificity: 33% PPV: 0.44 NPV: 1.00 PLR: 1.50 NLR: 0.00

Reference	**Agwu 1999[1]**
	AUC: NA
Source of funding	Not stated
Limitations	Risk of bias: Unclear methods of recruitment Indirectness: None
Comments	

Reference	**Choi 2002[3]**
Study type	Cross sectional/prospective diagnostic accuracy study
Study methodology	**Data source**: Hospital assessments **Recruitment**: patients with clinically suspected secondary adrenocortical insufficiency recruited from March 1997 to December 2000.
Number of patients	n = 72
Patient characteristics	**Age:** mean – 46 years; range (28-74 years) **Gender** (male to female ratio): 30:42 **Ethnicity**: Chinese **Setting**: Regional hospital (unclear if in- or out-patients) **Country**: Hong Kong **Inclusion criteria**: clinically suspected secondary adrenocortical insufficiency **Exclusion criteria**: contraindications for ITT such as epilepsy or ischaemic heart disease

Reference	Choi 2002[3]
	Underlying diagnoses: reasons for suspicion of dysfunction of the hypothalamic-pituitary-adrenocortical axis were nasopharyngeal carcinoma with radiotherapy to the pituitary region (n=22), iatrogenic Cushing's syndrome (n=20), non-functioning pituitary macroadenoma (n=14), empty sella syndrome (n=8), acromegaly (n=5), Sheehan's syndrome (n=1), Cooley's anaemia with secondary haemochromatosis (n=1), and prolactin-secreting pituitary macroadenoma (n=1). **Prior tests/treatment**: Patients receiving chronic steroid therapy for other medical illnesses were advised to cease steroid therapy 1 week before the tests.
Target condition(s)	Secondary adrenal insufficiency
Index test(s) and reference standard	**Index test** Morning fasting serum cortisol concentration (0900) **Threshold:** to maximise sensitivity or specificity **Reference standard** (insulin tolerance test) Performed by a specialty nurse supervised by a doctor after overnight fast. Actrapid insulin (NovoNordisk, Bagsvaerd, Denmark) 0.1 U/kg was given intravenously after a baseline blood sample was taken for glucose and cortisol assay. Diabetic patients receiving insulin treatment were given actrapid insulin 0.15 U/kg plus 50% of their morning dose of short-acting insulin. Bedside haemoglucostix (Surestep Plus; Lifescan, Milpitas, US) testing was performed at 15-minute intervals and blood was sampled for laboratory glucose assay. When hypoglycaemia developed (bedside haemoglucostix result of less than 2.2 mmol/L and occurrence of hypoglycaemic symptoms), a blood sample was taken for glucose and cortisol assay, and 50% dextrose solution 40 mL was given to the patient intravenously, followed by oral food. If hypoglycaemia had not occurred by 45 minutes, an additional dose of actrapid insulin 0.1 U/kg was given. Blood sampling for cortisol measurement was done at intervals of 15 minutes during the first 60 minutes, and thereafter at 30-minute intervals. **Threshold:** peak cortisol response of ≥550 nmol/L for adrenal sufficiency **Assays**: Cortisol was assayed by the chemiluminescence method (Bayer-Centaur, New York, US). The coefficient of variation for the assay was less than 5%. **Time** between measurement of index test and reference standard: 4-7 days
Statistical measures	**Index text (fasting morning cortisol)** Fasting morning cortisol range: 51 to 537 nmol/L. Threshold: ≥420 nmol/L – sensitivity 100% (no false negatives)

Reference	**Choi 2002[3]**
	Threshold: ≤112 nmol/L – specificity 100% (no false positives)
Source of funding	Not stated
Limitations	**Risk of bias:** Unclear methods of recruitment; time between index and reference tests 4-7 days **Indirectness**: None identified
Comments	

Reference	**Debono 2023[5]**
Study type	Cross sectional/prospective diagnostic accuracy study
Study methodology	**Data source**: Hospital assessments **Recruitment**: Patients were recruited by consecutive sampling at Sheffield Teaching Hospitals NHS Foundation Trust between November 2019 and December 2021.
Number of patients	n = 220 (208 available for primary outcome measure)
Patient characteristics	**Age:** 55.1–15.8 years **Gender** (male to female ratio): 106:102 **Ethnicity**: White: 90%, Asian: 5%, and the remaining patients were Black/Caribbean/African and multiracial. **Setting**: NHS hospital for ACTH and home-based salivary cortisone test **Country**: UK **Inclusion criteria**: Patients at high risk for adrenal insufficiency. All patients referred for an ACTH stimulation test to assess for a new diagnosis of adrenal insufficiency or for recovery from a previous diagnosis of adrenal insufficiency were considered for the study. Patients older than age 18 years with a high probability of either primary, secondary, or tertiary adrenal insufficiency as determined by the investigators were eligible for enrolment. **Exclusion criteria**: Patients were excluded who were unable to produce a suitable saliva sample; night shift workers; patients with known protein-losing disorders, known or suspected alcohol dependence, and known severe liver disease; patients with uncontrolled active infection; and patients taking oestrogens or those who were pregnant. Patients taking drugs that influence the hypothalamic-

Reference	Debono 2023[5]
	pituitary-adrenal axis (e.g., opioids) had their medications omitted on the day of testing, per routine clinical practice. In view of the coronavirus disease 2019 measures resulting in limited staff and fewer appointment slots to enable study tasks, some patients were excluded from study participation so as not to hinder their clinical care. **Reason for referral**: Patients who were dependent on any type of glucocorticoid, who had been on an oral glucocorticoid prednisolone-equivalent dose of 5 mg/d for 4 weeks, and who were referred for adrenal testing only after they had been weaned down to prednisolone 5 mg/d or equivalent or converted to physiologic doses of hydrocortisone 25 mg/d. Patients receiving any intermediate- or long-acting intramuscular or intra-articular glucocorticoid injections were recruited at least 3 months after their last injection. Patients with pituitary disease, such as tumours, inflammatory disease, or those with a history of cranial radiotherapy, were considered eligible for inclusion.
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index test: waking salivary cortisone, cortisol and baseline serum – (threshold: different cut offs reported)** Salivary sample upon waking using Salivette tubes containing synthetic swabs (Salivette Cortisol; Sarstedt). A total of 500 ml of saliva was necessary to ensure a good representative sample, and 50 ml was used for the assay. All patients were given written instructions with images and a video on how to collect their salivary sample. Patients were advised to refrain from smoking/vaping on the day of the test. Patients taking glucocorticoids were asked to omit these medicines the evening before and the day of the test until all samples were collected. Patients were allowed to follow their usual waking routines but were asked to collect their waking salivary sample the moment they got out of bed to commence the day and before cleaning teeth, eating, or drinking. **Thresholds:** salivary cortisone = 7 and 17nmol/l, salivary cortisol = 5 and 1 nmol/L, serum cortisol = 152 and 310 nmol/l **Reference standard** Standard dose ACTH stimulation test At the endocrine clinic an intravenous cannula was inserted, and baseline serum cortisol was measured. An ACTH stimulation test was performed with intravenous injection of 250 mg of Synacthen (Atnahs Pharma UK Limited), followed by a serum cortisol level blood draw at 30 minutes. Tests were performed by specialized endocrine nurses at the clinic. On completing the ACTH stimulation test, patients completed the final part of the questionnaire assessing their views on the ACTH stimulation test and salivary sample collection. Results of the ACTH stimulation test were interpreted by a consulting endocrinologist, or a specialized endocrine nurse who was unaware of the waking cortisol data. An a priori criterion of a peak cortisol level of 15.6 mg/dl (430 nmol/l) measured by immunoassay indicated adequate adrenal reserve, whereas those patients with levels less than this value were considered to have adrenal insufficiency according to current clinical practice in our centre. **Threshold:** peak cortisol level of ≥430 nmol/Lfor adrenal sufficiency **Assays**: Serum cortisol was analysed by immunoassay (Elecsys Cortisol II assay; Roche) and interpreted immediately at Sheffield Teaching Hospitals NHS Foundation Trust. An extra serum cortisol sample was stored at -80C and, together with the salivary sample,

<table>
<tr><td>Reference</td><td>Debono 2023[5]</td></tr>
<tr><td></td><td>was then analysed and interpreted by liquid chromatography–tandem mass spectrometry (LC-MS/MS) as a batch at the end of the study in a different laboratory in Manchester University NHS Foundation Trust.

<u>Time:</u> later in the same day (median value = approx. 3 hours later as self-reported by patients)</td></tr>
<tr><td>2×2 table (calculated from reported sensitivity and negative predictive values)</td><td>
<table>
<tr><td colspan="3"><u>Index text (waking salivary cortisone – threshold 17 nmol/l)</u></td><td></td></tr>
<tr><td></td><td>Reference standard +
(ACTH stimulation test 430nmol/l)</td><td>Reference standard −
(ACTH stimulation test 430nmol/l)</td><td>Total</td></tr>
<tr><td>Index test +
(waking salivary cortisone – threshold 17 nmol/l)</td><td>88</td><td>51</td><td>140</td></tr>
<tr><td>Index test –
(waking salivary cortisone – threshold 17 nmol/l)</td><td>3</td><td>66</td><td>68</td></tr>
<tr><td>Total</td><td>91</td><td>117</td><td>208</td></tr>
<tr><td colspan="4">Index text (waking salivary cortisone – threshold 7 nmol/l)</td></tr>
<tr><td></td><td>Reference standard +
(ACTH stimulation test 430nmol/l)</td><td>Reference standard −
(ACTH stimulation test 430nmol/l)</td><td>Total</td></tr>
<tr><td>Index test +
(waking salivary cortisone – threshold 7 nmol/l)</td><td>67</td><td>4</td><td>70</td></tr>
<tr><td>Index test – (waking salivary cortisone – threshold 7 nmol/l)</td><td>24</td><td>113</td><td>138</td></tr>
<tr><td>Total</td><td>91</td><td>117</td><td>208</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Index text (waking salivary cortisol – threshold 5 nmol/l)</td></tr>
<tr><td></td><td>Reference standard +
(ACTH stimulation test 430nmol/l)</td><td>Reference standard −
(ACTH stimulation test 430nmol/l)</td><td>Total</td></tr>
<tr><td>Index test +
(waking salivary cortisol – threshold 5 nmol/l)</td><td>86</td><td>46</td><td>132</td></tr>
<tr><td>Index test – (waking salivary cortisol – threshold 5 nmol/l)</td><td>5</td><td>71</td><td>76</td></tr>
<tr><td>Total</td><td>91</td><td>117</td><td>208</td></tr>
<tr><td colspan="4"></td></tr>
</table>
</td></tr>
</table>

Reference	Debono 2023[5]

Index text (waking salivary cortisol – threshold 1 nmol/l)

	Reference standard + (ACTH stimulation test 430nmol/l)	Reference standard − (ACTH stimulation test 430nmol/l)	Total
Index test + (waking salivary cortisol – threshold 1 nmol/l)	47	4	50
Index test – (waking salivary cortisol – threshold 1 nmol/l)	44	113	158
Total	91	117	208

Index text (Baseline serum cortisol – threshold 310 nmol/l)

	Reference standard + (ACTH stimulation test 430nmol/l)	Reference standard − (ACTH stimulation test 430nmol/l)	Total
Index test + (waking salivary cortisol – threshold 1 nmol/l)	87	69	156
Index test – (waking salivary cortisol – threshold 1 nmol/l)	4	48	52
Total	91	117	208

Index text (Baseline serum cortisol – threshold 152 nmol/l)

	Reference standard + (ACTH stimulation test 430nmol/l)	Reference standard − (ACTH stimulation test 430nmol/l)	Total
Index test + (waking salivary cortisol – threshold 1 nmol/l)	59	6	65
Index test – (waking salivary cortisol – threshold 1 nmol/l)	32	111	143
Total	91	117	208

Statistical measures (some data calculated from reported	**Index text (waking salivary cortisone – threshold 17 nmol/l)** Sensitivity: 97% (95% CI, 91 to 99) reported in study Specificity: 56% PPV: 63% NPV: 96% (95% CI, 90 to 99) reported in study

Reference	Debono 2023[5]
sensitivity and negative predictive values)	PLR: 2.20 NLR: 0.05 AUC: 0.95 (95% CI, 0.92 to 0.97) reported in study **Index text (waking salivary cortisone – threshold 7 nmol/l)** Sensitivity: 73% Specificity: 97% (95% CI, 92 to 99) reported in study PPV: 95% (95% CI, 87 to 99) NPV: 0.83 PLR: 24.4 NLR: 0.28 Maximum thresholds - To achieve at least 99% sensitivity to exclude adrenal insufficiency and 99% specificity to confirm adrenal insufficiency, one would need to use waking salivary cortisone cutoffs of 25 nmol/Land <1 nmol/l, respectively. **Index text (waking salivary cortisol – threshold 5 nmol/l)** Sensitivity: 95% (95% CI, 88 to 99) reported in study Specificity: 61% PPV: 65% NPV: 94% (95% CI, 85 to 98) reported in study. PLR: 2.40 NLR: 0.09 AUC: 0.89 (95% CI, 0.85 to 0.94) reported in study. **Index text (waking salivary cortisol – threshold 1 nmol/l)** Sensitivity: 52% Specificity: 97% (95% CI, 92 to 100) reported in study PPV: 93% (95% CI, 80 to 99) reported in study. NPV: 72% PLR: 15.1 NLR: 0.50 **Index text (baseline serum cortisol – threshold 310 nmol/l)** Sensitivity: 96% (95% CI, 90 to 99) reported in study

Reference	**Debono 2023[5]**
	Specificity: 41% PPV: 56% NPV: 93% (95% CI, 84 to 98) reported in study. PLR: 1.64 NLR: 0.09 AUC: 0.90 (95% CI, 0.86 to 0.94) reported in study. **Index text (baseline serum cortisol – threshold 152 nmol/l)** Sensitivity: 65% Specificity: 95% (95% CI, 90 to 98) reported in study PPV: 91% (95% CI, 81 to 97) reported in study. NPV: 77% PLR: 13.0 NLR: 0.37 **Additional analyses:** Examined the percentage of ACTH stimulation tests that could have been avoided using waking salivary cortisone cutoffs as a screening test in a two-stage process. The ACTH stimulation test would have been avoided in 70% (154 of 220) of participants — 73 patients with waking salivary cortisone.
Source of funding	Funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit Programme (grant reference number, PBPG-1217-20007). Dr. Elder and Mr. Keevil are both supported by funding from UK Research and Innovation–Innovation Scholars secondments: Biomedical Sciences. The study was also supported by the NIHR Clinical Research Network Yorkshire and Humber IRAS 262618.
Limitations	**Risk of bias:** None identified. **Indirectness**: None identified.
Comments	

Reference	**De Lange 1993[4]**
Study type	Cross sectional diagnostic accuracy study
Study methodology	**Data source**: Hospital assessments **Recruitment**: consecutive patients who had received adequately performed insulin-induced hypoglycaemia and metyrapone tests

Reference	De Lange 1993[4]
Number of patients	n = 58
Patient characteristics	**Age:** range 17-73 years **Gender** (male to female ratio): 33:25 **Ethnicity**: not stated **Setting**: Regional hospital (unclear if in- or out-patients) **Country**: Netherlands **Inclusion criteria**: received adequately performed insulin-induced hypoglycaemia and metyrapone tests **Exclusion criteria**: clinical evidence of adrenal insufficiency or Cushing's syndrome **Underlying diagnoses**: 45 had pituitary adenoma (30 before treatment; 10 after hypophysectomy and 5 after X-ray therapy). 13 had proven or suspected HPA disease. **Prior tests/treatment**: Corticosteroid medication was stopped at least 72 hours before investigation.
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index test** Morning fasting plasma cortisol concentration (0900-1000) **Thresholds:** 160 and 260 nmol/L **Reference standard** (insulin tolerance test) Performed after overnight fast. Soluble insulin 0.15 U/kg was given intravenously. Blood sampling for glucose and cortisol measurement was taken before and at 0, 30, 45, 60 and 90 minutes after injection. **Threshold:** rise in plasma cortisol above 500 nmol/L for adrenal sufficiency (normal – n=33/58) **Assays**: Plasma cortisol was measured by RIA. Blood glucose was measured by a hexokinase method.

Reference	**De Lange 1993[4]**
	Time between measurement of index test and reference standard: 0-1 days (2 samples taken for index test, one on the day of ITT and one the following day)
Statistical measures	**Index text (fasting morning cortisol)** Threshold: <260 nmol/L – sensitivity = 96%; specificity = 64% Threshold: <160 nmol/L – sensitivity = 64%; specificity = 91%
Source of funding	Not stated
Limitations	**Risk of bias:** none identified. **Indirectness**: Basal cortisol was mean of morning measurements on 2 consecutive days
Comments	

Reference	**Deutschbein 2009[7]**
Study type	Cross sectional diagnostic accuracy study
Study methodology	Data source: Hospital assessments Recruitment: patients were investigated because of suspected or proven disease of the HPA axis (dates of recruitment not reported)
Number of patients	n = 77
Patient characteristics	Age, mean (SD): 44.2 (1.8) years. Gender (male to female ratio): 41:36 Ethnicity: Not reported Setting: hospital (unclear if inpatient or outpatient) Country: Germany Inclusion criteria: suspected or proven HPA axis disease

Reference	**Deutschbein 2009[7]**
	Exclusion criteria: No subject had to be excluded because of contraindications to insulin-induced hypoglycaemia. **Underlying diagnoses** Reason for referral: 65 patients had sellar masses, of whom 11 were not operated; all 54 surgically treated patients were tested at least 3 months after surgery (median postoperative interval: 6.3 months; range: 3 – 68 months). The remaining 12 patients showed an impaired secretion of various hormones but did not have any detectable tumours within the HPA. **Prior tests/treatment**: At the time of hormonal evaluation, female patients were neither on contraceptives nor oestrogens. Patients on chronic corticosteroid replacement therapy (generally 10–15 mg hydrocortisone per day) received their last dosage at 1400 h the day before testing, resulting in a drug restriction period of at least 18 h.
Target condition(s)	Secondary adrenal insufficiency
Index test(s) and reference standard	**Index tests: Basal serum and salivary cortisol** Basal salivary cortisol: saliva was collected by chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h and before blood withdrawal. Basal serum cortisol: venipuncture. **Threshold cut-off value:** Basal salivary cortisol - an optimal cut-off of 15.1 nmol/l, an upper cut-off of 21.1 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. Basal serum cortisol - ROC analysis revealed an optimal cut-off of 283 nmol/l, an upper cut-off of 470 nmol/l, and a lower cut-off of 103 nmol/Lrespectively. **Assay:** Serum cortisol was measured by a competitive immunoassay (Advia Centaur, Bayer, Germany). The lower detection limit of this assay was 5.5 nmol/l, and intra- and inter-assay coefficients of variation were less than 3.8 and 5.5%, respectively. Salivary cortisol was determined using a modification of a commercial radioimmunoassay (RIA) (GammaCoat, DiaSorin, USA), decreasing the sample volume from 200 to 100 µ l. The intra- and inter-assay coefficients of variation were 5.4 and 15.9 %, respectively. **Reference standard: Insulin tolerance test (ITT)** The ITT was performed between 0800 and 1000, using a peak cortisol cut-off point of 500 nmol/Lto distinguish between adrenal insufficient (AI) and adrenal sufficient (AS) patients. After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycaemia were required as evidence of sufficient stress. Samples for blood glucose and Se C were taken at 0, 15, 30, 45, 60, 90, and 120 min. **Threshold:** Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L(18 µg/dl).

<table>
<tr><td>Reference</td><td colspan="4">Deutschbein 2009[7]</td></tr>
<tr><td></td><td colspan="4">Time between measurement of index test and reference standard: ITT was performed wither immediately after basal samples or on a separate day. Average time lag not reported.</td></tr>
<tr><td rowspan="9">2×2 table (calculated from reported sensitivity, specificity and true positive and true negative rates)</td><td colspan="4">Morning salivary cortisol (threshold 15.1 nmol/l)</td></tr>
<tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr>
<tr><td>Index test +</td><td>35</td><td>16</td><td>51</td></tr>
<tr><td>Index test −</td><td>6</td><td>20</td><td>26</td></tr>
<tr><td>Total</td><td>41</td><td>36</td><td>77</td></tr>
<tr><td colspan="4">Morning serum cortisol (threshold 283 nmol/l)</td></tr>
<tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr>
<tr><td>Index test +</td><td>30</td><td>11</td><td>41</td></tr>
<tr><td>Index test −</td><td>11</td><td>25</td><td>36</td></tr>
<tr><td>Total</td><td>41</td><td>36</td><td>77</td></tr>
<tr><td>Statistical measures</td><td colspan="4"><u>Index text –</u> basal salivary cortisol cut off 15.1 nmol/l
Sensitivity 85%
Specificity 56%
AUC 0.74

An optimal cutoff of 15.1 nmol/l, an upper cutoff of 21.1 nmol / l, and a lower cutoff of 5.0 nmol / l were found.
By using the upper as well as the lower cutoff, a diagnosis corresponding to the ITT result was established in 21 of 77 patients (27 %), whereas 56 patients (73 %) required additional testing procedures.

<u>Index text –</u> basal serum cortisol (cut off 283 nmol/l)
Sensitivity 73%
Specificity 69%
AUC 0.75

An optimal cutoff of 283 nmol/l, an upper cutoff of 470 nmol / l, and a lower cutoff of 103 nmol / l were found.
283
By applying both the upper and the lower cutoff, a diagnosis corresponding to the ITT result was established in 18 of 77 patients (23 %), leaving 59 patients (77 %) for further evaluation.</td></tr>
<tr><td>Source of funding</td><td colspan="4">Not stated</td></tr>
<tr><td>Limitations</td><td colspan="4">Risk of bias: very serious (due to patient selection and patient flow)</td></tr>
</table>

Reference	**Deutschbein 2009[7]**
	Indirectness: none
Comments	Unclear whether the patient sample overlaps with Deutschbein 2009a

Reference	**Deutschbein 2009a[6]**
Study type	Prospective diagnostic accuracy study
Study methodology	Data source: Hospital assessments (unclear if inpatients or outpatients) Recruitment: Between 2005 and 2007 fifty-five patients were investigated because of suspected or proven disease of the HPA axis
Number of patients	n = 55
Patient characteristics	Age, mean (SD): 45.9 (2.1) years. Gender (male to female ratio): 26:29 Ethnicity: Not reported Setting: outpatients Country: Germany Inclusion criteria: patients with suspected secondary adrenal insufficiency Exclusion criteria: At the time of hormonal evaluation, female patients were neither on contraceptives nor oestrogens. Patients on chronic corticosteroid replacement therapy (generally 10–15 mg hydrocortisone per day) received their last dosage at 1400 h the day before testing, resulting in a drug restriction period of at least 18 h. No subject had to be excluded because of contraindications to insulin-induced hypoglycemia. Reason for referral: Eight subjects had a present neoplasia (4 prolactinomas, 2 somatotropic adenomas, 1 non-functioning adenoma, 1 meningioma), and 40 subjects (24 non-functioning adenomas, 8 somatotropic adenomas, 3 craniopharyngiomas, 3 prolactinomas, 2 meningiomas) were tested at least 3 months after surgical treatment (median interval: 4.8 months). Seven patients suffered from pituitary hormone impairment without detectable sellar tumors (3 secondary hypogonadisms, 3 congenital hormone insufficiencies, 1 Sheehan syndrome).

Reference	Deutschbein 2009a[6]
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index tests: Basal serum and salivary cortisol** Basal salivary cortisol: saliva was collected by chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h. Basal serum cortisol: after catheterisation of a superficial cubital vein and a recovery period of 15 min to avoid stress-induced bias, blood samples were directly obtained into serum tubes (Monovetten, Sarstedt, Germany). During High dose SST, serum and saliva samples were taken at 0, 30, 60, 90, and 120 min after i.v. application of 250 mg synthetic ACTH (Synacthen, Novartis). **Threshold cut-off value:** Basal salivary cortisol - an optimal cut-off of 7.6 nmol/l, an upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. Basal serum cortisol - ROC analysis revealed an optimal cut-off of 260 nmol/l, an upper cut-off of 382 nmol/l, and a lower cut-off of 103 nmol/Lrespectively. **Assay:** Saliva samples were measured by a modification of the 'GammaCoat' RIA for cortisol (DiaSorin, Stillwater, MN, USA), decreasing the sample volume from 200 to 100 ml. The intra-assay and inter-assay coefficients of variation were 5.4% and 15.9%, respectively. **Reference standard: Insulin tolerance test (ITT)** The ITT served as reference test, using a peak cortisol cut-off point of 500 nmol/Lto distinguish between adrenal insufficient (AI) and adrenal sufficient (AS) patients. After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycemia were required as evidence of sufficient stress. Samples for blood glucose and Se C were taken at 0, 15, 30, 45, 60, 90, and 120 min. **Threshold:** Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L(18 µg/dl). **Assay**: Serum and salivary cortisol were measured by electrochemiluminescence method (ECLIA) using Cobas e 411 analysers with commercially available Elecsys Cortisol II (second generation, monoclonal antibody) kits which have showed a good correlation with gas chromatography mass spectrometry (GC-MS). Serum cortisol levels were determined by competitive immunoassay, using commercial kits (Advia Centaur, Bayer). The analytical sensitivity of the assay was 5.5 nmol/l. Intra-assay variations as coefficient of variation for various cortisol values were 3.7% (107.1 nmol/l), 3.1% (155.3 nmol/l), 2.9% (391.0 nmol/l), 3.8% (759.6 nmol/l), and 3.0% (1025.0 nmol/l). Inter assay variations for the cortisol concentrations mentioned above were 5.5, 3.8, 3.1, 1.9, and 4.0%.

<table>
<tr><td>Reference</td><td colspan="4">Deutschbein 2009a[6]</td></tr>
<tr><td></td><td colspan="4">Time between measurement of index test and reference standard: The minimum and maximum intervals between both tests were 1 day and 17 days, respectively, with a median interval of 2 days (unclear if this was the basal values)</td></tr>
<tr><td rowspan="9">2×2 table (calculated from reported sensitivity, specificity and true positive and true negative rates)</td><td colspan="4">Morning salivary cortisol (threshold 7.6 nmol/l)</td></tr>
<tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr>
<tr><td>Index test +</td><td>16</td><td>4</td><td>20</td></tr>
<tr><td>Index test −</td><td>14</td><td>21</td><td>35</td></tr>
<tr><td>Total</td><td>30</td><td>25</td><td>55</td></tr>
<tr><td colspan="4">Morning serum cortisol (threshold 260 nmol/l)</td></tr>
<tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr>
<tr><td>Index test +</td><td>22</td><td>7</td><td>29</td></tr>
<tr><td>Index test −</td><td>8</td><td>18</td><td>26</td></tr>
<tr><td>Total</td><td>30</td><td>25</td><td>55</td></tr>
<tr><td>Statistical measures</td><td colspan="4">Index text – basal salivary cortisol cut off 7.6 nmol/l
Sensitivity 53
Specificity 83
AUC 0.74

An upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. The use of both the upper and lower thresholds allowed a diagnosis in 19 of 55 patients (35%). If the remaining 36 subjects were then investigated by the high-dose short Synacthen test (HDT), the combination of upper and lower cut-offs diagnosed 3 (HDT periods 0–60and 0–120 min), 4 (HDT period 0–30 min), and 5 additional patients (HDT period 0–90 min). So, the combination of basal cortisol and HDT allowed a diagnosis in 42–45% of patients.

Index text – basal serum cortisol (cut off 260 nmol/l)
Sensitivity 73
Specificity 72
AUC 0.78
An upper cut-off of 382 nmol/l, and a lower cut-off of 103 nmol/Lwere identified. By applying the upper as well as the lower cut-off, 18 of 55 patients (33%) were diagnosed by their basal SeC levels, leaving 37 subjects for
an additional evaluation.
Of these subjects then investigated by upper and lower serum cut-offs defined for each HDT period, an additional eight subjects were diagnosed during HDT period 0–60 min, and nine subjects during each of the other HDT periods (0–30, 0–90, and 0–120 min) respectively.
So, the combination of basal cortisol and HDT allowed a diagnosis in 47–49% of patients.</td></tr>
<tr><td>Source of funding</td><td colspan="4">This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.</td></tr>
</table>

Reference	**Deutschbein 2009a[6]**
Limitations	Risk of bias: very serious (due to patient selection and patient flow) Indirectness: none
Comments	Unclear whether the patient sample overlaps with Deutschbein 2009 Query as to accuracy of reporting for sensitivity and specificity for salivary cortisol

Reference	**George 2020[9]**
Study type	Prospective diagnostic accuracy study
Study methodology	Data source: Outpatient assessments Recruitment: Consecutive outpatients with suspected AI were prospectively recruited
Number of patients	n = 67
Patient characteristics	Age, mean (SD): AI present - 40.69 ± 18.3, AI absent - 32.99 ± 15.8 Gender (male to female ratio): 29:38 Ethnicity: Not reported Setting: Tertiary referral centre for endocrine disorders Country: India Inclusion criteria: All patients with clinical suspicion of AI (anorexia, nausea, weight loss, unexplained hyponatraemia, progressive hyperpigmentation, history of pituitary macroadenoma or surgery) who attended the endocrine outpatient clinic and underwent SST during the study period were included. Exclusion criteria: Patients with a history of exogenous steroid intake for indications other than AI were excluded due to possible ethical issues related to worsening of primary disease condition while stopping the medication for study purpose. Patients with oral ulcers or poor oral hygiene were excluded. Patients who were already on replacement doses of steroids were advised to stop the medication (hydrocortisone for 24 hours and prednisolone for 72 hours) under close inpatient monitoring. Reason for referral: patients with clinical suspicion of AI (anorexia, nausea, weight loss, unexplained hyponatraemia, progressive hyperpigmentation, history of pituitary macroadenoma or surgery)

Reference	**George 2020[9]**
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index tests: Basal salivary cortisol** All tests were initiated between 0800 and 0900 hours. Patients were instructed to refrain from brushing their teeth, smoking, eating or drinking except water for at least 60 min before and during sampling. Drinking water was not allowed from 10 minutes before salivary sample collection. Saliva was collected in plain plastic containers (CML Biotech, Kochi) using passive drooling method until a sufficient volume (approximately 2 ml) was obtained. Salivary samples were collected at 0 (Basal), 60 (APSST 60) and 120 (APSST 120) minutes for measurement of salivary cortisol. APSST was done three days after the SST by IM administration of 30 units of porcine corticotrophin (Acton Prolongatum). The three-day interval was chosen for logistical reasons of doing the tests in outpatient department and to avoid possible potentiation of ACTH action interfering with the second stimulatory test. **Threshold cut-off value:** 14.1 nmol/L - optimal cut-off values for diagnosis of AI based on SST with a diagnostic cut-off value of <500nmol/L. 3.0 nmol/L cut-off values for diagnosis of AI based on SST with a diagnostic cut-off value of 400.1 nmol/L. **Assay:** Serum and salivary cortisol were measured by electrochemiluminescence method (ECLIA) using Cobas e 411 analysers with commercially available Elecsys Cortisol II (second generation, monoclonal antibody) kits which have showed a good correlation with gas chromatography mass spectrometry (GC-MS). Untreated centrifuged saliva was used for the assay. The measuring range was 0.054-63.4 µg/dl. Intra-assay coefficient of variation (CV) at high and low concentrations were 2.5% and 6.1%, respectively. Intraassay variability at high and low concentration had a CV of 3.6% and 11.8%, respectively. The functional sensitivity (lower limit of quantification) of the assay was 3 nmol/L. **Reference standard: ACTH short Synacthen test** The SST was performed by administering 250 µg of Synacthen intramuscularly, and blood sampling was done at 0, 30 and 60 minutes. **Threshold:** Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L(18 µg/dl). **Assay**: Serum and salivary cortisol were measured by electrochemiluminescence method (ECLIA) using Cobas e 411 analysers with commercially available Elecsys Cortisol II (second generation, monoclonal antibody) kits which have showed a good correlation with gas chromatography mass spectrometry (GC-MS). **Time between measurement of index test and reference standard**: 3 days
2×2 table (calculated from reported	Basal salivary cortisol (cut off 14.1 nmol/L) with an SST diagnostic cut-off value of <500nmol/L

	Reference standard +	Reference standard −	Total
Index test +	30	30	60

Reference	**George 2020[9]**				
sensitivity, specificity and true positive and true negative rates)	Index test −	2	5	7	
	Total	32	35	67	
(calculated from reported sensitivity, specificity and true positive and true negative rates)	Basal salivary cortisol (cut off 3.0 nmol/L) with an SST diagnostic cut-off value of <400.1 nmol/L				
		Reference standard +	Reference standard −	Total	
	Index test +	26	5	31	
	Index test −	9	27	36	
	Total	35	32	67	
Statistical measures	Index text – basal salivary cortisol (cut off 14.1 nmol/L) with an SST diagnostic cut-off value of <500nmol/L Sensitivity 93 CI (86.8-99.2) Specificity 14.3 CI (5.7- 22.9) PPV 0.50 NPV 0.71 PLR 1.09 NLR 0.44 AUC 0.82 Youden's Index 0.74 Index text – basal salivary cortisol (cut off 3.0 nmol/L) with an SST diagnostic cut-off value of <400.1 nmol/L Sensitivity 73.9 (63.2-84.6) Specificity 85.3 (76.6-93.9) PPV 0.84 NPV 0.75 PLR 4.75 NLR 0.30 AUC 0.86 Youden's Index 0.59				
Source of funding	Not reported				
Limitations	Risk of bias: Serious (due to patient flow) Indirectness: none				
Comments	Index test was the APSST but for purposes of the review only the basal salivary cortisol values are being used				

Reference	**Kim 2020[11]**
Study type	Prospective diagnostic study
Study methodology	Data source: Hospital assessments (unclear if inpatients or outpatients) Recruitment: 120 patients were prospectively included from Seoul National University Hospital from April 2013 to January 2014 who underwent the SST
Number of patients	n = 120
Patient characteristics	Age, mean (range): AI: 58 (42-67), non-AI: 58 (43-67) Gender (male to female ratio): 50:70 Ethnicity: Korean Setting: outpatients national university hospital Country: Korea Inclusion criteria: people suspected of having primary or secondary AI. Exclusion criteria: To prevent confounding factors, subjects with infection or bleeding in the oral cavity, a diagnosis of severe liver or renal disease, oral contraceptive use, and pregnancy were excluded. Reasons for investigation: The subjects had a history of pituitary disease (n=67), adrenal disease (n=10), or suspected secondary AI with iatrogenic Cushing syndrome (n=43). Of these subjects, 31 patients underwent pituitary surgery and only one patient took a steroid replacement postoperatively.
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index test:** Basal salivary cortisol The subjects were instructed to avoid eating, drinking anything except water, brushing teeth, and smoking beginning 1 hour before the test. Basal serum and salivary cortisol samples were collected between 8:00 AM and 10:00 AM and the SST was performed subsequently. Saliva was collected by chewing an oral cotton swab (Salivette, Sarstedt, Germany) for 2 to 5 minutes and the samples were frozen at −20°C until analysis. Basal and stimulated serum blood samples were immediately centrifuged at 4°C for 15 minutes and the resulting serum was stored at −20°C until use.

<table>
<tr><td>Reference</td><td colspan="4">Kim 2020[11]</td></tr>
<tr><td></td><td colspan="4">Assay: Salivary cortisol was analysed using an enzyme immunoassay kit (EIA, Salimetrics Inc., State College, PA, USA). The intra-assay CV was 3.2% to 6.3%, and the inter-assay CV was 5.7% to 6.8%. The expected morning ranges of salivary cortisol derived using this kit were 3.1 to 22.4 nmol/L and 4.1 to 20.4 nmol/L in adult men and women aged 51 to 70, respectively.

<u>Reference standard</u>
Stimulated serum cortisol levels were measured at 30 and 60 minutes after intravenous administration of 250 μg of synthetic adrenocorticotropic hormone (ACTH1-24) (Synacthen, Novartis, Basel, Switzerland). Stimulated serum blood samples were immediately centrifuged at 4°C for 15 minutes and the resulting serum was stored at −20°C until use.

Threshold: The AI group was defined by a level of stimulated serum cortisol less than 496.8 nmol/L

<u>Assays</u>: Serum total cortisol was measured using a Packard Cobra Gamma Counter analyser with commercial radioimmunoassay (RIA) kits (CIS Bio International, Saclay, France; inter-assay coefficient of variation [CV], 4.7%; intra-assay CV, 4.2%).

Time between measurement of index test and reference standard: reference standard was immediately after the index test</td></tr>
<tr><td>2×2 table (calculated from reported sensitivity, specificity and true positive and true negative rates)</td><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr>
<tr><td></td><td>Index test +</td><td>29</td><td>23</td><td>52</td></tr>
<tr><td></td><td>Index test −</td><td>5</td><td>63</td><td>68</td></tr>
<tr><td></td><td>Total</td><td>34</td><td>86</td><td>120</td></tr>
<tr><td>Statistical measures</td><td colspan="4"><u>Index text –</u>
Optimal baseline salivary cortisol cut point of 3.2 nmol/L
Sensitivity 84.9
Specificity 73.5
PPV 0.56
NPV 0.92
PLR 3.20
NLR 0.21
AUC 0.822 (95% CI, 0.732–0.913)</td></tr>
<tr><td>Source of funding</td><td colspan="4">This research received a grant from Seoul National University Hospital (grant number 03-2013-0310).</td></tr>
<tr><td>Limitations</td><td colspan="4">Risk of bias: Unclear methods of recruitment (serious)</td></tr>
</table>

Reference	**Kim 2020[11]**
	Indirectness: none
Comments	

Reference	**Mak 2017[13]**
Study type	Prospective diagnostic accuracy
Study methodology	Data source: One hundred seventy-one Chinese patients referred for the investigation of adrenal insufficiency and underwent hospital assessment. Recruitment: 56 healthy subjects were recruited from hospital staff and their relatives to derive cut offs. 171 patients suspected of having AI from the clinical context constituted the patient group.
Number of patients	n = 171
Patient characteristics	Age, mean (SD): AI group: 56.9 (14.2), non-AI group: 56.6 (12.4) Gender (male to female ratio): 82:89 Ethnicity: Chinese Setting: Prospective study in a regional hospital in Hong Kong from January 2014 to September 2015. Country: Hong Kong Inclusion criteria: patients suspected of having AI Exclusion criteria: All subjects with acute hypocortisolism, hemodynamic instability at the time of testing, bleeding inside the oral cavity, pregnancy, or mental incapacity for informed consent were excluded. Reasons for investigation: nonspecific clinical features (n = 55) such as unexplained dizziness and fatigue, electrolyte disturbance (n = 18) such as hyponatremia, diseases or interventions involving the sellar and suprasellar regions (n = 124), previous administration of exogenous glucocorticoids (n = 32), and post adrenalectomy (n = 4)
Target condition(s)	Adrenal insufficiency

Reference	**Mak 2017[13]**
Index test(s) and reference standard	**Index test: basal serum cortisol, salivary cortisol, and cortisone during low dose short Synachten testing** Patients receiving chronic hydrocortisone replacement had to stop the drug for at least 3 days before the test. A saliva sample for steroid measurement was collected by placing the cotton tubes (Salivette) in the mouth, chewing for 2 to 3 minutes, and then carefully putting the Salivette into a plastic container without touching it with hands. No drinking, eating, using mouthwash, or brushing teeth was allowed within 30 minutes before the collection of saliva samples, and subjects were advised to refrain from smoking or eating liquorice within 24 hours. All subjects underwent the low-dose SST (LDSST) in the morning. Simultaneous saliva and serum samples were collected at baseline. A 1-mg bolus of Synacthen was then injected intravenously. At 30 and 60 minutes after the injection, 2 more pairs of saliva and serum samples were collected. The higher value of the tested parameters, at either 30 or 60 minutes, was regarded as the "peak" value. **Assay**: Salivary cortisol and cortisone were assayed with LC-MS/MS using the Waters Xevo TQ MS system (Waters, Milford, MA). The assay CV for salivary cortisol was ~5% to 7% across all ranges; that for salivary cortisone was ~7% to 10% across all ranges. The lower limit of detection was 0.5 nmol/L for both salivary cortisol and cortisone. **Reference standard:** post-LDSST peak serum cortisol. Cut off: 376 nmol/L All subjects underwent the low-dose SST (LDSST) in the morning. Simultaneous saliva and serum samples were collected at baseline. A 1-mg bolus of Synacthen was then injected intravenously. At 30 and 60 minutes after the injection, 2 more pairs of saliva and serum samples were collected. The higher value of the tested parameters, at either 30 or 60 minutes, was regarded as the "peak" value. **Threshold:** The value corresponding to the mean 2 SDs of the peak serum cortisol in this healthy cohort (376 nmol/L) was used as the gold standard **Assay:** Serum total cortisol was assayed with the competitive chemiluminescent microparticle immunoassay using the Abbott Architect i2000SR system (Abbott Laboratories, Abbott Park, IL). The assay coefficient of variation (CV) was 4.0% to 6.2% at low levels and 3.3% to 4.3% at high levels. Serum CBG was measured using a commercial human CBG enzyme-linked immunosorbent assay kit (BioVendor– Laboratorni Medicina, Brno, Czech Republic). **Time** between measurement of index test and reference standard: simultaneously
2×2 table (calculated from reported sensitivity, specificity and	**Basal salivary cortisol (threshold 1.7 nmol/L)**

	Reference standard +	Reference standard −	Total
Index test +	37	5	42
Index test −	22	107	129
Total	59	112	171

Reference	**Mak 2017[13]**				
true positive and true negative rates)					
(calculated from reported sensitivity, specificity and true positive and true negative rates)	**Basal salivary cortisone (threshold 12.5 nmol/L)**				
		Reference standard +	Reference standard −	Total	
	Index test +	43	13	56	
	Index test −	16	99	115	
	Total	59	112	171	
	Basal serum cortisol 170 nmol/L				
		Reference standard +	Reference standard −	Total	
	Index test +	45	13	58	
	Index test −	14	99	113	
	Total	59	112	171	
Statistical measures	Index text (Basal salivary cortisol) 1.7 nmol/L Sensitivity 62.7 Specificity 95.5 PPV 0.88 NPV 0.83 PLR 14.05 NLR 0.39 AUC 0.836 (0.766–0.906) Index text (Basal salivary cortisone) 12.5 nmol/L Sensitivity 72.9 Specificity 88.4 PPV 0.77 NPV 0.86 PLR 6.279 NLR 0.307 AUC 0.862 (0.797–0.926) Index text (Basal serum cortisol) 170 nmol/L Sensitivity 76.3 Specificity 88.4 PPV 0.78				

Reference	**Mak 2017[13]**
	NPV 0.88 PLR 6.571 NLR 0.268 AUC 0.903 (0.856–0.951)
Source of funding	This work was supported by Grant KCC/RC/G/1415-A03 from the Hospital Authority, Hong Kong.
Limitations	Risk of bias: unclear patient selection (serious) Indirectness: none
Comments	

Reference	**Schmidt 2003[8]**
Study type	Cross sectional diagnostic study
Study methodology	Data source: Hospital assessments (unclear if inpatient or outpatient) Recruitment: Fifty-four patients were evaluated at our department because of suspected disease of the HPA axis from July 2001 until July 2002
Number of patients	n = 54
Patient characteristics	Age, mean (SD): 46.6 5 ± 2.5 years. Gender (male to female ratio): 19:22 Ethnicity: Not reported Setting: Single centre, department of endocrinology Country: Germany Inclusion criteria: patients with suspected hypothalamic-pituitary-adrenal disease

Reference	**Schmidt 2003[8]**
	Exclusion criteria: Not reported. Underlying diagnoses: Thirty-five had a history of tumours in the pituitary area (14 nonfunctioning adenomas, seven somatotropic adenomas, four prolactinomas, one case of neurosarcoidosis, two cases of hypophysitis, six craniopharyngeomas, and one chordoma), one patient had diabetes insipidus, and in five cases disease was suspected because of clinical symptoms.
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index test** Basal cortisol. Unstimulated serum cortisol values between 0800 and 0900h were taken for comparison with the peak cortisol response to the dynamic test. **Threshold:** Twelve morning cortisol values of the total of 20 healthy volunteers were available. The mean basal cortisol value was 439.3 =/- 24.9 nmol/L (326.0–600.0 nmol/L). The lower limit of a normal basal cortisol was calculated as 267 nmol/L (mean 2 sd). ROC analysis suggested an optimal baseline cortisol cut point of 285 nmol/L. **Reference standard** Patients underwent an ITT between 0900 and 1030 h by injection of 0.15 IU/kg regular insulin (Actrapid, Novo Nordisk, Mainz, Germany) to achieve blood glucose levels less than 40 mg/dl and until symptoms of hypoglycaemia developed. Blood samples were taken at 0, 15, 30, 45, 60, 90, and 120 min. **Threshold:** Patients were defined as adrenal insufficient or sufficient based on their cortisol peak response to hypoglycaemia of less than 500 nmol/L or more than 500 nmol/L. **Time between measurement of index test and reference standard:** reference standard was immediately after the index test. **Assays**: Serum cortisol levels (nanomoles per litre) were assayed at each time point by competitive immunoassay (ADVIA Centaur System, Bayer, Fernwald, Germany). The lower detection limit was assessed to be 5.5 nmol/L (0.20 g/dl). Intraassay variations as coefficient of variation for various cortisol values were 3.69% (107.05 nmol/L), 3.09% (155.33 nmol/L), 2.89% (390.95 nmol/L), 3.82% (759.55 nmol/L), and 2.98% (1024.97 nmol/L). Inter assay variations for the above-mentioned cortisol concentrations were 5.45, 3.83, 3.07, 1.86, and 3.99%.
2×2 table (calculated from reported sensitivity, specificity and	

	Reference standard +	Reference standard −	Total
Index test +	20	8	28
Index test −	0	13	13
Total	20	21	41

Reference	Schmidt 2003[8]				
true positive and true negative rates)					
Statistical measures	Index text – Optimal baseline serum cortisol cut point of 285 nmol/l Sensitivity 1 Specificity 0.6 PPV 0.70 NPV 1 PLR 2.5 NLR 0 AUC 0.88				
Source of funding	Not reported				
Limitations	Risk of bias: serious (due to unclear methods of recruitment) Indirectness: none				
Comments					

Reference	Tan 2023[18]
Study type	Prospective diagnostic study
Study methodology	Data source: Outpatients assessment for adrenal insufficiency Recruitment: 42 subjects were recruited from the endocrine testing clinic at the Department of Endocrinology at a tertiary centre (August 2020 to January 2022) who were planned for adrenocorticotrophic hormone (ACTH) stimulation test (AST) for the evaluation of suspected AI.
Number of patients	n = 42
Patient characteristics	Age, mean (SD): AI group = 62.2 (14.6), non-AI group = 51.1 (16.4) Gender (male to female ratio): 26:16 Ethnicity: Not reported

Reference	Tan 2023[18]
	Setting: Outpatient endocrinology unit Country: Singapore Inclusion criteria: The study prospectively recruited 42 outpatients undergoing evaluation for AI. Exclusion criteria: conditions that may affect cortisol binding, including liver cirrhosis, advanced chronic kidney disease (estimated glomerular filtration rate < 30 mL/min/1.73m2 or end-stage kidney failure on renal replacement therapy), pregnancy, oral contraceptive medication use, active malignancies or eating disorders, or weight loss of >10% over the past 3 months. Subjects with recent dental procedures or oral bleeding and those unable to follow instructions or provide informed consent were also excluded. Aetiologies of AI: The aetiologies of AI were due to exogenous glucocorticoid use in 16, primary AI in 2, hypopituitarism in 1, and post-adrenalectomy for adrenal Cushing's syndrome in 1. Thirteen were prescribed regular hydrocortisone replacement, 7 were given standby hydrocortisone to be taken on sick days, and 1 was not prescribed hydrocortisone.
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index test: Morning salivary cortisol** Eligible patients underwent the AST, which was conducted in the outpatient endocrine unit between 08:00 and 10:00 by specialised nursing staff. Subjects on chronic hydrocortisone replacement were instructed to omit the medication on the morning of the test. They were also instructed not to eat, drink, or brush their teeth for 15 min prior to specimen collection and throughout the duration of the test. An intravenous cannula was inserted followed by the simultaneous collection of baseline plasma and saliva samples (0-min sample. Saliva specimens were collected using the SARSTEDT Salivette®. Subjects were instructed to place the SARSTEDT Salivette® into their mouth for 2 min to obtain at least 1.5 mL of saliva per specimen. About 250 µg of ACTH (Synacthen®, Novartis) was then injected intravenously, followed by the simultaneous collection of serum and salivary cortisol samples at 30 and 60 min. **Assay:** Salivary cortisol measurement was performed at the Mayo Medical Laboratories. Salivary cortisol was extracted from the specimen using online turbulent flow high-performance liquid chromatography and analysed by liquid chromatography tandem mass spectrometry (LC-MS/MS) using multiple reaction monitoring in positive mode. The detection limit was 0.11 nmol/L. The intra-assay coefficient of variation was 7.2% at 3.0 nmol/L, and the inter-assay coefficient of variation was 5.8% at 1.4 nmol/L. **Reference standard: serum cortisol after AST** Eligible patients underwent the AST, which was conducted in the outpatient endocrine unit between 08:00 and 10:00 by specialised nursing staff.

<table>
<tr><td>Reference</td><td colspan="4">Tan 2023[18]</td></tr>
<tr><td></td><td colspan="4">An intravenous cannula was inserted followed by the simultaneous collection of baseline plasma and saliva samples (0-min sample). Blood was collected in an EDTA tube and analysed upon receipt by the laboratory. About 250 µg of ACTH (Synacthen®, Novartis) was then injected intravenously, followed by the simultaneous collection of serum and salivary cortisol samples at 30 and 60 min.

Threshold: AI was diagnosed if peak serum cortisol levels failed to reach 500 nmol/L

Time between measurement of index test and reference standard: simultaneously

Assay: Serum cortisol was measured using the Beckman Coulter UniCel DxI 800 Access Immunoassay Systems. The detection limit was 11 nmol/L. The assay exhibited a total imprecision of <12% at approximately 138 nmol/L and <10% for higher concentrations of cortisol.</td></tr>
<tr><td>2×2 table (calculated from reported sensitivity, specificity and true positive and true negative rates)</td><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td></tr>
<tr><td></td><td>Index test (baseline cortisol) +</td><td>11</td><td>3</td><td>14</td></tr>
<tr><td></td><td>Index test −</td><td>10</td><td>18</td><td>28</td></tr>
<tr><td></td><td>total</td><td>21</td><td>21</td><td>42</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Statistical measures</td><td colspan="4"><u>Index text: baseline early morning salivary cortisol (cut off 2.7 nmol/L)</u>
Sensitivity 52.4
Specificity 85.7
PPV 0.79
NPV 0.64
PLR 3.66
NLR 0.56
AUC 0.661 (95% CI: 0.491–0.831)

Youden's index: The best 0-min salivary cortisol cut-off derived using Youden’s index was 2.7 nmol/L, with a sensitivity of 52.4% and a specificity of 85.7%.</td></tr>
<tr><td>Source of funding</td><td colspan="4">This work was funded by Pitch for Funds Award, SingHealth Duke-National University of Singapore Medicine Academic Clinical Programme Research Support Programme Grant (grant number 03/FY2018/P1/13-A28_FY2020PFF03).</td></tr>
<tr><td>Limitations</td><td colspan="4">Risk of bias: Unclear methods of recruitment and patient flow (very serious)
Indirectness: none</td></tr>
<tr><td>Comments</td><td colspan="4"></td></tr>
</table>

D.2 Diagnostic Thresholds for Referral

Reference	Choi 2002[3]
Study type	Cross sectional/prospective diagnostic accuracy study
Study methodology	**Data source**: Hospital assessments **Recruitment**: patients with clinically suspected secondary adrenocortical insufficiency recruited from March 1997 to December 2000.
Number of patients	n = 72
Patient characteristics	**Age:** mean – 46 years; range (28-74 years) **Gender** (male to female ratio): 30:42 **Ethnicity**: Chinese **Setting**: Regional hospital (unclear if in- or out-patients) **Country**: Hong Kong **Inclusion criteria**: clinically suspected secondary adrenocortical insufficiency **Exclusion criteria**: contraindications for ITT such as epilepsy or ischaemic heart disease **Underlying diagnoses**: reasons for suspicion of dysfunction of the hypothalamic-pituitary-adrenocortical axis were: nasopharyngeal carcinoma with radiotherapy to the pituitary region (n=22), iatrogenic Cushing's syndrome (n=20), non-functioning pituitary macroadenoma (n=14), empty sella syndrome (n=8), acromegaly (n=5), Sheehan's syndrome (n=1), Cooley's anaemia with secondary haemochromatosis (n=1), and prolactin-secreting pituitary macroadenoma (n=1). **Prior tests/treatment**: Patients receiving chronic steroid therapy for other medical illnesses were advised to cease steroid therapy 1 week before the tests.
Target condition(s)	Secondary adrenal insufficiency

Reference	**Choi 2002[3]**
Index test(s) and reference standard	**Index test** Morning fasting serum cortisol concentration (0900) **Threshold:** to maximise sensitivity or specificity **Reference standard** (insulin tolerance test) Performed by a specialty nurse supervised by a doctor after overnight fast. Actrapid insulin (NovoNordisk, Bagsvaerd, Denmark) 0.1 U/kg was given intravenously after a baseline blood sample was taken for glucose and cortisol assay. Diabetic patients receiving insulin treatment were given actrapid insulin 0.15 U/kg plus 50% of their morning dose of short-acting insulin. Bedside haemoglucostix (Surestep Plus; Lifescan, Milpitas, US) testing was performed at 15-minute intervals and blood was sampled for laboratory glucose assay. When hypoglycaemia developed (bedside haemoglucostix result of less than 2.2 mmol/L and occurrence of hypoglycaemic symptoms), a blood sample was taken for glucose and cortisol assay, and 50% dextrose solution 40 mL was given to the patient intravenously, followed by oral food. If hypoglycaemia had not occurred by 45 minutes, an additional dose of actrapid insulin 0.1 U/kg was given. Blood sampling for cortisol measurement was done at intervals of 15 minutes during the first 60 minutes, and thereafter at 30-minute intervals. **Threshold:** peak cortisol response of ≥550 nmol/L for adrenal sufficiency **Assays**: Cortisol was assayed by the chemiluminescence method (Bayer-Centaur, New York, US). The coefficient of variation for the assay was less than 5%. **Time** between measurement of index test and reference standard: 4-7 days
Statistical measures	**Index text (fasting morning cortisol)** Fasting morning cortisol range: 51 to 537 nmol/L. Threshold: ≥420 nmol/L – sensitivity 100% (no false negatives) Threshold: ≤112 nmol/L – specificity 100% (no false positives)
Source of funding	Not stated
Limitations	**Risk of bias:** Unclear methods of recruitment; time between index and reference tests 4-7 days **Indirectness**: None identified
Comments	

Reference	**Debono 2023[5]**
Study type	Cross sectional/prospective diagnostic accuracy study

Reference	**Debono 2023[5]**
Study methodology	**Data source**: Hospital assessments **Recruitment**: Patients were recruited by consecutive sampling at Sheffield Teaching Hospitals NHS Foundation Trust between November 2019 and December 2021.
Number of patients	n = 220 (208 available for primary outcome measure)
Patient characteristics	**Age:** 55.1–15.8 years **Gender** (male to female ratio): 106:102 **Ethnicity**: White: 90%, Asian: 5%, and the remaining patients were Black/Caribbean/African and multiracial. **Setting**: NHS hospital for ACTH and home-based salivary cortisone test **Country**: UK **Inclusion criteria**: Patients at high risk for adrenal insufficiency. All patients referred for an ACTH stimulation test to assess for a new diagnosis of adrenal insufficiency or for recovery from a previous diagnosis of adrenal insufficiency were considered for the study. Patients older than age 18 years with a high probability of either primary, secondary, or tertiary adrenal insufficiency as determined by the investigators were eligible for enrolment. **Exclusion criteria**: Patients were excluded who were unable to produce a suitable saliva sample; night shift workers; patients with known protein-losing disorders, known or suspected alcohol dependence, and known severe liver disease; patients with uncontrolled active infection; and patients taking oestrogens or those who were pregnant. Patients taking drugs that influence the hypothalamic-pituitary-adrenal axis (e.g., opioids) had their medications omitted on the day of testing, per routine clinical practice. In view of the coronavirus disease 2019 measures resulting in limited staff and fewer appointment slots to enable study tasks, some patients were excluded from study participation so as not to hinder their clinical care. **Reason for referral**: Patients who were dependent on any type of glucocorticoid, who had been on an oral glucocorticoid prednisolone-equivalent dose of 5 mg/d for 4 weeks, and who were referred for adrenal testing only after they had been weaned down to prednisolone 5 mg/d or equivalent or converted to physiologic doses of hydrocortisone 25 mg/d. Patients receiving any intermediate- or long-acting intramuscular or intra-articular glucocorticoid injections were recruited at least 3 months after their last injection. Patients with pituitary disease, such as tumours, inflammatory disease, or those with a history of cranial radiotherapy, were considered eligible for inclusion.
Target condition(s)	Adrenal insufficiency

Reference	**Debono 2023[5]**
Index test(s) and reference standard	**Index test: waking salivary cortisone, cortisol and baseline serum – (threshold: different cut offs reported)** Salivary sample upon waking using Salivette tubes containing synthetic swabs (Salivette Cortisol; Sarstedt). A total of 500 ml of saliva was necessary to ensure a good representative sample, and 50 ml was used for the assay. All patients were given written instructions with images and a video on how to collect their salivary sample. Patients were advised to refrain from smoking/vaping on the day of the test. Patients taking glucocorticoids were asked to omit these medicines the evening before and the day of the test until all samples were collected. Patients were allowed to follow their usual waking routines but were asked to collect their waking salivary sample the moment they got out of bed to commence the day and before cleaning teeth, eating, or drinking. **Thresholds:** salivary cortisone = 7 and 17nmol/l, salivary cortisol = 5 and 1 nmol/L, serum cortisol = 152 and 310 nmol/l **Reference standard** Standard dose ACTH stimulation test At the endocrine clinic an intravenous cannula was inserted, and baseline serum cortisol was measured. An ACTH stimulation test was performed with intravenous injection of 250 mg of Synacthen (Atnahs Pharma UK Limited), followed by a serum cortisol level blood draw at 30 minutes. Tests were performed by specialized endocrine nurses at the clinic. On completing the ACTH stimulation test, patients completed the final part of the questionnaire assessing their views on the ACTH stimulation test and salivary sample collection. Results of the ACTH stimulation test were interpreted by a consulting endocrinologist, or a specialized endocrine nurse who was unaware of the waking cortisol data. An a priori criterion of a peak cortisol level of 15.6 mg/dl (430 nmol/l) measured by immunoassay indicated adequate adrenal reserve, whereas those patients with levels less than this value were considered to have adrenal insufficiency according to current clinical practice in our centre. **Threshold:** peak cortisol level of ≥430 nmol/Lfor adrenal sufficiency **Assays**: Serum cortisol was analysed by immunoassay (Elecsys Cortisol II assay; Roche) and interpreted immediately at Sheffield Teaching Hospitals NHS Foundation Trust. An extra serum cortisol sample was stored at -80C and, together with the salivary sample, was then analysed and interpreted by liquid chromatography–tandem mass spectrometry (LC-MS/MS) as a batch at the end of the study in a different laboratory in Manchester University NHS Foundation Trust. **Time:** later in the same day (median value = approx. 3 hours later as self-reported by patients)
2×2 table (calculated from reported sensitivity and negative predictive values)	**Index text (waking salivary cortisone – threshold 17 nmol/l)**

	Reference standard + (ACTH stimulation test 430nmol/l)	**Reference standard −** (ACTH stimulation test 430nmol/l)	**Total**
Index test + (waking salivary cortisone – threshold 17 nmol/l)	88	51	140
Index test –	3	66	68

Reference	Debono 2023[5]			
	(waking salivary cortisone – threshold 17 nmol/l)			
	Total	91	117	208
	Index text (waking salivary cortisone – threshold 7 nmol/l)			
		Reference standard + (ACTH stimulation test 430nmol/l)	**Reference standard −** (ACTH stimulation test 430nmol/l)	**Total**
	Index test + (waking salivary cortisone – threshold 7 nmol/l)	67	4	70
	Index test – (waking salivary cortisone – threshold 7 nmol/l)	24	113	138
	Total	91	117	208
	Index text (waking salivary cortisol – threshold 5 nmol/l)			
		Reference standard + (ACTH stimulation test 430nmol/l)	**Reference standard −** (ACTH stimulation test 430nmol/l)	**Total**
	Index test + (waking salivary cortisol – threshold 5 nmol/l)	86	46	132
	Index test – (waking salivary cortisol – threshold 5 nmol/l)	5	71	76
	Total	91	117	208
	Index text (waking salivary cortisol – threshold 1 nmol/l)			
		Reference standard + (ACTH stimulation test 430nmol/l)	**Reference standard −** (ACTH stimulation test 430nmol/l)	**Total**
	Index test + (waking salivary cortisol – threshold 1 nmol/l)	47	4	50
	Index test – (waking salivary cortisol – threshold 1 nmol/l)	44	113	158
	Total	91	117	208
	Index text (Baseline serum cortisol – threshold 310 nmol/l)			

<table>
<tr><td>Reference</td><td colspan="4">Debono 2023[5]</td></tr>
<tr><td></td><td></td><td>Reference standard +
(ACTH stimulation test 430nmol/l)</td><td>Reference standard −
(ACTH stimulation test 430nmol/l)</td><td>Total</td></tr>
<tr><td></td><td>Index test +
(waking salivary cortisol – threshold 1 nmol/l)</td><td>87</td><td>69</td><td>156</td></tr>
<tr><td></td><td>Index test – (waking salivary cortisol – threshold 1 nmol/l)</td><td>4</td><td>48</td><td>52</td></tr>
<tr><td></td><td>Total</td><td>91</td><td>117</td><td>208</td></tr>
<tr><td></td><td colspan="4">Index text (Baseline serum cortisol – threshold 152 nmol/l)</td></tr>
<tr><td></td><td></td><td>Reference standard +
(ACTH stimulation test 430nmol/l)</td><td>Reference standard −
(ACTH stimulation test 430nmol/l)</td><td>Total</td></tr>
<tr><td></td><td>Index test +
(waking salivary cortisol – threshold 1 nmol/l)</td><td>59</td><td>6</td><td>65</td></tr>
<tr><td></td><td>Index test – (waking salivary cortisol – threshold 1 nmol/l)</td><td>32</td><td>111</td><td>143</td></tr>
<tr><td></td><td>Total</td><td>91</td><td>117</td><td>208</td></tr>
<tr><td></td><td colspan="4"></td></tr>
<tr><td>Statistical measures
(some data calculated from reported sensitivity and negative predictive values)</td><td colspan="4"><u>Index text (waking salivary cortisone – threshold 17 nmol/l)</u>
Sensitivity: 97% (95% CI, 91 to 99) reported in study
Specificity: 56%
PPV: 63%
NPV: 96% (95% CI, 90 to 99) reported in study.
PLR: 2.20
NLR: 0.05

AUC: 0.95 (95% CI, 0.92 to 0.97) reported in study.

<u>Index text (waking salivary cortisone – threshold 7 nmol/l)</u>
Sensitivity: 73%
Specificity: 97% (95% CI, 92 to 99) reported in study
PPV: 95% (95% CI, 87 to 99)
NPV: 0.83
PLR: 24.4</td></tr>
</table>

Reference	Debono 2023[5]
	NLR: 0.28 Maximum thresholds - To achieve at least 99% sensitivity to exclude adrenal insufficiency and 99% specificity to confirm adrenal insufficiency, one would need to use waking salivary cortisone cutoffs of 25 nmol/Land <1 nmol/l, respectively. **Index text (waking salivary cortisol – threshold 5 nmol/l)** Sensitivity: 95% (95% CI, 88 to 99) reported in study Specificity: 61% PPV: 65% NPV: 94% (95% CI, 85 to 98) reported in study. PLR: 2.40 NLR: 0.09 AUC: 0.89 (95% CI, 0.85 to 0.94) reported in study. **Index text (waking salivary cortisol – threshold 1 nmol/l)** Sensitivity: 52% Specificity: 97% (95% CI, 92 to 100) reported in study PPV: 93% (95% CI, 80 to 99) reported in study. NPV: 72% PLR: 15.1 NLR: 0.50 **Index text (baseline serum cortisol – threshold 310 nmol/l)** Sensitivity: 96% (95% CI, 90 to 99) reported in study Specificity: 41% PPV: 56% NPV: 93% (95% CI, 84 to 98) reported in study. PLR: 1.64 NLR: 0.09 AUC: 0.90 (95% CI, 0.86 to 0.94) reported in study. **Index text (baseline serum cortisol – threshold 152 nmol/l)** Sensitivity: 65% Specificity: 95% (95% CI, 90 to 98) reported in study

Reference	**Debono 2023[5]**
	PPV: 91% (95% CI, 81 to 97) reported in study. NPV: 77% PLR: 13.0 NLR: 0.37 **Additional analyses:** Examined the percentage of ACTH stimulation tests that could have been avoided using waking salivary cortisone cutoffs as a screening test in a two-stage process. The ACTH stimulation test would have been avoided in 70% (154 of 220) of participants — 73 patients with waking salivary cortisone.
Source of funding	Funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit Programme (grant reference number, PBPG-1217-20007). Dr. Elder and Mr. Keevil are both supported by funding from UK Research and Innovation–Innovation Scholars secondments: Biomedical Sciences. The study was also supported by the NIHR Clinical Research Network Yorkshire and Humber IRAS 262618.
Limitations	**Risk of bias:** None identified. **Indirectness**: None identified.
Comments	

Reference	**De Lange 1993[4]**
Study type	Cross sectional diagnostic accuracy study
Study methodology	**Data source**: Hospital assessments **Recruitment**: consecutive patients who had received adequately performed insulin-induced hypoglycaemia and metyrapone tests
Number of patients	n = 58
Patient characteristics	**Age:** range 17-73 years **Gender** (male to female ratio): 33:25 **Ethnicity**: not stated **Setting**: Regional hospital (unclear if in- or out-patients)

Reference	De Lange 1993[4]
	Country: Netherlands **Inclusion criteria**: received adequately performed insulin-induced hypoglycaemia and metyrapone tests. **Exclusion criteria**: clinical evidence of adrenal insufficiency or Cushing's syndrome **Underlying diagnoses**: 45 had pituitary adenoma (30 before treatment; 10 after hypophysectomy and 5 after X-ray therapy). 13 had proven or suspected HPA disease. **Prior tests/treatment**: Corticosteroid medication was stopped at least 72 hours before investigation.
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index test** Morning fasting plasma cortisol concentration (0900-1000) **Thresholds:** 160 and 260 nmol/L **Reference standard** (insulin tolerance test) Performed after overnight fast. Soluble insulin 0.15 U/kg was given intravenously. Blood sampling for glucose and cortisol measurement was taken before and at 0, 30, 45, 60 and 90 minutes after injection. **Threshold:** rise in plasma cortisol above 500 nmol/L for adrenal sufficiency (normal – n=33/58) **Assays**: Plasma cortisol was measured by RIA. Blood glucose was measured by a hexokinase method. **Time** between measurement of index test and reference standard: 0-1 days (2 samples taken for index test, one on the day of ITT and one the following day)
Statistical measures	**Index text (fasting morning cortisol)** Threshold: <260 nmol/L – sensitivity = 96%; specificity = 64% Threshold: <160 nmol/L – sensitivity = 64%; specificity = 91%
Source of funding	Not stated

Reference	**De Lange 1993[4]**
Limitations	**Risk of bias:** None identified. **Indirectness**: Basal cortisol was mean of morning measurements on 2 consecutive days
Comments	

Reference	**Deutschbein 2009[7]**
Study type	Cross sectional diagnostic accuracy study
Study methodology	Data source: Hospital assessments Recruitment: patients were investigated because of suspected or proven disease of the HPA axis (dates of recruitment not reported)
Number of patients	n = 77
Patient characteristics	Age, mean (SD): 44.2 (1.8) years. Gender (male to female ratio): 41:36 Ethnicity: Not reported Setting: hospital (unclear if inpatient or outpatient) Country: Germany Inclusion criteria: suspected or proven HPA axis disease Exclusion criteria: No subject had to be excluded because of contraindications to insulin-induced hypoglycaemia. **Underlying diagnoses** Reason for referral: 65 patients had sellar masses, of whom 11 were not operated; all 54 surgically treated patients were tested at least 3 months after surgery (median postoperative interval: 6.3 months; range: 3 – 68 months). The remaining 12 patients showed an impaired secretion of various hormones but did not have any detectable tumours within the HPA. **Prior tests/treatment**: At the time of hormonal evaluation, female patients were neither on contraceptives nor oestrogens. Patients on chronic corticosteroid replacement therapy (generally 10–15 mg hydrocortisone per day) received their last dosage at 1400 h the day before testing, resulting in a drug restriction period of at least 18 h.

<table>
<tr><td>Reference</td><td colspan="4">Deutschbein 2009[7]</td></tr>
<tr><td>Target condition(s)</td><td colspan="4">Secondary adrenal insufficiency</td></tr>
<tr><td>Index test(s) and reference standard</td><td colspan="4"><u>Index tests: Basal serum and salivary cortisol</u>
Basal salivary cortisol: saliva was collected by chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h and before blood withdrawal.
Basal serum cortisol: venipuncture.

<u>Threshold cut-off value</u>:
Basal salivary cortisol - an optimal cut-off of 15.1 nmol/l, an upper cut-off of 21.1 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated.
Basal serum cortisol - ROC analysis revealed an optimal cut-off of 283 nmol/l, an upper cut-off of 470 nmol/l, and a lower cut-off of 103 nmol/Lrespectively.

<u>Assay:</u>
Serum cortisol was measured by a competitive immunoassay (Advia Centaur, Bayer, Germany). The lower detection limit of this assay was 5.5 nmol/l, and intra- and inter-assay coefficients of variation were less than 3.8 and 5.5%, respectively. Salivary cortisol was determined using a modification of a commercial radioimmunoassay (RIA) (GammaCoat, DiaSorin, USA), decreasing the sample volume from 200 to 100 µ l. The intra- and inter-assay coefficients of variation were 5.4 and 15.9 %, respectively.

<u>Reference standard: Insulin tolerance test (ITT)</u>
The ITT was performed between 0800 and 1000, using a peak cortisol cut-off point of 500 nmol/Lto distinguish between adrenal insufficient (AI) and adrenal sufficient (AS) patients. After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycaemia were required as evidence of sufficient stress. Samples for blood glucose and Se C were taken at 0, 15, 30, 45, 60, 90, and 120 min.

<u>Threshold</u>: Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L (18 µg/dl).

Time between measurement of index test and reference standard: ITT was performed wither immediately after basal samples or on a separate day. Average time lag not reported.</td></tr>
<tr><td>Data table</td><td colspan="4">Morning salivary cortisol (threshold 15.1 nmol/l)
<table>
<tr><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Similar diagnoses by ITT and basal cortisol</td><td></td></tr>
<tr><td>Index test + (<5.0 nmol/l)</td><td>19%</td><td>1%</td><td></td><td></td></tr>
</table></td></tr>
</table>

Reference	**Deutschbein 2009[7]**				
	Index test unclear (5.0-21.1 nmol/l)	31%	38%		
	Index test − (>21.1 nmol/l)	3%	**8%**	**27%**	
	Morning serum cortisol (threshold 283 nmol/l)				
		Reference standard +	Reference standard −	Similar diagnoses by ITT and basal cortisol	
	Index test + (<103 nmol/l)	**17%**	1%		
	Index test unclear (103-470 nmol/l)	34%	39%		
	Index test − (>470 nmol/l)	3%	**6%**	**23%**	
Statistical measures	Upper and lower thresholds (with ≥ 95% specificity either for adrenal sufficiency or adrenal insufficiency) were calculated based on ROC analysis. **Index text – basal salivary cortisol (upper cutoff of 21.1 nmol/l, and a lower cutoff of 5.0 nmol/l)** By using the upper as well as the lower cutoff, a diagnosis corresponding to the ITT result was established in 21 of 77 patients (27 %), whereas 56 patients (73 %) required additional testing procedures. **Index text – basal serum cortisol (upper cutoff of 470 nmol/l, and a lower cutoff of 103 nmol/l)** By applying both the upper and the lower cutoff, a diagnosis corresponding to the ITT result was established in 18 of 77 patients (23 %), leaving 59 patients (77 %) for further evaluation.				
Source of funding	Not stated				
Limitations	Risk of bias: very serious (due to patient selection and patient flow) Indirectness: none				
Comments	Unclear whether the patient sample overlaps with Deutschbein 2009a				

Reference	**Deutschbein 2009a[6]**
Study type	Prospective diagnostic accuracy study
Study methodology	Data source: Hospital assessments (unclear if inpatients or outpatients) Recruitment: Between 2005 and 2007 fifty-five patients were investigated because of suspected or proven disease of the HPA axis
Number of patients	n = 55
Patient characteristics	Age, mean (SD): 45.9 (2.1) years. Gender (male to female ratio): 26:29 Ethnicity: Not reported Setting: outpatients Country: Germany Inclusion criteria: patients with suspected secondary adrenal insufficiency Exclusion criteria: At the time of hormonal evaluation, female patients were neither on contraceptives nor oestrogens. Patients on chronic corticosteroid replacement therapy (generally 10–15 mg hydrocortisone per day) received their last dosage at 1400 h the day before testing, resulting in a drug restriction period of at least 18 h. No subject had to be excluded because of contraindications to insulin-induced hypoglycaemia. Reason for referral: Eight subjects had a present neoplasia (4 prolactinomas, 2 somatotropic adenomas, 1 non-functioning adenoma, 1 meningioma), and 40 subjects (24 non-functioning adenomas, 8 somatotropic adenomas, 3 craniopharyngiomas, 3 prolactinomas, 2 meningiomas) were tested at least 3 months after surgical treatment (median interval: 4.8 months). Seven patients suffered from pituitary hormone impairment without detectable sellar tumours (3 secondary hypogonadisms, 3 congenital hormone insufficiencies, 1 Sheehan syndrome).
Target condition(s)	Adrenal insufficiency
Index test(s) and reference standard	**Index tests: Basal serum and salivary cortisol** Basal salivary cortisol: saliva was collected by chewing a specific cotton swab (Salivette, Sarstedt, Germany). Following an overnight fast, basal cortisol was collected between 0800 and 0900 h. Basal serum cortisol: after catheterisation of a superficial cubital vein and a recovery period of 15 min to avoid stress-induced bias, blood samples were directly obtained into serum tubes (Monovetten, Sarstedt, Germany).

Reference	Deutschbein 2009a[6]
	During High dose SST, serum and saliva samples were taken at 0, 30, 60, 90, and 120 min after i.v. application of 250 mg synthetic ACTH (Synacthen, Novartis). **Threshold cut-off value:** Basal salivary cortisol - an optimal cut-off of 7.6 nmol/l, an upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. Basal serum cortisol - ROC analysis revealed an optimal cut-off of 260 nmol/l, an upper cut-off of 382 nmol/l, and a lower cut-off of 103 nmol/Lrespectively. **Assay:** Saliva samples were measured by a modification of the 'GammaCoat' RIA for cortisol (DiaSorin, Stillwater, MN, USA), decreasing the sample volume from 200 to 100 ml. The intra-assay and inter-assay coefficients of variation were 5.4% and 15.9%, respectively. Serum cortisol levels were determined by competitive immunoassay, using commercial kits (Advia Centaur, Bayer). The analytical sensitivity of the assay was 5.5 nmol/l. Intra-assay variations as coefficient of variation for various cortisol values were 3.7% (107.1 nmol/l), 3.1% (155.3 nmol/l), 2.9% (391.0 nmol/l), 3.8% (759.6 nmol/l), and 3.0% (1025.0 nmol/l). Inter assay variations for the cortisol concentrations mentioned above were 5.5, 3.8, 3.1, 1.9, and 4.0%. **Reference standard: Insulin tolerance test (ITT)** The ITT served as reference test, using a peak cortisol cut-off point of 500 nmol/Lto distinguish between adrenal insufficient (AI) and adrenal sufficient (AS) patients. After administration of insulin, both a serum glucose nadir below 40 mg/dl and symptoms of hypoglycaemia were required as evidence of sufficient stress. Samples for blood glucose and Se C were taken at 0, 15, 30, 45, 60, 90, and 120 min. **Threshold:** Cortisol sufficiency was confirmed if the peak serum cortisol was more than 500 nmol/L(18 µg/dl). **Assay**: Serum and salivary cortisol were measured by electrochemiluminescence method (ECLIA) using Cobas e 411 analysers with commercially available Elecsys Cortisol II (second generation, monoclonal antibody) kits which have showed a good correlation with gas chromatography mass spectrometry (GC-MS). **Time between measurement of index test and reference standard**: The minimum and maximum intervals between both tests were 1 day and 17 days, respectively, with a median interval of 2 days (unclear if this was the basal values)

Reference	Deutschbein 2009a[6]
Statistical measures	Index text – basal salivary cortisol An upper cut-off of 17.5 nmol/l, and a lower cut-off of 5.0 nmol/Lwere calculated. The use of both the upper and lower thresholds allowed a diagnosis in 19 of 55 patients (35%). If the remaining 36 subjects were then investigated by the short Synacthen test (SST), the combination of upper and lower cut-offs diagnosed 3 (HDT periods 0–60 and 0–120 min), 4 (HDT period 0–30 min), and 5 additional patients (HDT period 0–90 min). So, the combination of basal cortisol and HDT allowed a diagnosis in 42–45% of patients. Index text – basal serum cortisol An upper cut-off of 382 nmol/l, and a lower cut-off of 103 nmol/Lwere identified. By applying the upper as well as the lower cut-off, 18 of 55 patients (33%) were diagnosed by their basal serum cortisol levels, leaving 37 subjects for an additional evaluation. Of these subjects then investigated by upper and lower serum cut-offs defined for each SST period, an additional eight subjects were diagnosed during SST period 0–60 min, and nine subjects during each of the other SST periods (0–30, 0–90, and 0–120 min) respectively. So, the combination of basal cortisol and SST allowed a diagnosis in 47–49% of patients.
Source of funding	This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.
Limitations	Risk of bias: very serious (due to patient selection and patient flow) Indirectness: none
Comments	Unclear whether the patient sample overlaps with Deutschbein 2009 Query as to accuracy of reporting for sensitivity and specificity for salivary cortisol

Appendix E Forest plots

E.1 Diagnostic Tests Forest plots

E.1.1 Coupled sensitivity and specificity forest plots.

Figure 2: Sensitivity and specificity of serum cortisol for diagnosing AI in people with suspected AI and

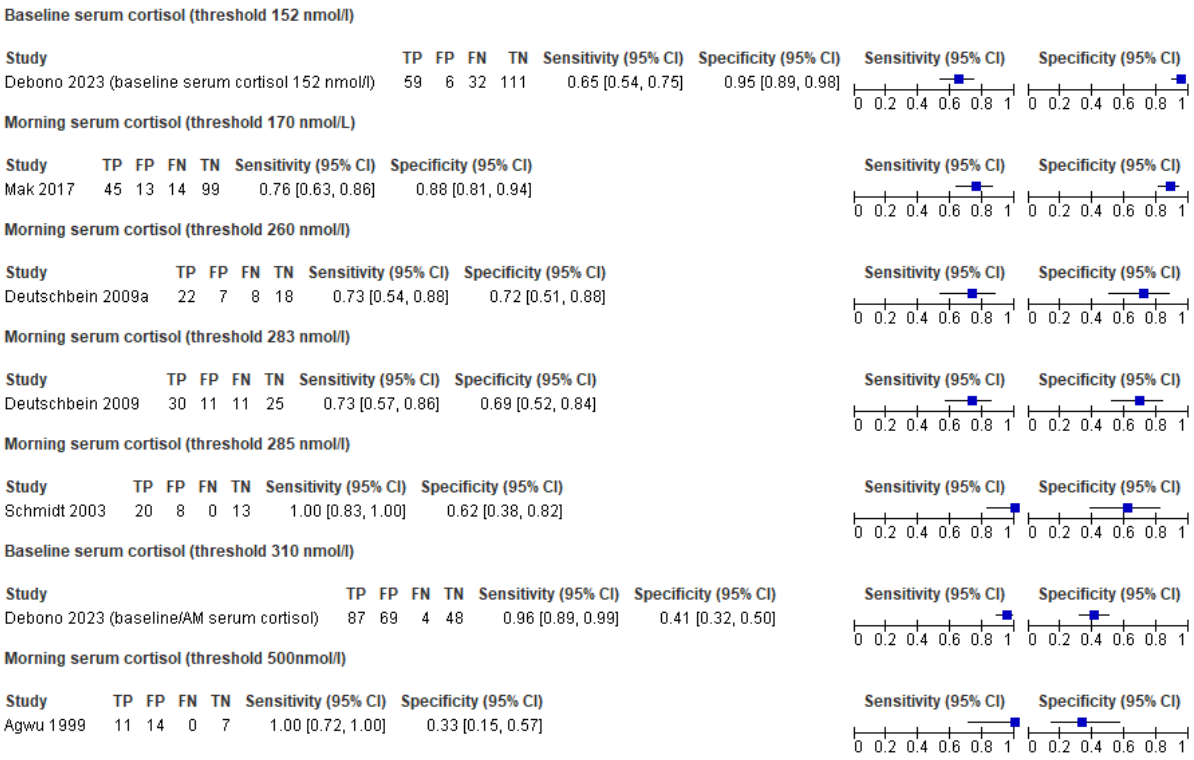


Figure 3: Sensitivity and specificity of salivary cortisol for diagnosing AI in people with suspected AI

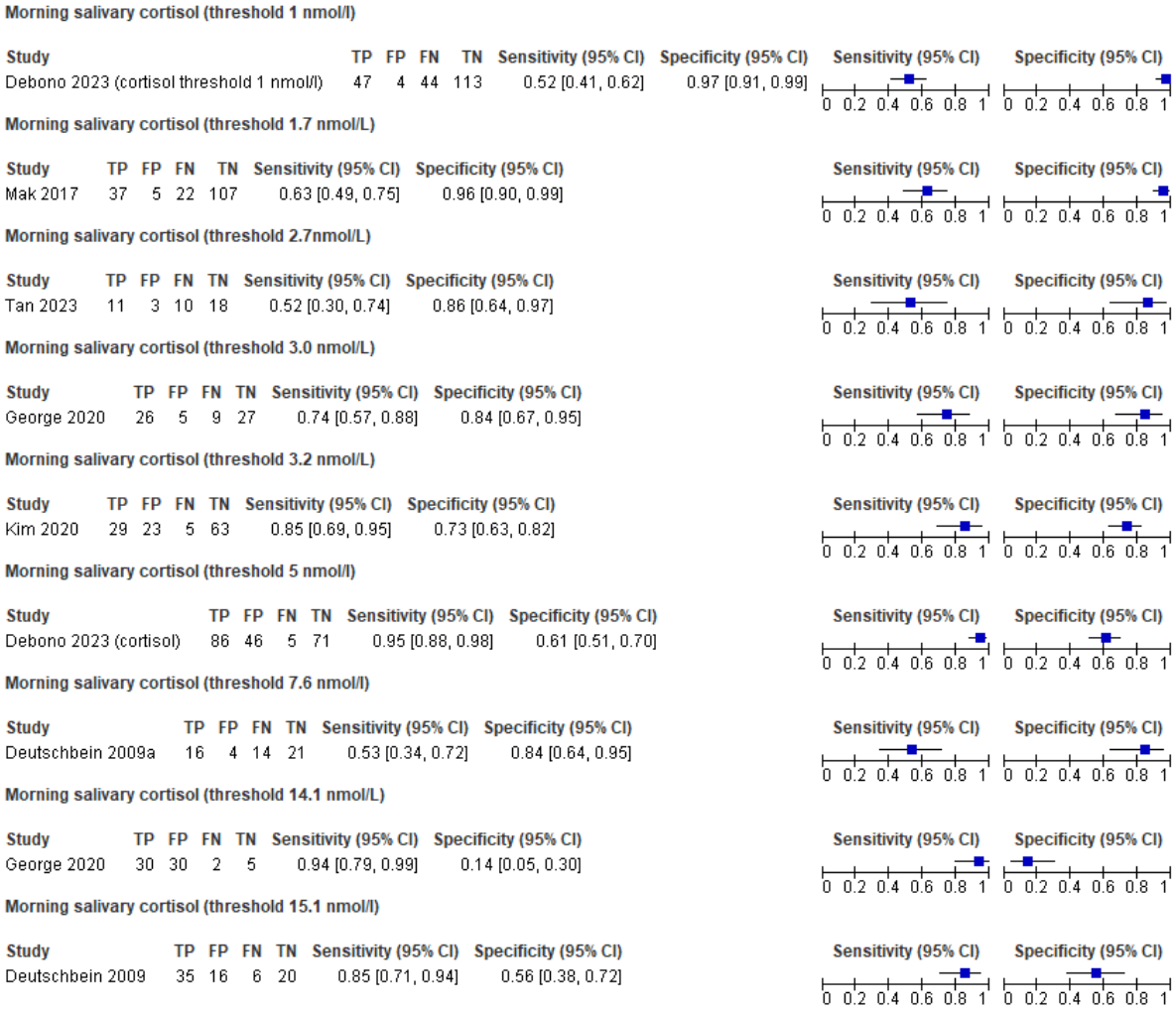
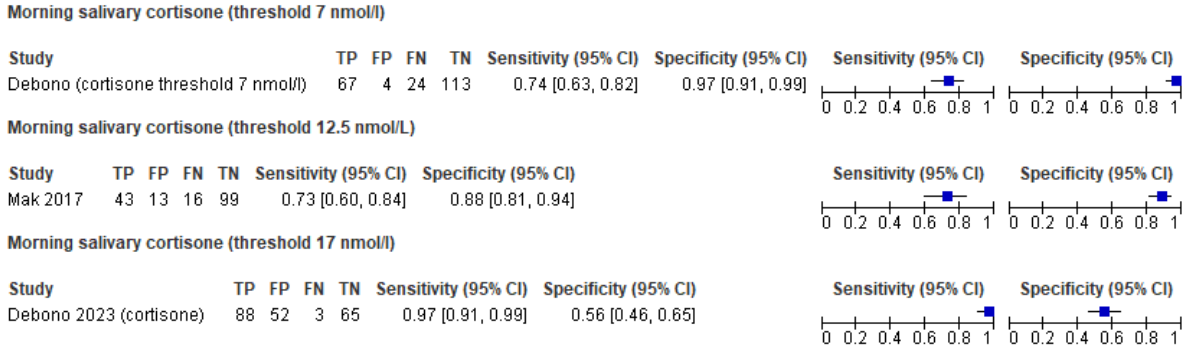


Figure 4: Sensitivity and specificity of salivary cortisone (threshold 12.5 nmol/L) for diagnosing AI in people with suspected AI



E.1.2 Analyses by study (multiple index tests in single population)

***Note**: these data are duplicated in the analyses by index test above.*

Figure 5: Sensitivity and specificity of serum and salivary cortisol for diagnosing secondary AI in people with HPA axis disorders

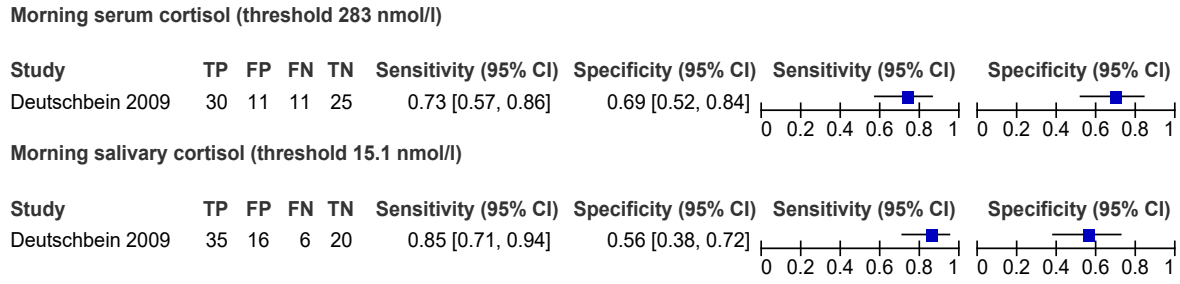


Figure 6: Sensitivity and specificity of serum and salivary cortisol for diagnosing secondary AI in people with HPA axis disorders

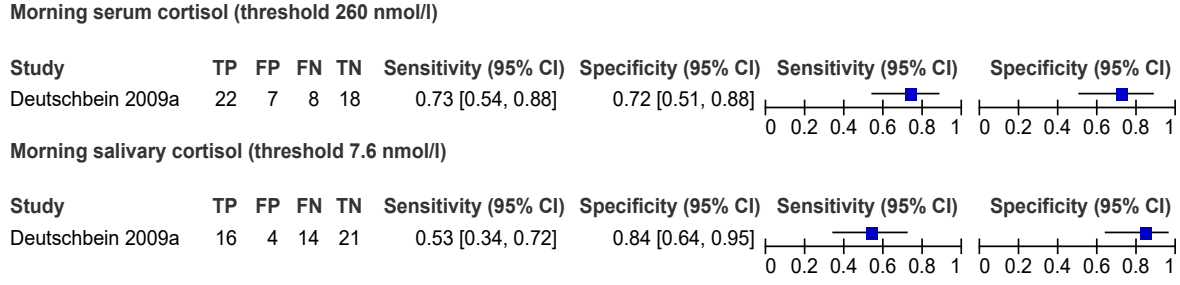


Figure 7: Sensitivity and specificity of serum cortisol, salivary cortisol and salivary cortisone for diagnosing AI in people with suspected AI

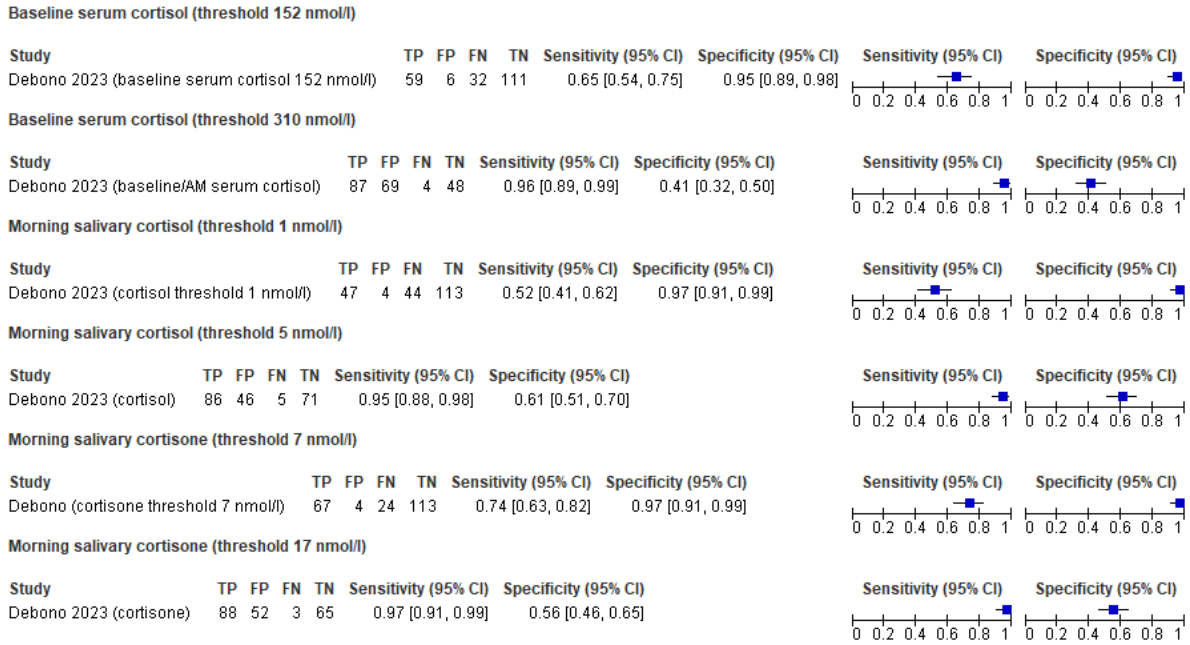
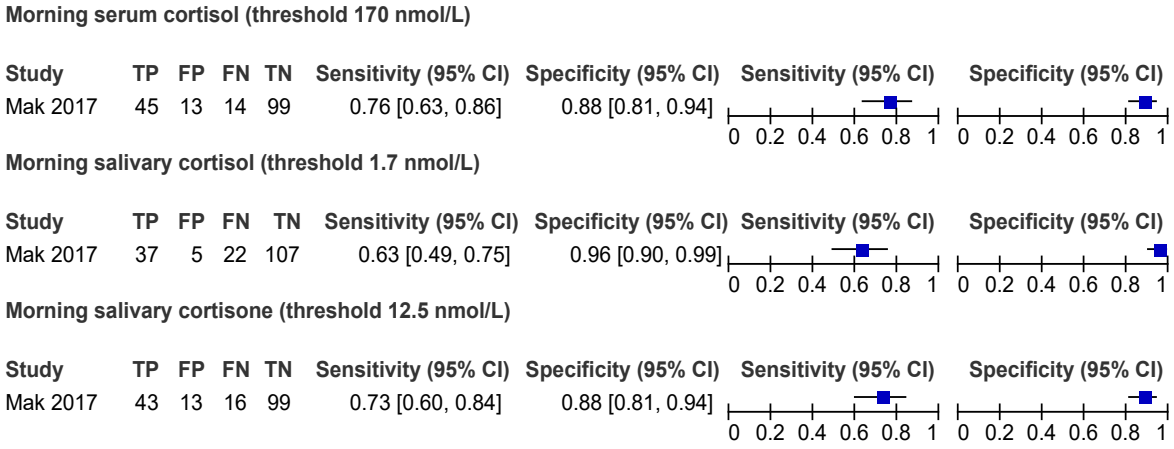


Figure 9: Sensitivity and specificity of serum cortisol, salivary cortisol and salivary cortisone for diagnosing AI in people with suspected AI

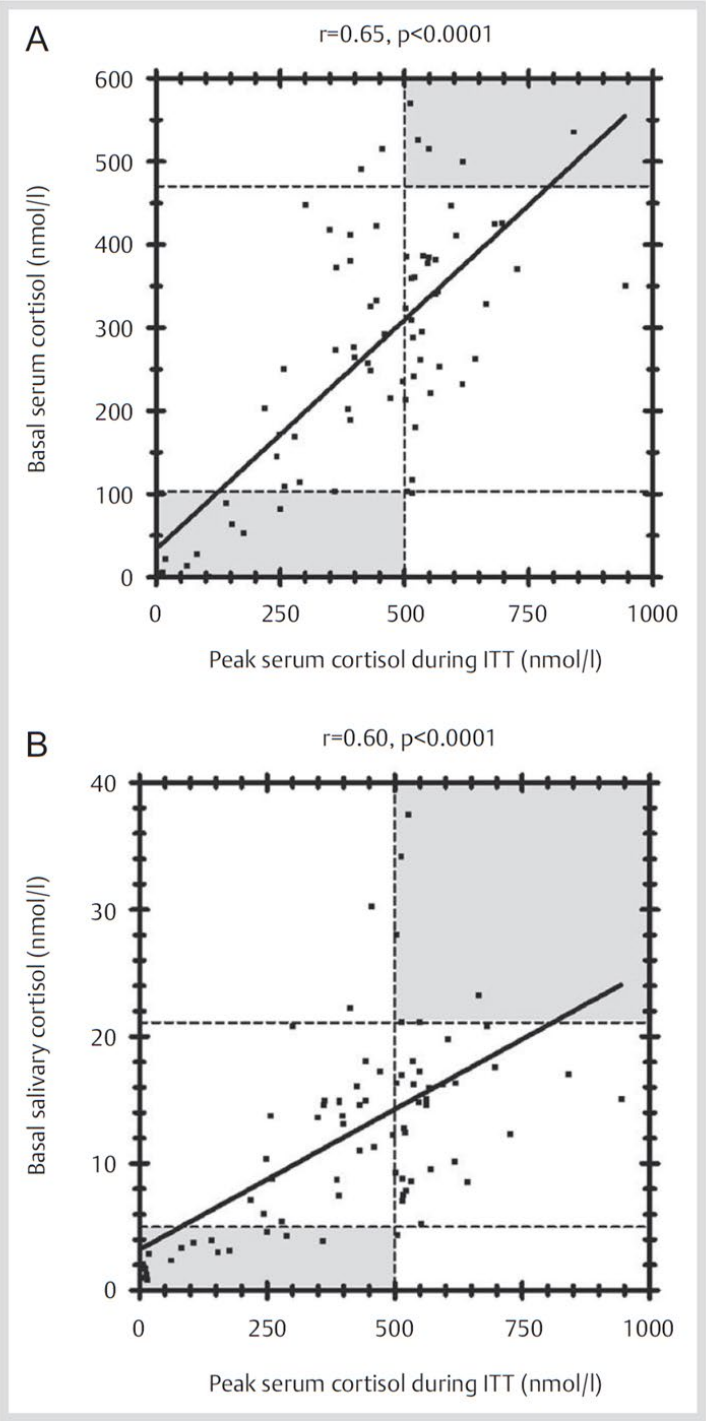


E.2 Diagnostic thresholds for referral forest plots

Insufficient data were available from the included studies to produce forest plots.

E.3 Index test and reference standard scatter plots

Figure 8: Individual serum cortisol peak values during ITT plotted against matched basal serum (A) and salivary (B) cortisol levels.

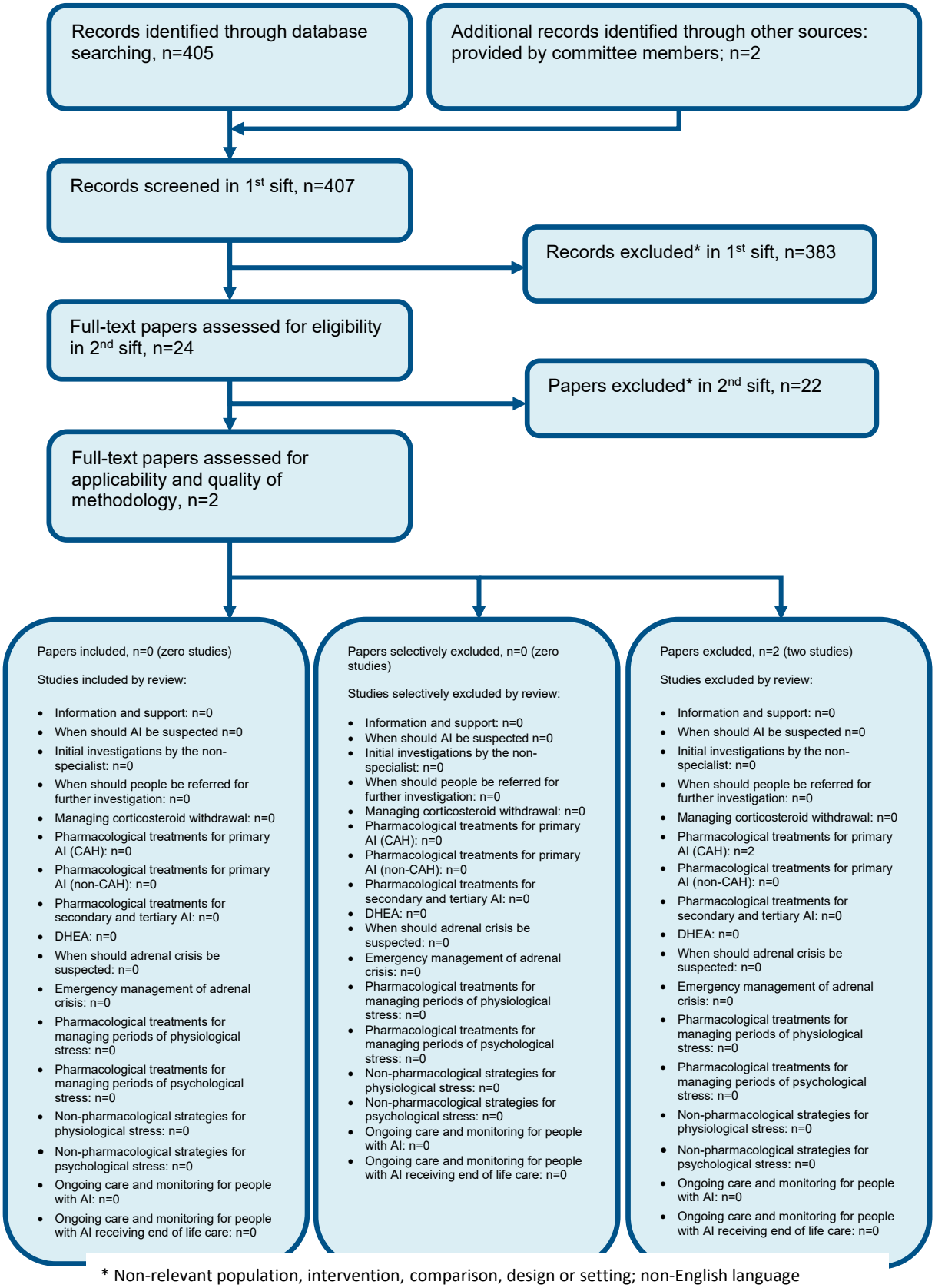


Upper and lower cutoffs for the index tests and the peak cortisol cut point of 500 nmol/Lin the reference standard are shown as broken lines. Dots within the shaded areas represent those who were similarly diagnosed by the index and reference tests.

Source: Deutschbein et al 2009

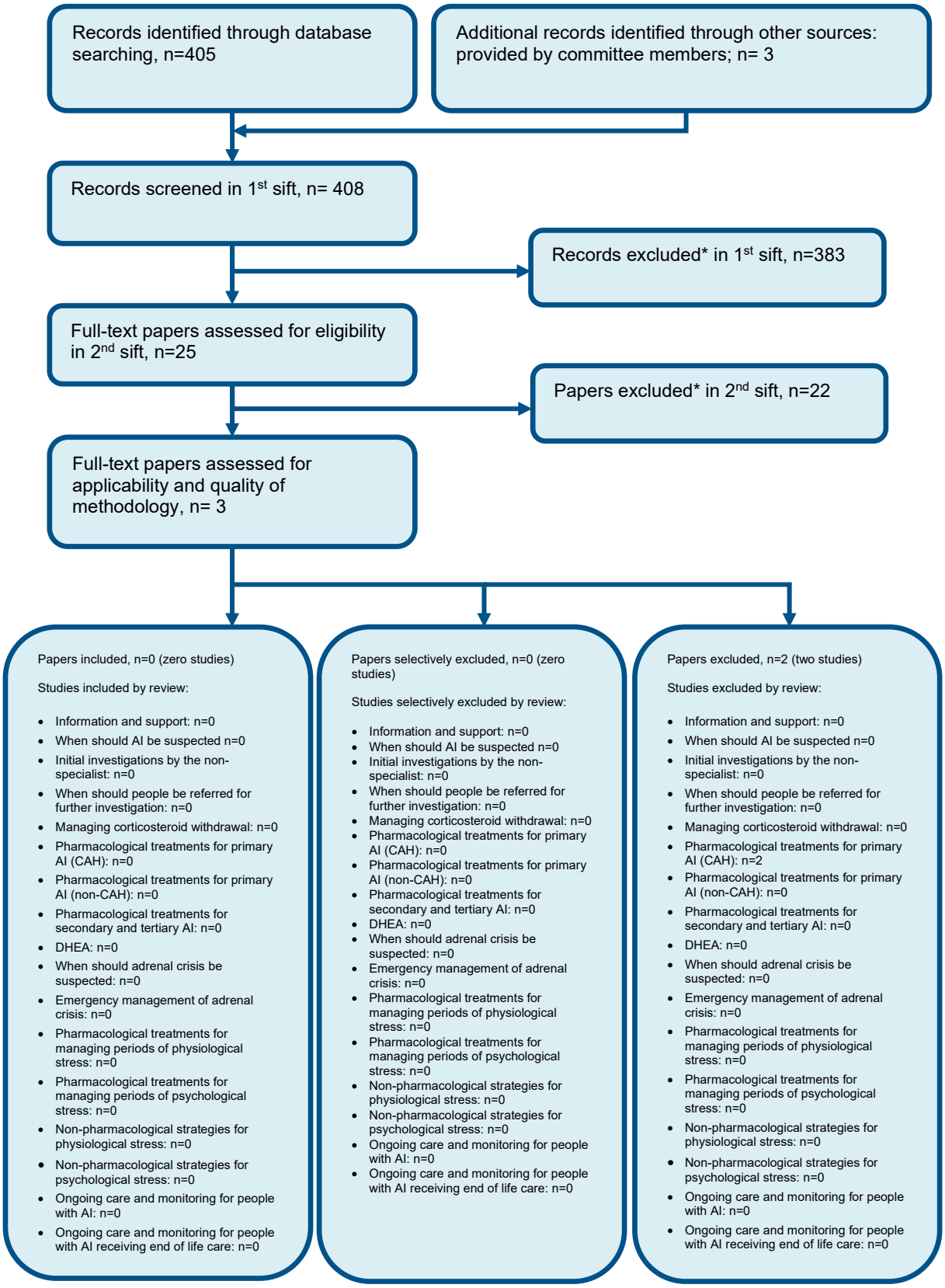
Appendix F Economic evidence study selection

F.1 Diagnostic tests



* Non-relevant population, intervention, comparison, design or setting; non-English language

F.2 Diagnostic thresholds for referral



* Non-relevant population, intervention, comparison, design or setting; non-English language

Appendix G Economic evidence tables

G.1 Diagnostic tests

None

G.2 Diagnostic thresholds for referral

None.

Appendix H Health economic model

H.1 Diagnostic tests

No original economic modelling was undertaken for this review question.

H.2 Diagnostic thresholds for referral

No original economic modelling was undertaken for this review question.

Appendix I Excluded studies

I.1 Clinical studies- Diagnostic tests

Table 20: Studies excluded from the clinical review

Study	Exclusion reason
Abdu, T A, Elhadd, T A, Neary, R et al. (1999) Comparison of the low dose short Synacthen test (1 microg), the conventional dose short Synacthen test (250 microg), and the insulin tolerance test for assessment of the hypothalamo-pituitary-adrenal axis in patients with pituitary disease. The Journal of clinical endocrinology and metabolism 84(3): 838-43	- Study does not contain any relevant index tests
Abdu, T.A.M. and Clayton, R.N. (2000) The low-dose Synacthen test for the assessment of secondary adrenal insufficiency. Current Opinion in Endocrinology and Diabetes 7(3): 116-121	- Review article but not a systematic review
Abeed, N.N.A.N., Mohamad, W.M.I.W., Yahya, N. et al. (2022) ACCURACY OF RANDOM SERUM CORTISOL IN DIAGNOSING SECONDARY ADRENAL INSUFFICIENCY. Journal of the ASEAN Federation of Endocrine Societies 37(supplement2): 12	- Conference abstract
Abraham, Smita Baid, Abel, Brent S, Sinaii, Ninet et al. (2015) Primary vs secondary adrenal insufficiency: ACTH-stimulated aldosterone diagnostic cut-off values by tandem mass spectrometry. Clinical endocrinology 83(3): 308-14	- Study does not contain a relevant reference standard Plasma ACTH stimulation test - Population not relevant to this review protocol People with known AI
Agha, A., Tomlinson, J.W., Clark, P.M. et al. (2006) The long-term predictive accuracy of the short Synacthen (corticotropin) stimulation test for assessment of the hypothalamic-pituitary-adrenal axis. Journal of Clinical Endocrinology and Metabolism 91(1): 43-47	- Retrospective - Study does not contain diagnostic accuracy data
Albert, L., Profitos, J., Sanchez-Delgado, J. et al. (2019) Salivary cortisol determination in ACTH stimulation test to diagnose adrenal insufficiency in patients with liver cirrhosis. International Journal of Endocrinology 2019: 7251010	- Population not relevant to this review protocol Known AI and healthy controls.
Ambrogio, Alberto G, Danesi, Leila, Baldini, Marina et al. (2018) Low-dose Synachten test with measurement of salivary cortisol in adult patients with beta-thalassemia major. Endocrine 60(2): 348-354	- Study does not report sensitivity or specificity Youden's index only
Ambrosi, B, Barbetta, L, Re, T et al. (1998) The one microgram adrenocorticotropin test in the assessment of hypothalamic-pituitary-adrenal function. European journal of endocrinology 139(6): 575-9	- Study does not contain any relevant index tests
Amin, H., Wynne-Edwards, K., Amin, P. et al. (2017) Is the Correlation between Salivary Cortisol and Serum Cortisol Reliable Enough to Enable Use of Salivary Cortisol Levels in	- Study does not contain diagnostic accuracy data correlation only

Study	Exclusion reason
Preterm Infants?. American Journal of Perinatology 34(13): 1302-1305	
Arregger, Alejandro L, Cardoso, Estela M L, Tumilasci, Omar et al. (2008) Diagnostic value of salivary cortisol in end stage renal disease. Steroids 73(1): 77-82	- Population not relevant to this review protocol critically ill - end stage renal disease
Arregger, Alejandro L, Cardoso, Estela M L, Zucchini, Alfredo et al. (2014) Adrenocortical function in hypotensive patients with end stage renal disease. Steroids 84: 57-63	- Population not relevant to this review protocol End stage renal disease
Atluri, Sridevi, Sarathi, Vijaya, Goel, Amit et al. (2022) Long-acting Porcine Sequence ACTH (Acton Prolongatum) Stimulation Test is a Reliable Alternative Test as Compared to the Gold Standard Insulin Tolerance Test for the Diagnosis of Adrenal Insufficiency. Indian journal of endocrinology and metabolism 26(1): 38-43	- Population not relevant to this review protocol Includes some with known AI. - Study design not relevant to this review protocol Recruitment method not clear
Bancos, Irina, Erickson, Dana, Bryant, Sandra et al. (2015) PERFORMANCE OF FREE VERSUS TOTAL CORTISOL FOLLOWING COSYNTROPIN STIMULATION TESTING IN AN OUTPATIENT SETTING. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 21(12): 1353-63	- Study does not report sensitivity or specificity AUC only
Bangar, V. and Clayton, R.N. (1998) How reliable is the short Synacthen test for the investigation of the hypothalamic-pituitary-adrenal axis?. European Journal of Endocrinology 139(6): 580-583	- Retrospective
Butt, Muhammad Imran, Alzuhayri, Nouf, Amer, Lama et al. (2020) Comparing the utility of 30- and 60-minute cortisol levels after the standard short Synacthen test to determine adrenal insufficiency: A retrospective cross-sectional study. Medicine 99(43): e22621	- Retrospective
Ceccato, Filippo, Selmin, Elisa, Antonelli, Giorgia et al. (2021) Low-dose short Synacthen test with salivary cortisol in patients with suspected central adrenal insufficiency. Endocrine connections 10(9): 1189-1199	- Retrospective
Cemeroglu, Ayse Pinar, Kleis, Lora, Postellon, Daniel C et al. (2011) Comparison of low-dose and high-dose cosyntropin stimulation testing in children. Pediatrics international : official journal of the Japan Pediatric Society 53(2): 175-80	- Study does not contain diagnostic accuracy data
Cetinkaya, Semra; Ozon, Alev; Yordam, Nursen (2007) Diagnostic value of salivary cortisol in children with abnormal adrenal cortex functions. Hormone research 67(6): 301-6	- Study does not contain any relevant index tests
Cheung, K.K.-T., So, W.-Y., Ma, R.C.-W. et al. (2015) Spot and morning cortisol in comparison to low dose short Synacthen test. Journal of the ASEAN Federation of Endocrine Societies 30(2): 147-154	- Data not reported in an extractable format or a format that can be analysed Text discussion confuses the interpretation of sensitivity and specificity for ruling in or out the condition; and threshold values for serum cortisol appear incorrect with low values linked

Study	Exclusion reason
	to maximum sensitivity and high values to maximum specificity.
Chitale, Aditi, Musonda, Patrick, McGregor, Alan M et al. (2013) Determining the utility of the 60 min cortisol measurement in the short Synacthen test. Clinical endocrinology 79(1): 14-9	- Retrospective
Chng, E., Lam, S., Hawkins, R. et al. (2014) The use of short Synacthen test in patients on exogenous steroids use in diagnosing adrenal insufficiency. Endocrine Reviews 35(suppl3)	- Conference abstract
Cho, Hwa Y, Kim, Jung H, Kim, Sang W et al. (2014) Different cut-off values of the insulin tolerance test, the high-dose short Synacthen test (250 mug) and the low-dose short Synacthen test (1 mug) in assessing central adrenal insufficiency. Clinical endocrinology 81(1): 77-84	- Study does not contain any relevant index tests
Colling, Caitlin, Nachtigall, Lisa, Biller, Beverly M K et al. (2022) The biochemical diagnosis of adrenal insufficiency with modern cortisol assays: Reappraisal in the setting of opioid exposure and hospitalization. Clinical endocrinology 96(1): 21-29	- Retrospective
Contreras, L.N., Arregger, A.L., Tumilasci, O. et al. (2006) Salivary steroids in response to ACTH: A less invasive approach to assess adrenal function in hypotensive patients with chronic renal failure. Endocrinologist 16(1): 30-35	- Study does not contain diagnostic accuracy data
Cornes, Michael P, Ashby, Helen L, Khalid, Yasmeen et al. (2015) Salivary cortisol and cortisone responses to tetracosactrin (Synacthen). Annals of clinical biochemistry 52(pt5): 606-10	- Study does not contain diagnostic accuracy data correlation only
Cortez, Samuel, Arbelaez, Ana Maria, Wallendorf, Michael et al. (2023) Peak Serum Cortisol Cutoffs to Diagnose Adrenal Insufficiency Across Different Cortisol Assays in Children. Journal of clinical research in pediatric endocrinology	- Study does not contain an intervention relevant to this review protocol
de Vries, Friso, Lobatto, Daniel J, Bakker, Leontine E H et al. (2020) Early postoperative HPA-axis testing after pituitary tumor surgery: reliability and safety of basal cortisol and CRH test. Endocrine 67(1): 161-171	- Retrospective
Dichtel, L.E., Schorr, M., De Assis, C.L. et al. (2017) Plasma free cortisol vs. Total cortisol in healthy individuals and in states of high and low cortisol binding globulin, including oral contraceptive use, cirrhosis and critical illness: implications for diagnosing adrenal insufficiency. Endocrine Reviews 38(3supplement1)	- Conference abstract
Dildar, S., Khan, A.H., Jaffri, S.A. et al. (2023) Clinical Utility of 30- and 60-min Serum Cortisol Values in Cosyntropin Stimulation Test for	- Conference abstract

Study	Exclusion reason
Diagnosis of Adrenal Insufficiency. Medical Forum Monthly 34(4): 22-24	
Dineen, Rosemary, Mohamed, Ahmed, Gunness, Anjuli et al. (2020) Outcomes of the short Synacthen test: what is the role of the 60 min sample in clinical practice?. Postgraduate medical journal 96(1132): 67-72	- Retrospective
Dluhy, R.G.; Himathongkam, T.; Greenfield, M. (1974) Rapid ACTH test with plasma aldosterone levels. Improved diagnostic discrimination. Annals of Internal Medicine 80(6): 693-696	- Study does not contain a relevant reference standard plasma ACTH test
Dorin, Richard I; Qualls, Clifford R; Crapo, Lawrence M (2003) Diagnosis of adrenal insufficiency. Annals of internal medicine 139(3): 194-204	- Review article but not a systematic review
Fede, Giuseppe, Spadaro, Luisa, Privitera, Graziella et al. (2015) Hypothalamus-pituitary dysfunction is common in patients with stable cirrhosis and abnormal low dose Synacthen test. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver 47(12): 1047-51	- Population not relevant to this review protocol selected those with AI.
Ferrante, Emanuele, Morelli, Valentina, Giavoli, Claudia et al. (2012) Is the 250 mug ACTH test a useful tool for the diagnosis of central hypoadrenalism in adult patients with pituitary disorders?. Hormones (Athens, Greece) 11(4): 428-35	- Study design not relevant to this review protocol Cohort study with 6 weeks between index and reference standard test
Fragoso Perozo, A F D, Fontes, R, Lopes, F P et al. (2023) Morning serum cortisol role in the adrenal insufficiency diagnosis with modern cortisol assays. Journal of endocrinological investigation	- Retrospective
Gasco, Valentina, Bima, Chiara, Geranzani, Alice et al. (2021) Morning Serum Cortisol Level Predicts Central Adrenal Insufficiency Diagnosed by Insulin Tolerance Test. Neuroendocrinology 111(12): 1238-1248	- Retrospective
Giordano, R, Picu, A, Bonelli, L et al. (2008) Hypothalamus-pituitary-adrenal axis evaluation in patients with hypothalamo-pituitary disorders: comparison of different provocative tests. Clinical endocrinology 68(6): 935-41	- Study does not contain any relevant index tests
Gleeson, H.K., Walker, B.R., Seckl, J.R. et al. (2003) Ten years on: Safety of short Synacthen tests in assessing adrenocorticotropin deficiency in clinical practice. Journal of Clinical Endocrinology and Metabolism 88(5): 2106-2111	- Retrospective
Goggans, F C, Wilson, W R Jr, Gold, M S et al. (1984) Comparison of the dexamethasone suppression test and the cortisol suppression index. The American journal of psychiatry 141(5): 698-700	- Study does not contain any relevant index tests - Population not relevant to this review protocol

Study	Exclusion reason
Gonc, E Nazli; Kandemir, Nurgun; Kinik, Sibel T (2003) Significance of low-dose and standard-dose ACTH tests compared to overnight metyrapone test in the diagnosis of adrenal insufficiency in childhood. Hormone research 60(4): 191-7	- Study does not contain a relevant reference standard metyrapone test
Gonc, E Nazli, Ozon, Z Alev, Alikasifoglu, Ayfer et al. (2011) Is basal serum 17-OH progesterone a reliable parameter to predict nonclassical congenital adrenal hyperplasia in premature adrenarche?. The Turkish journal of pediatrics 53(3): 274-80	- Retrospective
Goto, Masahiro; Shibata, Nao; Hasegawa, Yukihiro (2016) Efficacy of single serum cortisol reading obtained between 9 AM and 10 AM as an index of adrenal function in children treated with glucocorticoids or synthetic adrenocorticotropic hormone. Clinical pediatric endocrinology : case reports and clinical investigations : official journal of the Japanese Society for Pediatric Endocrinology 25(3): 83-9	- Retrospective
Grassi, G, Morelli, V, Ceriotti, F et al. (2020) Minding the gap between cortisol levels measured with second-generation assays and current diagnostic thresholds for the diagnosis of adrenal insufficiency: a single-center experience. Hormones (Athens, Greece) 19(3): 425-431	- Study does not contain diagnostic accuracy data Correlation between different assays
Gruvstad, Eva, Hedner, Lars Pavo, Hoglund, Peter et al. (2014) Comparison of methods for evaluation of the suppressive effects of prednisolone on the HPA axis and bone turnover: changes in s-DHEAS are as sensitive as the ACTH test. International journal of clinical pharmacology and therapeutics 52(1): 15-26	- Study design not relevant to this review protocol - Population not relevant to this review protocol
Gundgurthi, Abhay, Garg, M K, Dutta, M K et al. (2013) Intramuscular ACTH stimulation test for assessment of adrenal function. The Journal of the Association of Physicians of India 61(5): 320-4	- Study does not contain diagnostic accuracy data insufficient information to calculate accuracy data.
Hassan, Z., Nabi, S., Hussain, W. et al. (2021) Validation of Glucagon Stimulation Test in Establishing GH and ACTH Deficiency in Hypopituitarism. European Journal of Molecular and Clinical Medicine 8(4): 2005-2013	- Study design not relevant to this review protocol diagnostic accuracy for GH deficiency - Study does not contain any relevant index tests glucagon stimulation test
Javorsky, Bradley R, Raff, Hershel, Carroll, Ty B et al. (2021) New Cutoffs for the Biochemical Diagnosis of Adrenal Insufficiency after ACTH Stimulation using Specific Cortisol Assays. Journal of the Endocrine Society 5(4): bvab022	- Retrospective
Jayakumari, C., George, G.S., Nair, A. et al. (2017) ACTH stimulation test with long acting ACTH preparation for the diagnosis of adrenal insufficiency. Indian Journal of Endocrinology and Metabolism 21(8supplement1): 62	- Conference abstract

Study	Exclusion reason
Kadiyala, R, Kamath, C, Baglioni, P et al. (2010) Can a random serum cortisol reduce the need for short Synacthen tests in acute medical admissions?. Annals of clinical biochemistry 47(pt4): 378-80	- Retrospective
Kalaria, R.T., Agarwal, M., Kaur, S. et al. (2020) ANNALS EXPRESS: Hypothalamic-pituitary-adrenal (HPA) axis suppression a The value of salivary cortisol and cortisone in assessing HPA recovery. Annals of clinical biochemistry: 4563220961745	- Duplicate reference
Kamrath, Clemens and Boehles, Hansjosef (2010) The low-dose ACTH test does not identify mild insufficiency of the hypothalamnic-pituitary-adrenal axis in children with inadequate stress response. Journal of pediatric endocrinology & metabolism : JPEM 23(11): 1097-104	- Retrospective
Karpman, Matthew S, Neculau, Madalina, Dias, Valerian C et al. (2013) Defining adrenal status with salivary cortisol by gold-standard insulin hypoglycemia. Clinical biochemistry 46(15): 1442-6	- Study does not contain any relevant index tests
Kazlauskaite, Rasa, Evans, Arthur T, Villabona, Carmen V et al. (2008) Corticotropin tests for hypothalamic-pituitary- adrenal insufficiency: a metaanalysis. The Journal of clinical endocrinology and metabolism 93(11): 4245-53	- Systematic review used as source of primary studies
Kline, G A; Buse, J; Krause, R D (2017) Clinical implications for biochemical diagnostic thresholds of adrenal sufficiency using a highly specific cortisol immunoassay. Clinical biochemistry 50(9): 475-480	- Study does not contain any relevant index tests Comparing different assays
Kumar, Rajeev; Carr, Peter; Wassif, Ws (2022) Diagnostic performance of morning serum cortisol as an alternative to short Synacthen test for the assessment of adrenal reserve; a retrospective study. Postgraduate medical journal 98(1156): 113-118	- Retrospective
Langelaan, M.L.P., Kisters, J.M.H., Oosterwerff, M.M. et al. (2018) Salivary cortisol in the diagnosis of adrenal insufficiency: Cost efficient and patient friendly. Endocrine Connections 7(4): 560-566	- Retrospective
Laureti, S, Arvat, E, Candeloro, P et al. (2000) Low dose (1 microg) ACTH test in the evaluation of adrenal dysfunction in pre-clinical Addison's disease. Clinical endocrinology 53(1): 107-15	- Study does not contain any relevant index tests - Study design not relevant to this review protocol
Lee, May-Tze, Won, Justin Ging-Shing, Lee, Ting-I et al. (2002) The relationship between morning serum cortisol and the short ACTH test in the evaluation of adrenal insufficiency. Zhonghua yi xue za zhi = Chinese medical journal; Free China ed 65(12): 580-7	- Retrospective
Liu, Meng-Si, Lou, Yuan, Chen, Huan et al. (2022) Performance of DHEAS as a Screening	- Study does not contain any relevant index tests DHEAS

Study	Exclusion reason
Test for Autonomous Cortisol Secretion in Adrenal Incidentalomas: A Prospective Study. The Journal of clinical endocrinology and metabolism 107(5): e1789-e1796	- Study does not contain a relevant reference standard Dexamethasone suppression test
Lomenick, Jefferson P and Smith, W Jackson (2007) Low-dose adrenocorticotropic hormone stimulation testing in term infants. Journal of pediatric endocrinology & metabolism : JPEM 20(7): 773-9	- Retrospective
Mackenzie, S.D. and Gibb, F.W. (2016) Identification and validation of criteria for the use of random serum cortisol as a screening test for adrenal insufficiency. Endocrine Reviews 37(2supplement1)	- Conference abstract
Mackenzie, Scott D, Gifford, Robert M, Boyle, Luke D et al. (2019) Validated criteria for the interpretation of a single measurement of serum cortisol in the investigation of suspected adrenal insufficiency. Clinical endocrinology 91(5): 608-615	- Retrospective
Maguire, Ann M, Biesheuvel, Cornelis J, Ambler, Geoffrey R et al. (2008) Evaluation of adrenal function using the human corticotrophin-releasing hormone test, low dose Synacthen test and 9am cortisol level in children and adolescents with central adrenal insufficiency. Clinical endocrinology 68(5): 683-91	- Study does not contain a relevant reference standard plasma ACTH test
Manosroi, Worapaka, Atthakomol, Pichitchai, Buranapin, Supawan et al. (2020) 30-Minute Delta Cortisol Post-ACTH Stimulation Test and Proposed Cut-Off Levels for Adrenal Insufficiency Diagnosis. The journal of medical investigation : JMI 67(12): 95-101	- Retrospective - Study does not contain any relevant index tests
Manosroi, Worapaka, Phimphilai, Mattabhorn, Khorana, Jiraporn et al. (2019) Diagnostic performance of basal cortisol level at 0900-1300h in adrenal insufficiency. PloS one 14(11): e0225255	- Retrospective
Mansoor, S, Islam, N, Siddiqui, I et al. (2007) Sixty-minute post-Synacthen serum cortisol level: a reliable and cost-effective screening test for excluding adrenal insufficiency compared to the conventional short Synacthen test. Singapore medical journal 48(6): 519-23	- Study does not contain diagnostic accuracy data
Mathara Diddhenipothage, Shani A D, Beck, Katharina J, Loo, Helen et al. (2023) "A morning cortisol is the most effective clinical predictor of short Synacthen test outcome": A tertiary care centre experience. Clinical endocrinology 99(2): 142-151	- Exclude- retrospective
Montes-Villarreal, Juan, Perez-Arredondo, Luis Alberto, Rodriguez-Gutierrez, Rene et al. (2020) SERUM MORNING CORTISOL AS A SCREENING TEST FOR ADRENAL INSUFFICIENCY. Endocrine practice : official journal of the American College of	- Retrospective

Study	Exclusion reason
Endocrinology and the American Association of Clinical Endocrinologists 26(1): 30-35	
Munro, Vicki, Elnenaei, Manal, Doucette, Steve et al. (2018) The effect of time of day testing and utility of 30 and 60minute cortisol values in the 250mcg ACTH stimulation test. Clinical biochemistry 54: 37-41	- Retrospective - Study does not contain any relevant index tests
Nakhleh, Afif, Saiegh, Leonard, Shehadeh, Naim et al. (2022) Screening for non-classic congenital adrenal hyperplasia in women: New insights using different immunoassays. Frontiers in endocrinology 13: 1048663	- Study does not contain any relevant index tests 17-OHP - Study does not contain a relevant reference standard Diagnosis of non-classic congenital adrenal hyperplasia
Ng, Sze May; Agwu, Juliana Chizo; Dwan, Kerry (2016) A systematic review and meta-analysis of Synacthen tests for assessing hypothalamic-pituitary-adrenal insufficiency in children. Archives of disease in childhood 101(9): 847-53	- Systematic review used as source of primary studies
O'Grady, Michael J, Hensey, Conor, Fallon, Miriam et al. (2013) Lack of sensitivity of the 1-mug low-dose ACTH stimulation test in a paediatric population with suboptimal cortisol responses to insulin-induced hypoglycaemia. Clinical endocrinology 78(1): 73-8	- Retrospective
Ortiz-Flores, Andres E, Santacruz, Elisa, Jimenez-Mendiguchia, Lucia et al. (2018) Role of sampling times and serum cortisol cut-off concentrations on the routine assessment of adrenal function using the standard cosyntropin test in an academic hospital from Spain: a retrospective chart review. BMJ open 8(5): e019273	- Retrospective
Ospina, Naykky Singh, Al Nofal, Alaa, Bancos, Irina et al. (2016) ACTH Stimulation Tests for the Diagnosis of Adrenal Insufficiency: Systematic Review and Meta-Analysis. The Journal of clinical endocrinology and metabolism 101(2): 427-34	- Study does not contain any relevant index tests ACTH stimulation test as the index test
Panamonta, O., Kirdpon, W., Sungsahachart, D. et al. (2003) Adrenocorticotropin stimulation test in congenital adrenal hyperplasia: Comparison between standard and low dose test. Journal of the Medical Association of Thailand 86(7): 634-640	- Population not relevant to this review protocol
Papierska, Lucyna, Rabijewski, Michal, Migda, Bartosz et al. (2022) Evaluation of plasma ACTH in the metyrapone test is insufficient for the diagnosis of secondary adrenal insufficiency. Frontiers in endocrinology 13: 1004129	- Study does not contain any relevant index tests
Patel, R S, Wallace, A M, Hinnie, J et al. (2001) Preliminary results of a pilot study investigating the potential of salivary cortisol measurements to detect occult adrenal suppression secondary to steroid nose drops. Clinical otolaryngology and allied sciences 26(3): 231-4	- Study does not contain diagnostic accuracy data

Study	Exclusion reason
Patel, Rajan S, Shaw, Steve R, McIntyre, Halena E et al. (2004) Morning salivary cortisol versus short Synacthen test as a test of adrenal suppression. Annals of clinical biochemistry 41(pt5): 408-10	- Study does not contain diagnostic accuracy data
Perogamvros, Ilias, Owen, Laura J, Keevil, Brian G et al. (2010) Measurement of salivary cortisol with liquid chromatography-tandem mass spectrometry in patients undergoing dynamic endocrine testing. Clinical endocrinology 72(1): 17-21	- Population not relevant to this review protocol critically ill - end stage renal disease
Perton, F T, Mijnhout, G S, Kollen, B J et al. (2017) Validation of the 1 mug short Synacthen test: an assessment of morning cortisol cut-off values and other predictors. The Netherlands journal of medicine 75(1): 14-20	- Retrospective
Ramadoss, Vijay, Lazarus, Katharine, Prevost, Andrew Toby et al. (2021) Improving the Interpretation of Afternoon Cortisol Levels and SSTs to Prevent Misdiagnosis of Adrenal Insufficiency. Journal of the Endocrine Society 5(11): bvab147	- Retrospective
Rose, S R, Lustig, R H, Burstein, S et al. (1999) Diagnosis of ACTH deficiency. Comparison of overnight metyrapone test to either low-dose or high-dose ACTH test. Hormone research 52(2): 73-9	- Study does not contain a relevant reference standard Metyrapone test
Sbardella, E., Isidori, A.M., Woods, C.P. et al. (2017) Baseline morning cortisol level as a predictor of pituitary-adrenal reserve: a comparison across three assays. Clinical Endocrinology 86(2): 177-184	- Retrospective
Schindhelm, R K; van de Leur, J J C M; Rondeel, J M M (2010) Salivary cortisol as an alternative for serum cortisol in the low-dose adrenocorticotropic hormone stimulation test?. Journal of endocrinological investigation 33(2): 92-5	- Study does not contain any relevant index tests
Smolyar, D, Tirado-Bernardini, R, Landman, R et al. (2003) Comparison of 1-micro g and 250-micro g corticotropin stimulation tests for the evaluation of adrenal function in patients with acquired immunodeficiency syndrome. Metabolism: clinical and experimental 52(5): 647-51	- Study does not contain any relevant index tests
Steiner, H, Bahr, V, Exner, P et al. (1994) Pituitary function tests: comparison of ACTH and 11-deoxy-cortisol responses in the metyrapone test and with the insulin hypoglycemia test. Experimental and clinical endocrinology 102(1): 33-8	- Retrospective - Population not relevant to this review protocol
Struja, Tristan, Briner, Leonie, Meier, Aline et al. (2017) DIAGNOSTIC ACCURACY OF BASAL CORTISOL LEVEL TO PREDICT ADRENAL INSUFFICIENCY IN COSYNTROPIN TESTING: RESULTS FROM AN OBSERVATIONAL COHORT STUDY WITH 804 PATIENTS.	- Retrospective

Study	Exclusion reason
Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 23(8): 949-961	
Suliman, Abdulwahab M, Smith, Thomas P, Labib, Mourad et al. (2002) The low-dose ACTH test does not provide a useful assessment of the hypothalamic-pituitary-adrenal axis in secondary adrenal insufficiency. Clinical endocrinology 56(4): 533-9	- Study does not contain a relevant reference standard overnight metyrapone test
Thaler, L M and Blevins, L S Jr (1998) The low dose (1-microg) adrenocorticotropin stimulation test in the evaluation of patients with suspected central adrenal insufficiency. The Journal of clinical endocrinology and metabolism 83(8): 2726-9	- Review article but not a systematic review
Tolkin, Lior; Vidberg, Michal; Munter, Gabriel (2022) Basal serum cortisol levels predict a normal response to the Synacthen stimulation test in hospitalised patients. Internal medicine journal 52(1): 105-109	- Retrospective
Toru, B.K.; Canat, M.M.; Altuntas, Y. (2023) New Cut-off Value for Low-Dose Acth Stimulation Test in the Diagnosis of Adrenal Insufficiency. Haseki Tip Bulteni 61(3): 146-153	- Exclude- retrospective
Ueland, Grethe A, Methlie, Paal, Oksnes, Marianne et al. (2018) The Short Cosyntropin Test Revisited: New Normal Reference Range Using LC-MS/MS. The Journal of clinical endocrinology and metabolism 103(4): 1696-1703	- Study does not contain any relevant index tests comparing different assays of the same test
Ulhaq, Imran, Ahmad, Tauseef, Khoja, Adeel et al. (2019) Morning cortisol as an alternative to Short Synecthan test for the diagnosis of primary adrenal insufficiency. Pakistan journal of medical sciences 35(5): 1413-1416	- Retrospective
Vaiani, Elisa, Lazzati, Juan Manuel, Ramirez, Pablo et al. (2019) The Low-Dose ACTH Test: Usefulness of Combined Analysis of Serum and Salivary Maximum Cortisol Response in Pediatrics. The Journal of clinical endocrinology and metabolism 104(10): 4323-4330	- Study does not contain any relevant index tests Does not report basal cortisol values.
Vaiani, Elisa, Maceiras, Mercedes, Chaler, Eduardo et al. (2014) Central adrenal insufficiency could not be confirmed by measurement of basal serum DHEAS levels in pubertal children. Hormone research in paediatrics 82(5): 332-7	- Study does not contain any relevant index tests
Weintrob, N, Sprecher, E, Josefsberg, Z et al. (1998) Standard and low-dose short adrenocorticotropin test compared with insulin-induced hypoglycemia for assessment of the hypothalamic-pituitary-adrenal axis in children with idiopathic multiple pituitary hormone deficiencies. The Journal of clinical endocrinology and metabolism 83(1): 88-92	- Study does not contain any relevant index tests

Study	Exclusion reason
Yalovitsky, Guy, Shaki, David, Hershkovitz, Eli et al. (2023) Comparison of glucagon stimulation test and low dose ACTH test in assessing hypothalamic-pituitary-adrenal (HPA) axis in children. Clinical endocrinology 98(5): 678-681	- Retrospective - Study does not contain any relevant index tests
Younas, Alveena, Ali, Asif, Nawaz, Muhammad Asif et al. (2019) Comparative evaluation of 30 and 60 minutes cortisol levels during short Synacthen test for diagnosis of adrenal insufficiency. JPMA. The Journal of the Pakistan Medical Association 69(11): 1628-1631	- Article could not be accessed
Zarkovic, M, Ciric, J, Stojanovic, M et al. (1999) Optimizing the diagnostic criteria for standard (250-microg) and low dose (1-microg) adrenocorticotropin tests in the assessment of adrenal function. The Journal of clinical endocrinology and metabolism 84(9): 3170-3	- Population not relevant to this review protocol Includes those with AI and controls in the accuracy analysis.
Zha, Li, Li, Jieli, Krishnan, Subhashree Mallika et al. (2022) New Diagnostic Cutoffs for Adrenal Insufficiency After Cosyntropin Stimulation Using Abbott Architect Cortisol Immunoassay. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 28(7): 684-689	- Study does not contain any relevant index tests
Zollner, Ekkehard W, Lombard, Carl, Galal, Ushma et al. (2011) Hypothalamic-pituitary-adrenal axis suppression in asthmatic children on inhaled and nasal corticosteroids: is the early-morning serum adrenocorticotropic hormone (ACTH) a useful screening test?. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology 22(6): 614-20	- Study does not contain a relevant reference standard Post-metyrapone test
Zueger, Thomas, Jordi, Marlen, Laimer, Markus et al. (2014) Utility of 30 and 60 minute cortisol samples after high-dose synthetic ACTH-1-24 injection in the diagnosis of adrenal insufficiency. Swiss medical weekly 144: w13987	- Study does not contain diagnostic accuracy data

I.2 Clinical studies- diagnostic threshold for referral

Table 21: Studies excluded from the clinical review

Study	Exclusion reason
Abdu, T A, Elhadd, T A, Neary, R et al. (1999) Comparison of the low dose short Synacthen test (1 microg), the conventional dose short Synacthen test (250 microg), and the insulin tolerance test for assessment of the hypothalamo-pituitary-adrenal axis in patients with pituitary disease. The Journal of clinical endocrinology and metabolism 84(3): 838-43	- Study does not contain any relevant index tests
Abdu, T.A.M. and Clayton, R.N. (2000) The low-dose Synacthen test for the assessment of	- Review article but not a systematic review

Study	Exclusion reason
secondary adrenal insufficiency. Current Opinion in Endocrinology and Diabetes 7(3): 116-121	
Abeed, N.N.A.N., Mohamad, W.M.I.W., Yahya, N. et al. (2022) ACCURACY OF RANDOM SERUM CORTISOL IN DIAGNOSING SECONDARY ADRENAL INSUFFICIENCY. Journal of the ASEAN Federation of Endocrine Societies 37(supplement2): 12	- Conference abstract
Abraham, Smita Baid, Abel, Brent S, Sinaii, Ninet et al. (2015) Primary vs secondary adrenal insufficiency: ACTH-stimulated aldosterone diagnostic cut-off values by tandem mass spectrometry. Clinical endocrinology 83(3): 308-14	- Study does not contain a relevant reference standard Plasma ACTH stimulation test - Population not relevant to this review protocol People with known AI
Agha, A., Tomlinson, J.W., Clark, P.M. et al. (2006) The long-term predictive accuracy of the short Synacthen (corticotropin) stimulation test for assessment of the hypothalamic-pituitary-adrenal axis. Journal of Clinical Endocrinology and Metabolism 91(1): 43-47	- Retrospective - Study does not contain diagnostic accuracy data
Albert, L., Profitos, J., Sanchez-Delgado, J. et al. (2019) Salivary cortisol determination in ACTH stimulation test to diagnose adrenal insufficiency in patients with liver cirrhosis. International Journal of Endocrinology 2019: 7251010	- Population not relevant to this review protocol Known AI and healthy controls.
Ambrogio, Alberto G, Danesi, Leila, Baldini, Marina et al. (2018) Low-dose Synachten test with measurement of salivary cortisol in adult patients with beta-thalassemia major. Endocrine 60(2): 348-354	- Study does not report sensitivity or specificity Youden's index only
Ambrosi, B, Barbetta, L, Re, T et al. (1998) The one microgram adrenocorticotropin test in the assessment of hypothalamic-pituitary-adrenal function. European journal of endocrinology 139(6): 575-9	- Study does not contain any relevant index tests
Amin, H., Wynne-Edwards, K., Amin, P. et al. (2017) Is the Correlation between Salivary Cortisol and Serum Cortisol Reliable Enough to Enable Use of Salivary Cortisol Levels in Preterm Infants?. American Journal of Perinatology 34(13): 1302-1305	- Study does not contain diagnostic accuracy data correlation only
Arregger, Alejandro L, Cardoso, Estela M L, Tumilasci, Omar et al. (2008) Diagnostic value of salivary cortisol in end stage renal disease. Steroids 73(1): 77-82	- Population not relevant to this review protocol critically ill - end stage renal disease
Arregger, Alejandro L, Cardoso, Estela M L, Zucchini, Alfredo et al. (2014) Adrenocortical function in hypotensive patients with end stage renal disease. Steroids 84: 57-63	- Population not relevant to this review protocol End stage renal disease
Atluri, Sridevi, Sarathi, Vijaya, Goel, Amit et al. (2022) Long-acting Porcine Sequence ACTH (Acton Prolongatum) Stimulation Test is a Reliable Alternative Test as Compared to the Gold Standard Insulin Tolerance Test for the Diagnosis of Adrenal Insufficiency. Indian journal of endocrinology and metabolism 26(1): 38-43	- Population not relevant to this review protocol Includes some with known AI - Study design not relevant to this review protocol Recruitment method not clear

Study	Exclusion reason
Bancos, Irina, Erickson, Dana, Bryant, Sandra et al. (2015) PERFORMANCE OF FREE VERSUS TOTAL CORTISOL FOLLOWING COSYNTROPIN STIMULATION TESTING IN AN OUTPATIENT SETTING. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 21(12): 1353-63	- Study does not report sensitivity or specificity AUC only
Bangar, V. and Clayton, R.N. (1998) How reliable is the short Synacthen test for the investigation of the hypothalamic-pituitary-adrenal axis?. European Journal of Endocrinology 139(6): 580-583	- Retrospective
Butt, Muhammad Imran, Alzuhayri, Nouf, Amer, Lama et al. (2020) Comparing the utility of 30- and 60-minute cortisol levels after the standard short Synacthen test to determine adrenal insufficiency: A retrospective cross-sectional study. Medicine 99(43): e22621	- Retrospective
Ceccato, Filippo, Selmin, Elisa, Antonelli, Giorgia et al. (2021) Low-dose short Synacthen test with salivary cortisol in patients with suspected central adrenal insufficiency. Endocrine connections 10(9): 1189-1199	- Retrospective
Cemeroglu, Ayse Pinar, Kleis, Lora, Postellon, Daniel C et al. (2011) Comparison of low-dose and high-dose cosyntropin stimulation testing in children. Pediatrics international : official journal of the Japan Pediatric Society 53(2): 175-80	- Study does not contain diagnostic accuracy data
Cetinkaya, Semra; Ozon, Alev; Yordam, Nursen (2007) Diagnostic value of salivary cortisol in children with abnormal adrenal cortex functions. Hormone research 67(6): 301-6	- Study does not contain any relevant index tests
Cheung, K.K.-T., So, W.-Y., Ma, R.C.-W. et al. (2015) Spot and morning cortisol in comparison to low dose short Synacthen test. Journal of the ASEAN Federation of Endocrine Societies 30(2): 147-154	- Data not reported in an extractable format or a format that can be analysed Text discussion confuses the interpretation of sensitivity and specificity for ruling in or out the condition; and threshold values for serum cortisol appear incorrect with low values linked to maximum sensitivity and high values to maximum specificity.
Chitale, Aditi, Musonda, Patrick, McGregor, Alan M et al. (2013) Determining the utility of the 60 min cortisol measurement in the short Synacthen test. Clinical endocrinology 79(1): 14-9	- Retrospective
Chng, E., Lam, S., Hawkins, R. et al. (2014) The use of short Synacthen test in patients on exogenous steroids use in diagnosing adrenal insufficiency. Endocrine Reviews 35(suppl3)	- Conference abstract
Cho, Hwa Y, Kim, Jung H, Kim, Sang W et al. (2014) Different cut-off values of the insulin tolerance test, the high-dose short Synacthen test (250 mug) and the low-dose short Synacthen test (1 mug) in assessing central adrenal insufficiency. Clinical endocrinology 81(1): 77-84	- Study does not contain any relevant index tests

Study	Exclusion reason
Colling, Caitlin, Nachtigall, Lisa, Biller, Beverly M K et al. (2022) The biochemical diagnosis of adrenal insufficiency with modern cortisol assays: Reappraisal in the setting of opioid exposure and hospitalization. Clinical endocrinology 96(1): 21-29	- Retrospective
Contreras, L.N., Arregger, A.L., Tumilasci, O. et al. (2006) Salivary steroids in response to ACTH: A less invasive approach to assess adrenal function in hypotensive patients with chronic renal failure. Endocrinologist 16(1): 30-35	- Study does not contain diagnostic accuracy data
Cornes, Michael P, Ashby, Helen L, Khalid, Yasmeen et al. (2015) Salivary cortisol and cortisone responses to tetracosactrin (Synacthen). Annals of clinical biochemistry 52(pt5): 606-10	- Study does not contain diagnostic accuracy data correlation only
de Vries, Friso, Lobatto, Daniel J, Bakker, Leontine E H et al. (2020) Early postoperative HPA-axis testing after pituitary tumor surgery: reliability and safety of basal cortisol and CRH test. Endocrine 67(1): 161-171	- Retrospective
Dichtel, L.E., Schorr, M., De Assis, C.L. et al. (2017) Plasma free cortisol vs. Total cortisol in healthy individuals and in states of high and low cortisol binding globulin, including oral contraceptive use, cirrhosis and critical illness: implications for diagnosing adrenal insufficiency. Endocrine Reviews 38(3supplement1)	- Conference abstract
Dineen, Rosemary, Mohamed, Ahmed, Gunness, Anjuli et al. (2020) Outcomes of the short Synacthen test: what is the role of the 60 min sample in clinical practice?. Postgraduate medical journal 96(1132): 67-72	- Retrospective
Dluhy, R.G.; Himathongkam, T.; Greenfield, M. (1974) Rapid ACTH test with plasma aldosterone levels. Improved diagnostic discrimination. Annals of Internal Medicine 80(6): 693-696	- Study does not contain a relevant reference standard plasma ACTH test
Dorin, Richard I; Qualls, Clifford R; Crapo, Lawrence M (2003) Diagnosis of adrenal insufficiency. Annals of internal medicine 139(3): 194-204	- Review article but not a systematic review
Fede, Giuseppe, Spadaro, Luisa, Privitera, Graziella et al. (2015) Hypothalamus-pituitary dysfunction is common in patients with stable cirrhosis and abnormal low dose Synacthen test. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver 47(12): 1047-51	- Population not relevant to this review protocol selected those with AI.
Ferrante, Emanuele, Morelli, Valentina, Giavoli, Claudia et al. (2012) Is the 250 mug ACTH test a useful tool for the diagnosis of central hypoadrenalism in adult patients with pituitary disorders?. Hormones (Athens, Greece) 11(4): 428-35	- Study design not relevant to this review protocol Cohort study with 6 weeks between index and reference standard test

Study	Exclusion reason
Fragoso Perozo, A F D, Fontes, R, Lopes, F P et al. (2023) Morning serum cortisol role in the adrenal insufficiency diagnosis with modern cortisol assays. Journal of endocrinological investigation	- Retrospective
Gasco, Valentina, Bima, Chiara, Geranzani, Alice et al. (2021) Morning Serum Cortisol Level Predicts Central Adrenal Insufficiency Diagnosed by Insulin Tolerance Test. Neuroendocrinology 111(12): 1238-1248	- Retrospective
Giordano, R, Picu, A, Bonelli, L et al. (2008) Hypothalamus-pituitary-adrenal axis evaluation in patients with hypothalamo-pituitary disorders: comparison of different provocative tests. Clinical endocrinology 68(6): 935-41	- Study does not contain any relevant index tests
Gleeson, H.K., Walker, B.R., Seckl, J.R. et al. (2003) Ten years on: Safety of short Synacthen tests in assessing adrenocorticotropin deficiency in clinical practice. Journal of Clinical Endocrinology and Metabolism 88(5): 2106-2111	- Retrospective
Goggans, F C, Wilson, W R Jr, Gold, M S et al. (1984) Comparison of the dexamethasone suppression test and the cortisol suppression index. The American journal of psychiatry 141(5): 698-700	- Study does not contain any relevant index tests - Population not relevant to this review protocol
Gonc, E Nazli; Kandemir, Nurgun; Kinik, Sibel T (2003) Significance of low-dose and standard-dose ACTH tests compared to overnight metyrapone test in the diagnosis of adrenal insufficiency in childhood. Hormone research 60(4): 191-7	- Study does not contain a relevant reference standard metyrapone test
Gonc, E Nazli, Ozon, Z Alev, Alikasifoglu, Ayfer et al. (2011) Is basal serum 17-OH progesterone a reliable parameter to predict nonclassical congenital adrenal hyperplasia in premature adrenarche?. The Turkish journal of pediatrics 53(3): 274-80	- Retrospective
Goto, Masahiro; Shibata, Nao; Hasegawa, Yukihiro (2016) Efficacy of single serum cortisol reading obtained between 9 AM and 10 AM as an index of adrenal function in children treated with glucocorticoids or synthetic adrenocorticotropic hormone. Clinical pediatric endocrinology : case reports and clinical investigations : official journal of the Japanese Society for Pediatric Endocrinology 25(3): 83-9	- Retrospective
Grassi, G, Morelli, V, Ceriotti, F et al. (2020) Minding the gap between cortisol levels measured with second-generation assays and current diagnostic thresholds for the diagnosis of adrenal insufficiency: a single-center experience. Hormones (Athens, Greece) 19(3): 425-431	- Study does not contain diagnostic accuracy data Correlation between different assays
Gruvstad, Eva, Hedner, Lars Pavo, Hoglund, Peter et al. (2014) Comparison of methods for evaluation of the suppressive effects of	- Study design not relevant to this review protocol

Study	Exclusion reason
prednisolone on the HPA axis and bone turnover: changes in s-DHEAS are as sensitive as the ACTH test. International journal of clinical pharmacology and therapeutics 52(1): 15-26	- Population not relevant to this review protocol
Gundgurthi, Abhay, Garg, M K, Dutta, M K et al. (2013) Intramuscular ACTH stimulation test for assessment of adrenal function. The Journal of the Association of Physicians of India 61(5): 320-4	- Study does not contain diagnostic accuracy data. insufficient information to calculate accuracy data.
Hassan, Z., Nabi, S., Hussain, W. et al. (2021) Validation of Glucagon Stimulation Test in Establishing GH and ACTH Deficiency in Hypopituitarism. European Journal of Molecular and Clinical Medicine 8(4): 2005-2013	- Study design not relevant to this review protocol diagnostic accuracy for GH deficiency - Study does not contain any relevant index tests glucagon stimulation test
Javorsky, Bradley R, Raff, Hershel, Carroll, Ty B et al. (2021) New Cutoffs for the Biochemical Diagnosis of Adrenal Insufficiency after ACTH Stimulation using Specific Cortisol Assays. Journal of the Endocrine Society 5(4): bvab022	- Retrospective
Jayakumari, C., George, G.S., Nair, A. et al. (2017) ACTH stimulation test with long acting ACTH preparation for the diagnosis of adrenal insufficiency. Indian Journal of Endocrinology and Metabolism 21(8supplement1): 62	- Conference abstract
Kadiyala, R, Kamath, C, Baglioni, P et al. (2010) Can a random serum cortisol reduce the need for short Synacthen tests in acute medical admissions?. Annals of clinical biochemistry 47(pt4): 378-80	- Retrospective
Kalaria, R.T., Agarwal, M., Kaur, S. et al. (2020) ANNALS EXPRESS: Hypothalamic-pituitary-adrenal (HPA) axis suppression a The value of salivary cortisol and cortisone in assessing HPA recovery. Annals of clinical biochemistry: 4563220961745	- Duplicate reference
Kamrath, Clemens and Boehles, Hansjosef (2010) The low-dose ACTH test does not identify mild insufficiency of the hypothalamnic-pituitary-adrenal axis in children with inadequate stress response. Journal of pediatric endocrinology & metabolism : JPEM 23(11): 1097-104	- Retrospective
Karpman, Matthew S, Neculau, Madalina, Dias, Valerian C et al. (2013) Defining adrenal status with salivary cortisol by gold-standard insulin hypoglycemia. Clinical biochemistry 46(15): 1442-6	- Study does not contain any relevant index tests
Kazlauskaite, Rasa, Evans, Arthur T, Villabona, Carmen V et al. (2008) Corticotropin tests for hypothalamic-pituitary- adrenal insufficiency: a metaanalysis. The Journal of clinical endocrinology and metabolism 93(11): 4245-53	- Systematic review used as source of primary studies
Kline, G A; Buse, J; Krause, R D (2017) Clinical implications for biochemical diagnostic thresholds of adrenal sufficiency using a highly	- Study does not contain any relevant index tests Comparing different assays

Study	Exclusion reason
specific cortisol immunoassay. Clinical biochemistry 50(9): 475-480	
Kumar, Rajeev; Carr, Peter; Wassif, Ws (2022) Diagnostic performance of morning serum cortisol as an alternative to short Synacthen test for the assessment of adrenal reserve; a retrospective study. Postgraduate medical journal 98(1156): 113-118	- Retrospective
Langelaan, M.L.P., Kisters, J.M.H., Oosterwerff, M.M. et al. (2018) Salivary cortisol in the diagnosis of adrenal insufficiency: Cost efficient and patient friendly. Endocrine Connections 7(4): 560-566	- Retrospective
Laureti, S, Arvat, E, Candeloro, P et al. (2000) Low dose (1 microg) ACTH test in the evaluation of adrenal dysfunction in pre-clinical Addison's disease. Clinical endocrinology 53(1): 107-15	- Study does not contain any relevant index tests - Study design not relevant to this review protocol
Lee, May-Tze, Won, Justin Ging-Shing, Lee, Ting-I et al. (2002) The relationship between morning serum cortisol and the short ACTH test in the evaluation of adrenal insufficiency. Zhonghua yi xue za zhi = Chinese medical journal; Free China ed 65(12): 580-7	- Retrospective
Liu, Meng-Si, Lou, Yuan, Chen, Huan et al. (2022) Performance of DHEAS as a Screening Test for Autonomous Cortisol Secretion in Adrenal Incidentalomas: A Prospective Study. The Journal of clinical endocrinology and metabolism 107(5): e1789-e1796	- Study does not contain any relevant index tests DHEAS - Study does not contain a relevant reference standard Dexamethasone suppression test
Lomenick, Jefferson P and Smith, W Jackson (2007) Low-dose adrenocorticotropic hormone stimulation testing in term infants. Journal of pediatric endocrinology & metabolism : JPEM 20(7): 773-9	- Retrospective
Mackenzie, S.D. and Gibb, F.W. (2016) Identification and validation of criteria for the use of random serum cortisol as a screening test for adrenal insufficiency. Endocrine Reviews 37(2supplement1)	- Conference abstract
Mackenzie, Scott D, Gifford, Robert M, Boyle, Luke D et al. (2019) Validated criteria for the interpretation of a single measurement of serum cortisol in the investigation of suspected adrenal insufficiency. Clinical endocrinology 91(5): 608-615	- Retrospective
Maguire, Ann M, Biesheuvel, Cornelis J, Ambler, Geoffrey R et al. (2008) Evaluation of adrenal function using the human corticotrophin-releasing hormone test, low dose Synacthen test and 9am cortisol level in children and adolescents with central adrenal insufficiency. Clinical endocrinology 68(5): 683-91	- Study does not contain a relevant reference standard plasma ACTH test
Manosroi, Worapaka, Atthakomol, Pichitchai, Buranapin, Supawan et al. (2020) 30-Minute Delta Cortisol Post-ACTH Stimulation Test and Proposed Cut-Off Levels for Adrenal	- Retrospective - Study does not contain any relevant index tests

Study	Exclusion reason
Insufficiency Diagnosis. The journal of medical investigation : JMI 67(12): 95-101	
Manosroi, Worapaka, Phimphilai, Mattabhorn, Khorana, Jiraporn et al. (2019) Diagnostic performance of basal cortisol level at 0900-1300h in adrenal insufficiency. PloS one 14(11): e0225255	- Retrospective
Mansoor, S, Islam, N, Siddiqui, I et al. (2007) Sixty-minute post-Synacthen serum cortisol level: a reliable and cost-effective screening test for excluding adrenal insufficiency compared to the conventional short Synacthen test. Singapore medical journal 48(6): 519-23	- Study does not contain diagnostic accuracy data
Montes-Villarreal, Juan, Perez-Arredondo, Luis Alberto, Rodriguez-Gutierrez, Rene et al. (2020) SERUM MORNING CORTISOL AS A SCREENING TEST FOR ADRENAL INSUFFICIENCY. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 26(1): 30-35	- Retrospective
Munro, Vicki, Elnenaei, Manal, Doucette, Steve et al. (2018) The effect of time of day testing and utility of 30 and 60minute cortisol values in the 250mcg ACTH stimulation test. Clinical biochemistry 54: 37-41	- Retrospective - Study does not contain any relevant index tests
Nakhleh, Afif, Saiegh, Leonard, Shehadeh, Naim et al. (2022) Screening for non-classic congenital adrenal hyperplasia in women: New insights using different immunoassays. Frontiers in endocrinology 13: 1048663	- Study does not contain any relevant index tests 17-OHP - Study does not contain a relevant reference standard Diagnosis of non-classic congenital adrenal hyperplasia
Ng, Sze May; Agwu, Juliana Chizo; Dwan, Kerry (2016) A systematic review and meta-analysis of Synacthen tests for assessing hypothalamic-pituitary-adrenal insufficiency in children. Archives of disease in childhood 101(9): 847-53	- Systematic review used as source of primary studies
O'Grady, Michael J, Hensey, Conor, Fallon, Miriam et al. (2013) Lack of sensitivity of the 1-mug low-dose ACTH stimulation test in a paediatric population with suboptimal cortisol responses to insulin-induced hypoglycaemia. Clinical endocrinology 78(1): 73-8	- Retrospective
Ortiz-Flores, Andres E, Santacruz, Elisa, Jimenez-Mendiguchia, Lucia et al. (2018) Role of sampling times and serum cortisol cut-off concentrations on the routine assessment of adrenal function using the standard cosyntropin test in an academic hospital from Spain: a retrospective chart review. BMJ open 8(5): e019273	- Retrospective
Ospina, Naykky Singh, Al Nofal, Alaa, Bancos, Irina et al. (2016) ACTH Stimulation Tests for the Diagnosis of Adrenal Insufficiency: Systematic Review and Meta-Analysis. The	- Study does not contain any relevant index tests ACTH stimulation test as the index test

Study	Exclusion reason
Journal of clinical endocrinology and metabolism 101(2): 427-34	
Panamonta, O., Kirdpon, W., Sungsahachart, D. et al. (2003) Adrenocorticotropin stimulation test in congenital adrenal hyperplasia: Comparison between standard and low dose test. Journal of the Medical Association of Thailand 86(7): 634-640	- Population not relevant to this review protocol
Papierska, Lucyna, Rabijewski, Michal, Migda, Bartosz et al. (2022) Evaluation of plasma ACTH in the metyrapone test is insufficient for the diagnosis of secondary adrenal insufficiency. Frontiers in endocrinology 13: 1004129	- Study does not contain any relevant index tests
Patel, R S, Wallace, A M, Hinnie, J et al. (2001) Preliminary results of a pilot study investigating the potential of salivary cortisol measurements to detect occult adrenal suppression secondary to steroid nose drops. Clinical otolaryngology and allied sciences 26(3): 231-4	- Study does not contain diagnostic accuracy data
Patel, Rajan S, Shaw, Steve R, McIntyre, Halena E et al. (2004) Morning salivary cortisol versus short Synacthen test as a test of adrenal suppression. Annals of clinical biochemistry 41(pt5): 408-10	- Study does not contain diagnostic accuracy data
Perogamvros, Ilias, Owen, Laura J, Keevil, Brian G et al. (2010) Measurement of salivary cortisol with liquid chromatography-tandem mass spectrometry in patients undergoing dynamic endocrine testing. Clinical endocrinology 72(1): 17-21	- Population not relevant to this review protocol critically ill - end stage renal disease
Perton, F T, Mijnhout, G S, Kollen, B J et al. (2017) Validation of the 1 mug short Synacthen test: an assessment of morning cortisol cut-off values and other predictors. The Netherlands journal of medicine 75(1): 14-20	- Retrospective
Ramadoss, Vijay, Lazarus, Katharine, Prevost, Andrew Toby et al. (2021) Improving the Interpretation of Afternoon Cortisol Levels and SSTs to Prevent Misdiagnosis of Adrenal Insufficiency. Journal of the Endocrine Society 5(11): bvab147	- Retrospective
Rose, S R, Lustig, R H, Burstein, S et al. (1999) Diagnosis of ACTH deficiency. Comparison of overnight metyrapone test to either low-dose or high-dose ACTH test. Hormone research 52(2): 73-9	- Study does not contain a relevant reference standard Metyrapone test
Sbardella, E., Isidori, A.M., Woods, C.P. et al. (2017) Baseline morning cortisol level as a predictor of pituitary-adrenal reserve: a comparison across three assays. Clinical Endocrinology 86(2): 177-184	- Retrospective
Schindhelm, R K; van de Leur, J J C M; Rondeel, J M M (2010) Salivary cortisol as an alternative for serum cortisol in the low-dose adrenocorticotropic hormone stimulation test?. Journal of endocrinological investigation 33(2): 92-5	- Study does not contain any relevant index tests

Study	Exclusion reason
Smolyar, D, Tirado-Bernardini, R, Landman, R et al. (2003) Comparison of 1-micro g and 250-micro g corticotropin stimulation tests for the evaluation of adrenal function in patients with acquired immunodeficiency syndrome. Metabolism: clinical and experimental 52(5): 647-51	- Study does not contain any relevant index tests
Steiner, H, Bahr, V, Exner, P et al. (1994) Pituitary function tests: comparison of ACTH and 11-deoxy-cortisol responses in the metyrapone test and with the insulin hypoglycemia test. Experimental and clinical endocrinology 102(1): 33-8	- Retrospective - Population not relevant to this review protocol
Struja, Tristan, Briner, Leonie, Meier, Aline et al. (2017) DIAGNOSTIC ACCURACY OF BASAL CORTISOL LEVEL TO PREDICT ADRENAL INSUFFICIENCY IN COSYNTROPIN TESTING: RESULTS FROM AN OBSERVATIONAL COHORT STUDY WITH 804 PATIENTS. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 23(8): 949-961	- Retrospective
Suliman, Abdulwahab M, Smith, Thomas P, Labib, Mourad et al. (2002) The low-dose ACTH test does not provide a useful assessment of the hypothalamic-pituitary-adrenal axis in secondary adrenal insufficiency. Clinical endocrinology 56(4): 533-9	- Study does not contain a relevant reference standard overnight metyrapone test
Thaler, L M and Blevins, L S Jr (1998) The low dose (1-microg) adrenocorticotropin stimulation test in the evaluation of patients with suspected central adrenal insufficiency. The Journal of clinical endocrinology and metabolism 83(8): 2726-9	- Review article but not a systematic review
Tolkin, Lior; Vidberg, Michal; Munter, Gabriel (2022) Basal serum cortisol levels predict a normal response to the Synacthen stimulation test in hospitalised patients. Internal medicine journal 52(1): 105-109	- Retrospective
Ueland, Grethe A, Methlie, Paal, Oksnes, Marianne et al. (2018) The Short Cosyntropin Test Revisited: New Normal Reference Range Using LC-MS/MS. The Journal of clinical endocrinology and metabolism 103(4): 1696-1703	- Study does not contain any relevant index tests comparing different assays of the same test
Ulhaq, Imran, Ahmad, Tauseef, Khoja, Adeel et al. (2019) Morning cortisol as an alternative to Short Synecthan test for the diagnosis of primary adrenal insufficiency. Pakistan journal of medical sciences 35(5): 1413-1416	- Retrospective
Vaiani, Elisa, Lazzati, Juan Manuel, Ramirez, Pablo et al. (2019) The Low-Dose ACTH Test: Usefulness of Combined Analysis of Serum and Salivary Maximum Cortisol Response in Pediatrics. The Journal of clinical endocrinology and metabolism 104(10): 4323-4330	- Study does not contain any relevant index tests Does not report basal cortisol values.

Study	Exclusion reason
Vaiani, Elisa, Maceiras, Mercedes, Chaler, Eduardo et al. (2014) Central adrenal insufficiency could not be confirmed by measurement of basal serum DHEAS levels in pubertal children. Hormone research in paediatrics 82(5): 332-7	- Study does not contain any relevant index tests
Weintrob, N, Sprecher, E, Josefsberg, Z et al. (1998) Standard and low-dose short adrenocorticotropin test compared with insulin-induced hypoglycemia for assessment of the hypothalamic-pituitary-adrenal axis in children with idiopathic multiple pituitary hormone deficiencies. The Journal of clinical endocrinology and metabolism 83(1): 88-92	- Study does not contain any relevant index tests
Yalovitsky, Guy, Shaki, David, Hershkovitz, Eli et al. (2023) Comparison of glucagon stimulation test and low dose ACTH test in assessing hypothalamic-pituitary-adrenal (HPA) axis in children. Clinical endocrinology 98(5): 678-681	- Retrospective - Study does not contain any relevant index tests
Younas, Alveena, Ali, Asif, Nawaz, Muhammad Asif et al. (2019) Comparative evaluation of 30 and 60 minutes cortisol levels during short Synacthen test for diagnosis of adrenal insufficiency. JPMA. The Journal of the Pakistan Medical Association 69(11): 1628-1631	- Article could not be accessed
Zarkovic, M, Ciric, J, Stojanovic, M et al. (1999) Optimizing the diagnostic criteria for standard (250-microg) and low dose (1-microg) adrenocorticotropin tests in the assessment of adrenal function. The Journal of clinical endocrinology and metabolism 84(9): 3170-3	- Population not relevant to this review protocol Includes those with AI and controls in the accuracy analysis.
Zha, Li, Li, Jieli, Krishnan, Subhashree Mallika et al. (2022) New Diagnostic Cutoffs for Adrenal Insufficiency After Cosyntropin Stimulation Using Abbott Architect Cortisol Immunoassay. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 28(7): 684-689	- Study does not contain any relevant index tests
Zollner, Ekkehard W, Lombard, Carl, Galal, Ushma et al. (2011) Hypothalamic-pituitary-adrenal axis suppression in asthmatic children on inhaled and nasal corticosteroids: is the early-morning serum adrenocorticotropic hormone (ACTH) a useful screening test?. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology 22(6): 614-20	- Study does not contain a relevant reference standard Post-metyrapone test
Zueger, Thomas, Jordi, Marlen, Laimer, Markus et al. (2014) Utility of 30 and 60 minute cortisol samples after high-dose synthetic ACTH-1-24 injection in the diagnosis of adrenal insufficiency. Swiss medical weekly 144: w13987	- Study does not contain diagnostic accuracy data

I.3 Health Economic studies

I.3.1 Diagnostic tests

None

I.3.2 Diagnostic thresholds for referral

None.

Appendix J Recommendations for research – full details

What is the clinical and cost-effectiveness of salivary cortisone or cortisol to identify people with adrenal insufficiency?

J.1 Why this is important

J.1.1 Rationale for the recommendation for research

Importance to 'patients' or the population	*Unrecognised and/or untreated adrenal insufficiency leads inevitably to adrenal crisis and death. Currently many healthcare professionals will use a short Synacthen test as a gold-standard test to assess adrenal function. However, this is expensive as it is costed as a day case procedure, and time intensive for people undergoing assessment and staff alike. Time is about 2 hours for hospital staff and cost is in the region of £400. There are research data showing that other methods of assessing adrenal function are equally sensitive, but this has not yet been validated in large populations or with a health economic analysis. Comparative methodologies to be investigated would be 9am serum cortisol, and early morning salivary cortisol and cortisone. I addition patients on oral oestrogen would have to stop these products for 4-6 weeks to assess adrenal function by serum cortisol measurement or Synacthen test. As salivary cortisol/cortisone is free cortisol/cortisone this is not impacted by oral oestrogen effects on circulating CBG so may be a more cost effecting a safer option than stopping oral oestrogen (Rees JCEM 2023).
Relevance to NICE guidance	Whilst writing this guideline it is clear that widespread testing for adrenal function has the potential to have an impact of the health economy. There are nearly 1 million prescriptions for exogenous glucocorticoids in the NHS each year and some of these patients are at risk of adrenal insufficiency. Therefore, we need to ensure there is a cost effective and clinically safe way of assessing adrenal function in all populations at risk. The knowledge from this research would then be incorporated into next version of the guideline. These populations are described in the guideline.
Relevance to the NHS	Whilst deaths from adrenal insufficiency /adrenal crisis are rare they are totally preventable. By simplifying the cost and person power needed for testing for adrenal insufficiency this would benefit people undergoing testing. As the proposed methods include salivary measurements then there is additional benefit to people undergoing testing and staff, as patients can do sample collection in their own home, widening access and making testing acceptable to children and hard to reach populations.
National priorities	There is a National Patient Safety and Learning report (NHSE 2020) stating that all health care providers need to identify patients under their care with adrenal insufficiency. This comes under the remit of CQC to monitor and so is of importance for all health and care organisations.
Current evidence base	Currently there is new evidence looking at salivary cortisol and cortisone as a method to assess adrenal function. (DeBono NEJM Evidence 2023) Methodology here is an ambulatory early morning saliva sample taken at home which is then sent back to hospitals. There is about a 5% rate of inadequate samples sent back, and also some risk of contamination of salivary samples from blood in the mouth, or if a patient is using oral products with hydrocortisone in them. The methodology is using mass spectrometry (LC-MS/MS) which is available in only a few NHS trusts so there needs to be data to validate that this is worthwhile before endorsing

	roll out as the main method of testing for adrenal insufficiency. There is a cost to set up this service and also of buying LC-MS/MS machines for about £100,000. These data also look at 9am serum cortisol which also has good sensitivity and specificity with defined cut offs. 9am serum cortisol is relatively easy but does mean attendance at a fixed time for a blood test, and minute to minute variation owing to the pulsatile nature of cortisol secretion. The extant study was in 200 patients so a low number and selected from a secondary care population so a high-risk population. Therefore, the data set needs to be a lot larger, and include patients seen in primary care with different risks. By determining the real-world utility these tests, and their health resource use implications, we would determine which would be best use of NHS resource whilst diagnosing people with adrenal insufficiency with the highest sensitivity. Section xxx of guideline shows the limited amount of evidence and only 1 good study using these 3 ways of measuring cortisol (deBono 2023) As described above, current testing with a Synacthen test at a cost of around £400 is cost effective. Deutschbein 1993 also showed utility of salivary cortisol although in 55 patients so again more data is needed to justify expense of advocating new methodology (LC_MS/MS) across the NHS for measuring salivary cortisol /cortisone.
Equality considerations	A single blood test or home saliva testing is much easier for patients. Groups finding it hard to travel would benefit from home testing (eg. mobility issues, agoraphobia). People with learning disability may find it easier to have a single blood test rather than a day case admission- this may need a best interest meeting for those who do not have capacity to make decisions about their care. A single blood or saliva test may obviate the need for this.

J.1.2 Modified PICO table

Population	People at risk of adrenal insufficiency as described in Section 2.3 of guideline
Intervention	Salivary cortisol vs salivary cortisone vs 8-9am serum cortisol
Comparator	Between the 3 groups
Outcome	Cortisol cut offs as described by deBONO et al 2023 NEJM evidence
Study design	Cross-sectional study design
Timeframe	TBC
Additional information	None