Clinical Research Article

Improving the Interpretation of Afternoon Cortisol Levels and SSTs to Prevent Misdiagnosis of Adrenal Insufficiency

Vijay Ramadoss,[1] Katharine Lazarus,[1,2] Andrew Toby Prevost,[3] Tricia Tan,[1,2,4] Karim Meeran,[1,2] and Sirazum Choudhury[1,2,4]

[1]Department of Endocrinology, Imperial College Healthcare NHS Trust, London, UK; [2]Division of Diabetes, Endocrinology and Metabolism, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK; [3]Nightingale-Saunders Clinical Trials and Epidemiology Unit, King's College London, London, UK; and [4]Department of Clinical Biochemistry, Northwest London Pathology, London, UK

ORCiD numbers: 0000-0003-1723-0796 (A. T. Prevost); 0000-0001-5873-3432 (T. Tan); 0000-0002-7112-2756 (K. Meeran); 0000-0003-2429-005X (S. Choudhury).

Abstract

Background: Adrenal Insufficiency (AI), especially iatrogenic-AI, is a treatable cause of mortality. The difficulty in obtaining 9 AM cortisol levels means samples are taken at sub-optimal times, including a substantial proportion in the afternoon. Low afternoon cortisol levels often provoke short Synacthen tests (SSTs). It is important that this does not lead to patients misdiagnosed with AI, exposing them to the excess mortality and morbidity of inappropriate steroid replacement therapy.

Methods: This retrospective study collected 60 178 cortisol results. Medical records, including subsequent SSTs of initial cortisol results measured after midday were reviewed.

Results: Receiver operating characteristic analysis (area under the curve: 0.89) on 6531 suitable cortisol values showed that a limit of <201.5 nmol/L achieved a sensitivity and specificity of 95.6% and 72.6%, while a limit of <234 nmol/L had a sensitivity of 100% and a specificity of 59.5%. Out of 670 SSTs, 628 patients passed. Of these, 140 would have otherwise failed if only their 30-min cortisol was assessed without the 60-min value. A 30- and 60-min SST cortisol cutoff of 366.5 nmol/L and 418.5 nmol/L, respectively, can achieve a sensitivity of >95% on the Abbott analyser platform.

Conclusion: An afternoon cortisol >234 nmol/L excludes AI on Abbott analyser platforms. In patients who have an afternoon cortisol <234 nmol/L, including both 30- and 60-min SST cortisol values prevents unnecessary glucocorticoid replacement therapy in 22.3% of individuals in this study. The Abbott analyser SST cortisol cutoffs used to define AI should be 366.5 nmol/L and 418.5 nmol/L at 30 and 60 min, respectively. All patients remained well subsequently with at least 1-year longitudinal follow-up.

Adrenal insufficiency (AI) is associated with an increased mortality, principally due to cardiovascular disease, malignancy, and infection [1]. There is growing evidence that this observation is fueled by excess glucocorticoid replacement from intrinsically suboptimal replacement regimens [2-4]. It is therefore important that patients are not incorrectly labeled with AI, thereby exposing them to inappropriate lifelong glucocorticoid replacement and the accompanying excess mortality and morbidity.

AI is a differential diagnosis that is commonly considered despite its rarity in a multitude of clinical presentations, especially as affected individuals can present with nonspecific signs and symptoms such as tiredness, weight loss, hypotension, nausea, and vomiting [5]. A study of 216 patients has demonstrated the significant difficulty physicians and patients face in establishing a diagnosis [6]. Over 50% of males and 70% of females in a German population took greater than 6 months to obtain a diagnosis, with 20% being diagnosed after 5 years and two thirds receiving a diagnosis only after having consulted 3 different physicians. It is important not to mislabel those who have nonspecific features as AI, but also important to correctly diagnose the rare patient with true AI, in whom appropriate glucocorticoid replacement therapy is lifesaving. A 9 AM cortisol should be requested in the first instance. A value of greater than 336 nmol/L (on Abbott platforms, but may vary according to different analyzers) suggests that a short Synacthen test (SST) using tetracosactide acetate is not required [7].

There is often practical difficulty in obtaining 9 AM cortisol levels in patients. Demand for 9 AM phlebotomy slots can mean that some patients have to wait beyond the specified time. Blood tests are usually ordered after a physician review possibly in outpatient clinic in the afternoon, and patients may opt to be bled at the same time rather than experience the inconvenience of attending the hospital on a separate occasion the next morning. This often leads to cortisol samples being collected for analysis at inappropriate times, such as during late morning phlebotomy rounds in admissions to hospital or in the afternoon at outpatient clinic settings. While morning cortisol levels have been rigorously studied in the diagnosis of AI [8-10], there is currently no harmonized reference interval for an untimed cortisol level, and it is unlikely that one will be ever be defined. The observed tendency to order SSTs on the basis of afternoon cortisol results in the present study is likely to stem from concern about missing patients with undiagnosed AI and subsequently missing a potential adrenal crisis. Although the accepted practice of performing an early morning cortisol should be the first step and can preclude the need for a formal SST [5], many patients find it more convenient to complete the SST instead of having further blood tests on another occasion. There is no evidence that this practice has a significant impact on preventing adrenal crises and subsequent death in undiagnosed patients. Studies that stratify hospital admissions for crises consistently show increased incidences for those with primary AI over secondary AI, possibly because the former are unable to synthesize aldosterone [11].

The act of performing afternoon cortisols may be driving the increasing number of SSTs performed, when low cortisol results are returned. As the interpretation of SSTs is largely unchanged since the 1960s despite advances in assay techniques, this tendency to overinvestigate with SSTs will in turn lead to more failed SSTs and apparent increase in AI prevalence. This is especially problematic in patients with an otherwise low index of suspicion for AI in the context of equivocal afternoon cortisol results. The consequence of this is the misdiagnosis of patients with AI and the subsequent excess risk of mortality in this population.

A recent retrospective analysis of 2700 patients demonstrated that an afternoon cortisol value cutoff value of less than 250 nmol/L correlated with suboptimal SST outcomes and achieved over 96% sensitivity and 37.7% specificity in subsequent diagnosis of AI [12]. In practice, this cutoff does not account for the number of false positives, which in turn may still lead to a large number of unnecessary confirmatory SST tests. At present, at Imperial College Healthcare National Health Service (NHS) Trust, an afternoon value of less than 100 nmol/L is labeled as a "critical result" and prompts a review by our biochemistry department with the prospect of a further call-out to a clinician as per Royal College of Pathology guidelines and International Organisation for Standardization (ISO) 15189 standards [13,14]. Many of these "critical results" lead to investigation and no subsequent evidence for AI. The aim of this study was to clarify the diagnostic utility of cortisol levels taken in patients after midday.

Methods

Design

This is a retrospective study looking at cortisol levels taken from patients at Imperial College Healthcare NHS Trust. All cortisol blood test results between May 12,

2016 and February 29, 2020 were retrieved from the data repository held by Northwest London Pathology. The data were filtered for all cortisol samples collected between 12:00 PM and 11:59 PM. All related cortisol values were extracted and collated with the afternoon result to permit a fuller review. Patient's age and time of afternoon cortisol level were recorded. SST data were sought for patients with eligible afternoon cortisol levels and were subsequently analyzed. Patients with their first recorded cortisol sample taken in the morning before 12 PM, previous known diagnosis of AI, samples taken as part of a dynamic function tests (other than to assess the adrenal axis for AI), laboratory quality samples, research samples, duplicate entries and patients on oral oestrogens due to their effect of elevating cortisol binding globulin, were excluded (Fig. 1A). Each patient with an afternoon cortisol of ≤100 nmol/L was reviewed by clinician V.R. and classified according to whether a diagnosis of AI was later applied (true positives) or not applied (false positives). Equivocal or ambiguous cases were further reviewed by a specialist (S.C., T.T., or K.M.) before classification. Where insufficient information was available or there was ambiguity, the patient's full electronic clinical record was reviewed. A value of 100 nmol/L was used as the cutoff as this is the local action limit used by Imperial College Healthcare NHS Trust and Northwest London Pathology.

The proportion of patients with an afternoon cortisol >100 nmol/L who were subsequently diagnosed with AI (false negatives) or not (true negatives) are shown in Figure 1B. Each patient was reviewed by V.R. similarly to the true positives and false negatives.

Where SST data were available, patients' baseline, 30-min, and 60-min values were noted and further analysis performed. The SST data and patient histories were comprehensively reviewed and characterized to assess the performance of the existing institutional pass criteria for SSTs and to elucidate up to date cutoff values. Patients who met the criteria to pass an SST at 60 min but not 30 min were reviewed to investigate whether the practice of using 60-min cortisol values is detrimental to patient care. Patients were classified as either "biochemical pass at 30 min" (cortisol value ≥450 nmol/L and a cortisol rise of ≥150 nmol/L from baseline at 30 min), "biochemical pass at 60 min" (cortisol value ≥450 nmol/L and a cortisol rise of ≥150 nmol/L from baseline at 60 min), "failed SST" (leading to diagnosis of AI), or "clinical pass" (based on clinical assessment in patients who had no evidence of AI). These were patients who did not reach 450 nmol/L but had no clinical features of AI and did not require steroid replacement. The decision to pass was made in these cases by consensus of consultant endocrinologists

and biochemists who clinically assessed the patients at results meetings. The investigators in this study further reviewed the longitudinal electronic patient notes for a minimum of 1 year to ensure that these patients were not subsequently diagnosed with-, or treated for- AI. The majority of these patients constituted borderline fails at SST. This study was conducted and registered as a local audit of practice (ref: ASM-030).

Assay Methodology

Plasma adrenocorticotropin (ACTH) was quantified using the Siemens Immulite platform. Intra- and interassay coefficient of variation is <10%. The lower limit of quantification is <5 ng/L.

Serum cortisol was measured on either Abbott Architect or Alinity analyzers. The intra- and interassay coefficients of variation for the Architect and Alinity platforms are <5.5% and <6.2% and <4.3% and <5.1%, respectively. The lower limit of quantification is <28 nmol/L. Imperial College Healthcare NHS Trust encompasses multiple hospitals, which were initially using the Abbott Architect platform. A phased transition across all sites to the Alinity platform occurred in 2018 to 2019. During this transition, the Alinity analyzers were validated against the previous Architect analyzers in compliance with ISO15189.

The results generated on the Alinity analyzers are directly comparable to the previous Architect analyzers, and as such, there has been no change to the local reference ranges or cutoff values employed for cortisol.

Statistical Analysis

The patients were split into 4 groups based on their initial afternoon cortisol level and subsequent diagnosis: (1) true positive (cortisol ≤100 nmol/L and diagnosed with AI), (2) false positive (cortisol ≤100 nmol/L with no evidence of AI), (3) true negative (cortisol >100 nmol/L with no evidence of AI), and (4) false negative (cortisol >100 nmol/L and diagnosed with AI).

Data were received from the biochemistry repository as a Microsoft Excel Spreadsheet. The data were filtered in Microsoft Excel 365 and exported to Prism 9 (GraphPad Software) for further statistical analysis.

For each receiver operating characteristic (ROC) curve that was drawn, an area under the curve was calculated as a summary measure of performance using Prism 9. Data were checked for normality using the Shapiro-Wilk test and were found to be inconsistent with the normal distribution assumption and so nonparametric methods were used. Data for the true-negative and false-positive groups included over 100
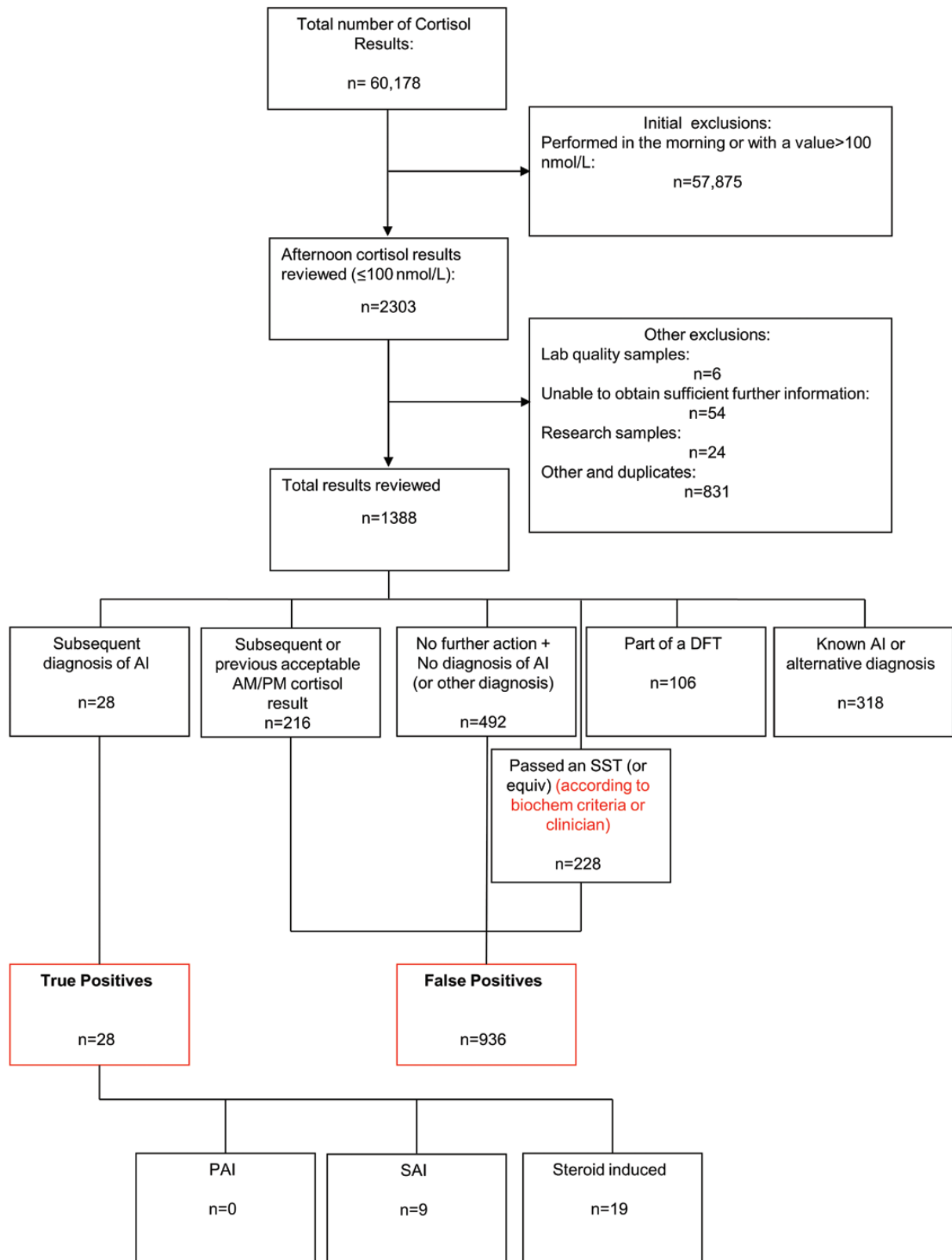
Figure 1. (A) STROBE diagram outlining the characterization of patients with afternoon cortisol value ≤100 nmol/L. (B) STROBE diagram outlining the characterization of patients with afternoon cortisol value >100 nmol/L. Abbreviations: AI, adrenal insufficiency; DFT, dynamic function testing; PAI, primary adrenal insufficiency; SAI, secondary adrenal insufficiency; SST, short Synacthen test.
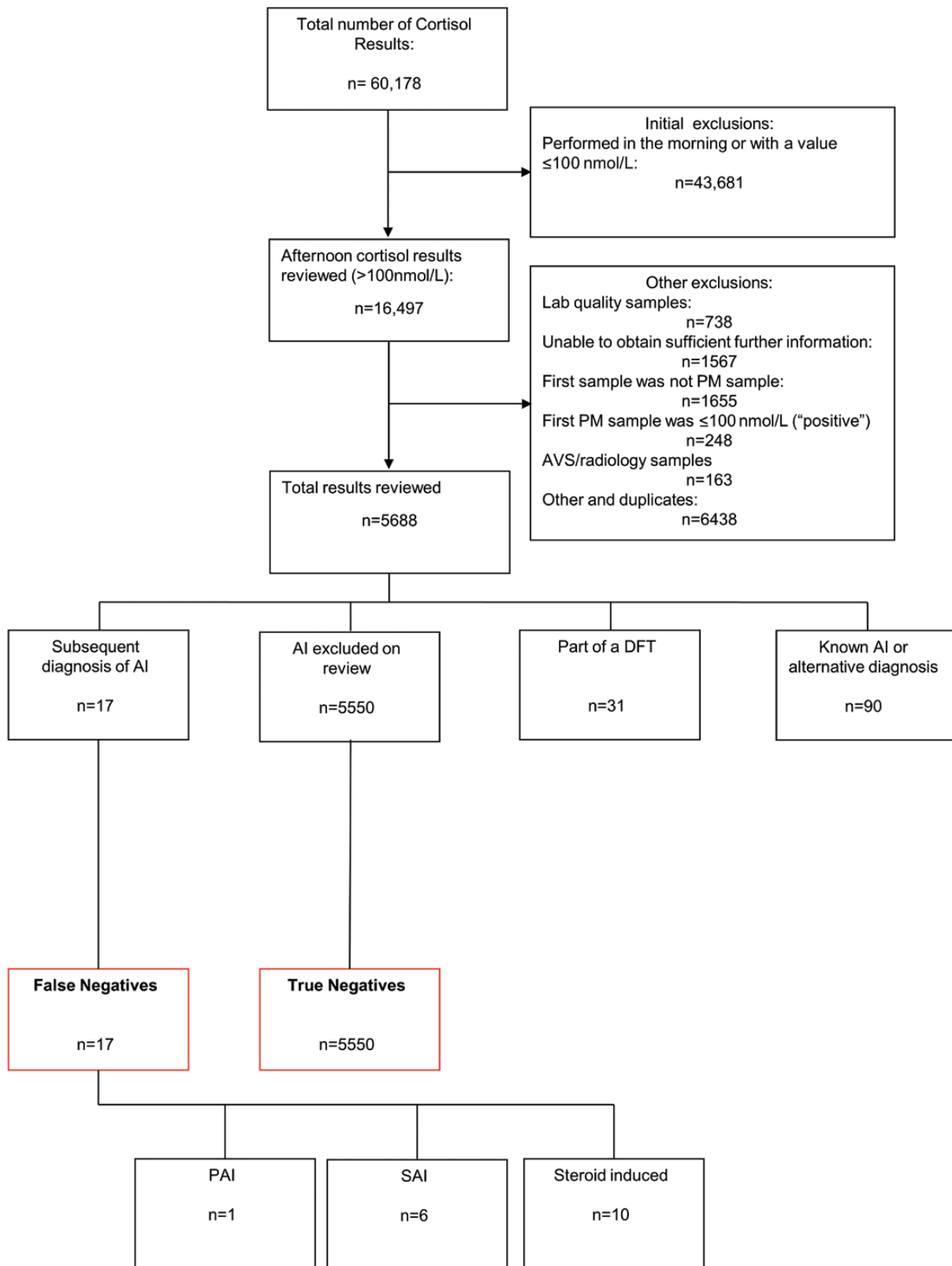
Total number of Cortisol Results:

n= 60,178

Initial exclusions:
Performed in the morning or with a value ≤100 nmol/L:

n=43,681

Afternoon cortisol results reviewed (>100nmol/L):

n=16,497

Other exclusions:
Lab quality samples:

n=738

Unable to obtain sufficient further information:

n=1567

First sample was not PM sample:

n=1655

First PM sample was ≤100 nmol/L ("positive")

n=248

AVS/radiology samples

n=163

Other and duplicates:

n=6438

Total results reviewed

n=5688

Subsequent diagnosis of AI

n=17

AI excluded on review

n=5550

Part of a DFT

n=31

Known AI or alternative diagnosis

n=90

False Negatives

n=17

True Negatives

n=5550

PAI

n=1

SAI

n=6

Steroid induced

n=10

Figure 1. Continued.

samples. To exclude the possibility of a Type 1 error on the Shapiro-Wilk test for these groups, Q-Q plots were generated and visually inspected to confirm the need for a nonparametric approach. Correlations were estimated using Spearman's rank correlation method. SST baseline, 30-min, and 60-min values were paired and compared using Wilcoxon test. Unpaired

data were assessed using the Mann-Whitney U test. Statistical significance was reported at an alpha level of 5%.

Results

A total of 60 178 cortisol samples were analyzed, of which 1388 results were reviewed in patients with an initial afternoon cortisol ≤100 nmol/L, and 5688 results were reviewed in patients with an initial afternoon cortisol >100 nmol/L. In patients with an initial afternoon cortisol ≤100 nmol/L (Fig. 1A), a further 106 patients were excluded as their cortisol levels were part of a dynamic function test, and 318 patients had a known diagnosis of AI or alternative diagnosis recorded on their electronic medical record. In the remaining 964 patients, 28 patients were subsequently diagnosed with AI following an appropriate dynamic function test. Zero patients were diagnosed with primary AI, 9 were diagnosed with secondary AI, and 19 had tertiary (steroid-induced) AI following long-term steroid use (Table 1).

In patients with an initial afternoon cortisol >100 nmol/L (Fig. 1B), a further 31 patients were excluded as their cortisol levels were part of a dynamic function test, and 90 patients had known AI or an appropriate alternative diagnosis. In the remaining 5567 patients, 17 patients were subsequently diagnosed with AI. One patient was diagnosed with primary AI, 6 were diagnosed with secondary AI, and 10 had a new diagnosis of tertiary (steroid-induced) AI upon review of their electronic medical records. (Table 1).

Diagnosis of AI

After exclusion, 6531 patients' afternoon cortisol values were analyzed (Table 2). Of these samples, 5609 (85.9%) were collected between 12 PM and 11:59 PM (Fig. 2). In patients with an initial afternoon cortisol ≤100 nmol/L, 28 patients were later diagnosed with AI while 937 patients did not have clinical or biochemical evidence of AI. In patients with an initial afternoon cortisol >100 nmol/L, 17 patients were diagnosed with AI while 5552 patients had not been diagnosed with AI at the point of review. The

Table 1. Classification of adrenal insufficiency in patients who had an initial afternoon cortisol ≤100 nmol/L and >100 nmol/L

	Primary AI	Secondary AI	Tertiary	Total
Cortisol ≤100 nmol/L	0	9	19	28
Cortisol >100 nmol/L	1	6	10	17
Total	1	15	29	45

Abbreviation: AI, adrenal insufficiency.

available ACTH values measured and the specific cause of AI in the false-negative and true-positive populations have been summarized (Table 3) The mean (±SD) afternoon cortisol values (Table 4) in the true-positive, false-negative, true-negative, and false-positive groups were 62.1 (±22.6) nmol/L, 196.5 (±95.9) nmol/L, 364.7 (±471.5) nmol/L, and 70.1 (±22.4) nmol/L, respectively. The median (IQR) follow up duration was 35 (24) months, in this study.

ROC Analysis of Afternoon Cortisols

An afternoon cortisol cutoff value of ≤100 nmol/L was assessed as the current benchmark, demonstrating a sensitivity of 62.2% (28/45) and a specificity of 85.6% (5550/6486) in diagnosing AI. The positive predictive value (PPV) and negative predictive value was 2.9% and 99.7%, respectively.

ROC analysis (Fig. 3) showed that an afternoon cortisol value of <234 nmol/L provided a sensitivity of 100% and a specificity of 59.5% while an afternoon cortisol value of <201.5 nmol/L provided a sensitivity and specificity of 95.6% and 72.6%, respectively, in diagnosing AI. This indicates that an afternoon cortisol exceeding 234 nmol/L can be used to confidently exclude AI. In addition, an afternoon cortisol value of <53.5 nmol/L

Table 2. Classifies the patient's cortisol values under the criteria for true positive, false positive, true negative and false negative

	AI diagnosed (n)	AI not diagnosed (n)	Total (n)
Cortisol ≤100 nmol/L	True positive (28)	False positive (936)	Positive tests (964)
Cortisol >100 nmol/L	False negative (17)	True negative (5550)	Negative tests (5567)
Total	With disease (45)	Without disease (6486)	Total patients (6531)

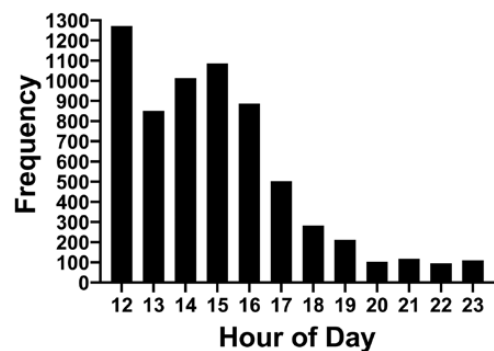Abbreviation: AI, adrenal insufficiency.

Figure 2. Frequency distribution histogram of the hour of sample collection of the 6531 cortisol results analyzed.

Table 3. Breakdown of causes of adrenal insufficiency by type

Group/type	ACTH (ng/L)	Causes
True positive		
Primary	—	NA
Secondary	9.4 ± 7.0	1× pituitary apoplexy
	(n = 7)	2× immunotherapy induced hypopituitarism
		2× macroadenoma
		3× idiopathic hypopituitarism
		1× undetermined pituitary pathology
Tertiary	19.5 ± 27.1 (n = 4)	1× rheumatological (autoimmune)
		2× psoriatic arthropathy
		7× asthma/ respiratory illness
		1×x renal pathology
		2× IBD
		2× unclear indication for long term steroids
		1× neurological diagnosis
		1× autoimmune thyroiditis
		2× other malignancy
False negative		
Primary	274.0	1× histoplasmosis
	(n = 1)	
Secondary	15.0 ± 11.0	2× nonfunctioning adenoma
	(n = 5)	1× subclinical Cushing's
		2× Sheehan's syndrome
		1× macroadenoma
Tertiary	11.3 ± 9.4	1× rheumatological (autoimmune)
	(n = 6)	2× asthma/respiratory illness
		5× renal pathology
		1× IBD
		1× Langerhans histiocytosis

Data are for all patients diagnosed with AI from Fig. 1A and 1B. ACTH values are Reported as mean ± SD and are from early morning, where a result was available. Abbreviation: ACTH, adrenocorticotropin; AI, adrenal insufficiency; IBD, inflammatory bowel disease.

Table 4. Descriptive data of patient's follow-up period and afternoon cortisol values

Group	n	Follow-up, months median (IQR)	PM cortisol, nmol/L mean (±SD)	PM cortisol, nmol/L median (IQR)
True positive (cortisol ≤100 nmol/L, diagnosed with AI)	28	20.5 (23.25)	62.1 (±22.6)	43 (70.5)
False negative (cortisol >100 nmol/L, diagnosed with AI)	17	39 (16)]	196.5 (±95.9)	137 (76)
True negative (cortisol >100 nmol/L, not diagnosed with AI)	5550	35 (25)	364.7 (±471.5)	281 (162.3)
False positive (cortisol ≤100 nmol/L, not diagnosed with AI)	936	33 (23)	70.1 (±22.4)	70 (47)

Abbreviations: AI, adrenal insufficiency; IQR, interquartile range.

provided a sensitivity and specificity of 37.8% and 95.0%, respectively. There were 341 samples with a result of <53.5 nmol/L, of which 17 were later diagnosed with AI.

Correlation Analysis of Afternoon Cortisol With SST Responses

SST data were available in 670 patients out of 6531 (10.3%) patients reviewed in total (Table 5). In these patients, initial afternoon cortisol samples were correlated with SST values at

baselines, 30 min, and 60 min. (Fig. 4A-D). In patients diagnosed with AI (the true-positive and false-negative groups), there was no significant correlation between initial afternoon cortisol and SST values at 0, 30, and 60 min. SST values at 0, 30, and 60 min are strongly correlated in both these groups.

Afternoon Cortisol Levels Predict Magnitude of Response to Synacthen in Healthy Individuals

In patients with AI (Fig. 5A), there is a significant difference in initial afternoon cortisol and stimulated SST
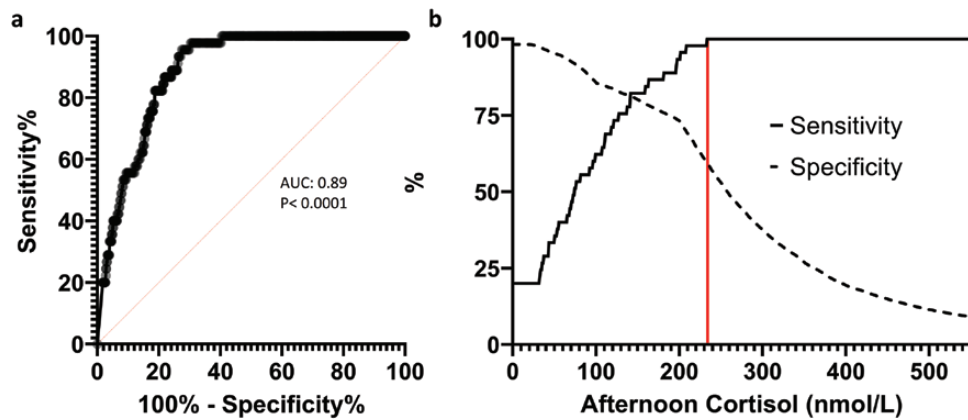
Figure 3. (A) Receiver operating characteristics curve for initial afternoon cortisol values in the diagnosis of adrenal insufficiency. (B) Sensitivity and specificity profile of afternoon cortisol values. The red line demarcates a value of 234 nmol/L, at which a sensitivity of 100% is achieved, suggesting that adrenal insufficiency can be excluded if afternoon cortisol values exceed this cutoff.

Table 5. Descriptive Data of Patients Who had Received an SST in Addition to Providing an Afternoon Cortisol Values

Group	With SST data, n	PM cortisol, nmol/L mean (± SD)	SST T = 0 cortisol nmol/L mean (± SD)	SST T = 30 cortisol, nmol/L mean (± SD)	SST T = 60 cortisol, nmol/L mean (± SD)
True tositive (cortisol ≤100 nmol/L, diagnosed with AI)	24	38.1 (±35.8)	91.9 (±90.7)	216.8 (±85.4)	261.5 (±94.5)
False negative (cortisol >100 nmol/L, diagnosed with AI)	15	192.3 (±113.5)	148.3 (±113.1)	224.7 (±120.6)	292.0 (±131.7)
True negative (cortisol >100 nmol/L, not diagnosed with AI)	406	260.2 (±171.5)	317.4 (±144.1)	565.9 (±145.9)	656.2 (±180.5)
False positive (cortisol ≤100 nmol/L, not diagnosed with AI)	225	61.3 (±29.7)	196.3 (±113.7)	469.7 (±115.1)	558.9 (±113.2)

Patients without recorded SST cortisol values were excluded.

Abbreviations: AI, adrenal Insufficiency; SST, short Synacthen test.

cortisol values. The significant difference from baseline SST cortisol to 30- and 60-min cortisol values are preserved. In healthy individuals (non-AI), there was a significant difference between all comparisons between initial afternoon cortisol, baselines SST, 30-min, and 60-min values (Fig. 5B).

Considering the patients groups subsequently diagnosed with AI, there is no significant difference seen between the true-positive and false-negative groups when comparing baseline, 30-min, or 60-min cortisol values in the SST.

In healthy patients (true negative and false positive), the false-positive patients had a significantly lower initial afternoon cortisol and SST values at baseline, 30 min, and 60 min. Despite the lower SST baseline, these patients were still able to mount a sufficient response to pass their SST. This demonstrates that afternoon cortisol value can indicate the magnitude of response to Synacthen stimulation but does not predict whether the patient will pass or fail the SST. The convergence of the

AI patient's SST results regardless of the initial afternoon cortisol *vs* the divergence of the healthy individuals according to initial afternoon cortisol is best visualized using an SST profile (Fig. 6).

Passing the SST at T = 30 Min *vs* T = 60 Min

Out of the 670 patients that had an SST, 628 patients passed. These 628 patients were further subdivided as pass at 30 min by biochemical criteria, pass at 60 min by biochemical criteria, and clinical pass. In our study population, 140 patients passed their SST at 60 min and not 30 min by the strict biochemical criteria (cortisol value ≥450 nmol/L and a cortisol rise of ≥150 nmol/L from baseline). This made up 22.3% of patients who passed the SST (Table 6, Fig. 7). The clinical records of the 140 patients who had a biochemical SST pass at 60 min but would have otherwise failed at 30 min were reviewed. All 140 cases were reviewed by the authors after a minimum 1-year follow-up. In 138 out of
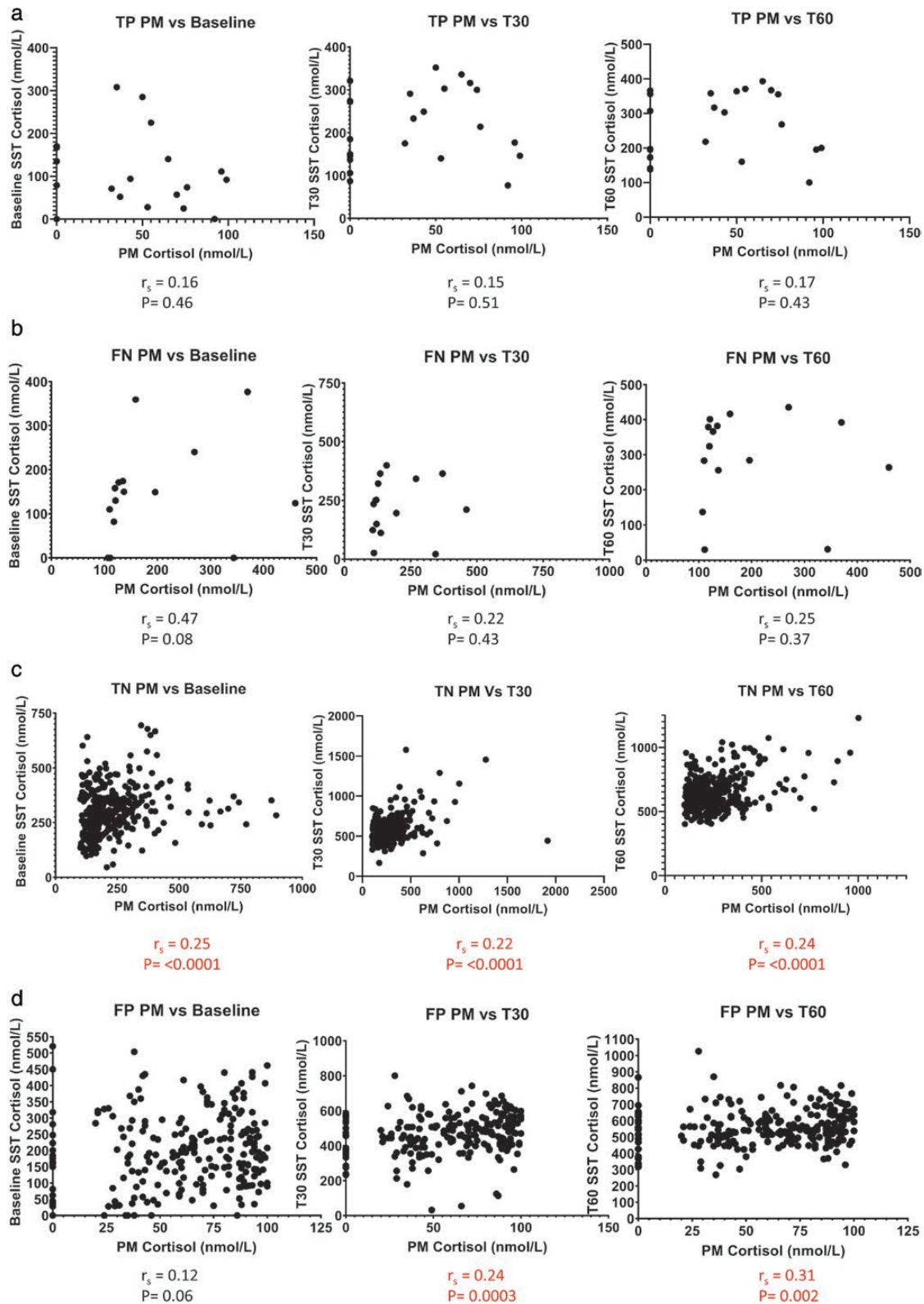
Figure 4. (A) Scatter plot for patients in the true-positive group. (B) Scatter plot for patients in the false-negative group. (C) Scatter plot for patients in the true-negative group. For all panels, afternoon cortisol values are plotted against SST baseline, 30 min, and 60 min values. (D) Scatter plot for patients in the false-positive group. For all panels, afternoon cortisol values are plotted against SST baseline, 30 min, and 60 min values. Abbreviations: FN, false negative; FP, false positive; SST, short Synacthen test; T30, time 30 min; T60, time 60 min; TN, true negative; TP, true positive.
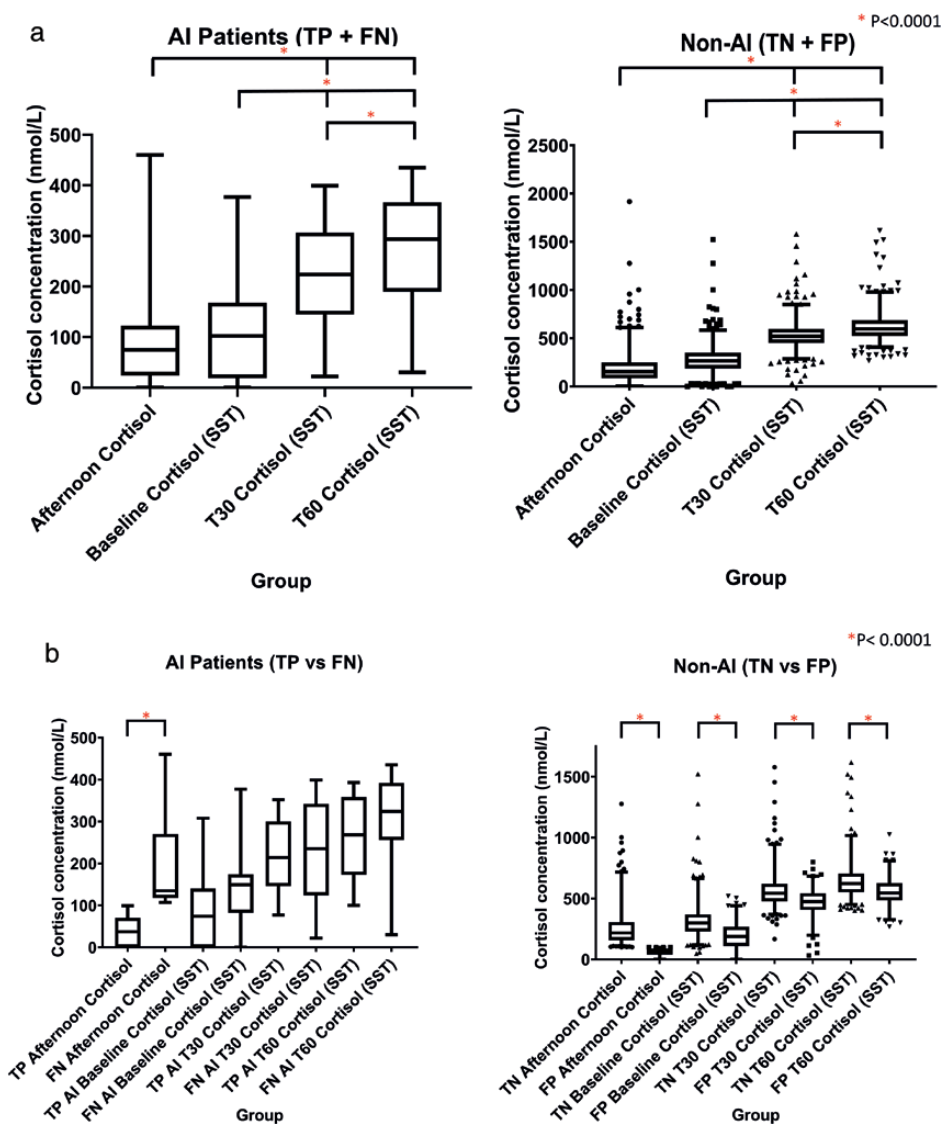
Figure 5. (A) Box and whisker plot demonstrating the difference between cortisol values at different timepoints. Whiskers illustrate the 2.5 to 97.5 centiles of the data. The box demarcates the first and third quartiles, and the middle line represents the median. (B) Box and whisker plot demonstrating the effect of afternoon cortisol level on response to the short Synacthen tests. Whiskers illustrate the 2.5 to 97.5 centiles of the data. The box demarcates the first and third quartiles, and the middle line represents the median. Abbreviations: AI, adrenal insufficiency; FN, false negative; FP, false positive; SST, short Synacthen test; T30, time 30 min; T60, time 60 min; TN, true negative; TP, true positive.

these 140 patients, no further steroids were prescribed in their follow-up, demonstrating that they did not have signs and symptoms of AI. In the 2 patients started on replacement glucocorticoids, 1 had regrowth of a pituitary tumour with loss of their hypothalamic-pituitary-adrenal axis, and the other patient was inappropriately started on glucocorticoid replacement, which was subsequently discontinued.

A further analysis was performed to represent centers that do not include the cortisol increment from baseline in their SST pass criteria. When the cortisol increment of ≥150 nmol/L was omitted at 30 min, 105 patients failed at 30 min who would have otherwise passed at 60 min. This represented 16.7% of patients who passed the SST.

ROC Analysis of SST Data

ROC curves were generated for SST baseline, SST 30-min, and SST 60-min cortisol values, having elucidated whether each patient was subsequently diagnosed with AI or not (Fig. 8A-8C). Afternoon cortisol values for this cohort were reanalyzed in a separate ROC curve (Fig. 8D). At baseline, a cutoff of <359.5 nmol/L yielded a sensitivity of 97.4% and specificity of 22.6%. Optimum sensitivity for 30-min cortisol values was achieved at <366.5 nmol/L, with a sensitivity and specificity of 97.4% and 93.37%, respectively. At 60 min, a cortisol value of <418.5 nmol/L produced 97.4% sensitivity and 93.1% specificity. An afternoon value of <462 nmol/L produced a sensitivity of 100% with
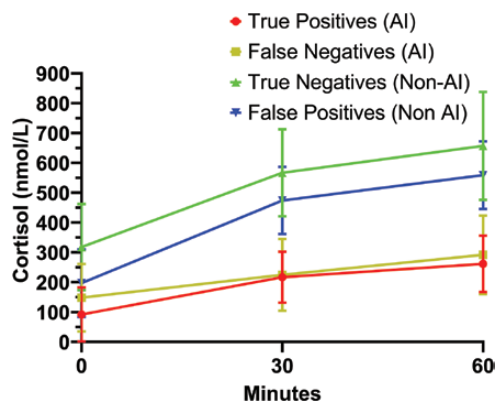
Figure 6. Short Synacthen test (SST) profiles (mean ± SD) of patients according to initial afternoon cortisol. True positives: initial afternoon cortisol of ≤100 nmol/L and subsequently diagnosed with AI. False negatives: initial afternoon cortisol of >100 nmol/L and subsequently diagnosed with AI. True negatives: initial afternoon cortisol of >100 nmol/L in healthy individual. False positives: initial afternoon cortisol of ≤100 nmol/L in healthy individuals.

Table 6. Break down of the results from patients who had passed their short Synacthen test

	n	%
Total number of passed SSTs	628	
T30 pass by biochemical criteria	414	65.9
T60 pass by biochemical criteria	554	88.2
Passed at T60 but not T30	140	22.3
Clinical passes	74	11.8

The following criteria were defined in the study methodology: biochemical pass at 30 mins (cortisol value ≥ 450 nmol/L and a cortisol rise of ≥ 150 nmol/L from baseline at 30 min), biochemical pass at 60 min (cortisol value ≥ 450 nmol/L and a cortisol rise of ≥ 150 nmol/L from baseline at 60 min), and clinical pass (did not meet biochemical criteria as stated but documented as passed by clinician).
Abbreviations: SST, short Synacthen test; T30, time 30 min; T60, time 60 min.

a specificity of 4.8%, while a value of <28.5 nmol/L provided a sensitivity of 23.7% and specificity of 95.1%.

Discussion

This retrospective analysis of over 60 000 cortisol samples suggests that an afternoon cortisol value of ≤100 nmol/L has a sensitivity of 62.2%, specificity of 85.6%, PPV of 2.9%, and negative predictive value of 99.7% for the diagnosis of AI. As AI is a rare disease, the low PPV is expected and reinforces the fact that cortisol values alone cannot be used as a screening test in AI. However, a cortisol value of <53.5 nmol/L provided a specificity of 95.0% in our study population. By instituting a call-out policy of <53.5 nmol/L for cortisol samples collected after noon, there is likely to be a noticeable reduction in call-outs with an acceptable
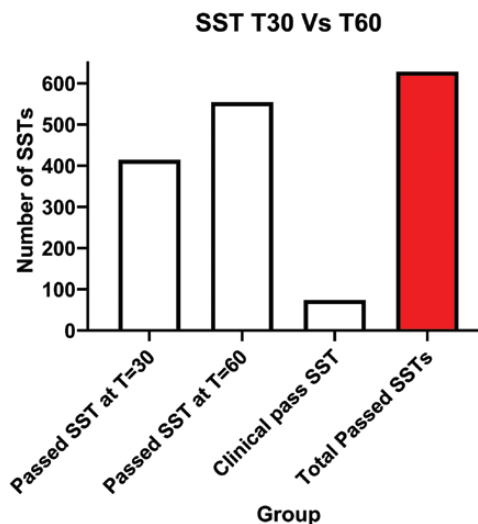

Figure 7. Bar graph demonstrating the number of patients that passed their short Synacthen test (SST) based on each criterion defined in the methodology. Biochemical pass at 30 min (cortisol value ≥ 450 nmol/L and a cortisol rise of ≥ 150 nmol/L from baseline at 30 min), biochemical pass at 60 min (cortisol value ≥ 450 nmol/L and a cortisol rise of ≥150 nmol/L from baseline at 60 min), and clinical pass (did not meet biochemical criteria as stated but documented as passed by consensus of clinicians and biochemists, with no diagnosis of AI on longitudinal review).

level of false positives (individuals with cortisol levels below 53.5 nmol/L, who do not have AI).

Cortisol has a physiological diurnal rhythm, peaking in the morning and with levels lowest at midnight [15]. As such, the current standard is to measure cortisol values at its peak value between 8 and 9 AM in individuals with normal circadian rhythm. With a myriad of factors affecting cortisol levels in the body including the pulsatile rhythm of its secretion, stress, and differing sleep-wake cycles, it is difficult to establish an accurate reference range outside this time window and even more challenging for an afternoon cortisol value [16,17]. Currently, a large proportion of cortisol samples are still taken at less appropriate times, particularly during the afternoon when it is often more practical for both patients and clinicians. A retrospective observational study analyzed untimed cortisol samples in patients who had subsequent SSTs and concluded that no single cutoff point in baseline cortisol was both adequately sensitive and specific [18]. Therefore, they identified a cortisol value <420 nmol/L as their threshold for requesting progression to an SST. The study is confounded by the use of an arbitrary 30-min SST cutoff of 550 nmol/L to exclude AI, instead of a Roche E170 platform specific cutoff. The use of such a high threshold for untimed cortisol will undoubtedly lead to a significant burden in SSTs to detect a very rare disease.

Physicians understandably hold a low threshold for referring patients with low untimed cortisol values for SST as
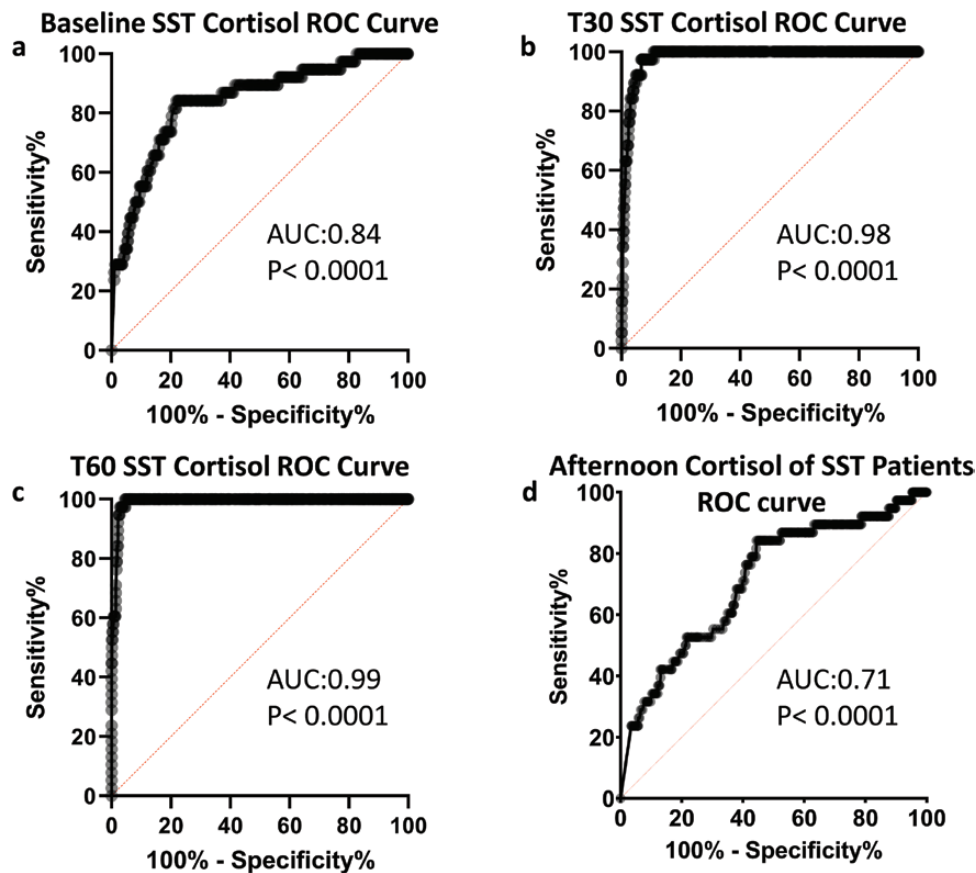
Figure 8. Receiver operating characteristics (ROC) curve for (A) short Synacthen test (SST) baseline cortisol values, (B) SST 30-min cortisol values, (C) SST 60-min cortisol values, and (D) afternoon cortisols for patient with SSTs.

the risk of missing a patient with primary AI could be fatal. Starting glucocorticoid replacement in these individuals is of the utmost importance. In this study, however, none of the patients with an afternoon cortisol ≤100 nmol/L were diagnosed with primary AI. It is this cohort of primary AI patients who are at risk of a true Addisonian crisis [19]. Secondary AI patients still have normal mineralocorticoid function as the renin-angiotensin-aldosterone-system is independent of pituitary function, and data have shown that their risk of a crisis is lower than in primary AI cohorts [11]. It is likely that when there is a strong clinical suspicion of primary AI, untimed cortisols are avoided in favor of 9 AM cortisols; therefore, primary AI patients were not captured in our data set with the exception of 1 individual, albeit with an afternoon cortisol >100 nmol/L. This patient was being screened with an SST and concurrently demonstrated an elevated ACTH of 257 ng/L (reference range <30 ng/L at 9 AM). In this context, afternoon cortisol measurements may still be of value in cases of secondary and tertiary AI, as cases were still picked up. Although ACTH was not commonly requested with cortisol in the afternoons, it is important to note that elevated levels of ACTH in secondary and tertiary cases may be a precursor to resumption of adrenal function. For this reason,

it is essential that ACTH quantification is done at the time of morning SSTs to allow for a fuller assessment of the hypothalamic-pituitary-adrenal axis.

Individuals with false-positive results (PM cortisol ≤100 nmol/L but not AI) were still able to mount an appropriate SST response despite significantly lower baseline SST values compared to those in the true-negative group (PM cortisol >100 nmol/L and not AI). This demonstrates the ability to mount a sufficient glucocorticoid response in times of physiological stress such as during surgery or sepsis. The variability among the eucortisolemic healthy population is greatly underappreciated [20].

The current diagnostic criteria set for an SST may vary locally but is principally based on the seminal work by Plumpton and Besser [21]. As highlighted by Khoo et al, the cut-offs for stimulation of cortisol secretion during dynamic testing derived from this original study are not applicable in the modern context since the cortisol assays in the 1960s were prone to interference from other glucocorticoids, reliable ACTH quantification did not exist in the era, and surgical approaches have evolved to produce significantly less physiological stress [22]. As such, the diagnostic criteria have yet to be updated with the improvements in technology. Recent data have explored the utility of 60-min SST

samples, with some groups challenging the need for 30-min SST samples [23,24]. Studies have reported that without the 60-min sample, up to 5% of patients could be inappropriately started on long-term steroids. In our study population, this was closer to 22.3%. In the patients who had a biochemical pass at 60 mins but not at 30 mins, 138 out of 140 did not require any further steroids. In the remaining 2, neither constituted a missed diagnosis. There is therefore a need to ensure patients are not diagnosed with AI on the basis on a 30-min cortisol value and to further investigate outcomes in patients passing an SST based on the 60-min cortisol value alone.

By comprehensively reviewing the medical records associated with 670 SSTs performed at Imperial College Healthcare NHS Trust, this study provides a longitudinal picture of whether these patients subsequently were diagnosed with AI or not. ROC analysis of the SST data available has strongly indicated that the diagnostic cutoffs used for 30- and 60-min SST cortisol values should be modified to improve performance on Abbott Architect and Alinity platforms. The local cutoffs currently employed demand an increase of 150 nmol/L to a level greater than >450 nmol/L. It is important to consider the context that this center uses an Abbott platform to measure cortisol levels and that there is paucity of evidence available in the literature for Abbott analyzers. The data from this study indicates that 30-min and 60-min decision levels for SSTs should be set at 366.5 nmol/L and 418.5 nmol/L, respectively, to ensure a sensitivity of at least 95%. The proposed 60-min cutoff is in keeping with other groups who suggest a lower reference limit for stimulated cortisol on the Abbott Architect of 420 nmol/L [25]. This is slightly higher than the suggested lower limit of 403 nmol/L for the newer generation of Roche analyzers [26].

Limitations

This study was limited by the retrospective nature of the data collection and the approach to filtering. The principal aim was to investigate the utility of untimed cortisol levels that are performed, particularly in the afternoon. Consequently, the SST data that have been reviewed only include patients who had their first cortisol measurement after midday, possibly excluding a number of SSTs that were not represented in the data set. The afternoon cutoff of >234 nmol/L has been derived from a large number of cortisol tests performed in a large institution, where the clinical records were assessed to exclude a subsequent diagnosis of AI. As not all of the patients received SSTs or equivalent dynamic function tests to definitively exclude AI, there is a chance that the rate of AI is underrepresented in this study, as patients may still later be diagnosed with AI after the follow-up period.

As a data collection study encompassing the entirety of cortisol results in a large institution, this study did not consider the individual clinician's index of suspicion prior to ordering the initial afternoon cortisol. The cutoff values derived in this study may still require expert endocrinology interpretation in the context of the clinical scenario. For example, in the context of a family history of Addison's disease with a longitudinally rising ACTH, an afternoon cortisol of >234 nmol/L may not necessarily preclude the need for a subsequent SST. Given the large volume of data, consideration was not given to patients who had received glucocorticoids in the previous 24 h that may have temporarily suppressed cortisol. These patients were not excluded. Further, afternoon cortisol values may still be needed for pragmatic reasons, given the impracticality of performing morning cortisols.

The performance of untimed cortisols in their predictive value may have been detrimentally affected by including samples that were collected from 6 PM to 11:59 PM. On balance, these samples were included because out-of-hours biochemistry services do not typically have the autonomy to treat these results differently to those obtained from 12 PM to 5:59 PM (Fig. 2).

The analytical approach of this study, in the absence of a prospective design, hinders the ability to assess the 150 nmol/L increment that is described in literature to characterize a pass in an SST. It is not universal practice to consider the increment in the interpretation of an SST with many centers now omitting it.

Conclusion

This study demonstrates that healthy individuals with low afternoon cortisol values also show significantly low SST baseline and peak cortisol values compared to healthy individuals with afternoon cortisol >100 nmol/L. However, these low afternoon cortisol values did not correlate with baseline SST levels or SST stimulated levels and therefore are not predictive of AI. Moreover, a single low afternoon cortisol (≤100 nmol/L) did not lead to a diagnosis of a primary AI in our study population. Although this does not include afternoon cortisol values that were ordered after a clinician expressed clinical suspicion, it opens the possibility of reducing the current guidelines for low cortisol callouts in our local trust from 100 nmol/L to 53.5 nmol/L. Further, afternoon cortisol values of >234nmol/L achieved sensitivity of 100%, indicating that cortisol measurements above this level exclude AI. There is further scope to lower the Abbott Architect and Alinity platform specific cutoffs for 30-min and 60-min SST cortisols. The results of this study advocate the inclusion of 60-min SST stimulated cortisol levels when assessing patients to avoid mislabeling up

to 22.3% of individuals with AI and the detrimental effects that accompany inappropriate glucocorticoid replacement.

Additional Information

Correspondence: Sirazum Choudhury, MBBS BSc, Division of Diabetes, Endocrinology and Metabolism, 6th Floor Commonwealth Building, Hammersmith Hospital, Imperial College London, Du Cane Road, London W12 0NN, UK. E-mail: s.choudhury@imperial.ac.uk.

Disclosure Summary: The authors have nothing to disclose.

Data Availability: Restrictions apply to the availability of some or all data generated or analyzed during this study to preserve patient confidentiality or because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided.

References

1. Bergthorsdottir R, Leonsson-Zachrisson M, Odén A, Johannsson G. Premature mortality in patients with Addison's disease: a population-based study. *J Clin Endocrinol Metab*. 2006;**91**(12):4849-4853.
2. Behan LA, Carmody D, Rogers B, et al. Low-dose hydrocortisone replacement is associated with improved arterial stiffness index and blood pressure dynamics in severely adrenocorticotrophin-deficient hypopituitary male patients. *Eur J Endocrinol*. 2016;**174**(6):791-799.
3. Sherlock M, Reulen RC, Alonso AA, et al. ACTH deficiency, higher doses of hydrocortisone replacement, and radiotherapy are independent predictors of mortality in patients with acromegaly. *J Clin Endocrinol Metab*. 2009;**94**(11):4216-4223.
4. Zueger T, Kirchner P, Herren C, et al. Glucocorticoid replacement and mortality in patients with nonfunctioning pituitary adenoma. *J Clin Endocrinol Metab*. 2012;**97**(10):E1938-E1942.
5. Oelkers W. Adrenal insufficiency. *N Engl J Med*. 1996;**335**(16):1206-1212.
6. Bleicken B, Hahner S, Ventz M, Quinkler M. Delayed diagnosis of adrenal insufficiency is common: a cross-sectional study in 216 patients. *Am J Med Sci*. 2010;**339**(6):525-531.
7. Sbardella E, Isidori AM, Woods CP, et al. Baseline morning cortisol level as a predictor of pituitary-adrenal reserve: a comparison across three assays. *Clin Endocrinol (Oxf)*. 2017;**86**(2):177-184.
8. Arlt W, Allolio B. Adrenal insufficiency. *Lancet*. 2003;**361**(9372):1881-1893.
9. Hägg E, Asplund K, Lithner F. Value of basal plasma cortisol assays in the assessment of pituitary-adrenal insufficiency. *Clin Endocrinol (Oxf)*. 1987;**26**(2):221-226.
10. Yip CE, Stewart SA, Imran F, et al. The role of morning basal serum cortisol in assessment of hypothalamic pituitary-adrenal axis. *Clin Invest Med*. 2013;**36**(4):E216-E222.
11. Rushworth RL, Torpy DJ, Falhammar H. Adrenal crises: perspectives and research directions. *Endocrine*. 2017;**55**(2):336-345.
12. Mackenzie SD, Gifford RM, Boyle LD, Crane MS, Strachan MWJ, Gibb FW. Validated criteria for the interpretation of a single measurement of serum cortisol in the investigation of suspected adrenal insufficiency. *Clin Endocrinol (Oxf)*. 2019;**91**(5):608-615.
13. Croal B. The communication of critical and unexpected pathology results. Royal College of Pathologists. Published October 2017. Accessed January 31, 2021. https://www.rcpath.org/uploads/assets/bb86b370-1545-4c5a-b5826a2c431934f5/The-communication-of-critical-and-unexpected-pathology-results.pdf
14. International Organization for Standardization. *Medical Laboratories—Requirements for Quality and Competence* (ISO 15189). ISO; 2012.
15. Howlett TA. An assessment of optimal hydrocortisone replacement therapy. *Clin Endocrinol (Oxf)*. 1997;**46**(3):263-268.
16. Young EA, Abelson J, Lightman SL. Cortisol pulsatility and its role in stress regulation and health. *Front Neuroendocrinol*. 2004;**25**(2):69-76.
17. Nui S, Chung M, Chen C, Hegney D, O'Brien A, Chou K. The effect of shift rotation on employee cortisol profile, sleep quality, fatigue, and attention level: a systematic review. *J Nurs Res*. 2011;**19**(1):68-81.
18. Kadiyala R, Kamath C, Baglioni P, Geen J, Okosieme OE. Can a random serum cortisol reduce the need for short synacthen tests in acute medical admissions? *Ann Clin Biochem*. 2010;**47**(Pt 4):378-380.
19. de Herder WW, van der Lely AJ. Addisonian crisis and relative adrenal failure. *Rev Endocr Metab Disord*. 2003;**4**(2):143-147.
20. Hahner S, Loeffler M, Bleicken B, et al. Epidemiology of adrenal crisis in chronic adrenal insufficiency: the need for new prevention strategies. *Eur J Endocrinol*. 2010;**162**(3):597-602.
21. Plumpton FS, Besser GM. The adrenocortical response to surgery and insulin-induced hypoglycaemia in corticosteroid-treated and normal subjects. *Br J Surg*. 1969;**56**(3):216-219.
22. Khoo B, Boshier PR, Freethy A, et al. Redefining the stress cortisol response to surgery. *Clin Endocrinol (Oxf)*. 2017;**87**(5):451-458.
23. Chitale A, Musonda P, McGregor AM, Dhatariya KK. Determining the utility of the 60 min cortisol measurement in the short Synacthen test. *Clin Endocrinol (Oxf)*. 2013;**79**(1):14-19.
24. Kumar R, Carr P, Moore K, Rajput Z, Ward L, Wassif WS. Do we need 30 min cortisol measurement in the short Synacthen test: a retrospective study. *Postgrad Med J*. 2020;**96**(1138):467-472.
25. El-Farhan N, Pickett A, Ducroq D, et al. Method-specific serum cortisol responses to the adrenocorticotrophin test: comparison of gas chromatography-mass spectrometry and five automated immunoassays. *Clin Endocrinol (Oxf)*. 2013;**78**(5):673-680.
26. Javorsky BR, Raff H, Carroll TB, et al. New cutoffs for the biochemical diagnosis of adrenal insufficiency after ACTH stimulation using specific cortisol assays. *J Endocr Soc*. 2021; **5**(4):bvab022.