BASELINE MORNING CORTISOL LEVEL AS A PREDICTOR OF PITUITARY-ADRENAL RESERVE: A COMPARISON ACROSS THREE ASSAYS

Emilia Sbardella[1,2], Andrea M. Isidori[2], Conor P. Woods[1], Nicola Argese[3], Jeremy W Tomlinson[1], Brian Shine[4], Bahram Jafar-Mohammadi[1] and Ashley B Grossman[1]

[1]Department of Endocrinology, Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Oxford, UK, OX3 7LE, UK

[2] Department of Experimental Medicine, Sapienza University of Rome, Rome, Italy, Viale Regina Elena, 324, 00161, Rome Italy

[3] Department of Endocrinology, S.S. Annunziata Hospital, Taranto, Italy

[4] Department of Clinical Biochemistry, John Radcliffe Hospital, University of Oxford, Oxford, UK, OX3 9DU, UK

Short Title: Morning cortisol as a predictor of HPA reserve in 3 assays

Key words: Short Synacthen Test, cortisol, adrenal insufficiency, Synacthen, ACTH test

Word Count: 3262 words

Number of Figures and Tables: 4 tables (one of them Supplementary table) and 1 Figure

Corresponding Author:

Prof. Ashley Grossman
Department of Endocrinology,
Oxford Centre for Diabetes, Endocrinology and Metabolism,
Churchill Hospital,
Oxford OX3 7LE, UK
Tel +44-1865-857308
Fax: +44-1865-857311
Email: ashley.grossman@ocdem.ox.ac.uk

Acknowledgement: Nothing to declare.

Disclosure: The authors have nothing to disclose.

Abbreviations: short Synacthen test (SST); adrenal insufficiency (AI); adrenal sufficiency (AS); positive predictive value (PPV); glucocorticoid (GC)

ABSTRACT

Context: The short ACTH stimulation test (250μg) is the dynamic test most frequently used to assess adrenal function. It is possible that a single basal cortisol could be used to predict the dynamic response, but research has been hampered by the use of different assays and thresholds.

Objective: To propose a morning baseline cortisol criterion of three of the most commonly-used modern cortisol immunoassays - *Advia Centaur* (Siemens), *Architect* (Abbott) and the *Roche Modular System* (Roche) - that could predict adrenal sufficiency.

Design: Observational, retrospective cross-sectional study at two centres

Patients and Measurements: Retrospective analysis of the results of 1019 SSTs with the *Advia Centaur,* 449 SSTs with the *Architect,* and 2050 SSTs with the *Roche Modular System assay*. Serum cortisol levels were measured prior to injection of 250μg *Synacthen* and after 30 minutes. Overall, we were able to collate data from a total of 3518 SSTs in 3571 patients.

Results: Using receiver-operator curve analysis, baseline cortisol levels for predicting passing the SST with 100% specificity were 358 nmol/l for *Siemens*, 336 nmol/l for *Abbott* and 506 nmol/l for *Roche*. Utilising these criteria: 589, 158 and 578 SSTs respectively for Siemens, Abbott and Roche immunoassays could have been avoided.

Conclusions: We have defined assay-specific morning cortisol levels that are able to predict the integrity of the hypothalamo-pituitary-adrenal axis. We propose that this represents a valid tool for the initial assessment of adrenal function and has the potential to obviate the need for dynamic testing in a significant number of patients.

INTRODUCTION

Careful evaluation of adrenal function is essential in patients with disease that can affect the hypothalamo-pituitary-adrenal (HPA) axis. Laboratory evaluation is recommended in patients suspected to have primary adrenocortical insufficiency (AI) and in those at risk of developing AI following prolonged glucocorticoid (GC) therapy, hypothalamo-pituitary disease and related surgery or radiation, and in patients with other proven disorders of the HPA axis.

The established ‘gold standard’ test for the assessment of the HPA axis is to measure the cortisol response to insulin-induced hypoglycaemia (1-4). However, while in extensive use and with a good safety record, this test remains potentially dangerous for the patient, and is highly demanding in terms of the experienced medical supervision required (5). It is contraindicated in patients with cerebrovascular diseases, epilepsy, ischaemic heart disease and severe metabolic disorders, and is not recommended in young children or patients with concomitant serious disease (6, 7).

Stimulation of the adrenal with exogenous ACTH (*Synacthen, Cosyntropin*) at the conventional dose of 250μg of *Synacthen* has been described as a reliable screening method in patients with suspected adrenal impairment due to hypothalamic or pituitary dysfunction, and has been validated against insulin-induced hypoglycaemia (ITT) (5, 8-18). In general, other than in patients with an acute insult to the HPA axis as immediatey after pituitary surgery (19-21), this test has been shown to correlate well with more complex testing procedures, is rapid and simple to perform, and has been claimed in extensive use to show an absence of severe side effects or contraindications.

However, the SST is associated with a small risk of hypersensitivity reactions and with the need for medical observation following the administration of *Synacthen*: the *Society for Endocrinology* has recommended that it only be perfomed in units where immediate resuscitation facilities are available. As an alternative, random serum cortisol measurements are considered to be unreliable as there is a diurnal rhythm of cortisol secretion, and low levels may simply be a physiological response to circadian rhythmicity. However, such a rhythm peaks at around 08.00-09.00h, and previous authors have demonstrated the presence of a strong correlation between such basal morning (08.00-09.00h) serum cortisol and maximal-stimulated cortisol levels during the ITT (22), and also between basal cortisol levels and peak levels following ACTH administration (8, 23). Thus, in previous studies, with a basal morning cortisol of less than 100 nmol/L or more than 500 nmol/L the SST was of little added value (24). Nevertheless, in other studies basal cortisol measurements

in the range 300-500 nmol/l have been suggested as rendering full stimulation with ACTH unnecessary (19, 22, 25, 26). Unfortunately, many of these studies are hampered by the use of differing assays and thus varying thresholds for defining a normal response, with criteria for one assay used to define normality in a different assay. This variation between assays is critical, and was emphasised in a recent study of a comparison of the normative responses to *Synacthen* using a variety of different cortisol measurements (27).

In a previous study, the data from a single centre were analysed using a single assay. We have now conducted a retrospective study analysing SSTs performed at two major centres, the Oxford Centre for Diabetes, Endocrinology and Metabolism, (OCDEM) in Oxford, UK, between January 2011 and April 2016, and at the Centre for Endocrinology, Diabetes and Metabolism at the University Hospital of Birmingham, UK, between January 2008 and December 2012, in order to assess the predictive value of baseline morning serum cortisol in determining the outcome of SST to evaluate adrenal reserve using a variety of different assays, and looking at the effect of the timing of the sampling. Data from the cohort of patients undergoing SST analysed by the *Roche* assay have been presented previously (23). In this paper, a discrete subset of this cohort was analysed where specific indications for performing the SST were documented paralleling the clinical data in patient samples analysed by the *Centaur* and *Architect* assays. We therefore have attempted to broaden the use of single morning cortisol samples and provide normative values which can be more generally applicable to different centres.

METHODS

STUDY POPULATION AND ASSAYS

This study was an observational, retrospective cross-sectional analysis of all SSTs performed in two secondary/tertiary care centres across all medical specialities with three different cortisol assays: *Siemens ADVIA Centaur immunoassay analyser* (Siemens Healthcare Diagnostics, Frimley, UK), *Abbott Architect i-2000 immunoassay analyser* (Abbott Diagnostics, Maidenhead, UK) and the *Roche Modular System* (Roche, Mannheim, Germany).

Overall, we were able to collate data from a total of 3957 out-patient SSTs in 3571 patients.

We collected and analysed the results of 1375 SSTs performed in 1050 patients at OCDEM, Oxford University Hospitals NHS Foundation Trust, UK, between January 2011 and December 2014 when cortisol samples were analysed with *Advia Centaur,* the results of 532 SSTs performed in 471 patients at OCDEM, Oxford University Hospitals NHS Foundation Trust, UK between January 2015 and April 2016 when cortisol samples were analysed with *Architect*, and the results of SSTs performed in 2050 patients at the Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, UK, between January 2008 and December 2012 with the *Roche Modular System*.

All results were collected from the electronic medical record system.

The indication for the SST was derived from clinical fields and clinical letters to general practitioners. We excluded SSTs performed without any clear indication or

with a non-specified indication; thus, we were able to analyse the results of 1019 SSTs performed with *Advia Centaur*, 449 with *Architect*, and *2050* with Roche Modular system, a total of 3518.

All SSTs were performed in the morning (08.30h–12.00h), but for the data from Oxford, (*Advia Centaur* and *Architect*) only those tests started in the time range 08.30-10.00h were included in the principal analysis. However, we performed a secondary analysis including only SSTs initiated after 10.00h comprising 443 SSTs analysed by *Advia Centaur* and 109 SSTs analysed by *Architect* (we did not analyse the latter data due to the small sample size).

With regard to the *Roche Modular System* data, all SSTs were initiated between 08.30h and 12.00h, but the exact time of the beginning of the test was not recorded and thus we were unable to perform a secondary analysis based on the precise time of beginning of the test in this cohort.

Patients on the oral contraceptive pill or other oestrogen replacement were required to stop the treatment at least 6 weeks before the test. Baseline serum cortisol levels were measured prior to injection of 250μg *Synacthen* (*Synacthen* 250μg, Questcor Operations Limited, Dublin, Ireland) for the groups of Siemens and Abbott assays and 250μg *Synacthen* (Alliance Pharmaceuticals, Chippenham, UK) for the group of Roche assays, intramuscularly; for the Abbott assay *Synacthen* was injected intramuscularly until December 2015 and then, from January to April 2016, intravenously. Blood was sampled for serum cortisol at baseline and after 30 minutes. The 30 minute response to intramuscular or intravenous *Synacthen* has

been shown to be equivalent (28). After administration of *Synacthen*, the patients were observed for 15 minutes for signs of any allergic reaction.

Serum cortisol analysed by *Advia Centaur* showed an inter-assay imprecision of 10.5% at 83 nmol/L, 6.0% at 524 nmol/L, and 7.0% at 904 nmol/L; by *Architect* of 5.6% at 72 nmol/L 2.2% at 433 mol/L and 2.4 at 667 mol/L; and by the *Roche Modular System* of <8% for levels between 76 and 925 nmol/l.

The interpretation of the SST is based on the 30-minutes serum cortisol where an adequate response to *Synacthen* for *Advia Centaur* was defined as >450 nmol/l (27), for *Architect* as >430 nmol/l (27), for the *Roche Modular System* as >550 nmol/l (27).

STATISTICAL METHODS

Data presented are expressed as means ± standard deviations (SD) for continuous variables and as counts (%) for categorical variables. For comparisons of single variables, T-tests were used while, for analyses involving multiple comparisons, the one-way ANOVA or, for non-parametric data, the Mann-Whitney test, were used to determine statistical significance. Frequencies were compared using the chi-squared test.

We performed receiver-operating characteristic (ROC) curves to evaluate the diagnostic performance of basal cortisol as a predictor of adrenal sufficiency (AS)

analysing sensitivity and specificity for each possible test threshold/cut-off, and we used the area under the ROC curve to express the overall diagnostic accuracy of the index criterion. In particular, we have reported different thresholds of both specificity and sensitivity: 95%, 99%, 100%. Subsequent to this, separate analyses were performed for data for different subgroups of cases.

P<0.05 was considered indicative of a statistically significant difference. Statistical analyses were performed using SPSS (version 17, Chicago, IL, USA) and the GraphPad Prism 6.0 software package (GraphPad Software, Inc. La Jolla, CA, USA).

RESULTS

SHORT SYNACTHEN TEST RESULTS

Advia Centaur (Siemens) assay

A total of 1019 patients were included, 416 (40.8%) males. The mean age was 51.7±19 years with an age range of 12–96 years.

Overall, 133/1019 patients (13.1%) had adrenal insufficiency (AI) as defined by a serum cortisol <450 nmol/l at 30 min.

Results of the SSTs according to different indications for the performance of the test are reported in Table 1.

Architect (Abbott) assay

449 patients were included, 195 (43.4%) males. The mean (SD) age was 51.7±18.3 years and an age range 18–95 years.

Overall, 89/449 patients (19.8%) had adrenal insufficiency as defined by a serum cortisol <430 nmol/l at 30 min.

Results of the SSTs according to different indications for the performance of the test are reported in Table 1.

Roche Modular System (Roche) assay

A total of 2050 patients were included, 910 (44.4%) males. The mean (SD) age was 55.7±19.2 years with an age range 18–100 years.

Overall, 435/2050 patients (21.2%) had adrenal insufficiency as determined by a serum cortisol <550 nmol/l at 30 min.

Results of the SSTs according to different indications for the performance of the test are reported in Table 1.

ACCURACY OF BASELINE MORNING CORTISOL IN PREDICTING SST RESULTS

We performed ROC curve analyses with the aim of finding a baseline cortisol level able to predict an accurate value for passing the SST for each of the three cortisol assays described (Table 2 and Figure 1).

Advia Centaur (Siemens) assay

Baseline serum cortisol in the 1019 SSTs analysed correlated significantly with the levels 30 minutes after *Synacthen* administration (Spearman's r=0.67 p<0.001). A ROC curve performed on these data set showed that a baseline cortisol ≥358 nmol/l had a specificity of 100% for predicting passing the SST while a baseline cortisol ≤56 nmol/l had a sensitivity of 100% for predicting failure (AUC: 0.960, 95% CI 0.947-0.973). Utilising these criteria, 589 SSTs (57.8% of all SSTs performed) could have been avoided.

Architect (Abbott) assay

Baseline serum cortisol in the 449 SSTs analysed correlated significantly with the levels 30 min after *Synacthen* administration (Spearman's r=0.66 p<0.001). A serum cortisol concentration ≥336 nmol/l provided 100% specificity for predicting a normal SST response while a baseline cortisol ≤83 nmol/l gave 100% sensitivity for predicting failure (AUC: 0.872, 95% CI 0.831-0.913). With these cut-off values, 158 SSTs (35.2% of all SSTs performed) could have been avoided.

Roche Modular system (Roche) assay

Baseline serum cortisol in the 2050 SSTs analysed correlated significantly with the levels 30 min after ACTH stimulation (Spearman's r=0.73 p<0.001). A ROC curve performed on this data set showed that a baseline cortisol ≥506 nmol/l had a

specificity of 100% for predicting a normal SST and a baseline cortisol ≤102 nmol/l was 99% sensitive for predicting failure (AUC: 0.899, 95% CI 0.882-0.916). Utilising these cut-offs, 578 SSTs (28.2% of all SSTs done) could have been avoided.

We then analysed the accuracy of the basal cortisol for each independent group of pathologies and test indications. The AUC with 95% CI 0.947-0.973 and threshold cortisol values for specificities and sensitivities for each cortisol assay in all patients and according to different subgroups of indication for the SST are shown in Table 2, and then analysed according to age, sex, menopausal status in Supplementary Table 1. Considering the different indications for the performance of the SST in the three cortisol assays analysed, there was no clear trend for different criteria to be used contingent upon a specific pathology except for the data on patients on GCs treatment in the *Roche* group, as previously reported (23).

THE TIMING OF MORNING CORTISOL MEASUREMENT AND ITS ABILITY TO PREDICT AN INTACT ADRENAL RESERVE

In order to determine if the timing of the morning cortisol measurement influenced its ability to predict adrenal reserve, we performed separate analyses of the samples anaysed by the *Advia Centaur* assay for those tests initiated before or after 10.00h (Table 3).

A total of 443 patients were included in this subgroup analysis of tests initiated after 10.00h, of whom 197 (44.5%) were males. Overall, 37/443 patients (8.4%) had

adrenal insufficiency as determined by a serum cortisol <450 nmol/l at 30 minutes. Baseline serum cortisol in the 443 SSTs analysed correlated significantly with the levels 30 minutes after ACTH stimulation (Spearman's r=0.66, p<0.001). A ROC curve performed on this data set showed that a baseline cortisol ≥376 nmol/l had a specificity of 100% for predicting passing the SST and a baseline cortisol ≤35 nmol/l had a sensitivity of 100% for predicting failure (AUC: 0.944, 95% CI 0.914-0.975).

We calculated Positive Predictive Value (PPV) of baseline cortisol with given 100% specificity cut-offs for *Advia Centaur* performed before and after 10.00h. We calculated a PPV of 27.7% for baseline cortisol performed before 10.00h and 18.6% after 10.00h with *Advia Centaur* (Siemens).

AGE AND SEX SUBGROUP DIFFERENTIATION

Differentiation of baseline cortisol values according to age and sex subgroups are shown in Supplementary Table 1. There was no difference between baseline cortisol in male and female subjects in the three assays. Patients above 70 years of age showed a significantly higher mean baseline cortisol compared to younger patients both in *Advia Centaur* (mean cortisol of 351.7 in patients under 70 years vs 410.9 nmol/l in patients above 70 years, p=<0.001) and *Roche* (mean cortisol of 315.5 in patients under 70 years vs 401.4 nmol/l in patients above 70 years, p=<0.001) assays, but not in the *Architect* assay. Furthermore, post-menopausal women had a higher baseline cortisol compared to pre-menopausal women in the

Advia Centaur (mean cortisol of 380 vs 347 nmol/l, p=0.02) and *Roche* assays (mean cortisol of 364 vs 310 nmol/l, p=0.007), but not in the *Architect* assay.

ECONOMIC ASPECTS

The costs for *Synacthen* have varied over time, and in the UK this cost has recently increased substantially: currently, the price to the National Health Service for a single *Synacthen* ampoule is £45.71, with additional costs for nursing and medical supervison (varying with level of experience).

In the three populations analysed we calculated that the number of SSTs that could have been saved according to the given specificity and sensitivity baseline cortisol thresholds would have been 589 for *Advia Centaur* (*Siemens*), 158 for *Architect* (*Abbott*) and 578 for the *Roche Modular system* (*Roche*). Given these numbers, performing the basal cortisol and using the proposed cut-offs could lead to a significant cost saving.

DISCUSSION

In the present study we have confirmed that there is a strong correlation between basal morning cortisol and the response to *Synacthen* administration, as previously described (8, 23, 26). For the first time we have evaluated three commonly-employed cortisol assays and large patient data-base and we suggest basal levels of serum cortisol which strongly predict a normal response to *Synacthen*, and thus could be used to avoid unnecessary dynamic testing. Our results also confirm the

variability in cortisol assessment dependent on the assay used (23), and thus the inherent inaccuracy in specifying normative data on cortisol levels without specify the assay used.

Previous studies have shown that a basal cortisol of <100 nmol/l is highly predictive of adrenal insufficiency (AI), (22, 26, 29, 30) but there is no consensus in the literature as to a level of basal morning cortisol that can accurately predict adrenal sufficiency (20, 22, 30, 31), at least in part due to the use of different cortisol assays.

Interestingly, a previous study (23) showed that baseline morning cortisol levels that predict the outcome of SST are significantly lower in patients exposed to glucocorticoid (GC) therapy (baseline cortisol of 410 nmol/l showed 100% specificity for predicting passing the SST and a serum cortisol concentration of 34 nmol/l had 100% sensitivity for predicting failure) and even more in the subgroup of patients taking inhaled GCs (baseline morning cortisol of 348 nmol/l gave 100% specificity for predicting passing the SST and of 34 nmol/l showed 100% sensitivity for predicting failure) compared to all patients that had SST performed for all indications (serum cortisol concentration of 506 nmol/l was 100% specific for predicting passing the SST and of 107 nmol/l was 99% sensitive for predicting failure).

Our results confirm previous data concerning the usefulness of morning cortisol level in the assessment of adrenal reserve (23), but this is the first study that

analysed morning cortisol level of three commonly used cortisol assay to predict SST outcome. We recommend that to assess the HPA axis one should perform a morning basal cortisol as the first diagnostic test when suspecting adrenal insufficiency. Considering that AI is a condition with serious consequences if undiagnosed, we have investigated the basal cortisol with a 100% specificity in excluding failure to the SST, and hence AI. Thus, a morning basal cortisol level ≥358 nmol/l for *Advia Centaur* (Siemens), ≥336 nmol/l for *Architect* (Abbott) and ≥506 nmol/l for the *Roche Modular System* (Roche), the last one as previously reported (23), were highly predictive of adrenal sufficiency in this cohort of more than 3000 SSTs. These criteria appeared to be independent of disease or indication for the SST, except for the data on patients on GC treatment in the *Roche* assay as previously reported (23). However, it should be noted that the *Advia Centaur* assay has undergone several 'realignments', although it is unclear as to whether this has affected its threshold levels. In addition, the Roche assay has been replaced by Roche II, for which preliminary dta show broad alignment with the *Architect* assay, and thus a normative threshold in the region of 420-430nmol/L.

Overall, using a baseline morning cortisol as the first step and utilising the proposed cut-offs, 29% of SST could have been avoided leading to a significant decrease of unnecessary tests for patients, work and efforts for nursing and medical staff and relevant cost savings.

Limitations

The principal limitations of this study are the retrospective nature of data collection and the institutional referral bias that may lead to a potential positive selection bias that may over-estimate the prevalence rate of severe conditions and the percentage of patient failing the SST. Furthermore, a substantial number of the patients from Birmingham, using the *Roche* assay, underwent the SST after 10.00h. It should be noted, however, that the normative values for the SST are independent of the time of day.

However, even though most patients will be diagnosed with high accuracy using these morning baseline cortisol criteria, clinical judgement is essential in deciding when to totally rely on such values. Furthermore, while the SST is extremely useful in defining an adequate reserve of the HPA axis in most situations, there may still be situations where a complete test of the HPA axis is necessary, and then the ITT or related investigation should be used. It may aso be noted that while the SST closely correlates with the ITT, the latter essentially tests the whole HPA axis and thus represents a more appropriate biological response (32).

In conclusion, the morning basal cortisol is an easy to perform, cheap and safe test and represents a valid tool for the initial assessment of adrenal function and of HPA axis integrity. It reduces the use of unnecessary dynamic function tests with an associated reduction in cost and risks of adverse reactions for patient. Thus, we recommend the use of the morning baseline cortisol as the first test in the evaluation of patients with suspected adrenal insufficiency, and we provide

normative values from a large cohort of patients evaluating three commonly used cortisol assays to minimise the use of the SST and save time and resources.

REFERENCES

1. Greenwood FC, Landon J, Stamp TC. The plasma sugar, free fatty acid, cortisol, and growth hormone response to insulin. I. In control subjects. J Clin Invest. 1966;45(4):429-36.
2. Jacobs HS, Nabarro JD. Tests of hypothalamic-pituitary-adrenal function in man. Q J Med. 1969;38(152):475-91.
3. Edwards CR, Besser GM. Diseases of the hypothalamus and pituitary gland. Clin Endocrinol Metab. 1974;3(3):475-505.
4. Arlt W, Allolio B. Adrenal insufficiency. Lancet. 2003;361(9372):1881-93.
5. Hurel SJ, Thompson CJ, Watson MJ, Harris MM, Baylis PH, Kendall-Taylor P. The short Synacthen and insulin stress tests in the assessment of the hypothalamic-pituitary-adrenal axis. Clin Endocrinol (Oxf). 1996;44(2):141-6.
6. Jones SL, Trainer PJ, Perry L, Wass JA, Bessser GM, Grossman A. An audit of the insulin tolerance test in adult subjects in an acute investigation unit over one year. Clin Endocrinol (Oxf). 1994;41(1):123-8.
7. Gleeson HK, Walker BR, Seckl JR, Padfield PL. Ten years on: Safety of short synacthen tests in assessing adrenocorticotropin deficiency in clinical practice. J Clin Endocrinol Metab. 2003;88(5):2106-11.
8. Agha A, Tomlinson JW, Clark PM, Holder G, Stewart PM. The long-term predictive accuracy of the short synacthen (corticotropin) stimulation test for assessment of the hypothalamic-pituitary-adrenal axis. J Clin Endocrinol Metab. 2006;91(1):43-7.

9. Stokes FJ, Bailey LM, Ganguli A, Davison AS. Assessment of endogenous, oral and inhaled steroid cross-reactivity in the Roche cortisol immunoassay. Ann Clin Biochem. 2014;51(Pt 4):503-6.

10. Stewart PM, Corrie J, Seckl JR, Edwards CR, Padfield PL. A rational approach for assessing the hypothalamo-pituitary-adrenal axis. Lancet. 1988;1(8596):1208-10.

11. Lindholm J, Kehlet H, Blichert-Toft M, Dinesen B, Riishede J. Reliability of the 30-minute ACTH test in assessing hypothalamic-pituitary-adrenal function. J Clin Endocrinol Metab. 1978;47(2):272-4.

12. Lindholm J, Kehlet H. Re-evaluation of the clinical value of the 30 min ACTH test in assessing the hypothalamic-pituitary-adrenocortical function. Clin Endocrinol (Oxf). 1987;26(1):53-9.

13. Hagg E, Hassel U, Lithner F, Asplund K. [Hypothalmo-pituitary-adrenal function--can a short ACTH-test replace insulin tolerance test?]. Lakartidningen. 1981;78(45):4027-30.

14. Kane KF, Emery P, Sheppard MC, Stewart PM. Assessing the hypothalamo-pituitary-adrenal axis in patients on long-term glucocorticoid therapy: the short synacthen versus the insulin tolerance test. QJM. 1995;88(4):263-7.

15. Courtney CH, McAllister AS, Bell PM, McCance DR, Leslie H, Sheridan B, et al. Low- and standard-dose corticotropin and insulin hypoglycemia testing in the assessment of hypothalamic-pituitary-adrenal function after pituitary surgery. J Clin Endocrinol Metab. 2004;89(4):1712-7.

16. Abdu TA, Elhadd TA, Neary R, Clayton RN. 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. J Clin Endocrinol Metab. 1999;84(3):838-43.

17. Bangar V, Clayton RN. How reliable is the short synacthen test for the investigation of the hypothalamic-pituitary-adrenal axis? Eur J Endocrinol. 1998;139(6):580-3.

18. Stewart PM, Clark P. The short Synacthen test: is less best? Clin Endocrinol (Oxf). 1999;51(2):151-2.

19. Varadhan L, Nayak AU, Mukherjee A, Jose B, Varughese GI. Can a baseline morning cortisol predict outcome of short Synacthen test in an endocrine unit in an outpatient setting. Clin Endocrinol (Oxf). 2015;82(2):309-11.

20. Karaca Z, Tanriverdi F, Atmaca H, Gokce C, Elbuken G, Selcuklu A, et al. Can basal cortisol measurement be an alternative to the insulin tolerance test in the assessment of the hypothalamic-pituitary-adrenal axis before and after pituitary surgery? Eur J Endocrinol. 2010;163(3):377-82.

21. Mukherjee JJ, de Castro JJ, Kaltsas G, Afshar F, Grossman AB, Wass JA, et al. A comparison of the insulin tolerance/glucagon test with the short ACTH stimulation test in the assessment of the hypothalamo-pituitary-adrenal axis in the early post-operative period after hypophysectomy. Clin Endocrinol (Oxf). 1997;47(1):51-60.

22. Hagg E, Asplund K, Lithner F. Value of basal plasma cortisol assays in the assessment of pituitary-adrenal insufficiency. Clin Endocrinol (Oxf). 1987;26(2):221-6.

23. Woods CP, Argese N, Chapman M, Boot C, Webster R, Dabhi V, et al. Adrenal suppression in patients taking inhaled glucocorticoids is highly prevalent and management can be guided by morning cortisol. Eur J Endocrinol. 2015;173(5):633-42.

24. Le Roux CW, Meeran K, Alaghband-Zadeh J. Is a 0900-h serum cortisol useful prior to a short synacthen test in outpatient assessment? Ann Clin Biochem. 2002;39(Pt 2):148-50.

25. Grinspoon SK, Biller BM. Clinical review 62: Laboratory assessment of adrenal insufficiency. J Clin Endocrinol Metab. 1994;79(4):923-31.

490 26. Yo WS, Toh LM, Brown SJ, Howe WD, Henley DE, Lim EM. How good is a morning
491 cortisol in predicting an adequate response to intramuscular synacthen stimulation? Clin
492 Endocrinol (Oxf). 2014;81(1):19-24.
493 27. El-Farhan N, Pickett A, Ducroq D, Bailey C, Mitchem K, Morgan N, et al. Method-specific
494 serum cortisol responses to the adrenocorticotrophin test: comparison of gas
495 chromatography-mass spectrometry and five automated immunoassays. Clin Endocrinol
496 (Oxf). 2013;78(5):673-80.
497 28. Wang TW, Wong MS, Smith JF, Howlett TA. The use of the short tetracosactrin test for
498 the investigation of suspected pituitary hypofunction. Ann Clin Biochem. 1996;33 (Pt 2):112-
499 8.
500 29. Watts NB, Tindall GT. Rapid assessment of corticotropin reserve after pituitary
501 surgery. JAMA. 1988;259(5):708-11.
502 30. Erturk E, Jaffe CA, Barkan AL. Evaluation of the integrity of the hypothalamic-pituitary-
503 adrenal axis by insulin hypoglycemia test. J Clin Endocrinol Metab. 1998;83(7):2350-4.
504 31. Kadiyala R, Kamath C, Baglioni P, Geen J, Okosieme OE. Can a random serum cortisol
505 reduce the need for short synacthen tests in acute medical admissions? Ann Clin Biochem.
506 2010;47(Pt 4):378-80.
507 32. Grossman A. Assessment of the HPA axis: another new test? Endocrine 2015;50:268-9.
508

509 # TABLES
510
511 Table 1: Results of SST

	Advia Centaur (Siemens)			Architect (Abbott)		
Indication	Total (n)	Pass % (n)	Fail % (n)	Total (n)	Pass % (n)	Fail % (n)
ALL	1019	86.9% (886)	13.1% (133)	449	80.2% (360)	19.8% (89)
Pituitary diseases	390	82.1% (320)	17.9% (70)	156	75% (117)	29% (35)
Other Tumours of the CNS	119	89.1% (106)	10.9% (13)	66	84.8% (56)	15.2% (10)

Adrenal Disease (Addison's Disease, CAH, Adenoma, Carcinoma, Metastases)	25	60% (15)	40% (10)	26	42.3% (11)	57.7% (15)
Co-Existent Autoimmune Disease (Type 1 Diabetes Mellitus, Thyroid Disease, Premature Ovarian Failure, Vitiligo)	58	96.6% (56)	3.4% (2)	13	100% (13)	
Chronic Steroid Treatment For Disease Different From Endocrine Conditions	40	30% (12)	70% (28)	32	43.8% (14)	56.3% (18)
Other (Hyponatraemia, Hyperkalaemia, Hypoglycaemia, Postural Hypotension, Syncope, Collapse Or Dizziness, Weight loss, Nausea, Vomiting, Diarrea or Abdominal Pain, Fatigue, Malaise)	387	97.4% (377)	2.6% (10)	156	95.5% (149)	4.5% (7)

Table 2 Accuracy of baseline cortisol in predicting SST's results in all patients and according to indication to perform SST

	Advia Centaur (Siemens)								**Architect (Abbott)**								
			Specificity			Sensitivity					Specificity			Sensitivity			
Indication	n	AUC 95% CI	95%	99%	100%	95%	99%	100%	n	AUC 95% CI	95%	99%	100%	95%	99%	100%	n
ALL pts	1019	0.960 0.947-0.973	277	344	358	185	123	56	449	0.872 0.831-0.913	295	330	336	158	116	83	2050
Pituitary diseases	390	0.965 0.946-0.984	245	287	358	168	122	83	156	0.925 0.876-0.975	277		330	157	106	96	692
Other Tumours of the CNS	119	0.964 0.927-1			289	162	102	81	66	0.903 0.820-0.986			297	215		70	315
Adrenal diseases	25	0.927 0.826-1			314	212		166	26	0.870 0.725-1			308			115	73
Co-Existent Autoimmune diseases	58	0.946 0.874-1			263	236	86	73	13		No pathological pts						113
Chronic Steroid treat	40	0.982 0.950-1	272		323			211	32	0.770 0.606-0.934			336			124	407
Other	387	0.896 0.830-0.963			345	199	135	48	156	0.918 0.825-1			251	158	118	113	452

Legend: Pituitary diseases comprises postoperative assessment after pituitary surgery (followed and not followed by radiotherapy), pituitary adenoma treated conservatively, other diseases affecting the pituitary gland such as hypophysitis, apoplexy, cyst
Adrenal disease comprises: Addison's disease, CAH, adenoma, carcinoma, metastases;
Co-existent autoimmune disease comprises: type 1 diabetes mellitus, thyroid disease, premature ovarian failure, vitiligo;
Chronic steroid treat is chronic steroid treatment for disease different from endocrine conditions;
Other comprises: hyponatraemia, hyperkalaemia, hypoglycaemia, postural hypotension, syncope, collapse, dizziness, weight loss, nausea, vomiting, diarrea, abdominal pain, fatigue, malaise. pts= patients

Table 3: Comparison between before 10.00h and after 10.00h cortisol in terms of results of SST, accuracy of baseline cortisol in predicting SST's results, False Positive (FP) and PPV (Positive Predicting Value)

	BEFORE 10.00h	**AFTER 10.00h**
n	1019	443
Pass % (n)	86.9% (886)	91.6% (406)
Fail % (n)	13.1% (133)	8.4% (37)
AUC (95% CI)	0.960 (0.947-0.973)	0.944 (0.914-0.975)
Specificity 95%	277 nmol/l	314 nmol/l
Specificity 99%	344 nmol/l	--
Specificity 100%	358 nmol/l	376 nmol/l
Sensitivity 95%	184.5 nmol/l	151 nmol/l
Sensitivity 99%	123 nmol/l	62 nmol/l
Sensitivity 100%	56 nmol/l	35 nmol/l
FP (n)	346	162
PPV (%)	27.7	18.6

Supplementary Table 1: Accuracy of baseline cortisol in predicting SST's results according to age and sex

		Advia Centaur (Siemens)				**Architect (Abbott)**		
		Specificity	Sensitivity			Specificity	Sensitivity	

	n	AUC 95% CI	95%	99%	100%	95%	99%	100%	n	AUC 95% CI	95%	99%	100%	95%	99%	100%	n
Age < 30	153	0.954 0.914-0.994			287.5	186.5	103.5	88	64	0.928 0.859-0.997			212.5	141		109.5	232
Age 30-49	316	0.952 0.926-0.978	244.5		277.5	166	98	55	139	0.781 0.662 0.900	318.5		329.5	168.5	126.5	82.5	578
Age 50-69	344	0.960 0.937-0.983	287.5	320	344	184.5	129.5	82	161	0.872 0.805-0.939	291.5		335.5	147.5	114.5	110.5	689
Age 70-100	206	0.983 0.965-1.000			357.5	262.5	148	120	85	0.939 0.889-0.990	278.5		328.5	226.5		82.5	551
Male	416	0.957 0.935-0.978	296	345.5	357.5	197.5	123.5	54	195	0.903 0.858-0.948	278		318.5	151.5	122.5	77.5	910
Female	603	0.966 0.951-0.981	250.5	285.5	287.5	178	120.5	72	254	0.833 0.762-0.904	328.5		335.5	160	114.5	92.5	1140
Pre-menopause	317	0.940 0.907-0.973	276.5		287.5	169.5	104.5	67.5	143	0.824 0.717-0.934	296.5		329.5	168.5	118	107.5	503
Post-menopause	286	0.985 0.974-0.997	248.5		260.5	197.5	126	82	111	0.836 0.740-0.931	329.5		335.5	123.5	110.5	92.5	637

Figure 1: Baseline serum cortisol as a predictor of SST outcome in three cortisol immunoassay. Baseline serum cortisol is graphed against the % likelihood of passing (specificity: continous line) or failing (sensitivity: dashed line) the SST.