BMJ Open

# Importance of sampling time and assay cut-offs for routine assessment of adrenal function: an observational longitudinal study.



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2017-019273 |
| Article Type: | Research |
| Date Submitted by the Author: | 22-Aug-2017 |
| Complete List of Authors: | Ortiz-Flores, Andrés; Hospital Universitario Ramón y Cajal, Endocrinology<br>Santacruz, Elisa; Hospital Universitario Ramon y Cajal<br>Jiménez-Mendiguchia, Lucía; Hospital Universitario Ramon y Cajal<br>García-Cano, Ana; Hospital Universitario Ramon y Cajal<br>Nattero-Chávez, Lía; Hospital Universitario Ramon y Cajal<br>Escobar-Morreale, Héctor; Hospital Universitario Ramon y Cajal<br>Luque-Ramírez, Manuel; Hospital Universitario Ramon y Cajal |
| Primary Subject Heading: | Diabetes and endocrinology |
| Secondary Subject Heading: | Diagnostics, Patient-centred medicine |
| Keywords: | Adrenal disorders < DIABETES & ENDOCRINOLOGY, cortisol, biochemical diagnosis, specificity, immunoassay |
| | |

**Title** Importance of sampling time and assay cut-offs for routine assessment of adrenal function: an observational longitudinal study.

**Short title** Cosyntropin test in adrenal insufficiency.

**Authors** Andrés E. Ortiz-Flores, M.D.,[1] Elisa Santacruz-Cerdá, M.D,[1]Lucía Jiménez-Mendiguchia, M.D.,[2] Ana García-Cano, M.D.,[2] Lia Nattero Chávez, M.D.,[1] Héctor F. Escobar-Morreale, M.D., Ph.D.,[1,3] Manuel Luque-Ramírez, M.D., Ph.D.[1,3]

**Affiliations**

[1] Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Madrid (Spain).

[2] Department of Clinical Biochemistry, Hospital Universitario Ramón y Cajal, Madrid (Spain).

[3] Diabetes, Obesity and Human Reproduction Research Group, Hospital Universitario Ramón y Cajal &Universidad de Alcalá & Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS) & Centro de Investigación Biomédica en Red Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain.

**Word count:** 3,506**; Word count for abstract:** 288; **Figures:** 4; **Tables:** 2; **Number of references:** 21.

**CORRESPONDENCE & REPRINTS:** Andres E. Ortiz-Flores, MD. Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Carretera de Colmenar km 9.1, E-28034 Madrid, Spain. Phone/Fax: +34-91 336 9029. E-mail: andres_ortiz_f@yahoo.com

## ABSTRACT

**Objectives** Serum cortisol concentrations after adrenal stimulation with a high-dose cosyntropin bolus is the test of choice for diagnosis of primary and non-acute central adrenal insufficiency. We aim to: i) assess the role of 30-min and 60-min sample timing, and the importance of assay-specific normative cut-offs concentrations for adrenal insufficiency diagnosis, ii) to estimate specificity and positive predictive value of 30-min and 60-min sampling time, and iii) to establish an assay-specific lower limit of normality of serum cortisol concentrations after cosyntropin stimulation.

**Design and Setting:** Observational retrospective study performed in a tertiary-level Spanish hospital between 2011 and 2015.

**Participants and interventions:** Two groups were evaluated: i) A main study group that m 406 patients in whom serum cortisol was measured at 30 and 60 minutes after cosyntropin stimulation, and ii) a confirmative group that included 153 women with a normal hypothalamic-pituitary-adrenal axis in whom a cosyntropin test was conducted for other reasons. Diagnostic agreement between sampling times was analysed considering classic (500 nmol/l) and assay-specific serum cortisol cut-off concentrations.

**Results** Diagnostic agreement was greater when applying assay-specific cut-off values instead of those derived from the literature. For suspected primary adrenal insufficiency, serum cortisol measured 30-min after cosyntropin administration was enough to make a diagnosis in over 95% of cases, without missing any necessary treatment. For central adrenal insufficiency suspicion, 60-min cortisol concentrations were more specific, establishing diagnosis in over 97% of cases.

**Conclusions:** Assay-specific cut-off cortisol concentrations instead of classic literature values improve the diagnostic accuracy of the cosyntropin test. For primary adrenal insufficiency, 30-

min subnormal serum cortisol responses are a reliable marker of adrenal function. On the contrary, when secondary or tertiary adrenal insufficiency is suspected, a 60-min cortisol measurement improves the diagnostic accuracy of the test.

**Strengths and limitations:**

A large series of subjects with suspected AI evaluated with a standardized dynamic study, in whom, a systematic review of their clinical recordings was performed.

The present study enhances the importance of the use of local normative thresholds for adrenal function assessment, situation than in the clinical practice is rarely considered among physicians.

Our present results may not be extrapolable to other populations in whom SC has been measured with different immunoassays that would require different local normative data.

Analysis of specificity and positive predictive value has not been challenged against a biochemical gold-standard in most cases and, we have not been able to establish false negative rates, sensitivity and negative predictive values

**Keywords:** Adrenal insufficiency; biochemical diagnosis; cortisol; immunoassay; specificity.

## Introduction

The laboratory diagnosis of adrenal insufficiency (AI) at the clinical setting lies on the finding of an inappropriately low morning circulating serum cortisol (SC) or subnormal SC responses to adrenal stimulation (1). Classic adrenal stimulation protocol consists of SC measurements at 30 and 60 minutes after a single 250 μg intravenous bolus or intramuscular injection of tetracosactide (cosyntropin). A SC value 30 and/or 60-min after cosyntropin administration ≥ 500 nmol/l (18 μg/dl) is considered as a normal response (2). This protocol, also known as a short standard high-dose test (SST), is the dynamic exploration of choice for primary AI diagnosis (1, 3) and it is also used for non-acute central AI (4, 5). In critically ill patients, SST may be performed to rule out a functional form of AI −critical illness-related corticosteroid insufficiency− in subjects showing sustained refractory hypotension and no response to vasopressors (2, 6). Although a definite biochemical definition for this condition is lacking, a SC increase above 248 nmol/l (9 μg/dl) in response to cosyntropin is associated with a good prognosis in septic patients (7).

The most appropriate sampling time for SC during the SST is controversial. SC measurements 30 min after SST have been validated against a "gold standard" such as an insulin tolerance test (8). Thus, some authors suggest that a single 30 min SC is enough to establish or rule out a clinically significant AI (4, 9, 10). Other studies show that a 60 min sample may avoid unnecessary overdiagnosis (11-13). Recent clinical practice guidelines recommend further research to clarify whether 60 min SC sampling might be more specific for AI diagnosis (3, 14).

Liquid chromatography/mass spectrometry techniques are currently recommended for the accurate measurement of circulating steroids. But in most centres, clinical routine still relies on automated immunoassays for SC measurement (15). Considering that the classic cut-off value

for the SST was established for SC as measured by older radioimmunoassays, and that immunochemiluminescent assays differ in antibody specificity with radioimmunoassays (16), establishing local assay-specific cut-off values is of paramount importance to properly classify SC responses to cosyntropin (3, 16, 17). This fact is not a spurious one because despite being well established that local assay-specific lower limits of normality (LLN) should be used for dynamic assessments of the HPA axis (3), in our experience many physicians are still using classic cut-offs in their routine practice. Also, other factors that may influence SC measurement include the stimulation of hepatic synthesis and secretion of cortisol binding globulin by oestrogens, sex and several non-glucocorticoid drugs (18).

To provide new insights into still open questions, our study's aims were: i) to assess the concordance between 30 and 60 min SC concentrations after SST at the clinical setting; ii) to estimate the diagnostic agreement between both sampling times when using literature or assay- and sex-specific cut-offs values, taking into account the origin of AI; iii) to estimate the specificity (Sp) and positive predictive value (PPV) of 30 and 60 min sampling times while taking into account the origin of AI; and iv) to confirm assay-specific LLN for SC concentration after SST in a group of subjects with a normal hypothalamic-pituitary-adrenal (HPA) function.

## Subjects and methods

From January 1, 2011 to December 31, 2015 we conducted a longitudinal observational study, performed in a third-level Spanish hospital, where we assessed SC responses to SST in two study populations:

i) A subgroup of adult subjects (n = 451) in whom 0, 30 and 60 min SC concentrations were assayed during a SST conducted at the clinical setting for suspected AI (main study population).

ii) A group of women with normal HPA axis (n = 153) prospectively recruited from our Reproductive Endocrinology clinic during the study of functional hyperandrogenism whom 0 and 30 min SC concentrations were obtained during a SST performed for routine screening of non-classic congenital adrenal hyperplasia by a local study protocol including SC values at those sampling times (confirmative group). Non-classic congenital adrenal hyperplasia screening was negative in all cases.

Before conducting the study, we obtained approval from the local ethics committee. All women from our Reproductive Endocrinology clinic had previously signed an informed consent form for the inclusion of a selection of coded clinical variables in an electronic database for clinical research purposes that included the SC measurements presented here.

*Main study population*

Basal and stimulated SC values were extracted from the electronic database of our Department of Clinical Biochemistry. We collected a minimum dataset in an electronic case form from the clinical records of the patients including age, sex, weight, height, laboratory measurements at the dates when the SST was conducted such as circulating electrolytes, glomerular filtration rate and basal ACTH concentrations, clinical suspicion of primary or central AI, other dynamic tests performed for the evaluation of adrenal function, history of pituitary disease, time from hypothalamic-pituitary insult to SC determination, administration of drugs that may interfere with the HPA axis, and the immunoassay used for SC assay. Baseline characteristics of study population are shown in **Table 1**.

We considered a clinical suspicion of primary AI in cases when the patient had some known adrenal disease, had required mineralocorticoid supplementation latter in their follow-up, had received drugs that may interfere with cortisol biosynthesis, had not clinical suspicion of any

hypothalamic-pituitary condition, and had not developed such a condition latter in time. Conversely, we suspected a central AI in subjects known to suffer from hypothalamic-pituitary disease, or had received drugs that may suppress the HPA axis.

We excluded from analysis those subjects stimulated with cosyntropin doses other than 250 µg (n = 4), subjects aged below 18-year old (n = 35), and subjects from whom we could not obtain enough information from their clinical records as to establish a reason for conducting a SST (n = 6). Therefore, the study group finally included in the analyses consisted of 406 subjects.

*Confirmative group*

The results of SST from 153 premenopausal women with a normal HPA axis aged from 14 to 42 years old were included. Three women showed clearly subnormal SC responses and were consequently excluded from the study: in two, the HPA axis suppressive effect of progestins administered during 10 days before the SST with the aim of inducing a withdrawal bleed could justify the abnormal results; in the other case, we could not establish the cause of the subnormal response with certainty because the patient was lost to follow-up.

*Assays*

During the study period, two immunoassays were used in our centre: i) *Siemens Immulite 2000© Cortisol Immunoassay System* from 2011 to July 1, 2013 (immunoassay 1) and ii) *Abbot Laboratories Diagnostics Division Architect© Cortisol Immunoassay System* from 2013 August 1, 2013 to December 31, 2015 (immunoassay 2).

*Agreement among 30 and 60 min sampling times*

We analysed diagnostic agreement between the 30 and 60 min SC using two different stimulated SC LLN: i) the classic ≥ 500 nmol/l (3), and ii) sex and assay-specific cut-off values taking into account concomitant COC use if necessary (14). For immunoassay 1, the reported LLN (2.5$^{th}$ percentile) was 470 nmol/l (17 µg/dl) in men and women, and 690 nmol/l (25 µg/dl) for women taking COC. For immunoassay 2, the LLNs were 441 nmol/l (16 µg/dl) for men, 414 nmol/l (15 µg/dl) for women, and 579 nmol/l (21 µg/dl) for women taking COC (16).

*Statistical analysis*

Data are shown as mean ± standard deviation or 95% confidence interval (CI), median (minimum-maximum), and raw numbers (percentage) as needed. Normal distribution of continuous variables was assessed by the Kolmogorov-Smirnov test for one sample after a two-step approach for transforming skewed variables if necessary (19). Comparisons among continuous variables were performed by the Student's t test or a repeated-measure ANOVA. Comparisons among categorical variables were performed by Fisher's exact or $\chi^2$ tests as appropriate. Pearson's analysis was used to correlate 30 and 60-min sampling times. Consistency and absolute agreement among both point times of SST were determined by their intraclass correlation coefficient (ICC) with a two-factor and random-effect model. Quantitative agreement was graphically assessed by Bland-Altman plots. Biochemical agreement in the diagnosis of normal or subnormal adrenal was assessed by using the kappa (κ) coefficient. True positives (TP) were defined as SSTs showing subnormal cortisol responses at both time points. True negatives (TN) were defined as SSTs showing a normal cortisol response at both time points in patients who did not need glucocorticoid replacement during their follow-up, did not suffer an adrenal crisis, and did not have any other functional HPA test with a subnormal response if

available. False positives (FP) were SSTs showing a subnormal response in only one of the sampling times but not in the other. We calculated Sp and PPV [Sp = TN / (TN + FP) and PPV = TP / (TP + FP)] for each sampling times after the SST. A *P* value < 0.05 was considered statistically significant.

## Results

*Main study population*

Of 406 cases included (261 tested by immunoassay 1 and 145 tested by immunoassay 2), 168 SSTs were performed to rule out primary AI; 226 to rule out central AI; and 12 SST were performed in critically ill patients. Regarding all SSTs as a whole, SC concentrations at 30 and 60 min after SST increased when compared to unstimulated values (**Figure 1A**), and SC concentrations at both 30 and 60 min showed a very strong linear correlation (**Figure 1B**). The ICC among both sampling times showed a very good consistence index (0.948; 95%CI: 0.937 – 0.957) and a good absolute agreement (0.899, 95%CI: 0.476 – 0.962), although according to the 95%CI lower limit only qualify as fair. The Bland-Altman plot (**Figure 1C**) showed a good agreement between 30 and 60 min sampling times, with a tendency towards greater differences with increasing mean values of stimulated SC and only ~ 5% of extreme differences among both times.

Diagnostic agreement among both times according to classic and to assay-specific cut-off values is shown in **Figure 2.** With a classic cut-off point, 42 cases (10.3%) had a subnormal response at 30 min, whereas it was ≥ 500 nmol/l at 60 min. On the contrary, the response was normal at 30 min in 5 cases (1.2%) but subnormal at 60 min. Using sex- and assay-specific values, 37 cases (9.1%) had a subnormal response at 30 min but normal at 60 min. In 7 cases (1.7%), the response was normal at 30 min and subnormal at 60 min.

The analysis of the diagnostic agreement as a function of central or primary AI suspicion is shown in **Figure 3**. As a rule, agreement among sampling times of the SST was better when primary AI was suspected compared with central AI suspicions. When using classic cut-off values to rule out primary AI, 8 cases (4.8%) showed a subnormal response at 30 min that reached normal concentrations at 60 min, whereas no subject with a normal response at 30 min had a subnormal response at 60 min. Using sex- and assay-specific cut-off values, in 7 cases (4.2%) the response was subnormal at 30 min but reached normal concentrations at 60 min. Five of them showed a subnormal SC response after SST that was very close to reaching the cut-off value. In these subjects, the differences between the cut-off value and the stimulated SC ranged from 22 to 39 nmol/l (0.8 to 1.4 µg/dl), very small concentrations that are in fact included within the coefficient of variation of the assay (18,19), thereby suggesting no clinical consequences. The two remaining patients showed peak SC concentrations of 303 and 360 nmol/l (11 and 13 µg/dl) at the 30 min sampling time: one had received oral glucocorticoid replacement therapy that did not preclude the patient of responding to cosyntropin by showing a SC of 470 nmol/l (17 µg/dl) at the 60 min sample, and the other subject was submitted to SST because of the presence of bilateral adrenal hyperplasia and did not show any signs or symptoms of AI nor suffered an adrenal crisis during follow-up. None of the SSTs showing normal responses at 30 min had a subnormal response at 60 min.

When central AI was suspected and a classic cut-off point was applied, 33 cases (14.6%) had a normal response at 60 min and subnormal at 30 min. Only 3 subjects (1.3%) presented with the opposite situation. Regarding specific cut-offs, 30 cases (13.3%) had a normal response at 60 min but subnormal at 30 min, and in only 5 cases (2.2%) the contrary occurred. These 5 subjects had been evaluated in the context of withdrawal of prolonged glucocorticoid therapy during the first year after a pituitary injury (surgery and/or pituitary radiotherapy). Three of them

showed a complete recovery of their HPA axis throughout the follow-up period, whereas in the other two patients, who had received pituitary radiotherapy, the subnormal response to cosyntropin was maintained over time.

The Sp and PPV for different sampling times and cut-off values used here are shown in **Table 2**. SC concentrations at 60 min had a higher Sp and PPV compared with 30 min measurements, particularly when central AI was suspected. Nonetheless, the Sp of the determination at 30 min was as high as 95% when SST had been performed to rule out primary disease both when applying classic or sex- and assay-specific cut-off values.

We observed discordant results between classic and specific cut-off values in 50 cases. In 47 of these subjects the subnormal response observed considering the classic cut-off value was normal if a sex- and assay-specific cutoff was applied. In 7 of them, SST was performed to rule out a primary AI and the remaining 40 SSTs were performed to rule out a central AI. Glucocorticoid replacement was started in 18 cases, and no subject presented with signs or symptoms of chronic or acute AI. Of the 50 discordant SSTs, 3 were conducted in women under estrogenic therapy and presented a normal response according to the classic cut-off value, but subnormal when considering a sex- and assay-specific cut-off, yet none of them required glucocorticoid therapy.

Finally, no discordant results were observed among sampling times and between classic and sex- and assay-specific cut-off values in critically ill patients.

*Confirmative group*

Of these women (n = 150, 97 tested by immunoassay 1 and 53 tested by immunoassay 2), 30 (20%) presented with a subnormal response to SST according to classic cut-off values, yet this figure was reduced to 3 (2%) when sex- and assay-specific cut-off values were used

[observed agreement: 82%; κ: 0.151 (95%CI: 0.066-0.235)]. The 3 women showing subnormal response during SST using a sex- and assay-specific cut-off value showed stimulated SC concentrations of 342 nmol/l (12.4 µg/dl), 353 nmol/l (12.8 µg/dl) and 372 nmol/l (13.5 µg/dl), whereas the LLNs (2.5$^{th}$ percentile) of SC concentrations at 30 min sampling time of SST were 441 and 414 nmol/l for immunoassays 1 and 2, respectively **(Figure 4)**. None of these female controls developed any HPA disease during their follow-up**.**

## DISCUSSION

AI is a clinical condition associated with a high morbidity and mortality. Unstimulated early morning SC values below 138 nmol/l (5 µg/dl) show a high PPV for AI, whereas concentrations over 500 nmol/l predict a normal adrenal response. However, values between 138 and 500 nmol/l are considered indeterminate and require adrenal stimulation to confirm or rule out a diagnosis (1-3).

Our data shows that both 30 and 60 min sampling times during SST have an adequate index of consistency, but the same is not true in terms of absolute agreement, particularly if central AI is suspected. Overall, a single determination at 60 min during the SST appears to have the higher Sp and PPV for the diagnosis of subjects presenting with either primary or central AI. In consonance, after evaluating retrospectively 73 subjects, Zueger et al.(20) reported that sampling at 30 min of the SST did not provide any additional diagnostic advantage over performing a single determination at 60 min of the test. Although similar results have been also reported by others (12, 13), these studies did not take into account the primary or central origin of AI and did not apply sex- and assay specific cut-off values, a fact of paramount importance because of the considerable influence that cortisol immunoassays exerts on the final values observed after SST (16, 17).

Our results also indicate that SC measurement at 30 min during the SST, when using sex- and assay-specific cut-off values, are enough to rule out clinically relevant primary AI since only 4.2% of patients in this particular situation showed a subnormal response at 30 min followed by normal response at 60 min. Furthermore, these subjects presented with stimulated SC concentrations which were very close to the cut-off point values, to the extent that the differences with these normal limits may be explained by the intrinsic variability of the commercial immunoassays. Even more important from a practice point of view, these subjects did not require replacement therapy during their follow-up, did not suffer an acute adrenal crisis, and were not diagnosed with any adrenal condition strongly suggesting that their HPA function was actually normal.

However, in line with abovementioned studies, the 60-min sampling time appears to be more specific than 30 min measurements when central AI is suspected. In such a case, 12% of subjects presenting with a subnormal response at 30 min had a normal response at 60 min, avoiding unnecessary treatments in them. Although a subnormal 30-min response in patients with suspicion of secondary AI may not translate adrenal replacement needs in a non-critical scenario, it is more than likely that many physicians feel more convenient with a stimulated value over the LLN for not beginning that therapy, and 60-min sampling time is mildly better in gauging that aim. Furthermore, the need of relying mostly on 60-min cortisol responses to cosyntropin when a central AI is also supported by the fact that in 2 out of 5 cases showing a subnormal response at 60 min but normal values at 30 min, an AI was confirmed latter in their follow-up due to former pituitary radiotherapy.

Our present findings also reinforce the need of sex and assay-specific cut-off values to interpret the results of the SST, in agreement with recent clinical guidelines (3). The use of such cut-offs leads to a reduction in FP results, higher Sp and PPV, less discordant results among

sampling times, and fewer unnecessary treatments [20 patients (5%) could have been treated unnecessarily if a classic cut-off value was applied for diagnosis]. The reliability of sex- and assay-specific cut-off values was confirmed in our population of premenopausal women with normal HPA axis, in whom these cut-offs were more appropriate than relying on classic values to assess the functionality of their HPA axis. In this population, the LLNs for 30 min stimulated SC were very close to those reported for each immunoassay by the manufacturers, yet reinforcing the need to establish local normative data in order to improve the diagnostic accuracy of cortisol measurements during SSTs (16).

Among the strengths of our study, we would highlight the large series of subjects suspected of AI who were evaluated with a standardized dynamic study, and the systematic review of subjects' clinical recordings that followed such evaluations. However, we are aware of several weaknesses derived from the observational and retrospective design of the study, making impossible to rule out information bias. Our best efforts might have not been enough to avoid misclassification of patients according to the suspicion of primary or central AI. Also, the administration of supraphysiological doses of cosyntropin does not permit ruling out partial deficiencies either, particularly in those suspected of central HPA defects. Also, and even considering the large sample of subjects included in our study, our present results may not be extrapolable to other populations in whom SC has been measured with different immunoassays that would require different local normative data. Moreover, analysis of Sp and PPV has not been challenged against a biochemical gold-standard in most cases and, we have not been able to establish false negative rates, sensitivity and negative predictive values. Nonetheless, besides those assessments had been unethical in most cases, the lack of a laboratory gold-standard such as an insulin tolerance test did not override our results, since from a practical point of view, we are looking for patients needing replacement therapy and not for those with a partial AI who do

not require any treatment. Lastly, we could not rule out entirely pre-treatment with progestogens in the context of induction of withdrawal bleeding in our confirmative population. Because these drugs might exert a mild suppressive effect on the HPA axis (18, 21), their administration in a few cases could have, at least in theory, lowered stimulated SC values.

## CONCLUSIONS

Compared with the use of classic cut-off values derived from the literature, application of sex- and assay-specific cut-off values of SC responses to cosyntropin results into higher Sp and PPV for establishing a diagnosis of AI, thereby avoiding unnecessary treatments. Measurement of stimulated SC at 30 min after SST may suffice for the correct diagnosis of primary AI, yet 60 min measurements might be preferable when central AI is suspected.

**Conflict of interest**: None.

**Funding**: This work has been supported by a grant Fondo de Investigación Sanitaria (PI1400649) from Instituto de Salud Carlos III, Spanish Ministry of Economy and Competitiveness. M.L.-R. has a local grant for clinical research from the Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS). CIBERDEM is also an initiative of Instituto de Salud Carlos III, partially supported by Fondo Europeo de Desarrollo Regional FEDER. There were no other sources of funding.

**Data sharing statement:** Individual participant data that underlie the results reported in this article, after deidentification, so as the study protocol would be available immediately after publication to anyone who wishes to access the data to achieve aims in the approved proposal

and for individual participant data meta-analysis. Proposals should be directed to andres_ortiz_f@yahoo.com or to manuel.luque@salud.madrid.org. To gain access, data requestors will need to sign a data access agreement

**Contributorship statement:** A.O.-F. y M.L.-R. designed the protocol, and performed the statistical analysis. A.O.-F. y E.S.-C. reviewed the clinical data using the electronic or written records if necessary. A.G.-C. and L.J.-M. performed the electronic search of serum cortisol samples. A.O.-F. y M.L.-R. wrote the first draft of the study. All the authors, including L.N.-C. and H.F.E.-.M, reviewed the manuscript before its submission and contributed to intellectual content. All the authors have accepted responsibility for the entire content of the manuscript and approved the final submission**.**

## REFERENCES


1. Charmandari E, Nicolaides NC, Chrousos GP. Adrenal insufficiency. Lancet. 2014;383(9935):2152-67.

2. Neary N, Nieman L. Adrenal insufficiency: etiology, diagnosis and treatment. Curr Opin Endocrinol Diabetes Obes. 2010;17(3):217-23.

3. Bornstein SR, Allolio B, Arlt W, Barthel A, Don-Wauchope A, Hammer GD, et al. Diagnosis and Treatment of Primary Adrenal Insufficiency: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2016;101(2):364-89.

4. Grossman AB. Clinical Review#: The diagnosis and management of central hypoadrenalism. J Clin Endocrinol Metab. 2010;95(11):4855-63.

5. 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.

6. Sprung CL, Annane D, Keh D, Moreno R, Singer M, Freivogel K, et al. Hydrocortisone therapy for patients with septic shock. N Engl J Med. 2008;358(2):111-24.

7. Annane D, Sebille V, Troche G, Raphael JC, Gajdos P, Bellissant E. A 3-level prognostic classification in septic shock based on cortisol levels and cortisol response to corticotropin. Jama. 2000;283(8):1038-45.

8. 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.

9. 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.

10. Gonzalbez J, Villabona C, Ramon J, Navarro MA, Gimenez O, Ricart W, et al. Establishment of reference values for standard dose short synacthen test (250 microgram), low dose short synacthen test (1 microgram) and insulin tolerance test for assessment of the hypothalamo-pituitary-adrenal axis in normal subjects. Clin Endocrinol (Oxf). 2000;53(2):199-204.

11. Longui CA, Vottero A, Harris AG, Chrousos GP. Plasma cortisol responses after intramuscular corticotropin 1-24 in healthy men. Metabolism. 1998;47(11):1419-22.

12. Mansoor S, Islam N, Siddiqui I, Jabbar A. 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 Med J. 2007;48(6):519-23.

13. 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):14-9.

14. de Miguel Novoa P, Vela ET, Garcia NP, Rodriguez MM, Guerras IS, Martinez de Salinas Santamaria Mde L, et al. Guidelines for the diagnosis and treatment of adrenal insufficiency in the adult. Endocrinol Nutr. 2014;61 Suppl 1:1-35.

15. Turpeinen U, Hamalainen E. Determination of cortisol in serum, saliva and urine. Best Pract Res Clin Endocrinol Metab. 2013;27(6):795-801.

16. El-Farhan N, Pickett A, Ducroq D, Bailey C, Mitchem K, Morgan N, 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):673-80.

17. Klose M, Lange M, Rasmussen AK, Skakkebaek NE, Hilsted L, Haug E, et al. Factors influencing the adrenocorticotropin test: role of contemporary cortisol assays, body composition, and oral contraceptive agents. J Clin Endocrinol Metab. 2007;92(4):1326-33.

18. Kontula K, Paavonen T, Luukkainen T, Andersson LC. Binding of progestins to the glucocorticoid receptor. Correlation to their glucocorticoid-like effects on in vitro functions of human mononuclear leukocytes. Biochem Pharmacol. 1983;32(9):1511-8.

19. Templeton G. A two step approach for transforming continuous variables to normal: Implications and recommendations for IS Research. Communications of the Association for Information Systems. 2011;28(4):41-58.

20. Zueger T, Jordi M, Laimer M, Stettler C. Utility of 30 and 60 minute cortisol samples after high-dose synthetic ACTH-1-24 injection in the diagnosis of adrenal insufficiency. Swiss Med Wkly. 2014;144(23):w13987.

21. Bornstein SR. Predisposing factors for adrenal insufficiency. N Engl J Med. 2009;360(22):2328-39.

**TABLE 1.** Baseline characteristics population as a function as clinical suspicion of primary adrenal disease, secondary adrenal disease or critical patient.

| | **Suspicion of primary AI** (n = 168) | | **Suspicion of Central AI** (n= 226) | | **Critically-illpatients** (n = 12) | |
|---|---|---|---|---|---|---|
| Sex | **Female** (n = 105) | **Male** (n = 63) | **Female** (n =141) | **Male** (n = 85) | **Female** (n = 5) | **Male** (n = 7) |
| Age *(years)* | 53 ± 19 | 59 ± 15 | 55 ± 16 | 55 ± 13 | 53 ± 21 | 45 ± 23 |
| Weight*(kg)* | 60 ± 12 | 71 ± 15 | 72 ± 14 | 84 ± 17 | 59 ± 13 | 66 ± 14 |
| BMI *($kg/m^2$)* | 24 ± 5 | 24 ± 5 | 28 ± 5 | 29 ± 5 | 23 ± 4 | 25 ± 4 |
| Na*(mmol/l)* | 138 ± 4 | 137 ± 5 | 139 ± 2 | 140 ± 4 | 141 ± 8 | 141 ± 9 |
| K*(mmol/l)* | 4.3 ± 0.6 | 4.4 ± 0.7 | 4.1 ± 0.4 | 4.1 ± 0.3 | 4.2 ± 0.7 | 4.2 ± 0.9 |
| Ca (*mmol/l)* | 2.3 ± 0.1 | 2.3 ± 0.2 | 2.4 ± 0.1 | 2.3 ± 0.1 | 2.2 ± 0.2 | 2.1 ± 0.2 |
| Cr*(µmol/l)* | 62 (44 – 1114) | 80 (44 – 1158) | 62 (18 – 875) | 80 (44 – 150) | 141 (35 – 856) | 97 (53 – 283) |
| GFR (MDRD) *($ml/min/1.73m^2$)* | 88 (4 – 137) | 85 (4 – 183) | 84 (5 – 361) | 93 (43 – 163) | 37 (4 –184) | 78 (20 – 132) |

*Abbreviations, BMI, body mass index; F, female; GFR, glomerular filtration rate; M, male; sCa, total serum calcium; sCr, serum creatinine; sK, serum potassium;sNA, serum sodium.*

Data are presented as mean ± standard deviation or median (minimum-maximum) according to their distribution. To convert serum calcium to conventional system units, multiply by 4 (result in mg/dl); to convert serum creatinine to conventional system units by 0.0113 (result in mg/dl).

**TABLE 2.** Specificity and positive predictive value (PPV) for diagnosis of adrenal insufficiency, after short high-dose cosyntropin test, according to serum cortisol cut-off points (classic and assay-specific), and as a function of suspected level of the defect.

| | Classic cut-off values | | | | | | Sex- and assay-specific cut-off values | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Global | | Clinical suspicion | | | | Global | | Clinical suspicion | | | |
| | | | Primary AI | | Central AI | | | | Primary AI | | Central AI | |
| **Sampling time (min)** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** |
| **Specificity (%)** | 87 | 98 | 95 | 100 | 79 | 98 | 89 | 98 | 95 | 100 | 83 | 97 |
| **PPV (%)** | 69 | 95 | 74 | 100 | 67 | 96 | 67 | 92 | 75 | 100 | 61 | 90 |

## Legend to figures

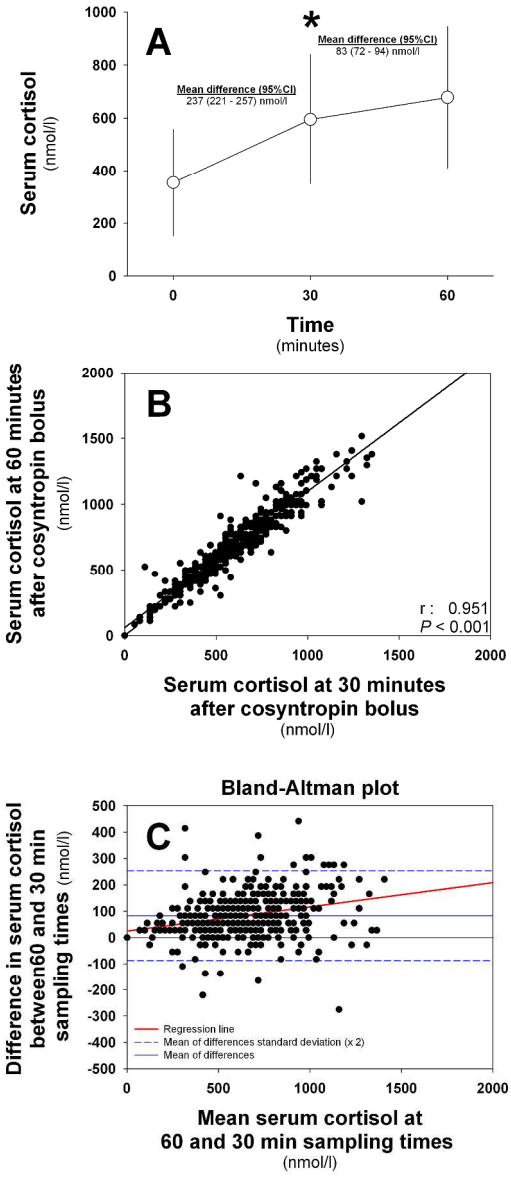
**Figure 1.** Panel A: Mean difference among 30 and 60 min time points after cosyntropin stimulation. Data are shown as mean ± standard deviation, and mean difference (95%CI). Comparisons among time points were performed by a repeated-measure ANOVA addressing main effects by a Bonferroni's confidence interval adjustment. Panel B: Pearson's correlation analysis between serum cortisol values at 30 and 60 min time points. Panel C: Bland-Altman plot. Solid black line represents the perfect agreement among both time points. Solid blue line is the mean of differences among both time points. Dashed blue lines are 2 standard deviation of the mean of differences. Solid red line is the regression line of mean differences.To convert serum cortisol to metric units, multiply by 0.03625 (result in µg/dl).

**Figure 2.**Subgroups of patients according to serum cortisol response after cosyntropin stimulation as a function of classic and assay-specific cut-offs. Figures on top of bars indicate the number of patients included in the subgroups. Diagnostic agreement is shown as the percentage of observed agreements and kappa coefficients (95%CI).

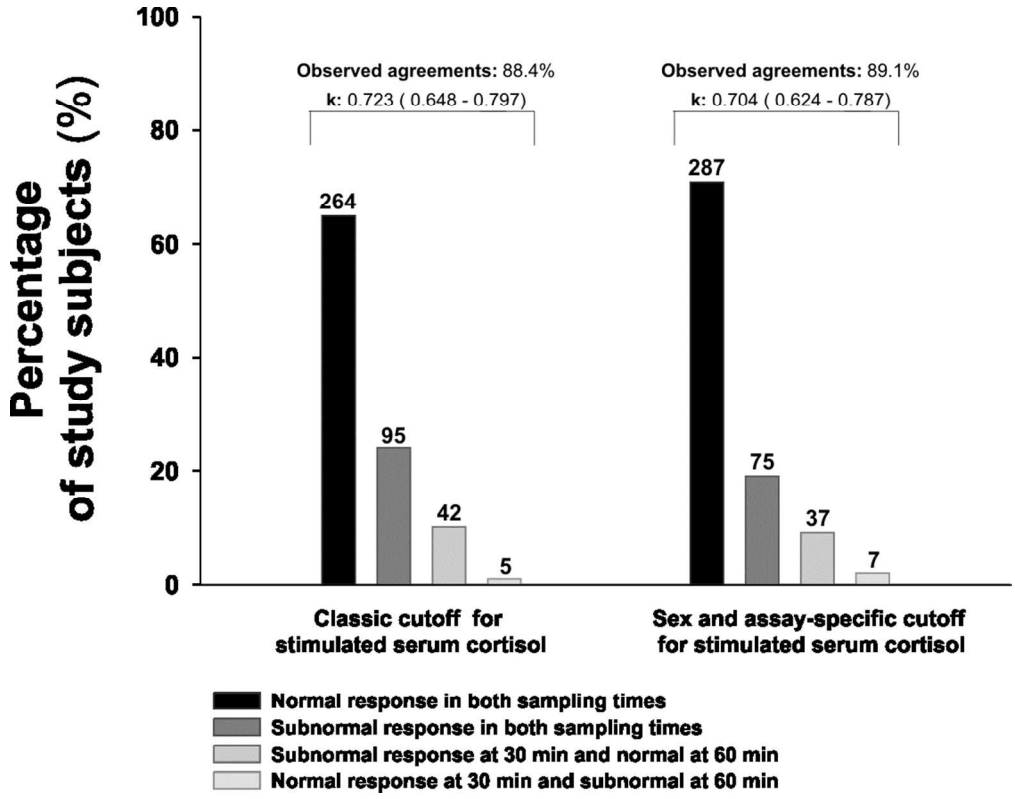
**Figure 3.** Subgroups of patients according to serum cortisol response after cosyntropin stimulation as a function of cut-off values and suspected primary or central AI. Figures on top of the bars show the number of patients included in the different subgroups. Diagnostic agreement is shown as the percentage of observed agreements and kappa coefficients (95%CI).

**Figure 4.** Descriptive statistics and distribution of 30 min stimulated serum cortisol measurement in a population of premenopausal healthy women with evidence of normal HPA axis function. The boundary of the box closest to zero indicates the 25$^{th}$ percentile, the solid and

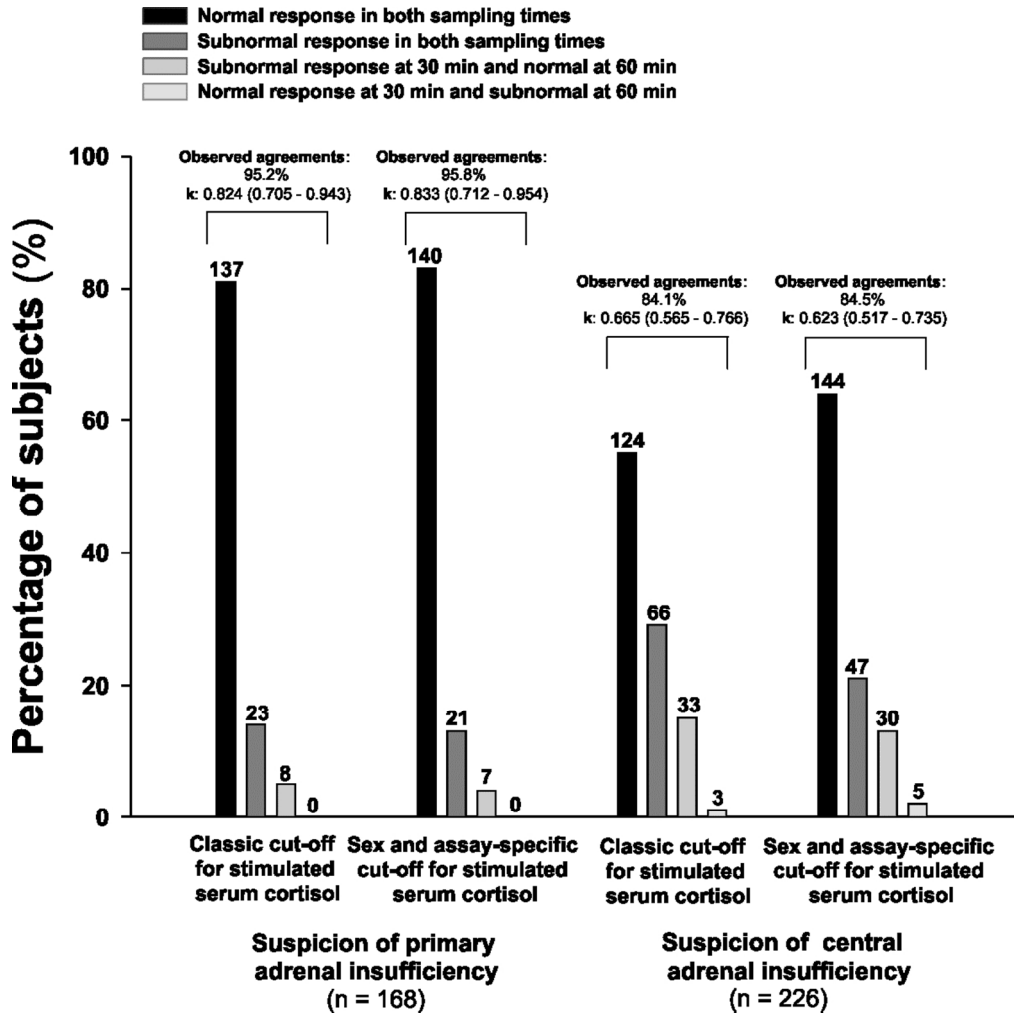
long dash lines within the box marks the median and mean, respectively, and the boundary of the farthest from zero indicates the 75$^{th}$percentile. Whiskers above and below the box indicate the 90$^{th}$ and 10$^{th}$ percentiles. The dashed red line indicates the lower limit of normality (2.5$^{th}$ percentile) for each immunoassay.



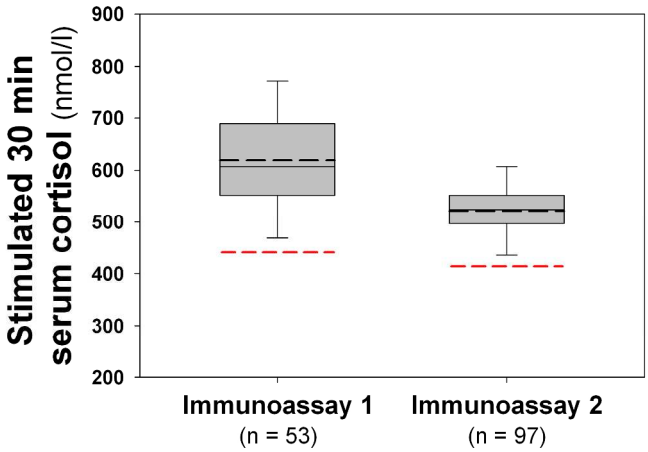

176x410mm (300 x 300 DPI)




190x149mm (150 x 150 DPI)




201x202mm (150 x 150 DPI)




209x296mm (300 x 300 DPI)

# STROBE Statement—checklist of items that should be included in reports of observational studies

**Title:** Importance of sampling time and assay cut-offs for routine assessment of adrenal function: an observational longitudinal study.

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** **Page:** 1 | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract. **Page 1** |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found. **Page 2** |
| **Introduction** | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported. **Page 4** |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses. **Page 5** |
| **Methods** | | |
| Study design | 4 | Present key elements of study design early in the paper. **Page 5** |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection. **Page 5** |
| Participants | 6 | (*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants. **Page: 5-6** |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable**. Page: 5-7** |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group: **Page: 5-7** |
| Bias | 9 | Describe any efforts to address potential sources of bias**. Page 5-7** |
| Study size | 10 | Explain how the study size was arrived at. **Page: 5-7** |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why. **Page: 5-7** |
| Statistical methods **Page: 8-9** | 12 | (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed<br>*Case-control study*—If applicable, explain how matching of cases and controls was addressed<br>*Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses |

Continued on next page

| | | |
|---|---|---|
| **Results** | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **Page: 5-6** |
| | | (b) Give reasons for non-participation at each stage **Page: 5-6** |
| | | (c) Consider use of a flow diagram **Not considered.** |
| Descriptive data **Page 5 - 6** | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. **Page 5 - 6** |
| | | (b) Indicate number of participants with missing data for each variable of interest. **Page: 5-6** |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures **Page: 5-6** |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **Page 9-11** |
| | | (*b*) Report category boundaries when continuous variables were categorized **Page 9-11** |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period. **Not relevant** |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Page 9-11** |
| **Discussion** | | |
| Key results | 18 | Summarise key results with reference to study objectives **Page 12 - 15** |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Page 14** |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Page 14** |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results **Page 15** |
| **Other information** | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Page 15** |

Hospital Universitario Ramón y Cajal

SaludMadrid

Comunidad de Madrid

# PROYECTO DE INVESTIGACIÓN

# EVALUACIÓN DE LA PRUEBA DE ESTIMULACIÓN CORTA CON 1-24 CORTICOTROPINA EN EL DIAGNÓSTICO DE LA INSUFICIENCIA SUPRARRENAL



**Co-IPs**

Manuel Luque Ramírez

Andrés Eduardo Ortiz Flores

Servicio de Endocrinología y Nutrición

**Abril 2016**

Hospital Universitario Ramón y Cajal

SaludMadrid

Comunidad de Madrid

# INDICE

Hospital Universitario Ramón y Cajal
SaludMadrid
Comunidad de Madrid

## RESUMEN (Objetivos y Metodología)

1. **OBJETIVOS:**

   Valorar la calidad asistencial de la prueba de estimulación corta con 1-24 ACTHsintética en nuestro centro hospitalario, en el diagnóstico de pacientes con sospecha de insuficiencia suprarrenaltanto primaria como secundaria, analizando los valores de cortisol sérico en el tiempo 30 y 60 tras el inicio del estímulo, y valorar si son equivalentes o no; de igual manera, buscamos establecer el límite inferior de la normalidad con mayor sensibilidad y/o especificidad para el diagnóstico de insuficiencia suprarrenal. Finalmente buscamos conocer el valor predictivo de esta prueba en nuestro centro hospitalario

2. **METODOLOGÍA**

   Es un estudio de diseño retrospectivo, observacional, en el cual se recolectaránaquellos datos correspondientes a las determinaciones de cortisol plasmático no estimulado, y en los tiempos 30 y 60 tras el inicio del estímulo con 250 µg de 1-24 ACTH sintética, prueba indicada tanto para el screening como para el diagnóstico de insuficiencia suprarrenal primaria y central. Para esto se realizará una revisión de aquellos datos que constan en el registro electrónico del Servicio de Análisis Clínicos-Sección de Hormonas (años 2011 a 2015) (n: 536).Luego de esto se obtendrán aquellos datos clínicos y bioquímicos relevantes, mediante el análisis de la historia clínica electrónica, o del archivo en papel en caso de ser necesario. En segundo lugar, buscaremos determinar el punto de corte tras estímulo con mayor sensibilidad y/o especificidad para el diagnóstico de insuficiencia suprarrenal, y finalmente buscaremos determinar la existencia o no de concordancia entre los tiempos 30 y 60 tras estímulo con 1-24 ACTH, buscando de esta manera valorar la calidad asistencial de este test en el diagnóstica de esta entidad clínica.

Hospital Universitario Ramón y Cajal
SaludMadrid
Comunidad de Madrid

## INTRODUCCIÓN

La insuficiencia suprarrenal (IS) es un trastorno que refleja una síntesis deficiente de glucocorticoides, ya sea por un defecto primario a nivel de las glándulas suprarrenales, o por una secreción inadecuada de corticotropina (ACTH) por la hipófisis que produce una atrofia suprarrenal secundaria (1). Presenta una elevada mortalidad, por lo que es importante, ante la presencia de clínica compatible con la misma, realizar un adecuado diagnóstico, el cual debe ser confirmado mediante las pruebas de laboratorio correspondientes. De ahí la necesidad de disponer de métodos de alta sensibilidad y especificidad que permitan establecer un diagnóstico y tratamiento precoz.

El diagnóstico de IS se basa en la demostración de una producción de cortisol inadecuadamente baja (1). La determinación de cortisol sérico no estimulado a las 8:00 AM permite establecer una aproximación diagnóstica, dado que cifras por debajo de 5 µg/dl sugieren una alta probabilidad de IS, mientras que valores por encima de 18 - 20 µg/dl predicen una respuesta normal en las pruebas de estímulo (hipoglucemia insulínica o prueba de estimulación corta con ACTH sintética). Valores entre 5 a 18 µg/dl se consideran indeterminados y precisan de pruebas que confirmen el diagnóstico (1-2).

La prueba de estimulación corta (EC), se realiza mediante la determinación de los niveles de cortisol sérico a los 30 minutosy/o 60 minutos de la administraciónde 250 µg de 1-24 ACTH sintética. Cualquier valor por encima de 18 - 20 µg/dl tras la prueba de estímulo, es considerado como una respuesta normal (2). Es la prueba de elección para el diagnóstico de IS primaria (1,2).

Por otra parte, esta prueba también se emplea en el diagnóstico de IS secundaria a una producción endógena insuficiente de ACTH, que conducirá a atrofia suprarrenal y, por lo tanto, a una respuesta pobre en la producción de cortisol (3,4). Sin embargo, no se recomienda su utilización en los primeros momentos tras el insulto hipotálamo-hipofisario, ya que requiere tiempo (4 a 6 semanas) para que se produzca una alteración en la respuestaal estimular la reserva suprarrenal (3, 4).

Debido a que algunos pacientes con insuficiencia suprarrenal central parcial presentan una respuesta normal tras el EC con 250 µg de 1-24 ACTH, pero anormal tras otras pruebas de estímulo, algunos autores han propuesto la realización de la EC con 1 µg (5). No obstante, un porcentaje significativo de individuos sanos presentan respuestas "patológicas" con este estímulo (6), lo que sugiere la existencia de problemas técnicos que influyen a la hora de realizar el EC con 1 µg (6), como una administración incompleta de la dosis bien por errores en la dilución del producto, o bien por adherencia de la ACTH a las paredes del catéter (6). Por este motivo, y dado el mayor grado de evidencia disponible, actualmente también se recomienda la utilización del EC con 250 µg de 1-24 ACTH para el diagnóstico de IS secundaria (1-4).

Finalmente, en las unidades de cuidados intensivos, tanto pediátrica como de adultos, el EC con 1-24 ACTH se utiliza, generalmente en pacientes con sepsis grave, que presentan hipotensión sostenida con falta de respuesta al tratamiento vasopresor o inotrópico, sospechando por lo tanto en un déficit relativo de glucocorticoides, entidad conocida como insuficiencia suprarrenal relativa o del paciente crítico (2,8). No existe una clara definición clínica o bioquímica de IS relativa, por lo que hasta fecha actual se considera como diagnóstico un incremento del pico de cortisol en el tiempo 30 o 60 inferior a 9 µg/dl respecto a su valor basal (7,8).

Hospital Universitario Ramón y Cajal
SaludMadrid
Comunidad de Madrid

# JUSTIFICACIÓN DEL PRESENTE PROYECTO DE INVESTIGACIÓN

Actualmente, existen pocos estudios que avalen la necesidad de realizar la determinación de cortisol simultáneamente en los tiempos 30 y 60 min. tras el estímulo con 1-24 ACTH, y aunque ciertos autores sugieren innecesaria la determinación del tiempo 60 (3, 9-10), no existe evidencia suficiente que lo justifique, puesto que este momento de extracción se ha demostrado como más apropiado en otros estudios (11-12). No obstante, las Guías Clínicas más recientemente publicadas sugieren la determinación del cortisol sérico a los 30 minutos de la estimulación (13-14), aunque reconocen que evaluar la posibilidad de que la determinación a los 60 minutos sea más específica requiere ser investigado (14). Estas controversias podrían estar relacionadas con los métodos de determinación de cortisol, vía de administración del estímulo (acceso intravenoso o intramuscular), y la evaluación conjunta de pacientes con déficit primario o central, en los que el momento más adecuado para la determinación del cortisol sérico tras estimulación puede ser también diferente. Pese a estos datos, o como consecuencia de los mismos, durante los últimos años en nuestro Servicio se realizan ambas determinaciones (tiempos 30 y 60 min tras estímulo).

Junto con el momento más adecuado para la extracción de cortisol sérico, otro aspecto importante es la determinación del punto de corte de mayor sensibilidad y especificidad a la hora de establecer el diagnóstico de insuficiencia suprarrenal, dada las importantes implicaciones clínicas de este diagnóstico, y que debería ensayo-específico, puesto que el empleo de diferentes anticuerpos de detección influye en el límite de normalidad tras el estímulo con 1-24 ACTH (14-15). Aunque como ya se ha comentado, clásicamente se considera un valor superior a 18-20 µg/dl como "normal", existen pocos datos en población de nuestro entorno que evalúen el valor predictivo positivo de este punto de corte (10), y en nuestro centro en los últimos años se ha cambiado el ensayo para la determinación de cortisol, lo que ya ha demostrado variar el límite de normalidad en estudios previos (15).

Hospital Universitario Ramón y Cajal
SaludMadrid
Comunidad de Madrid

Por lo tanto, con el objetivo de evaluar y mejorar la calidad asistencial en nuestro Servicio en el diagnóstico de la insuficiencia suprarrenal mediante la estimulación con 1-24 ACTH, parece pertinente la revisión sistemática de nuestros resultados que permitiría determinar el momento para la determinación de cortisol y punto de corte más adecuado, mejorandola eficiencia de esta prueba diagnóstica. A modo de ejemplo, la ausencia de diferenciasentre ambos tiempos de evaluación podría disminuir el tiempo y número de determinaciones con la consiguiente reducción de costes económicos, mejorar la calidad de atención del paciente y evitar posibles tratamientos innecesarios.

## HIPÓTESIS DE TRABAJO

1) La determinación de cortisol plasmático a los 30 minutos del estímulo con 1-24 ACTH es lo suficientemente sensible para detectar a aquellos pacientes con insuficiente función corticosuprarrenal, obviando la necesidad de la determinación a los 60 minutos.

2) Algunos pacientes con insuficiencia suprarrenal central podrían presentar una respuesta tardía al estímulo con 1-24 ACTH que precisaría la determinación a los 60 minutos tras estímulo.

3) El límite inferior de la normalidad utilizado para el diagnóstico de insuficiencia suprarrenal tras el estímulo con 1-24 ACTH en nuestro centro (18-20 mcg/dl), no es adecuado con el inmunoensayo utilizado actualmente, sobreestimando el diagnóstico de insuficiencia suprarrenal.

## OBJETIVOS ESPECÍFICOS

3. Valorar la calidad asistencial del EC con 1-24 ACTH en nuestro centro hospitalario en el diagnóstico de pacientes con sospecha de IS tanto primaria como secundaria.

4. Establecer una correlación entre los hallazgos de laboratorio obtenidos, con la sospecha clínica que motivó el estudio en estos pacientes, valorando así la eficacia y la calidad asistencial.

5. Conocer el valor predictivo positivo de esta prueba en nuestro centro hospitalario, analizando losresultados patológicos tras la prueba de estímulo, la existencia o no de IS y la necesidad de tratamiento sustitutivo con glucocorticoides.


6. Analizar los valores de cortisol sérico obtenidos en el tiempo 30 y 60 tras el inicio del estímulo, compararlos y determinar si ambos resultados son equivalentes o no, para poder justificar la necesidad de realizar o no ambos tiempos.

7. Establecer el límite inferior de la normalidad con mayor sensibilidad y/o especificidad para el diagnóstico de insuficiencia suprarrenal en nuestro medio.

# MATERIAL Y MÉTODOS

## Diseño del estudio y variables de análisis

Se trata de unestudio observacional retrospectivo en el que se recogerán los datos correspondientes a las determinaciones de cortisol plasmático de aquellos pacientes con determinación basal, y en los tiempos 30 y 60 minutos tras estímulo con 1-24 ACTH solicitados para el despistaje-diagnóstico de insuficiencia suprarrenal en nuestro centro, y disponibles en el registro electrónico del Servicio de Análisis Clínicos-Sección de Hormonas (años 2011 a 2015) (n: 536). Junto con los valores de cortisol plasmático en los 3 tiempos, los IPs o investigadores colaboradores accederán a la historia clínica electrónica, y si es preciso al archivo en papel con el objetivo de obtener un mínimo conjunto de datos que permita la consecución de los objetivos del estudio:

- Edad en el momento de la determinación.
- Sexo.
- Peso y Talla.
- Na, K y Calcio plasmáticos en el momento del estímulo.
- Cr y TFGe.
- Concentraciones de ACTH basal si están disponibles.
- Motivo de la solicitud. Sospecha de insuficiencia suprarrenal 1ª / Central / no determinada.
- Antecedentes de patología hipotálamo-hipofisaria. Si existen tiempo de evolución desde el insulto hipotálamo-hipofisario y estudio de función suprarrenal.
- Antecedentes de patología autoinmune poliglandular: tiroiditis, diabetes mellitus tipo 1, etc...
- Antecedentes o presencia de tratamiento estrogénico o gestación.

Hospital Universitario Ramón y Cajal
SaludMadrid
Comunidad de Madrid

- Antecedentes o presencia de tratamiento con mitotane, ketoconazol, metopirona, fenitoina o carbamacepina.
- Antecedentes de tratamiento glucocorticoideo previo al estudio. Dosis, preparado, tiempo de administración, tiempo de evolución desde su suspensión.
- Realización de otros estudios del eje hipotálamo-hipófisis-suprarrenal: hipoglucemia insulínica, prueba de estímulo con glucagón, prueba de estímulo con CRH, prueba de supresión con metopirona y resultado de la misma.
- Tipo de inmunoensayo: Immulite ® / Architect ®.
- Necesidad de tratamiento glucocorticoideo ± mineralocorticoideo por déficit de producción posterior al estudio con la prueba de estimulación.
- Presencia de cuadros clínicos compatibles con crisis suprarrenal aguda.
- Diagnóstico final de funcionalidad del eje hipotálamo-hipófisis-suprarrenal.

Para determinar el punto de corte tras estímulo con mayor sensibilidad y/o especificidad para el diagnóstico de insuficiencia suprarrenal se definirá como pacientes con eje hipotálamo-hipófisis-suprarrenal normal desde el punto de vista clínico a aquellos pacientes que cumplan todos los siguientes criterios:

i) Ausencia de necesidad de administración glucocorticoideo ± mineralocorticoideo transitorio o crónico con la indicación de tratamiento sustitutivo de la función cortiocosuprarrenal.

ii) Ausencia de cuadros clínicos compatibles con crisis suprarrenal aguda.

iii) Ausencia de hiperpigmentación cutánea de etiología no filiada.

iv) Ausencia de hiponatremia o hiperpotasemia de origen no filiado.

v) Ausencia de hipoglucemia en pacientes sin diabetes mellitus de origen no filiado.

vi) Ausencia de pérdida de peso no justificada de origen no filiado.

vii) Ausencia de cuadros compatibles con hipotensión ortostática de origen no filiado.

viii) Ausencia de elevación de ACTH 2 veces por encima del límite superior de la normalidad (> 300 pg/ml).

ix) Ausencia de demostración de insuficiencia suprarrenal parcial o completa en alguna otra prueba de valoración del eje hipotálamo-hipófisis-suprarrenal.

Para establecer la concordancia entre la determinación de las concentraciones de cortisol plasmático entre los tiempos 30 y 60 minutos se utilizará como límite inferior de la normalidad:

i) Un valor de cortisol plasmático tras estímulo con 1-24 ACTH ≥ 18 µg/dl (punto de corte clásico).

ii) Un valor de cortisol plasmático tras estímulo con 1-24 ACTH específico para inmunoensayo, sexo y si está presente, ingesta de ACO combinados (15).

Hospital Universitario Ramón y Cajal
SaludMadrid
Comunidad de Madrid

## Presentación y análisis de resultados

Los datos se presentan en forma de media ± desviación típica o IC95%, mediana (rango intercuatílico) (valor mínimo-máximo), número crudos (porcentaje), de acuerdo a la naturaleza y distribución de las mismos, que se analizará mediante la prueba de Kolmogorov-Smirnov para una muestra. Las comparaciones entre variables continuas se realizarán mediante la prueba U de Mann-Whitney o *t* de Student según la distribución de las mismas. Las comparaciones entre variables cualitativas se realizarán mediante la prueba exacta de Fisher o $\chi^2$ según sea necesario. La correlación entre los tiempos de determinación 30 y 60 minutos se analizará mediante análisis de Pearson, aplicando transformación logarítmica de las variables si es necesario para asegurar la normalidad de las mismas. La concordancia entre los tiempos de determinación de cortisol plasmático se analizará mediante el coeficiente de correlación intraclase y los métodos de representación gráfica de Bland-Altman. La concordancia entre el diagnóstico bioquímico de funcionalidad suprarrenal normal tras estímulo en los tiempos 30 y 60 minutos se analizará mediante la determinación del coeficiente κ. Los puntos de corte tras estimulación con mayor sensibilidad y especificidad para el diagnóstico de insuficiencia suprarrenal se evaluarán mediante curvas ROC. Se considerará estadísticamente significativo un valor $p < 0,05$. Los análisis se realizarán con el paquete estadístico SPSS 15.0.

## Limitaciones del estudio

Las limitaciones del estudio provienen fundamentalmente de su diseño observacional y retrospectivo, con la consiguiente pérdida de datos no recogidos en la historia clínica. Para compensar por estas potenciales pérdidas, los análisis se realizarán tanto con los valores disponibles como mediante análisis de imputación múltiple. Por otro lado, otra de las limitaciones del estudio es el establecimiento del patrón oro con el que comparar los resultados de cortisol plasmático tras estimulación en términos de integridad del eje hipotálamo-hipófisis-suprarrenal. No obstante, como se han especificado previamente hemos establecido una serie de criterios clínicos y bioquímicos, que garantizan con un grado muy elevado de seguridad la ausencia de una insuficiencia suprarrenal completa, y razonablemente, una insuficiencia suprarrenal parcial clínicamente significativa.

## Seguridad

El estudio no supone ningún riesgo para los pacientes incluidos en el mismo, dado que no implica ninguna intervención dado su carácter observacional y retrospectivo.

## Plan de Trabajo

La recogida de datos del Servicio de Análisis Clínicos y el conjunto mínimo de datos de la historia clínica detallado previamente, se realizará por el Equipo Investigador desde el 1 de abril de 2016 al 30 de junio de 2016. El análisis de los mismos se realizará del 1 de julio de 2016 al 31 de septiembre de 2016. La comunicación de resultados del 1 de octubre al 31 de diciembre de 2016.


## Medios disponibles para la realización del proyecto

## Equipo Investigador

El Equipo Investigador está compuesto como co-investigadores principales de Manuel Luque Ramírez, Facultativo Especialista de Área y Andrés Eduardo Ortiz Flores, M.I.R. de 3º año del Servicio de Endocrinología y Nutrición.

Como investigadores colaboradores participarán en la adquisición, análisis de datos y reporte de resultados, Elisa Santacruz Cerdá, M.I.R. de 3º del Servicio de Endocrinología y Nutrición, Ana García Cano y Lucía Jiménez Mendiguchia, Facultativos Especialistas de Área del Servicio de Análisis Clínicos.

## Medios Materiales

El presente estudio se realizará con el equipo informático y software disponible en los Servicios de Análisis Clínicos y Endocrinología y Nutrición, no precisando la adquisición de ningún equipo ni material fungible.

## Memoria Económica

El presente estudio no tiene presupuestado ningún gasto ni va a incurrir en ningún coste extraordinario para el centro, puesto que los investigadores implicados realizarán el trabajo de revisión de historias clínicas fuera de su horario habitual, y el software informático preciso para la realización del mismo está ya disponible, como se ha detallado previamente.

## Aspectos Éticos

**Declaración de cumplimiento, cumplimiento de los requisitos éticos y regulatorios.**

Este estudio se realizará con arreglo al protocolo, a los principios establecidosen la Declaración de Helsinki, a las directrices de buenas prácticas clínicas (BPC) del Comité Internacional de Armonización y a la ley de Investigación Biomédica (14/2007, de 3 de julio).

El protocolo del estudio y los documentos que demuestran la cualificación del investigador son remitidos al Comité Ético de Investigación Clínica para su revisión ética y su aprobación con arreglo a las normativas locales, antes del inicio del estudio.

Las modificaciones de la realización del estudio o de los análisis previstos se documentarán en una enmienda de protocolo y/o del plan de análisis estadístico.

**Confidencialidad de los pacientes**

Los investigadores preservarán la confidencialidad de todos los pacientes que participen en el estudio, con arreglo a las BPC, Declaración de Helsinki y a la legislación local (Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal).

La información obtenida y difundida por la puesta en marcha del presente estudio es considerada confidencial y deberá ser tratada en todo momento como tal. Las pacientes del estudio se identificarán con un código numérico tanto en el cuaderno de recogida de datos (CRD) en una base informatizada. Sólo aquellos datos de la historia clínica que estén relacionados con el estudio, variables especificadas previamente, serán objeto de comprobación. Esta comprobación se hará en presencia del Investigador Principal / Investigadores Colaboradores, responsables de garantizar la confidencialidad de todos los datos de las historias clínicas pertenecientes a las pacientes participantes en el estudio. Los datos recogidos para el estudio estarán identificados mediante un código y solo el investigador principal / colaboradores podrán relacionar dichos datos con el paciente y con su historia clínica.

El tratamiento de los datos se hará con las medidas de seguridad establecidas en cumplimiento de la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal. Tanto los investigadores responsables del ensayo clínico, como un representante de las Autoridades Sanitarias y el CEIC del Hospital Universitario Ramón y Cajal tendrán acceso a la información registrada a lo largo del estudio. En la publicación de los resultados del estudio no se revelará la identidad de los participantes.

**Consideraciones acerca del consentimiento informado (CI).**

El presente proyecto de investigación no contempla en ningún caso entrevistar a los sujetos a los que se practicó la prueba de estimulación con 1-24 ACTH, además de adoptar un método de disociación seguro para evitar manejar datos personales durante el análisis de los resultados. Su carácter retrospectivo, en pacientes que en muchos casos no tienen seguimiento posterior en el centro, y su objetivo de evaluación de calidad asistencial, hace que en consideración de los investigadores del mismo no sea precisa la solicitud de consentimiento informado específico, más allá del consentimiento genérico que otorga el paciente por su atención en un Hospital Universitario del Servicio Madrileño de Salud. No obstante, este punto del protocolo debe ser valorado favorablemente por el CEIC del Centro.

**Compensación a los investigadores**

Se trata de un estudio de Promoción Interna, que no cuenta con ninguna subvención a cargo de fondos públicos, y en el que no se contempla compensación económica a los investigadores participantes en el estudio.

**Difusión de resultados**

Los investigadores se comprometen a publicar los resultados derivados del presente estudio, independientemente de los resultados y conclusiones del mismo, respetando siempre la confidencialidad de la identidad de los sujetos participantes.

## BIBLIOGRAFIA


1. Charmandari E *et al.* ***Adrenal insufficiency***. Lancet **2015**; 383: 2152-67.
2. Neary N & NiemanL. ***Adrenal Insufficiency- etiology, diagnosis and treatment***Curr Opin Endocrinol Diabetes Obes **2010**; 17: 217–223
3. Grossmann A. ***The diagnosis and management of central hypoadrenalism***. J Clin Endocrinol Metab**2010**; 95: 4855–63.
4. Agha A *et al.****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: 43–7.
5. Kazlauskaite R *et al.****Corticotropin tests for hypothalamic-pituitary-adrenal insufficiency: a metaanalysis***. J Clin Endocrinol Metab **2008**; 93: 4245–53.
6. Wade M *et al.* ***Technical details influence the diagnostic accuracy of the 1 microg ACTH stimulation test***. Eur J Endocrinol **2010**; 162: 109–13.
7. Loriaux DL & Fleseriu M. ***Relative adrenal insufficiency***. Curr Opin Endocrinol Diabetes Obes **2009**; 16: 392–400.
8. Sprung CL *et a*l. ***Hydrocortisone therapy for patients with septic shock***. N Engl J Med **2008**; 358: 111–24.
9. Hurel SJ *et al.* ***The short Synacthen and insulin stress tests in the assessment of the hypothalamic-pituitary-adrenal axis***. Clin Endocrinol (Oxf) **1996**; 44: 141-6.
10. Gonzálbez J *et al.* ***Establishment of reference values for standard dose short synacthen test (250 microgram), low dose short synacthen test (1 microgram) and insulin tolerance test for assessment of the hypothalamo-pituitary-adrenal axis in normal subjects***. Clin Endocrinol (Oxf)**2000**; 53: 199-204.
11. Longui CA *et al.* ***Plasma cortisol responses after intramuscular corticotrophin 1-24 in healthy men***. Metabolism **1998**; 47: 1419-22.

12. Mansoor S *et al*. ***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 Med J **2007**; 48: 519-23.

13. de Miguel P*et al*. ***Guía para el diagnóstico y tratamiento de la insuficiencia suprarrenal en el adulto***. Endocrinol Nutr **2014**; 61: 1-35.

14. Borstein SR *et al*. ***Diagnosis and treatment of primary adrenal insufficiency: an Endocrine Society Clinical Practice Guideline***. J Clin Endocrinol Metab **2016**; 101: 364-89.

15. El-Farhan N *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: 673-80.

BMJ Open

# 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: a cross-sectional study.



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2017-019273.R1 |
| Article Type: | Research |
| Date Submitted by the Author: | 15-Dec-2017 |
| Complete List of Authors: | Ortiz-Flores, Andrés; Hospital Universitario Ramón y Cajal, Endocrinology<br>Santacruz, Elisa; Hospital Universitario Ramon y Cajal<br>Jiménez-Mendiguchia, Lucía; Hospital Universitario Ramon y Cajal<br>García-Cano, Ana; Hospital Universitario Ramon y Cajal<br>Nattero-Chávez, Lía; Hospital Universitario Ramon y Cajal<br>Escobar-Morreale, Héctor; Hospital Universitario Ramon y Cajal<br>Luque-Ramírez, Manuel; Hospital Universitario Ramon y Cajal |
| Primary Subject Heading: | Diabetes and endocrinology |
| Secondary Subject Heading: | Diagnostics, Patient-centred medicine |
| Keywords: | Adrenal disorders < DIABETES & ENDOCRINOLOGY, cortisol, biochemical diagnosis, specificity, immunoassay |
| | |

**Title** 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: a cross-sectional study.

**Short title** Cosyntropin test and adrenal insufficiency

**Authors** Andrés E. Ortiz-Flores, M.D.,[1] Elisa Santacruz-Cerdá, M.D,[1] Lucía Jiménez-Mendiguchia, M.D.,[2] Ana García-Cano, M.D.,[2] Lia Nattero-Chávez, M.D.,[1] Héctor F. Escobar-Morreale, M.D., Ph.D.,[1,3] Manuel Luque-Ramírez, M.D., Ph.D.[1,3]

**Affiliations**

[1] Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Madrid (Spain).

[2] Department of Clinical Biochemistry, Hospital Universitario Ramón y Cajal, Madrid (Spain).

[3] Diabetes, Obesity and Human Reproduction Research Group, Hospital Universitario Ramón y Cajal &Universidad de Alcalá & Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS) & Centro de Investigación Biomédica en Red Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain.

**Word count:** 5094**; Word count for abstract:** 300; **Figures:** 5; **Tables:** 3; **Number of references:** 26.

**CORRESPONDENCE & REPRINTS**

Andres E. Ortiz-Flores, MD. Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Carretera de Colmenar Km 9.1, E-28034 Madrid, Spain. Phone/Fax: +34-91 336 9029.

E-mail: andres_ortiz_f@yahoo.com

## ABSTRACT

**Objectives** With the final aim of validating the use of a single post-stimulus sampling protocol for the cosyntropin test (SST) in our institution, our primary objectives were: i) to determine the concordance between 30 and 60 min serum cortisol measurements (SC) during the SST; ii) to evaluate the diagnostic agreement between both sampling times when using classic or assay- and sex-specific cut-offs values for SC. Secondary objectives included :i) estimating the specificity and positive predictive value of 30 and 60 min sampling times while considering the suspected origin of adrenal insufficiency; iv) to obtain assay-specific cut-offs for SC after SST in a group of subjects with a normal hypothalamic-pituitary-adrenal (HPA) axis.

**Design and setting** Cross-sectional study conducted in an Academic Hospital between 2011 and 2015.

**Participants and interventions** Two groups were evaluated: i) a main study group including 370 patients in whom SC was measured at 30 and 60 minutes during the SST; and ii) a confirmative group that included 150 women presenting with a normal HPA axis in whom the SST was conducted to rule out late onset congenital adrenal hyperplasia. Diagnostic agreement between both sampling times was assessed by considering both classic (500 nmol/l) and sex- and assay-specific SC cut-off concentrations.

**Results** Diagnostic agreement between both sampling times was greater when applying sex- and assay-specific cut-off values instead of classic cut-offs. For suspected primary adrenal insufficiency, SC measured at 30-min was enough to make a diagnosis in over 95% of cases, without missing any necessary treatment. When the suspicion was central adrenal insufficiency, the 60 min SC measurement was more specific, establishing diagnosis in over 97% of cases.

**Conclusions:** The use of sex- and assay-specific SC cut-off values improve the diagnostic accuracy of SST for the evaluation of suspected adrenal insuficiency. For primary disease, a subnormal SC response at 30 min is a reliable marker of adrenal dysfunction. On the contrary, when central adrenal insufficiency is suspected, the 60 min SC measurement improves the diagnostic accuracy of the test.

**Strengths and limitations:**

- We assessed a very large series of well-characterized subjects with a suspicion of adrenal insufficiency and a minimum clinical follow up of 12 months after the cosyntropin test.
- We used a pre-test distinction between primary and central adrenal insufficiency based on clinical data.
- We used a local cohort of women with definitely normal cortisol secretion to validate our findings.
- Our results were not challenged against a biochemical gold-standard and, therefore, false negative rates, sensitivity, and negative predictive values were not established.
- The confirmatory group was comprised only by premenopausal women, and cosyntropin-stimulated SC concentrations were only obtained at the 30 min sampling time in these subjects.

**Keywords:** Adrenal insufficiency; biochemical diagnosis; cosyntropin test; immunoassay; reference values; sampling times; serum cortisol; specificity.

## Introduction

The laboratory diagnosis of adrenal insufficiency (AI) at the clinical setting relies on the finding of an inappropriately low morning circulating serum cortisol (SC) or subnormal SC responses to adrenal stimulation [1]. However, the diagnosis of AI diagnosis should not be made according only to laboratory tests, since analytical results must always be interpreted in the context of the whole clinical picture of the individual patient [1–3]. The most widely used adrenal stimulation protocol consists of measuring SC in samples obtained 30 and 60 min after a single 250 μg intravenous bolus or intramuscular injection of tetracosactide (cosyntropin). The normal response consists of a SC value ≥ 500 nmol/l (18 μg/dl) 30 at any time after cosyntropin administration. This protocol, also known as a short standard high-dose test (SST), is the dynamic exploration of choice for primary AI diagnosis [1,3] and it is also used for non-acute central AI [4,5]. In critically ill patients, SST may be performed to rule out a functional form of AI −critical illness-related corticosteroid insufficiency− in subjects showing sustained refractory hypotension and no response to vasopressor drugs [2,6]. Clinical guidelines suggest that this condition may be best diagnosed by a random SC below 276 nmol/l (10 μg/dl) or when the increase in SC after cosyntropin is less than 248 nmol/l (9 μg/dl) [7,8].

The issue of which sampling time - 30 min or 60 min – of the SST is the most appropriate is controversial. The 30 min SC measurements have been validated against a "gold standard" such as the insulin tolerance test (ITT) [9]. Hence, some authors [4,10,11] suggest that a single SC measurement 30 min after cosyntropin administration is enough to establish or rule out clinically relevant AI. Other studies show that a 60 min sample may avoid unnecessary overdiagnosis [12–14]. Recent clinical practice guidelines recommend further research to clarify whether 60 min SC might be more specific than 30 min measurements for AI diagnosis [3,15].

Even though liquid chromatography/mass spectrometry techniques are currently recommended for the accurate measurement of circulating steroids, in most centres clinical routine still relies on automated immunoassays for SC [16]. Considering that the classic cut-off value for the SST was established for SC as measured by older radioimmunoassays, and that immunochemiluminescent assays differ in antibody specificity with these earlier assays [17], establishing local assay-specific cut-off values is of paramount importance to properly classify SC responses to cosyntropin [3,17,18]. This issue is not inconsequential because, despite the recommendation of using local assay-specific lower limits of normality (LLN) for the dynamic assessment of the hypothalamic-pituitary-adrenal (HPA) axis [3], in our experience many physicians still apply classic cut-off values in their routine practice. Also, other factors that may influence SC measurement include the stimulation of hepatic synthesis and secretion of cortisol binding globulin by oestrogens, sex and several non-glucocorticoid drugs [18,19].

To provide new insights into these still open questions, and while validating the use of a single post-stimulus sampling protocol for the routine cosyntropin test (SST) in our institution, our primary goals were: i) to assess the concordance between 30 and 60 min SC concentrations after cosyntropin stimulation at the clinical setting; ii) to estimate the diagnostic agreement between both sampling times when using classic cut-offs derived from the literature or assay- and sex-specific cut-offs values, taking into account the suspected origin of AI. As secondary objectives, we aimed to :i) estimate the specificity (Sp) and positive predictive value (PPV) of 30 and 60 min sampling times while taking into account the origin of AI; and ii) confirm assay-specific LLN for SC concentration after cosyntropin in a group of subjects with a normal HPA function.

## Subjects and methods

From January 1, 2011 to December 31, 2015 we conducted a cross-sectional study in an academic hospital from Spain. We assessed SC responses during a SST in two study populations:

i) Main study population: Four hundred fifty one adults in whom SC concentrations at 0, 30 and 60 min during a SST conducted at the clinical setting for suspected AI.

ii) Confirmative group: One hundred fifty three women with normal HPA axis recruited from our Reproductive Endocrinology clinic during the study of functional hyperandrogenism in whom SC concentrations were obtained at 0 and 30 min during a SST performed for the routine screening of non-classic congenital adrenal hyperplasia (NCAH). NCAH had been ruled out in all those women because cosyntropin-stimulated 17-hydroxyprogesterone and 11-deoxycortisol concentrations were below 10 ng/ml and 21 ng/ml, respectively [20]. None of these women were using combined contraceptives or any other hormonal therapy at the time of sampling.

Before conducting the study, we obtained approval from the local ethics committee. All women from our Reproductive Endocrinology clinic had previously signed an informed consent form for the inclusion of a selection of coded clinical variables in an electronic database for clinical research purposes that included the SC measurements presented here.

*Main study population*

Basal and stimulated SC values were extracted from the electronic database of our Department of Clinical Biochemistry. We collected a minimum dataset in an electronic case form from the clinical records of the patients including age, sex, weight, height, laboratory measurements at the dates when the SST was conducted such as circulating electrolytes, glomerular filtration rate and basal ACTH concentrations at the time of SST, clinical suspicion of primary or central AI, other dynamic tests performed for the evaluation of adrenal function,

history of pituitary disease, time from hypothalamic-pituitary insult to SC determination, administration of drugs that may interfere with the HPA axis, time of follow-up, and the immunoassay used for SC assay. Baseline characteristics of study population are shown in Table 1.

We considered a clinical suspicion of potential primary AI in cases when the patients were known to have adrenal disease, had required mineralocorticoid supplementation during follow-up, had received drugs that may interfere with cortisol biosynthesis, had not clinical suspicion of any hypothalamic-pituitary condition, and had not developed such a condition later in time. Conversely, we suspected a potential central AI in subjects known to suffer from hypothalamic-pituitary disease, had received drugs that may suppress the HPA axis, or when her/his refering physician reported a clinical suspicion of central AI in the clinical record. All patients included here had a minimum follow-up of 12 months after obtaining the SST.

We excluded from analysis: i) seven subjects submitted to dynamic tests other than SST such as the insulin tolerance test (n = 2), corticotrophin-releasing hormone test (n = 2), oral glucose tolerance test (n = 2) and glucagon stimulation test (n = 1); ii) thirty six subjects aged below 18 years; iii) twenty subjects with a follow-up shorter than 12 months; iv) twelve subjects in whom critically-ill related AI was suspected; and v) six subjects from whom we could not obtain enough information from their clinical records as to explain the reason for conducting a SST.Therefore, the study group finally included in the analyses consisted of 370subjects.

*Confirmative group*

The results of SST from 153 premenopausal women with a normal HPA axis aged from 14 to 42 years old were included. Three women who showed a clearly subnormal SC response were excluded from the analysis. In two of these women the suppressive effect on the HPA axis of the

progestins administered during 10 days before the SST with the aim of inducing a withdrawal vaginal bleeding could justify the abnormal results; in the other case, we could not establish the cause of the subnormal response with certainty because the patient was lost to follow-up.

*Assays*

During the study period, two immunoassays were used in our centre: i) from 2011 to July 1, 2013 the *Siemens Immulite 2000© Cortisol Immunoassay System* (immunoassay 1) was used and had 6.0% and 7.8% intra- and inter-assay coefficient of variation (CV) respectively; and ii) from 2013 August 1, 2013 to December 31, 2015, the *Abbot Laboratories Diagnostics Division Architect© Cortisol Immunoassay System* (immunoassay 2) was used, showing 3.2% and 3.4% intra- and inter-assay CVs, respectively.

*Analysis of the agreement between the 30 and 60 min sampling times*

We analysed diagnostic agreement between the 30 and 60 min SC in patients of the main study population – in the confirmation subgroup the 60 min measurement was not obtained – considering two different LLN for cosyntropin-stimulated SC: i) the classic ≥ 500 nmol/l (3), and ii) sex- and assay-specific cut-off values taking into also account the use of combined oral contraceptives (COC) by 8 women [18]. For immunoassay 1, the reported LLN (2.5$^{th}$ percentile) was 470 nmol/l (17 µg/dl) in men and women, and 690 nmol/l (25 µg/dl) for women taking COC. For immunoassay 2, the LLNs were 441 nmol/l (16 µg/dl) for men, 414 nmol/l (15 µg/dl) for women, and 579 nmol/l (21 µg/dl) for women taking COC [17].

*Statistical analysis*

Data are shown as mean ± standard deviation or 95% confidence interval (CI), median (minimum-maximum), and raw numbers (percentage) as appropriate. The normal distribution of continuous variables was assessed by the Kolmogorov-Smirnov test for one sample after a two-step approach for transforming skewed variables if necessary [21]. Comparisons among continuous variables were performed by repeated-measures ANOVA. Comparisons among categorical variables were performed by Fisher's exact or $\chi^2$ tests as appropriate. Pearson's analysis served to correlate SC at 30 and 60 min samples. Consistency and absolute agreement among both point times of SST were determined by their intra-class correlation coefficient (ICC) with a two-factor and random-effect model. Quantitative agreement was graphically assessed by Bland-Altman plots. Biochemical agreement in the diagnosis of normal or subnormal adrenal was assessed by using the kappa (κ) coefficient. True positives (TP) were defined as SSTs showing subnormal cortisol responses at both time points in patients who required adrenal replacement therapy. True negatives (TN) were defined as SSTs showing a normal cortisol response at both time points in patients who did not need glucocorticoid replacement during their follow-up, did not suffer an adrenal crisis, and, when submitted to other dynamic HPA test, showed normal responses. False positives (FP) for one of the sampling times consisted of the finding of a subnormal response in one of the sampling times but not in the other. We calculated Sp and PPV [Sp = TN / (TN + FP) and PPV = TP / (TP + FP)] for each SC sampling times during the SST. A $P$ value < 0.05 was considered statistically significant.

## Results

*Main study population*

Of 370 SSTs including 30 and 60 min sampling times, SC was assayed by immunoassay 1 in 227 cases and by immunoassay 2 in the remaining 143 tests. Basal and cosyntropin-stimulated SC concentrations, ACTH levels when available, and the median duration of follow-up in patients with either normal or insufficient responses are shown in **Table 2**.

SC concentrations when patients in the main study group were analyzed as a whole are represented in **Figure 1A.** SC concentrations at 30 and 60 min during the SST increased when compared to baseline values **(Figure 1, panel A)**, and showed a very strong linear correlation **(Figure 1, panel B)**. Baseline SC concentrations correlated with 30 min SC measurements (r = 0.735, $P = 0.001$), and with 60 min SC values (r = 0.660, $P = 0.001$).

Similar results were observed when analyzing separately the 150 SSTs performed with the aim of to ruling out primary AI (correlation between baseline SC and 30 min SC: r = 0.720, $P = 0.001$), and correlation between baseline SC and 60 min SC: r = 0.640, $P = 0.001$) and the 220 SSTs conducted to exclude central AI (correlation between baseline SC and 30 min SC: r = 0.723, $P = 0.001$, and correlation between baseline SC and 60 min SC: r = 0.644 ($P = 0.001$).

The ICC among SC concentrations as assayed at both sampling times showed a very good consistence index (0.940; 95%CI: 0.928 – 0.952) and a good absolute agreement (0.889, 95%CI: 0.465 – 0.957), even though the latter only qualifies as fair according to the lower limit of the 95%CI. The Bland-Altman plot **(Figure 1, panel C)** showed a good agreement between SC assayed at 30 and 60 min, with a slight tendency towards greater percentage differences with decreasing mean values of stimulated SC.

**Figure 2 and Table 2** show SC concentrations as a function of the clinical suspicion and whether or not the result of the SST was normal. The diagnostic agreement among both sampling times according to classic and to sex – and assay-specific cut-off values is shown in **Figure 3**. Disagreements between both sampling times were as follows. When relying on the classic SC

cut-off point (≥ 500 nmol/l), 39 cases (10.5%) had a subnormal response at 30 min that reached normal values at 60 min whilst, in 3 patients (0.8%), a normal response at 30 min ended being subnormal at 60 min. Using sex- and assay-specific values, 34 cases (9.2%) showed subnormal responses at 30 min but normal SC concentrations at 60 min, whereas in 5 cases (1.3%), the response was normal at 30 min but subnormal at 60 min.

The analysis of the diagnostic agreement as a function of the suspicion of primary versus central AI is shown in **Figure 4**. As a rule, agreement among both sampling times of the SST was better when primary AI was suspected compared with a suspicion of central AI. When using classic cut-off values to rule out primary AI, 7 cases (4.7%) showed a subnormal response at 30 min that reached normal concentrations at 60 min, whereas no subject with a normal response at 30 min had a subnormal response at 60 min. Using sex- and assay-specific cut-off values, in 6 cases (4.0%) the response was subnormal at 30 min but reached normal concentrations at 60 min. Four of them showed a subnormal SC responses to cosyntropin that were very close to the cut-off value. In these subjects, the differences between the cut-off value and the stimulated SC ranged from 22 to 39 nmol/l (0.8 to 1.4 µg/dl), very small concentrations that are, in fact, included within the CV of the assays, thereby suggesting no clinical relevance. The two remaining patients showed peak SC concentrations of 320 and 364 nmol/l (11,6 and 13,2µg/dl) at the 30 min sampling time: one had received oral glucocorticoid replacement therapy that did not preclude the patient of responding to cosyntropin by showing a SC of 470 nmol/l (17 µg/dl) at the 60 min sample, and the other subject was submitted to SST because of the presence of bilateral adrenal hyperplasia and did not show any signs or symptoms of AI nor suffered an adrenal crisis during follow-up. None of the SSTs showing normal responses at 30 min had a subnormal response at 60 min.

When central AI was suspected and the classic cut-off point was applied, 32 cases (14.5%) had a normal SC response at 60 min but a subnormal SC value at 30 min. Only 3 subjects (1.4%) presented with the opposite situation. Using sex- and assay-specific cut-off cocnentrations, 28 cases (12.7%) showed a normal response at 60 min but a subnormal result at 30 min, yet in only 5 cases (2.3%) the contrary occurred. These 5 subjects had been evaluated in the context of withdrawal of prolonged glucocorticoid therapy during the first year after a pituitary insult (surgery and/or pituitary radiotherapy). Three of them showed a complete recovery of their HPA axis throughout the follow-up period, whereas in the other two patients, who had received pituitary radiotherapy, the subnormal response to cosyntropin was maintained over time.

The Sp and PPV for different sampling times and cut-off values used here are shown in Table 3. SC concentrations at 60 min had a higher Sp and PPV compared with 30 min measurements, particularly when central AI was suspected. Nonetheless, the Sp of the determination at 30 min was as high as 95% when SST had been performed to rule out primary disease both when applying classic or sex- and assay-specific cut-off values.

We observed discordant results between classic and sex-and assay-specific cut-off concentrations in 50 cases. In 47 of these subjects, a subnormal response using the classic cut-off value turned into a normal response had sex- and assay-specific cut-offs been used. In 7 of them, SST was performed to rule out primary AI and in the remaining 40 subjects the SSTs were conducted to rule out central AI. Glucocorticoid replacement was started in 18 cases, and no subject presented with signs or symptoms of chronic or acute AI. In addition, from the 50 discordant SSTs, 3 were conducted in women under estrogenic therapy and presented a normal response according to the classic cut-off value, but subnormal when considering sex- and assay-specific cut-offs, yet none of them required glucocorticoid therapy.

*Confirmative group*

Thirty (20%) of these women presented with a subnormal response to SST according to classic cut-off values, yet this figure was reduced to only 3 (2%) when sex- and assay-specific cut-off values were used [observed agreement: 82%; κ: 0.151 (95%CI: 0.066-0.235)]. The three women showing a subnormal response during SST using a sex- and assay-specific cut-off value showed stimulated SC concentrations of 342 nmol/l (12.4 µg/dl), 353 nmol/l (12.8 µg/dl) and 372 nmol/l (13.5 µg/dl), whereas the LLNs (2.5$^{th}$ percentile) of SC concentrations at 30 min sampling time of SST were 436 nmol/l (15.8 µg/dl) and 411 nmol/l (14.9 µg/dl) for immunoassays 1 and 2, respectively. The 5$^{th}$ percentiles for both immunoassays were 450 nmol/l (16.3 µg/dl) and 414 nmol/l (15.0 µg/dl), respectively, showing minimal differences with the LLNs **(Figure 5).** None of these female controls developed any HPA disease during their follow-up**.**

## DISCUSSION

AI is a clinical condition associated with a high morbidity and mortality. Unstimulated early morning SC values below 138 nmol/l (5 µg/dl) show a high PPV for AI, whereas concentrations over 500 nmol/l predict a normal adrenal response. However, values between 138 and 500 nmol/l are considered indeterminate and require adrenal stimulation to confirm or rule out a diagnosis, always in consonance with the clinical picture [1–3].

Baseline SC concentrations showed stronger linear correlations with cosyntropin-stimulated SC levels at 30 and 60 min samples of the SST, in agreement with previous reports [22]. Our data also show that both 30 and 60 min SC measurments during a SST have an adequate index of consistency, but the same is not true in terms of absolute agreement,

particularly when a central AI is suspected. Furthermore, a single determination at 60 min during the SST appears to have the higher Sp and PPV for the diagnosis of subjects presenting with either primary or central AI. In consonance, after evaluating retrospectively 73 subjects, Zueger et al. [23] reported that sampling at 30 min of the SST did not provide any additional diagnostic advantage over performing a single determination at 60 min of the test. Although similar results have been also reported by others [13,14], these studies did not take into account the primary or central origin of AI and did not apply sex- and assay specific cut-off values, a fact of paramount importance because of the considerable influence that cortisol immunoassays exerts on the final values observed after cosyntropin-stimulation [17,18].

Our results also indicate that SC measurement at 30 min during the SST, when using sex- and assay-specific cut-off values, are enough to rule out clinically relevant primary AI since only 4% of patients in this particular situation showed a subnormal response at 30 min followed by normal response at 60 min. Furthermore, these subjects presented with stimulated SC concentrations which were very close to the cut-off concentrations, to the extent that the differences with these normal limits may be explained by the analytical variability of thes commercial immunoassays used here. Even more important from a clinical point of view, none of these subjects required replacement therapy during their follow-up, suffered an acute adrenal crisis, nor were diagnosed with any adrenal condition during follow-up, strongly suggesting that their HPA function was actually normal at the time the SST was performed. The use of sex- and assay-specific cut-off values appears to be essential, since other authors have suggested that some healthy individual may have a delayed response to SST using classic reference values [24].

On the other hand, 60 min samples appears to be more specific than 30 min measurements when central AI is suspected. In such a case, 12.7% of the subjects presenting with a subnormal response at 30 min actually had a normal response at 60 min, avoiding

unnecessary treatments in them. Although a subnormal response 30 min after cosyntropin-stimulation in patients with suspicion of secondary AI may not translate into the need of adrenal replacement in a non-critical scenario, it is likely that most physicians would feel more confident with not starting replacement therapy after obtaining a cosyntropin-stimulated SC concentration above the LLN, favoring the use of 60 min samples over 20 min determinations for this particular reason. Furthermore, relying mostly on 60 min SC responses to cosyntropin when suspecting a central origin of AI is also supported by the fact that, in 2 out of the 5 patients in our series who showed a subnormal response at 60 min preceded by normal SC values at 30 min, AI was actually confirmed during follow-up because of former pituitary radiotherapy.

Our present findings also reinforce the need of sex- and assay-specific cut-off values to interpret the results of the SST, in agreement with recent clinical guidelines[3]. The use of such cut-off values lead in our study to a reduction in FP results, higher Sp and PPV, less discordant results among sampling times of the SST, and fewer unnecessary treatments [20 patients (5%) could have been treated unnecessarily if classic cut-off values were applied for diagnosis]. The reliability of sex- and assay-specific cut-off values was confirmed in our population of premenopausal women with normal HPA axis, in whom these cut-offs were more appropriate than relying on classic values to assess the functionality of their HPA axis. In this population, the LLNs for stimulated SC at 30 min were very close to those reported for each immunoassay by the manufacturers, which relied on the 2.5$^{th}$ percentile [17], yet reinforcing the need to establish local normative data in order to improve the diagnostic accuracy of cortisol measurements during SSTs [17,25].

Among the strengths of our study, we would highlight the large series of subjects suspected of suffering AI who were evaluated with a standardized dynamic study, and the careful review of subjects' medical records that followed such evaluations. However, we are aware of

several weaknesses derived from the observational and retrospective design of the study, making impossible to rule out information bias. Our best efforts might have not been enough to avoid misclassification of patients according to the suspicion of primary or central AI. Also, the administration of supraphysiological doses of cosyntropin does not permit ruling out partial deficiencies either, particularly in those suspected of central HPA defects. Also, and even considering the large sample of subjects included in our study, our present results may not be extrapolable to other populations in whom SC has been measured with different immunoassays that would require specific local normative data. Moreover, analysis of Sp and PPV has not been challenged against a biochemical gold-standard in most cases and, as a consequence, we have not been able to establish false negative rates, sensitivity and negative predictive values. Nonetheless, besides those assessments had been unethical in most cases, the lack of a laboratory gold-standard such as an ITT did not override our results, since from a practical point of view, we are looking for patients needing replacement therapy and not for those with a partial AI who do not require any treatment. Another limitation was that the confirmation group is not fully representative of our main study population since was only comprised of premenopausal women and stimulated SC was only available at the 30 min sampling time. Lastly, we could not rule out entirely pre-treatment with progestogens in the context of induction of withdrawal bleeding in our confirmative population. Because these drugs might exert a mild suppressive effect on the HPA axis [19,26], their administration in a few cases could have, at least in theory, lowered stimulated SC values.

## CONCLUSIONS

Compared with the use of classic cut-off values derived from the literature, application of sex- and assay-specific cut-off values of SC responses to cosyntropin results into higher Sp and

PPV for establishing a diagnosis of AI, thereby avoiding unnecessary treatments. Measurement of stimulated SC at 30 min after cosyntropin-stimulation may suffice for the correct diagnosis of primary AI, yet 60 min measurements might be preferable when central AI is suspected.

**Conflict of interest**: None.

**Funding**: This work has been supported by grant PI1400649) from Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Spanish Ministry of Economy and Competitiveness. M.L.-R. has a local grant for clinical research from the Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS). CIBERDEM is also an initiative of Instituto de Salud Carlos III, partially supported by FondoEuropeo de Desarrollo Regional FEDER. There were no other sources of funding.

**Data sharing statement:** Individual participant data that underlie the results reported in this article, after deidentification, so as the study protocol would be available immediately after publication to anyone who wishes to access the data to achieve aims in the approved proposal and for individual participant data meta-analysis. Proposals should be directed to andres_ortiz_f@yahoo.com or to manuel.luque@salud.madrid.org. To gain access, data requestors will need to sign a data access agreement.

**Contributorship statement:** A.O.-F. y M.L.-R. designed the protocol, and performed the statistical analysis. A.O.-F. y E.S.-C. reviewed the clinical data using the electronic or written

records if necessary. A.G.-C. and L.J.-M. performed the electronic search of serum cortisol samples. A.O.-F. y M.L.-R. wrote the first draft of the study. All the authors, including L.N.-C. and H.F.E.-.M, reviewed the manuscript before its submission and contributed to intellectual content. All the authors have accepted responsibility for the entire content of the manuscript and approved the final submission.

## REFERENCES

1. Charmandari E, Nicolaides NC, Chrousos GP. Adrenal insufficiency. The Lancet. 2014;383(9935):2152-67.

2. Neary N, Nieman L. Adrenal Insufficiency- etiology, diagnosis and treatment. Curr Opin Endocrinol Diabetes Obes. 2010;17(3):217-23.

3. Bornstein SR, Allolio B, Arlt W, Barthel A, Don-Wauchope A, Hammer GD, et al. Diagnosis and Treatment of Primary Adrenal Insufficiency: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2016;101(2):364-89.

4. Grossman AB. The Diagnosis and Management of Central Hypoadrenalism. J Clin Endocrinol Metab. 2010;95(11):4855-63.

5. 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.

6. Sprung CL, Annane D, Keh D, Moreno R, Singer M, Freivogel K, et al. Hydrocortisone Therapy for Patients with Septic Shock. N Engl J Med. 2008;358(2):111-24.

7 Marik PE, Pastores SM, Annane D, *et al.*Recommendations for the diagnosis and management of corticosteroid insufficiency in critically ill adult patients: consensus statements from an international task force by the American College of Critical Care Medicine. *Crit Care Med* 2008;36:1937–49.

8. Annane D, Sébille V, Troché G, Raphaël J-C, Gajdos P, Bellissant E. A 3-Level Prognostic Classification in Septic Shock Based on Cortisol Levels and Cortisol Response to Corticotropin. JAMA. 2000;283(8):1038-45.

9. 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.

10. 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.

11. Gonzálbez, Villabona, Ramón, Navarro, Giménez, Ricart, et al. Establishment of reference values for standard dose short synacthen test (250 µg), low dose short synacthen test (1 µg) and insulin tolerance test for assessment of the hypothalamo–pituitary–adrenal axis in normal subjects. Clin Endocrinol (Oxf). 2000;53(2):199-204.

12. Longui CA, Vottero A, Harris AG, Chrousos GP. Plasma cortisol responses after intramuscular corticotropin 1–24 in healthy men. Metab - Clin Exp. 1998;47(11):1419-22.

13. Mansoor S, Islam N, Siddiqui I, Jabbar A. 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 Med J. 2007;48(6):519-23.

14. 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):14-9.

15. de Miguel Novoa P, Vela ET, García NP, Rodríguez MM, Guerras IS, Martínez de Salinas Santamaría M de LÁ, et al. Guidelines for the diagnosis and treatment of adrenal insufficiency in the adult. Endocrinol Nutr Organo Soc Espanola Endocrinol Nutr. 2014;61 (Suppl 1):1-35.

16. Turpeinen U, Hämäläinen E. Determination of cortisol in serum, saliva and urine. Best Pract Res Clin Endocrinol Metab. 2013;27(6):795-801.

17. El-Farhan N, Pickett A, Ducroq D, Bailey C, Mitchem K, Morgan N, 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):673-80.

18. Klose M, Lange M, Rasmussen AK, Skakkebæk NE, Hilsted L, Haug E, et al. Factors Influencing the Adrenocorticotropin Test: Role of Contemporary Cortisol Assays, Body Composition, and Oral Contraceptive Agents. J Clin Endocrinol Metab. 2007;92(4):1326-33.

19. Kontula K, Paavonen T, Luukkainen T, Andersson LC. Binding of progestins to the glucocorticoid receptor: Correlation to their glucocorticoid-like effects on in vitro functions of human mononuclear leukocytes. Biochem Pharmacol. 1983;32(9):1511-8.

20. Escobar-Morreale HF, Sanchón R, San Millán JL. A prospective study of the prevalence of nonclassical congenital adrenal hyperplasia among women presenting with hyperandrogenic symptoms and signs. J Clin Endocrinol Metab. 2008;93(2):527-33.

21. Templeton G. A Two-Step Approach for Transforming Continuous Variables to Normal: Implications and Recommendations for IS Research. Commun Assoc Inf Syst. 2011;28(1). Available at: http://aisel.aisnet.org/cais/vol28/iss1/4

22 Dorin RI, Qiao ZG, Bouchonville M, *et al.*Characterization of Cortisol Secretion Rate in Secondary Adrenal Insufficiency. *J Endocr Soc* 2017;1:945–56.

23. Zueger T, Jordi M, Laimer M, *et al.*Utility of 30 and 60 minute cortisol samples after high-dose synthetic ACTH-1-24 injection in the diagnosis of adrenal insufficiency. *Swiss Med Wkly* 2014;144:w13987.

24. Karaca Z, Lale A, Tanriverdi F, *et al.* The comparison of low and standard dose ACTH and glucagon stimulation tests in the evaluation of hypothalamo-pituitary-adrenal axis in healthy adults. *Pituitary* 2011;14:134–40.

25. Clark PM, Neylon I, Raggatt PR, *et al.* Defining the normal cortisol response to the short Synacthen test: implications for the investigation of hypothalamic-pituitary disorders. *Clin Endocrinol (Oxf)* 1998;49:287–92.

26. Bornstein SR. Predisposing Factors for Adrenal Insufficiency. N Engl J Med. 2009;360(22):2328-39.

**TABLE 1.** Baseline characteristics population as a function as clinical suspicion of primary or secondary adrenal disease

| | **Clinical suspicion of primary AI** (n = 150) | | **Clinical suspicion of central AI** (n= 220) | |
|---|---|---|---|---|
| Sex | **Female** (n = 98) | **Male** (n = 52) | **Female** (n =139) | **Male** (n = 81) |
| Age *(years)* | 53 ± 19 | 56 ± 15 | 55 ± 16 | 55 ± 13 |
| Weight*(kg)* | 59 ± 12 | 72 ± 14 | 72 ± 14 | 84 ± 16 |
| BMI *(kg/m$^2$)* | 24 ± 5 | 24 ± 5 | 28 ± 5 | 29 ± 5 |
| Na *(mmol/l)* | 138 ± 3 | 137 ± 5 | 139 ± 2 | 140 ± 4 |
| K *(mmol/l)* | 4.3 ± 0.6 | 4.4 ± 0.7 | 4.1 ± 0.4 | 4.1 ± 0.3 |
| Ca (*mmol/l)* | 2.3 ± 0.1 | 2.3 ± 0.2 | 2.4 ± 0.1 | 2.3 ± 0.1 |
| Cr *(µmol/l)* | 62 (44 – 1114) | 80 (44 – 1158) | 62 (18 – 875) | 80 (44 – 150) |
| eGFR (MDRD) *(ml/min/1.73m$^2$)* | 87 (4 – 137) | 86 (4 – 183) | 84 (5 – 361) | 93 (43 – 163) |

*Abbreviations, BMI, body mass index; Ca, total serum calcium; Cr, serum creatinine; eGFR, estimated glomerular filtration rate; K, serum potassium; Na, serum sodium.*

Data are presented as mean ± SD or median (minimum-maximum) as appropriate.

**TABLE 2.** Basal and cosyntropin-stimulated serum cortisol concentrations as a function of the presence of a normal or abnormal result during the SST, and mean follow-up of the patients in each subgroup. From the whole sample, ACTH measurements were available for 342 samples.

| | Normal responses at both times (n = 307) | Confirmed primary AI (n = 18) | Confirmed secondary AI (n = 45) |
|---|---|---|---|
| **Basal ACTH (pmol/l)** | 4 (1 – 43) | 6 (1 – 71) * | 3 (1 - 11) |
| **Basal SC (nmol/l)** | 386 ± 166 | 165 ± 110 | 138 ± 83 |
| **SC at 30 min (nmol/l)** | 662 ± 193 | 248 ± 110 | 276 ± 110 |
| **SC at 60 min (nmol/l)** | 745 ± 221 | 304 ± 138 | 304 ± 110 |
| **Follow-up (months)** | 37 ± 17 | 43 ± 18 | 36 ± 15 |

Data are presented as mean ± SD or median (minimum-maximum) as appropriate. To convert SC to metric units, multiply nmol/l by 0.03625 (result in µg/dl). To convert ACTH to metric units, multiply pmol/l by 4.54545 (result in pg/ml). * Despite not having any hypothalamic-pituitary condition at diagnosis or throughout their follow-up, and not having received drugs that suppress the HPA axis, seven patients with clinical suspicion of primary disease who required replacement therapy presented with normal ACTH levels. Three of them had begun glucocorticoid therapy at the time of SST. In another 4 cases, there is a strong suspicion of that was the case, although the possibility of an inadequate sample proccesing also existed (i.e.: sample transport at room temperature). *Abbreviations, ACTH, adrenocorticotropin hormone; AI, adrenal insufficiency; SC, serum cortisol.*

**TABLE 3.** Specificity and positive predictive value (PPV) of the short high-dose cosyntropin test, for the diagnosis of adrenal insufficiency (AI), according to serum cortisol cut-off concentrations (classic and sex- and assay-specific), and as a function of the suspected origin of the disease.

| | Classic cut-off values | | | | | | Sex- and assay-specific cut-off values | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Global | | Clinical suspicion | | | | Global | | Clinical suspicion | | | |
| | | | Primary AI | | Central AI | | | | Primary AI | | Central AI | |
| **Sampling time (min)** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** |
| **Specificity (%)** | 86 | 99 | 95 | 100 | 79 | 98 | 89 | 98 | 96 | 100 | 84 | 97 |
| **PPV (%)** | 68 | 97 | 74 | 100 | 66 | 96 | 65 | 93 | 75 | 100 | 61 | 90 |

## Legend to figures

**Figure 1. Panel A,** Serum cortisol levels at different sampling times. Data are shown as mean (95%CI) and mean differences (MD) (95%CI). Comparisons among time points were performed by a repeated-measure ANOVA addressing main effects by a Bonferroni's confidence interval adjustment. * *P* value < 0.001. **Panel B,** Pearson's correlation analysis between serum cortisol values at 30 and 60 min sampling times. Solid red line represents the simple linear regression and dotted black lines represent the 95%CI of the regression line. **Panel C**, Bland-Altman plot. Solid black line represents the perfect agreement among both time points. Solid blue line is the mean of the percentage difference among both sampling times, and dashed blue lines are ± 2 standard deviation (SD) of that mean. Solid red line is the regression line of the percentage differences.

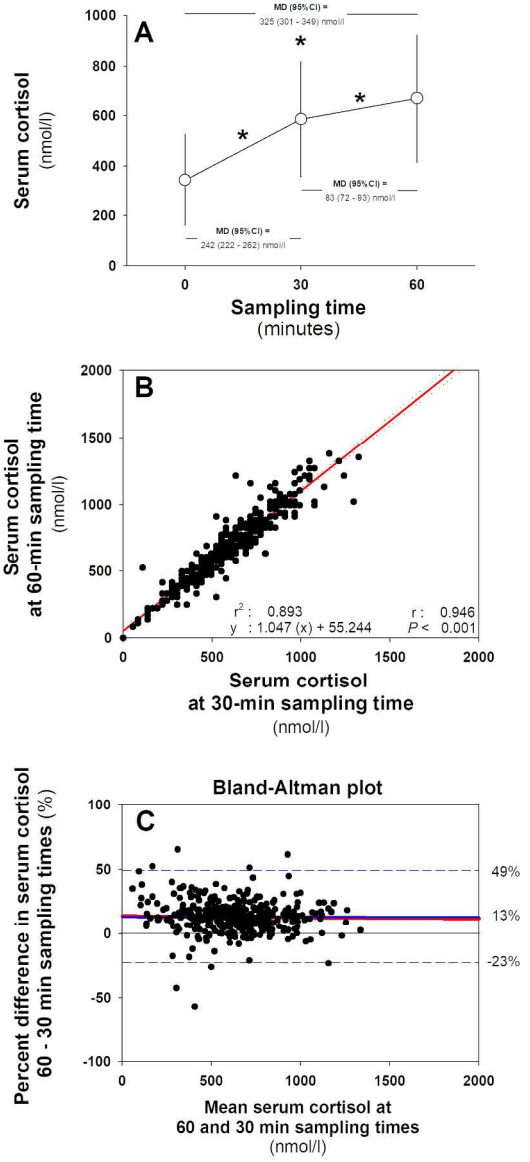
**Figure 2**. Baseline and stimulated serum cortisol concentrations as a function of clinical suspicion and response to cosyntropin test. Data are shown as mean and 95%CI.*Abbreviatures: AI: Adrenal Insufficiency*

**Figure 3.** Subgroups of patients according to serum cortisol responses to cosyntropin-stimulation as a function of classic and sex- and assay-specific cut-offs. Figures on top of the bars indicate the number of patients included in each subgroup. Diagnostic agreement is shown as the percentage of observed agreements and kappa coefficients (95%CI).

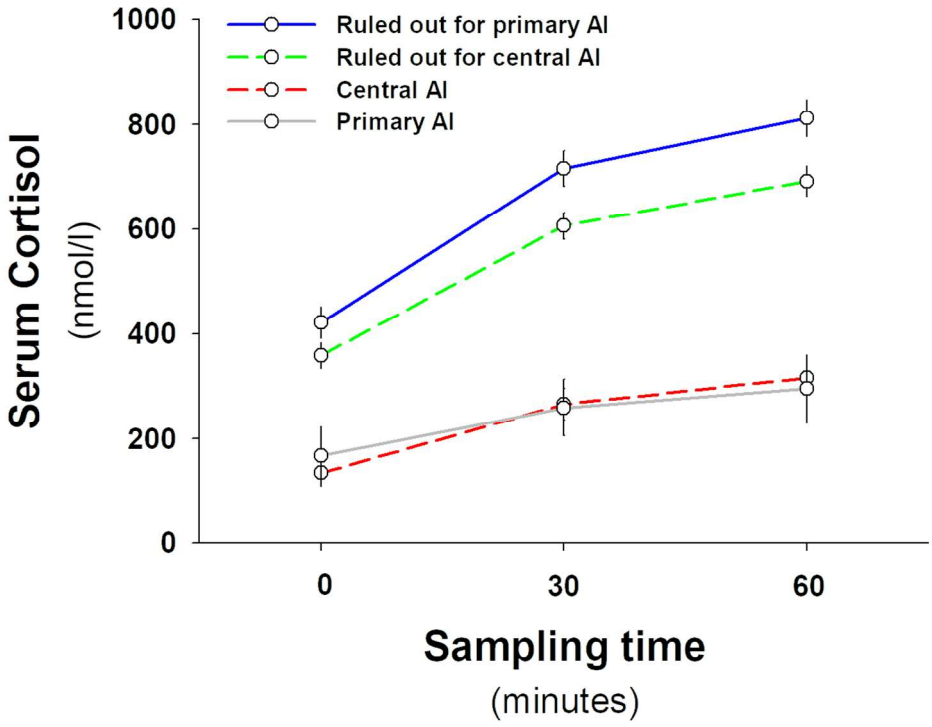
**Figure 4.** Subgroups of patients according to serum cortisol responses to cosyntropin-stimulation as a function of cut-off values and clinical suspicion of primary or central AI. Figures on top of

the bars show the number of patients included in the different subgroups. Diagnostic agreement is shown as the percentage of observed agreements and kappa coefficients (95%CI).

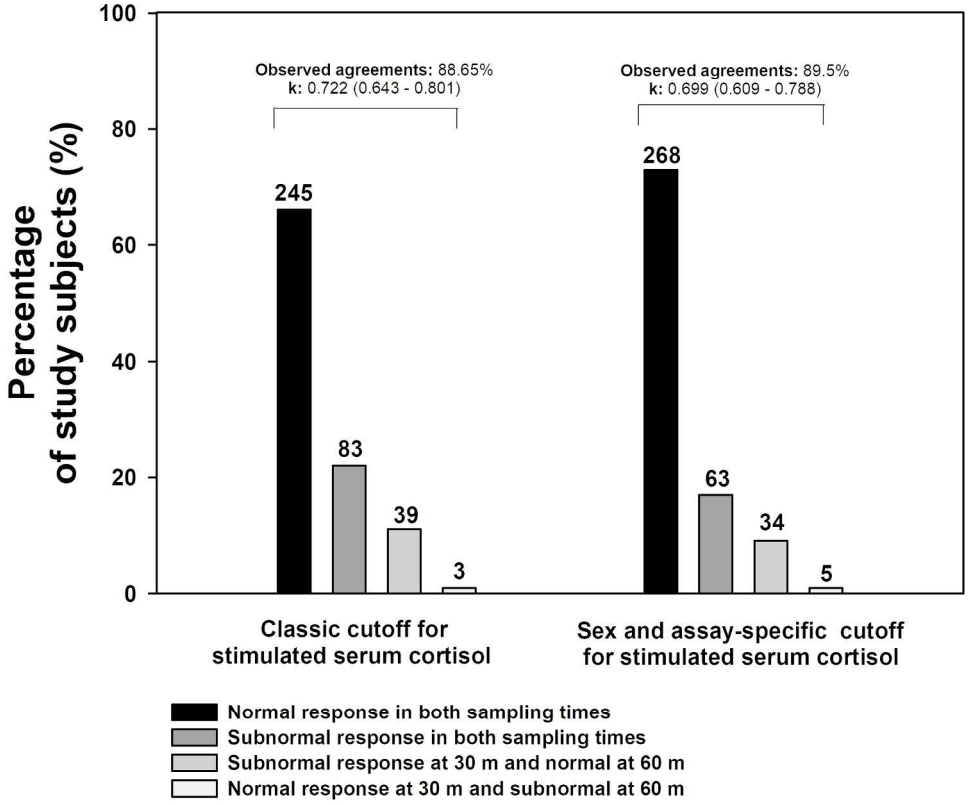
**Figure 5.** Descriptive statistics and distribution of 30 min cosyntropin-stimulated serum cortisol concentrations in a population of premenopausal healthy women with evidence of normal HPA axis function. The boundary of the box closest to zero indicates the 25$^{th}$ percentile, the solid and long dash lines within the box marks the median and mean, respectively, and the boundary of the farthest from zero indicates the 75$^{th}$ percentile. Whiskers above and below the box indicate the 90$^{th}$ and 10$^{th}$ percentiles. Black circles represent the 5$^{th}$ percentile and the dashed red line indicates the lower limit of normality (2.5$^{th}$ percentile) for each immunoassay.



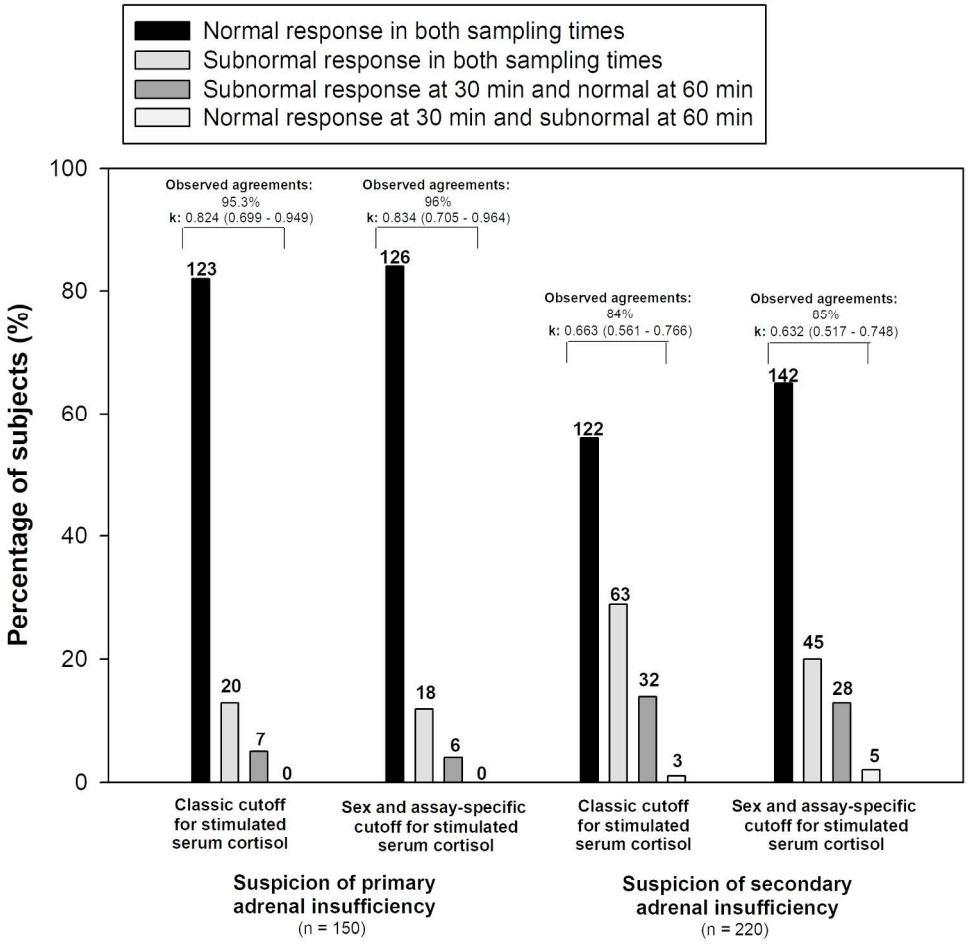

199x282mm (300 x 300 DPI)



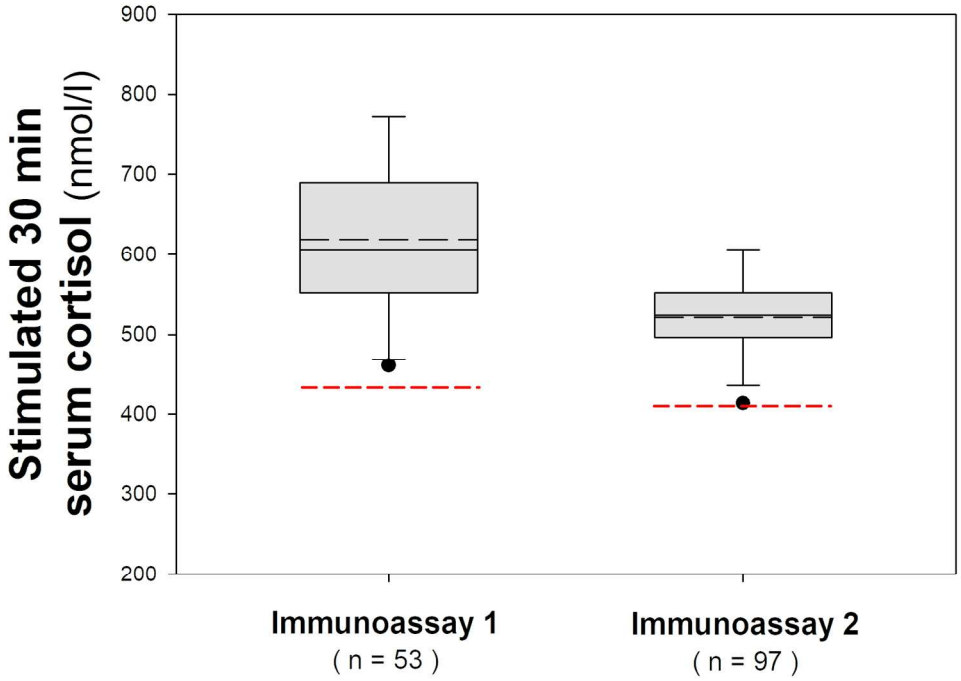

118x97mm (300 x 300 DPI)




214x180mm (300 x 300 DPI)




208x195mm (300 x 300 DPI)




160x120mm (300 x 300 DPI)

# STROBE Statement—checklist of items that should be included in reports of observational studies

**Title:** Importance of sampling time and assay cut-offs for routine assessment of adrenal function: an observational longitudinal study.

| | Item No | Recommendation |
|---|---|---|
| **Title and abstract** **Page:** 1 | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract. **Page 1** |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found. **Page 2** |
| **Introduction** | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported. **Page 4** |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses. **Page 5** |
| **Methods** | | |
| Study design | 4 | Present key elements of study design early in the paper. **Page 5** |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection. **Page 5** |
| Participants | 6 | (*Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants. **Page: 5-6** |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable**. Page: 5-7** |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group: **Page: 5-7** |
| Bias | 9 | Describe any efforts to address potential sources of bias**. Page 5-7** |
| Study size | 10 | Explain how the study size was arrived at. **Page: 5-7** |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why. **Page: 5-7** |
| Statistical methods **Page: 8-9** | 12 | (*a*) Describe all statistical methods, including those used to control for confounding |
| | | (*b*) Describe any methods used to examine subgroups and interactions |
| | | (*c*) Explain how missing data were addressed |
| | | (*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy |
| | | (*e*) Describe any sensitivity analyses |

Continued on next page

| | | |
|---|---|---|
| **Results** | | |
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **Page: 5-6** |
| | | (b) Give reasons for non-participation at each stage **Page: 5-6** |
| | | (c) Consider use of a flow diagram **Not considered.** |
| Descriptive data **Page 5 - 6** | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. **Page 5 - 6** |
| | | (b) Indicate number of participants with missing data for each variable of interest. **Page: 5-6** |
| | | *Cross-sectional study*—Report numbers of outcome events or summary measures **Page: 5-6** |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **Page 9-11** |
| | | (*b*) Report category boundaries when continuous variables were categorized **Page 9-11** |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period. **Not relevant** |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **Page 9-11** |
| **Discussion** | | |
| Key results | 18 | Summarise key results with reference to study objectives **Page 12 - 15** |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **Page 14** |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **Page 14** |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results **Page 15** |
| **Other information** | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based **Page 15** |

BMJ Open

# 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.

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2017-019273.R2 |
| Article Type: | Research |
| Date Submitted by the Author: | 23-Jan-2018 |
| Complete List of Authors: | Ortiz-Flores, Andrés; Hospital Universitario Ramón y Cajal, Endocrinology<br>Santacruz, Elisa; Hospital Universitario Ramon y Cajal<br>Jiménez-Mendiguchia, Lucía; Hospital Universitario Ramon y Cajal<br>García-Cano, Ana; Hospital Universitario Ramon y Cajal<br>Nattero-Chávez, Lía; Hospital Universitario Ramon y Cajal<br>Escobar-Morreale, Héctor; Hospital Universitario Ramon y Cajal<br>Luque-Ramírez, Manuel; Hospital Universitario Ramon y Cajal |
| Primary Subject Heading: | Diabetes and endocrinology |
| Secondary Subject Heading: | Diagnostics, Patient-centred medicine |
| Keywords: | Adrenal disorders < DIABETES & ENDOCRINOLOGY, cortisol, biochemical diagnosis, specificity, immunoassay |
| | |

**Title** 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.

**Short title** Cosyntropin test and adrenal insufficiency

**Authors** Andrés E. Ortiz-Flores, M.D.,[1] Elisa Santacruz-Cerdá, M.D,[1] Lucía Jiménez-Mendiguchia, M.D.,[2] Ana García-Cano, M.D.,[2] Lia Nattero-Chávez, M.D.,[1] Héctor F. Escobar-Morreale, M.D., Ph.D.,[1,3] Manuel Luque-Ramírez, M.D., Ph.D.[1,3]

**Affiliations**

[1] Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Madrid (Spain).

[2] Department of Clinical Biochemistry, Hospital Universitario Ramón y Cajal, Madrid (Spain).

[3] Diabetes, Obesity and Human Reproduction Research Group, Hospital Universitario Ramón y Cajal &Universidad de Alcalá & Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS) & Centro de Investigación Biomédica en Red Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Madrid, Spain.

**Word count:** 5515 (including tables and figure legends)**; Word count for abstract:** 300; **Figures:** 5; **Tables:** 3; **Number of references:** 26.

**CORRESPONDENCE & REPRINTS**

Andres E. Ortiz-Flores & Manuel Luque-Ramírez. Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Carretera de Colmenar Km 9.1, E-28034 Madrid, Spain. Phone/Fax: +34-91 336 9029.

E-mail: andres_ortiz_f@yahoo.com / manuel.luque@salud.madrid.org

## ABSTRACT

**Objectives** Aiming to validate the use of a single post-stimulus sampling protocol for cosyntropin test (SST) in our institution, our primary objectives were: i) to determine the concordance between 30 and 60-min serum cortisol (SC) measurements during SST; ii) to evaluate diagnostic agreement between both sampling times when using classic or assay- and sex-specific SC cut-offs values. Secondary objectives included: i) estimating specificity and positive predictive value of 30 and 60 min sampling times while considering the suspected origin of adrenal insufficiency (AI); iv) to obtain assay-specific cut-offs for SC after SST in a group of subjects with normal hypothalamic-pituitary-adrenal (HPA) axis.

**Design and setting** Retrospective chart review study conducted at an Spanish Academic Hospital from 2011 to 2015

**Participants and interventions** Two groups were evaluated: i) a main study group including 370 patients in whom SC was measured at 30 and 60 minutes during SST; and ii) a confirmative group that included 150 women presenting with a normal HPA axis in whom SST was conducted to rule out late onset congenital adrenal hyperplasia. Diagnostic agreement between both sampling times was assessed by considering both classic (500 nmol/l) and assay-specific SC cut-off concentrations.

**Results** Diagnostic agreement between both sampling times was greater when applying sex- and assay-specific cut-off values instead of classic cut-offs. For suspected primary AI, 30-min SC determination was enough to establish diagnosis in over 95% of cases, without missing any necessary treatment. When central AI is suspected, 60 min SC measurement was more specific, establishing diagnosis in over 97% of cases.

**Conclusions:** Sex- and assay-specific SC cut-off values improve diagnostic accuracy of SST. For primary disease, a subnormal SC response at 30 min is a reliable marker of adrenal dysfunction. On the contrary, when central AI is suspected, 60-min SC measurement improves diagnostic accuracy of the test.

**Strengths and limitations:**

- We assessed a very large series of well-characterized subjects with a suspicion of adrenal insufficiency and a minimum clinical follow up of 12 months after the cosyntropin test.
- We used a pre-test distinction between primary and central adrenal insufficiency based on clinical data.
- We used a local cohort of women with definitely normal cortisol secretion to validate our findings.
- Our results were not challenged against a biochemical gold-standard and, therefore, false negative rates, sensitivity, and negative predictive values were not established.
- The confirmatory group was comprised only by premenopausal women, and cosyntropin-stimulated SC concentrations were only obtained at the 30 min sampling time in these subjects.

**Keywords:** Adrenal insufficiency; biochemical diagnosis; cosyntropin test; immunoassay; reference values; sampling times; serum cortisol; specificity.

## Introduction

The laboratory diagnosis of adrenal insufficiency (AI) at the clinical setting relies on the finding of an inappropriately low morning circulating serum cortisol (SC) or subnormal SC responses to adrenal stimulation [1]. However, the diagnosis of AI diagnosis should not be made according only to laboratory tests, since analytical results must always be interpreted in the context of the whole clinical picture of the individual patient [1–3]. The most widely used adrenal stimulation protocol consists of measuring SC in samples obtained 30 and 60 min after a single 250 μg intravenous bolus or intramuscular injection of tetracosactide (cosyntropin). The normal response consists of a SC value ≥ 500 nmol/l (18 μg/dl) 30 at any time after cosyntropin administration. This protocol, also known as a short standard high-dose test (SST), is the dynamic exploration of choice for primary AI diagnosis [1,3] and it is also used for non-acute central AI [4,5]. In critically ill patients, SST may be performed to rule out a functional form of AI −critical illness-related corticosteroid insufficiency− in subjects showing sustained refractory hypotension and no response to vasopressor drugs [2,6]. Clinical guidelines suggest that this condition may be best diagnosed by a random SC below 276 nmol/l (10 μg/dl) or when the increase in SC after cosyntropin is less than 248 nmol/l (9 μg/dl) [7,8].

The issue of which sampling time - 30 min or 60 min – of the SST is the most appropriate is controversial. The 30 min SC measurements have been validated against a "gold standard" such as the insulin tolerance test (ITT) [9]. Hence, some authors [4,10,11] suggest that a single SC measurement 30 min after cosyntropin administration is enough to establish or rule out clinically relevant AI. Other studies show that a 60 min sample may avoid unnecessary overdiagnosis [12–14]. Recent clinical practice guidelines recommend further research to clarify whether 60 min SC might be more specific than 30 min measurements for AI diagnosis [3,15].

Even though liquid chromatography/mass spectrometry techniques are currently recommended for the accurate measurement of circulating steroids, in most centres clinical routine still relies on automated immunoassays for SC [16]. Considering that the classic cut-off value for the SST was established for SC as measured by older radioimmunoassays, and that immunochemiluminescent assays differ in antibody specificity with these earlier assays [17], establishing local assay-specific cut-off values is of paramount importance to properly classify SC responses to cosyntropin [3,17,18]. When local validation is not feasible, published assay-specific cut-off values should be considered [17]. This issue is not inconsequential because, despite the recommendation of using local assay-specific lower limits of normality (LLN) for the dynamic assessment of the hypothalamic-pituitary-adrenal (HPA) axis [3], in our experience many physicians still apply classic cut-off values in their routine practice. Also, other factors that may influence SC measurement include the stimulation of hepatic synthesis and secretion of cortisol binding globulin by oestrogens, sex and several non-glucocorticoid drugs [18,19].

To provide new insights into these still open questions, and while validating the use of a single post-stimulus sampling protocol for the routine cosyntropin test (SST) in our institution, our primary goals were: i) to assess the concordance between 30 and 60 min SC concentrations after cosyntropin stimulation at the clinical setting; ii) to estimate the diagnostic agreement between both sampling times when using classic cut-offs derived from the literature or assay- and sex-specific cut-offs values, taking into account the suspected origin of AI. As secondary objectives, we aimed to: i) estimate the specificity (Sp) and positive predictive value (PPV) of 30 and 60 min sampling times while taking into account the origin of AI; and ii) confirm assay-specific LLN for SC concentration after cosyntropin in a group of subjects with a normal HPA function.

## Subjects and methods

We conducted a retrospective chart review study addressing SC responses during SST in two study populations from January 1, 2011 to December 31, 2015 at an academic hospital from Spain:

i) Main study population: Four hundred fifty one adults in whom SC concentrations at 0, 30 and 60 min during a SST conducted at the clinical setting for suspected AI.

ii) Confirmative group: One hundred fifty three women with normal HPA axis recruited from our Reproductive Endocrinology clinic during the study of functional hyperandrogenism in whom SC concentrations were obtained at 0 and 30 min during a SST performed for the routine screening of non-classic congenital adrenal hyperplasia (NCAH). NCAH had been ruled out in all those women because cosyntropin-stimulated 17-hydroxyprogesterone and 11-deoxycortisol concentrations were below 10 ng/ml and 21 ng/ml, respectively [20]. None of the women in the confirmative group was using combined contraceptives or any other hormonal therapy at the time of sampling.

Before conducting the study, we obtained approval from the local ethics committee. All women from our Reproductive Endocrinology clinic had previously signed an informed consent form for the inclusion of a selection of coded clinical variables in an electronic database for clinical research purposes that included the SC measurements presented here.

*Main study population*

Basal and stimulated SC values were extracted from the electronic database of our Department of Clinical Biochemistry. We collected a minimum dataset in an electronic case form from the clinical records of the patients including age, sex, weight, height, laboratory measurements at the dates when the SST was conducted such as circulating electrolytes,

glomerular filtration rate and basal ACTH concentrations at the time of SST, clinical suspicion of primary or central AI, other dynamic tests performed for the evaluation of adrenal function, history of pituitary disease, time from hypothalamic-pituitary insult to SC determination, administration of drugs that may interfere with the HPA axis, time of follow-up, and the immunoassay used for SC assay. Baseline characteristics of study population are shown in Table 1.

We considered a clinical suspicion of potential primary AI in cases when the patients were known to have adrenal disease, had required mineralocorticoid supplementation during follow-up, had received drugs that may interfere with cortisol biosynthesis, had not clinical suspicion of any hypothalamic-pituitary condition, and had not developed such a condition later in time. Conversely, we suspected a potential central AI in subjects known to suffer from hypothalamic-pituitary disease, had received drugs that may suppress the HPA axis, or when her/his referring physician reported a clinical suspicion of central AI in the clinical record. According to their clinical records, all patients included here had a minimum 12-month follow-up after obtaining the SST at any outpatient or in-patient facility of our centre. We actively reviewed these records looking for any latter diagnosis of AI.

We excluded from analysis: i) seven subjects submitted to dynamic tests other than SST such as the insulin tolerance test (n = 2), corticotrophin-releasing hormone test (n = 2), oral glucose tolerance test (n = 2) and glucagon stimulation test (n = 1); ii) thirty six subjects aged below 18 years; iii) twenty subjects with a follow-up shorter than 12 months; iv) twelve subjects in whom critically-ill related AI was suspected; and v) six subjects from whom we could not obtain enough information from their clinical records as to explain the reason for conducting a SST. Therefore, the study group finally included in the analyses consisted of 370 subjects.

*Confirmative group*

The results of SST from 153 premenopausal women with a normal HPA axis aged from 14 to 42 years old were included. Three women who showed a clearly subnormal SC response were excluded from the analysis. In two of these women the suppressive effect on the HPA axis of the progestins administered during 10 days before the SST with the aim of inducing a withdrawal vaginal bleeding could justify the abnormal results; in the other case, we could not establish the cause of the subnormal response with certainty because the patient was lost to follow-up.

*Assays*

During the study period, two immunoassays were used in our centre: i) from 2011 to July 1, 2013 the *Siemens Immulite 2000© Cortisol Immunoassay System* (immunoassay 1) was used and had 6.0% and 7.8% intra- and inter-assay coefficient of variation (CV) respectively; and ii) from 2013 August 1, 2013 to December 31, 2015, the *Abbot Laboratories Diagnostics Division Architect© Cortisol Immunoassay System* (immunoassay 2) was used, showing 3.2% and 3.4% intra- and inter-assay CVs, respectively. Plasma ACTH concentrations were measured by the *Siemens Immulite 2000© ACTH Immunoassay System* with an analytical sensitivity of 1.1 pmol/l, and intra- and interassay CVs below 10%. The upper limit of normality for healthy subjects was 10 pmol/l.

*Analysis of the agreement between the 30 and 60 min sampling times*

We analysed diagnostic agreement between the 30 and 60 min SC in patients of the main study population – in the confirmation subgroup the 60 min measurement was not obtained – considering two different LLN for cosyntropin-stimulated SC: i) the classic ≥ 500 nmol/l (3), and ii) sex- and assay-specific cut-off values derived from the estimated lower reference limit for the

SC response at 30-min to cosyntropin, taking also into account the concurrent use by 7 women of combined oral contraceptives (COC) [18]. For immunoassay 1, the reported LLN (2.5$^{th}$ percentile) was 470 nmol/l (17 µg/dl) in men and women, and 690 nmol/l (25 µg/dl) for women taking COC. For immunoassay 2, the LLNs were 441 nmol/l (16 µg/dl) for men, 414 nmol/l (15 µg/dl) for women, and 579 nmol/l (21 µg/dl) for women taking COC [17].

*Statistical analysis*

Data are shown as mean ± standard deviation or 95% confidence interval (CI), median (minimum-maximum), and raw numbers (percentage) as appropriate. The normal distribution of continuous variables was assessed by the Kolmogorov-Smirnov test for one sample after a two-step approach for transforming skewed variables if necessary [21]. Comparisons among continuous variables were performed by repeated-measures ANOVA. Comparisons among categorical variables were performed by Fisher's exact or $\chi^2$ tests as appropriate. Pearson's analysis served to correlate SC at 30 and 60 min samples. Consistency and absolute agreement among both point times of SST were determined by their intra-class correlation coefficient (ICC) with a two-factor and random-effect model. Quantitative agreement was graphically assessed by Bland-Altman plots. Biochemical agreement in the diagnosis of normal or subnormal adrenal was assessed by using the kappa (κ) coefficient. True positives (TP) were defined as SSTs showing subnormal cortisol responses at both time points in patients who required adrenal replacement therapy. True negatives (TN) were defined as SSTs showing a normal cortisol response at both time points in patients who did not need glucocorticoid replacement during their follow-up, did not suffer an adrenal crisis, and, when submitted to other dynamic HPA test, showed normal responses. False positives (FP) for one of the sampling times consisted of the

finding of a subnormal response in one of the sampling times but not in the other. We calculated Sp and PPV [Sp = TN / (TN + FP) and PPV = TP / (TP + FP)] for each SC sampling times during the SST. A *P* value < 0.05 was considered statistically significant.

## Results

*Main study population*

Of 370 SSTs including 30 and 60 min sampling times, SC was assayed by immunoassay 1 in 227 cases and by immunoassay 2 in the remaining 143 tests. Basal and cosyntropin-stimulated SC concentrations, ACTH levels when available, and the median duration of follow-up in patients with either normal or insufficient responses are shown in **Table 2**.

SC concentrations when patients in the main study group were analyzed as a whole are represented in **Figure 1A.** SC concentrations at 30 and 60 min during the SST increased when compared to baseline values **(Figure 1, panel A)**, and showed a very strong linear correlation **(Figure 1, panel B)**. Baseline SC concentrations correlated with 30 min SC measurements (r = 0.735, *P* = 0.001), and with 60 min SC values (r = 0.660, *P* = 0.001).

Similar results were observed when analyzing separately the 150 SSTs performed with the aim of ruling out primary AI (correlation between baseline SC and 30 min SC: r = 0.720, *P* = 0.001), and correlation between baseline SC and 60 min SC: r = 0.640, *P* = 0.001) and the 220 SSTs conducted to exclude central AI (correlation between baseline SC and 30 min SC: r = 0.723, *P* = 0.001, and correlation between baseline SC and 60 min SC: r = 0.644 (*P* = 0.001).

The ICC among SC concentrations as assayed at both sampling times showed a very good consistence index (0.940; 95%CI: 0.928 – 0.952) and a good absolute agreement (0.889, 95%CI: 0.465 – 0.957), even though the latter only qualifies as fair according to the lower limit of the 95%CI. The Bland-Altman plot **(Figure 1, panel C)** showed a good agreement between

SC assayed at 30 and 60 min, with a slight tendency towards greater percentage differences with decreasing mean values of stimulated SC.

**Figure 2 and Table 2** show SC concentrations as a function of the clinical suspicion and whether or not the result of the SST was normal. The diagnostic agreement among both sampling times according to classic and to sex – and assay-specific cut-off values is shown in **Figure 3**. Disagreements between both sampling times were as follows. When relying on the classic SC cut-off point (≥ 500 nmol/l), 39 cases (10.5%) had a subnormal response at 30 min that reached normal values at 60 min whilst, in 3 patients (0.8%), a normal response at 30 min ended being subnormal at 60 min. Using sex- and assay-specific values, 34 cases (9.2%) showed subnormal responses at 30 min but normal SC concentrations at 60 min, whereas in 5 cases (1.3%), the response was normal at 30 min but subnormal at 60 min.

The analysis of the diagnostic agreement as a function of the suspicion of primary versus central AI is shown in **Figure 4**. As a rule, agreement among both sampling times of the SST was better when primary AI was suspected compared with a suspicion of central AI. When using classic cut-off values to rule out primary AI, 7 cases (4.7%) showed a subnormal response at 30 min that reached normal concentrations at 60 min, whereas no subject with a normal response at 30 min had a subnormal response at 60 min. Using sex- and assay-specific cut-off values, in 6 cases (4.0%) the response was subnormal at 30 min but reached normal concentrations at 60 min. Four of them showed a subnormal SC responses to cosyntropin that were very close to the cut-off value. In these subjects, the differences between the cut-off value and the stimulated SC ranged from 22 to 39 nmol/l (0.8 to 1.4 µg/dl), very small concentrations that are, in fact, included within the CV of the assays, thereby suggesting no clinical relevance. The two remaining patients showed peak SC concentrations of 320 and 364 nmol/l (11,6 and 13,2µg/dl) at the 30 min sampling time: one had received oral glucocorticoid replacement therapy that did not

preclude the patient of responding to cosyntropin by showing a SC of 470 nmol/l (17 μg/dl) at the 60 min sample, and the other subject was submitted to SST because of the presence of bilateral adrenal hyperplasia and did not show any signs or symptoms of AI nor suffered an adrenal crisis during follow-up. None of the SSTs showing normal responses at 30 min had a subnormal response at 60 min.

When central AI was suspected and the classic cut-off point was applied, 32 cases (14.5%) had a normal SC response at 60 min but a subnormal SC value at 30 min. Only 3 subjects (1.4%) presented with the opposite situation. Using sex- and assay-specific cut-off concentrations, 28 cases (12.7%) showed a normal response at 60 min but a subnormal result at 30 min, yet in only 5 cases (2.3%) the contrary occurred. These 5 subjects had been evaluated in the context of withdrawal of prolonged glucocorticoid therapy during the first year after a pituitary insult (surgery and/or pituitary radiotherapy). Three of them showed a complete recovery of their HPA axis throughout the follow-up period, whereas in the other two patients, who had received pituitary radiotherapy, the subnormal response to cosyntropin was maintained over time.

The Sp and PPV for different sampling times and cut-off values used here are shown in Table 3. SC concentrations at 60 min had a higher Sp and PPV compared with 30 min measurements, particularly when central AI was suspected. Nonetheless, the Sp of the determination at 30 min was as high as 95% when SST had been performed to rule out primary disease both when applying classic or sex- and assay-specific cut-off values.

We observed discordant results between classic and sex-and assay-specific cut-off concentrations in 50 cases. In 47 of these subjects, a subnormal response using the classic cut-off value turned into a normal response had sex- and assay-specific cut-offs been used. In 7 of them, SST was performed to rule out primary AI and in the remaining 40 subjects the SSTs were

conducted to rule out central AI. Glucocorticoid replacement was started in 18 cases, and no subject presented with signs or symptoms of chronic or acute AI. In addition, from the 50 discordant SSTs, 3 were conducted in women under estrogenic therapy and presented a normal response according to the classic cut-off value, but subnormal when considering sex- and assay-specific cut-offs, yet none of them required glucocorticoid therapy.

*Confirmative group*

Thirty (20%) of these women presented with a subnormal response to SST according to classic cut-off values, yet this figure was reduced to only 3 (2%) when sex- and assay-specific cut-off values were used [observed agreement: 82%; κ: 0.151 (95%CI: 0.066-0.235)]. The three women showing a subnormal response during SST using a sex- and assay-specific cut-off value showed stimulated SC concentrations of 342 nmol/l (12.4 µg/dl), 353 nmol/l (12.8 µg/dl) and 372 nmol/l (13.5 µg/dl), whereas the LLNs (2.5$^{th}$ percentile) of SC concentrations at 30 min sampling time of SST were 436 nmol/l (15.8 µg/dl) and 411 nmol/l (14.9 µg/dl) for immunoassays 1 and 2, respectively. The 5$^{th}$ percentiles for both immunoassays were 450 nmol/l (16.3 µg/dl) and 414 nmol/l (15.0 µg/dl), respectively, showing minimal differences (~10%) with the LLNs previously described **(Figure 5).** None of these female controls developed any HPA disease during their follow-up**.**

We performed a sensitivity analysis of the results in the main study population, after excluding women taking oral contraceptive therapy, using the LLNs derived from the women with a normal HPA axis that composed our confirmatory group. Both sampling times showed a similar agreement than that observed earlier when using LLNs derived from the literature [observed agreement: 92%; κ: 0.724 (95%CI: 0.632-0.816)]. In the whole group of subjects, 4 out of 286 individuals (1.4%) with a normal response at 30-min sampling time showed a

subnormal response at 60-min. Conversely, 26 out of 77 subjects (34%) with a subnormal response at 30-min had a normal response at 60-min. Then, we analyzed those data as a function of the suspected reason for screening AI. Supporting our previous findings, agreement among both SST sampling times was better when primary AI was suspected [observed agreement: 97%; κ: 0.846 (95%CI: 0.714-0.977)] compared with a suspicion of central AI [observed agreement: 89%; κ: 0.667 (95%CI: 0.548-0.785)], data being almost the same observed in Figure 4.

## DISCUSSION

AI is a clinical condition associated with a high morbidity and mortality. Unstimulated early morning SC values below 138 nmol/l (5 μg/dl) show a high PPV for AI, whereas concentrations over 500 nmol/l predict a normal adrenal response. However, values between 138 and 500 nmol/l are considered indeterminate and require adrenal stimulation to confirm or rule out a diagnosis, always in consonance with the clinical picture [1–3].

Baseline SC concentrations showed stronger linear correlations with cosyntropin-stimulated SC levels at 30 and 60 min samples of the SST, in agreement with previous reports [22]. Our data also show that both 30 and 60 min SC measurements during a SST have an adequate index of consistency, but the same is not true in terms of absolute agreement, particularly when a central AI is suspected. Furthermore, a single determination at 60 min during the SST appears to have the higher Sp and PPV for the diagnosis of subjects presenting with either primary or central AI. In consonance, after evaluating retrospectively 73 subjects, Zueger et al. [23] reported that sampling at 30 min of the SST did not provide any additional diagnostic advantage over performing a single determination at 60 min of the test. Although similar results have been also reported by others [13,14], these studies did not take into account the primary or central origin of AI and did not apply sex- and assay specific cut-off values, a fact of paramount

importance because of the considerable influence that cortisol immunoassays exerts on the final values observed after cosyntropin-stimulation [17,18].

Our results also indicate that SC measurement at 30 min during the SST, when using sex- and assay-specific cut-off values, are enough to rule out clinically relevant primary AI since only 4% of patients in this particular situation showed a subnormal response at 30 min followed by normal response at 60 min. Furthermore, these subjects presented with stimulated SC concentrations which were very close to the cut-off concentrations, to the extent that the differences with these normal limits may be explained by the analytical variability of these commercial immunoassays used here. Even more important from a clinical point of view, none of these subjects required replacement therapy during their follow-up, suffered an acute adrenal crisis, nor were diagnosed with any adrenal condition during follow-up, strongly suggesting that their HPA function was actually normal at the time the SST was performed. The use of sex- and assay-specific cut-off values appears to be essential, since other authors have suggested that some healthy individual may have a delayed response to SST using classic reference values [24].

On the other hand, 60 min samples appear to be more specific than 30 min measurements when central AI is suspected. In such a case, 12.7% of the subjects presenting with a subnormal response at 30 min actually had a normal response at 60 min, avoiding unnecessary treatments in them. Although a subnormal response 30 min after cosyntropin-stimulation in patients with suspicion of secondary AI may not translate into the need of adrenal replacement in a non-critical scenario, it is likely that most physicians would feel more confident with not starting replacement therapy after obtaining a cosyntropin-stimulated SC concentration above the LLN, favoring the use of 60 min samples over 20 min determinations for this particular reason. Furthermore, relying mostly on 60 min SC responses to cosyntropin when suspecting a central origin of AI is also supported by the fact that, in 2 out of the 5 patients in our series who showed

a subnormal response at 60 min preceded by normal SC values at 30 min, AI was actually confirmed during follow-up because of former pituitary radiotherapy.

Our present findings also reinforce the need of sex- and assay-specific cut-off values to interpret the results of the SST, in agreement with recent clinical guidelines[3]. The use of such cut-off values lead in our study to a reduction in FP results, higher Sp and PPV, less discordant results among sampling times of the SST, and fewer unnecessary treatments [20 patients (5%) could have been treated unnecessarily if classic cut-off values were applied for diagnosis]. The reliability of sex- and assay-specific cut-off values was confirmed in our population of premenopausal women with normal HPA axis, in whom these cut-offs were more appropriate than relying on classic values to assess the functionality of their HPA axis. In this population, the LLNs for stimulated SC at 30 min were very close to those reported for each immunoassay by the manufacturers, which relied on the 2.5$^{th}$ percentile [17], yet reinforcing the need to establish local normative data in order to improve the diagnostic accuracy of cortisol measurements during SSTs [17,25].

Among the strengths of our study, we would highlight the large series of subjects suspected of suffering AI who were evaluated with a standardized dynamic study, and the careful review of subjects' medical records that followed such evaluations. However, we are aware of several weaknesses derived from the observational and retrospective design of the study, making impossible to rule out information bias. Our best efforts might have not been enough to avoid misclassification of patients according to the suspicion of primary or central AI. Also, the administration of supraphysiological doses of cosyntropin does not permit ruling out partial deficiencies either, particularly in those suspected of central HPA defects. Another limitation is that published assay-specific normative value used in our study derived from SC sampling at 30-min [17]. Thus, the possibility exists that SC sampling at 60-min may require its own normative

cut-off. Also, and even considering the large sample of subjects included in our study, our present results may not be extrapolable to other populations in whom SC has been measured with different immunoassays that would require specific local normative data. Moreover, analysis of Sp and PPV has not been challenged against a biochemical gold-standard in most cases and, as a consequence, we have not been able to establish false negative rates, sensitivity and negative predictive values. Nonetheless, besides those assessments had been unethical in most cases, the lack of a laboratory gold-standard such as an ITT did not override our results, since from a practical point of view, we are looking for patients needing replacement therapy and not for those with a partial AI who do not require any treatment. Another limitation was that the confirmation group is not fully representative of our main study population since was only comprised of premenopausal women and stimulated SC was only available at the 30 min sampling time. Lastly, we could not rule out entirely pre-treatment with progestogens in the context of induction of withdrawal bleeding in our confirmative population. Because these drugs might exert a mild suppressive effect on the HPA axis [19,26], their administration in a few cases could have, at least in theory, lowered stimulated SC values, precluding the generation of local normative data from their results. Instead, we had to rely on published assay-specific cut-off values for this reason.

## CONCLUSIONS

To assist clinical judgement, and compared with the use of classic cut-off values derived from the literature, application of sex- and assay-specific cut-off values of SC responses to cosyntropin results into higher Sp and PPV for establishing a diagnosis of AI, thereby avoiding unnecessary treatments. Measurement of stimulated SC at 30 min after cosyntropin-stimulation

may suffice for supporting a clinical diagnosis of primary AI, yet 60 min measurements might be preferable when central AI is suspected.

**Conflict of interest**: None.

**Funding**: This work has been supported by grant PI1400649) from Fondo de Investigación Sanitaria, Instituto de Salud Carlos III, Spanish Ministry of Economy and Competitiveness. M.L.-R. has a local grant for clinical research from the Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS). CIBERDEM is also an initiative of Instituto de Salud Carlos III, partially supported by Fondo Europeo de Desarrollo Regional FEDER. There were no other sources of funding.

**Data sharing statement:** Individual participant data that underlie the results reported in this article, after deidentification, so as the study protocol would be available immediately after publication to anyone who wishes to access the data to achieve aims in the approved proposal and for individual participant data meta-analysis. Proposals should be directed to andres_ortiz_f@yahoo.com or to manuel.luque@salud.madrid.org. To gain access, data requestors will need to sign a data access agreement.

**Authorship statement:** A.O.-F. y M.L.-R. designed the protocol and performed the statistical analysis. A.O.-F. y E.S.-C. reviewed the clinical data using the electronic or written records if necessary. A.G.-C. and L.J.-M. performed the electronic search of serum cortisol samples. A.O.-

F. y M.L.-R. wrote the first draft of the study. All the authors, including L.N.-C. and H.F.E.-.M, reviewed the manuscript before its submission and contributed to intellectual content. All the authors have accepted responsibility for the entire content of the manuscript and approved the final submission**.**

## REFERENCES


1. Charmandari E, Nicolaides NC, Chrousos GP. Adrenal insufficiency. The Lancet. 2014;383(9935):2152-67.
2. Neary N, Nieman L. Adrenal Insufficiency- etiology, diagnosis and treatment. Curr Opin Endocrinol Diabetes Obes. 2010;17(3):217-23.
3. Bornstein SR, Allolio B, Arlt W, Barthel A, Don-Wauchope A, Hammer GD, et al. Diagnosis and Treatment of Primary Adrenal Insufficiency: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2016;101(2):364-89.
4. Grossman AB. The Diagnosis and Management of Central Hypoadrenalism. J Clin Endocrinol Metab. 2010;95(11):4855-63.
5. 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.
6. Sprung CL, Annane D, Keh D, Moreno R, Singer M, Freivogel K, et al. Hydrocortisone Therapy for Patients with Septic Shock. N Engl J Med. 2008;358(2):111-24.
7 Marik PE, Pastores SM, Annane D, *et al.*Recommendations for the diagnosis and management of corticosteroid insufficiency in critically ill adult patients: consensus statements from an international task force by the American College of Critical Care Medicine. *Crit Care Med* 2008;36:1937–49.
8. Annane D, Sébille V, Troché G, Raphaël J-C, Gajdos P, Bellissant E. A 3-Level Prognostic Classification in Septic Shock Based on Cortisol Levels and Cortisol Response to Corticotropin. JAMA. 2000;283(8):1038-45.

9. 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.

10. 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.

11. Gonzálbez J, Villabona C, Ramón J, Navarro MA, Giménez O, Ricart W, et al. Establishment of reference values for standard dose short synacthen test (250 μg), low dose short synacthen test (1 μg) and insulin tolerance test for assessment of the hypothalamo–pituitary–adrenal axis in normal subjects. Clin Endocrinol (Oxf). 2000;53(2):199-204.

12. Longui CA, Vottero A, Harris AG, Chrousos GP. Plasma cortisol responses after intramuscular corticotropin 1–24 in healthy men. Metab - Clin Exp. 1998;47(11):1419-22.

13. Mansoor S, Islam N, Siddiqui I, Jabbar A. 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 Med J. 2007;48(6):519-23.

14. 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):14-9.

15. de Miguel Novoa P, Vela ET, García NP, Rodríguez MM, Guerras IS, Martínez de Salinas Santamaría Mde L, et al. Guidelines for the diagnosis and treatment of adrenal insufficiency in the adult. Endocrinol Nutr Organo Soc Espanola Endocrinol Nutr. 2014;61 (Suppl 1):1-35.

16. Turpeinen U, Hämäläinen E. Determination of cortisol in serum, saliva and urine. Best Pract Res Clin Endocrinol Metab. 2013;27(6):795-801.

17. El-Farhan N, Pickett A, Ducroq D, Bailey C, Mitchem K, Morgan N, 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):673-80.

18. Klose M, Lange M, Rasmussen AK, Skakkebæk NE, Hilsted L, Haug E, et al. Factors Influencing the Adrenocorticotropin Test: Role of Contemporary Cortisol Assays, Body Composition, and Oral Contraceptive Agents. J Clin Endocrinol Metab. 2007;92(4):1326-33.
19. Kontula K, Paavonen T, Luukkainen T, Andersson LC. Binding of progestins to the glucocorticoid receptor: Correlation to their glucocorticoid-like effects on in vitro functions of human mononuclear leukocytes. Biochem Pharmacol. 1983;32(9):1511-8.
20. Escobar-Morreale HF, Sanchón R, San Millán JL. A prospective study of the prevalence of nonclassical congenital adrenal hyperplasia among women presenting with hyperandrogenic symptoms and signs. J Clin Endocrinol Metab. 2008;93(2):527-33.
21. Templeton G. A Two-Step Approach for Transforming Continuous Variables to Normal: Implications and Recommendations for IS Research. Commun Assoc Inf Syst. 2011;28(1). Available at: http://aisel.aisnet.org/cais/vol28/iss1/4
22 Dorin RI, Qiao ZG, Bouchonville M, *et al.*Characterization of Cortisol Secretion Rate in Secondary Adrenal Insufficiency. *J Endocr Soc* 2017;1:945–56.
23. Zueger T, Jordi M, Laimer M, *et al.*Utility of 30 and 60 minute cortisol samples after high-dose synthetic ACTH-1-24 injection in the diagnosis of adrenal insufficiency. *Swiss Med Wkly* 2014;144:w13987.
24. Karaca Z, Lale A, Tanriverdi F, *et al.* The comparison of low and standard dose ACTH and glucagon stimulation tests in the evaluation of hypothalamo-pituitary-adrenal axis in healthy adults. *Pituitary* 2011;14:134–40.

25. Clark PM, Neylon I, Raggatt PR, *et al.* Defining the normal cortisol response to the short Synacthen test: implications for the investigation of hypothalamic-pituitary disorders. *Clin Endocrinol (Oxf)* 1998;49:287–92.

26. Bornstein SR. Predisposing Factors for Adrenal Insufficiency. N Engl J Med. 2009;360(22):2328-39.

**TABLE 1.** Baseline characteristics of the main study population as a function of the clinical suspicion of primary or central adrenal disease.

| | **Clinical suspicion of primary AI** (n = 150) | | | | **Clinical suspicion of central AI** (n = 220) | | | |
|---|---|---|---|---|---|---|---|---|
| | Assay 1 | | Assay 2 | | Assay 1 | | Assay 2 | |
| Sex | Women (n = 70) | Men (n = 36) | Women (n = 28 ) | Men (n = 16) | Women (n = 75) | Men (n = 46) | Women (n = 64) | Men (n = 35 ) |
| Age *(years)* | 52 ± 19 | 58 ± 14 | 55 ± 18 | 51 ± 14 | 54 ± 14 | 57 ± 13 | 56 ± 18 | 54 ± 13 |
| Weight*(kg)* | 59 ± 13 | 69 ± 14 | 59 ± 9 | 78 ± 14 | 73 ± 14 | 84 ± 12 | 72 ± 13 | 83 ± 20 |
| BMI *($kg/m^2$)* | 24 ± 5 | 24 ± 4 | 23 ± 4 | 26 ± 5 | 29 ± 6 | 29 ± 3 | 28 ± 5 | 29 ± 6 |
| Na *(mmol/l)* | 138 ± 3 | 137 ± 5 | 138 ± 4 | 138 ± 4 | 139 ± 2 | 139 ± 4 | 140 ± 2 | 140 ± 3 |
| K *(mmol/l)* | 4.3 ± 0.6 | 4.5 ± 0.8 | 4.1 ± 0.5 | 4.2 ± 0.3 | 4 ± 0.3 | 4.1 ± 0.4 | 4.2 ± 0.3 | 4.2 ± 0.4 |
| Ca (*mmol/l)* | 2.4 ± 0.1 | 2.3 ± 0.2 | 2.3 ± 0.1 | 2.4 ± 0.1 | 2.4 ± 0.1 | 2.3 ± 0.1 | 2.4 ± 0.1 | 2.4 ± 0.1 |
| Cr *(µmol/l)* | 62 (44 – 1114) | 80 (44 – 1158) | 71 (53 – 230) | 80 (62 - 115) | 62 (44 – 875) | 71 (53 -150) | 71 (18 – 97) | 71 (44 – 141) |
| eGFR (MDRD) *($ml/min/1.73m^2$)* | 88 (4 – 137) | 80 (4 – 183) | 77 (20 – 110) | 98 (57 – 125) | 90 (5 - 144) | 95 (43 – 154) | 81 (48 – 361) | 91 (44 – 163) |
| ACTH *(pmol/l)* | 3 (1 – 16) | 5 (1 - 21) | 4 (1 – 25) | 6 (1 – 230) | 3 (1 – 28) | 4 (1 – 17) | 4 (1 – 43) | 5 (1 – 19) |

*Abbreviations, BMI, body mass index; Ca, total serum calcium; Cr, serum creatinine; eGFR, estimated glomerular filtration rate; K, serum potassium; Na, serum sodium.* Data are presented as mean ± SD or median (minimum-maximum) as appropriate.

**TABLE 2.** Basal and cosyntropin-stimulated serum cortisol concentrations as a function of the presence of a normal or abnormal result during the SST, and mean follow-up of the patients in each subgroup. From the whole sample, ACTH measurements were available for 342 samples.

| | Normal responses at both times (n = 307) | Confirmed primary AI (n = 18) | Confirmed secondary AI (n = 45) |
|---|---|---|---|
| **Basal ACTH (pmol/l)** | 4 (1 – 43) | 6 (1 – 71) * | 3 (1 - 11) |
| **Basal SC (nmol/l)** | 386 ± 166 | 165 ± 110 | 138 ± 83 |
| **SC at 30 min (nmol/l)** | 662 ± 193 | 248 ± 110 | 276 ± 110 |
| **SC at 60 min (nmol/l)** | 745 ± 221 | 304 ± 138 | 304 ± 110 |
| **Follow-up (months)** | 37 ± 17 | 43 ± 18 | 36 ± 15 |

Data are presented as mean ± SD or median (minimum-maximum) as appropriate. To convert SC to metric units, multiply nmol/l by 0.03625 (result in µg/dl). To convert ACTH to metric units, multiply pmol/l by 4.54545 (result in pg/ml). * Despite not having any hypothalamic-pituitary condition at diagnosis or throughout their follow-up, and not having received drugs that suppress the HPA axis, seven patients with clinical suspicion of primary disease who required replacement therapy presented with normal ACTH levels. Three of them had begun glucocorticoid therapy at the time of SST. In another 4 cases, there is a strong suspicion of that was the case, although the possibility of an inadequate sample processing also existed (i.e.: sample transport at room temperature). *Abbreviations, ACTH, adrenocorticotropin hormone; AI, adrenal insufficiency; SC, serum cortisol.*

**TABLE 3.** Specificity and positive predictive value (PPV) of the short high-dose cosyntropin test, for the diagnosis of adrenal insufficiency (AI), according to serum cortisol cut-off concentrations (classic and sex- and assay-specific), and as a function of the suspected origin of the disease.

| | Classic cut-off values | | | | | | Sex- and assay-specific cut-off values | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Global | | Clinical suspicion | | | | Global | | Clinical suspicion | | | |
| | | | Primary AI | | Central AI | | | | Primary AI | | Central AI | |
| **Sampling time (min)** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** | **30** | **60** |
| **Specificity (%)** | 86 | 99 | 95 | 100 | 79 | 98 | 89 | 98 | 96 | 100 | 84 | 97 |
| **PPV (%)** | 68 | 97 | 74 | 100 | 66 | 96 | 65 | 93 | 75 | 100 | 61 | 90 |

## Legend to figures

**Figure 1. Panel A,** Serum cortisol levels at different sampling times. Data are shown as mean (95%CI) and mean differences (MD) (95%CI). Comparisons among time points were performed by a repeated-measure ANOVA addressing main effects by a Bonferroni's confidence interval adjustment.*$P$ value < 0.001. **Panel B,** Pearson's correlation analysis between serum cortisol values at 30 and 60 min sampling times. Solid red line represents the simple linear regression and dotted black lines represent the 95%CI of the regression line. **Panel C**, Bland-Altman plot. Solid black line represents the perfect agreement among both time points. Solid blue line is the mean of the percentage difference among both sampling times, and dashed blue lines are ± 2 standard deviation (SD) of that mean. Solid red line is the regression line of the percentage differences.

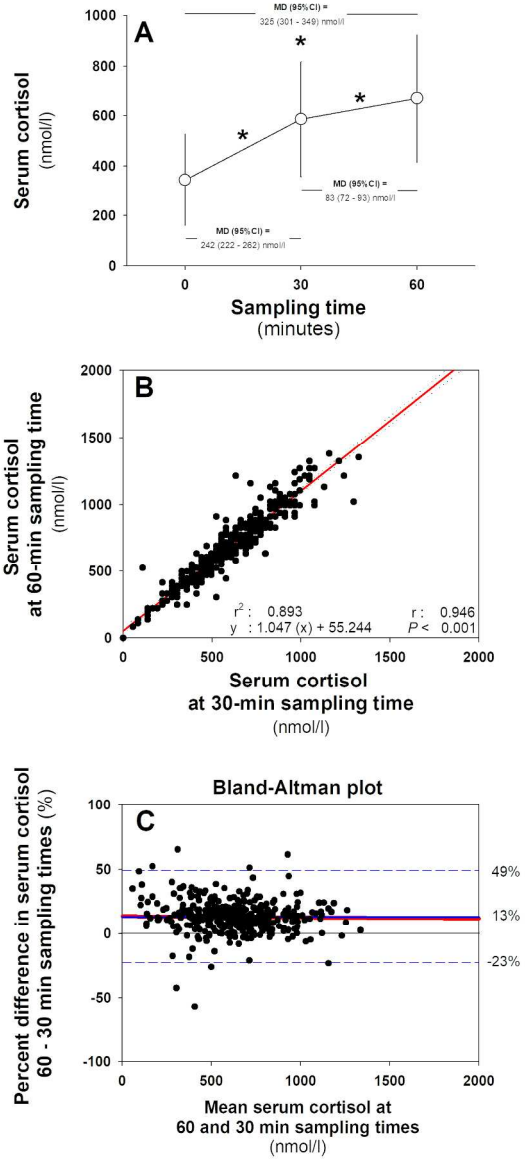
**Figure 2**. Baseline and stimulated serum cortisol concentrations as a function of clinical suspicion and response to cosyntropin test. Data are shown as mean and 95%CI.*Abbreviatures: AI: Adrenal Insufficiency*

**Figure 3.** Subgroups of patients according to serum cortisol responses to cosyntropin-stimulation as a function of classic and sex- and assay-specific cut-offs. Figures on top of the bars indicate the number of patients included in each subgroup. Diagnostic agreement is shown as the percentage of observed agreements and kappa coefficients (95%CI).

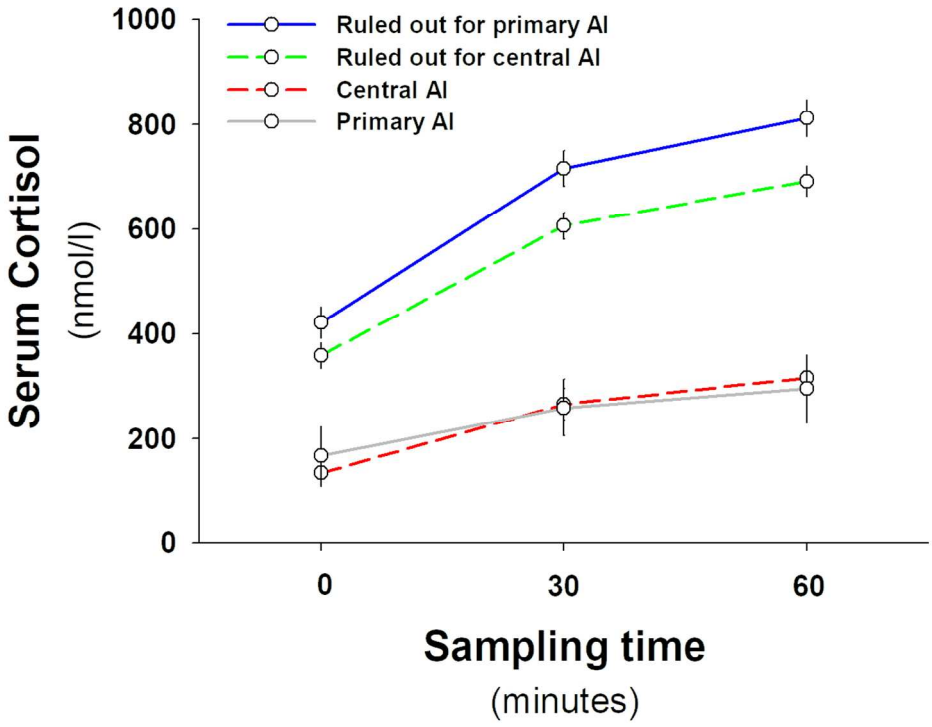
**Figure 4.** Subgroups of patients according to serum cortisol responses to cosyntropin-stimulation as a function of cut-off values and clinical suspicion of primary or central AI. Figures on top of

the bars show the number of patients included in the different subgroups. Diagnostic agreement is shown as the percentage of observed agreements and kappa coefficients (95%CI).

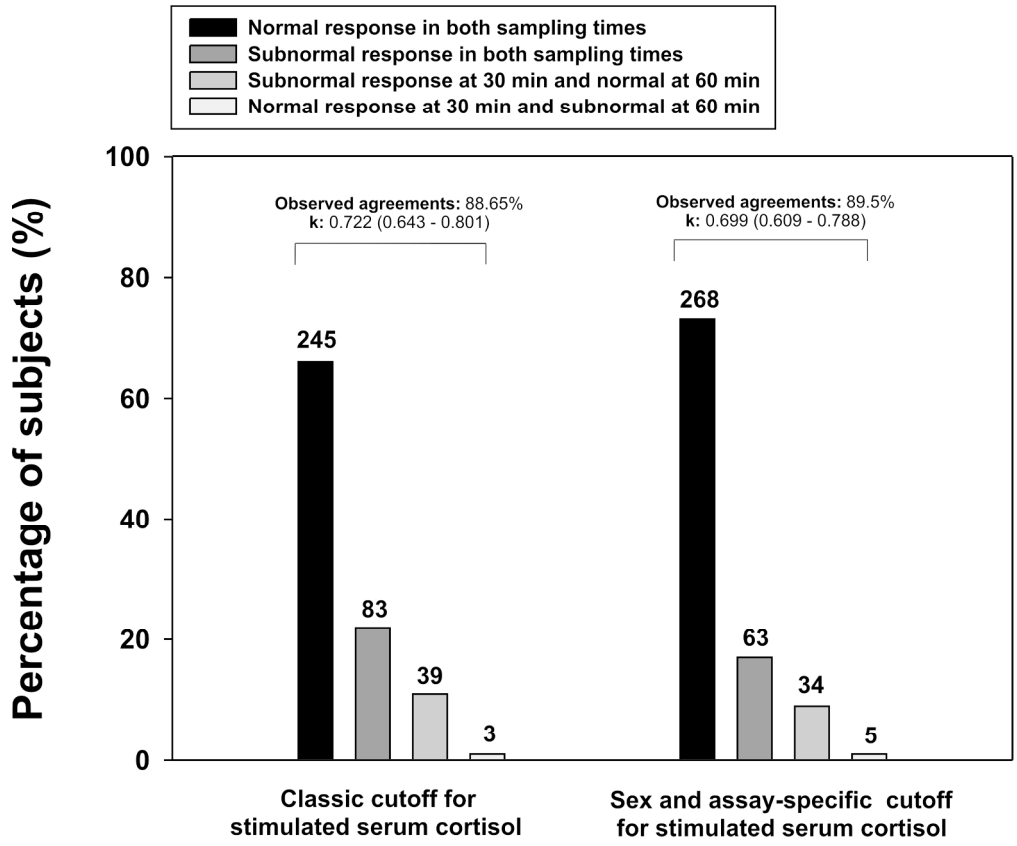
**Figure 5.** Descriptive statistics and distribution of 30 min cosyntropin-stimulated serum cortisol concentrations in a population of premenopausal healthy women with evidence of normal HPA axis function. The boundary of the box closest to zero indicates the 25$^{th}$ percentile, the solid and long dash lines within the box marks the median and mean, respectively, and the boundary of the farthest from zero indicates the 75$^{th}$ percentile. Whiskers above and below the box indicate the 90$^{th}$ and 10$^{th}$ percentiles. Black circles represent the 5$^{th}$ percentile and the dashed red line indicates the lower limit of normality (2.5$^{th}$ percentile) for each immunoassay.

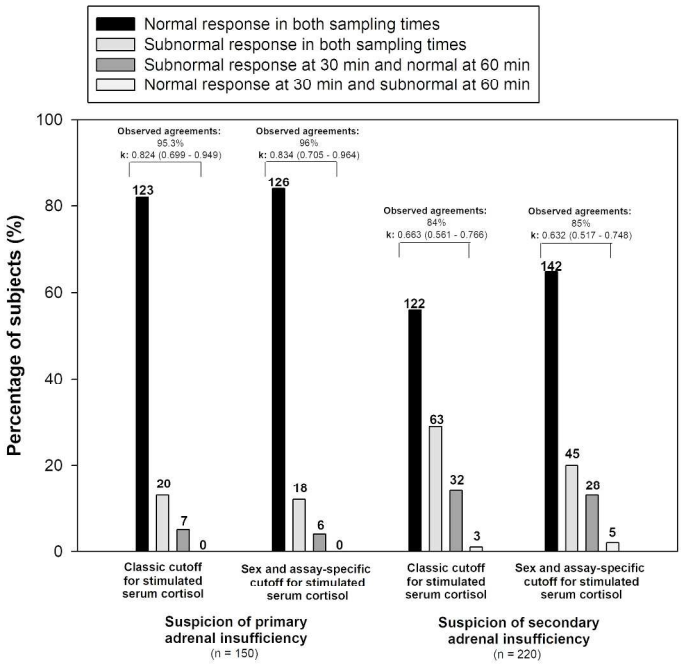



199x282mm (300 x 300 DPI)

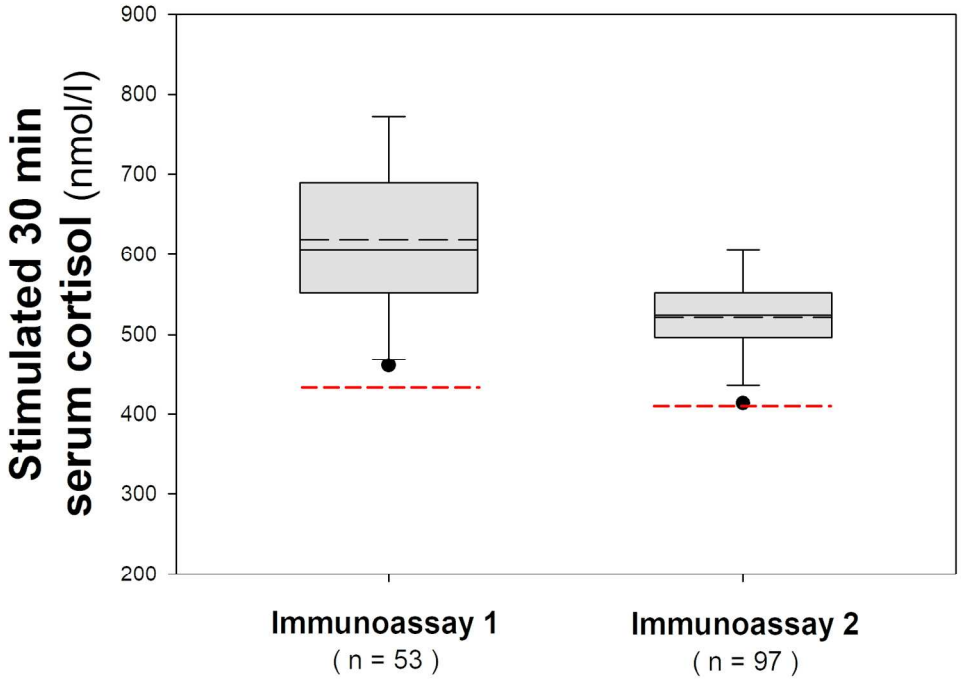



118x97mm (300 x 300 DPI)






296x209mm (300 x 300 DPI)




160x120mm (300 x 300 DPI)

**The STARD 2015 list***

| Section and topic | No | Item |
|---|---|---|
| **Title or abstract:** | | |
| Page 3, Line 47 to 49 | 1 | Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC) |
| **Abstract:** | | |
| Page 3 to 4 | 2 | Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts) |
| **Introduction:** | | |
| Page 5 to 6 | 3 | Scientific and clinical background, including the intended use and clinical role of the index test |
| Page 6, line 109 | 4 | Study objectives and hypotheses |
| **Methods** | | |
| **Study design:** Page 7, Line 121 | 5 | Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study) |
| **Participants** Page 7 | 6 | Eligibility criteria |
| Page 7, Line 124-133; Page 8 Line 139 and page 9 line 168 | 7 | On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry) |

| Section and topic | No | Item |
|---|---|---|
| Page 7, line 122 | 8 | Where and when potentially eligible participants were identified (setting, location, and dates) |
| Page 7, line 124 - 133 | 9 | Whether participants formed a consecutive, random, or convenience series |
| **Test methods** Page 9 - 10 | 10a | Index test, in sufficient detail to allow replication |
| Page 9 - 10 | 10b | Reference standard, in sufficient detail to allow replication |
| Page 9 – 10 | 11 | Rationale for choosing the reference standard (if alternatives exist) |
| Page 10, line 193-196 | 12a | Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory |
| Page 10, line 193-196 | 12b | Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory |
| Page 20, line 130 | 13a | Whether clinical information and reference standard results were available to the performers or readers of the index test |
| Page 20, line 130 | 13b | Whether clinical information and index test results were available to the assessors of the reference standard |
| **Analysis:** Page 9, Line 187 | 14 | Methods for estimating or comparing measures of diagnostic accuracy |
| Methods section, page 8, line 150-159. Therefore, no indeterminate data were avoided. | 15 | How indeterminate index test or reference standard results were handled |
| N/A (no missing data of the index text) | 16 | How missing data on the index test and reference standard were handled |

| Section and topic | No | Item |
|---|---|---|
| Page 10 | 17 | Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory |
| N/A | 18 | Intended sample size and how it was determined |
| **Results** | | |
| Flow not performed. Description of the flow of participants at page 11 | 19 | Flow of participants, using a diagram |
| Table 1, page 25. Also in results, page 11 | 20 | Baseline demographic and clinical characteristics of participants |
| Figure 1, page 31 and result section, page 11 | 21a | Distribution of severity of disease in those with the target condition |
| N/A. No alternative diagnosis were studied | 21b | Distribution of alternative diagnoses in those without the target condition |
| N/A. | 22 | Time interval and any clinical interventions between index test and reference standard |
| **Test results,** Figure page 31, figure 2 page 32, figure 3 page 33 | 23 | Cross tabulation of the index test results (or their distribution) by the results of the reference standard |
| Table 3, page 28 | 24 | Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals) |
| N/A. Our index text / reference standard has no adverse events to report. | 25 | Any adverse events from performing the index test or the reference standard |
| **Discussion** | | |
| Page 17, line 376 | 26 | Study limitations, including sources of potential bias, statistical uncertainty, and generalisability |

| Section and topic | No | Item |
|---|---|---|
| Page 16 and page 18 (Conclusions) | 27 | Implications for practice, including the intended use and clinical role of the index test |
| **Other information** | | |
| N/A, it is a retrospective chart review. | 28 | Registration number and name of registry |
| N/A. Study protocol was attached to editors as supplementary file. | 29 | Where the full study protocol can be accessed |
| Page 19 and 20 | 30 | Sources of funding and other support; role of funders |