

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

Andrés E Ortiz-Flores,^{1,2} Elisa Santacruz,¹ Lucía Jiménez-Mendiguchia,³ Ana García-Cano,³ Lia Nattero-Chávez,¹ Héctor F Escobar-Morreale,^{1,2} Manuel Luque-Ramírez^{1,2}

To cite: Ortiz-Flores AE, Santacruz E, Jiménez-Mendiguchia L, *et al.* Role of sampling times and serum cortisol cut-off concentrations on the routine assessment of adrenal function using the standard cosyntropin test in an academic hospital from Spain: a retrospective chart review. *BMJ Open* 2018;**8**:e019273. doi:10.1136/bmjopen-2017-019273

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-019273>).

Received 22 August 2017
Revised 23 January 2018
Accepted 9 April 2018



For numbered affiliations see end of article.

Correspondence to

Dr Andrés E Ortiz-Flores;
andres_ortiz_f@yahoo.com

ABSTRACT

Objectives Aiming to validate the use of a single poststimulus sampling protocol for cosyntropin test short standard high-dose test (SST) in our institution, our primary objectives were (1) to determine the concordance between 30 and 60 min serum cortisol (SC) measurements during SST; and (2) to evaluate the diagnostic agreement between both sampling times when using classic or assay-specific and sex-specific SC cut-off values. The secondary objectives included (1) estimating the specificity and positive predictive value of 30 and 60 min sampling times while considering the suspected origin of adrenal insufficiency (AI); and (2) obtaining assay-specific cut-off values for SC after SST in a group of subjects with normal hypothalamic–pituitary–adrenal (HPA) axis.

Design and setting This is a retrospective chart review study conducted at a Spanish academic hospital from 2011 to 2015.

Participants and interventions Two groups were evaluated: (1) a main study group including 370 patients in whom SC was measured at 30 and 60 min during SST; and (2) 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-specific and assay-specific cut-off values instead of the classic cut-off values. For suspected primary AI, 30 min SC determination was enough to establish a diagnosis in over 95% of cases, without missing any necessary treatment. When central AI is suspected, 60 min SC measurement was more specific, establishing a diagnosis in over 97% of cases.

Conclusions Sex-specific and assay-specific SC cut-off values improve the 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

Strengths and limitations of this study

- We assessed a very large series of well-characterised subjects with a suspicion of adrenal insufficiency and a minimum clinical follow up of 12 months after the cosyntropin test.
- We used a pretest 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 comprised premenopausal women only, and cosyntropin-stimulated serum cortisol concentrations were only obtained at the 30 min sampling time in these subjects.

central AI is suspected, 60 min SC measurement improves the diagnostic accuracy of the test.

INTRODUCTION

The laboratory diagnosis of adrenal insufficiency (AI) in 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 should not be made according to laboratory tests only, 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 $\mu\text{g}/\text{dL}$) 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 $\mu\text{g}/\text{dL}$) or when the increase in SC after cosyntropin is less than 248 nmol/L (9 $\mu\text{g}/\text{dL}$).^{7,8}

The issue of which sampling time—30 min or 60 min—of 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 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 (LLNs) for the dynamic assessment of the hypothalamic–pituitary–adrenal (HPA) axis,³ in our experience many physicians still apply the 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 poststimulus sampling protocol for the routine cosyntropin test (SST) in our institution, our primary goals were (1) to assess the concordance between 30 and 60 min SC concentrations after cosyntropin stimulation in the clinical setting; and (2) to estimate the diagnostic agreement between both sampling times when using the classic cut-off values derived from the literature or assay-specific

and sex-specific cut-off values, taking into account the suspected origin of AI. As secondary objectives, we aimed to (1) 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 (2) 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 1 January 2011 to 31 December 2015 at an academic hospital in Spain: (1) main study population: 451 adults in whom SC concentrations were obtained at 0, 30 and 60 min during an SST conducted in the clinical setting for suspected AI; and (2) confirmative group: 153 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 an SST performed for routine screening of non-classic congenital adrenal hyperplasia (NCAH). NCAH had been ruled out in all these 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 were using combined contraceptives or any other hormonal therapy at the time of sampling.

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 the Department of Clinical Biochemistry. We collected a minimum data set in an electronic case form from the clinical records of patients, including age, sex, weight, height, laboratory measurements on the dates when SST was conducted such as circulating electrolytes, glomerular filtration rate and basal adrenocorticotrophic hormone (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. The baseline characteristics of study population are shown in [table 1](#).

We considered a clinical suspicion of potential primary AI in cases where patients were known to have adrenal disease, had required mineralocorticoid supplementation during follow-up, had received drugs that may interfere with cortisol biosynthesis, had no clinical suspicion of any hypothalamic–pituitary condition and had not

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

Sex	Clinical suspicion of primary AI (n=150)				Clinical suspicion of central AI (n=220)			
	Assay 1		Assay 2		Assay 1		Assay 2	
	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±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/1.73 m 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)

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

ACTH, adrenocorticotropic hormone; AI, adrenal insufficiency; BMI, body mass index; Ca, total serum calcium; Cr, serum creatinine; eGFR, estimated glomerular filtration rate; K, serum potassium; MDRD, Modification of diet in renal disease; Na, serum sodium.

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 of 12-month follow-up after obtaining SST at any outpatient or inpatient facility of our centre. We actively reviewed these records looking for any later diagnosis of AI.

We excluded the following from the analysis: (1) 7 subjects submitted to dynamic tests other than SST, such as ITT (n=2), corticotrophin-releasing hormone test (n=2), oral glucose tolerance test (n=2) and glucagon stimulation test (n=1); (2) 36 subjects aged below 18 years; (3) 20 subjects with a follow-up shorter than 12 months; (4) 12 subjects in whom critically ill related AI was suspected; and (5) 6 subjects from whom we could not obtain enough information from their clinical records as to explain the reason for conducting an SST. Therefore, the study group included in the analyses finally consisted of 370 subjects.

Confirmative group

The results of SST from 153 premenopausal women aged 14–42 years old with normal HPA axis 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 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: (1) from 2011 to 1 July 2013, the Siemens Immulite 2000 Cortisol Immunoassay System (immunoassay 1) was used and had 6.0% and 7.8% intra-assay and interassay coefficients of variation (CVs), respectively; and (2) from 1 August 2013 to 31 December 2015, the Abbott Laboratories Diagnostics Division Architect Cortisol Immunoassay System (immunoassay 2) was used, showing 3.2% and 3.4% intra-assay 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-assay 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 the diagnostic agreement between the 30 and 60 min SC in the patients of the main study population—in the confirmation subgroup the 60 min measurement was not obtained—considering two different LLNs for cosyntropin-stimulated SC: (1) the classic ≥ 500 nmol/L³, and (2) sex-specific and assay-specific cut-off values derived from the estimated lower reference limit for the SC response at 30 min to cosyntropin, also taking into account the concurrent use by seven women of combined oral contraceptives (COC).¹⁸ For immunoassay 1, the reported LLNs (2.5th percentile) were 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±SD or 95% 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 analysis of variance. Comparisons among categorical variables were performed by Fisher's exact or χ^2 tests as appropriate. Pearson's analysis served to correlate SC at 30 and 60 min samples. Consistency and absolute agreement among both time points 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 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 of the SC sampling times during 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 analysed as a whole are presented in figure 1A. SC concentrations at 30 and 60 min during SST increased when compared with baseline values (figure 1A) and showed a very strong linear correlation (figure 1B). 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 analysing 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

Table 2 Basal and cosyntropin-stimulated serum cortisol concentrations as a function of the presence of a normal or abnormal result during SST, and mean follow-up of the patients in each subgroup

	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

From the whole sample, ACTH measurements were available for 342 samples.

Data are presented as mean±SD or median (minimum–maximum) as appropriate. To convert SC to metric units, multiply nmol/L by 0.03625 (results in $\mu\text{g/dL}$). To convert ACTH to metric units, multiply pmol/L by 4.54545 (results 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 hypothalamic–pituitary–adrenal 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 four cases, there is a strong suspicion that was the case, although the possibility of an inadequate sample processing also existed (ie, sample transport at room temperature). ACTH, adrenocorticotropic hormone; AI, adrenal insufficiency; SC, serum cortisol; SST, short standard high-dose test.

correlation between baseline SC and 60 min SC: $r=0.644$, $p=0.001$).

The ICC among the SC concentrations as assayed at both sampling times showed a very good consistency index (0.940; 95% CI 0.928 to 0.952) and a good absolute agreement (0.889, 95% CI 0.465 to 0.957), even though the latter only qualifies as fair according to the lower limit of the 95% CI. The Bland-Altman plot (figure 1C) showed a good agreement between the SCs 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 SST was normal. The diagnostic agreement among both sampling times according to classic and to sex-specific 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, while in three patients (0.8%) a normal response at 30 min ended being subnormal at 60 min. Using sex-specific and assay-specific values, 34 cases (9.2%) showed

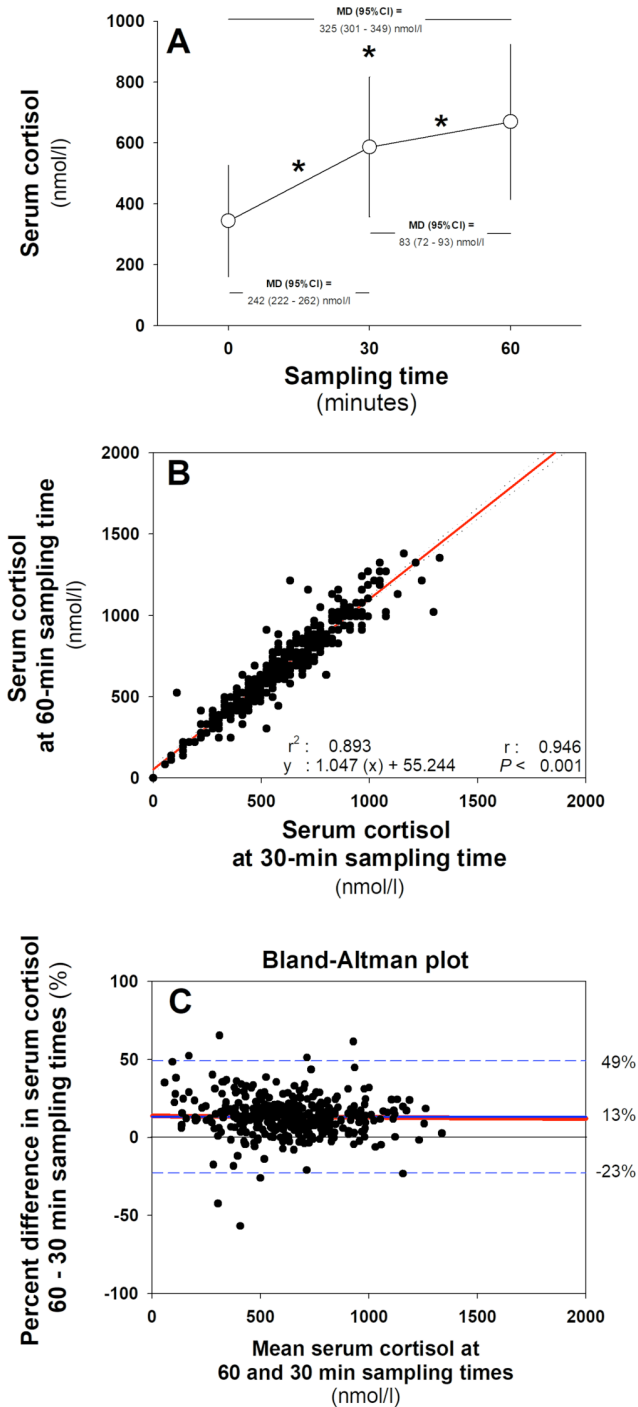


Figure 1 (A) Serum cortisol levels at different sampling times. Data are shown as mean (95% CI) and mean differences (MD) (95% CI). Comparisons among time points were performed by a repeated-measures analysis of variance addressing the main effects by a Bonferroni CI adjustment ($*p < 0.001$). (B) Pearson's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

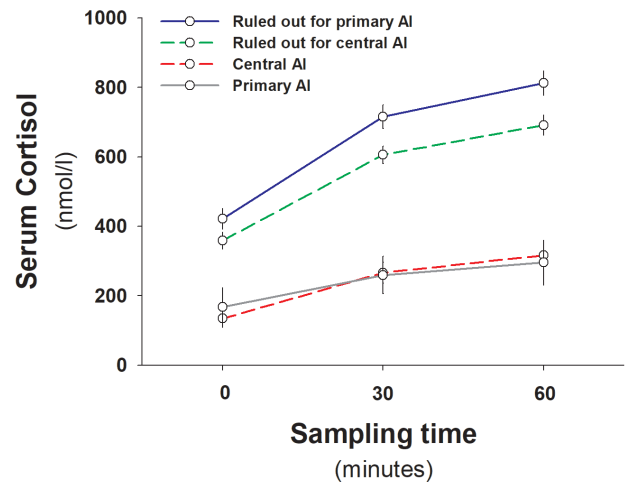


Figure 2 Baseline and stimulated serum cortisol concentrations as a function of clinical suspicion and response to cosyntropin test. Data are shown as mean and 95% CI. AI, adrenal insufficiency.

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 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, seven 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-specific and assay-specific cut-off values, in six cases (4.0%) the response was subnormal at 30 min but reached normal concentrations at 60 min. Four of them showed 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 (0.8–1.4 $\mu\text{g}/\text{dL}$), 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 13.2 $\mu\text{g}/\text{dL}$) at the 30 min sampling time: one had received oral glucocorticoid replacement therapy, which did not preclude the patient from responding to cosyntropin by showing an SC of 470 nmol/L (17 $\mu\text{g}/\text{dL}$) 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 suffer 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.

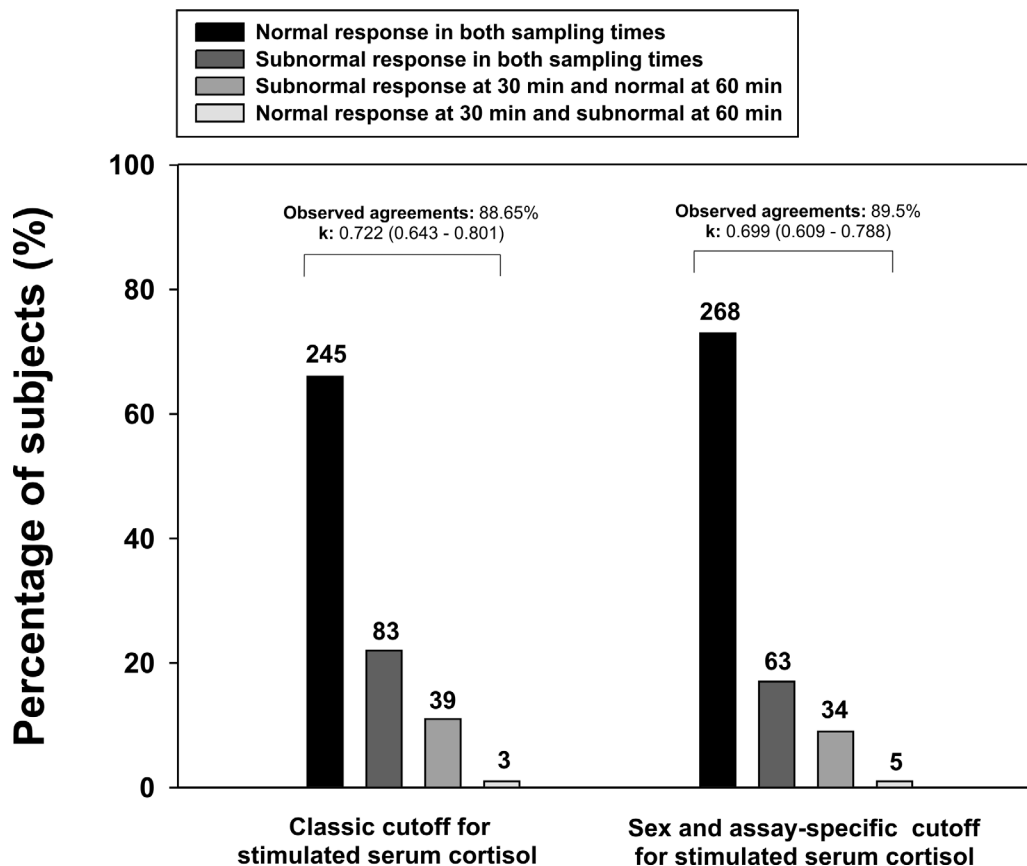


Figure 3 Subgroups of patients according to serum cortisol responses to cosyntropin stimulation as a function of classic and sex-specific and assay-specific cut-off values. Figures on top of the bars indicate the number of patients included in each subgroup. Diagnostic agreement is shown as the percentage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

Only three subjects (1.4%) presented with the opposite situation. Using sex-specific 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 five 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 the cut-off values used here are shown in table 3. SC concentrations at 60 min had 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 the classic or sex-specific and assay-specific cut-off values.

We observed discordant results between the classic and the sex-specific 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-specific and assay-specific cut-off values been used. In seven of them, SST was performed to rule

out primary AI, and in the remaining 40 subjects SST was 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 oestrogenic therapy and presented a normal response according to the classic cut-off value, but subnormal when considering sex-specific and assay-specific cut-off values, yet none of them required glucocorticoid therapy.

Confirmative group

Thirty (20%) of these women presented with a subnormal response to SST according to the classic cut-off values, yet this figure was reduced to only three (2%) when sex-specific and assay-specific cut-off values were used (observed agreement: 82%; κ : 0.151 (95% CI 0.066 to 0.235)). The three women showing a subnormal response during SST using sex-specific and assay-specific cut-off values showed stimulated SC concentrations of 342 nmol/L (12.4 μ g/dL), 353 nmol/L (12.8 μ g/dL) and 372 nmol/L (13.5 μ g/dL), whereas the LLNs (2.5th percentile) of SC concentrations at 30 min sampling time of SST were 436 nmol/L (15.8 μ g/dL) and 411 nmol/L (14.9 μ g/dL) for immunoassays 1 and 2, respectively. The fifth percentiles for both immunoassays were 450 nmol/L (16.3 μ g/dL) and 414 nmol/L (15.0 μ g/dL), respectively, showing minimal

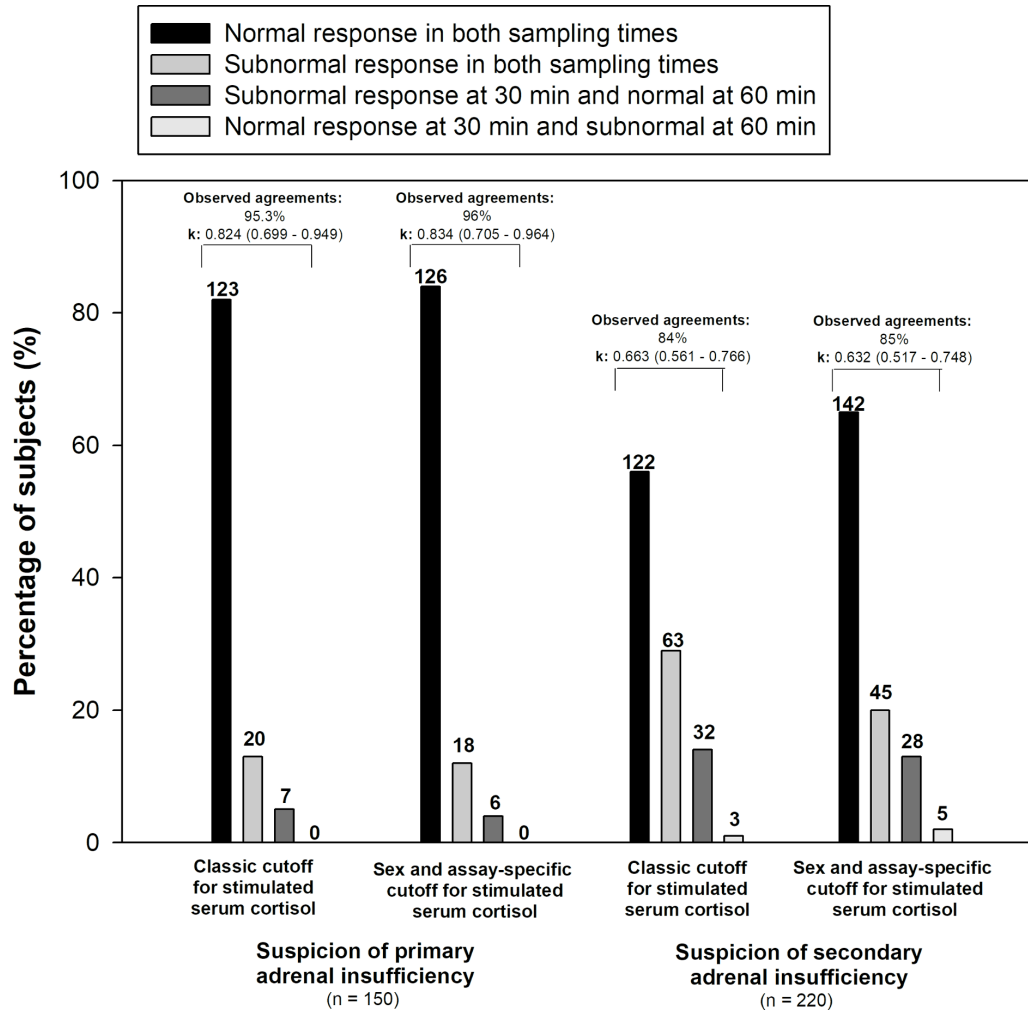


Figure 4 Subgroups of patients according to serum cortisol responses to cosyntropin stimulation as a function of cut-off values and clinical suspicion of primary or central adrenal insufficiency. Figures on top of the bars show the number of patients included in the different subgroups. Diagnostic agreement is shown as the percentage of observed agreements and kappa coefficients (95% CI).

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 to 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 analysed those data as a function of the suspected reason for screening AI. Supporting our previous findings, agreement between

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-specific and assay-specific), and as a function of the suspected origin of the disease

	Classic cut-off values						Sex-specific 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

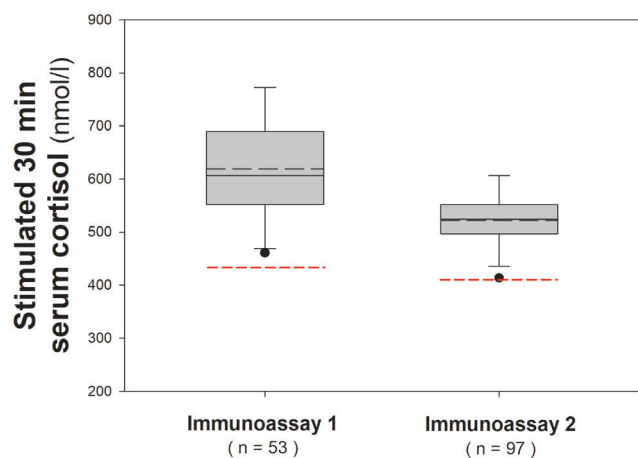


Figure 5 Descriptive statistics and distribution of 30 min cosyntropin-stimulated serum cortisol concentrations in a population of premenopausal healthy women with evidence of normal hypothalamic–pituitary–adrenal axis function. The boundary of the box closest to zero indicates the 25th percentile, the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

both SST sampling times was better when primary AI was suspected (observed agreement: 97%; κ : 0.846 (95% CI 0.714 to 0.977)) compared with a suspicion of central AI (observed agreement: 89%; κ : 0.667 (95% CI 0.548 to 0.785)), with the data being almost the same as observed in figure 4.

DISCUSSION

AI is a clinical condition associated with high morbidity and mortality. Unstimulated early morning SC values below 138 nmol/L (5 μ g/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 SST, in agreement with previous reports.²² Our data also show that both 30 and 60 min SC measurements during 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 SST appears to have 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 SST did not provide any additional diagnostic advantage

over performing a single determination at 60 min of the test. Although similar results have also been reported by others,^{13 14} these studies did not take into account the primary or central origin of AI and did not apply sex-specific 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 measurements at 30 min during SST, when using sex-specific 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 a 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 the 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 SST was performed. The use of sex-specific and assay-specific cut-off values appears to be essential, since other authors have suggested that some healthy individuals 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 a need for 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, favouring the use of 60 min samples over 30 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 two out of the five 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 for sex-specific and assay-specific cut-off values to interpret the results of SST, in agreement with recent clinical guidelines.³ The use of such cut-off values leads, in our study, to a reduction in FP results, higher Sp and PPV, less discordant results among sampling times of SST, and fewer unnecessary treatments (20 patients (5%) could have been treated unnecessarily if the classic cut-off values were applied for diagnosis). The reliability of sex-specific and assay-specific cut-off values was confirmed in our

population of premenopausal women with normal HPA axis, in whom these cut-off values were more appropriate than relying on the 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 2.5th 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 standardised 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 it impossible to rule out information bias. Our best efforts might not have been enough to avoid misclassification of patients according to the suspicion of primary or central AI. Also, the administration of suprphysiological 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 values 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 extrapolated 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 having been unethical in most cases, the lack of a laboratory gold standard such as an ITT did not over-ride 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 it only comprised premenopausal women and stimulated SC was only available at the 30 min sampling time. Lastly, we could not entirely rule out pretreatment 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-specific and assay-specific cut-off values of SC responses to cosyntropin results in 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 in supporting a clinical diagnosis of primary AI, yet 60 min measurements might be preferable when central AI is suspected.

Author affiliations

¹Department of Endocrinology and Nutrition, Hospital Universitario Ramón y Cajal, Madrid, Spain

²Diabetes, 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

³Department of Clinical Biochemistry, Hospital Universitario Ramon y Cajal, Madrid, Spain

Contributors AEO-F, ML-R designed the protocol and performed the statistical analysis. AEO-F, ES-C reviewed the clinical data using the electronic or written records if necessary. AG-C and LJ-M performed the electronic search of serum cortisol samples. AEO-F, ML-R wrote the first draft of the study. All the authors, including LN-C and HFE-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.

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. ML-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.

Competing interests None declared.

Patient consent Obtained.

Ethics approval We obtained approval from the local ethics committee (Ramón y Cajal University Hospital, Ethics Committee) before conducting the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Individual participant data that underlie the results are 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 the 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 requesters will need to sign a data access agreement.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Charmandari E, Nicolaidis NC, Chrousos GP. Adrenal insufficiency. *Lancet* 2014;383:2152–67.
2. Neary N, Nieman L. Adrenal insufficiency: etiology, diagnosis and treatment. *Curr Opin Endocrinol Diabetes Obes* 2010;17:217–23.
3. Bornstein SR, Allolio B, Arlt W, et al. Diagnosis and Treatment of Primary Adrenal Insufficiency: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2016;101:364–89.
4. Grossman AB. Clinical review#: the diagnosis and management of central hypoadrenalism. *J Clin Endocrinol Metab* 2010;95:4855–63.

5. Agha A, Tomlinson JW, Clark PM, *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.
6. Sprung CL, Annane D, Keh D, *et al.* Hydrocortisone therapy for patients with septic shock. *N Engl J Med* 2008;358:111–24.
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, *et al.* A 3-level prognostic classification in septic shock based on cortisol levels and cortisol response to corticotropin. *JAMA* 2000;283:1038–45.
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* 1987;26:53–9.
10. Hurel SJ, Thompson CJ, Watson MJ, *et al.* The short Synacthen and insulin stress tests in the assessment of the hypothalamic-pituitary-adrenal axis. *Clin Endocrinol* 1996;44:141–6.
11. González J, Villabona C, Ramón 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* 2000;53:199–204.
12. Longui CA, Vottero A, Harris AG, *et al.* Plasma cortisol responses after intramuscular corticotropin 1-24 in healthy men. *Metabolism* 1998;47:1419–22.
13. Mansoor S, Islam N, Siddiqui I, *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.
14. Chitale A, Musonda P, McGregor AM, *et al.* Determining the utility of the 60 min cortisol measurement in the short synacthen test. *Clin Endocrinol* 2013;79:14–19.
15. de Miguel Novoa P, Vela ET, García NP, *et al.* Guidelines for the diagnosis and treatment of adrenal insufficiency in the adult. *Endocrinol Nutr* 2014;61(Suppl 1):1–34.
16. Turpeinen U, Hämäläinen E. Determination of cortisol in serum, saliva and urine. *Best Pract Res Clin Endocrinol Metab* 2013;27:795–801.
17. El-Farhan N, Pickett A, Ducroq D, *et al.* Method-specific serum cortisol responses to the adrenocorticotrophin test: comparison of gas chromatography-mass spectrometry and five automated immunoassays. *Clin Endocrinol* 2013;78:673–80.
18. Klose M, Lange M, Rasmussen AK, *et al.* Factors influencing the adrenocorticotropin test: role of contemporary cortisol assays, body composition, and oral contraceptive agents. *J Clin Endocrinol Metab* 2007;92:1326–33.
19. Kontula K, Paavonen T, Luukkainen T, *et al.* Binding of progestins to the glucocorticoid receptor. Correlation to their glucocorticoid-like effects on in vitro functions of human mononuclear leukocytes. *Biochem Pharmacol* 1983;32:1511–8.
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;93:527–33.
21. Templeton G. A two-step approach for transforming continuous variables to normal: implications and recommendations for its research. *Commun Assoc Inf Syst* 2011;28.
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* 1998;49:287–92.
26. Bornstein SR. Predisposing factors for adrenal insufficiency. *N Engl J Med* 2009;360:2328–39.