Clinical Immunology (2011) 140, 291–299

available at www.sciencedirect.com

Clinical Immunology

www.elsevier.com/locate/yclim

# Measuring adrenal autoantibody response: Interlaboratory concordance in the first international serum exchange for the determination of 21-hydroxylase autoantibodies☆

Alberto Falorni[a,*,1], Shu Chen[b,1], Renato Zanchetta[c,1], Liping Yu[d,1], Claudio Tiberti[e,1], Maria Luisa Bacosi[a,1], Jadwiga Furmaniak[b,1], Vittorio Bini[a,1], Francesco Dotta[f,1], George S. Eisenbarth[d,1], Bernard Rees Smith[b,1], Corrado Betterle[c,1]

[a] *Department of Internal Medicine, Section of Internal Medicine and Endocrine & Metabolic Sciences, University of Perugia, 06126 Perugia, Italy*
[b] *FIRS Laboratories, RSR Ltd., Llanishen, Cardiff, CF14 5DU, UK*
[c] *Endocrine Unit, Department of Medical and Surgical Sciences, University of Padua, 35122 Padua, Italy*
[d] *Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Aurora, CO 80045, USA*
[e] *Department of Clinical Sciences, Sapienza University, 00161 Rome, Italy*
[f] *Department of Internal Medicine and Endocrine and Metabolic Sciences, University of Siena, 53100 Siena, Italy*

Received 21 March 2011; accepted with revision 21 April 2011
Available online 28 April 2011

**KEYWORDS**
Adrenal insufficiency;
Autoimmune Addison's disease;
Autoimmune polyendocrine syndrome;
Radioimmunoassay;
Standardization

**Abstract** 21-hydroxylase autoantibodies (21OHAb) are the gold standard immune marker to identify patients with clinical or subclinical autoimmune Addison's disease (AAD). No assessment of interlaboratory concordance has been made for 21OHAb measurement. Serum samples from 267 patients with primary adrenal insufficiency and from 83 healthy control subjects were distributed to four independent laboratories that determined presence and titer of 21OHAb, by using radiobinding assays with either *in vitro* translated $^{35}$S-radiolabelled or $^{125}$I-radiolabelled autoantigen. Cohen's κ of inter-rater agreement ranged from 0.857 to 0.983, showing a very good concordance of the positive/negative score among the four laboratories. Passing–Bablok regression showed a good agreement of 21OHAb titers arranged by ranks, but important discrepancies emerged at the Bland–

*Abbreviations* 21OHAb, 21-hydroxylase autoantibodies; AAD, Autoimmune Addison's Disease; ACA, Adrenal Cortex Autoantibodies; AHC, adrenal hypoplasia congenita; ALD, adrenoleukodystrophy; APS, autoimmune polyendocrine syndrome; AUC, area under ROC curve; CI, confidence interval; IAN, Italian Addison Network; ICC, intra-class correlation coefficient; PAI, primary adrenal insufficiency; RC, repeatability coefficient; ROC, receiver-operating characteristic; TBC, tuberculosis; T1DM, type 1 diabetes mellitus.

☆ *Disclosure statement:* RSR Ltd is a manufacturer of medical diagnostics including kits for 21OH autoantibodies.

* Corresponding author at: Department of Internal Medicine, Section of Internal Medicine and Endocrine & Metabolic Sciences, Via E. Dal Pozzo, 06126 Perugia, Italy. Fax: +39 075 5730855.
*E-mail address:* alberto.falorni@unipg.it (A. Falorni).

[1] On behalf of the Italian Addison Network.

1521-6616/$ - see front matter 
doi:10.1016/j.clim.2011.04.012

Altman plot, as the repeatability coefficient was much higher than the laboratory cut-offs, which indicates that results from different laboratories cannot be used interchangeably. A standardization international program for 21OHAb measurement is strongly needed.


## 1. Introduction

Primary adrenal insufficiency (PAI) affects approximately 1 in 7000 individuals [1,2]. The primitive adrenocortical deficit has an heterogeneous etiopathogenesis and may develop through several distinct mechanisms, including autoimmunity, infiltrative adrenalitis, adrenoleukodystrophy (ALD), genetic disorders, metastasis, adrenal hemorrhage, surgery, sepsis, infections or toxic agents [3,4]. Autoimmune Addison's Disease (AAD) is caused by an autoimmune process responsible for the selective destruction of adrenal cortex cells [5,6]. Although AAD is the result of a T-cell mediated process, adrenal autoimmunity is made evident by the appearance of circulating adrenal cortex autoantibodies (ACA), that represent the best immune marker to identify patients with AAD [5,6]. ACA have little or no pathogenetic role [5–8], but their detection in human serum is clinically useful for both the etiological classification of PAI [9] and the identification of subjects at high-risk for future development of clinical AAD [10–16].

From 1963 until the middle of the 1990s, ACA were exclusively detected by means of the indirect immunofluorescence technique on cryostatic sections of human or animal adrenal glands [17]. The identification of the steroidogenic enzyme 21-hydroxylase as the main autoantigen identified by ACA [18,19] led to the development of sensitive and specific immunoassays for the detection of 21-hydroxylase autoantibodies (21OHAb) in human serum [20–22]. Several subsequent studies have demonstrated that 21OHAb is the gold standard immune marker for diagnosis of clinical and pre-clinical AAD [5,6,10–16,20–29]. In Europe, from 1974 to 2010, over 2000 patients with PAI have been reported in the literature as having been evaluated for adrenal autoantibodies [5,6]. The prevalence of adrenal autoantibodies ranged from 44.5 to 94% in different studies [5,6].

The Italian Addison Network (IAN) has developed a comprehensive flow-chart for the etiological classification of PAI which takes into consideration immunological, biochemical and imaging data [9]. In this flow-chart, the analysis of adrenal autoantibodies, and more specifically of 21OHAb, plays a major role to discriminate autoimmune from non-autoimmune forms of the disease [9]. Studies on subjects with organ-specific autoimmune diseases have shown that not only the presence, but also levels, of 21OHAb may have clinical relevance in the estimates of future risk for development of the clinical signs of AAD [13].

In the IAN study already mentioned [9], a comparison of 21OHAb assays in two independent laboratories and of ACA in two other independent laboratories was performed. The concordance rate of the two laboratories that performed 21OHAb assays was higher than that of the two laboratories that performed ACA assays [9]. As an extension of the previous IAN study, a novel study was planned in which four independent laboratories (two from Italy and two from other countries) performed 21OHAb assays in a large series of samples from patients with PAI and healthy control subjects. The aim of the present study was specifically that of comparing 21OHAb results generated by different laboratories using their own internal standard sera and in-house calculated cut-offs, to define the need for future standardization programs aimed at harmonizing the methods and identifying a common standard serum. Accordingly, serum samples were consecutively collected from PAI patients irrespective of disease duration or etiology.

**Table 1** 21OHAb positivity in different clinical forms of PAI, in 4 different laboratories.

| | PG | RM | CAR | DEN |
|---|---|---|---|---|
| Idiopathic | 168/205 (82%) | 168/205 (82%) | 173/205 (84.4%) | 170/205 (82.9%) |
| APS1 | 14/15 (93.3%) | 14/15 (93.3%) | 13/15 (86.7%) | 14/15 (93.3%) |
| ALD | 0/8 | 0/8 | 2/8 (25%) | 0/8 |
| AHC | 0/1 | 0/1 | 0/1 | 0/1 |
| Post-sepsis | 0/1 | 0/1 | 0/1 | 0/1 |
| Post-surgical | 0/1 | 1/1 | 0/1 | 0/1 |
| Post-TBC | 0/36 | 0/36 | 1/36 (2.8%) | 1/36 (2.8%) |
| Healthy control subjects | 0/83 | 0/83 | 2/83 (2.4%) | 0/83 |

AHC: adrenal hypoplasia congenita; ALD: adrenoleukodystrophy; APS1: autoimmune polyendocrine syndrome type 1; Post-TBC: post-tuberculosis primary adrenal insufficiency.

## 2. Materials and methods

### 2.1. Serum samples

Serum samples from 267 patients (107 males and 160 females) with PAI and from 83 healthy control subjects (35 males and 48 females, median age 39 years, range 21–55 years) were collected by the IAN, as previously described [9], and stored at −70 °C until subsequently used for the present study. At the time of sample collection, the median patient age was 49 years (range, 6–87 years) and the median disease duration was 5 years (range, 0–53 years). In all patients, clinical symptoms and signs of PAI were associated with low basal cortisol (<3 μg/dl) and high basal ACTH (>100 pg/ml) levels. For statistical purposes, PAI patients were subdivided into two groups: a) clinically idiopathic/APS1 (n=220; this group included 15 cases with autoimmune polyendocrine syndrome type 1), and b) with demonstrated non-autoimmune etiology (such as post-infiltrative adrenalitis, n=36, ALD, n=8, adrenal hypoplasia

**Table 2** Cohen's κ of inter-rater agreement among the four laboratories.

| | RM | CAR | DEN |
|---|---|---|---|
| PG | 0.983<br>95% CI: 0.963–1.000 | 0.880<br>95% CI: 0.830–0.929 | 0.971<br>95% CI: 0.946–0.996 |
| RM | | 0.862<br>95% CI: 0.809–0.915 | 0.954<br>95% CI: 0.923–0.986 |
| CAR | | | 0.885<br>95% CI: 0.836–0.934 |

congenita, n=1, sepsis, n=1 or surgical adrenalectomy, n=1). All patients gave their written informed consent for this study.

All sera were coded by the laboratory of the University of Perugia, Italy and redistributed, blindly, to all the participating laboratories, as frozen aliquots shipped in dry ice. Altogether, four independent laboratories were involved in this study for testing 21OHAb by different techniques: Perugia (Italy) (lab code: PG), Rome (Italy) (lab code: RM), Denver (CO, USA) (lab code: DEN) and Cardiff (UK) (lab code: CAR).

## 2.2. 21OHAb assays

Three laboratories (PG, RM and DEN) used a radiobinding assay with *in vitro* translated recombinant human $^{35}$S-21OH and immunoprecipitation of the immunocomplexes using protein A-Sepharose beads [21]. Autoantibody levels were evaluated by counting the immunoprecipitated radioactivity in a multiwell-plate scintillation counter (Top Count, Canberra Packard) and expressed as a relative index based on the analysis of one positive and two negative standard sera in each assay (21OHAb index) using the formula: (cpm sample − mean cpm negative controls) / (cpm positive control − mean cpm negative controls). The cut-offs defined by the three laboratories were 0.060, 0.060 and 0.150 for PG, RM and DEN, respectively. Our laboratory in Perugia reported previously an intra-assay CV of 7–12% [21].

Our laboratory in Cardiff used a radiobinding assay with $^{125}$I and the immunoprecipitation of immunocomplexes by protein A [22]. The radiobinding was counted on a gammacounter. The cut off was established as 1 unit/mL (arbitrary units). Inter-assay precision on 25 determinations of the same serum sample was mean of 7.5 units/mL with SD=0.49 and CV=6.58%. Intra-assay precision obtained with 25 determinations of the same serum sample was mean of 5.7 units/mL with SD=0.32 and CV of 5.6%.

## 2.3. Statistical methods

The degree of concordance for the dichotomous variable presence/absence of 21OHAb between two different laboratories was estimated using the Cohen's κ-test of interrater agreement [30] (κ-value ranging from 0 to 1), with Fleiss–Cuzick extension [31] when appropriate. Interpretation of strength of agreement was made according to the Landis and Koch's gradation: <0.2 = poor, 0.2–0.4 = fair, 0.4–0.6 = moderate, 0.6–0.8 = good, 0.8–1.0 = very good [32]. Passing–Bablok regression analysis [33] was performed for each participating laboratory using the median rank of each sample in the four laboratories and the rank of each sample in each laboratory as variables. This type of analysis does not depend on the assignation of each variable as x or y and enables calculation of the regression slope and y axis intercept, which are then used to determine if they differ significantly from the ideal slope=1 and intercept=0. Linearity was evaluated by CUSUM test which rejects the null hypothesis of parameter constancy whenever the range of cumulated ordinary least squares residuals becomes too large [34].

The Bland–Altman plot technique [35], or difference plot technique, was used to compare quantitative 21OHAb assay results in two independent laboratories. This method was used only for the three assays that expressed 21OHAb levels as a relative index (21OHAb index). In this graphical method, the differences between the two techniques were plotted against the averages of the two techniques. Intra-class correlation coefficient (ICC value ranging from 0 to 1), defined as the proportion of variance of an observation due to between-subject variability in the true scores, assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. The type of ICC was 2,1 [36]. A high ICC indicates that there is little variation between the scores given to each item by the raters. Kendall's τ correlation coefficient between intra-subject standard deviation and intra-subject mean, used to assess the interdependence of these two parameters, were also calculated. The repeatability coefficient (RC) was used as a precision measure which represents the value below which the absolute difference between two repeated test results may be expected to lie with a probability of 95%.

Diagnostic sensitivity (as determined in patients with clinically idiopathic PAI/APS1, n=220) and specificity (as determined in healthy control subjects, n=83) were estimated by using receiver-operating characteristic (ROC) curves with cut-offs corresponding to the best combination of high sensitivity and high specificity. Differences in area under ROC curves (AUC) were tested with modified Z-test [37]. In all analysis, a $p$ value<0.05 was considered significant. Statistical analyses were performed using Predictive Analytic Software (PASW) release 17.0.2, (SPSS Inc., Chicago, USA, 2009), and StatsDirect version 2.7.2 (StatsDirect Ltd, Altrincham, Cheshire, UK, 2008).

# 3. Results

Prevalence of 21OHAb in different forms of PAI is reported in Table 1. 21OHAb were detected in 82–84.4% patients with idiopathic PAI, 86.7–93.3% patients with APS1, 0 to 25% patients with ALD and 0 to 2.8% patients with post-TBC PAI.

One serum from a post-surgical PAI patient was found negative in three assays and positive in one assay. On the other hand, the serum from a patient with post-sepsis PAI and the serum from a patient with adrenal hypoplasia congenita were found negative in all four assays. Positivity of 21OHAb in 83 healthy control subjects ranged from 0 to 2.4%. To evaluate qualitative concordance among the four autoantibody assays, namely classification of subjects as positive or negative, we used the Cohen's $\kappa$ of inter-rater agreement. The general agreement over the two categories with four raters per subject (evaluated by using the Fleiss–Cuzick extension) was very good ($\kappa = 0.923$, 95% CI: 0.880–0.965, $p < 0.0001$). A very good agreement was also observed in the comparisons of each laboratory with each other (Table 2).

Globally, 175 samples were found positive and 151 were found negative in all the four laboratories (representing 93% of the total samples having a 100% concordance in classification as positive or negative).

To evaluate diagnostic sensitivity and specificity of the four 21OHAb assays, ROC curves were also generated (Fig. 1), and AUC for each laboratory was calculated. AUC was high in all the four laboratories ranging from 0.900 (95% CI: 0.865–0.935) to 0.969 (95% CI: 0.952–0.987). Statistically significant differences in AUC were observed between DEN and PG ($p = 0.002$), DEN and CAR ($p < 0.001$) and between RM and CAR (0.011).

To analyze quantitative results, conversion of 21OHAb values as ranks was needed, because one of the laboratories expressed its results using different measure units. When rank in each laboratory was plotted against median rank in the four laboratories, no particular trend related to single laboratories was observed (Fig. 2). A clear subdivision in two distinct subpopulations (positive and negative values) was evident. The two subpopulations showed a different pattern, as negative results (left side of the graph) appeared more dispersed than positive results (right side of the graph).

Passing–Bablok test was then performed for each laboratory. No significant deviation from linearity was observed in any of the tested laboratories (CUSUM test, $p > 0.10$). Slope of the regression line resulted 0.967 (95% CI: 0.927–1.009) for PG, 0.990 (95% CI: 0956–1.024) for RM, 1.021 (95% CI: 0.980–1.062) for CAR and 1.025 (95% CI: 1.000–1.057) for DEN. Intercept resulted 13.11 (95% CI: 3.165–23.33) for PG, 5.450 (95% CI: −0.8095 to 12.0263) for RM, −4.08 (95% CI: −13.92–5.636) for CAR and −5.612 (95% CI: −12.44–0) for DEN. PG data were more dispersed than those generated in the other laboratories, as 0 was not included in the 95% CI of the intercept for that laboratory. We hypothesized that this was likely caused by a higher dispersion of negative values than in the other laboratories. To test this hypothesis, we repeated the Passing–Bablok analysis taking into consideration only the positive results (Fig. 3). When negative samples were excluded, slope of the regression line was 0.980 (95% CI: 0.936–1.023) for PG, 1.000 (95% CI: 0.938–1.062) for RM, 0.944 (95% CI: 0.876–1.006) for CAR and 1.000 (95% CI: 0.959–1.053) for DEN. On the other hand,

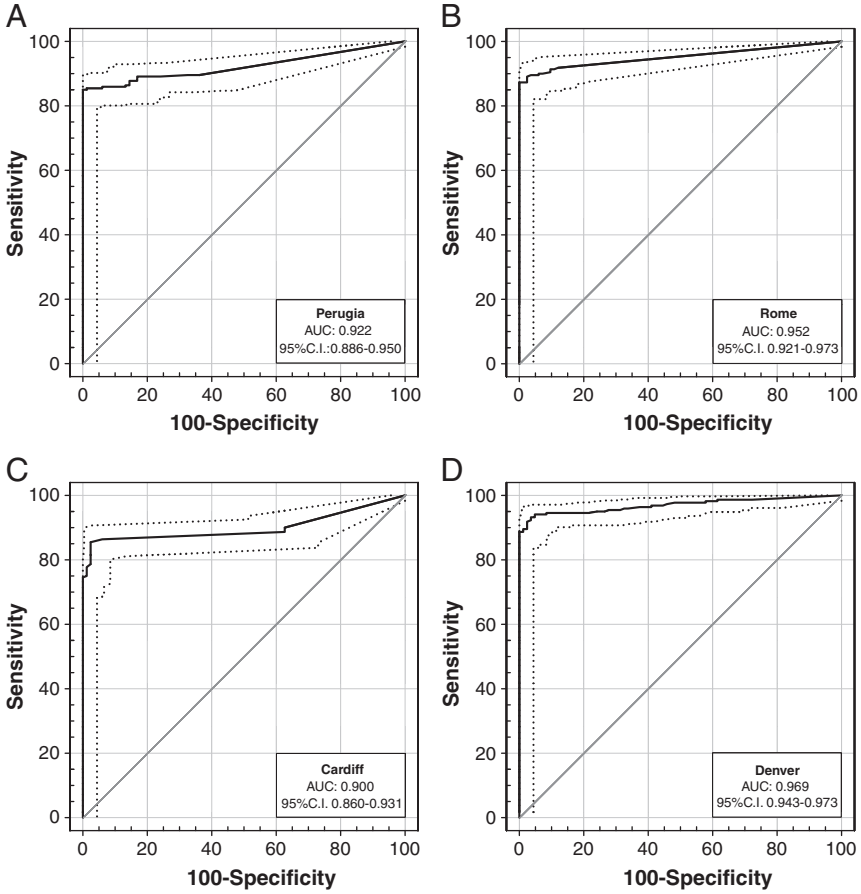


**Figure 1** Receiving-operator-characteristic (ROC) curves for PG (panel A), RM (panel B), CAR (panel C) and DEN (panel D). AUC: Area under the ROC curve.

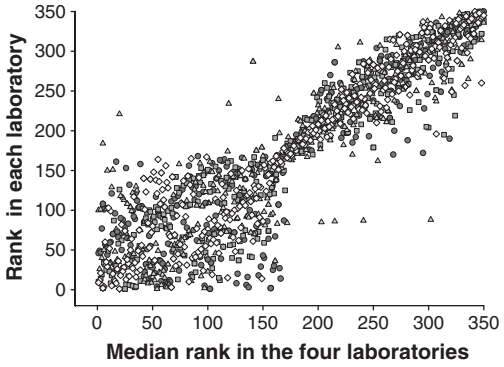


**Figure 2** 21OHAb data generated by PG (circles), RM (squares), CAR (triangles) and DEN (diamonds) expressed as median rank in the four laboratories (x axis) *vs.* rank in the single laboratory (y axis).

intercept results were 9.687 (95% CI: −1.454–22.10) for PG, 2.000 (95% CI: −12.5000 to 17.9538) for RM, 16.72 (95% CI: 0.770–35.01) for CAR and 1.000 (95% CI: −11.45–10.76) for DEN. Accordingly, analysis of positive results was associated with an improvement of the regression line of PG (which supports the hypothesis that most of the dispersion was due to negative values, in that laboratory), but not of CAR, for which a significant deviation from linearity appeared ($p<0.05$).

Agreement of the continuous 21OHAb index among PG, RM and DEN was tested using the Bland–Altman plot analysis. Although all the three intra-class correlation coefficients were rather high, being 0.709 for the comparison between PG and RM, 0.765 for the comparison between PG and DEN and 0.849 for the comparison between RM and DEN (which indicates that high levels in one laboratory tend to correspond to high levels also in the other), 95% limits of agreement of 21OHAb index were very broad, ranging from −0.675 to 0.888 for the comparison between PG and RM (Fig. 4A), from −0.708 to 0.832 for the comparison between PG and DEN (Fig. 4B) and from −0.219 to 0.555 for the comparison between RM and DEN (Fig. 4C). In all cases, cut-offs of the three laboratories were included within the 95% limits of agreement, which indicates that differences between laboratories may potentially have clinical relevance and the different 21OHAb tests cannot be used interchangeably. Similarly, the repeatability coefficient was much higher than the three laboratory cut-offs, corresponding to a 21OHAb index of 0.808 for the comparison between PG and RM, 0.779 for the comparison between PG and DEN and 0.508 for the comparison between RM and DEN. Correlation between mean values in the same subject and the correspondent standard deviation was highly significant ($p<0.0001$) in all three comparisons (Figs. 4D–F), which indicates that the error was strongly dependent on the size of the measure, being higher for samples with higher 21OHAb index.

## 4. Discussion

21OHAb is the best single immune marker useful to both identify patients with clinical AAD and subjects at high-risk to develop clinical and biochemical signs of adrenal insufficiency

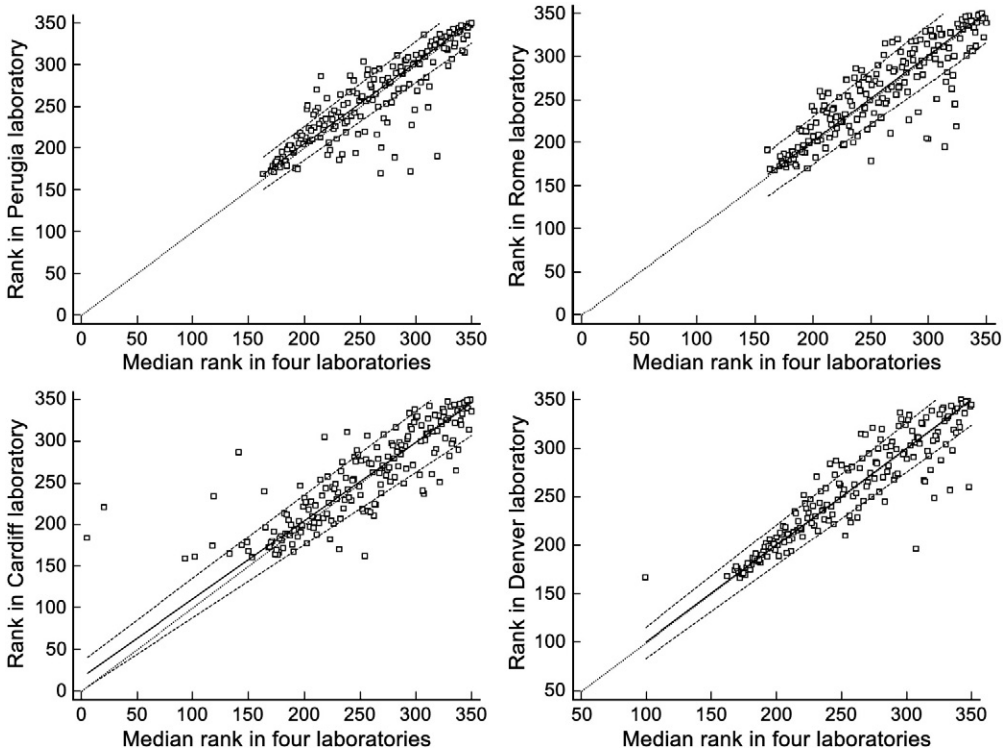


**Figure 3** Passing–Bablok regression test for PG, RM, CAR and DEN. Data are expressed as median rank in the four laboratories (x axis) *vs.* rank in the single laboratory (y axis).

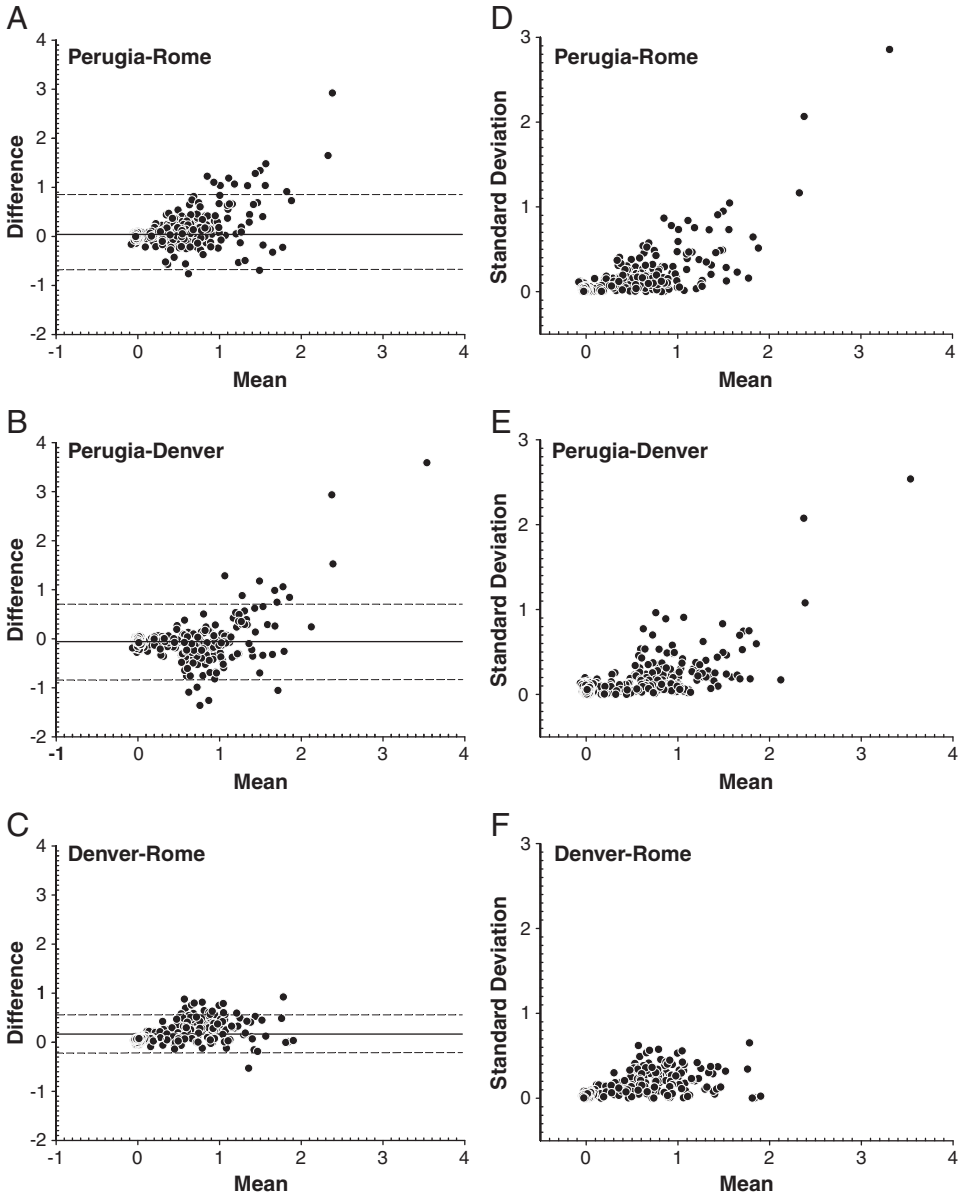


**Figure 4** Bland–Altman plot analysis: agreement plot (panel A, B and C) and repeatability plot (panels D, E, F).

[5,6,9–29]. More specifically, the unequivocal diagnosis of AAD is based on the detection of 21OHAb in serum samples. Accordingly, imprecisions in adrenal autoantibody tests may have important effects on clinical management of PAI, as a low diagnostic specificity may determine an erroneous classification of ALD patients as AAD, or, in the case of a low diagnostic sensitivity, individuals with pre-clinical AAD may go undiagnosed until an Addisonian crisis occurs. Hence, it is important that currently available tests for determination of 21OHAb be compared and standardized. Although 21OHAb have been tested in several studies in patients with PAI or other diseases, no attempt at standardizing this test has yet been made. The long-term experience of several workshops for the standardization of islet autoantibody determinations [38–41] is highly valuable as a model for the future organization of standardization programs for 21OHAb and other steroid-cell autoantibodies. A complicating factor is that diagnosis of AAD depends on the detection of adrenal autoantibodies and is not independently formulated on pure clinical grounds, as it is in the case of T1DM. Accordingly, in our present study, we tested diagnostic sensitivity in patients with clinically idiopathic PAI, compared to normal control subjects, after exclusion of known non-autoimmune causes of the disease. This may have resulted in an underestimation of the actual diagnostic sensitivity of 21OHAb for AAD. In addition, patients with long-term disease duration were also enrolled in our study, and this may further have reduced diagnostic sensitivity, as we cannot exclude that some samples scored negative because of the disappearance of previously present autoantibodies. However, the aim of the present study was not to estimate the frequency of this marker in newly diagnosed AAD patients, but exclusively that of comparing the autoantibody results among different laboratories, independently from etiological classification of PAI and disease duration, and the above limitations have little impact

on the observed results. In addition, the study was not aimed at identifying a common standard serum to be used in laboratories throughout the world, but at providing initial information on the agreement for 21OHAb assays in selected research laboratories with long-term experience in the field. More specifically, we did not attempt to harmonize the assays before the samples were tested, nor to address traceability using common materials, as our study intended to use the data as they were generated by the laboratories using internal standards and in-house calculated cut-offs. Accordingly, we did not calculate a common study cut-off, because this would mean attempting to standardize the different methods, which was beyond the scope of our present investigation and is more appropriate with a larger number of participating laboratories.

In a previous IAN study [9], 21OHAb and ACA assays were each compared in two independent Italian laboratories. That study demonstrated the higher inter-laboratory concordance and higher diagnostic accuracy of 21OHAb for the diagnosis of AAD, as compared to ACA [9]. Subsequently, a large series of serum samples from patients with PAI and healthy control subjects collected by IAN were randomized and blindly distributed to four laboratories, which have been very active in adrenal and other autoantibody research in the past. Our present study represents the first international serum exchange for comparison of 21OHAb determination.

Statistical analyses were aimed at evaluating concordance on two distinct aspects of 21OHAb determination: qualitative results (positive/negative score) and quantitative data (21OHAb titer). However, not all the laboratories expressed 21OHAb titer using similar scale units and some analyses could not be performed for all laboratories.

The Cohen's $\kappa$-test showed very good inter-laboratory agreement on positive/negative scores, which confirms that diagnostic accuracy of 21OHAb for AAD is extremely high, as shown by previous publications from the participating laboratories as well as from other groups. To evaluate diagnostic sensitivity and specificity of the four assays we constructed ROC curves, using the results obtained with sera from idiopathic/APS1 AAD patients and from healthy control subjects, and we calculated the area under each curve (AUC). All four ROC curve results were very good and the AUC was higher than 0.9 in all cases (Fig. 1). Again, as previously discussed, each calculated AUC might not be totally accurate, as diagnostic sensitivity was determined using patients for which an autoimmune origin could not be demonstrated in all cases. Nevertheless, one aim of the study was to compare the AUCs generated by the different laboratories, and a good agreement was observed with this analysis. The results of our study are in line with those observed for other organ-specific autoantibodies tested using similar assay types [40]. A slightly lower AUC was observed for CAR as compared to RM and DEN, which is mostly due to a somewhat lower diagnostic specificity for that laboratory as compared to the other three participating laboratories, even though it showed the highest diagnostic sensitivity (Table 1). However, the simple calculation of the area under the ROC curve may lead to erroneous conclusions as this value may be influenced by variables (such as dispersion of negatively scored samples) that have little impact on clinical application. Accordingly, the statistically significant differences observed in the ROC analysis among different laboratories do not necessarily correspond to actual differences in diagnostic sensitivity and specificity reported by each laboratory (Table 1).

To compare quantitative data, two separate statistical analyses were performed. In the first series of analysis, the results from each laboratory were plotted *vs.* the global results of the four laboratories. As the participating laboratories expressed their data using different scale units, samples were transformed in ascending rank (from the lowest to the highest) in each laboratory and the median rank from all laboratories was also determined. This type of analysis has previously been used also to compare data for islet autoantibodies from different laboratories [38]. It was evident that a large dispersion of the values below the cut-offs was present among the different laboratories (Fig. 2). More specifically, PG had the highest dispersion of the values of the samples scored as "negative" for 21OHAb. Passing–Bablok regression analysis [33] was chosen because it does not depend on the assignation of each variable as x or y and enables calculation of the regression slope and y axis intercept. Using this test, all the laboratories had regression slopes and y axis intercepts which were not significantly different from 1 to 0, respectively, with the exception of PG, for which 95% confidence intervals of the intercept did not include 0, mostly because of the already mentioned high dispersion of samples scored negative. When the analysis was repeated taking into consideration only samples scored positive for 21OHAb, a significant deviation from linearity was detected only for CAR. This lack of linearity of CAR was however caused by only 2 samples considered positive by this lab and clearly negative for the remaining three (Fig. 3C). Globally, all the laboratories performed highly satisfactorily in the Passing–Bablok analysis and the observed small deviations from the expected, ideal line were due either to the negative samples (for which the actual 21OHAbindex has no clinical relevance) or to a very limited number of samples discordant with other laboratories.

The final analysis aimed at comparing 21OHAbindex between two laboratories. CAR was not included in the Bland–Altman plot test because of the different scale units used. This analysis revealed major discrepancies among the three laboratories that may potentially have clinical relevance. In all comparisons, the repeatability coefficient was much higher than the three laboratory cut-offs, thus showing that the results generated by the three laboratories cannot be used interchangeably. Interestingly, the higher the 21OHAbindex the higher was the observed difference among laboratories. This phenomenon may in part reduce the clinical impact of the inter-laboratory discrepancies, as one may expect that for samples with low 21OHAbindex a lower discrepancy might exist, but the lack of agreement in 21OHAbindex requires additional, larger international programs of standardization of 21OHAb determination.

## 5. Conclusions

We have reported the results of the first international serum exchange for the determination of 21OHAb. The four participating laboratories showed a high agreement as positive/negative score, but some major discrepancies emerged for the quantitation of 21OHAb titer. The organization of a standardization study aimed at identifying a standard serum that may enable expression of results in

common international units is strongly recommended. The identification of the international standard serum will also be instrumental for the harmonization of the different assays and the calculation of a uniform cut-off level.

## Role of the funding sources

AF is chairing a Sub-Committee on Organ-Specific Autoantibodies of the IUIS Committee on Quality Assessment and Standardization (www.iuisonline.org) and received an unconditioned support from IUIS. The study was in part supported by EU FP7, Grant number 201167, Euradrenal to AF and CB. The funding sources had no direct role in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, or in the ultimate decision to submit the paper for publication.

## Acknowledgments

The following members of the Italian Addison Network contributed to the collection of data and blood samples from patients with primary adrenal insufficiency: B. Ambrosi (Milan), G. Angeletti (Perugia), G. Arnaldi (Ancona), E. Arvat (Turin), G. Basta (Perugia), P. Beck-Peccoz (Milan), A. Bellastella (Naples), A. Bizzarro (Naples), M. Boscaro (Ancona), F. Cavagnini (Milan), F. Calcinaro (Perugia), R. Celleno (Perugia), C. Dal Pra (Padua), A. De Bellis (Naples), E. Ghigo (Turin), R. Giordano (Turin), L. Iorio (Milan), S. Laureti (Perugia), F. Loré (Siena), M. Mannelli (Florence), F. Mantero (Padua), S. Marzotti (Perugia), F. Pecori Giraldi (Milan), R. Perniola (Lecce), F. Santeusanio (Perugia), M. Terzolo (Orbassano), P. Toja (Milan), M. Torlontano (S. Giovanni Rotondo), V. Toscano (Rome), V. Trischitta (S. Giovanni Rotondo).

## References

[1] S. Laureti, L. Vecchi, F. Santeusanio, A. Falorni, Is the prevalence of Addison's disease underestimated? J. Clin. Endocrinol. Metab. 84 (1999) 1762.
[2] K. Løvås, E.S. Husebye, High prevalence and increasing incidence of Addison's disease in western Norway, Clin. Endocrinol. (Oxf.) 56 (2002) 787–791.
[3] W. Oelkers, Adrenal insufficiency, N. Engl. J. Med. 335 (1996) 1206–1212.
[4] S. Laureti, F. Santeusanio, A. Falorni, Recent advances in the diagnosis and therapy of primary adrenal insufficiency, Curr. Med. Chem. Immun. End. Metab. Agents 2 (2002) 251–272.
[5] C. Betterle, C. Dal Pra, F. Mantero, R. Zanchetta, Autoimmune adrenal insufficiency and autoimmune polyendocrine syndromes: autoantibodies, autoantigens, and their applicability in diagnosis and disease prediction, Endocr. Rev. 23 (2002) 327–364.
[6] A. Falorni, A. Brozzetti, F. Calcinaro, S. Marzotti, F. Santeusanio, Recent advances in adrenal autoimmunity, Exp. Rev. Clin. Endocrinol. Metab. 4 (2009) 333–348.
[7] M. Boscaro, C. Betterle, M. Volpato, F. Fallo, J. Furmaniak, B. Rees-Smith, N. Sonino, Hormonal responses during various phases of autoimmune adrenal failure: no evidence for 21-hydroxylase enzyme activity block in vivo, J. Clin. Endocrinol. Metab. 181 (1996) 2801–2804.
[8] C. Betterle, C. Dal Pra, B. Pedini, R. Zanchetta, M.P. Albergoni, S. Chen, J. Furmaniak, B. Rees-Smith, Assessment of adrenocortical function and autoantibodies in a baby born to a mother with autoimmune polyglandular syndrome type 2, J. Endocrinol. Invest. 27 (2004) 618–621.
[9] A. Falorni, S. Laureti, A. De Bellis, R. Zanchetta, C. Tiberti, G. Arnaldi, V. Bini, P. Beck-Peccoz, A. Bizzarro, F. Dotta, F. Mantero, A. Bellastella, C. Betterle, F. Santeusanio, Italian Addison Network Study: update of diagnostic criteria for the etiological classification of primary adrenal insufficiency, J. Clin. Endocrinol. Metab. 89 (2004) 1598–1604.
[10] C. Betterle, M. Volpato, B. Rees Smith, J. Furmaniak, S. Chen, N.A. Greggio, M. Sanzari, F. Tedesco, B. Pedini, M. Boscaro, F. Presotto, I. Adrenal cortex and steroid 21-hydroxylase autoantibodies in adult patients with organ-specific autoimmune diseases: markers of low progression to clinical Addison's disease, J. Clin. Endocrinol. Metab. 82 (1997) 932–938.
[11] C. Betterle, M. Volpato, B. Rees Smith, J. Furmaniak, S. Chen, R. Zanchetta, N.A. Greggio, B. Pedini, M. Boscaro, F. Presotto, II. Adrenal cortex and steroid 21-hydroxylase autoantibodies in children with organ-specific autoimmune diseases: markers of high progression to clinical Addison's disease, J. Clin. Endocrinol. Metab. 82 (1997) 939–942.
[12] K.W. Brewer, V.S. Parziale, G.S. Eisenbarth, Screening patients with insulin-dependent diabetes mellitus for adrenal insufficiency, N. Engl. J. Med. 337 (1997) 202.
[13] S. Laureti, A. De Bellis, V.I. Muccitelli, F. Calcinaro, A. Bizzarro, R. Rossi, A. Bellastella, F. Santeusanio, A. Falorni, Levels of adrenocortical autoantibodies correlate with the degree of adrenal dysfunction in subjects with preclinical Addison's disease, J. Clin. Endocrinol. Metab. 83 (1998) 3507–3511.
[14] A. De Bellis, A. Falorni, S. Laureti, S. Perrino, C. Coronella, F. Forini, E. Bizzarro, A. Bizzarro, G. Abbate, A. Bellastella, Time course of 21-hydroxylase antibodies and long-term remission of subclinical autoimmune adrenalitis after corticosteroid therapy: case report, J. Clin. Endocrinol. Metab. 86 (2001) 675–678.
[15] J.M. Barker, A. Ide, C. Hostetler, L. Yu, D. Miao, P.R. Fain, G.S. Eisenbarth, P.A. Gottlieb, Endocrine and immunogenetic testing in individuals with type 1 diabetes and 21-hydroxylase autoantibodies: Addison's disease in a high-risk population, J. Clin. Endocrinol. Metab. 90 (2005) 128–134.
[16] G. Coco, C. Dal Pra, F. Presotto, M.P. Albergoni, C. Canova, B. Pedini, R. Zanchetta, S. Chen, J. Furmaniak, B. Rees Smith, F. Mantero, C. Betterle, Estimated risk for developing autoimmune Addison's disease in patients with adrenal cortex autoantibodies, J. Clin. Endocrinol. Metab. 91 (2006) 1637–1645.
[17] C. Betterle, C. Scalici, F. Presotto, B. Pedini, L. Moro, F. Rigon, F. Mantero, The natural history of adrenal function in autoimmune patients with adrenal autoantibodies, J. Endocrinol. 117 (1988) 467–475.
[18] O. Winqvist, F.A. Karlsson, O. Kämpe, 21-hydroxylase, a major autoantigen in idiopathic Addison's disease, Lancet 339 (1992) 1559–1562.
[19] J. Bednarek, J. Furmaniak, N. Wedlock, Y. Kiso, A. Baumann-Antczak, S. Fowler, H. Krishnan, J.A. Craft, B. Rees-Smith, Steroid 21-hydroxylase is a major autoantigen involved in adult onset autoimmune Addison's disease, FEBS Lett. 309 (1992) 51–55.
[20] J. Colls, C. Betterle, M. Volpato, L. Prentice, B. Rees Smith, J. Furmaniak, Immunoprecipitation assay for autoantibodies to steroid 21-hydroxylase in autoimmune adrenal diseases, Clin. Chem. 41 (1995) 375–380.
[21] A. Falorni, A. Nikoshkov, S. Laureti, E. Grenbäch, A.L. Hulting, G. Casucci, F. Santeusanio, P. Brunetti, H. Luthman, Å. Lernmark, High diagnostic accuracy for idiopathic Addison's disease with a sensitive radiobinding assay for autoantibodies against recombinant human 21-hydroxylase, J. Clin. Endocrinol. Metab. 80 (1995) 2752–2755.
[22] H. Tanaka, M.S. Perez, M. Powell, J.F. Sanders, J. Sawicka, S. Chen, L. Prentice, T. Asawa, C. Betterle, M. Volpato, B. Rees Smith, J. Furmaniak, Steroid 21-hydroxylase autoantibodies:

measurements with a new immunoprecipitation assay, J. Clin. Endocrinol. Metab. 82 (1997) 1440–1446.
[23] A. Falorni, S. Laureti, A. Nikoshkov, M.L. Picchio, B. Hallengren, C.L. Vandewalle, F.K. Gorus, Belgian Diabetes Registry, C. Tortoioli, H. Luthman, P. Brunetti, F. Santeusanio, 21-hydroxylase autoantibodies in adult patients with endocrine autoimmune diseases are highly specific for Addison's disease, Clin. Exp. Immunol. 107 (1997) 341–346.
[24] J. Seissler, M. Schott, H. Steinbrenner, P. Peterson, W.A. Scherbaum, Autoantibodies to adrenal cytochrome P450 antigens in isolated Addison's disease and autoimmune polyendocrine syndrome type II, Exp. Clin. Endocrinol. Diabetes 107 (1999) 208–213.
[25] C. Betterle, M. Volpato, B. Pedini, S. Chen, B. Rees Smith, J. Furmaniak, Adrenal-cortex autoantibodies and steroid-producing cells autoantibodies in patients with Addison's disease: comparison of immunofluorescence and immunoprecipitation assays, J. Clin. Endocrinol. Metab. 84 (1999) 618–622.
[26] R. Do Carmo Silva, C.E. Kater, S.A. Dib, S. Laureti, F. Forini, A. Cosentino, A. Falorni, Autoantibodies against recombinant human steroidogenic enzymes 21-hydroxylase, side-chain cleavage and 17alpha-hydroxylase in Addison's disease and autoimmune polyendocrine syndrome type III, Eur. J. Endocrinol. 142 (2000) 187–194.
[27] A.G. Myhre, D.E. Undlien, K. Løvås, S. Uhlving, B.G. Nedrebø, K.J. Fougner, T. Trovik, J.I. Sørheim, E.S. Husebye, Autoimmune adrenocortical failure in Norway; autoantibodies and human leukocyte antigen class II associations related to clinical features, J. Clin. Endocrinol. Metab. 87 (2002) 618–623.
[28] M.M. Erichsen, K. Løvås, B. Skinningsrud, A.B. Wolff, D.E. Undlien, J. Svartberg, K.J. Fougner, T.J. Berg, J. Bollerslev, B. Mella, J.A. Carlson, H. Erlich, E.S. Husebye, Clinical, immunological, and genetic features of autoimmune primary adrenal insufficiency: observations from a Norwegian registry, J. Clin. Endocrinol. Metab. 94 (2009) 4882–4890.
[29] M. Fichna, P. Fichna, M. Gryczyńska, J. Walkowiak, M. Zurawek, J. Sowiński, Screening for associated autoimmune disorders in Polish patients with Addison's disease, Endocrine 37 (2010) 349–360.
[30] D.G. Altman, Practical statistics for medical research, Chapman, Hall, London, 1995, pp. 403–409.
[31] J.L. Fleiss, J. Cuzick, The reliability of dichotomous judgements: unequal numbers of judges per subject, Appl. Psychol. Meas. 3 (1979) 537–542.
[32] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174.
[33] H. Passing, W. Bablok, A new biometrical procedure for testing the equality of measurements from two different analytical methods, J. Clin. Chem. Clin. Biochem. 21 (1983) 709–720.
[34] R.H. Brown, J. Durbin, J.M. Evans, Tecniques for testing the constancy of regression relationship over time, JRSS ser B 37 (1975) 149–163.
[35] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet I (1986) 307–310.
[36] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability, Psychol. Bull. 86 (1979) 420–427.
[37] J.A. Hanley, B.J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases, Radiology 148 (1983) 839–843.
[38] C.F. Verge, D. Stenger, E. Bonifacio, P.G. Colman, C. Pilcher, P.J. Bingley, G.S. Eisenbarth, Combined use of autoantibodies (IA-2 autoantibody, GAD autoantibody, insulin autoantibody, cytoplasmic islet cell antibodies) in type 1 diabetes, Diabetes 47 (1998) 1857–1866.
[39] P.J. Bingley, E. Bonifacio, P.W. Mueller, Participating Laboratories, Diabetes Antibody Standardization Program: first assay proficiency evaluation, Diabetes 52 (2003) 1128–1136.
[40] C. Törn, P.W. Mueller, M. Schlosser, E. Bonifacio, P.J. Bingley, Participating Laboratories, Diabetes Antibody Standardization Program: evaluation of assays for autoantibodies to glutamic acid decarboxylase and islet antigen-2, Diabetologia 51 (2008) 846–852.
[41] M. Schlosser, P.W. Mueller, C. Törn, E. Bonifacio, P.J. Bingley, Participating Laboratories, Diabetes Antibody Standardization Program: evaluation of assays for insulin autoantibodies, Diabetologia 53 (2010) 2611–2620.