Serum cortisol

DOI:
10.1373/clinchem.2016.255034

Document Version
Final published version

Clinical Chemistry 62:9
1220–1229 (2016)

Serum Cortisol: An Up-To-Date Assessment of Routine Assay Performance

James M. Hawley,[1*] Laura J. Owen,[1,2] Stephen J. Lockhart,[1] Phillip J. Monaghan,[3] Annie Armston,[4] Carrie A. Chadwick,[5] Heather Wilshaw,[5] Maisa Freire,[6] Leslie Perry,[7] and Brian G. Keevil[1,2]

BACKGROUND: Accurate serum cortisol quantification is required for the correct diagnosis and management of adrenal pathologies. Presently, most laboratories use immunoassay to measure serum cortisol with proficiency schemes demonstrating a wide dispersion of results. Here, we investigate the effects of sex, matrix, and antibody specificity on serum cortisol quantification in 6 routine assays.

METHODS: Surplus serum was obtained before disposal and the following cohorts were created: males, nonpregnant females, pregnant patients, and patients prescribed either metyrapone or prednisolone. Samples were anonymized and distributed to collaborating laboratories for cortisol analysis by 6 routine assays. Cortisol was also measured in all samples using an LC-MS/MS candidate reference measurement procedure (cRMP); cortisol-binding globulin (CBG) was measured in the nonpregnant and pregnant female cohorts.

RESULTS: Considerable inter- and intraassay variation was observed across the male and nonpregnant female cohorts relative to the cRMP. Four immunoassays underrecovered cortisol in the pregnancy cohort, and CBG was found to be significantly higher in this cohort than in the nonpregnant females. In the metyrapone and prednisolone cohorts, all immunoassays overestimated cortisol. The first generation Roche E170 and Siemens Centaur XP were particularly prone to overestimation. In all cohorts the routine LC-MS/MS assay aligned extremely well with the cRMP.

CONCLUSIONS: Despite the clinical importance of serum cortisol, the performance of routine immunoassays remains highly variable. Accurate quantification is compromised by both matrix effects and antibody specificity. Underpinning this study with a cRMP has highlighted the deficiencies in standardization across routine cortisol immunoassays.

Serum cortisol measurement is recommended for the investigation of adrenal insufficiency *(1)* and Cushing syndrome *(2)*. Interpretation of these investigations typically relies on historic, single-value diagnostic cutoffs with minimal consideration of assay-specific biases *(3)*. In addition, cortisol quantification aids the assessment of adrenal reserve following chronic exogenous glucocorticoid therapy *(4, 5)*. In the management of Cushing syndrome, cortisol concentrations are used to help guide the appropriate dosing of metyrapone, a potent 11β-hydroxylase inhibitor *(6)*. Thus, assay-dependent biases and/or nonspecific interferences can produce inaccurate results that are difficult to interpret. Therefore, accurate serum cortisol measurement is essential for the correct diagnosis and management of adrenal pathologies.

External quality assessment (EQA)[8] schemes allow laboratories to assess their method performance relative to that of other users. Currently, the majority of laboratories use automated commercial immunoassays for cortisol quantification. Although variation for individual methods generally remains within narrow limits, the variation across all methods is considerably wider. This frequently produces a wide dispersion of results with a sex-specific bimodal distribution (see the Data File 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol62/issue9). Although sample heterogeneity and noncommutability may contribute to this dispersion, adherence to consensus guidelines allows EQA providers to minimize these effects *(7)*.

The frequency and consistency of this variability have prompted investigation into its underlying cause.

[1] University Hospital South Manchester, Manchester, UK; [2] Manchester Academic Health Science Centre, Manchester, UK; [3] The Christie Pathology Partnership, The Christie NHS Foundation Trust, Manchester, UK; [4] Southampton Hospital, Southampton, UK; [5] Aintree Hospital, Liverpool, UK; [6] St. Georges' University Hospital, London, UK; [7] Croydon University Hospital, London, UK.

* Address correspondence to this author at: Clinical Biochemistry Department, Clinical Sciences Building, University Hospital South Manchester, Southmoor Rd., Manchester M23 9LT, UK. Fax +44–161-291–2927; e-mail: james.hawley@uhsm.nhs.uk.

Received January 20, 2016; accepted May 17, 2016.
Previously published online at DOI: 10.1373/clinchem.2016.255034

[8] Nonstandard abbreviations: EQA, external quality assessment; RMP, reference measurement procedure; OCP, oral contraceptive pill; cRMP, candidate reference measurement procedure; UHSM, University Hospital South Manchester; CBG, cortisol-binding globulin; ANS, 8-anilino-1-napthalene sulfonic acid.

Comparisons of cortisol results to a GC-MS reference measurement procedure (RMP) have previously demonstrated that commercial immunoassays are prone to both sex-specific biases and matrix effects in select patient populations *(8)*. Furthermore, mean cortisol concentrations at baseline and 30 min post-synacthen were found to be assay specific and exhibit a consistently negative bias in women taking the oral contraceptive pill (OCP) *(9)*. Immunoassay cortisol measurements in patients taking metyrapone have also been reported to display a positive bias relative to a routine LC-MS/MS assay *(10, 11)*.

The performance and standardization of hormone assays are coming under increasing scrutiny. The Endocrine Society has now established minimum performance criteria for steroid assays, which recognize that metrological traceability is not currently available for all assays, yet remains an important goal *(12–14)*. The Partnership for Accurate Testing of Hormones has been established to address this issue in select assays *(15)*. However, despite its clinical importance, the poor performance of routine cortisol assays has yet to be considered.

The objectives of this study were to provide an up-to-date assessment of the performance of routine cortisol methods relative to an LC-MS/MS candidate RMP (cRMP) through investigating sex- and matrix-associated biases and investigating antibody specificity in patients prescribed prednisolone and metyrapone.

Methods

SAMPLE COLLECTION AND PREPARATION

Samples were collected into plastic 5-mL serum gel tubes (Becton Dickinson) at the University Hospital South Manchester (UHSM). Following routine analysis, samples were stored at 4 °C and retrieved within 24 h. Samples were subsequently anonymized, aliquoted into plastic 1.5 mL cryogenic tubes (Clinicon) and frozen at −20 °C before inhouse analysis or distribution to a collaborating laboratory. The following cohorts were created: (*a*) males, (*b*) nonpregnant females, (*c*) pregnant females, (*d*) patients prescribed metyrapone and (*e*) patients prescribed prednisolone.

PATIENT COHORTS

Control cohorts consisted of 51 males and 45 nonpregnant females. Before anonymization, each sample was checked against patient records. Patients prescribed exogenous glucocorticoids, hormone replacement therapy or any formulation of the OCP were excluded. The mean (range) age of the male cohort was 54 years (range, 35–81 years) and the nonpregnant female cohort was 60 years (27–85 years). Further subdivision of the nonpregnant female cohort yielded 17 presumed premenopausal patients (<40 years of age) and 21 presumed postmenopausal patients (>60 years of age). In addition, samples received for females 10–20 weeks pregnant (n = 72), routine prednisolone analysis (n = 42), and LC-MS/MS cortisol analysis on patients taking metyrapone (n = 27) were processed as detailed above. All information was obtained in accordance with UHSM Trust policies.

ASSAYS

Cortisol was quantified in all samples using an LC-MS/MS cRMP *(16)*. Briefly, all samples were analyzed in triplicate over 3 days with the mean result from each day's analysis averaged to provide the final result. Interassay %CVs for this method were 1.3%, 1.2%, and 1.1% at concentrations of 132, 520, and 772 nmol/L respectively. The 5 automated commercial cortisol immunoassays were: (*a*) Abbott Architect, (*b*) Beckman Access, (*c*) Roche E170 Generation I, (*d*) Roche E170 Generation II and (*e*) Siemens Centaur XP. All samples were also measured using a routine LC-MS/MS assay *(17)*. An overview of the methodologies and performance characteristics for each assay is provided in online Supplemental Data File 2. Interassay imprecision for all routine assays ranged from 1.6% to 7.5% over cortisol concentrations of 84 to 990 nmol/L; a complete summary of assay-specific imprecision is available in online Supplemental Data File 3.

Prednisolone was measured in the prednisolone cohort using a published LC-MS/MS method *(18)*. Samples with a prednisolone concentration <30 μg/L were discarded. Cortisol-binding globulin (CBG) was measured in the nonpregnant and pregnant female cohorts using a commercial ELISA in accordance with the manufacturer's instructions (Biovendor).

DATA ANALYSIS

Analyze-It version 2.30 was used for data analysis. For each assay, performance relative to the cRMP was assessed through Passing–Bablok regression analysis, Bland–Altman bias plots and Pearson correlation coefficients. Deviance of the cRMP results in each cohort from gaussian distribution was established by the Shapiro–Wilk test where a P value <0.05 was deemed significant. When the results for the cRMP and comparator assays followed a gaussian distribution, they were compared using a dependent samples t-test. When the cRMP and/or the comparator assay results significantly deviated from gaussian, the nonparametric Wilcoxon test was used. Samples with cortisol concentrations below an assay's limit of quantification were excluded from analysis for that respective cohort.

Results

The cohort-specific performance characteristics for the comparator assays relative to the cRMP are described

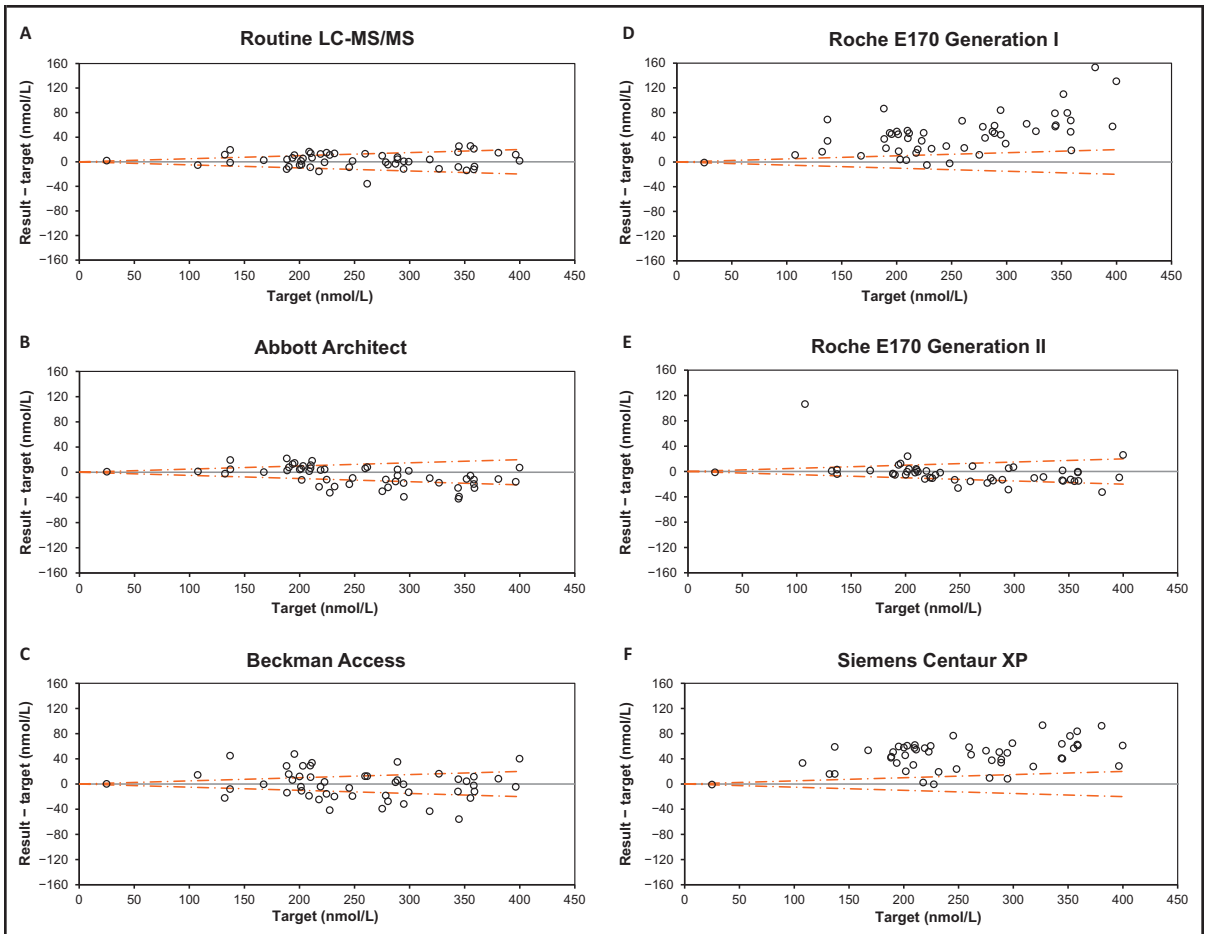

Fig. 1. Bland–Altman bias plots depicting the performance of the assays investigated relative to the cRMP in the male cohort (n = 51).
Dashed lines represent the 5% (95% CI) measurement uncertainty of the cRMP.

below. (To convert nmol/L to μg/dL, divide the cortisol concentration by 27.5.)

MALES

Analysis of this cohort by the cRMP generated normally distributed results ranging from 24.8 to 399.8 nmol/L. Routine assay performance relative to the cRMP is presented as Bland–Altman bias plots in Fig. 1 and summarized in online Supplemental Data File 4, Table 1.

Comparison of the mean bias across these assays produced 2 distinct groups. The routine LC-MS/MS, Abbott Architect, Beckman Access, and second generation Roche E170 all generated a mean bias between −1.5% and −1.2%. In each case, 95% CIs encompassing 0% supported the absence of bias. In contrast, the first generation Roche E170 and Siemens Centaur XP displayed a positive bias relative to the target concentrations. With an overall mean bias of 17.0%, yet a proportional bias of 24.0%, it is apparent from Fig. 1D that bias increases with concentration for the first generation Roche E170 assay. This relationship likely contributes to the negative constant bias of −19.45 nmol/L. Conversely, in the case of the Siemens Centaur XP, the overall mean bias of 18.5% is the cumulative effect of a 13.0% proportional and a 15.59-nmol/L constant bias. Accordingly, results produced by these assays were significantly different from those of the cRMP ($P < 0.0001$).

NONPREGNANT FEMALES

Analysis of this cohort by the cRMP yielded normally distributed results ranging from 56.5 to 602.7 nmol/L. Routine assay performance relative to the cRMP is presented as Bland–Altman bias plots in Fig. 2 and summarized in online Supplemental Data File 4, Table 2.

Again, comparison of the mean bias produced 2 separate groups. The routine LC-MS/MS, Abbott Architect, Beckman Access, and second generation Roche E170 all generated results with a mean bias ranging from −4.9%

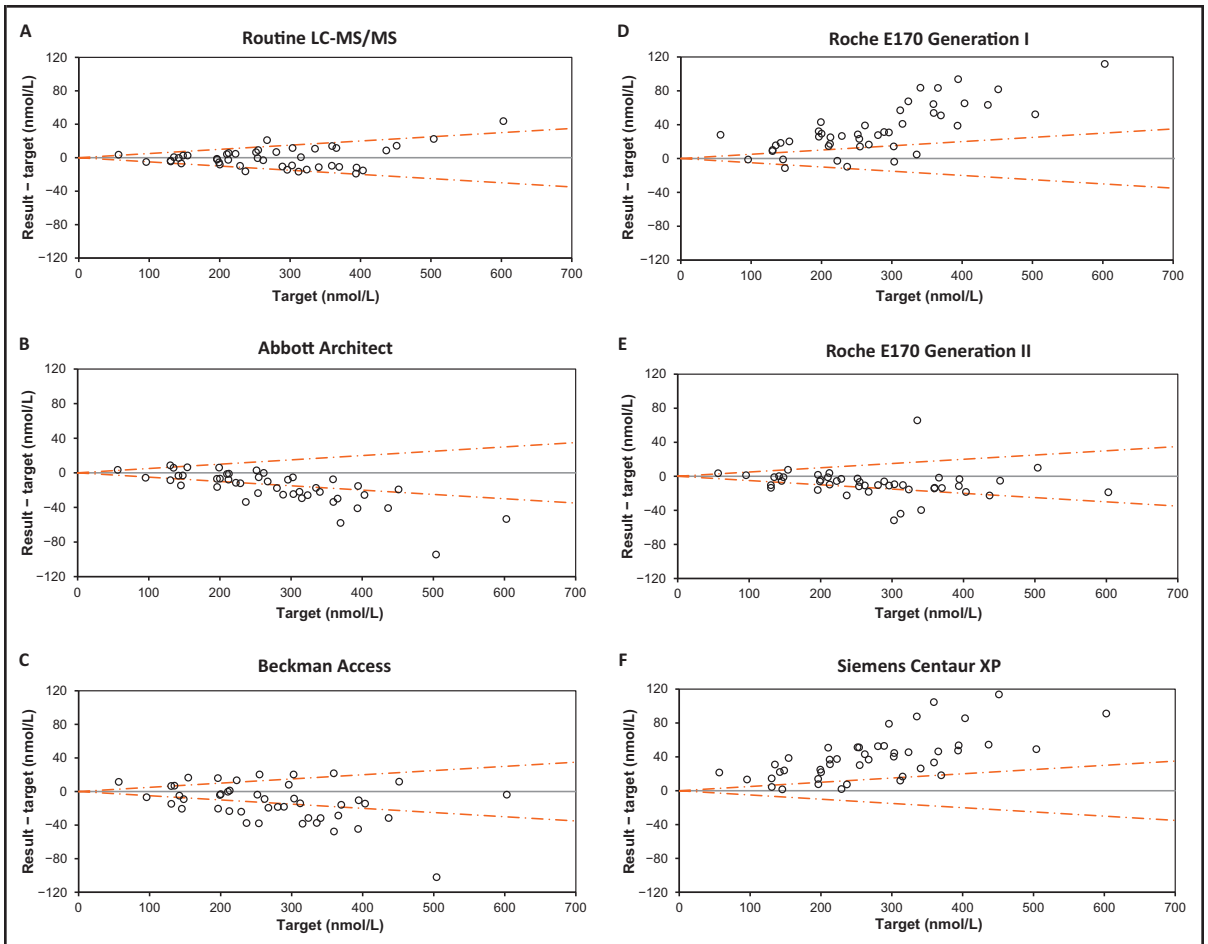

Fig. 2. Bland-Altman bias plots showing the performance of the assays investigated relative to the cRMP in the nonpregnant female cohort (n = 45).

Dashed lines represent the 5% (95% CI) measurement uncertainty of the cRMP.

to −0.4%. However, of these, only the routine LC-MS/MS assay possessed 95% CIs encompassing 0%. The 95% CIs of the other assays all fell below 0% indicating a slight negative bias. This likely contributed to the significant differences observed between results of these assays and the cRMP ($P < 0.01$). Conversely, the routine LC-MS/MS assay results were not found to differ significantly from those of the cRMP ($P = 0.9284$). The first generation Roche E170 and Siemens Centaur XP displayed respective mean biases of 11.8 and 14.6%. Like the male cohort, both assays generated 95% CIs that were above 0% identifying a positive bias. This is further supported as each assay produced results significantly different from those of the cRMP ($P < 0.01$).

PREGNANCY

As determined by the cRMP, this cohort generated normally distributed results ranging from 129.0 to 1067.2 nmol/L. Routine assay performance relative to the cRMP is displayed as Bland–Altman bias plots in Fig. 3 and summarized in online Supplemental Data File 4, Table 3.

Of the assays investigated, the routine LC-MS/MS results were found to be closest to those of the cRMP ($P = 0.0653$). Although all immunoassay results were statistically significantly different from the target concentrations ($P < 0.0001$), the performance characteristics of the second generation Roche compared well to the cRMP. Despite a slight negative mean bias of −2.6%, Passing–Bablok regression analysis produced a slope and intercept with 95% CIs encompassing the cRMP target parameters. Conversely, with respective mean biases of −5.5%, −23.3%, and −25.0% equating to −36.7, −143.6, and −156.6 nmol/L, the Siemens Centaur XP, Beckman Access, and Abbott Architect displayed markedly negative bias relative to the cRMP. The performance of these assays in the pregnant female cohort was dis-

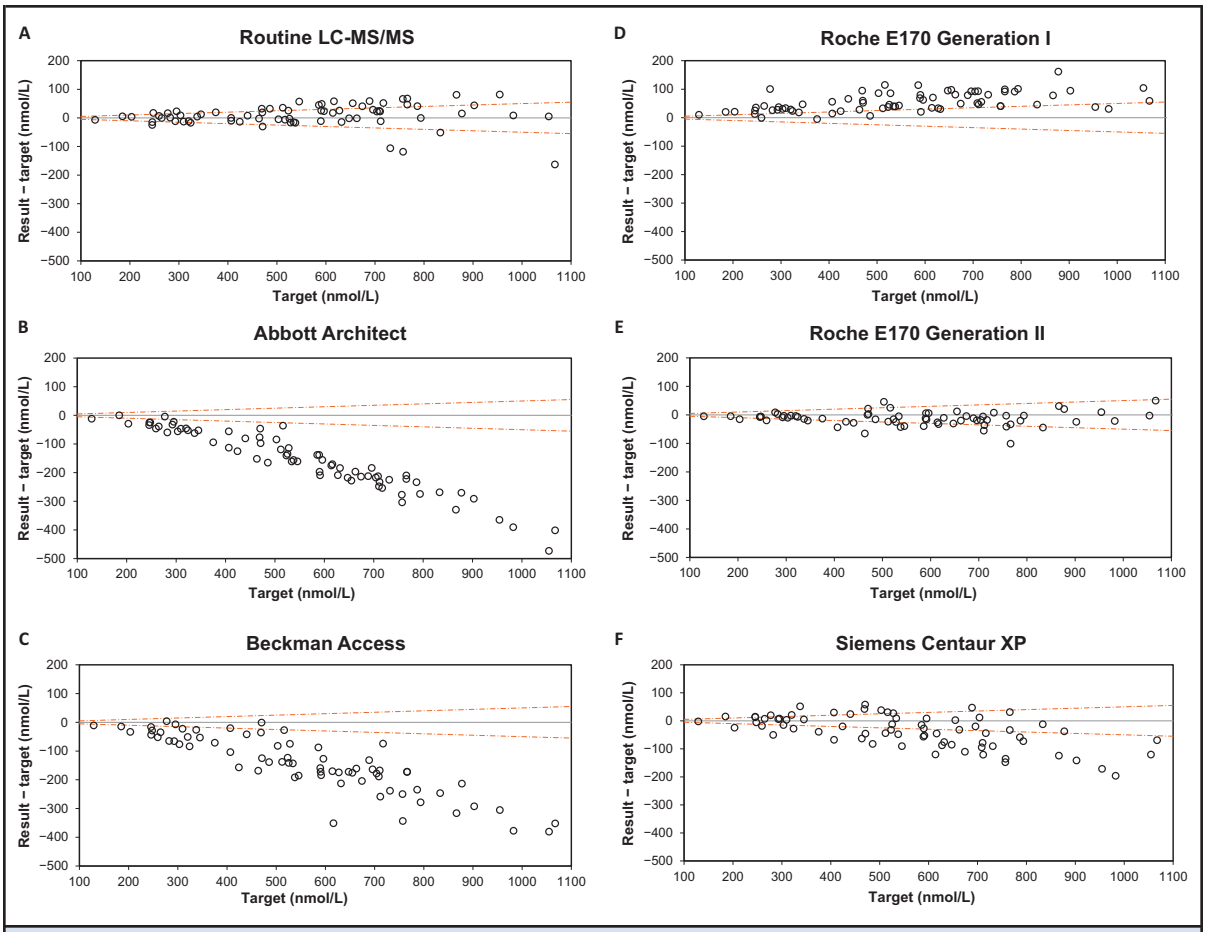

Fig. 3. Bland–Altman bias plots displaying the performance of the assays investigated relative to the cRMP in the pregnant cohort (n = 72).

Dashed lines represent the 5% (95% CI) measurement uncertainty of the cRMP.

tinctly different from that observed in the nonpregnant female cohort, as is visually apparent when Fig. 2, B, C, and F, are compared to their corresponding panels in Fig. 3. With a mean bias of 10.2%, the first generation Roche E170 performed similarly both in the pregnant female and nonpregnant female cohorts. Despite this, the Passing–Bablok characteristics were markedly different since the 95% CIs for the 2 slopes shared no overlap, and the intercept shifted from negative in the nonpregnant cohort to positive in the pregnant cohort.

METYRAPONE

Analysis of this cohort by the cRMP yielded nongaussian results ranging from 91.3 to 535.3 nmol/L. Routine assay performance relative to the cRMP is presented as Bland–Altman bias plots in Fig. 4 and summarized in online Supplemental Data File 4, Table 4.

Of the immunoassays, mean bias ranged from 16.4% to 95.0% with Passing–Bablok regression parameters ranging from 0.72 to 1.37 (slope) and 34.02 to 127.82 nmol/L (intercept). Collectively, these performance characteristics resulted in an overall positive bias in each immunoassay that is clearly evident in Fig. 4, B–F. Wilcoxon analysis confirmed the immunoassay results were highly statistically significantly different from those of the cRMP ($P < 0.0001$). In contrast, the differences between the routine LC-MS/MS and cRMP results were not as marked ($P = 0.0152$).

Within the cohort, 3 samples were evidently outliers as, with target concentrations >350 nmol/L, they were separated from the majority of results (Fig. 4). In the treatment of Cushing syndrome, target cortisol concentrations for patients prescribed metyrapone should fall between 150 and 300 nmol/L *(6)*. Results higher than this could therefore represent either patients that have only recently commenced therapy, poor compliance or severe disease such as adrenocortical carcinoma.

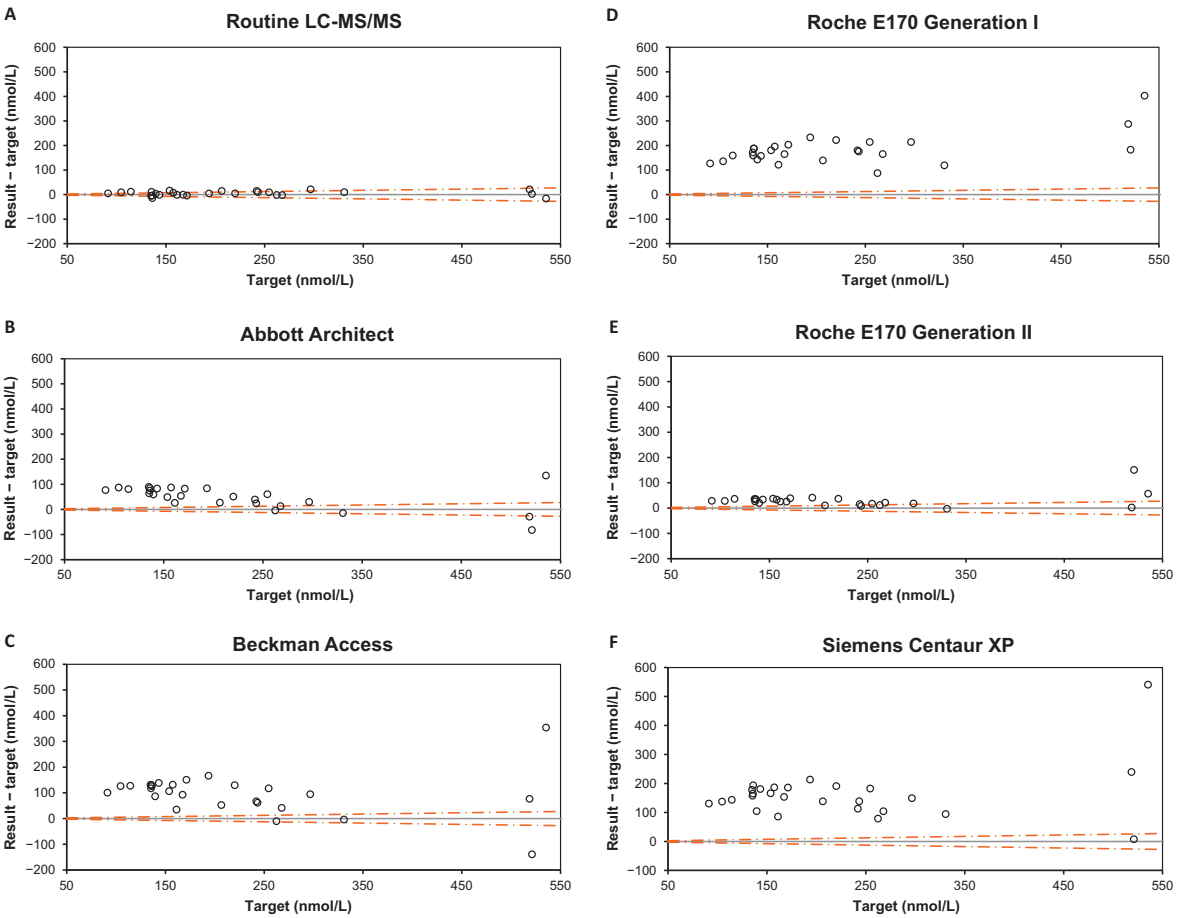

Fig. 4. Bland–Altman bias plots displaying the performance of the assays investigated relative to the cRMP in the metyrapone cohort (n = 27).

Dashed lines represent the 5% (95% CI) measurement uncertainty of the cRMP.

PREDNISOLONE

LC-MS/MS quantification confirmed the presence of prednisolone in all samples. The mean (range) prednisolone concentration was 245 μg/L (range 41 to 590 μg/L). Cortisol analysis by the cRMP generated nongaussian results ranging from 2.6 to 99.3 nmol/L. Routine assay performance relative to the cRMP is presented as Bland–Altman bias plots in Fig. 5 and summarized in online Supplemental Data File 4, Table 5.

Whereas the sample distribution was found to be nongaussian for the cRMP and several routine assays, the results of the Abbott Architect and first generation Roche E170 were not found to significantly differ from a normal distribution. In all cases, the nonparametric Wilcoxon test was used for statistical analysis. Of the assays investigated, all produced results that were statistically significantly different from those of the cRMP ($P < 0.0001$).

Despite this, the routine LC-MS/MS assay exhibited superior performance compared to the immunoassays. Although the mean bias for LC-MS/MS was 15.5% this equated to only 3.1 nmol/L and all Passing–Bablok regression parameters produced 95% CIs that included their respective targets. Furthermore, only 24/42 samples were quantifiable by the LC-MS/MS assay; the remaining samples were correctly identified as having a cortisol concentration below the limit of quantification of the assay.

In contrast, the immunoassays exhibited mean biases ranging from 690.9% to 8169.4%, corresponding to 56.0 to 658.6 nmol/L. In addition, Passing–Bablok slope characteristics were positive relative to the line of identity for each assay except the Siemens Centaur XP, which failed the CUSUM linearity test ($P < 0.01$) thereby, invalidating regression analysis. Although linearity was detected between the first generation Roche E170 and

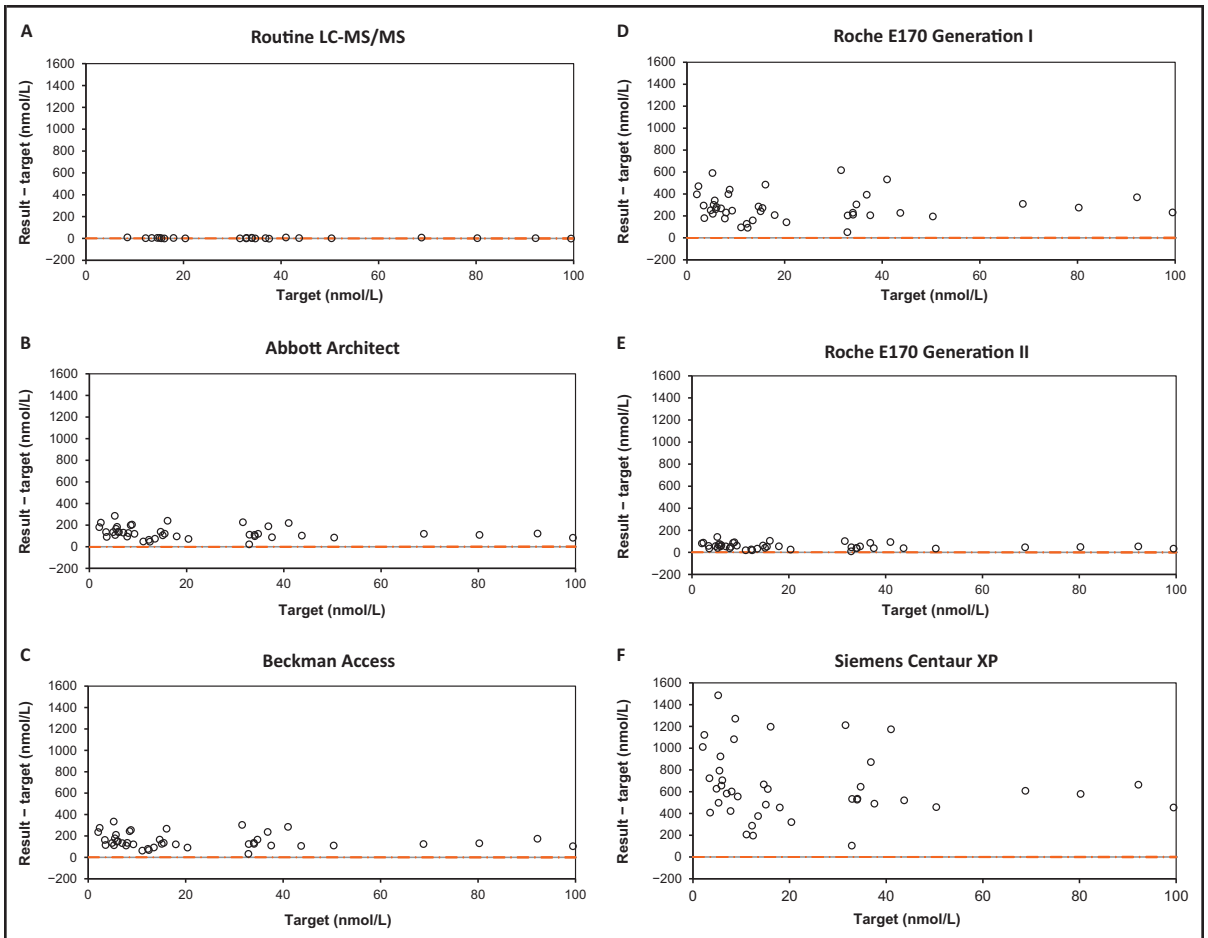

Fig. 5. Bland-Altman bias plots displaying the performance of the assays investigated relative to the cRMP in the prednisolone cohort (n = 42).

Dashed lines represent the 5% (95% CI) measurement uncertainty of the cRMP.

cRMP, the software was unable to provide 95% CIs for the slope and intercept.

CORTISOL-BINDING GLOBULIN

The mean (95% CIs) CBG concentrations for the nonpregnant female and pregnant female cohorts were 27.23 mg/L (26.02–28.45) and 52.02 mg/L (47.78–56.26), respectively. These data are plotted against cortisol concentrations for each cohort in Fig. 6. An independent *t*-test determined these CBG distributions to be statistically different ($P < 0.0001$). Subdivision of the nonpregnant female cohort into their presumed pre- and postmenopausal cohorts provided mean (95% CIs) CBG concentrations of 29.37 (27.65–31.09) and 25.92 mg/L (23.64–27.08), respectively. An independent *t*-test demonstrated weak but statistically significant differences between these subgroups ($P = 0.0029$).

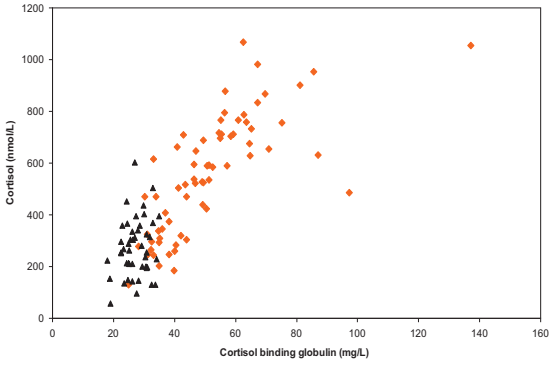

Fig. 6. The distribution of cortisol concentrations (nmol/L) relative to cortisol binding globulin concentrations (mg/L) in nonpregnant healthy females (▲) and pregnant females (◆).

Discussion

Underpinning this study with a cRMP has ensured that each method was compared to a higher-order measurement procedure. This certifies metrological traceability to SI units thereby, providing a platform for meaningful statistical analysis.

In contrast to EQA schemes and previous studies *(8)*, this study benefits from using single donor samples for comparison. Although pooling samples can provide sufficient volume for large-scale studies, it can also alter the sample matrix and potentially bias results. Our approach instead provides a true representation of the performance of each assay for individual samples. In addition, to the best of our knowledge, this is the first study to use a cRMP to comprehensively assess the performance of routine assays for the measurement of cortisol in serum obtained from patients taking either metyrapone or prednisolone. In all cases, variation due to noncommutability and heterogeneity has been minimized through consistent sample preparation and storage before analysis. Hence, assay-specific causes are likely to be the largest contributing factors to any variation observed. Although limited by an absence of particularly high cortisol concentrations in the male and nonpregnant female cohorts, this study is representative of these populations, with high concentrations difficult to obtain without sample adulteration.

Our investigation into sex-specific bias has found that considerable interassay variation exists across the routine commercial immunoassays relative to the cRMP. In addition, intraassay performance is variable across the male and nonpregnant female cohorts. Although the Beckman Access and second generation Roche E170 exhibited good agreement with the cRMP in the male cohort, all immunoassays demonstrated significant differences with the cRMP in the nonpregnant female cohort. In both cohorts, the Abbott Architect, first generation Roche E170 and Siemens Centaur XP displayed significant differences with the cRMP. Whereas the Abbott Architect consistently underestimated cortisol, the first generation Roche E170 and Siemens Centaur XP consistently overestimated it.

We also observed that pregnancy had a striking effect across the Abbott Architect, Beckman Access, first generation Roche and Siemens Centaur XP assays. Considerable underestimation was demonstrated in these assays when compared both directly to the cRMP and relatively to their respective performance characteristics in the nonpregnant female cohort. This underestimation was exacerbated with increasing cortisol concentrations, a relationship that was seemingly related to increased CBG concentrations.

Previous work has speculated that cortisol underrecovery may be attributed to increased CBG concentrations secondary to pregnancy *(8)* or OCP medication *(9)*. These conditions are associated with a high estrogenic state that serves to stimulate CBG synthesis. We have confirmed that CBG concentrations in the pregnant cohort are significantly greater than that of the nonpregnant female cohort. This supports the hypothesis that the underrecovery observed in patients with increased CBG concentrations is secondary to the inability of the assay to completely liberate cortisol from its binding proteins. Furthermore, we have observed that the performance characteristics for each immunoassay differ across male and nonpregnant female cohorts. Subdivision of the nonpregnant female cohort into presumed pre- and postmenopausal groups demonstrated that CBG concentrations were modestly different between these 2 groups. This infers that CBG concentrations decline postmenopause, possibly secondary to the attenuated estrogenic stimulus. We speculate that the effect of CBG may be pronounced when there are only small differences in concentration. This may explain the observed difference in performance between the male and nonpregnant female cohorts and the bimodal distribution observed across female samples in EQA schemes. Further work to substantiate the relationship between CBG concentration and cortisol quantification in larger sample sizes is indicated. It is possible that each assay has a specific CBG concentration above which cortisol quantification becomes increasingly challenging.

Accurate cortisol measurement relies on the complete recovery of cortisol from its protein-bound fractions. In general, information on how this is achieved by commercial immunoassays remains proprietary. Hence, without knowledge of the specific approach, it is difficult to speculate on individual assays. Nevertheless, it has been reported that manufacturers may displace cortisol from CBG by adjusting temperature, lowering pH or alternatively using agents such as salicylate or 8-anilino-1-napthalene sulfonic acid (ANS) *(19)*. However, the efficiency of ANS is influenced by the concentration of binding proteins. Hence, the concentration of ANS required for healthy male and nonpregnant female samples may be inadequate for samples with increased CBG *(19)*. Therefore, it is conceivable that immunoassays may rely on agents such as ANS that, when present in insufficient quantities, fail to completely free cortisol from CBG. As CBG is significantly higher in estrogenic states, further work is required to assess the effects of other medications and conditions associated with increased estrogen concentrations.

Whereas underrecovery is attributed to matrix effects such as increased CBG concentrations, we hypothesize that the over-recovery in the first generation Roche E170 and Siemens Centaur XP is secondary to nonspecific cross-reactivity. We have investigated selectivity

through experiments with serum obtained from patients prescribed metyrapone and prednisolone.

Metyrapone reversibly inhibits 11-β hydroxylase, preventing oxidation of 11-deoxycortisol to cortisol. This attenuates negative feedback on the hypothalamic–pituitary–adrenal axis, resulting in the accumulation of cortisol precursors *(6)*. These precursors share structural homology with cortisol, necessitating that immunoassays require highly specific antibodies to ensure selectivity. Our results indicate all 5 immunoassays over recover cortisol relative to the cRMP in the metyrapone cohort. Of these, the first generation Roche E170 and Siemens Centaur XP were particularly prone to over-estimation, likely secondary to antibody cross–reactivity with cortisol precursors *(10, 11)*. Our observations support recent Endocrine Society guidelines that advise serum cortisol is measured by LC-MS/MS in patients prescribed metyrapone *(6)*.

Prednisolone and prednisone are the most commonly prescribed exogenous glucocorticoids *(20)*. For patients not on lifelong treatment, their adrenal reserve should be ascertained before altering dosing regimens *(4)*. This provides a requirement for assays to differentiate cortisol from exogenous glucocorticoids and their metabolites, to guarantee the integrity of the short synacthen test.

Our results showed that all 5 immunoassays overestimate cortisol in the prednisolone cohort. This is likely because reagent antibodies fail to differentiate cortisol from prednisolone and its metabolites. While our data suggest that the first generation Roche E170 and Siemens Centaur XP are markedly affected by cross-reactivity, the Abbott Architect, Beckman Access and second generation Roche E170 are also subject to considerable overrecovery.

Importantly, our observations have potential implications for patient care. Cortisol under or overrecovery can affect the interpretation of both the short synacthen test and the overnight dexamethasone suppression test such that patients may not be provided with the correct diagnosis or appropriate follow–up. Similarly, for patients prescribed metyrapone, immunoassay quantification may lead to an inappropriate dosing regimen and may ultimately mask adrenal insufficiency.

This study was conducted during the release of the second generation Roche E170 assay. Given our results, it is plausible that this assay benefits from a more specific antibody than its predecessor. Certainly, the overrecovery exhibited in the male, nonpregnant and pregnant female cohorts has improved, with better alignment to the cRMP. Furthermore, the overrecovery in the metyrapone and prednisolone cohorts, although still undesirable, is now considerably reduced. The use of a more specific antibody is further substantiated by the improved percentage cross-reactivities relative to the first generation assay (see online Supplemental Data File 2).

In all cohorts, the routine LC-MS/MS assay and cRMP aligned well. When the 5% measurement uncertainty of the cRMP is applied, nearly all routine LC-MS/MS results fell within this and could therefore be considered statistically unbiased.

Not all laboratories possess the resources and expertise for LC-MS/MS measurements. Hence, there remains a requirement for routine assays to provide metrologically traceable results. From the evidence presented here, it is apparent that this requirement is not currently satisfied. Indeed, many of the assays investigated failed to achieve a bias $<10.3\%$, the current desirable limit as specified by Westgard *(21)*. Although calibrators may be traceable to certified reference materials, this does not necessarily translate to patient samples. Consequently, results are not comparable across assay, location or time.

In summary, routine serum cortisol immunoassay performance remains highly variable. Although there is increasing awareness of the impact of assay-specific biases on result interpretation, it is unrealistic to expect that all users are familiar with the limitations of their assay. The case for sex- and assay-specific cutoffs has been made *(9)*. Although these may be helpful in the short-term, reagent lot changes, recalibration and assay reformulation may increase result uncertainty and thus not represent a long-term solution.

The recent drive towards standardization is welcome *(15)* and we reason that given the limitations evidenced here, serum cortisol assays should be considered for inclusion in this initiative.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: B.G. Keevil, Abbott Diagnostics.
Expert Testimony: None declared.
Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, and final approval of manuscript.

Acknowledgments: J.M. Hawley, L.J. Owen, and B.G. Keevil thank Abbott for the supply of reagents for the analysis of cortisol at UHSM.

References

1. Bornstein SR, Allolio B, Arlt W, Barthel A, Don-Wauchope A, Hammer GD, Husebye ES, et al. Diagnosis and treatment of primary adrenal insufficiency: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2016;101:364–89.
2. Nieman LK, Biller BM, Findling JW, Newell-Price J, Savage MO, Stewart PM, Montori VM. The diagnosis of Cushing's syndrome: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2008;93:1526–40.
3. Chatha KK, Middle JG, Kilpatrick ES. National UK audit of the Short Synacthen Test. Ann Clin Biochem 2010; 47:158–64.
4. Bancos I, Hahner S, Tomlinson J, Arlt W. Diagnosis and management of adrenal insufficiency. Lancet Diabetes Endocrinol 2015;3:216–26.
5. Dinsen S, Baslund B, Klose M, Rasmussen AK, Friis-Hansen L, Hilsted L, Feldt-Rasmussen U. Why glucocorticoid withdrawal may sometimes be as dangerous as the treatment itself. Eur J Intern Med 2013;24:714–20.
6. Nieman LK, Biller BMK, Findling JW, Murad MH, Newell-Price J, Savage MO, Tabarin A. Treatment of Cushing's syndrome: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2015;100: 2807–31.
7. CLSI. Characterization and qualification of commutable reference materials for laboratory medicine; approved guideline. Wayne (PA): CLSI; 2010. CLSI document C53-A.
8. Dodd AJ, Ducroq DH, Neale SM, Wise MP, Mitchem KL, Armston A, et al. The effect of serum matrix and gender on cortisol measurement by commonly used immunoassays. Ann Clin Biochem 2014;51:379–85.
9. 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;78:673–80.
10. Monaghan PJ, Owen LJ, Trainer PJ, Brabant G, Keevil BG, Darby D. Comparison of serum cortisol measurement by immunoassay and liquid chromatography-tandem mass spectrometry in patients receiving the 11 beta-hydroxylase inhibitor metyrapone. Ann Clin Biochem 2011;48:441–6.
11. Owen LJ, Halsall DJ, Keevil BG. Cortisol measurement in patients receiving metyrapone therapy. Ann Clin Biochem 2010;47:573–5.
12. Handelsman DJ, Wartofsky L. Requirement for mass spectrometry sex steroid assays in the journal of clinical endocrinology and metabolism. J Clin Endocrinol Metab 2013;98:3971–3.
13. Wierman ME, Auchus RJ, Haisenleder DJ, Hall JE, Handelsman D, Hankinson S, et al. The new instructions to authors for the reporting of steroid hormone measurements [Editorial]. Mol Endocrinol 2014;28:1917.
14. Monaghan PJ, Keevil BG, Stewart PM, Trainer PJ. Case for the wider adoption of mass spectrometry-based adrenal steroid testing, and beyond. J Clin Endocrinol Metab 2014;99:4434–7.
15. PATH. Partnership for the accurate testing of hormones. http://www.hormoneassays.org (Accessed January 2016).
16. Hawley JM, Owen LJ, MacKenzie F, Mussell C, Cowen S, Keevil BG. Candidate reference measurement procedure for the quantification of total serum cortisol with LC-MS/MS. Clin Chem 2016;62:262–9.
17. Owen LJ, Adaway JE, Davies S, Neale S, El-Farhan N, Ducroq D, et al. Development of a rapid assay for the analysis of serum cortisol and its implementation into a routine service laboratory. Ann Clin Biochem 2013;50: 345–52.
18. Owen LJ, Gillingwater S, Keevil BG. Prednisolone measurement in human serum using liquid chromatography tandem mass spectrometry. Ann Clin Biochem 2005;42:105–11.
19. Bertholf RL, Jialal I, Winter WE. The adrenal cortex. In: Burtis CA, Ashwood ER, Bruns DE, editors. Tietz textbook of clinical chemistry and molecular diagnostics, 5th ed. St. Louis: Elsevier; 2012. p 1847–904.
20. Forss M, Batcheller G, Skrtic S, Johannsson G. Current practice of glucocorticoid replacement therapy and patient-perceived health outcomes in adrenal insufficiency – a worldwide patient survey. BMC Endocr Disord 2012;12:8.
21. Westgard QC. Desirable biological variation database specifications. https:// www.westgard.com/biodatabase1.htm (Accessed January 2016).