



Published in final edited form as:

Steroids. 2014 April ; 82: 7–13. doi:10.1016/j.steroids.2013.12.005.

High variability in serum estradiol measurements in men and women

Hubert W. Vesper^{a,*}, Julianne C. Botelho^a, Meghan L. Vidal^a, Yasamin Rahmani^a, Linda M. Thienpont^b, and Samuel P. Caudill^a

^aCenters for Disease Control and Prevention, 4770 Buford Highway, NE F25, Atlanta 30341-3724, Georgia

^bUniversity of Ghent, Harelbekestraat 72, Ghent B-9000, Belgium

Abstract

To reduce the variability in estradiol (E2) testing and to assure better patient care, standardization of E2 measurements has been recommended. This study aims to assess the accuracy and variability of E2 measurements performed by 11 routine immunological methods and 6 mass spectrometry methods using single donor serum materials and to compare the results to a reference method. The contribution of calibration bias, specificity or matrix effects, and imprecision on the overall variability of individual assays was evaluated.

This study showed substantial variability in serum E2 measurements in samples from men and pre- and post-menopausal women. The mean bias across all samples, for each participant, ranged between -2.4% and 235%, with 3 participants having a mean bias of over 100%. The data suggest that calibration bias is the major contributor to the overall variability for nine assays.

The analytical performances of most assays measuring E2 concentrations do not meet current needs in research and patient care. Three out of 17 assays would meet performance criteria derived from biological variability of $\pm 12.5\%$ bias at concentrations ≥ 20 pg/mL, and a maximum allowable bias of ± 2.5 pg/mL at concentrations < 20 pg/mL. The sensitivity differs highly between assays. Most assays are not able to measure E2 levels below 10 pg/mL. Standardization, specifically calibration to a common standard by using panels of individual patient samples, can reduce the observed variability and improve the utility of E2 levels in clinical settings.

Keywords

Estradiol; Assay comparison; Analytical variability

1. Introduction

Measurement of estradiol (E2) in serum provides important information for patient care, research and public health. In patient care, it is currently used for the differential diagnosis

*Corresponding author. Tel.: +995 (770) 488 4191; fax: +995 (404) 683 2467.

Appendix A. Supplementary data: Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.steroids.2013.12.005>.

of amenorrhea, for the assessment of fertility and to monitor follicle stimulation therapy. Further, E2 measurements provide information about response of patients to treatments such as those using aromatase inhibitors or gonadotropin-releasing hormone agonists to treat certain cancers, and those using steroid hormones as part of hormone replacement therapy.

Research studies suggest E2 levels are promising biomarkers for assessing breast cancer risk in postmenopausal women. Epidemiological studies found that the relative risk for breast cancer in postmenopausal women with high E2 levels ranges from 1.5 to 3.0 [1–5]. Other research reported associations of serum E2 levels with chronic diseases such as cardiovascular disease [6,7], cognitive function [8], and fracture risk [9]. The use of estradiol as a drug, as well as the association of the effects of estradiol levels in blood on human health, is subject to extensive research [10].

Many of these current and emerging applications require accurate and reliable measurements of E2 at low concentrations typically observed in postmenopausal women, men and children. E2 in serum is commonly measured by enzyme-based immunoassays without prior isolation from serum ('direct assays') [11]. More recently, analytical methods using liquid chromatography coupled with tandem mass spectrometry (HPLC/MS/MS) are increasingly used for measuring E2 in patient care and research [12].

While estradiol measurements are widely used, high inaccuracy and variability in estradiol measurements, especially at low serum E2 concentrations has been reported [19,20,21,22]. As one consequence, formulation of common clinical decision levels (i.e., identifying women at increased risk for breast cancer or men for osteoporotic fractures) cannot be formulated. The Endocrine Society in its clinical practice guideline on osteoporosis in men points out that low estradiol levels are associated with increased fracture risk and accelerated bone loss in older men. However, it does not recommend measurement of estradiol because of the lack of easily available, accurate assays [13]. The importance of estradiol testing in research and patient care, as well as the need for reliable and accurate estradiol measurements, are emphasized in a recent Endocrine Society Position Statement [14].

To reduce the variability in E2 testing and to assure better patient care, standardization of E2 measurements was recommended [14,19,20]. The CDC is addressing this need in its Hormone Standardization Program. This work is performed in collaboration with the Partnership for Accurate Testing of Hormones (PATH). As a first step towards standardization, this study aims to assess the accuracy and variability of estradiol measurements performed by routine immunological methods and mass spectrometry methods using single donor serum materials and an established reference method for comparisons.

2. Materials and methods

A panel of commercially prepared and deidentified single donor sera from 40 healthy donors (15 men, age: 26–77 years, E2 concentration range: 11.9–30.8 pg/mL, and 25 women age: 21–76 years, E2 concentration range: 2.5–285 pg/mL) was obtained from Solomon Research Park (Seattle, WA). The collection protocol for approved by the local IRB board and

documentation was reviewed and approved by the CDC Human Subjects coordinator. The sera were processed according to the Clinical and Laboratory Standards Institute (CLSI) protocol C37A [15] to assure a similar quality as regular patient samples.

Target values were assigned using an isotope dilution gas chromatography/mass spectrometry reference measurement procedure (RMP) and were verified with an HPLC/MS/MS candidate reference measurement procedure [16,17]. The expanded uncertainty of the RMP was 2.6% and 5.3% for concentrations ≥ 5 and <5 pg/mL, respectively. Seventeen participants (11 immunoassays and 6 mass spectrometry-based assays) contributed to this study. Of the eleven immunoassays, nine were performed by the assay manufacturers and two by clinical laboratories (Table 1). Participants were asked to analyze each sample in duplicate on three days. A run order was provided which required participants to run samples in a forward then reverse order. Each participant was instructed to follow their own quality assurance procedures.

For each sample, the overall variability of individual results reported by all participants was expressed by the coefficient of variation (CV), by the ratio between the highest and lowest result, and by using Box-and-Whisker plots. After removing outliers following the CLSI protocol EP9A2 [18], regression parameters were estimated from measurement results obtained by the assay and RMP using a weight of the inverse of variance of assay results ($1/\text{var}_{\text{Assay Result}}$). Four assays (Assays 2, 7, 8, 11) required a quadratic fit while for all other assays a linear fit model was used. Bias plot analysis was performed using Analyze-it software (Analyze-it Software, Ltd., Leeds, UK, version 2.26). Measurement bias against the reference target values were assessed using bias plots and box-whisker plots.

Assay imprecision, expressed as the percent coefficient of variation (CV), was calculated for each sample using both the individual measurements reported on all days (overall imprecision, CV_A) and the mean from daily measurements (between-day variability).

We developed a model to estimate the contributions of calibration bias, imprecision and other factors such as specificity and specimen matrix effects (further described as 'sample effects') on the overall variability of an assay. We first used weighted regression analysis (weight = $1/\text{var}_{\text{Assay Result}}$) to model the relationship between the results reported for each sample and the reference values assigned by the RMP. We then calculated the weighted sums of squares of the various sources of error: calibration bias, specimen effect, and imprecision. We defined the calibration bias as the difference between predicted values and their corresponding target values. The specimen effect was defined as the difference between the sample means and their predicted values obtained from the regression equation. The imprecision was defined as the difference between the individual values and their means. The overall variability comprises the sum of these 3 weighted sums of squares (for formulas and further descriptions see the Supplemental file).

3. Results

When comparing the reference values of the individual donor samples with the reportable range of the individual assays, only four assays had a reportable range (approximately 2–

5000 pg/mL; Table 1) that covered the target concentration (approximately 2.5–285 pg/mL; Supplemental Table 1) of all the samples used in this study. Although the reference values for a number of samples were outside the reportable range of some assays, values inside the reportable range were reported for these samples overestimating the actual E2 concentrations in these samples (Table 2). All assays reported results on samples with target values greater than 30 pg/mL.

The variability in individual results reported on a single sample by all assays, expressed as the ratio between the highest and lowest results, was on average 6 (range: 4–13, Fig. 1). At concentrations greater than 30 pg/mL, the highest and lowest values differed on average by a factor of 5. The CV for individual reported results on each sample, across laboratories was on average 50% (range: 36–83%). The overall variability among assays did not change notably when using only the first replicate from each day ($n = 51$) or only the first measurement reported by each participant ($n = 17$). When removing three immunoassays (8,9,10) that had a mean measurement bias of over 100%, results differed on average by a factor of 3 (range: 2–6), and the average overall CV was 24% (range: 11–58%). The median interquartile range of individual reported results among immunoassays was about 5 times higher than among mass spectrometry assays. When excluding the three immunoassays with extremely high mean biases, the interquartile range was for most samples twice as high for immunoassays than for mass spectrometry-based assays (median interquartile ranges: 3.6 pg/mL for mass spectrometry-based assays and 7.25 pg/mL for immunoassays).

Comparing the reported results against the reference values using weighted regression analysis, four methods showed a non-linear relationship to the RMP. The linear slopes ranged between 0.61 and 2.23 (Table 2, Supplemental Fig. 2). The mean bias across all samples for each participant ranged between –2.4% and 235% with 3 participants having a mean bias of over 100% (Fig. 2). The mean bias of the mass spectrometry assays ranged between –2.4% and 22.5% and for immunoassays from 2.6% to 235% (2.6–37.5%, excluding the 3 assays with more than 100% mean bias). All assays showed an increase in bias with decreasing E2 concentration. This increase is more pronounced with immunoassays than mass spectrometry assays (Fig. 3).

The median imprecision was less than 10% CV for most assays. The imprecision on individual samples, however, was as high as 75% CV (Table 2). Moreover, the imprecision does slightly increase with decreasing E2 concentration and is similar for immunoassays and mass spectrometry assays (Supplementary Fig. 1).

The overall assay variability, calculated as the sum of the weighted sums of squares of the various sources of error (i.e., calibration bias, specimen effect, and imprecision) relative to the ideal line obtained using weighted regression analysis, was high (over 100) for 7 assays (Table 3). It was lowest for most mass spectrometry assays. The major contributor to the overall variability was calibration bias for nine assays; sample effects for five assays; and imprecision for three assays. For two mass spectrometry-based assays and three immunoassays, specimen effects are the major contributors to the overall variability.

4. Discussion

This study aimed at assessing the variability of E2 measurements performed in research and patient care. The between-assay variability, expressed as the ratio between the highest and lowest reported result and as the bias distribution for individual samples, is substantial especially at concentrations typically observed in postmenopausal women and men. It is similar in magnitude to those described in earlier studies [19–22], suggesting that no substantial improvements in variability occurred over the past years.

An European Menopause and Andropause Society (EMAS) position statement recommends confirming diagnosis of premature ovarian failure with E2 levels below 50 pmol/L (14 pg/mL) along with other parameters [23]. In this study, one sample with a target value of 14.1 pg/mL (Supplemental Table 1), had values reported between 9.4 and 64.8 pg/mL with many assays overestimating the actual E2 concentration. Furthermore, reference ranges for E2 in men were suggested with the lower level of normal to be 28 pmol/L (7.6 pg/mL) [24]. One sample with a target value of 7.21 pg/mL had values reported ranging between 5 and 38.5 pg/mL. The between-assay variability determined in this study does not allow for a consistent diagnosis of premature ovarian failure, as suggested by EMAS, nor does it allow for the reliable detection of abnormal estradiol levels in men.

One source for the high between-assay variability, especially at low E2 concentrations, appears to be lack of specificity. All assays show an increase in measurement bias with decreasing E2 target concentrations, suggesting that compounds other than E2 contribute to the measurement result. Furthermore, several assays report E2 values within the assay's reportable range for certain samples even though the actual target values are below the reportable range (Table 2). These observations are less pronounced with mass spectrometry-based assays than with immunoassays, suggesting that mass spectrometry-based assays in general have a greater specificity than immunoassays.

The sensitivity of the investigated assays differs highly among assays. The reportable range for most assays does not reach below 10 pg/mL and those assays reaching below 10 pg/mL, show a high positive measurement bias at these low E2 concentrations. E2 levels in patients responding to aromatase inhibitor therapy typically have E2 levels below 10 pg/mL [25]. Thus, most assays would not be sufficiently sensitive to monitor patients on aromatase inhibitor therapy or may falsely indicate that the goal of such a therapy has not been reached, leading to an inappropriate change in therapy. These findings support similar observations [26].

Our new model estimates the contribution of different sources of variability to the overall variability of an assay. It estimates the contribution of calibration bias, sample specific effects and assay impression. Results obtained with this model suggest that for the investigated E2 concentration range (2.5–285 pg/ml), calibration bias is the major contributor to the overall variability for most assays. This source of variability can be minimized through appropriate calibration to a higher metrological reference and verification of measurement accuracy.

We identified eight assays where recalibration may not lead to significant improvements in assay performance. For these assays, standardization efforts need to also focus on minimizing sources for assay imprecision and sample-related effects such as specificity. Our findings suggest that problems in assay specificity affect all assays independent of the assay technology. The presence and impact of interfering compounds on E2 measurements can vary among patients. Therefore, improvements in variability due to interfering compounds can be achieved most efficiently by using multiple single-donor patient samples for calibration verification. Such samples are provided and used in the CDC Hormone Standardization Program [27].

To assure measurement accuracy and precision meet current needs in patient care and research, common analytical performance goals for allowable bias and imprecision are needed. Analytical performance criteria derived from data on biological variability were calculated to be $\pm 12.5\%$ for minimal desirable bias [28,29]. However, these performance criteria were derived from premenopausal women and did not include data from postmenopausal women.

Epidemiologic studies in postmenopausal women investigate health effects and disease risks by comparing women with E2 levels in the higher to those in the lower percentile ranges. To appropriately categorize women's E2 concentrations, the bias across assays should be low enough to consistently distinguish E2 levels in the higher and lower quartiles, i.e., to identify women potentially at increased risk for breast cancer. Using data from 9 studies on postmenopausal women [1,31], the median width of quartiles was 2.5 pg/mL (95% CI- 2.2–3.4 pg/mL) (Supplemental Table 2). Thus, to reliably distinguish between highest and lowest quartiles across assays, a maximum allowable bias of 2.5 pg/mL would be desirable. At a concentration of 20 pg/mL, an allowable bias of 2.5 pg/mL corresponds to a 12.5 % bias. Therefore, the maximum allowable bias would be ± 2.5 pg/mL at concentrations ≤ 20 pg/mL and $\pm 12.5\%$ at concentrations >20 pg/mL. Using different performance criteria at high and low E2 concentrations was proposed earlier by an expert group [30]. However, these criteria did not consider clinical use of E2 measurements in either postmenopausal women or women on aromatase inhibitor therapy.

The bias criteria derived with this proposed approach reflects the allowable bias for an individual patient measurement that is comprised of calibration bias and bias due to non-specificity or matrix-effects. Therefore, the accuracy across a number of individual patient samples needs to be assessed to determine whether an assay is sufficiently accurate. Criteria commonly used in proficiency testing/external quality assurance programs require participants to meet the target for 80% of the reported results. Three participants (Assay 14, 15, 16- Supplemental Table 3) would meet these proposed bias criteria for 80% of the reported results.

This study focused on E2 measurements at low concentrations typically observed in men, pre- and post-menopausal women, because of the clinical importance of E2 measurements in these populations; e.g., cancer risk assessments, cancer treatments, and clinical diagnosis. Further studies are needed to assess the measurement performance at high concentrations typically needed in assisted reproductive medicine.

In summary, this study showed substantial and clinically relevant variability in serum E2 measurements in samples from men, and pre- and post-menopausal women. Most assays do not have the analytical sensitivity and specificity to reliably measure E2 levels in men and in postmenopausal women, especially to monitor specific cancer therapies. Thus, the analytical performances of most E2 assays do not meet current needs in research and patient care. The CDC Hormone Standardization Program intends to improve E2 measurements across assays by providing panels of individual donor sera to participants for accuracy assessment, calibration, and to help resolve bias related to specificity and sample matrix.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the Division of Laboratory Sciences at the National Center for Environmental Health and the Division of Cancer Prevention and Control at the National Center for Chronic Disease Prevention and Health Promotion. Professional societies and organizations, such as the Endocrine Society, the American Association of Clinical Endocrinologist, and the American Association of Clinical Chemistry have provided additional support. The authors also acknowledge the hormone standardization research team at CDC: Dr. Yuesong Wang, Jacob Farris, Hans C Cooper, Matthew Gatcombe, Christopher Shacklady, and Gabrielle Gay. Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

References

1. Key T, Appleby P, Barnes I, Reeves G. Endogenous hormones and breast cancer collaborative group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst.* 2002; 94:606–16. [PubMed: 11959894]
2. Kaaks R, Rinaldi S, Key TJ, Berrino F, Peeters PH, Biessy C, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Enocr Rlat Cancer.* 2005; 12:1071–82.
3. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J Natl Cancer Inst.* 2004; 96:1856–65. [PubMed: 15601642]
4. Key TJ, Appleby PN, Reeves GK, Roddam AW, Helzlsouer KJ, Alberg AJ, et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer.* 2011; 105:709–22. [PubMed: 21772329]
5. Zhang X, Tworoger SS, Eliassen AH, Hankinson SE. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. *Breast Cancer Res Treat.* 2013; 137:883–92. [PubMed: 23283524]
6. Lee JS, Yaffe K, Lui LY, Cauley J, Taylor B, Browner W, et al. Prospective study of endogenous circulating estradiol and risk of stroke in older women. *Arch Neurol.* 2010; 67:195–201. [PubMed: 20142527]
7. Chen Y, Zeleniuch-Jacquotte A, Arslan AA, Wojcik O, Toniolo P, Shore RE, et al. Endogenous hormones and coronary heart disease in postmenopausal women. *Atherosclerosis.* 2011; 216:414–9. [PubMed: 21367421]
8. Laughlin GA, Kritz-Silverstein D, Barrett-Connor E. Endogenous oestrogens predict 4-year decline in verbal fluency in postmenopausal women: the Rancho Bernardo study. *Clin Endocrinol (Oxf).* 2010; 72:99–106. [PubMed: 19508596]
9. Roddam AW, Appleby P, Neale R, Dowsett M, Folkerd E, Tipper S, et al. Association between endogenous plasma hormone concentrations and fracture risk in men and women: the EPIC-Oxford prospective cohort study. *J Bone Miner Metab.* 2009; 27:485–93. [PubMed: 19333682]

10. [accessed 24.06.2013] ClinicalTrials.gov search for estradiol estradiol | Recruiting | Exclude Unknown. Available at: <<http://clinicaltrials.gov>
11. http://www.cap.org/apps/cap.portal?_nfpb=true?cntvwrPtl_t_action Override=%2Fportlet%2FcontentViewer%2Fshow&_windowLabel=cntvwrPtl_t&cntvwrPtl_t%2FactionForm.contentReference%2FD=cap_today%2F0611%2F0611_CAPTODAY_ImmunoanalyzersGuide.html&_state=maximized&_pageLabel=cntvwr; [accessed 26.06.2013].
12. Stanczyk FZ, Clarke NJ. Advantages and challenges of mass spectrometry assays for steroid hormones. *J Steroid Biochem Mol Biol.* 2010; 121:491–5. [PubMed: 20470886]
13. Watts NB, Adler RA, Bilezikian JP, Drake MT, Eastell R, Orwoll ES, et al. Osteoporosis in men: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab.* 2012; 97:1802–22. [PubMed: 22675062]
14. Rosner W, Hankinson SE, Sluss PM, Vesper HW, Wierman ME. Challenges to the measurement of estradiol: an endocrine society position statement. *J Clin Endocrinol Metab.* 2013; 98:1376–87. [PubMed: 23463657]
15. Clinical Laboratory Standards Institute. Preparation and validation of commutable frozen human serum pools as secondary reference materials for cholesterol measurement procedures. Wayne, PA: Clinical Laboratory Standards Institute; 1999. CLSI document C-37A
16. Thienpont LM, Patrick GV, Van Brussel KA, De Leenheer AP. Estradiol-17 β quantified in serum by isotope dilution-gas chromatography-mass spectrometry: reversed-phase C18 high-performance liquid chromatography compared with immune-affinity chromatography for sample pretreatment. *Clin Chem.* 1988; 34:2066–9. [PubMed: 3168218]
17. Botelho JC, Shacklady C, Vesper HW. Development of a candidate reference measurement procedure for the determination of estradiol in human serum. *Clin Chem.* 2011; 57(Suppl):A1.
18. Clinical Laboratory Standards Institute. Method comparison and bias estimation using patient samples. Wayne, PA: Clinical Laboratory Standards Institute; 2002. CLSI document EP9
19. Coucke W, Devleeschouwer N, Libeer JC, Schiettecatte J, Martin M, Smits J. Accuracy and reproducibility of automated estradiol-17 β and progesterone assays using native serum samples: results obtained in the Belgian external assessment scheme. *Hum Reprod.* 2007; 22:3204–9. [PubMed: 18025029]
20. Yang DT, Owen WE, Ramsay CS, Xie H, Roberts WL. Performance characteristics of eight estradiol immunoassays. *Am J Clin Pathol.* 2004; 122:332–7. [PubMed: 15362362]
21. Reinsberg J, Bätz O, Bertsch T, Bewarder N, Deschner W, Drescher V, et al. Precision and long-term stability of different estradiol immunoassays assessed in a multi-center quality control study. *Clin Lab.* 2009; 55:201–6. [PubMed: 19728553]
22. Middle JG. External quality assessment of steroid hormone assays in the United Kingdom. *Ann 1st Super Sanita.* 1991; 27:459–466.
23. Vujovic S, Brincat M, Erel T, Gambacciani M, Lambrinoudaki I, Moen MH, et al. EMAS position statement: managing women with premature ovarian failure. *Maturitas.* 2010; 67:91–3. [PubMed: 20605383]
24. Yeap BB, Alfonso H, Chubb SA, Handelsman DJ, Hankey GJ, Norman PE, et al. Reference ranges and determinants of testosterone, dihydrotestosterone, and estradiol levels measured using liquid chromatography-tandem mass spectrometry in a population-based cohort of older men. *J Clin Endocrinol Metab.* 2012; 97:4030–9. [PubMed: 22977273]
25. Nagao T, Kira M, Takahashi M, Honda J, Hirose T, Tangoku A, et al. Serum estradiol should be monitored not only during the peri-menopausal period but also the post-menopausal period at the time of aromatase inhibitor administration. *World J Surg Oncol.* 2009; 7:88. [PubMed: 19909552]
26. Jaque J, Macdonald H, Brueggmann D, Patel SK, Azen C, Clarke N, et al. Deficiencies in immunoassay methods used to monitor serum estradiol levels during aromatase inhibitor treatment in postmenopausal breast cancer patients. *SpringerPlus.* 2013; 2:5. [PubMed: 23520572]
27. CDC. [accessed 26.06.2013] <http://www.cdc.gov/labstandards/hs.html>
28. Fraser CG, Kallner A, Kenny D, Petersen PH. Introduction: strategies to set global quality specifications in laboratory medicine. *Scand J Clin Lab Invest.* 1999; 59:477–8. [PubMed: 10667682]

29. Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest*. 1999; 59:491–500. [PubMed: 10667686]
30. Thienpont L. Meeting report: first and second estradiol international workshop. *Clin Chem*. 1996; 42:112–4.
31. Key TJ. Endogenous oestrogens and breast cancer risk in premenopausal and postmenopausal women. *Steroids*. 2011; 76:812–5. [PubMed: 21477610]

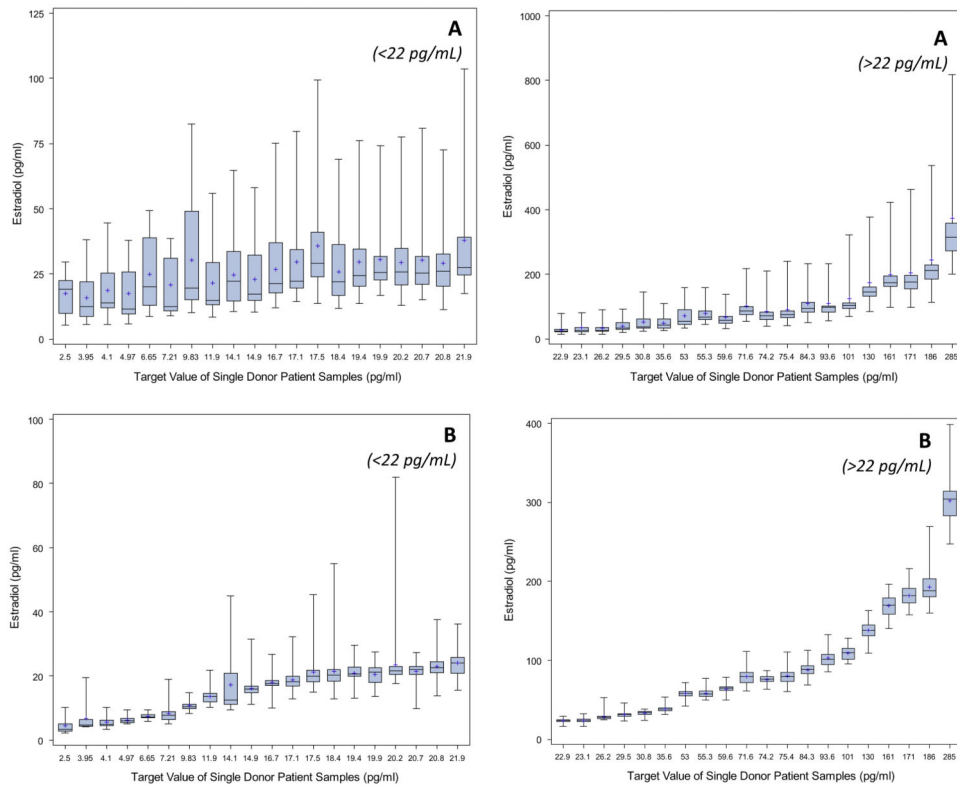


Fig. 1. Box-and-Whisker plot of distribution of individual results by sample and assay technology (Panel A: Immunoassays, Panel B: Mass spectrometry assays).

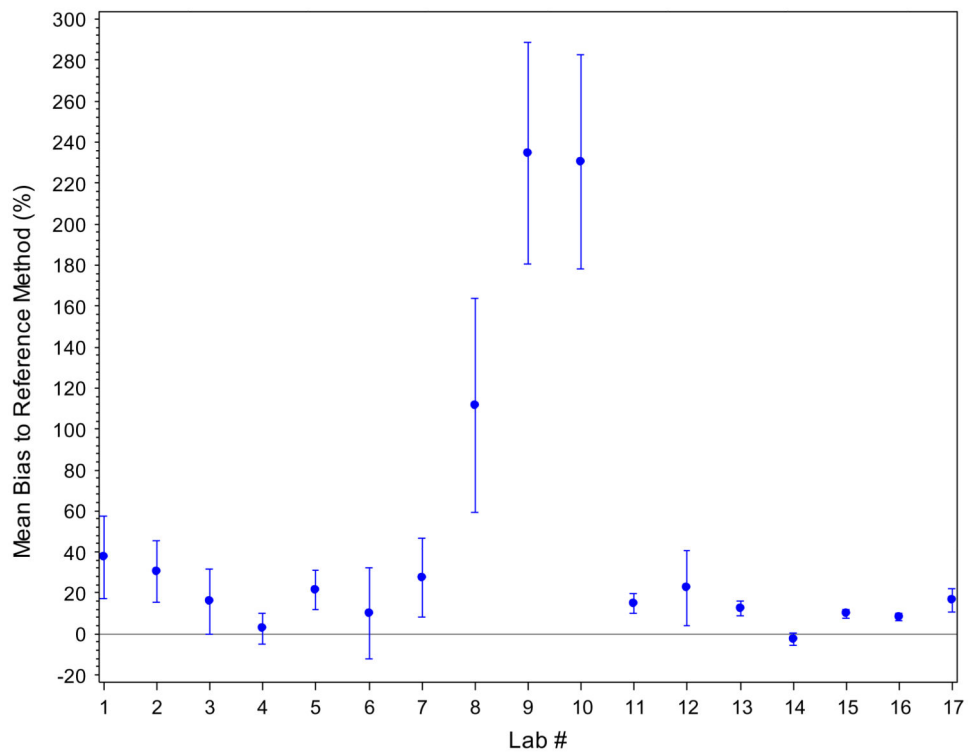


Fig. 2. Mean percent bias (95% CI) of each participating laboratory.

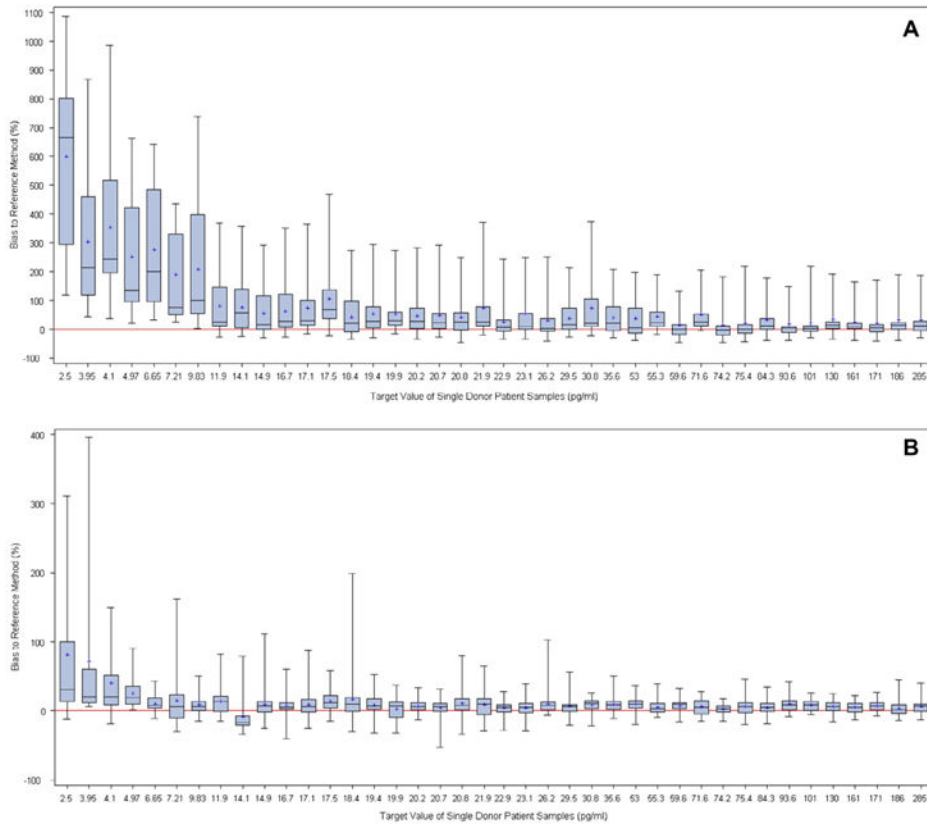


Fig. 3. Bias distribution of individual results by sample and assay technology (Panel A: Immunoassays, Panel B: Mass spectrometry assays).

Table 1

Characteristics of participating assays.

LabID#	Lab type	Assay type ^a	Sample volume used (lL)	Reportable range (pg/mL) ^b	Limit of detection (pg/mL) ^b
1	Manufacturer	IA	75	25.0–3250	25
2	Manufacturer	IA	35	5.00–4300	5
3	Manufacturer	IA	80	11.8–3000	11.8
4	Manufacturer	IA	80	11.8–3000	11.8
5	Manufacturer	IA	12	5.00–1500	5
6	Clinical laboratory	IA	25	6.81–3814	6.36
7	Manufacturer	IA	35	20.0–4800	20
8	Manufacturer	IA	35	20.0–4800	20
9	Manufacturer	IA	25	6.36–3814	6.36
10	Manufacturer	IA	25	6.36–3814	6.36
11	Clinical laboratory	IA	35	20.0–4300	5
12	Clinical laboratory	MS	200	2.00–5000	2
13	Clinical laboratory	MS	1000	P1.00	1
14	Clinical laboratory	MS	200	1.00–2000	0.5
15	Clinical laboratory	MS	500	10.0–600	4.31
16	Clinical laboratory	MS	350	0.50–4000	0.5
17	Clinical laboratory	MS	200	9.80–5000	4.9

^a IA, immunoassay; MS, mass spectrometry assay.

^b As provided by participant.

Table 2

Regression analysis and mean bias to the reference method using single donor sera.

LabID#	Intercept (95% CI)	Slope_X (95% CI)	Slope_X2 (95% CI)	Mean percent bias (95% CI)%	Median imprecision ^a (min-max)%CV	Number of results reported on samples with target values outside the reportable range
1	16.69 (10.00–23.39)	0.8062 (0.6996–0.9128)		37.5 (17.6–57.5)	6 (3–15)	55
2	5.11 (2.99–7.23)	0.9446 (0.8574–1.0318)	0.00099 (0.00047–0.0015)	30.5 (15.6–45.4)	5 (1–21)	17
3	3.58 (0.63–6.53)	0.9327 (0.869–0.9964)		15.9 (0.1–31.7)	7 (2–15)	8
4	-3.48 (-7.60–0.65)	1.123 (1.0407–1.2053)		2.65 (-4.66–9.96)	8 (2–21)	3
5	3.76 (2.06–5.47)	0.9855 (0.9545–1.0165)		21.4 (11.9–30.9)	7 (1–18)	13
6	7.01 (5.37–8.65)	0.6137 (0.5672–0.6602)		10.1 (-12.2–32.4)	6 (2–14)	30
7	14.23 (10.13–18.33)	0.6297 (0.4662–0.7932)	0.00189 (0.00086–0.00291)	27.5 (8.5–46.4)	13 (5–27)	19
8	17.7 (14.60–20.80)	0.8378 (0.6884–0.9873)	0.00137 (0.00033–0.0024)	111 (59.3–163)	15 (6–29)	58
9	15.29 (9.61–20.96)	2.2362 (2.0614–2.4111)		235 (181–289)	9 (4–21)	21
10	16.1 (10.74–21.47)	2.1687 (2.0221–2.3153)		230 (178–283)	16 (4–30)	21
11	5.07 (1.95–8.19)	0.932 (0.8315–1.0325)	0.00093 (0.00042–0.00145)	14.7 (9.88–19.6)	5 (1–13)	24
12	2.58 (1.21–3.94)	0.9926 (0.9559–1.0293)		22.5 (4.19–40.8)	8 (3–41)	0
13	1.23 (0.42–2.04)	1.0514 (1.02–1.0828)		12.4 (8.77–15.9)	7 (4–17)	0
14	0.44 (-0.08–0.96)	0.9400 (0.9205–0.9595)		-2.41 (-5.38–0.56)	6 (2–57)	0
15	0.11 (-0.56–0.78)	1.0957 (1.0776–1.1139)		9.79 (7.93–11.66)	5 (2–12)	6
16	-0.45 (-0.95–0.04)	1.1083 (1.0904–1.1262)		8.31 (6.50–10.13)	3 (1–21)	0
17	-0.12 (-2.37–2.14)	1.1599 (1.1053–1.2144)		16.39 (10.84–21.94)	14 (7–75)	1

^aDetermined for each sample from duplicate measurements on three days (*n* = 6).

Table 3

Overall assay variability and proportions due to calibration bias, specificity/matrix effects, and imprecision.

Lab ID	Overall variability ^a	Proportion of calibration bias (%) ^b	Proportion of non-specificity/matrix effects (%) ^b	Proportion of imprecision (%) ^b
1	308	45	50	6
2	60	73	21	6
3	49	14	75	11
4	95	20	63	17
5	33	33	47	20
6	119	86	11	3
7	154	35	17	49
8	333	74	8	18
9	2903	93	5	2
10	3316	91	4	4
11	68	71	20	9
12	23	25	40	35
13	12	47	23	30
14	5	34	23	43
15	18	73	8	19
16	9	80	12	8
17	118	36	16	48

^aSum of the weighted sums of squares of the various sources of error (calibration bias, specificity/matrix effects, and imprecision), after using a weighted regression (weight = $1/\text{varAssay Result}$) to model the relationship between the results reported for each sample and the reference values assigned by the RMP.

^bProportion of individual factor on the overall variability.