COLLEGE of AMERICAN PATHOLOGISTS

Calcium

Measurement of total calcium is widely used for both the diagnosis and the monitoring of a range of conditions related to the bones, heart, nerves, and kidneys. Total calcium measurements include protein-bound calcium and ionized calcium. Calcium concentrations in blood are highly regulated by the endocrine system, with concentrations ranging in healthy adults between 8.6-10.2 mg/dL[1]. Thus, the samples used in this Survey are in a range typically observed in healthy individuals.

Total calcium is commonly measured by spectrophotometric determination with metallochromic indicators (i.e., cresolphthalein complexone or arsenazo III dye) and by indirect potentiometry, which utilizes an electrode selective for calcium. A maximum imprecision (expressed as %CV) of 0.9% and a maximum bias against a reference method of 0.7% have been proposed as analytical performance criteria for serum calcium[2].

Due to the small number of participants, no instrument peer group statistics could be performed. Therefore, participants were combined into three method groups based on the principal measurement approach: spectrophotometric methods using arsenazo III dye (four different instrument methods from different manufacturers), cresolphthalein complexone (two different methods from Roche Diagnostics), and potentiometric methods (one method from Beckman Coulter). Because reference values were not available in this Survey for serum calcium, measurement bias was assessed by comparing the means of each of the three groups against each other.

The potentiometric method maintains the highest level of precision in all four samples (1.4-1.7%CV), while the methods using arsenazo III dye show overall similar %CV values. The imprecision of the methods using cresolphthalein complexone is approximately two times that of the potentiometric method. None of the three method groups meet the suggested performance criterion for imprecision.

The differences among group means range from -0.2% to 2.8%, with the methods using cresolphthalein complexone and ion selective electrodes having the closest agreement (Table 1). The difference between these two methods meets the suggested bias criterion for 3 out of the 4 samples, while the difference between the other methods' means is two to three times higher than the suggested measurement bias. Across all four samples, the differences among method means appear similar, which suggests that these differences are mainly due to variances in assay calibration rather than assay specificity. Further studies are needed to verify these assertions. The data suggest that the observed measurement bias can be minimized by calibration to a common reference. Further investigations are needed to identify the causes of the high imprecision observed within these assays.

Table 1: Difference between method means in percent:

	ABS-01	ABS-02	ABS-03	ABS-05
Arsenazo III dye method vs. Cresolphthalein method	1.7	2.0	1.8	2.3
Cresolphthalein method vs. ion selective electrode method	0.4	0.8	0.3	-0.2
Arsenazo III dye method vs. ion selective electrode method	2.2	2.8	2.1	2.1

Cortisol

Serum cortisol levels are used to aid in both the diagnosis of Cushing's syndrome and the discrimination between primary and secondary adrenal insufficiency. About 90% of circulating cortisol is bound to serum proteins such as albumin and cortisol-binding globulin. Serum cortisol concentrations show diurnal variation with a peak in the morning and a nadir in the afternoon. In morning samples taken from healthy adults, concentrations roughly range between 5-25 µg/dL. The samples used in this Survey cover the lower and upper range of concentrations commonly observed in healthy adults.

The measurement of serum cortisol is mainly performed by RIA and chemiluminescent immunometric technologies. High-performance liquid chromatography with tandem mass spectrometry is becoming increasingly common as an alternate technology. Analytical performance goals for serum cortisol were proposed to be 12.5%CV for imprecision and 12.5% for bias relative to a reference method[3].

Because no reference values were available for the survey samples, the median values of each peer group were compared against each other in order to assess measurement bias. The smallest differences are observable between the Beckman UniCel Dxl and the Abbott Architect I assays, while the highest differences are observable upon comparison of any method against the Roche cobas e600series/E170 assays, which consistently measure higher than the other methods. Nine out of the 24 comparisons show a bias within the suggested bias criterion (Table 2). The differences among methods are similar across samples, thus suggesting that the observed differences are due mainly to dissimilarities in method calibration. The levels of imprecision are well below the suggested imprecision criterion for all methods except the Siemens ADVIA Centaur/XP with sample ABS-02.

Table 2: Differences between method medians in percent from a pairwise comparison of methods:

	ABS-01	ABS-02	ABS-03	ABS-05
Abbott Architect I System vs. Beckman UniCel Dxl	-6	-2	-6	-1
Abbott Architect I System vs. Roche cobas e600series/E170	-33	-26	-33	-29
Abbott Architect I System vs. Siemens ADVIA Centaur/XP	-17	-12	-18	-18
Beckman UniCel Dxl vs. Roche cobas e600series/E170	-27	-24	-26	-28
Beckman UniCel Dxl vs. Siemens ADVIA Centaur/XP	-12	-10	-11	-17
Roche cobas e600series/E170 vs. Siemens ADVIA Centaur/XP	15	14	15	11

Overall, the participating cortisol methods meet suggested performance criteria for imprecision. The data suggest that the observed measurement bias can be minimized by calibration to a common reference.

Estradiol

Measurement of estradiol in serum aids in the diagnosis of diseases, such as amenorrhea, and in the assessment of fertility; measurements are also used to monitor follicle stimulation therapy. Additionally, estradiol measurements provide information about patient response to treatments such as those using aromatase inhibitors or gonadotropin-releasing hormone agonists. Circulating estradiol, like testosterone, is mainly bound to sex hormone binding globulin and albumin. Serum estradiol levels in premenopausal women change during the course of the menstrual cycle: a concentration of about 50 pg/mL or less presents in the preovulatory period, while concentration levels reach about 150-500 pg/mL in the second half of the follicular phase. In

postmenopausal women, estradiol levels are roughly 30 pg/mL or less, while in men, values are roughly between 10-50 pg/mL. Based on these ranges, the samples used in this Survey have serum estradiol concentrations that are typically observed in healthy men, pre-menopausal women, and postmenopausal women.

Serum estradiol measurements are mainly performed using immunoassays (RIA and chemiluminescent). Mass spectrometry coupled with either high-performance liquid chromatography or gas chromatography is viable as an alternate technology, but it is less frequently utilized for this analyte. Analytical performance goals for serum estradiol were proposed to be 11.4%CV for imprecision and 8.3% for bias as compared to a reference method[4].

For sample ABS-05, the sample with the highest estradiol concentration, three out of four peer groups meet the suggested precision criterion (Abbott Architect I System, Roche cobas e600series/E170, Siemens ADVIA Centaur/XP eE2). None of the peer groups meet the precision criterion with the other three samples. Each peer group has at least two samples that meet the suggested bias criterion, except for the Siemens ADVIA Centaur/XP eE2 group, which has only one sample meeting the criterion (Table 3). Except for the Beckman UniCel Dxl group, the bias is either consistently positive or negative, suggesting that calibration is one contributor to the observed measurement bias. However, within each peer group, the magnitude of bias across the four samples is very different, which suggests that specificity is another relevant cause of the observed measurement bias.

Table 3: Measurement bias of peer group median to the reference method in percent:

	ABS-01	ABS-02	ABS-03	ABS-05
Abbott Architect I System	-17.0	-12.9	-0.3	-7.7
Beckman UniCel Dxl	-5.1	-5.3	14.0	16.0
Roche cobas e600series/E170	6.7	2.3	5.4	20.1
Siemens ADVIA Centaur/XP eE2	14.6	17.4	22.5	0.0

In summation, most peer groups only meet the suggested imprecision criterion of 11.4%CV in the samples with elevated estradiol levels (ABS-05: 169 pg/mL). Differences in calibration and assay specificity seem to contribute to the observed measurement bias; both sources of bias can be further assessed and minimized using panels of single donor patient samples as described in the section discussing the testosterone results.

Sex hormone binding globulin (SHBG) and free testosterone

Sex hormone binding globulin (SHBG) is primarily ordered in conjunction with total testosterone to evaluate the status of free testosterone, which is considered the biological, active form of testosterone. In most, but not all, clinical conditions, a measurement of total testosterone is adequate for the evaluation of a patient. Because SHBG concentrations can be influenced by many factors (e.g., obesity, testosterone treatment polycystic ovary syndrome, aging, pregnancy, and estrogen therapy), there are clinical situations in which measured concentrations of total testosterone may not reflect the free or bioavailable concentrations. In these circumstances, a test for assessing bioavailable or free testosterone can be helpful in clinical decision-making[5]. Normal ranges for free testosterone are about 50-200 pg/mL and 1.0-8.5 pg/mL for healthy men and women, respectively[1]. The samples used in this Survey cover the aforementioned concentration range commonly observed in healthy adults.

The measurement of SHBG is performed by immunoassays. Analytical performance goals were proposed to be 6.1 %CV for imprecision and 11.1% for bias relative to a reference method[4]. Free testosterone is either calculated

using both the total testosterone and SHBG measurement results, or it is measured using direct assays. For the former calculation, different formulas are used. For free testosterone, only a performance criterion for precision was suggested, the value of which was proposed to be 4.7 %CV[4].

Because of insufficient data, peer group statistics were calculated for only one of the two SHBG and free testosterone peer groups. The precision for the Roche cobas e600series/E170 group is well below the suggested criterion. No reference method for SHBG exists, and target values are assigned to these samples; therefore, the difference between the medians of the two peer groups was calculated to assess measurement bias (Table 4). The difference between both peer group medians is below the suggested bias criterion. In contrast, free testosterone imprecision is above 70% in all samples, and the median value for the calculated free testosterone is seven to nine times higher than median values for the direct assays. Thus, results for free testosterone are not comparable within and among peer groups.

Considering the good agreement of the SHBG measurements, the high variability in calculated free testosterone could be explained with the variability in total testosterone measurements and with the use of different formulas to calculate free testosterone; there are four different formulas suggested for calculating free testosterone. A study investigating the impact of analytical variability of testosterone, SHBG, and the use of different formulas on free testosterone calculation reported that variations between commercially available total testosterone immunoassays have a strong impact on calculation of free testosterone, with total testosterone assays contributing to 82.2% of overall variance, compared with 13.7% for the calculation algorithms and 4.1% for the SHBG assays[6].

Table 4: Bias of SHBG measurements among peer group medians:

	Bias (%)
ABS-01	9.3
ABS-02	1.1
ABS-03	2.9
ABS-05	7.0

In summary, while SHBG measurement performance meets the suggested performance criteria, free testosterone measurements are highly variable. Variability in total testosterone measurements and the use of different formulas could explain the profound variability in free testosterone results. Further studies to better identify and minimize the sources of variability are warranted.

Testosterone

Serum testosterone tests aid in the diagnosis of a variety of conditions, such as hypogonadism, polycystic ovary syndrome, certain tumors, and androgen deficiency in men. About 60% of circulating testosterone is bound to sex hormone binding globulin and about 38% to albumin. Serum testosterone shows diurnal patterns with a peak in the morning and a nadir in the afternoon. Serum concentrations fall in an approximate range of 300-1,200 ng/dL and 30-95 ng/dL in men and women, respectively[7]. The samples used in this Survey have concentrations commonly measured in healthy men and women.

Serum testosterone measurements are mainly performed using immunoassays (RIA and chemiluminescent) and mass spectrometry coupled with either high-performance liquid chromatography or gas chromatography.

Analytical performance goals for serum testosterone were proposed to be 5.3%CV for imprecision and 6.4% for bias relative to a reference method[8].

Due to the small number of participants, means and standard deviations could not be calculated for all peer groups; for this reason, peer group medians were used in this assessment. Among the peer groups where imprecision could be calculated, only the Roche cobas e600series/E170 peer group meets the imprecision criterion in the two samples with high testosterone values (ABS-01: 865 ng/dL, ABS-02: 558 ng/dL). Only 3 peer groups (Beckman UniCel Dxl, Mass Spectrometry, and Siemens ADVIA Centaur/XP) have a single sample that passes the suggested bias criterion of 6.4% (Table 5). The measurement biases within the peer groups (across the four samples) are very inconsistent, suggesting that measurement specificity, rather than variation in calibration, is the major source of bias. The observed bias is negative for most samples, which indicates a lack of analyte recovery. Incomplete removal of testosterone from sex hormone binding globulin could lead to incomplete recovery and may explain in part the negative bias observed in this Survey. Sample ABS-05 has the lowest sex hormone binding globulin concentrations and shows consistently positive biases across all peer groups, which seems to support this hypothesis.

Table 5: Measurement bias of peer group median to the reference method in percent:

	ABS-01	ABS-02	ABS-03	ABS-05
Abbott Architect I System	-8.2	6.6	-17.6	34.0
Beckman UniCel Dxl	-34.9	-25.4	-4.3	80.4
Mass Spectrometry	-21.0	-11.1	-30.9	3.1
Roche cobas e600series/E170	-17.6	-9.0	-20.2	28.9
Siemens ADVIA Centaur/XP	-31.3	-24.6	3.7	64.9

In summation, most assays do not meet current performance goals for imprecision and bias. The data suggest assay specificity to be one major contributor to the observed measurement bias; sources contributing to the observed high imprecision need to be further investigated. Additional information about the contribution of assay specificity on measurement bias can be obtained by analyzing panels of samples obtained from different donors and also with target values assigned by a reference method. Measuring individual donor samples would provide information about the frequency of interferences in different samples and about the impacts such interferences have on measurement bias.

Thyroid Stimulating Hormone

Thyroid Stimulating Hormone (TSH) tests are used to evaluate thyroid function and/or symptoms of hyperthyroidism or hypothyroidism. Normal values range from approximately 0.4-4.0 mIU/L [7], and therefore, the samples used in this Survey are within the range commonly observed in healthy individuals.
TSH measurements are performed by immunoassays. Analytical performance goals were proposed to be 9.7 %CV for imprecision and 7.8% for bias relative to a reference method[4].

The imprecision within peer groups is well below the proposed imprecision criterion for samples. The Roche cobas e600series/E170 group consistently maintains the highest median values for four samples, while the Abbott Architect I System group presents the lowest median values for three of the four samples. When comparing the medians of each peer group, 17 out of the 40 comparisons show differences that meet the suggested criterion for measurement bias (Table 6). However, the biases within each comparison appear highly variable across the four

samples, except for the comparison of the two Siemens peer groups. This suggests that assay specificity is one major contributor to the observed measurement bias. These findings seem consistent with findings from another study[9].

Table 6: Differences between method medians in percent from a pairwise comparison of methods

	ABS-01	ABS-02	ABS-03	ABS-05
Group 1 vs. Group 2	-7.8	-20.6	-9.0	10.4
Group 1 vs. Group 3	-19.4	-30.7	-12.3	-20.6
Group 1 vs. Group 4	-11.5	-18.8	-11.5	-7.6
Group 1 vs. Group 5	-5.9	-16.3	-9.8	-3.2
Group 2 vs. Group 3	-11.6	-10.3	-3.4	-30.9
Group 2 vs. Group 4	-3.7	1.8	-2.5	-17.9
Group 2 vs. Group 5	1.9	4.3	-0.9	-13.6
Group 3 vs. Group 4	7.9	12.1	0.8	13.1
Group 3 vs. Group 5	13.5	14.6	2.5	17.4
Group 4 vs. Group 5	5.6	2.5	1.7	4.4

Group 1: Abbott Architect I System
Group 2: Beckman UniCel DxL 3rd Generation
Group 3: Roche cobas e600series/E170
Group 4: Siemens ADVIA Centaur/XP
Group 5: Siemens ADVIA Centaur/XP TSH3-Ultra

In summary, the assays in this Survey meet the suggested performance criteria for precision. The performance criteria for measurement bias are not uniformly met, which is probably due to problems related to assay specificity. Suitable reference materials are currently not available, and a harmonization program conducted by the International Federation for Clinical Chemistry and Laboratory Medicine (IFCC) was established to address the described variability and provide materials that can be used to better assess problems affecting measurement bias, such as differences in calibration and specificity.

References:

1. Tietz Textbook of Clinical Chemistry and Molecular Diagnostics. Burtis CA, Ashwood ER, Bruns DE, eds. 5th edition, St. Louis: Elsevier Saunders; 2011.
2. Stöckl D et al. Desirable Routine Analytical Goals for Quantities Assayed in Serum. Eur J Clin Chem Clin Biochem. 1995; 33:157-169.
3. Klee GG. Establishment of Outcome-Related Analytic Performance Goals. Clin Chem. 2004;56:714-722.
4. Ricós C, et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest. 1999;59:491-500.
5. Swerdloff Rs, el al. Free testosterone measurement by the analog displacement direct assay: old concerns and new evidence. Clin Chem. 2008;54:458-60.
6. Sartorius G, et al. Predictive accuracy and sources of variability in calculated free testosterone estimates. Ann Clin Biochem. 2009;46:137-43.
7. Henry JB. Clinical diagnosis and management by laboratory methods. eds. 20th edition, Philadelphia, PA: WB Saunders; 2001.
8. Yun YM et al. Performance criteria for testosterone measurements based on biological variation in adult males: recommendations from the Partnership for the Accurate Testing of Hormones. Clin Chem. 2012;58:1703-10.

9. Thienpont LM et a. Report of the IFCC Working Group for Standardization of Thyroid Function Tests, part 1: Thyroid-stimulating hormone. Clin Chem 2010;56:902–11.

Hubert Vesper, PhD
Chemistry Resource Committee