Clinical Chemistry 60:3
541–548 (2014)

Interlaboratory Agreement of Insulin-like Growth Factor 1 Concentrations Measured by Mass Spectrometry

Holly D. Cox,[1] Filipe Lopes,[2] Getachew A. Woldemariam,[3] Jessica O. Becker,[4] Mark C. Parkin,[2] Andreas Thomas,[5] Anthony W. Butch,[3] David A. Cowan,[2] Mario Thevis,[5] Larry D. Bowers,[6] and Andrew N. Hoofnagle[4*]

BACKGROUND: Insulin-like growth factor 1 (IGF-1)[7] is a key mediator of growth hormone (GH) action and a well-characterized biomarker of GH abuse. Current immunoassays for IGF-1 suffer from poor concordance between platforms, which makes comparison of results between laboratories difficult. Although previous work has demonstrated good interlaboratory imprecision of LC-MS/MS methods when plasma is supplemented with purified proteins, the interlaboratory imprecision of an endogenous protein in the nanogram-per-milliliter concentration range has not been reported.

METHODS: We deployed an LC-MS/MS method to quantify serum IGF-1 in 5 laboratories using 5 different instruments and analyzed 130 healthy human samples and 22 samples from patients with acromegaly. We determined measurement imprecision (CV) for differences due to instrumentation, calibration curve construction, method of calibration, and reference material.

RESULTS: Instrument-dependent variation, exclusive of digestion, across 5 different instrument platforms was determined to be 5.6%. Interlaboratory variation was strongly dependent on calibration. Calibration materials from a single laboratory resulted in less variation than materials made in individual laboratories (CV 5.2% vs 12.8%, respectively). The mean imprecision for 152 samples between the 5 laboratories was 16.0% when a calibration curve was made in each laboratory and 11.1% when a single-point calibration approach was used.

CONCLUSIONS: The interlaboratory imprecision of serum IGF-1 concentrations is acceptable for use of the assay in antidoping laboratories and in standardizing results across clinical laboratories. The primary source of variability is not derived from the sample preparation but from the method of calibration.

Current immunoassays for insulin-like growth factor 1 (IGF-1) have poor agreement between platforms, which may be caused by several factors. The assays use different antibodies, different sample extraction methods to remove high-affinity binding proteins, and different approaches to calibration *(1, 2)*. When 5 immunoassay methods were compared to the former gold standard, the Nichols Advantage assay, the slopes of the regression lines varied between 0.527 and 1.00 *(3)*. Better interlaboratory agreement of IGF-1 concentrations is needed for clinical and antidoping applications *(4)*. Serum IGF-1 concentrations are used for the diagnosis and treatment of growth hormone (GH) disorders. Because each immunoassay platform has different reference ranges for IGF-1, it is difficult to provide specific guidelines for GH disease management. For antidoping applications, serum concentrations of IGF-1 and the N-terminal propeptide of type III procollagen are used in combination to detect GH abuse *(5–9)*. This use of these 2 biomarkers was recently implemented in the 2012 Olympic Games in London and will soon be adopted in all 33 World Anti-Doping Agency (WADA)-accredited laboratories worldwide. In this approach, the results of each assay are algorithmically combined into a single score. Calculated decision limits include the imprecision of each assay. Given the intraindividual variability of IGF-1 *(10–12)* and the goal to collect longitudinal data from individual athletes in mul-

[1] Sports Medicine Research and Testing Laboratory, Salt Lake City, UT; [2] Drug Control Centre, Analytical and Environmental Science Division, King's College London, London, UK; [3] David Geffen School of Medicine at UCLA, Los Angeles, CA; [4] Department of Laboratory Medicine, University of Washington, Seattle, WA; [5] Center for Preventive Doping Research, German Sport University Cologne, Cologne, Germany; [6] United States Anti-Doping Agency, Colorado Springs, CO.
* Address correspondence to this author at: Department of Laboratory Medicine, Box 357110, University of Washington, Seattle, WA 98195. Fax 206-598-6189; e-mail ahoof@u.washington.edu.

Received April 19, 2013; accepted November 14, 2013.
Previously published online at DOI: 10.1373/clinchem.2013.208538

[7] Nonstandard abbreviations: IGF-1, insulin-like growth factor-1; GH, growth hormone; LC-MS/MS, liquid chromatography-tandem mass spectrometry; SRM, selected reaction monitoring; CAP, College of American Pathologists; SISCAPA, Stable Isotope Standards and Capture by Anti-Peptide Antibodies.

tiple laboratories as part of the "athlete biological passport" *(13)*, successful detection of GH misuse necessitates adequate interlaboratory agreement of serum concentrations of IGF-1 and the N-terminal propeptide of type III procollagen.

For small molecules, liquid chromatography-tandem mass spectrometry (LC-MS/MS) has proven to be an accurate and specific method of quantification *(14)*. LC-MS/MS could also circumvent the disadvantages associated with using immunoassays to measure proteins in serum and might allow better interlaboratory agreement *(15)*. Methods have been developed to measure IGF-1 by mass spectrometry. For example, intact IGF-1 has been enriched by use of immunoaffinity purification followed by LC-MS/MS or matrix-assisted laser desorption ionization MS *(16, 17)*. More recently, a method to measure intact IGF-1 by use of high-resolution MS was reported for clinical use *(18, 19)*. It is possible that LC-MS/MS could provide better agreement between IGF-1 concentrations measured in multiple laboratories. To deploy an LC-MS/MS method in multiple laboratories, it is desirable to have a selected reaction monitoring (SRM)-based method that can take advantage of existing instruments already present in most laboratories (typically triple-quadrupole instruments). Because IGF-1 is a low-abundance serum protein, an enrichment or serum depletion method is required to achieve a useful limit of quantification. The sample preparation method must also completely dissociate and remove high-affinity binding proteins that complex with IGF-1 in serum and interfere with quantification *(20)*. Previous methods have used acetonitrile depletion of serum to measure IGF-1 peptides after trypsin digestion *(21, 22)*.

The IGF-1 Working Group, assembled by the Partnership for Clean Competition Research Consortium, deployed a similar method that used acid-acetonitrile protein precipitation, trypsin digestion, and LC-MS/MS for the measurement of serum IGF-1. The goal of the current study was to determine the interlaboratory variability of this method when performed on 5 different instrument platform configurations in 5 different laboratories in 3 different countries. The assay was calibrated by use of external calibration materials, and sources of variation were identified and quantified. We evaluated interlaboratory imprecision by use of healthy human serum samples and samples from patients with acromegaly, which contain high concentrations of IGF-1.

Materials and Methods

HUMAN SAMPLES

We obtained frozen aliquots of healthy human serum samples directly from Solomon Park Research Laboratory (Kirkland, WA). Serum samples from patients with acromegaly were obtained under a research ethics committee–approved protocol from Oxford University Hospital (Oxford, UK), and frozen aliquots were sent to each laboratory.

Table 1. Interlaboratory comparison of a commercial immunoassay for IGF-1.

Samples	Slope[a]	Intercept[a]	SD[b]	$S_{y\|x}$[c]	r^2
Healthy	0.60	18.3	87.8	20.2	0.66
Acromegaly	1.00	−48.3	139.4	19.7	0.86
All	0.98	−67.0	161.7	24.8	0.88

[a] Deming regression of IGF-1 measurements by IMMULITE 1000 in 2 different laboratories.
[b] SD of results from the population is presented for comparison with $S_{y|x}$.
[c] SD of the residuals calculated as the perpendicular to the regression line.

IMMUNOASSAY

We measured IGF-1 by use of the Siemens IMMULITE 1000. The assay is a solid-phase, enzyme-linked chemiluminescent immunometric assay that uses a monoclonal anti–IGF-1 antibody linked to beads and an enzyme-linked polyclonal anti–IGF-1 antibody in solution. The immunoassay of both healthy and acromegaly serum samples was performed according to the kit manufacturer's instructions in 2 laboratories with different kit lot numbers. One laboratory used reagent kit lot 0328 and low and high adjuster lots LGFL0128 and LGFH0128, and the other laboratory used reagent kit lots 0326 and 0328 and lots LGFL0128 and LGFH0127 for the low and high adjusters, respectively.

LC-MS/MS ASSAY

A detailed standard operating procedure is included in Supplemental Data, which accompanies the online version of this article at http://www.clinchem.org/content/vol60/issue3. Briefly, 100 μL serum was incubated in 200 μL of 1% acetic acid in the presence of 2 μg/mL IGF-2 and 200 ng/mL isotope-labeled (^{15}N) IGF-1 internal standard. Proteins were then precipitated with acetonitrile. The supernatant was dried by use of centrifugal evaporation, reconstituted, reduced with dithiothreitol, alkylated with iodoacetamide, and digested overnight with trypsin.

Each laboratory used different LC-MS/MS instrumentation, which included 4 triple-quadrupole instruments and 1 quadrupole time-of-flight instrument. Online Supplemental Table 1 provides complete details of the instrumentation used and relevant instrument settings. Two peptides corresponding to amino acids 1–21 (T1) and 22–36 (T2) of IGF-1 were monitored. Beyond specification of the peptides, all other

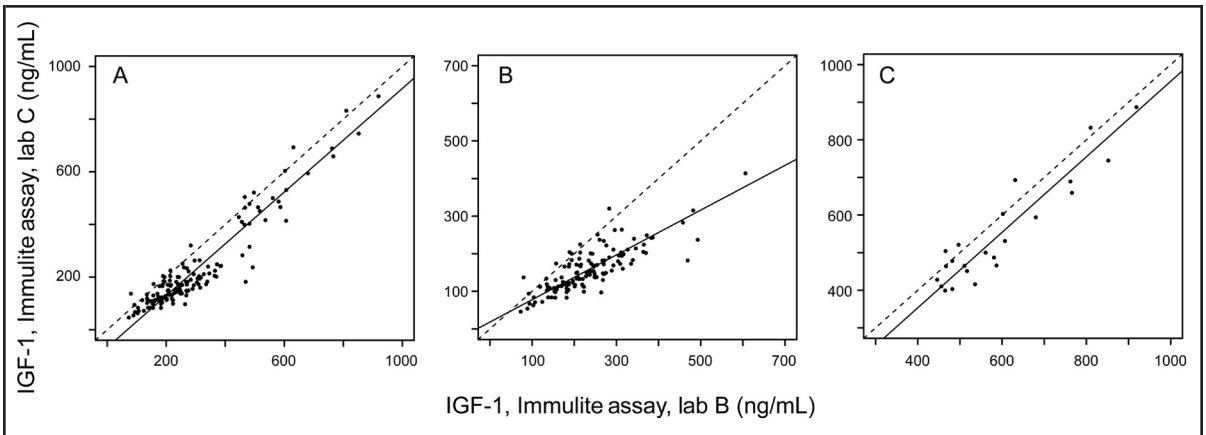

Fig. 1. Comparison of immunoassay results. Deming regression of IGF-1 concentrations measured by the same immunoassay in 2 different laboratories for all samples (A), samples from healthy individuals (B), and samples from acromegaly patients (C).

instrumentation parameters, including peptide charge state and product ions selected for SRM transitions, were optimized independently for the instrument in each laboratory. Matrix interference, ion suppression, and coeluting interfering peaks often depend on the HPLC conditions and the SRM transitions selected for measurement; the laboratories were responsible for evaluating chromatography and absence of interferents before participation in the study.

DATA COLLECTION AND ANALYSIS

We measured the chromatographic peak areas for quantifier ion transitions for each IGF-1 and [^{15}N]-IGF-1 peptide and calculated the peak area ratios. We determined the concentration of IGF-1 by use of the peak area ratio and the equation of the line from the nonweighted linear regression of the calibration curve ($r^2 > 0.95$). The IGF-1 concentration was determined for each peptide, and mean of the 2 concentrations was calculated.

For the interlaboratory comparison of 152 samples (130 healthy, 22 acromegaly), we divided the samples into 8 batches. Calibration curve samples were processed with each batch and used to determine the concentration of IGF-1. For recalibration by use of a single human calibrator, we used the sample with the lowest between-site variability in each batch to adjust the measured concentration at each site. The intersite variability was then recalculated as the CV for each sample and means calculated separately for the healthy and acromegaly samples. The data from each of the single point calibrators was removed from the calculation of between-site variability. We compared the serum IGF-1 concentrations measured by LC-MS/MS in 5 laboratories by use of both calibration methods or as measured by immunoassay by Deming regression analysis using the R statistics language (http://www.r-project.org) and the MethComp package. $S_{y|x}$ was calculated as the standard deviation of perpendicular distance of each point to the Deming regression curve.

Results

INTERLABORATORY COMPARISON OF AN IMMUNOASSAY METHOD

To provide a benchmark against which we could compare LC-MS/MS assays to 1 another, we first compared the same automated immunoassay, the Siemens IMMULITE 1000, run in 2 different laboratories each with a different kit lot number. When IGF-1 was measured in the serum samples from healthy individuals and from patients with acromegaly with high concentrations of IGF-1 (Fig. 1), the Deming regression for the healthy samples had a slope of 0.60 with a Pearson correlation coefficient (r^2) of 0.66 and $S_{y|x}$ of 20.2. For the acromegaly samples, the slope was better (1.0), with a Pearson correlation coefficient (r^2) of 0.86 and $S_{y|x}$ of 19.7.

VARIATION DUE TO INSTRUMENTATION

It was unclear how much variability would be due to the use of different instrument platforms in each laboratory. Within the IGF-1 Working Group, 5 different HPLC systems were used, including 1 ultrahigh-pressure HPLC system and 1 nano-ultrahigh-pressure HPLC system. In addition, mass spectrometers included 1 quadrupole-time-of-flight and 4 triple-quadrupole instruments. To evaluate variability due to instrumentation, 4 human serum samples were processed according to the established protocol in a single laboratory and aliquots of the 4 prepared samples were

Table 2. IGF-1 measurement imprecision due to LC-MS/MS instrumentation.

Sample	T1 peptide, CV, %[a]	T2 peptide, CV, %[a]	Peptide, mean CV, %[b]
1	5.7	11.9	6.6
2	4.6	15.4	5.9
3	11.2	8.7	6.0
4	3.9	8.4	4.0
Mean CV, %[c]	6.4	11.1	5.6

[a] Protein concentration was determined using only 1 peptide, and the imprecision was calculated for each sample across 4 laboratories.
[b] Protein concentration was determined for each peptide and then averaged, and the imprecision for each sample was calculated across the laboratories.
[c] Average CV was calculated from the CV of all 4 samples.

Table 3. Comparison of a central calibrator vs a local calibrator.

Sample	Mean concentration, ng/mL	Central calibrator, CV, %[a]	Local calibrator, CV, %[b]
1	85	4.0	16.4
2	90	5.6	19.3
3	210	6.1	10.6
4	356	6.9	5.5
5	179	3.5	11.8
Mean CV, %[c]		5.2	12.8

[a] One laboratory generated a single calibration curve and distributed it.
[b] Each laboratory generated its own calibration curve using rat plasma and reference material.
[c] Average CV was calculated from the CV of all 4 samples.

sent for measurement in 3 additional laboratories. IGF-1 concentrations were determined by use of the T1 and T2 peptides. Imprecision was improved when the mean of the measured concentrations of the 2 peptides was used to determine the protein concentration (Table 2). Despite the variety of instrumentation used for IGF-1 measurement, the mean imprecision between 4 laboratories was <6%.

VARIATION DUE TO THE CALIBRATION CURVE

IGF-1 concentrations were determined by use of a calibration curve made from rat serum fortified with the First WHO International Standard for human recombinant IGF-1 (02/254). To determine the influence of external calibration material on interlaboratory precision, we evaluated the variation due to formulation of the calibration curve in each laboratory (local calibrator) compared to variation when the calibration curve was made in a single laboratory (central calibrator). Five healthy serum samples were measured in each laboratory, and the concentration of IGF-1 was determined by use of both local and central calibration curves. The mean imprecision of the 5 serum samples was 5.2% when a central calibration curve was used and 12.8% when a local calibration curve was used (Table 3).

VARIATION IN MAKING THE CALIBRATION CURVES

The greater imprecision in the measurement observed for the local calibration curves vs the central calibration curve could result from the variability of the amount of material in each vial, the variability of the reconstitution step (which includes swirling to dissolve as much of the lyophilized material in the vial as possible), and the variability of pipetting to make the final calibration material. Reverse pipetting was used at each site to minimize this effect; however, future studies should examine gravimetric approaches as well. Each vial of the WHO international standard from the National Institutes of Biological Standards and Controls contains 8.5 μg IGF-1, with an expanded uncertainty (95%) range of 7.73–9.23 μg ($\pm$8.8%, CV 4.5%), as reported by the manufacturer. To determine the variation due to the reconstitution and pipetting steps, a single laboratory (laboratory E) estimated the within-vial and between-vial variation by measuring the concentration of IGF-1 in 7 human serum samples determined by use of calibration curves made from 3 different vials of IGF-1 reference material in each of 3 separate batches. The mean intravial variation (which included variability due to preparation of the calibration curve, sample extraction, and LC-MS/MS analysis) across 3 days was 7.3%, whereas the mean total intervial variation (which included each of the contributions to the intravial variation and the variability in the amount of material in each vial at the time of manufacturing and the variability in the reconstitution step) was 10.3%. This was slightly higher than the predicted intervial variation calculated by use of the sum of squares of the measured intravial variation (7.3%) and the reported intervial variation (4.5%) of 8.6%, suggesting that reconstitution and pipetting steps increased the variation of the assay by approximately 5.7%, and that reference material provided in a liquid form might help reduce this variation.

USE OF SINGLE-POINT CALIBRATOR

Given the observed variation in the interlaboratory measurements that might be contributed by the reconstitution of the reference material and the making of the calibration curves, we evaluated the use of a single-point human calibrator. The IGF-1 concentrations in 6 healthy human serum samples were determined in 3 laboratories by use of 1 healthy human serum sample

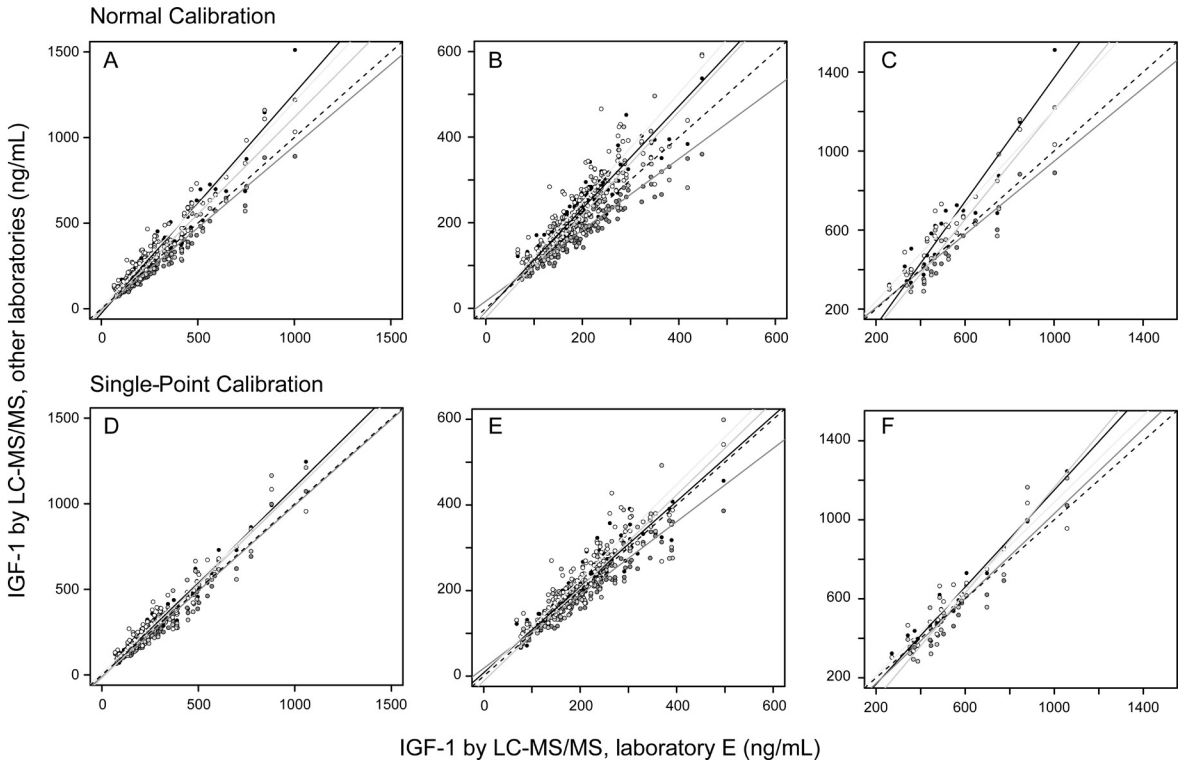

Fig. 2. Comparison of LC-MS/MS results. Deming regression of IGF-1 concentrations measured by LC-MS/MS in each of 4 laboratories onto concentrations measured by LC-MS/MS in laboratory E.

The data obtained by use of local calibration curves are presented for all samples (A), samples from healthy individuals (B), and samples from acromegaly patients (C). Also shown are the data by use of a single-point calibration approach for all samples (D), samples from healthy individuals (E), and samples from acromegaly patients (F).

as a calibrator or by use of a central calibration curve. The mean imprecision for the 6 human samples by use of the single-point human calibrator was 4.5%, whereas the imprecision by use of the central calibration curve was 7.7%. The results suggested that use of a single-point human calibrator might reduce the interlaboratory variability associated with the use of spiked-matrix external calibration curves.

INTERLABORATORY COMPARISON OF THE LC-MS/MS METHOD

After identifying the sources of potential variation between the laboratories, the IGF-1 concentration of 152 samples was measured in all 5 laboratories in 8 batches over several weeks. The samples measured were the same as those measured by immunoassay. The IGF-1 concentrations were determined by use of locally produced 5-point calibration curves. The laboratory with results closest to the mean of all laboratories, laboratory E, was selected as the reference laboratory shown on the *x* axis, and the data for the remaining 4 laboratories was plotted on the *y* axis (Fig. 2). The equations of the Deming regression analyses are listed in online Supplemental Table 2. The corresponding Bland–Altman analysis is shown in online Supplemental Fig. 1 and online Supplemental Table 3. For the healthy samples, the Deming slopes ranged between 0.83 and 1.27, the intercepts between −21.39 and 16.96, and the Pearson correlation coefficients between 0.782 and 0.868. The mean CV for all of the healthy samples compared between the 5 laboratories was 16.3%.

Growing adolescents and athletes misusing GH can have serum IGF-1 concentrations above the range typically observed in healthy adults. It was therefore important to test the interlaboratory agreement of the method on human samples containing high concentrations of IGF-1. Serum samples were obtained from patients with acromegaly, in which the measurement of IGF-1 is important for disease diagnosis and disease management following treatment with pegvisomant, a GH receptor antagonist. Deming regression analysis of the results from the 5 laboratories is shown in Fig. 2C and online Supplemental Table 2. The Dem-

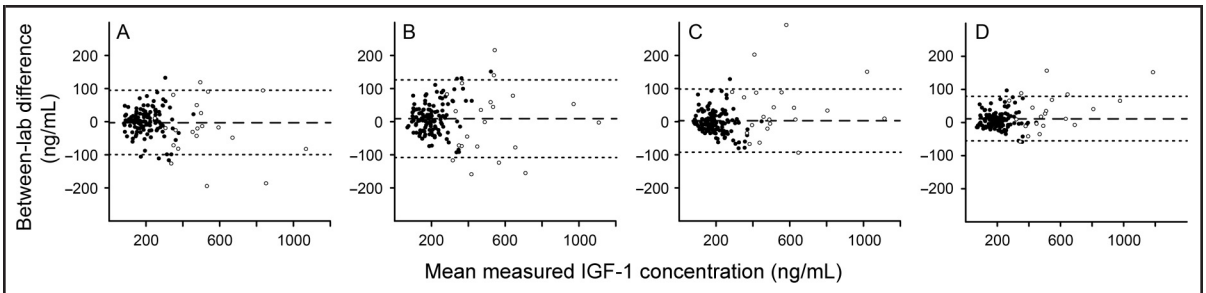

Fig. 3. Bland–Altman plots of interlaboratory comparisons.

Data from each of the other 4 laboratories (A–D) are plotted as the difference from laboratory E vs the mean concentration observed for the 2 laboratories. For each comparison, healthy individuals are shown with ●, acromegaly patients are shown with ○, mean differences are shown by a dashed line, and 2-SD ranges are shown by dotted lines.

ing regression analysis of the acromegaly samples was generally similar to that for the healthy patients, with Pearson correlation coefficients (r^2) between 0.783 and 0.884. The mean CV for the acromegaly samples was 14.5%.

SINGLE-POINT CALIBRATION APPROACH

We then tested whether interlaboratory precision would be improved with the use of a single-point human calibrator approach. Deming regression analysis is shown in Fig. 2, and the regression data are listed in online Supplemental Table 2. The corresponding Bland–Altman analysis is shown in Fig. 3 and online Supplemental Table 3. The mean CV of the healthy samples decreased from 16.3% to 11.3%, and the mean CV for the acromegaly samples decreased from 14.5% to 9.6%. For both populations combined, the mean CV decreased from 16.0% to 11.1%. Comparison of the Bland–Altman analysis showed a decrease in the bias for all of the laboratories (see online Supplemental Table 3), supporting the hypothesis that native human calibrators can help reduce variability in quantitative clinical proteomics assays that use trypsin digestion.

Discussion

The goal of the IGF-1 Working Group assembled by the Partnership for Clean Competition Research Consortium was to evaluate an LC-MS/MS-based assay for serum IGF-1 for use in antidoping laboratories and clinical laboratories around the world. Part of the motivation behind the project was the known variability between immunoassays and even between the same immunoassay performed at multiple sites *(23–28)*, including for IGF-1 *(3)*. Those reports were supported by the data available from the College of American Pathologists (CAP) proficiency testing program for IGF-1 (BGS survey), which from September 2011 to March 2013 demonstrated interlaboratory variability up to 33.5% CV. This level of variability is higher than the goal of 15% total allowable error suggested by the Royal College of Pathologists of Australasia and can make programmatic and clinical use of results across laboratories challenging. The mean interlaboratory imprecision observed in this study (11.1%–16.0%) could be advantageous in this respect.

In antidoping laboratories, the improved imprecision across laboratories has profound implications. On the basis of CAP proficiency testing, our results show that we should be able to reduce the standard deviation between antidoping laboratories by at least half (i.e., 33.5% to <16.0%). Assuming that the population distribution of IGF-1 results is described by the square root of the sum of squares of interindividual variability (44% *(10)*) and interlaboratory variability, an IGF-1 result that was previously 3.72 SD above the mean of a reference population (likely to happen by chance in <1 in 10 000 cases), would effectively now be more than 4.39 SD above the mean (likely to happen by chance in <1 in 100 000 cases). By the same logic, because the GH-2000 scores will be calculated by use of results from laboratories around the world, improved interlaboratory imprecision will effectively reduce apparent intraindividual variability and make the athlete biological passport a much more sensitive and specific method for detecting doping in sports *(13)*. Finally, the deployment of an MS assay across antidoping laboratories will obviate the difficult process of transforming decision thresholds, which is necessary when using new immunoassays (the result of manufacturer-discontinued immunoassays) or manufacturer-modified immunoassays *(1, 6, 7)*.

Another motivating factor for the use of LC-MS/MS in the measurement of IGF-1 is the specificity of the approach. The ability of the liquid chromato-

graph to resolve peptides on the basis of size and hydrophobicity, and the ability of the mass spectrometer to accurately isolate and detect precursor and product ions with specific mass-to-charge ratios, allows for 3 independent dimensions of specificity *(29, 30)*. The use of multiple peptides from the same protein increases specificity further, verified when the protein concentrations inferred from >1 directly detected peptide are similar. This is in contrast to immunoassays, in which the concentration of the protein is determined indirectly by use of signal generated from labeled antibodies—a signal that can be generated in the absence of analyte *(15)*.

Our results extend the findings of other groups that have demonstrated the potential for good interlaboratory imprecision when measuring peptides in plasma by use of LC-MS/MS methods. For example, when unfractionated plasma was digested with trypsin and supplemented with predigested, purified proteins and stable isotope–labeled peptide internal standards, the interlaboratory imprecision was <32% *(31, 32)*. More recently, interlaboratory imprecision was evaluated for the SISCAPA (Stable Isotope Standards and Capture by Anti-Peptide Antibodies) method of protein quantification *(33)*. When unfractionated plasma was supplemented with protein standards, digested, and quantified by use of the SISCAPA method, the imprecision was <14%. The work demonstrated that the use of a stable isotope–labeled protein internal standard rather than peptide standards significantly improved the imprecision of the method. Neither of these interlaboratory comparisons was designed to assess the interlaboratory variability of the measurement of endogenous protein concentration in native human samples. Our experiments, which aimed to measure the concentration of native IGF-1 in human serum, demonstrate that calibration of the peak area ratios, which are variable between laboratories *(31, 32)*, can reduce interlaboratory variability, and further support the concept of using LC-MS/MS measurements of serum proteins for the international standardization of clinical protein assays *(34–36)*.

In this work, we tried to identify some of the sources of interlaboratory variability in a method to quantify IGF-1 peptides in serum. Variation from the sample preparation steps was minimized through the use of [^{15}N]-IGF-1 protein, which has been demonstrated previously to improve interlaboratory agreement *(33)*. The use of different instrumentation contributed to a low level of variability, while reconstitution of the reference material and method of calibration were identified as the primary source of interlaboratory variation. Most striking was the improvement in agreement observed upon the use of a single-point human calibrator rather than a calibration curve constructed in each laboratory.

The use of a limited number of calibration adjusters is not unusual in the clinical laboratory; many immunoassays use this approach, which assumes a constant shape of the nonlinear calibration curve. Additionally, the use of native human serum to calibrate assays is also common, as evidenced by standard reference materials produced by NIST *(37)*. Although not commonly used in laboratory developed tests, the use of a single-point native human serum calibrator was recently demonstrated to reduce assay variability in the trypsin-digestion LC-MS/MS measurement of serum apolipoprotein A-I and apolipoprotein B *(38)*. There are at least 2 reasons that traceable native serum calibrators could be beneficial in LC-MS/MS protein assays. First, they would avoid the measurable variability associated with the reconstitution of purified reference material that we observed here. Second, the calibrator protein is present in its native state, with naturally occurring posttranslational modifications and interactions with other serum proteins, lipids, and other small molecules. Indeed, calibration with native serum was superior to calibration with spiked purified apolipoproteins in an animal serum matrix *(38)*. A consensus statement from a meeting of experts in the field of GH and IGF-1 research has recently suggested distribution of such material, as QC samples, for improved concordance of IGF-1 immunoassays *(39)*.

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: D. Cowan, King's College London; L.D. Bowers, US Anti-Doping Agency; A.N. Hoofnagle, *Clinical Chemistry*, AACC.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: This study was supported by the Partnership for Clean Competition, US Anti-Doping Agency.
Expert Testimony: L.D. Bowers, PGA Tour Arbitration.
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, or preparation or approval of manuscript.

Acknowledgments: The authors thank Dr. Nishan Guha for providing the acromegaly samples and the US Anti-Doping Agency for providing the ^{15}N-labeled IGF-1 internal standard.

References

1. Cowan DA, Bartlett C. Laboratory issues in the implementation of the marker method. Growth Horm IGF Res 2009;19:357–60.
2. Frystyk J, Freda P, Clemmons DR. The current status of IGF-I assays: a 2009 update. Growth Horm IGF Res 2010;20:8–18.
3. Krebs A, Wallaschofski H, Spilcke-Liss E, Kohlmann T, Brabant G, Volzke H, Nauck M. Five commercially available insulin-like growth factor I (IGF-I) assays in comparison to the former nichols advantage IGF-I in a growth hormone treated population. Clin Chem Lab Med 2008;46:1776–83.
4. Clemmons DR. Clinical laboratory indices in the treatment of acromegaly. Clin Chim Acta 2011; 412:403–9.
5. Erotokritou-Mulligan I, Guha N, Stow M, Bassett EE, Bartlett C, Cowan DA, et al. The development of decision limits for the implementation of the GH-2000 detection methodology using current commercial insulin-like growth factor-I and amino-terminal pro-peptide of type III collagen assays. Growth Horm IGF Res 2012;22:53–8.
6. Guha N, Erotokritou-Mulligan I, Bartlett C, Cowan DA, Bassett EE, Stow M, et al. The effects of a freeze-thaw cycle and pre-analytical storage temperature on the stability of insulin-like growth factor-I and pro-collagen type III N-terminal propeptide concentrations: implications for the detection of growth hormone misuse in athletes. Drug Test Anal 2012;4:455–9.
7. Guha N, Erotokritou-Mulligan I, Burford C, Strobridge G, Brigg J, Drake T, et al. Serum insulin-like growth factor-I and pro-collagen type III N-terminal peptide in adolescent elite athletes: Implications for the detection of growth hormone abuse in sport. J Clin Endocrinol Metab 2010;95: 2969–76.
8. Holt RI, Erotokritou-Mulligan I, McHugh C, Bassett EE, Bartlett C, Fityan A, et al. The GH-2004 project: The response of IGF1 and type III procollagen to the administration of exogenous GH in non-caucasian amateur athletes. Eur J Endocrinol 2010;163:45–54.
9. Powrie JK, Bassett EE, Rosen T, Jorgensen JO, Napoli R, Sacca L, et al. Detection of growth hormone abuse in sport. Growth Horm IGF Res 2007;17:220–6.
10. Erotokritou-Mulligan I, Eryl Bassett E, Cowan DA, Bartlett C, Milward P, Sartorio A, et al. The use of growth hormone (GH)-dependent markers in the detection of GH abuse in sport: physiological intra-individual variation of IGF-I, type 3 procollagen (P-III-P) and the GH-2000 detection score. Clin Endocrinol 2010;72:520–6.
11. Kniess A, Ziegler E, Thieme D, Muller RK. Intra-individual variation of GH-dependent markers in athletes: comparison of population based and individual thresholds for detection of GH abuse in sports. J Pharm Biomed Anal 2013;84:201–8.
12. Velloso CP, Aperghis M, Godfrey R, Blazevich AJ, Bartlett C, Cowan D, et al. The effects of two weeks of recombinant growth hormone administration on the response of IGF-I and N-terminal pro-peptide of collagen type III (P-III-NP) during a single bout of high resistance exercise in resistance trained young men. Growth Horm IGF Res 2013;23:76–80.
13. Sottas PE, Vernec A. Current implementation and future of the athlete biological passport. Bioanalysis 2012;4:1645–52.
14. Grant RP. High throughput automated LC-MS/MS analysis of endogenous small molecule biomarkers. Clin Lab Med 2011;31:429–41.
15. Hoofnagle AN, Wener MH. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. J Immunol Methods 2009;347:3–11.
16. Bredehoft M, Schanzer W, Thevis M. Quantification of human insulin-like growth factor-1 and qualitative detection of its analogues in plasma using liquid chromatography/electrospray ionisation tandem mass spectrometry. Rapid Commun Mass Spectrom 2008;22:477–85.
17. Nelson RW, Nedelkov D, Tubbs KA, Kiernan UA. Quantitative mass spectrometric immunoassay of insulin like growth factor 1. J Proteome Res 2004; 3:851–5.
18. Bystrom CE, Sheng S, Clarke NJ. Narrow mass extraction of time-of-flight data for quantitative analysis of proteins: determination of insulin-like growth factor-1. Anal Chem 2011;83:9005–10.
19. Bystrom C, Sheng S, Zhang K, Caulfield M, Clarke NJ, Reitz R. Clinical utility of insulin-like growth factor 1 and 2; determination by high resolution mass spectrometry. PLoS One 2012;7:e43457.
20. Clemmons DR. IGF-I assays: current assay methodologies and their limitations. Pituitary 2007;10: 121–8.
21. Barton C, Kay RG, Gentzer W, Vitzthum F, Pleasance S. Development of high-throughput chemical extraction techniques and quantitative HPLC-MS/MS (SRM) assays for clinically relevant plasma proteins. J Proteome Res 2010;9:333–40.
22. Kay RG, Barton C, Velloso CP, Brown PR, Bartlett C, Blazevich AJ, et al. High-throughput ultra-high-performance liquid chromatography/tandem mass spectrometry quantitation of insulin-like growth factor-I and leucine-rich alpha-2-glycoprotein in serum as biomarkers of recombinant human growth hormone administration. Rapid Commun Mass Spectrom 2009;23:3173–82.
23. La'ulu SL, Roberts WL. Performance characteristics of five automated CA 19–9 assays. Am J Clin Pathol 2007;127:436–40.
24. Manley SE, Stratton IM, Clark PM, Luzio SD. Comparison of 11 human insulin assays: implications for clinical investigation and research. Clin Chem 2007;53:922–32.
25. Mongia SK, Rawlins ML, Owen WE, Roberts WL. Performance characteristics of seven automated CA 125 assays. Am J Clin Pathol 2006;125:921–7.
26. Rawlins ML, Roberts WL. Performance characteristics of six third-generation assays for thyroid-stimulating hormone. Clin Chem 2004;50:2338–44.
27. Schlumberger M, Hitzel A, Toubert ME, Corone C, Troalen F, Schlageter MH, et al. Comparison of seven serum thyroglobulin assays in the follow-up of papillary and follicular thyroid cancer patients. J Clin Endocrinol Metab 2007;92:2487–95.
28. Slev PR, Rawlins ML, Roberts WL. Performance characteristics of seven automated CA 15-3 assays. Am J Clin Pathol 2006;125:752–7.
29. Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 2012;9:555–66.
30. Strathmann FG, Hoofnagle AN. Current and future applications of mass spectrometry to the clinical laboratory. Am J Clin Pathol 2011;136: 609–16.
31. Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol 2009;27:633–41.
32. Hoofnagle AN. Quantitative clinical proteomics by liquid chromatography-tandem mass spectrometry: assessing the platform. Clin Chem 2010;56:161–4.
33. Kuhn E, Whiteaker JR, Mani DR, Jackson AM, Zhao L, Pope ME, et al. Interlaboratory evaluation of automated, multiplexed peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. Mol Cell Proteomics 2012;11: M111 013854.
34. Kaiser P, Akerboom T, Ohlendorf R, Reinauer H. Liquid chromatography-isotope dilution-mass spectrometry as a new basis for the reference measurement procedure for hemoglobin A1c determination. Clin Chem 2010;56:750–4.
35. Miller WG, Thienpont LM, Van Uytfanghe K, Clark PM, Lindstedt P, Nilsson G, Steffes MW. Toward standardization of insulin immunoassays. Clin Chem 2009;55:1011–8.
36. Seegmiller JC, Sviridov D, Larson TS, Borland TM, Hortin GL, Lieske JC. Comparison of urinary albumin quantification by immunoturbidimetry, competitive immunoassay, and protein-cleavage liquid chromatography-tandem mass spectrometry. Clin Chem 2009;55:1991–4.
37. Phinney KW, Bedner M, Tai SS, Vamathevan VV, Sander LC, Sharpless KE, et al. Development and certification of a standard reference material for vitamin D metabolites in human serum. Anal Chem 2012;84:956–62.
38. Agger SA, Marney LC, Hoofnagle AN. Simultaneous quantification of apolipoprotein A-I and apolipoprotein B by liquid-chromatography-multiple-reaction-monitoring mass spectrometry. Clin Chem 2010;56:1804–13.
39. Clemmons DR. Consensus statement on the standardization and evaluation of growth hormone and insulin-like growth factor assays. Clin Chem 2011;57:555–9.