

Gaps in the Traceability Chain of Human Growth Hormone Measurements

Sébastien Boulo,¹ Katja Hanisch,¹ Martin Bidlingmaier,² Cristian-Gabriel Arsene,³ Mauro Panteghini,⁴ Guy Auclair,¹ Catharine Sturgeon,⁵ Heinz Schimmel,¹ and Ingrid Zegers^{1*}

BACKGROUND: Human growth hormone (hGH) is measured for the diagnosis of secretion disorders. These measurements fall under the EU Directive 98/79/EC on in vitro diagnostic medical devices requiring traceability of commercial calibrator values to higher-order reference materials or procedures (*Off J Eur Communities* 1998 Dec 7;L 331:1–37). External quality assessment schemes show large discrepancies between results from different methods, even though most methods provide results traceable to the recommended International Standard (IS 98/574). The aim of this study was to investigate possible causes for these discrepancies.

METHODS: We investigated the commutability and recovery of hGH in reconstituted IS 98/574. We tested different reconstitution protocols and used 4 different serum matrices for spiking. These IS preparations were measured together with serum samples. We quantified hGH by 5 different methods in 4 different laboratories.

RESULTS: Results from the different methods correlated well for the serum samples. Mean discrepancies between results from different methods were $\leq 20\%$. None of the IS preparations was commutable for all the method comparisons. The recovery of hGH in preparations of IS 98/574 depended on the reconstitution protocol (>10 -fold differences) and background matrix (relative differences $\leq 17\%$ for different serum matrices).

CONCLUSIONS: The use of different protocols for reconstitution and spiking of hGH reference preparations affects quantification by immunoassays, potentially leading to a bias between commercial methods, despite

the use of calibrators with values claimed to be traceable to the same higher-order reference material.

© 2013 American Association for Clinical Chemistry

Human growth hormone (hGH)⁶ is a protein secreted by the pituitary gland. It regulates the increase in height during childhood and has other anabolic effects. An excess of hGH leads to acromegaly, whereas a deficiency causes growth failure and short stature. The biochemical diagnosis of GH deficiency is based on peak GH concentrations obtained during different stimulation tests. A nadir value after oral glucose load ($<1 \mu\text{g/L}$) is commonly used as decision limit to diagnose acromegaly, although this decision limit has been criticized for being too high (1–3).

Circulating hGH shows large heterogeneity due to alternative splicing, different posttranslational modifications, oligomerization, complex formation, and proteolytic processing. The 22-kDa hGH form represents $\leq 90\%$ of total circulating hGH (4, 5). The 20-kDa hGH (approximately 5%–10% of total circulating hGH) lacks residues 32–46 of the 22-kDa hGH form. These 2 forms are co-secreted in a fixed ratio (6–8). Dimers represent $\leq 10\%$ of circulating hGH. Fragments of hGH with 16 and 17 kDa and other forms are also present (approximately 2%) (9).

In vitro diagnostic (IVD) measurements of hGH are within the scope of EU Directive 98/79/EC on IVDs, which requires traceability of values assigned to commercial calibrators to higher-order reference materials or methods (10). Calibration of hGH immunoassays in terms of International Standard (IS) 98/574 has been recommended by the International Collaborative on Standardisation of Growth Hormone methods

¹ Institute for Reference Materials and Measurements (IRMM), Joint Research Centre, European Commission, Geel, Belgium; ² Endocrine Laboratory, Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Munich, Germany; ³ Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany; ⁴ Centre for Metrological Traceability in Laboratory Medicine, University of Milan, Milan, Italy; ⁵ UK National External Quality Assessment Service (NEQAS) Edinburgh, Department of Laboratory Medicine, Royal Infirmary, Edinburgh, UK.

* Address correspondence to this author at: European Commission, Joint Research Centre, IRMM, Retieseweg 111, 2440 Geel, Belgium. Fax +32-14-571-548;

e-mail ingrid.zegers@ec.europa.eu.

Received November 19, 2012; accepted March 7, 2013.

Previously published online at DOI: 10.1373/clinchem.2012.199489

⁶ Nonstandard abbreviations: hGH, human growth hormone; IVD, in vitro diagnostic; IS, International Standard; IRP, International Reference Preparation; EQA, external quality assessment; SRM, Standard Reference Material; ASTM, American Society for Testing and Materials; NEQAS, National External Quality Assessment Service; IRMM, Institute for Reference Materials and Measurements; GHBP, GH binding protein.

(7, 8, 11). IS 98/574 has been prepared from recombinant 22-kDa growth hormone (somatropin) and has an assigned value of 1.95 mg somatropin per ampoule. This hGH value was obtained by comparison of the 22-kDa form of hGH in IS 98/574 with that found in International Reference Preparation (IRP) 88/624 (12) by size-exclusion liquid chromatography. The concentration of hGH in a solution prepared with the content of 1 ampoule of IS 98/574 reconstituted in 2 g water was measured independently by mass spectrometry and found to be 1.09 (0.11) mg/g with 95% confidence ($\kappa = 2$), or a total amount of about 2.18 mg per ampoule (13).

Different immunoassays are available for IVD measurements of hGH. In most cases, the kit inserts state that the method gives values traceable to IS 98/574. However, data from external quality assessment (EQA) schemes show large differences between measurement results obtained with different IVDs, suggesting that other factors also contribute (14). The aim of this study was to investigate possible causes for these discrepancies, and in particular the influence of using different protocols when preparing matrix-based calibrators from hGH IS 98/574. We compared the immunoreactivity of preparations made from IS 98/574 to that of patient samples. This comparability, characterizing the so-called commutability (15), provides an indication of the potential degree of intermethod harmonization achievable when a given preparation is used for calibration of commercial procedures.

Materials and Methods

hGH

The hGH material used was the second WHO IS 98/574, containing the lyophilized 22-kDa form of hGH (somatropin) plus excipients (16).

SAMPLES

The serum matrices used in this study were hGH-depleted serum (ref. SF220-2, SCIPAC), human charcoal-stripped serum (ref. IPLA-SER7, Dunn Labortechnik), NIST Standard Reference Material (SRM) 971 male serum, and sheep serum (ref. S2263, Sigma-Aldrich). The density of each of these sera and of the different buffers used in this study were all measured by the American Society for Testing and Materials (ASTM) D 4052 method at 20 °C (17).

We selected serum samples from a large number of sets of 0.5-mL samples, stored at -70 °C. These sera were obtained from blood donations from healthy people (18). Selected serum samples had hGH concentrations between 0.1 and 12 $\mu\text{g/L}$ (mean values of results from the 5 immunoassays used).

PREPARATION OF MATERIALS SPIKED WITH IS 98/574

We designed the protocols for reconstitution, dilution, and spiking while keeping in mind practices used by IVD manufacturers to produce calibrators and the UK National External Quality Assessment Service (NEQAS) for preparing samples for distribution in EQA schemes. We performed reconstitution in different diluents to assess the importance of the buffer and the protein background separately. The final matrices used for the preparation of calibrators reflect the choices available for materials that are free (or almost free) of human GH but have a matrix comparable to that of human serum. Sheep serum is used to avoid interference by (human) GH binding protein (GHBP).

For all studies, we gave randomly assigned numbers to serum lots from individual patients and spiked materials. Serum samples were thawed and combined in sets with spiked samples on the day of shipment. Serum samples and spiked samples were filled in identical vials. In combination with the use of randomized numbers, this protocol ensured that all samples were treated equally during shipment and by the participating laboratories. Samples transport took <24 h and included cooling elements to ensure that the temperature remained at 2 °C–8 °C. Samples were stored at 2 °C–8 °C at each of the laboratories. hGH in serum has been shown to be stable at 4 °C for 8 days; therefore we requested that measurements be performed within 5 days of receipt of the samples.

Commutability study. For the commutability study, we made 13 preparations volumetrically to achieve a hGH concentration, after spiking, of approximately 7 $\mu\text{g/L}$ (Fig. 1). IS 98/574 was first reconstituted with MilliQ water, PBS (137 mmol/L NaCl, 2.7 mmol/L KCl, and 10 mmol/L phosphate buffer solution at pH 7.4), or PBS containing 5% BSA (PBS-BSA). After reconstitution, we made several serial dilutions of the respective solutions (3 times by a factor of 100), before the final spiking of the hGH solution into the different sera at 7 $\mu\text{g/L}$ (Fig. 1). We divided these preparations into 0.5-mL aliquots. All preparations were performed on ice.

Recovery study. For the recovery study, we made 10 preparations gravimetrically. IS 98/574 was reconstituted with PBS or PBS-BSA. After reconstitution, we performed 2 serial dilutions (1 by a factor of 250 and 1 by a factor of 40), spiked hGH into the different serum matrices at 5 $\mu\text{g/L}$, and then divided the samples into aliquots. To accurately determine the hGH mass fraction of each preparation, we measured the density of sera and reconstitution buffers. The results were 1.0052, 1.0174, 1.0214, 1.0196, 1.0249, and 1.0237 kg/L , respectively, for PBS, PBS-BSA, hGH-depleted serum, human charcoal-stripped serum, SRM 971

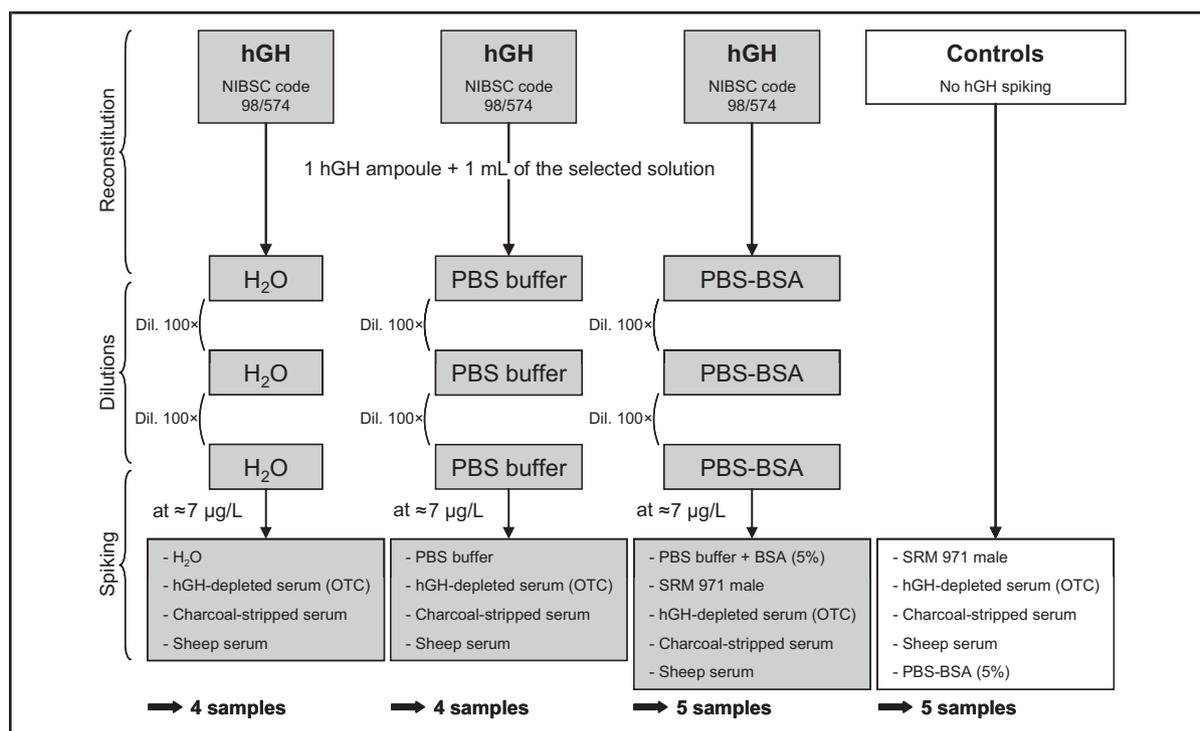


Fig. 1. Overview of the volumetric preparations.

Each hGH ampoule (WHO standard IS 98/574) was reconstituted in water, PBS, or PBS-BSA. Each reconstitution was then diluted into the solution used for reconstitution and subsequently spiked at about 7 µg/L into the different matrices and buffers (hGH-depleted serum, human charcoal-stripped serum, NIST SRM 971 male serum, sheep serum, PBS, and PBS-BSA). Five control matrices were also tested.

male serum, and sheep serum (SD <0.07 g/L for all measurements).

hGH MEASUREMENTS

We provided participating laboratories with detailed protocols. All measurements were performed in triplicate within 1 analytical run. The methods used were Immulite 2000 (Siemens Healthcare Diagnostics) at the Universitair Ziekenhuis Gent (Belgium); Cobas e411 (Roche Diagnostics) at the CIRME, University of Milan (Italy); Liaison (DiaSorin) and iSYS (IDS) at the Klinik der Universität München (Germany); and hGH-sensitive ELISA (Mediagnost) at Institute for Reference Materials and Measurements (IRMM) (Belgium). All the samples were measured once in 1 sequence, then measured in reverse order in a second sequence, and finally measured again as in the first sequence.

DATA ANALYSIS

We analyzed the equivalence of results from different methods by performing, for each pair of methods, a Passing-Bablok regression and determining the Pear-

son correlation coefficient with Analyse-it software. We assessed commutability by Deming regression on the measurement results obtained for the serum samples and verified whether the results for the hGH preparations were within the 95% prediction interval with Excel and R software (19).

LC-MS MEASUREMENTS

We performed LC-MS measurements for 3 serum samples according to the method described by Arsene et al. (20). With this isotope-dilution mass spectrometry method, hGH tryptic cleavage products YSFLQNPQT-SLCFSES IPTPSNR (T6) and LEDGSPR (T12) were quantified by LC-MS. T12 is present in both the 22- and 20-kDa forms of the growth hormone, whereas T6 is present only in the 22-kDa form. Samples were measured in duplicate. Measurements were performed using either labeled peptides or the fully [¹⁵N]-labeled 22-kDa form of growth hormone as internal standard. Results were obtained in micrograms per gram and converted to micrograms per liter using a serum density of 1.02 kg/L, the density of the sera used in this study.

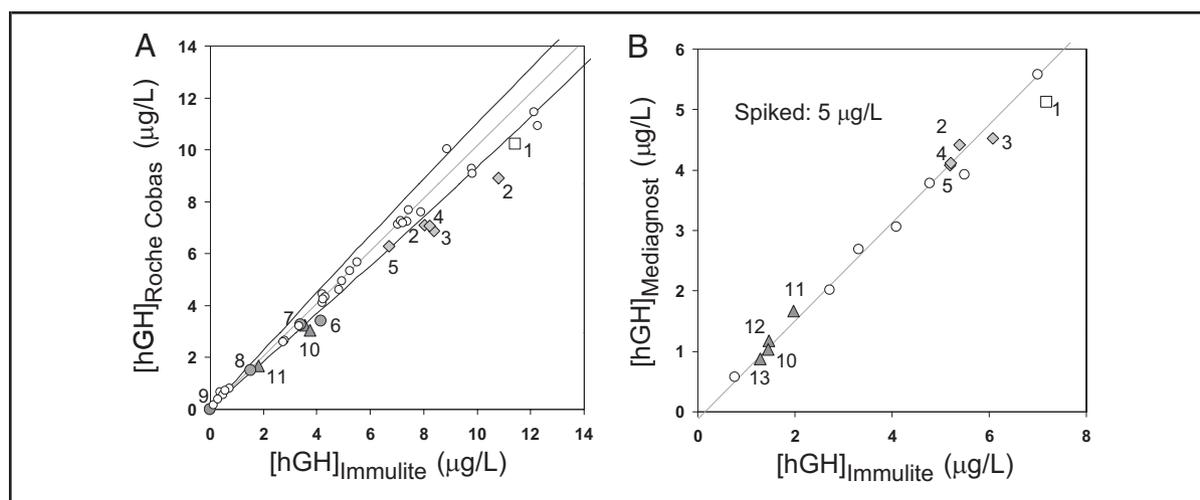


Fig. 2. Results from commutability (A) and recovery (B) studies for hGH.

Plotted are the measurement results for clinical samples and hGH preparations, as well as the regression line and the lines limiting the 95% prediction interval for the clinical samples. Patient samples are shown as filled circles (●). hGH preparations: □, reconstituted with PBS-BSA and spiked into PBS-BSA; ◇, reconstituted with PBS-BSA and spiked into serum (hGH-depleted serum 2, sheep serum 3, SRM 971 male 4, charcoal-stripped serum 5); ○, reconstituted with MilliQ water (sheep serum 6, charcoal-stripped serum 7, GH-depleted serum 8, MilliQ water 9); ▲, reconstituted with PBS and spiked into serum (sheep serum 10, GH-depleted serum 11, charcoal-stripped serum 12, SRM 971 male 13).

IMPACT OF GHBP

GHBP purified from natural sources (ProSpec) was spiked into the samples to produce final concentrations of 0, 10, 75, 250, and 500 µg/L. We then performed hGH measurements in a single analytical sequence, in triplicate, with the hGH-sensitive ELISA (Mediagnost). GHBP content was also measured in both serum samples and hGH preparations using the Active® Human Growth Hormone Binding Protein ELISA (DLS-10-48100, Diagnostic Systems Laboratories). The measurements were performed in 1 single run, in triplicate.

Results

COMMUTABILITY STUDY

For the commutability study, 13 preparations were measured together with 29 serum samples and appropriate control samples (buffer or serum), in 1 analytical run.

For those control preparations with hGH concentrations >0.05 µg/L, the value was corrected by subtraction: specifically, for the NIST SRM 971 material, where the measured hGH concentration was 0.24 µg/L, and for the human charcoal-stripped serum, where the measured hGH concentration was 0.07 µg/L (mean values from all measurements with all methods). The mean value for the control sample, calculated

per method, was subtracted from each value measured for a spiked material with that method. An example of the measurement results is presented in Fig. 2A. The data for all the different methods are shown in the Data Supplement, which accompanies the online version of this article at <http://www.clinchem.org/content/vol59/issue7>.

Method comparisons. All methods showed intraassay CVs <6.0% for the serum samples (Table 1), with the Roche method having the lowest mean CV of 1.7%. For each pair of methods, the correlation between the results (mean value calculated per serum

Table 1. Comparison of results obtained for clinical samples.^a

	Siemens	Roche	Mediagnost	DiaSorin	IDS	CV, %
Siemens		-0.19	NS ^b	NS	NS	4.06
Roche	0.98		-0.21	-0.19	-0.09	1.70
Mediagnost	0.81	0.82		NS	0.10	4.11
DiaSorin	0.78	0.80	0.96		NS	5.69
IDS	0.88	0.86	1.06	1.13		6.03

^a Data are slopes (bottom) and intercepts (top) of Passing-Bablok regression for the equation $A = \text{slope} \times B + \text{intercept}$, e.g., $[\text{hGH}]_{\text{Roche}} = 0.98 \times [\text{hGH}]_{\text{Siemens}} - 0.19$. Mean CV is calculated over all the serum samples.

^b NS, 95% CI of the regression covers 0.

Table 2. Results from the hGH recovery study.^a

Background matrix	Recovery value, %					
	hGH, $\mu\text{g/L}$	Immulite	Cobas e411	hGH-sensitive ELISA	DiaSorin	IDS
Reconstitution in PBS						
PBS	5.05	<1	0.60	<1	<1	<1
hGH-depleted serum	5.18	38	38	32	28	31
Charcoal-stripped serum	5.09	28	28	23	21	23
SRM 971 male	5.03	25	22	17	16	20
Sheep serum	5.12	28	24	20	15	18
Reconstitution in PBS-BSA						
PBS-BSA	4.94	145	135	103	89	106
hGH-depleted serum	5.01	107	98	88	81	89
Charcoal-stripped serum	5.03	103	94	81	82	88
SRM 971 male	4.91	106	90	83	80	92
Sheep serum	4.97	122	100	90	77	90

^a Comparison of different immunoassay measurement values to the theoretical hGH content (gravimetrically controlled). Data are the mean recovery values of triplicate measurements for each of the 5 methods.

sample) was evaluated. The Pearson correlation coefficients were 0.99 or higher, indicating very good correlation of results for individual patient samples despite potential differences in the selectivity of the methods. The slopes of the correlation lines were between 0.78 and 1.13, indicating that a large bias existed between the results from the different methods. The intercepts were either small (<0.2 $\mu\text{g/L}$) or insignificant.

Commutability. For the 5 methods, the different hGH preparations behaved in a manner that seemed similar to those of the serum samples. However, when commutability was assessed according to CLSI guide C53-A, comparing the reference material preparation with the 95% prediction interval for the serum samples (15), none of the preparations was commutable for all comparisons of methods (see online Supplemental Fig. 1, comparisons A to J). Preparations with low measured hGH concentrations (those reconstituted with water or PBS and spiked into hGH-depleted serum) were more often found to be commutable. Comparing the 10 different pairs of experiments, the preparations in which hGH was reconstituted with PBS-BSA and then spiked into hGH-depleted serum (see online Supplemental Fig. 1) showed the best commutability. This preparation was commutable for 8 of the 10 combinations of methods.

RECOVERY STUDY

To investigate the recovery of hGH in preparations of IS 98/574, all the reconstitution and spiking steps were

carefully controlled, and the amount of hGH in each preparation was determined with a low relative standard uncertainty (<0.07%). IS 98/574 was reconstituted with PBS or PBS-BSA and was spiked into the different matrices. The calculated final hGH concentrations of the gravimetric preparations were all close to the intended value, between 4.91 and 5.18 $\mu\text{g/L}$ (Table 2).

The 10 preparations were measured together with 9 serum samples and 6 control samples in 1 analytical run. hGH concentrations determined in the controls (mean value 0.25 $\mu\text{g/L}$ for NIST SRM 971 and 0.08 $\mu\text{g/L}$ for human charcoal-stripped serum) were subtracted from the values of the respective preparations.

The results from the recovery study confirmed that there were large differences in recovery of hGH in different preparations (Fig. 2B). Preparations of IS 98/574 with PBS for reconstitution gave hGH values between 1 and 2 $\mu\text{g/L}$ (Fig. 2B; see online Supplemental Fig. 2). In preparations where reconstitution had been performed with PBS-BSA, the hGH concentrations measured were between 4.5 and 7 $\mu\text{g/L}$, depending on the method and the background matrix into which the hGH was finally spiked (Fig. 2B). hGH reconstituted with PBS-BSA and spiked into the same buffer gave a high value, well above the actual concentration of approximately 5 $\mu\text{g/L}$.

The effect of using different background sera (hGH-depleted serum, charcoal-stripped serum, SRM 971, or sheep serum) was smaller. Differences between hGH concentrations measured in different back-

Table 3. Comparison of LC-MS and immunoassay techniques with respect to hGH quantification.^a

	Sample		
	A	B	C
[hGH] by LC-MS, $\mu\text{g/L}$			
Peptide T12	7.250 (0.023)	5.205 (0.010)	4.365 (0.019)
Peptide T6	6.348 (0.005)	4.458 (0.017)	3.675 (0.012)
$[\text{hGH}]_{\text{LC-MS T12}}/[\text{hGH}]_{\text{IA}}$	0.88	0.83	1.27
[GHBP], $\mu\text{g/L}$	42.2 (1.8)	21.2 (1.6)	24.7 (1.4)
Decrease in [hGH] upon addition of 10 $\mu\text{g/L}$ GHBP, % ^b	9.7	10.6	14.8
[hGH] by immunoassay, $\mu\text{g/L}$			
Siemens	9.78 (0.31)	7.14 (0.44)	4.21 (0.18)
Roche	9.29 (0.21)	7.26 (0.12)	4.10 (0.11)
Mediagnost	7.91 (0.17)	5.62 (0.06)	3.14 (0.07)
DiaSorin	6.90 (0.87)	5.60 (0.30)	2.87 (0.15)
IDS	7.99 (0.76)	6.18 (0.25)	3.25 (0.13)

^a Data are means (SD) unless noted otherwise. Data are the average values of quadruplicate measurements from LC-MS and triplicate measurements for each of the 5 methods.

^b As measured by hGH-sensitive ELISA.

grounds ranged from 2% to 17%, depending on the method. However, these differences were of the same order of magnitude as the discrepancies found between results from different methods. The different recoveries in different serum backgrounds could therefore contribute substantially to the observed between-method bias.

MASS SPECTROMETRY MEASUREMENTS

For 3 serum samples, hGH concentrations were measured by LC-MS. For each of these samples, the means and SDs were calculated from 2 measurements with labeled peptides and 2 measurements with labeled hGH. The SDs of the measurement results were very low (Table 3). From these data, the calculated mass fraction of hGH in the 20-kDa form was 12.4%, 14.4%, and 15.8% for samples A, B, and C, respectively. The comparison with immunoassay results for the same samples showed that the mean ratio between LC-MS and immunoassay results, calculated over the different methods, was 0.88 for sample A, 0.82 for sample B, and 1.24 for sample C. There was no constant ratio between immunoassay and LC-MS results.

GHBP MEASUREMENTS

In plasma, 50% of hGH molecules are expected to be bound to GHBP (9). GHBP may also have an impact on the detection of hGH by immunoassay, whereas for the LC-MS method it has been shown that recombinant GHBP does not interfere with measurements (20). Different concentrations of GHBP were there-

fore spiked into serum samples and other preparations to investigate the effect of the presence of GHBP. The presence of larger concentrations of GHBP reduced hGH values as measured by immunoassay by up to 38%, 33%, and 31% in serum samples A, B, and C, respectively. For preparations containing IS 98/574, the measured hGH concentration was reduced by 46% in the preparation made from the IS reconstituted in PBS-BSA and spiked into PBS-BSA, and by 44% in the preparation where the IS was spiked into hGH-depleted serum (see online Supplemental Fig. 3).

Discussion

The present study addressed the traceability of results from hGH measurements performed with commercial immunoassays. Measurement results are made traceable to a higher-order reference material or reference method with the aim of obtaining equivalent results. However, hGH values measured with different routine methods differ significantly (14), despite the fact that the values assigned to the calibrators (true for all 5 methods used in this study) are stated to be traceable to IS 98/574. Possible causes for the lack of harmonization were investigated in this study, as well as possible solutions.

Differences in selectivity of methods intended to measure the same quantity are common in clinical chemistry, because different methods use different reagents (at least) for the same target analyte. Antibodies

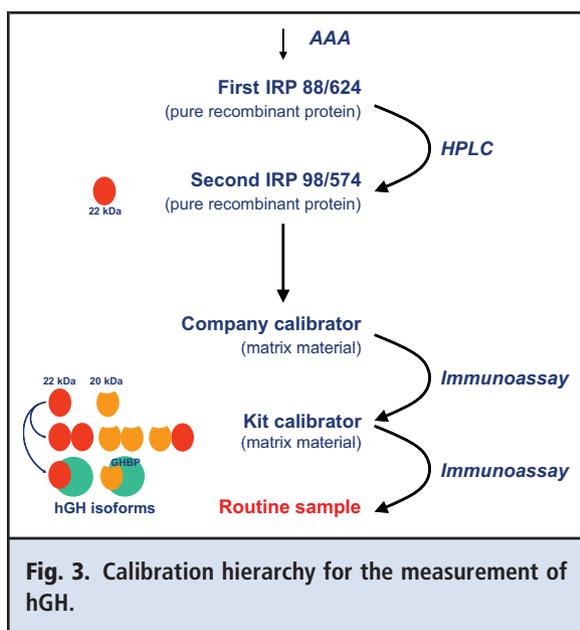


Fig. 3. Calibration hierarchy for the measurement of hGH.

can be monoclonal or polyclonal, derived from different species (e.g., goat, mouse, chicken), and raised against different targets (e.g., recombinant hGH, a mixture of isoforms, a single peptide). Our results showed that for the 5 methods evaluated here, the correlation of results for the serum samples was very good (correlation coefficients >0.99). This means that either these methods had no major differences in selectivity or they had different selectivities for analytes present in a constant ratio in the samples. The latter is more likely, because substantial differences in isoform recognition by the different antibodies used have been described (21).

Because the results from the different methods correlated, it should be possible to reduce discrepancies in measurement results by calibrating all methods with a common commutable calibrator or by data adjustment. Muller et al. (22) showed that the between-method CV for hGH measurements could be reduced from 24.3% to <14% by data adjustment. The mean between-method CV for the serum samples in our study was 17.7%. When calculating mean correction factors per method (by averaging correction factors calculated for individual samples) and applying this to the results for the serum samples, the between-method CV was reduced to 7.5%, comparable to what was found in a study on the potential use of a common harmonization sample (23, 24).

The most common currently available calibration hierarchy, or traceability chain, for hGH measurements is shown in Fig. 3. Clinical measurements are performed in a complex matrix, on a mixture of hGH forms. IS 98/574, referred to in traceability statements, is several steps up in the traceability chain and consists

of a purified single isoform without matrix. There are different problems in linking the IS to the clinical measurement results. First, the lack of a common reconstitution and spiking protocol leads to a situation where manufacturers may be using calibrators that are de facto different and produce uncontrolled bias. For example, the use of different serum backgrounds for spiking resulted in values for hGH that differed by $\leq 17\%$.

Second, none of the calibrator solutions produced from IS 98/574 according to the protocols presented here were commutable for all the combinations of methods. The distance of the results for the reconstituted IS 98/574 from the regression line for the serum samples provided a measure of the bias caused by a lack of commutability. This bias was of the same order of magnitude as the between-method bias. Therefore, it would not be sufficient to use a common reconstitution and spiking protocol for the production of a master calibrator from IS 98/574. Optimal harmonization requires that such a protocol ensures a commutable matrix preparation.

Third, since different protocols lead to different recoveries of hGH, the value obtained with an immunoassay would be protocol dependent. This could eventually be solved by use of an orthogonal method for measuring hGH, such as LC-MS. A reconstitution and spiking protocol yielding immunoassay and LC-MS results having the same ratio for the calibrator and serum samples could be selected. This would anchor the immunoassay results to the hGH concentration as determined by LC-MS, which should be independent of matrix used. The first results showed, however, that the ratio of hGH as measured by LC-MS and immunoassay was significantly different for the different samples. This indicates that a better understanding of the influence of quantities and parameters of immunoassays and LC-MS measurements—i.e., of what we are measuring—is required before LC-MS measurements can be used for anchoring the immunoassay results. A larger-scale study of samples measured both by LC-MS and immunoassay should provide more information.

One of the main influencing factors for hGH immunoassays is the presence of GHBP (5). Different serum samples have different GHBP concentrations, and different GHBP concentrations in samples change hGH values as measured by immunoassay (see online Supplemental Fig. 3). This is true not only for the method used in our study but also for other methods (25). GHBP was measured in the 3 serum samples to determine whether this could explain the large ratio of $[\text{hGH}]_{\text{LC-MS T12}}/[\text{hGH}]_{\text{IA}}$ for sample C. However, there was no simple correlation between immunoassay values, LC-MS values, and GHBP concentrations, since the GHBP concentrations were comparable in samples B and C.

In conclusion, the commutability and recovery studies reported here show that to standardize hGH measurements, it is not sufficient to recommend traceability to a higher-order reference material such as IS 98/574. Results from different immunoassays correlate well for serum samples, so it is possible in principle to systematically improve the equivalence of hGH measurement results. It has been proposed that correction factors be used to achieve this goal, at least for the present. However, this approach has some inherent disadvantages; it would be difficult to maintain standardization of hGH measurements and avoid drift in values. The production of a matrix reference material from pooled serum, and its use for calibration, should lead to equivalence of results. Standardization might also be achieved through a protocol for producing a commutable matrix calibrator from IS 98/574, but our results show that this is challenging. LC-MS methods could provide reference values. However, the limited correlation between the data from the group of immunoassay-based procedures and data from the LC-MS-based measurement procedure may indicate differences in selectivity. Therefore, the many factors influencing commercial measurement procedures and LC-MS need to be clearly understood to allow control of all sources of calibration bias and design a stable reference system on the basis of LC-MS.

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: M. Bidlingmaier, Immunodiagnostic Systems.

Stock Ownership: None declared.

Honoraria: M. Bidlingmaier, Siemens, Roche, and DiaSorin.

Research Funding: The work described in this paper was carried out within the EURAMET joint research project ClinBioTrace, which has received funding from the European Community's Seventh Framework Programme, ERA-NET Plus, under grant agreement no. 217257. hGH reagents used at CIRME (University of Milan) were a gift from Roche Diagnostics. U-¹⁵N-labeled 22-kDa-hGH used at Physikalisch-Technische Bundesanstalt (PTB) was kindly provided by Jens Breinholt (Novo Nordisk).

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, or preparation or approval of manuscript.

Acknowledgments: The authors thank Gavin O'Connor and Milena Quaglia (LGC) for a sample containing reconstituted hGH, as well as André Henrion (PTB) for useful discussions.

References

- Markkanen H, Pekkarinen T, Valimäki MJ, Alftan H, Kauppinen-Makelin R, Sane T, Stenman UH. Effect of sex and assay method on serum concentrations of growth hormone in patients with acromegaly and in healthy controls. *Clin Chem* 2006;52:468–73.
- Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. *J Clin Endocrinol Metab* 2000;83:3990–3.
- Assessment and long-term monitoring in patients with acromegaly: statement from a joint consensus conference of the Growth Hormone Research Society and the Pituitary Society. *J Clin Endocrinol Metab* 2004;89:3099–102.
- Popii V, Baumann G. Laboratory measurement of growth hormone. *Clin Chim Acta* 2004;350:1–16.
- Strasburger CJ, Bidlingmaier M. How robust are laboratory measures of growth hormone status? *Horm Res* 2005;64 Suppl 2:1–5.
- Tong J, D'Alessio D, Ramisch J, Davis HW, Stambrook E, Tschop MH, Bidlingmaier M. Ghrelin stimulation of growth hormone isoforms: parallel secretion of total and 20-kDa growth hormone and relation to insulin sensitivity in healthy humans. *J Clin Endocrinol Metab* 2012;97:3366–74.
- Trainer PJ, Barth J, Sturgeon C, Wieringaon G. Consensus statement on the standardisation of GH assays. *Eur J Endocrinol* 2006;155:1–2.
- Bidlingmaier M, Freda PU. Measurement of human growth hormone by immunoassays: current status, unsolved problems and clinical consequences. *Growth Horm IGF Res* 2010;20:19–25.
- De Palo EF, De Filippis V, Gatti R, Spinella P. Growth hormone isoforms and segments/fragments: molecular structure and laboratory measurement. *Clin Chim Acta* 2006;364:67–76.
- Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. *Off J Eur Communities* 1998 Dec 7;L 331:1–37. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31998L0079:20090807:EN:HTML> (Accessed June 2013).
- Clemmons DR. Consensus statement on the standardization and evaluation of growth hormone and insulin-like growth factor assays. *Clin Chem* 2011;57:555–9.
- Bristow AF, Jespersen AM. The Second International Standard for somatotropin (recombinant DNA-derived human growth hormone): preparation and calibration in an international collaborative study. *Biologicals* 2001;29:97–106.
- Pritchard C, Quaglia M, Mussell C, Burkitt WI, Parkes H, O'Connor G. Fully traceable absolute protein quantification of somatotropin that allows independent comparison of somatotropin standards. *Clin Chem* 2009;55:1984–90.
- RfB: Referenzinstitut für Bioanalytik. Ringversuche/HP – Hormonbestimmungen Gruppe 2/2009/ Serum 2009. <http://www.dgkl-rfb.de> (Accessed March 2013).
- CLSI. Characterization and qualification of commutable reference materials for laboratory medicine; approved guideline. Wayne (PA): CLSI; 2010. Document no. C53-A.
- WHO International Standard: somatotrophin (recombinant DNA-derived human growth hormone): NIBSC code: 98/574: instructions for use (version 4.0, dated 14/01/2008). Potters Bar (UK): National Institute for Biological Standards and Control; 2012. <http://www.nibsc.ac.uk/> (Accessed June 2013).
- ASTM D4052 - 11: ASTM D4052 - 11 standard test method for density, relative density, and API gravity of liquids by digital density meter. 2011. <http://www.astm.org/Standards/D4052> (Accessed June 2013).
- Zegers I, Schreiber W, Sheldon J, Blirup-Jensen S, Muñoz A, Merlini G, et al. Certification of proteins in the human serum: Certified Reference Material ERM®-DA470k/IFCC. Luxembourg: Office for Official Publications of the European Communities; 2008. <http://publications.jrc.ec.europa.eu/repository/handle/11111111/12304> (Accessed June 2013). Identifier: EUR 23431 EN.
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2012. <http://www.R-project.org> (Accessed March 2013).

-
20. Arsene CG, Henrion A, Diekmann N, Manolopoulou J, Bidlingmaier M. Quantification of growth hormone in serum by isotope dilution mass spectrometry. *Anal Biochem* 2010;401:228–35.
 21. Manolopoulou J, Alami Y, Petersenn S, Schopohl J, Wu Z, Strasburger CJ, Bidlingmaier M. Automated 22-kD growth hormone-specific assay without interference from pegvisomant. *Clin Chem* 2012;58:1446–56.
 22. Muller A, Scholz M, Blankenstein O, Binder G, Pfaffle R, Korner A, et al. Harmonization of growth hormone measurements with different immunoassays by data adjustment. *Clin Chem Lab Med* 2011;49:1135–42.
 23. Ross HA. Reporting growth hormone assay results in terms of one consensus recombinant standard preparation offers less than optimal reduction of between-method variation. *Clin Chem Lab Med* 2008;46:1334–5.
 24. Ross HA, Lentjes E, Menheere PP. The consensus statement on the standardization and evaluation of growth hormone and insulin-like growth factor assays lacks a recommendation to attempt efficacious harmonization. *Clin Chem* 2011;57:1463, author reply 1464.
 25. Hansen TK, Fisker S, Hansen B, Sorensen HH, Christiansen JS, Jorgensen JO, H OR. Impact of GHBP interference on estimates of GH and GH pharmacokinetics. *Clin Endocrinol* 2002;57:779–86.