Harmonization of growth hormone measurements with different immunoassays by data adjustment

Anne Müller[1], Markus Scholz[2], Oliver Blankenstein[3], Gerhard Binder[4], Roland Pfäffle[5], Antje Körner[5], Wieland Kiess[5], Annegret Heider[6], Martin Bidlingmaier[7], Joachim Thiery[1] and Jürgen Kratzsch[1,*]

[1] Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany
[2] Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
[3] Charité University of Medicine, Berlin, Germany
[4] Paediatric Endocrinology, University Children's Hospital, Tübingen, Germany
[5] Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany
[6] Central Laboratory, University of Halle (Saale), Halle, Germany
[7] Department of Internal Medicine (Endocrinology)-Innenstadt, University of Munich, Munich, Germany

Abstract

Background: The aim of our study was to evaluate the between-assay variability of commercially available immunoassays for the measurement of human growth hormone (hGH). In addition, we asked whether the comparability of the diagnosis of childhood onset growth hormone deficiency could be improved by adjusting hGH results by statistical methods, such as linear regression, conversion factors, and quantile transformation.
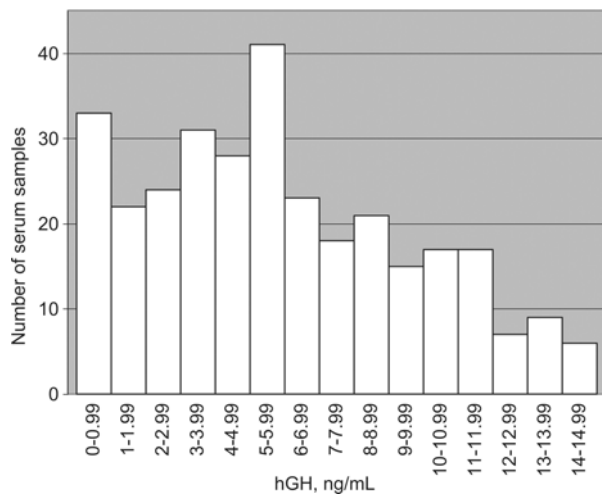**Methods:** In archived sera from 312 children and adolescents (age: 17 days–17 years) hGH values between 0.01 and 16.5 ng/mL were determined by using the following immunoassays: AutoDELFIA (PerkinElmer), BC-IRMA (Beckman-Coulter), ELISA (Mediagnost), IMMULITE 2000 (Siemens), iSYS (IDS), Liaison (DiaSorin), UniCel DxI 800 Access (BeckmanCoulter) and ''In house''-RIA (Tübingen).
Results: The assays differed in median hGH concentrations by as much as 5.44 ng/mL (Immulite), and as little as 2.67 ng/mL (BC-IRMA). The mean difference between assays ranged from 0.35 to 2.71 ng/mL, whereas several samples displayed differences up to 11.4 ng/mL. The best correlation (r = 0.992) was found between AutoDELFIA and Liasion, the lowest (r = 0.864) was between an in-house RIA and iSYS. The between-assay CV (mean ± SD) of values within the cut-off range was 24.3% ± 7.4%, resulting in an assay-dependent diagnosis of growth hormone deficiency (GHD) in more than 27% of patients. Yet, adjustment of this data by linear regression or a conversion factor reduced the CV below 14%, and the ratio of assay-dependent diagnoses below 8%. Using quantile transformation, the CV and ratio were reduced to 11.4% and < 1%, respectively.
Conclusions: hGH measurements using different assays vary significantly. Linear regression, conversion factors, or particularly quantile transformation are useful tools to improve comparability in the diagnostic procedure for the confirmation of GHD in childhood and adolescence.

Keywords: adolescents; children; growth hormone deficiency; immunoassay; short stature.

*Corresponding author: Prof. Dr. rer. nat. Jürgen Kratzsch, University of Leipzig, Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, Paul-List-Str. 13–15, 04103 Leipzig, Germany
Phone: +49 341 97 222 41, E-mail: kraj@medizin.uni-leipzig.de

Introduction

The prevalence for growth hormone deficiency (GHD) in children and adolescents is estimated to be between 1:4000 and 1:20,000 (1). After exclusion of other causes of short stature, growth velocity below the 25th percentile in combination with bone age retardation by more than one standard deviation (SD) and insulin-like growth factor-I (IGF-I), and/or IGF-binding protein (IGFBP)-3 concentrations below -1 SD score are sufficient for the strong suspicion of GHD according to the German diagnostic guidelines (2). A definite diagnosis is established by two stimulation tests with a human growth hormone (hGH) peak below an arbitrary cut-off of 5–10 ng/mL (2, 3) as decision cut-point. Measured hGH values can vary remarkably depending on the assay method used (4). Reasons for the variability of immunoassay results are the preparation of assay calibrators, specificity of antibodies, assay design (sandwich vs. competitive), matrix differences between standards and samples, and interferences with endogenous human growth hormone binding protein (hGHBP) (5). Inter-laboratory tests of the German Society of Clinical Chemistry and Laboratory Medicine with synthetic test samples showed a coefficient of variation (CV) for hGH measurements between 18% and 36% for more than 20 immunoassays that were used in 2009 (4). However, synthetic control samples may not reflect all the potential inter-

Figure 1   Number of serum samples in dependence on measured hGH levels by the IMMULITE 2000 fully automated system.

fering factors in an immunoassay. To evaluate the comparability of hGH immunoassay measurements in serum of children and adolescents, we tested seven commercially available methods frequently used in clinical practice, as well as an in-house radioimmunoassay (RIA). Then, we reduced the variability between these assays using three different mathematical harmonization procedures: linear regression analysis, conversion factors and quantile transformation. The aim of our study was to assess which adjusting procedure performs best in harmonizing growth hormone measurements and in improving comparability of GHD diagnosis.

Materials and methods

Samples

A total of 168 samples from male patients and 144 samples from female patients with suspected GHD ranging in age from 17 days to 17 years were collected from surplus sera from stimulation tests and from the evaluation of spontaneous hGH secretion at the pediatric departments of the universities of Leipzig, Tübingen, and Berlin.

The mean age $\pm$ SD of our patients was 8.00 ± 3.44 years (n = 312). The initially documented hGH concentrations of all sera were between 0 and 15 ng/mL (Figure 1).

Two hundred and sixty-two samples were from results of single blood collections, 23 derived from two pooled samples of patients having comparable hGH concentrations and similar age, 27 derived from two pooled samples from the same child containing similar hGH concentrations (mean differences $\pm$ standard deviation: 0.27 ± 0.38 ng/mL). Pools were made to achieve adequate volume for our analytical determination. The samples were stored at a minimum of $< -25°C$ until determination of hGH. The study was reviewed and approved by the Ethic Committee of the University of Leipzig.
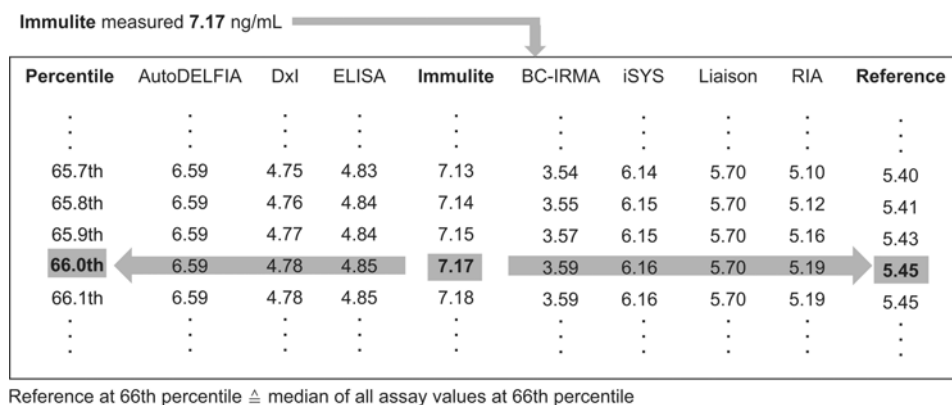
HGH measurement

HGH concentrations of serum samples were investigated by a multicentre study using eight different immunoassays. Five were fully automated (AutoDELFIA, PerkinElmer; IMMULITE 2000, Siemens; iSYS, IDS; Liaison, DiaSorin; UniCel DxI 800 Access, BeckmanCoulter) and three were performed using a manual procedure (BC-IRMA, BeckmanCoulter; ELISA, Mediagnost; ''In house''-RIA, Tübingen). All samples were processed between June 2008 and January 2010. Technical information about method-dependent calibrator and analytical specifications is given in Table 1. The only non-commercially available assay was a competitive hGH-

Table 1   Technical information for the 8 assays used for the measurement of hGH.

Method (Manufacturer)	Analytical sensitivity, ng/mL	Precision (coefficient of variation)	Specificity	Reference preparation	Conversion factor, ng×x =μU
AutoDELFIA (PerkinElmer)	0.012	0.6%–3.7% (3.2–13 ng/mL) 4.7%–13% (3.3–13.3 ng/mL)	No 20 kDa hGH	WHO 1. IS 80/505	2.6
DxI (Beckmann-Coulter)	0.002	2.1%–3.2% (2.5–14.1 ng/mL)	No 20 kDa hGH	WHO 2. IS 98/574	3
ELISA (Mediagnost)	0.0016	3.1%–6% (2.4–14.3 ng/mL)		WHO 2. IS 98/574	3
IMMULITE 2000 (Siemens)	0.01	4.2%–6.6% (2.6–17 ng/mL)		WHO 2. IS 98/574	3
BC-IRMA (Beckmann-Coulter)	0.01	5.7%–10.1% (0.6–5.2 ng/mL)		WHO 1. IS 88/624	3
iSYS (IDS)	0.05	2.7%–6.1% (1.7–26.8 ng/mL)	No 20 kD hGH	WHO 2. IS 98/574	3
Liaison (DiaSorin)	Max. 0.052	2.2%–3.1% (2.5–10.2 ng/mL)		WHO 2. IS 98/574	3
RIA (in house) Tübingen	0.05	2.2%–7.6% (2.5–10.2 ng/mL)		WHO 2. IS 98/574	3

DxI, UniCel DxI 800 Access; ELISA, Mediagnost ELISA E022; BC-IRMA, BC-BC-IRMA; iSYS, IDS-iSYS; RIA, In-house RIA Tübingen; Hb, haemoglobin; TG, triglycerides; hLH, human luteinizing hormone.

Immulite measured **7.17** ng/mL

Percentile	AutoDELFIA	DxI	ELISA	Immulite	BC-IRMA	iSYS	Liaison	RIA	Reference
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
65.7th	6.59	4.75	4.83	7.13	3.54	6.14	5.70	5.10	5.40
65.8th	6.59	4.76	4.84	7.14	3.55	6.15	5.70	5.12	5.41
65.9th	6.59	4.77	4.84	7.15	3.57	6.15	5.70	5.16	5.43
66.0th	6.59	4.78	4.85	**7.17**	3.59	6.16	5.70	5.19	**5.45**
66.1th	6.59	4.78	4.85	7.18	3.59	6.16	5.70	5.19	5.45
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.

Reference at 66th percentile $\triangleq$ median of all assay values at 66th percentile

Figure 2 Example calculation how a single GH measurement transforms when using quantile transformation.

RIA developed and employed at the Children's Hospital in Tübingen (6).

Statistical methods

The total number of measured growth hormone concentrations was not normally distributed. Therefore, we used Spearman's correlation statistics which was also interpreted as a measure of comparability between two assays.

Data analysis was performed using Microsoft Office Excel 2003. Linear regression analysis for the comparison of results from different methods was performed according to Passing-Bablok (7) using MedCalc (MedCalc Software, 9030 Mariakerke, Belgium).

As another mean to harmonize the data, we performed quantile normalization, a method adopted from microarray normalization (8). With this method, the percentiles of the empirical distributions of hGH concentrations were determined for each of the eight assays separately. Due to the lack of a gold-standard or reference scale, we defined a common ''reference'' of all assays by ''averaging'' the assay percentiles. More precisely, we calculated the percentiles of the common ''reference'' by taking the median of the matching percentiles of the eight assays. Finally, all single assay measurements were mapped to the common ''reference'' using their respective intra-assay percentiles. That is, a single measurement is mapped to the inter-assay median of the matching intra-assay percentile. Figure 2 illustrates the procedure in more detail. Calculations were

performed using the statistical software package R (www.r-project.org). The corresponding R-script and a running example can be found in the Annex.

To assess the effect of the different normalization methods used, we calculated the inter-assay coefficients of variation before and after the normalization steps. Different ranges (<1, $1–4.99$, $5–9.99$, ≥ 10) of hGH concentrations were used to assess the performance of the normalization methods for different groups of patients.

Results

Comparison of hGH levels measured

By using the eight different methods noted above for the measurement of hGH we obtained values ranging from 0.01 to 16.50 ng/mL (Table 2). The mean value of all measurements was 4.60 ng/mL, the median was 4.10 ng/mL. HGH results from the IMMULITE 2000 (median: 5.44 ng/mL) were found to be considerably higher than the overall median. The lowest hGH concentrations of all assays were from the BC-IRMA (median: 2.67 ng/mL). The medians of measurements by the AutoDELFIA (median: 4.95 ng/mL), Liaison (4.20 ng/mL), iSYS (4.55 ng/mL), RIA (3.90 ng/mL), ELISA (3.55 ng/mL), and DxI (3.46 ng/mL) were distributed

Table 2 Descriptive analysis for all samples (n = 312) measured by the different hGH assay methods.

	Mean ± SD, ng/mL	Median, ng/mL	Perc 25–75, ng/mL	Min–Max, ng/mL
AutoD	5.62 ± 3.84	4.95	2.44–8.58	0.04–16.5
DxI	3.91 ± 2.76	3.46	1.62–6.05	0.03–12.74
ELISA	3.91 ± 2.75	3.55	1.71–5.88	0.01–13.83
Immulite	5.90 ± 3.77	5.44	2.95–8.51	0.06–14.80
BC-IRMA	3.19 ± 2.36	2.67	1.39–5.00	0.01–11.30
iSYS	5.28 ± 3.52	4.55	2.71–7.40	0.14–15.71
Liaison	4.76 ± 3.16	4.20	2.20–7.10	0.10–15.20
RIA	4.25 ± 2.43	3.90	2.20–5.90	0.80–11.70
All assays	4.50 ± 3.18	3.99	1.99–6.50	0.01–16.50

AutoD, AutoDELFIA; DxI, UniCel DxI 800 Access; ELISA, Mediagnost ELISA E022. Immulite, IMMULITE 2000. BC-IRMA, Beckman-Coulter immunoradiometric assay; iSYS, IDS-iSYS; RIA, In-house RIA Tübingen; SD, standard deviation; x_{min}, minimum value; x_{max}, maximum value; Perc 25, 25th percentile; Perc 75, 75th percentile.

between the IMMULITE and BC-IRMA data. The maximum difference between individual measurements by different methods was 11.4 ng/mL. The mean differences of GH concentrations measured by specific assays ranged from 0.36 ng/mL (DxI vs. ELISA) to 2.71 ng/mL (IMMULITE vs. BC-IRMA).

Individual assay results from different methods were compared by correlation analysis. Coefficients of correlations varied between $r = 0.992$ for the comparison between AutoDELFIA with Liaison, and $r = 0.864$ for the comparison between iSYS and in-house RIA.

The median CV of all samples and methods in the potential clinically relevant cut-off range between 5 and 10 ng/mL ($n = 118$) was $24.3\% \pm 7.4\%$ (mean $\pm$ SD, see also Table 3). After adjustment for the biological potency of the international reference preparations used in each assay, we obtained a comparable CV of $24.3\% \pm 8.6\%$. After exclusion of methods not calibrated against the latest reference preparation of the WHO (98/574), the CV was calculated to be $17.3\% \pm 9.6\%$ (Table 3). Considering just the three methods that specifically measure the 22 kDa isoform of hGH by use of monoclonal antibodies, the CV was $24.1\% \pm 10.2\%$. For this calculation we had to consider carefully the different reference preparations used by the manufacturers. Thus, we multiplied all results of AutoDELFIA by the factor of 1.154, which was calculated as the ratio of the conversion factors 3 mU/L and 2.6 mU/L.

Procedures for the adjustment of data

In a next step, we aimed to adjust the different hGH concentrations of the individual assay methods to the results of the IMMULITE system which was arbitrarily defined as a reference standard (y) by linear regression analysis. The individual Passing-Bablok regression equations were, for AutoDELFIA $y = 1.034x + 0.207$; DxI $y = 1.387x + 0.356$; ELISA $y = 1.394x + 0.267$; BC-IRMA $y = 1.684x + 0.443$; iSYS $y = 1.109x + 0.043$, Liaison $y = 1.272x + 0.023$ and RIA $y = 1.571x - 0.971$. If these equations were applied to all

non-IMMULITE hGH measurements, we observed a distinctly decreased CV of the transformed data: $13.2\% \pm 7.6\%$ (Table 3).

The following step involved the calculation of conversion factors, which are another and much more practical tool for the linear transformation of hGH values, and therefore better comparable results. For this purpose, we selected all samples containing a hGH concentration in the putative range for the decision GHD or non-GHD between 5.0 and 10 ng/mL according to the Immulite assay ($n = 78$). Then, we calculated the ratio between the mean value of all measurements of all assays (as the arbitrary gold standard) and the mean value of all measurements from each individual assay. The mean concentration of these 78 samples measured by all assays was 6.27 ng/mL. The mean concentration of the same samples measured, for example, by the AutoDELFIA method was 7.73 ng/mL; accordingly, the conversion factor for the AutoDELFIA method was 0.81. The remaining conversion factors were calculated to be 1.13 for DxI; 1.11 for ELISA; 0.80 for IMMULITE 2000; 1.47 for BC-IRMA; 0.91 for iSYS; 0.97 for Liaison; 1.08 for RIA. The mean CV value after applying this transformation was $12.6\% \pm 7.6\%$ in the cut-off range between 5 and 10 ng/mL (Table 3).

Finally, we used quantile transformation as a novel normalization procedure. All hGH data were ranked and percentiles were determined for the individual assays. As a reference, the median of all assay values for a particular percentile was defined. The transformed value of any sample derived from its percentile after ranking and corresponds to the value of the reference at the same percentile. Applying this procedure, the mean CV value was reduced to $11.4\% \pm 7.6\%$ in the cut-off range between 5 and 10 ng/mL (Table 3).

Consequences for the diagnostic confirmation of GHD

The effect of different statistical procedures for the adjustment of hGH measurements on the diagnosis of GHD was assessed as shown in Table 4. If a hGH concentration of

Table 3 Coefficient of variation for all samples within a defined concentration range consistent with all 8 immunoassays.

CV (%) after using	hGH, ng/mL			
	<1	1–4.99	5–9.99	≥10
Raw data	68.2% ± 45.6% (n = 33)	28.2% ± 11.0% (n = 105)	24.3% ± 7.4% (n = 118)	24.0% ± 7.7% (n = 56)
Calibrator 98/574 only	63.7% ± 39.7% (n = 33)	23.4% ± 13.3% (n = 105)	17.3% ± 9.6% (n = 118)	19.9% ± 10.8% (n = 56)
No 20 kDa GH specific only	41.9% ± 36.1% (n = 33)	27.0% ± 13.0% (n = 105)	24.1% ± 10.2% (n = 118)	26.9% ± 10.0% (n = 33)
Linear regression	32.3% ± 29.0% (n = 33)	15.4% ± 11.7% (n = 105)	13.2% ± 7.6% (n = 118)	12.5% ± 8.3% (n = 56)
Conversion factor	69.1% ± 48.3% (n = 37)	17.7% ± 10.6% (n = 153)	12.6% ± 7.6% (n = 103)	15.9% ± 9.2% (n = 19)
Quantile transformation	34.6% ± 44.3% (n = 38)	13.4% ± 9.5% (n = 155)	11.4% ± 7.6% (n = 106)	13.0% ± 5.7% (n = 13)

Results for ''Calibrator 98/574 only'' contain values of those 5 immunoassays calibrated with IS 98/574. ''No 20 kDa hGH specific only'' contains data of the 3 immunoassays specific to the 22 kDa isoform of hGH (AutoDELFIA, UniCel DxI 800 Access, IDS-iSYS).

6 ng/mL would be defined as a cut-off (and all samples examined would be peak samples from stimulation tests), GHD would be diagnosed in 179 out of 312 children by the IMMULITE 2000 system, whereas the BC-IRMA would detect 268 cases. Accordingly, the diagnosis of 29% of the investigated children would be dependent on the immunoassay used. An increased cut-off of 8 ng/mL would not change this outcome substantially: 27% would have an assay-dependent diagnosis. This assay dependence was effectively reduced by adjustment of the hGH values measured using linear regression or conversion factor to below 8%, and by application of quantile normalization to below 1% for both cut-offs.

Discussion

This is, to the best of our knowledge, the most comprehensive study dealing with comparability of assays measuring hGH concentrations in children and adolescents. For eight frequently used hGH assays we calculated a median CV of 24.3% for concentrations between 5 and 10 ng/mL hGH. As the actual cut-off for the diagnosis of GHD in children and adolescents is within this concentration range, the diagnostic decision depends strongly on the accuracy and the precision of the laboratory method that is used. Accordingly, only a low degree of comparability was observed if these methods were used to diagnose GHD in our sera. To overcome this dilemma we investigated two different approaches: 1) harmonization by assay standardization (reference and antibody), 2) harmonization of assay results by statistical methods.

Regarding the assay standardization-based approach, we found that the restriction to only the single calibrator IS 98/574 led to an improved between-assay CV from 24.3% to 17.3%. These data are similar to findings reported by Ross (9) who demonstrated that a unique reference preparation decreased the between-assay CV of six hGH immunoassays from 25% to 15% when six samples were included in the statistical calculation. Additionally, according to Tanaka et al. (10), the restriction to a unique recombinant hGH standard (60 samples, six assays) resulted in a decrease of the between assay CV from 35% to 18%. This improvement led to the recent agreement that all manufacturers of hGH immunoassay kits in Japan need to use the same recombinant reference preparation for calibration purposes. Furthermore, neither transformation to international units (CV: 24.3%) nor considering only 22 kDa specific assays (CV: 24.1%), as proposed by Ebdrup et al. (11), Bidlingmaier and Freda (12) and others, demonstrated a more substantial reduction of variability in our measurements. Taking all these results together, the standardization approach improves the comparability of data significantly, but not satisfactorily. One may speculate that a lack of cross-reactivity with the 20 kDa hGH isoform alone, as it has been documented for three assays, is not sufficient for complete standardization of antigen-antibody interaction. Different levels of aggregates and cleavage of the 22 kDa monomer may be recognized differently even by monoclonal antibodies (13). Moreover, hGHBP may have a modulating effect on hGH results, although we did not find significant correlations between hGH and hGHBP concentrations in our samples (data not shown). Of particular concern, if the incubation period in the assay is relatively short, as for example 2 h in most of our methods, or if the affinity constant of the antibody to the 22 kDa isoform is only slightly higher than that of hGHBP, the probability for a confounding interaction by this binding protein could be high (11, 14).

To improve the comparability of our hGH results independently from the standardization of assay components, we searched for a statistical method that is able to harmonize hGH results: adjusting our data by linear regression led to a decrease in the CV from 24.3% to 13.2%. Linear regression has already been suggested in some papers for the harmonization of hGH. Tanaka et al. (10) concluded that correcting hGH values by this method could indeed enhance the equality of the diagnosis of GHD. We next assessed a conversion factor as a second statistical tool for linearly associated data. Its application resulted in satisfactory reduction of the CV

Table 4 Influence of different tools for data transformation on the diagnosis GHD (%) at two different cut-off points.

Cut-off	GHD according to raw data, %		GHD after linear regression, %		GHD after conversion factor, %		GHD after quantile transformation, %	
	6 ng/mL	8 ng/mL	6 ng/mL	8 ng/mL	6 ng/mL	8 ng/mL	6 ng/mL	8 ng/mL
AutoD	61.9	70.8	58.0	69.6	69.2	83.3	70.5	86.2
DxI	74.7	91.7	57.8	72.8	69.9	84.9	70.2	86.5
ELISA	76.0	92.0	57.8	72.8	71.8	84.9	70.5	86.5
Immulite	57.4	70.5	57.4	70.5	68.3	82.1	70.5	85.9
BC-IRMA	85.9	97.1	59.9	70.5	68.6	80.1	70.5	86.5
iSYS	65.4	78.5	59.3	73.7	67.9	82.4	70.5	86.5
Liaison	67.9	81.7	57.1	69.2	68.9	83.3	70.5	86.5
RIA	75.3	91.7	55.1	73.4	70.8	87.8	69.9	86.5
CV, %	13.0	12.3	2.5	2.5	1.9	2.8	0.3	0.3

AutoD, AutoDELFIA; DxI, UniCel DxI 800 Access; ELISA, Mediagnost ELISA E022; Immulite, IMMULITE 2000; BC-IRMA, Beckman-Coulter immunoradiometric assay; iSYS, IDS-iSYS; RIA, In-house RIA Tübingen; CV, coefficient of variation; Lin. Regr., linear regression; Quant. Transf., quantile transformation; GHD, growth hormone deficiency.

from 24.3% to 12.6%. Theoretically, a harmonization serum that enables establishment of a conversion factor for each individual assay run could further reduce this CV to below 10%, as shown recently for six samples by a panel of six methods (9).

However, the use of linear methods for data adjustment is debatable. Recently, Hauffa et al. investigated the use of linear transformation methods for the adjustment of hGH concentrations from different assays (6). HGH concentrations from this study were not linearly associated over the whole range up to approximately 100 mU/L, but linearity appeared to be given for the hGH concentration range relevant for the diagnosis of GHD. Amed et al. (15) investigated hGH in 101 samples from children in a comparison of five assay methods. They evidenced scatter in their results that suggests nonlinearity restricted to a concentration range from the detection limit to approximately 5 ng/mL hGH. In contrast to both papers, we observed largely linear associations of hGH concentrations in all our assays by the Passing-Bablok method for the range up to the decision threshold for the diagnosis GHD; between 0 and 8 ng/mL hGH, all assay comparisons demonstrated linearity in the data. This statement

was still valid for four comparisons if this range was extended to 10 ng/mL hGH (Figure 3, data not shown). Accordingly, data transformation by linear statistics could be achieved with an acceptable bias. However, the use of linear statistics for harmonization does indeed have the general disadvantage that bias could become noticeable with increasing deviation from the linear regression line.

Nevertheless, such methods seem to be a practical option with minimal effort for daily practice. As shown for the Netherlands, a central pool serum can be shipped and used for the calculation of run-dependent conversion factors (9). The ratio from its measuring value and the target concentration will be used for conversion of sample data.

A novel statistical approach used for hGH data harmonization was quantile transformation, a non-parametric method adopted from normalization of microarray data. The only precondition for this model was a sufficiently high number of measurements in order to estimate required percentiles and a monotony assumption regarding the assay data. That is, if a person has a higher value with one assay compared to a second person, it also should have a higher value with another assay. Both requirements were fulfilled by our data since
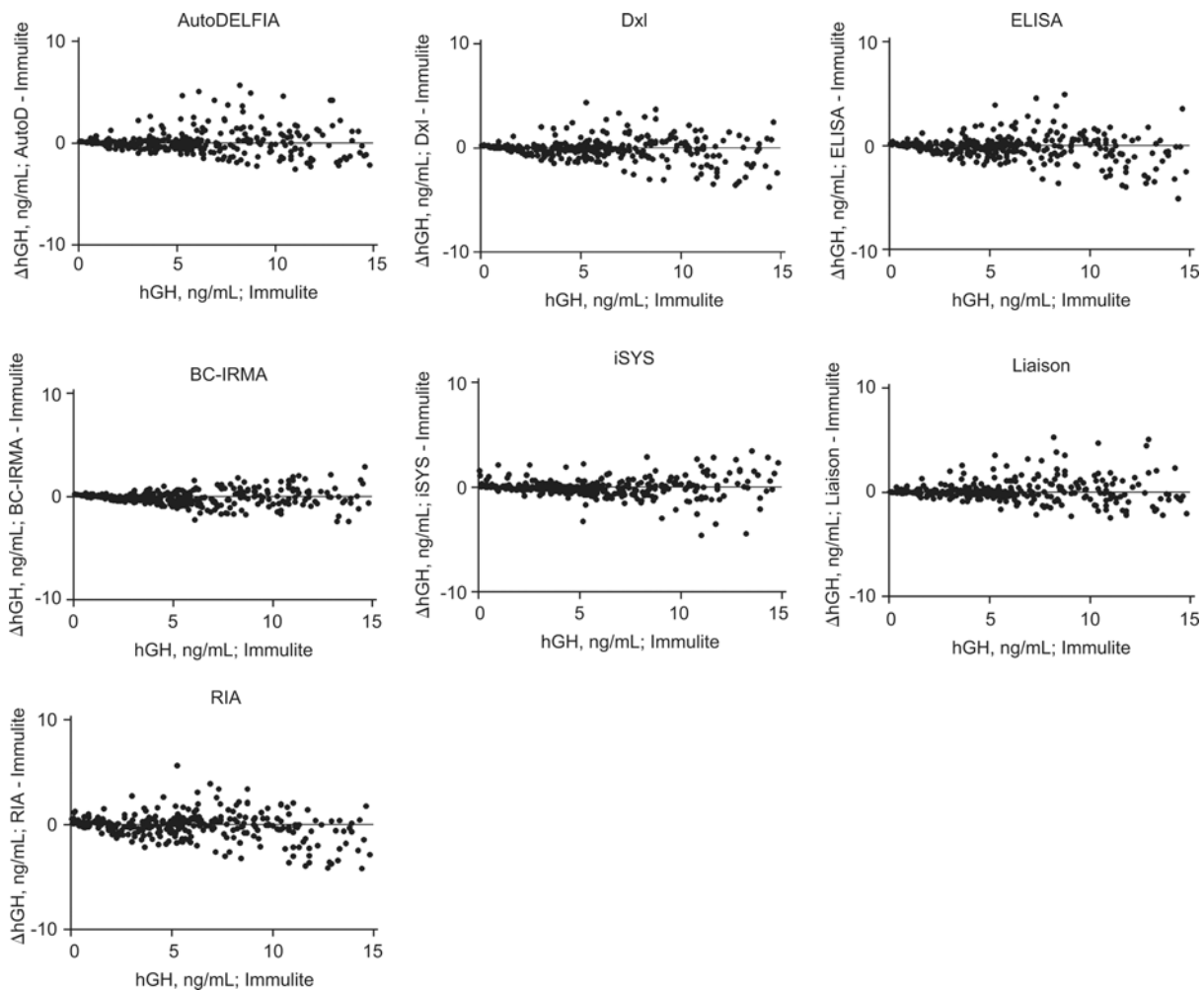
Figure 3 Passing-Bablok residual plots for the comparison between individual assays and the results of the IMMULITE 2000.

the relative standard errors of e.g., the diagnostically important 80th and 90th percentile were moderate in size for our sample size (between 2.7% and 5.6% depending on the assay), and the Spearman correlation was relatively high between all assays. Compared to the linear models, we demonstrated a comparable reduction of the CV from 24.3% to 11.4%. Moreover, the degree of comparability in the diagnostic decision was even better than for the linear procedures.

Despite such promising results, some restrictions have to be considered for the interpretation of our data: first, we developed a procedure for the harmonization of hGH concentrations, not a procedure for improving the accuracy in the measurement of hGH. For accurate measurement of hGH, fundamentally different approaches are needed. Non-immunological methods, such as mass spectrometry could offer an alternative as reference method or analytical routine platform in the future (16). Second, our suggestions cannot fully exclude a bias of comparability for the diagnosis of GHD. The only way to avoid this would be to measure hGH in one or more central laboratories using a single and well-suited analytical method as currently performed in Israel. Samples could be shipped to such central labs even as filter paper blood spots (17). However, such an agreement is currently unrealistic in larger countries. Third, the disadvantage of harmonizing methods is that they have to be adjusted for substitutions of antibody or other kit components resulting in changes of assay results for any individual method. In addition, we had to perform up to four freeze-thaw cycles before hGH was measured in some samples. However, as it is known that hGH is stable for five freeze-thaw cycles, we did not expect any bias in our results (18).

We conclude, that recent recommendations for the standardization of hGH immunoassays still lead to assay-dependent diagnostic decisions in approximately on one-third of patients. The use of statistical harmonization methods to adjust data, especially of quantile transformation, improves the comparability of hGH from different methods and allows a distinct reduction in the assay dependency of the laboratory based diagnosis of GHD in these subjects. The question of whether or not procedures for harmonization of hGH may be applied in routine laboratory analytics, has to be discussed by the responsive pediatric endocrine societies with regard to a critical balance between methodological advantages and disadvantages.

Acknowledgments

Conflict of interest statement

Annex

R-script:

```
#Example for application of quantile normalization

N<-300  #number of samples
#define gold standard -----------------------
gold<-exp(rnorm(N,sd=0.2))
#define alternative methods --------------------
x1<-gold*exp(rnorm(300,mean=0.3,sd=0.1))
#constant bias
x2<-gold^2*exp(rnorm(300,mean=0,sd=0.1))
#different scale 1
x3<-tanh(gold)*exp(rnorm(300,mean=0,sd=0.05))
#different scale 2

#plots before normalization
par(mfrow=c(2,3))
plot(gold,x1)
title(''before normalization'')
abline(0,1)
plot(gold,x2)
title(''before normalization'')
abline(0,1)
plot(gold,x3)
title(''before normalization'')
abline(0,1)

#data
x<-cbind(x1,x2,x3)

#quantile matrix, increase number of rows if needed
q<-matrix(0,nrow=1001,ncol=length(x[1,]))
qn<-x

#estimate quantiles
for (i in 1:length(x[1,]))
{
q[,i]<-quantile(x[,i],probs=seq(0,1,0.001), type=4)
#adjust seq in accordance to q, use adequate type of quantile estimation
}

qgold<-quantile(gold,probs=seq(0,1,0.001), type=4)

#normalization
for(j in 1:length(x[1,]))
{
ec<-ecdf(x[,j])
for (i in 1:length(x[,1]))
```

```
{
qn[i,j] <-qgold[round(ec(x[i,j]),digits=3)*1000+1]
}}

x1n <-qn[,1]
x2n <-qn[,2]
x3n <-qn[,3]

#plots after normalization
plot(gold,x1n)
title(''after normalization'')
abline(0,1)
plot(gold,x2n)
title(''after normalization'')
abline(0,1)
plot(gold,x3n)
title(''after normalization'')
abline(0,1)
```

References

1. Wales JK, Wit JM, Rogol AD. Pediatric endocrinology and growth. London: Saunders; 2003.
2. Binder G. Guideline of the german society of endocrinology/study group pediatric endocrinology/diabetology. Diagnostics of growth hormone deficiency in childhood and adolescence. Extensive version. See: http://www.uni-duesseldorf.de/AWMF/ll/. As at: Oct 29, 2008.
3. GH Research Society. 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;85:3990–3.
4. RfB: Ringversuche/HP – Hormonbestimmungen Gruppe 2/2009/Serum. Interlaboratory tests of the united german society of clinical chemistry and laboratory diagnostics. HP1/09, HP2/09, HP3/09 und HP4/09. See: http://www.dgkl-rfb.de. As at: Jun 19, 2010.
5. Strasburger CJ, Bidlingmaier M. How robust are laboratory measures of growth hormone status? Horm Res 2005;64:1–5.
6. Hauffa BP, Lehmann N, Bettendorf M, Mehls O, Dörr HG, Partsch CJ, et al. Central reassessment of GH concentrations measured at local treatment centers in children with impaired growth: consequences for patient management. Eur J Endocrinol 2004;150:291–7.
7. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part I. J Clin Chem Clin Biochem 1983;21:709–20.
8. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. Bioinformatics 2003;19:185–93.
9. Ross A. 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.
10. Tanaka T, Tachibana K, Shimatsu A, Katsumata N, Tsushima T, Hizuka N, et al. A nationwide attempt to standardize growth hormone assays. Horm Res 2005;64:6–11.
11. Ebdrup L, Fisker S, Sorensen HH, Ranke MB, Orskov H. Variety in growth hormone determinations due to use of different immunoassays and to the interference of growth hormone-binding protein. Horm Res 1999;51:20–6.
12. 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.
13. Baumann GP. Growth hormone isoforms. Growth Horm IGF Res 2009;19:333–40.
14. Hansen TK, Fisker S, Hansen B, Sørensen HH, Christiansen JS, Jørgensen JO, et al. Impact of GHBP interference on estimates of GH and GH pharmacokinetics. Clin Endocrinol (Oxf) 2002;57:779–86.
15. Amed S, Delvin E, Hamilton J. Variation in growth hormone immunoassays in clinical practice in canada. Horm Res 2008;69:290–4.
16. Thevis M, Bredehöft M, Kohler M, Schänzer W. Mass spectrometry-based analysis of IGF-1 and hGH. Handb Exp Pharmacol 2010;195:201–7.
17. Langkamp M, Weber K, Ranke MB. Human growth hormone measurement by means of a sensitive ELISA of whole blood spots on filter paper. Growth Horm IGF Res 2008;18:526–32.
18. Livesey JH, Hodgkinson SC, Roud HR, Donald RA. Effect of time, temperature and freezing on the stability of immunoreactive LH, FSH, TSH, growth hormone, prolactine and insulin in plasma. Clin Biochem 1980;13:151–5.