

Classification of Patients With GH Disorders May Vary According to the IGF-I Assay

Maria Mavromati, Emmanuelle Kuhn, Hélène Agostini, Sylvie Brailly-Tabard, Catherine Massart, Marie-Liesse Piketty, Armelle Arnoux, Jacques Young, Jean-Claude Souberbielle, Philippe Chanson

▶ **To cite this version:**

Maria Mavromati, Emmanuelle Kuhn, Hélène Agostini, Sylvie Brailly-Tabard, Catherine Massart, et al.. Classification of Patients With GH Disorders May Vary According to the IGF-I Assay. Journal of Clinical Endocrinology and Metabolism, 2017, 102 (8), pp.2844-2852. 10.1210/jc.2017-00202 . hal-01579404

HAL Id: hal-01579404
https://univ-rennes.hal.science/hal-01579404

Submitted on 10 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classification of patients with GH disorders may vary according to the IGF-I assay

Short title: Classification of patients and IGF-I assay

Maria Mavromati, Emmanuelle Kuhn, Hélène Agostini, Sylvie Brailly-Tabard, Catherine Massart, Marie-Liesse Piketty, Armelle Arnoux, Jacques Young, Jean-Claude Souberbielle, and Philippe Chanson.

Service d'Endocrinologie et des Maladies de la Reproduction and Centre de Référence des Maladies Endocriniennes Rares de la Croissance (M.M., E.K., J.Y., P.C.), Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires Paris-Sud, Hôpital de Bicêtre, Unité de Recherche Clinique (H.A, A.A.), and Service de Génétique Moléculaire, Pharmacogénétique et Hormonologie (S.B-T.), Le Kremlin-Bicêtre, F94275, France; Inserm 1185 (E.K., S.B-T., J.Y., P.C.), Fac Med Paris Sud, Université Paris-Saclay, Le Kremlin-Bicêtre, F-94276, France; and Laboratoire d'Hormonologie (C.M.), Centre Hospitalier Universitaire de Rennes, Centre d'Investigation Clinique Plurithématique, Inserm 1414, Hôpital Pontchaillou, Rennes, F29000, France; Service des Explorations Fonctionnelles (M-L.P., J.-C.S.), Assistance Publique-Hôpitaux de Paris, Hôpital Necker-Enfants Malades, Paris, F75015, France

Corresponding author and reprint requests:

Philippe Chanson, MD

Accepted manuscript

Service d'Endocrinologie et des Maladies de la Reproduction,

Hôpital de Bicêtre, 78 rue du Général Leclerc, 94275 Le Kremlin-Bicêtre, France

E-mail: philippe.chanson@bct.aphp.fr

This study has been registered as **ClinicalTrials.gov ID NCT01831648**

Abbreviations: BMI, body mass index; IGFBP, IGF binding protein; SDS, standard deviation score.

Key terms: IGF-I, Z-score, SD score, normative data, reference range, normal healthy population, acromegaly, growth hormone deficiency

Disclosure: The authors have no conflicts of interest to disclose

Word count (for text only): 3245

Abbreviated title: Patient classification with different IGF-I assays

Accepted manuscript

Summary

Context: IGF-I measurement is essential for the diagnosis and management of GH disorders. However, patient classification may vary substantially according to the assay technique.

Objective: We compared individual patient data and classifications obtained with six different IGF-I assay kits in a group of patients with various GH disorders.

Materials and Methods: In this cross-sectional study (ClinicalTrials.gov ID NCT01831648), we measured IGF-I with six immunoassays (iSYS, Liaison XL, Immulite, IGFI RIACT, MEDIAGNOST Elisa, MEDIAGNOST RIA) in 102 patients with active or treated acromegaly or GH deficiency. IGF-I normative data previously established for the same six assay kits were used to classify the patients (high, low or normal IGF-I levels), using both raw data and standard-deviation scores (SDS). Pair-wise concordance between assays was assessed with Bland-Altman plots and with the percentage of observed agreement and the weighted Kappa coefficient for categorized IGF-I SDS.

Results: We observed marked variability both across each individual's IGF-I raw data and across IGF-I SDS values obtained with each of the six immunoassays. Pairwise concordance between assay values, as assessed with the weighted kappa coefficient, ranged from 0.50 (moderate) to 0.81 (excellent).

Conclusion: Even when using normative data obtained in the same large population of healthy subjects and when using calculated IGF-I SD scores, agreement among IGF-I assay methods is only moderate to good. Differences in assay performance must be taken into account when evaluating and monitoring patients with GH disorders. This argues for the use of the same IGF-I assay for a given patient throughout follow-up.

Accepted manuscript

Introduction

Insulin-like growth factor I (IGF-I) measurement is of crucial importance for the diagnosis of acromegaly and growth hormone deficiency (GHD), as well as for treatment monitoring (1). The Endocrine Society clinical practice guidelines for acromegaly, and the Acromegaly Consensus Group, recommend IGF-I measurements rather than random GH values for diagnosis and treatment goals (2,3). In patients with GHD, IGF-I is also crucial for monitoring GH replacement therapy and for adjusting the GH dosage (4).

Accurate measurement of IGF-I is a complex issue, as the results depend on the type of analytical method. This variability can be attributed to differences in the calibration material, the epitope specificity of the different antibodies, and interference with IGF-I binding proteins (1,5). A universal calibrator is crucial for assay standardization. A recent consensus statement on the evaluation and standardization of IGF-I assays recommends the IS 02/254 WHO reference standard, a >97%-pure recombinant standard that has been well characterized by the NIBSC (6).

Even if they give different results, one would expect two different IGF-I assays to classify a given patient in the same way (high, normal or low values). However, even when using kits that are calibrated against the same international standard, and the same method to remove IGF-I binding proteins, patient classification in terms of IGF-I categories remains variable (7-10). We suspected that a potential reason for these discrepancies was the use of different reference values. Indeed, it is difficult to establish IGF-I normative data, as they depend on the choice of healthy reference population (6,11,12). While IGF-I values depend on many factors, such as gender, age, nutritional status, treatments (especially hormonal medications), diabetes and

Accepted manuscript

renal and hepatic failure, normative data used for the different IGF I kits were not obtained in the same, apparently healthy population. Furthermore, the distribution of IGF-I levels in healthy populations is non Gaussian, and transformations are thus necessary to obtain normal distributions and to calculate standard deviation scores (SDS). This prompted us to conduct the VARIETE study (VAleurs de Référence de l'IGF-I Et Transformation En Z-score) in order to establish normative reference values for six IGF-I immunoassays in the same healthy adult population, using the same statistical method to calculate SDS (13). We postulated that this would help to longitudinally assess disease control in a given patient using the IGF-I SDS, even if IGF-I was measured with more than one assay during follow up.

In the present study, we measured IGF-I with the same six kits in 102 patients with acromegaly or GH deficiency, and used the age- and gender-adjusted normative reference values from the VARIETE study to compare the raw data and SDS values obtained for each patient with each assay. We thus determined whether the patients' classifications were concordant.

Subjects and Methods

Study population

One hundred two patients (57 men and 45 women) belonging to the cohort of Service d'Endocrinologie et des Maladies de la Reproduction of Hôpitaux Universitaires Paris-Sud (Bicetre Hospital), Le Kremlin-Bicêtre, France, were enrolled in the study between December 2013 and March 2014. Fifty-six patients had acromegaly: 32 had a blood sample taken at diagnosis (n=11) or after incomplete surgery and before initiation of medical treatment (n=21); 24 patients were sampled during follow-up on medical treatment (cabergoline alone, n=1; somatostatin analog alone, n=10;

Accepted manuscript

pegvisomant alone, n=9; somatostatin analog and cabergoline, n=3; somatostatin analog and pegvisomant, n=1) but with variable disease control (either because of treatment modification, reinforcement or titration, or because they were resistant to medical treatment). Diagnosis of acromegaly was based on clinical criteria, unsuppressed GH in the oral glucose tolerance test, IGF-I elevation, and imaging or histologic proof of a somatotroph pituitary adenoma after surgery (2,14,15). Fourteen patients had GHD, either confirmed by a serum GH peak less than 5 μg/L after the insulin tolerance test (six patients) or strongly suggested by deficiencies in at least three other pituitary functions (4). Another 32 patients had other pituitary or endocrine disorders and were tested for suspected acromegaly or GHD. The patients' characteristics are summarized in **Table 1**. Each patient had a clinical examination and personal medical history, and was sampled at 8:00 AM after an overnight fast. One patient had serial IGF-I measurements with the six IGF-I assays (at diagnosis, after pituitary surgery and on medical treatment with somatostatin analogues). All the patients gave their written informed consent to participate in the study, which was approved by the Paris-Sud Ethics committee.

In each patient, IGF-I values were measured with the six assay kits (see below) used in the recently published VARIETE study (13). The main characteristics of the assays are shown in **Supplementary Table 1**.

Normative reference range

The normative reference data that we used to classify patients as having "normal", "high" or "low" IGF-I levels were obtained in the VARIETE study (13). In brief, this study was a cross-sectional, multicenter (24 centers), French nationwide cohort study (ClinicalTrials.gov ID NCT01831648) designed to develop reference normative sex- and age-adjusted IGF-I data for the adult general population, for each of the different

Accepted manuscript

assay techniques widely used in everyday clinical practice in France. The objective of this study was also to propose formulas for calculating IGF-I SDS, taking into account the non normal distribution of IGF-I levels in the healthy population. The study population consisted of 911 subjects (470 males), comprising 101, 118, 99, 98, 103, 102, 108, 97 and 85 subjects in the 18 –20, 21–23, 24 –26, 27–29, 30 –39, 40 –49, 50–59, 60–69, and 70–89 year age groups, respectively. Serum IGF-I was measured with the following 6 assay kits: iSYS (IDS; Boldon, UK), Liaison XL (Diasorin; Saluggia, Italy), Immulite (Siemens; Erlangen, Germany), IGFI RIACT (CIS BIO; Gif sur Yvette, France), MEDIAGNOST Elisa and MEDIAGNOST RIA (MEDIAGNOST; Reutlingen, Germany). IGF-I values were then matched in 3-year groups between 18 and 30 years of age and 10-year groups between 30 and 90 years, and mean and median values as well as the 2.5^{th} and 97.5^{th} percentiles were calculated. For each gender and age category, the distribution of measurements was normalized by means of sex- and age-specific Cox-Box power transformation, in order to calculate SDS. As men and women had significantly different IGF-I levels, curves were constructed separately using the LMS method.

The VARIETE study thus established age- and sex-specific adult normative data for the six commercial IGF-I assays, including the range of values from the 2.5th to the 97.5th percentile in mass units, and provided a formula for calculating SDS. A calculator available online (http://ticemed_sa.upmc.fr/sd_score/) or as an app (IGF-I_SD_score) downloadable for Android from Google Play and for iOS from Apple Store (free of charge) yields individual IGF-I SDS after entering the name of the assay, the individual's IGF-I value obtained with the assay, and the sex and age of the individual.

Statistics

Accepted manuscript

Data were analyzed with SAS software (Statistical Analysis System, version 9.4, SAS Institute, Cary, N.C., USA). We used scatter plots and Bland-Altman plots in order to assess pairwise concordance between assays, both for IGF-I raw values and SDS values. We classified the IGF-I results in three categories: high (SDS > +2), normal (SDS between -2 and +2) and low (SDS <-2), and evaluated pairwise agreement by means of the linearly weighted Kappa coefficient.

To interpret the kappa coefficient, we used the Fermanian scale (16,17), with kappa values >0.80, between 0.61 and 0.80, between 0.41 and 0.60, between 0.21 and 0.40, between 0.01 and 0.20, and <0.01 signifying almost perfect, substantial, moderate, fair, slight and poor agreement, respectively.

Results

Variability of individual IGF-SDS values according to the IGF-I assay

Variability between each individual's IGF-I SDS obtained with each of the 6 immunoassays is illustrated in **Figure 1** for the 57 male patients and the 45 female patients. Many patients were inconsistently classified, particularly when IGF-I values were close to the reference range.

In six prospectively followed patients with acromegaly, IGF-I was measured on three occasions (at diagnosis, after surgery and at follow up, generally under medical treatment) with between three and six of the IGF-I assays **(Figure 2)**. With the exception of one patient in whom two of the three assays used at diagnosis gave a high IGF-I SDS, IGF-I SDS were generally concordant in the elevated levels. In three out of six patients with borderline IGF-I SDS after surgery was either normal or moderately elevated, suggesting that the patient had persistent active disease. At

Accepted manuscript

follow up under treatment, when IGF-I SDS was borderline, some assays classified the patient as "controlled", although others gave a low SDS.

Percentages of patients classified as having normal, high or low IGF-I levels in the different IGF-I assays

The percentages of patients classified as having high (>+2), normal (between -2 and +2) and low (<-2) SDS values are shown in **Figure 3**. The ISYS and MEDIAGNOST RIA kits classified fewer patients as having "normal" levels (33% and 35%, respectively, *versus* 46-49% for the other assays) and, on the contrary, more patients as having "high" IGF-I values (54% and 51%, respectively, *versus* 30-39%). On the other hand, IGF-I RIACT and Liaison XL were more likely to classify the patients as having "low" IGF-I levels (20% and 23%, respectively, *versus* 13-16%).

Pairwise correlations between raw data and Z-scores obtained with the six IGF-I immunoassays

The results obtained with each IGF-I assay were compared with those obtained with each of the other five assays. Scatter plots and Bland-Altman plots based on raw values and SDS for each pair of assays are shown in **Supplemental Figure 1.**

Two examples of interassay comparisons are shown in **Figure 4**. The results obtained with iSYS and Mediagnost RIA were in good overall agreement, with no significant bias on Bland-Altman plots (Figure 4, A–D). Indeed, the discrepancy around the mean difference (average difference) line was quite stable when the average value increased, without very wide limits of agreement, and with consistent variability across the graph. In contrast, the results obtained with LIAISON XL and iSYS were not in good agreement, as the mean difference line was clearly different from zero and as iSYS tended to overestimate IGF-I values by comparison with Liaison XL, an

Accepted manuscript

effect that was accentuated as the average value increased, especially for raw data (**Figure 4, E–H).**

Pairwise assay concordances (weighted Kappa coefficient) for categorized IGF-1 SDS values are shown in Table 2. The best concordance was found between iSYS and MEDIAGNOST RIA, with a kappa coefficient of 0.81. Very good agreement was also observed between Immulite and MEDIAGNOST Elisa (kappa coefficient 0.77), MEDIAGNOST Elisa and IGF-I RIACT (kappa coefficient 0.77), and Immulite and Liaison XL (kappa coefficient 0.76), MEDIAGNOST Elisa and RIA (kappa coefficient 0.76), as well as between IGFI RIACT and iSYS or MEDIAGNOST RIA (kappa coefficient 0.71). The poorest concordance was observed between iSYS and Liaison XL (kappa coefficient 0.50), MEDIAGNOST RIA and Liaison XL (kappa coefficient 0.51), and iSYS and Immulite (kappa coefficient 0.55) (**Table 2**).

When we limited the assessment of concordance to the group of patients with acromegaly (n=56), the best pairwise agreement was again between iSYS and Mediagnost RIA, with a weighted kappa coefficient of 0.81, whereas the worst agreement was between Liaison XL and iSYS, and between Immulite and Mediagnost Elisa (kappa coefficient 0.41 for both).

Concordance between assays according to IGF-I SDS classes (high, normal, low).

We analyzed concordance according to IGF-I SDS classes (high, > +1.96, « normal », between -1.96 and +1.96, low, <1.96). The three classes were those obtained initially with Immulite assay. Due to the small numbers, it was not always possible to calculate kappa values for all comparisons. Thus we also calculated the concordance in terms of similar classification (as high, normal or low values) between assays. The results are indicated as the ratio of concordant to total results in the **Supplemental Tables 2-7**.

Accepted manuscript

For low values (SDS <1.96), assays are relatively concordant with only one or two patients (out of 11 to 14) who are discordantly classified by two assays (XLD-Liaison and iSYS, iSYS and Immulite, XLD-Liaison and Mediagnost Elisa, Immulite and Mediagnost Elisa).

For "normal" IGF-I SDS (between -1,96 and +1.96), concordances (as assessed by kappa values) are generally weak or poor. In general at least six patients out of around 40 are misclassified according to the assay which is used.

For high IGF-I SDS (>1.96), numbers used for comparisons are variable (n=16, 24, 31). In the majority of cases, only one or two assays give different classification. There are more than 3 misclassified patients when comparing XLD-Liaison and ISYS, XLD-Liaison and Immulite, XLD-Liaison and Mediagnost Elisa, XLD-Liaison and Mediagnost RIA, Mediagnost RIA and IGF-IRIACT. Finally, when assays give concordant results, they are more often in the high values of the techniques.

Discussion

Our results show significant variability among six commercial immunoassays for the determination of individual IGF-I SDS values and IGF-1 classification of 102 patients with various GH disorders, despite the use of normative reference intervals obtained, for each of the six assays, in the same, large, well-selected population of healthy French adults (13), as recommended by the Consensus Group on the Standardization and Evaluation of GH and IGF-I Assays (6).

Reliable normative reference intervals are necessary for the diagnosis of acromegaly and GHD, for the follow-up of patients with GH disorders, and for the detection of remission and recurrence of GH-related diseases. In 2011, a consensus statement on the standardization and evaluation of GH and IGF-I assays proposed the use of the

Accepted manuscript

international recombinant IGF-I calibration standard preparation 02/254, and emphasized the importance of antibody specificity, quality control analysis, and the elimination of interference with binding proteins. It also emphasized the importance of obtaining normative data based on a random selection of individuals from the background population, representing all age groups, after exclusion of individuals with poorly controlled diabetes, renal or hepatic failure, or taking medications (such as estrogen) that could affect outcome.

Based on this consensus statement, Bidlingmaier et al. published normative data for the iSYS IGF-I assay obtained in a cohort of 15 014 healthy subjects (18), while we recently proposed IGF-I reference intervals obtained with six widely used immunoassays in the same population of 911 healthy French adults aged from 18 to 92 years, as per the consensus recommendations (13). The inclusion criteria were strict, with careful clinical evaluation, a medical history-taking that included ongoing treatments, and exclusion of subjects receiving steroid hormones. In addition, separate curves were constructed for each sex, in view of significantly different IGF-I levels between men and women. Normative data ranged between percentiles 2.5 and 97.5 and were reported in mass units and SDS. Nevertheless, although we ensured the same pre-analytical conditions for all six immunoassays, and although four of the six assays were calibrated against the same international reference standard 02/254, concordance across the assays remained variable, both for raw data and IGF-I SDS (13).

To extend the results of our study of healthy individuals to the clinical setting, we created a group of patients of both sexes (57 males and 45 females) encompassing the whole spectrum of serum IGF-I levels, from very low (severe GHD) to very high (highly active acromegaly), representing the everyday practice of laboratories

Accepted manuscript

involved in IGF-I measurement. We therefore analyzed the concordance between the results obtained with each of the six assays in each of the 102 patients.

Pairwise agreement between the assays ranged from moderate to excellent. The best concordance was observed between iSYS and Mediagnost RIA. These two immunoassays, calibrated against the same international standard 02/254, classified fewer patients than the other four assays as "normal", and more patients as having "high" IGF-I serum levels. In the VARIETE study, the largest intercentile intervals and highest absolute values (in µg/L) were obtained with Immulite and IGF-I RIACT, the two immunoassays calibrated against the old standard IRP 87/518 (13). However, when using SDS in the present group of patients, instead of absolute mass values, these two immunoassays classified similar percentages of patients as having "normal" IGF-I levels as the Liaison XL assay and Mediagnost Elisa. Moreover, the three automated assays (Immulite, Liaison XL and iSYS) did not show excellent pairwise concordance: the pairs Liaison XL – iSYS and Immulite - iSYS exhibited only moderate agreement (kappa coefficients 0.50 and 0.55, respectively), and only the pair Immulite – Liaison XL showed substantial agreement (kappa coefficient 0.76).

This lack of concordance between certain assays has already been reported (7-10): one possible explanation was that the populations used to establish normative data were different, or that the quality of these normative data was suboptimal (too few patients studied, particularly in certain age ranges; bias, failure to select healthy subjects, with regard to concurrent treatments or medical conditions interfering with IGF-I measurement, etc.) (5,6,11,12,19). This is why we used the same large healthy population to establish normative data for the six immunoassays used here. Despite this, discordant results persisted between some assays, some pairs being clearly more discordant than others.

Accepted manuscript

Another possible explanation for the lack of concordance is a difference in the technical procedure (5,6,12,19). In the present study, the preanalytic procedure was exactly the same, and only the analytic procedure therefore differed. As underlined in our study of healthy volunteers, in which we also found such discordances (13), the most plausible explanation lies in the capacity of the assay to remove IGFBP, and the specificity and performance of the antibody. This may be particularly true for high IGF-I values, which are usually associated with high levels of IGFBP3.

These results confirm that, even when using normative values established in the same population of healthy subjects, IGF-I results obtained with different assays in a given individual, whether healthy (as in the VARIETE study), or having a GH-related disorder (as in the present study), are sometimes very different, potentially leading to patient misclassification.

It is crucial to understand the reasons behind differences in the results of commercial IGF-I immunoassays. Assays with similar characteristics must be used for the follow-up of a given patient. Assays that tend to overestimate or underestimate IGF-I values by comparison with other techniques must be clearly identified. Liquid chromatography (LC) tandem mass spectometry (MS) may or may not prove to be a valid alternative (20,21). Reference intervals obtained with LC-MS seem very similar to those obtained with immunoassays (22). However, tandem LC-MS is a time-consuming and complex method that requires expensive machines and technical expertise to control the many variables that can influence the results (23). Thus, despite their limitations, immunoassays will continue to be widely used, at least in the near future.

In conclusion, IGF-I levels obtained with six commercial IGF-I immunoassays widely used in clinical practice, and calculated IGF-I SDS, were quite variable in patients

Accepted manuscript

with GH-related disorders, despite the use of normative reference intervals obtained in the same large, well-defined population of French healthy adults. It is not possible, according to the results of this study, to recommend one assay or the other. From a practical point of view, very high levels or very low levels of IGF-I are generally concordant, whatever the assay which is used and classification of patients as having active acromegaly or severe GH deficiency is generally similar. On the contrary, when IGF-I levels are borderline, classification may differ from one assay to the other. This requires caution in interpretation of borderline IGF-I levels. In this context we do not recommend to follow a patient and to take therapeutic decisions based on IGF-I SDS calculated with one assay one day and another assay another day ! On the contrary, a given patient should preferably be monitored with the same IGF-I assay.

Acknowledgments

The study was supported by a grant from Programme Hospitalier de Recherche Clinique, French Ministry of Health, N° P081216 / IDRCB 2009-A00892-55, and by a grant from Fond National Suisse, P1GEP3_155694, to Dr Maria Mavromati. The Authors want to thank Céline Piedvache for her help in complementary statistical analysis.

Accepted manuscript

Legends of Figures

Figure 1.

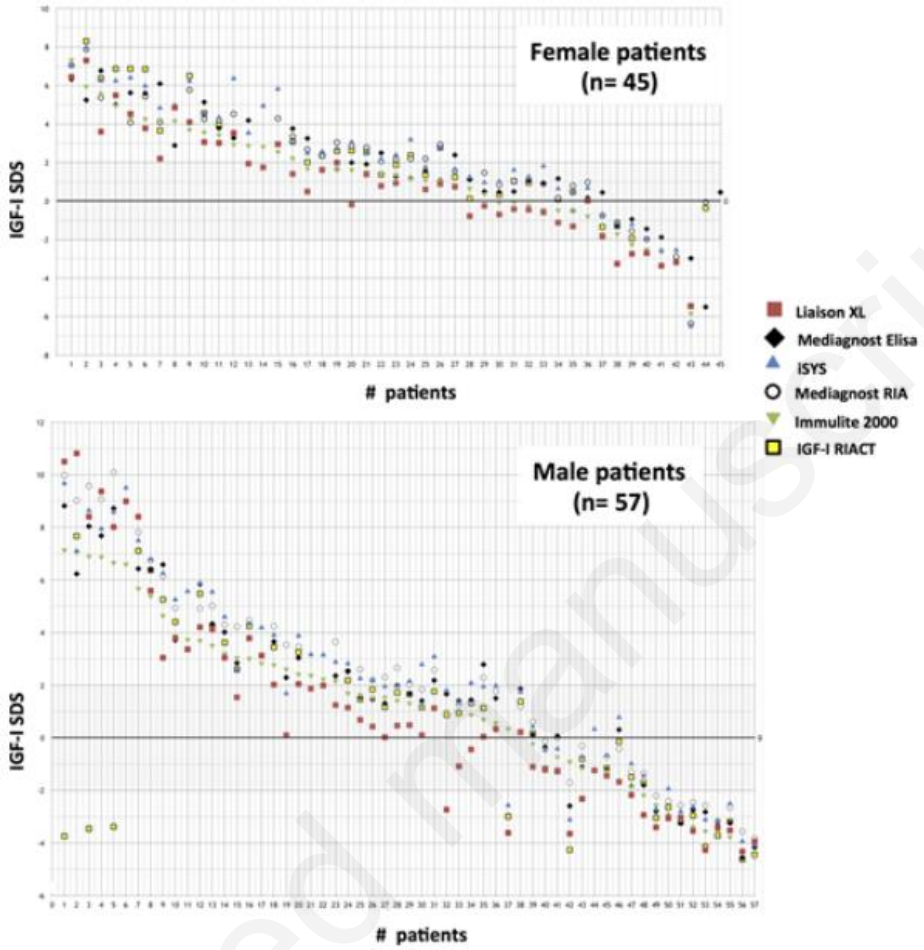


Figure 1

Variability among the six immunoassay SDS values for each of the 45 female patients (upper panel) and the 57 male patients (lower panel) with IGF-I disorder ranked by IGF-I Immulite 2000 decreasing value. Each assay is assigned a colored symbol.

Accepted manuscript

Figure 2.

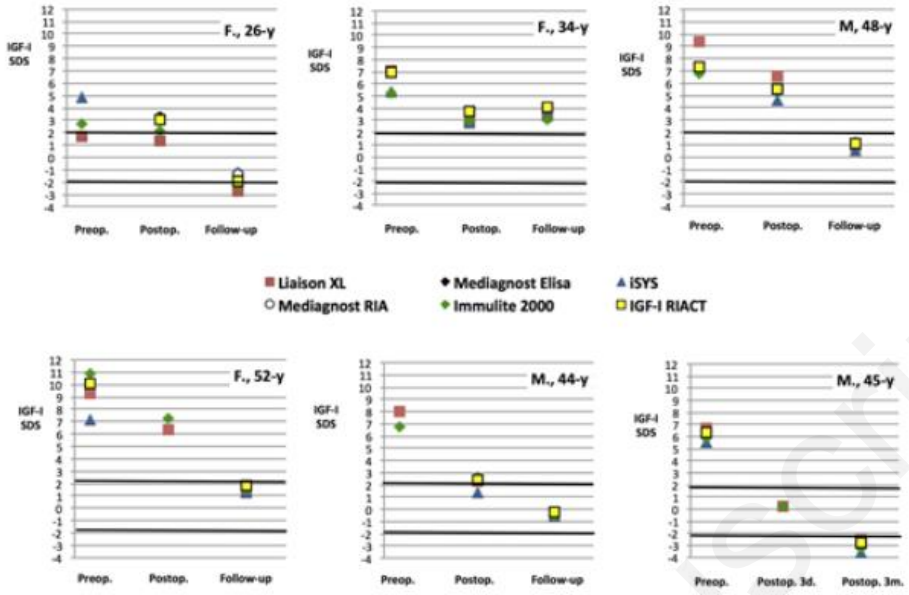


Figure 2

Variability of SDS values obtained with each of the six immunoassays in six patients with acromegaly, at three points of follow-up: diagnosis of acromegaly, immediately after surgery, and at follow up, generally under medical treatment. Horizontal lines represent the normal range from +2 to – 2 standard deviations. Each assay is assigned a different colored symbol.

Accepted manuscript

Figure 3.

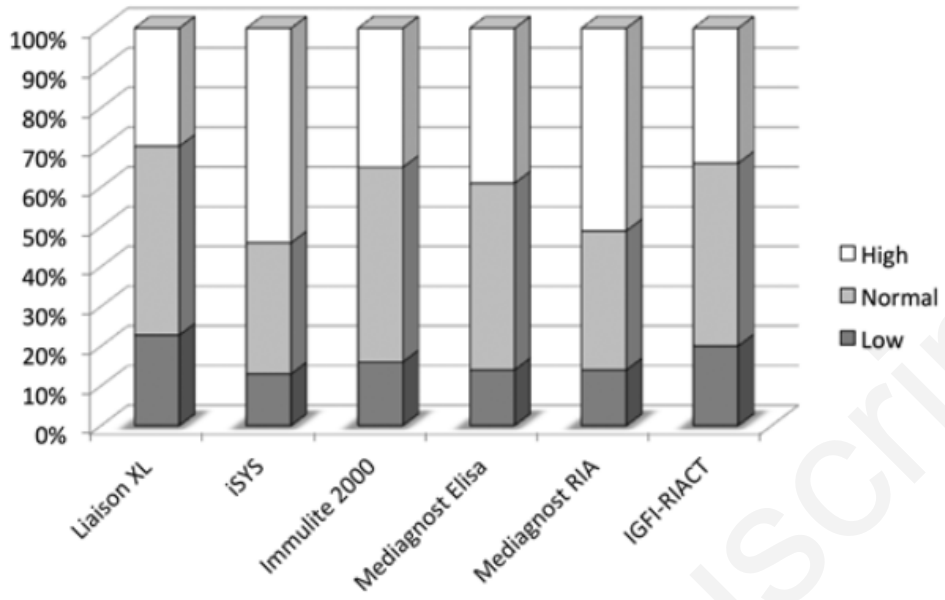


Figure 3

Percentages of patients with normal, low and high IGF-I levels according to each of the six IGF-I immunoassays.

Accepted manuscript

Figure 4.

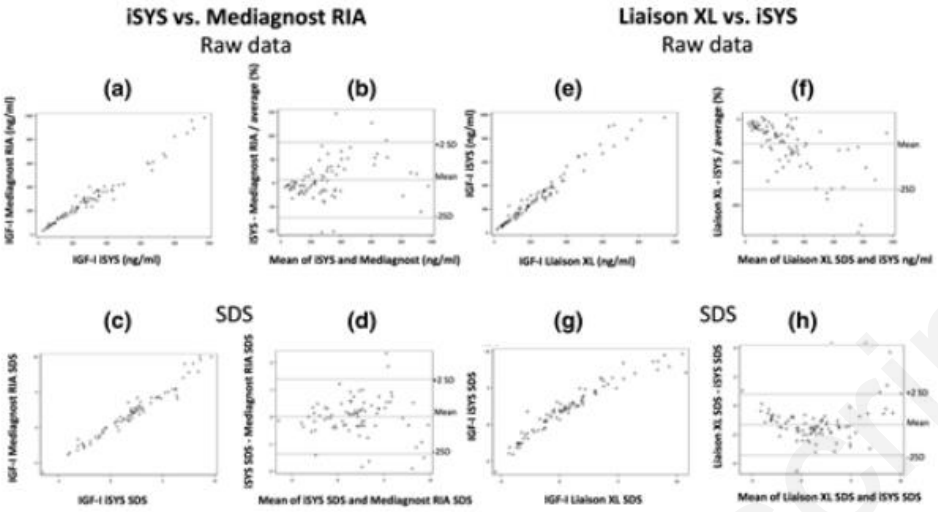


Figure 4

Comparisons between iSYS and Mediagnost RIA expressed as scatter plots (A) and Bland-Altman plots (B) for raw data, or scatter plots (C) and Bland-Altman plots (D) for SDS, showing excellent overall agreement between the two immunoassays. Comparisons between Liaison XL and iSYS expressed as scatter plots (E) and Bland-Altman plots (F) for raw data, or scatter plots (G) and Bland-Altman plots (H) for SDS, showing moderate overall agreement between these two immunoassays.

Accepted manuscript

Table 1

Characteristics of the 102 patients with various GH disorders in whom IGF-I was measured with the six IGF-I immunoassays.

Characteristics			**Males (N = 57)**	**Females (N = 45)**
Age			47.1 (19-72)	43 (24-78)
Type of GH disorder	Acromegaly (N = 56)	Under treatment	11	13
		Untreated	17	15
	GHD (N= 14)	GH-treated	0	1
		Untreated	11	2
	Suspicion of GH disorder (N = 32)		18	14

Accepted manuscript

Table 2 : Agreement between IGF-1 assay methods, expressed as weighted kappa coefficient.

Kappa coefficient	**Liaison XL**	**iSYS**	**Immulite 2000**	**Mediagnost Elisa**	**Mediagnost RIA**	**IGFI-RIACT**
Liaison XL		0.50	0.76	0.67	0.51	0.64
iSYS	0.50		0.55	0.69	0.81	0.71
Immulite 2000	0.76	0.55		0.77	0.62	0.65
Mediagnost Elisa	0.67	0.69	0.77		0.76	0.77
Mediagnost RIA	0.51	0.81	0.62	0.76		0.71
IGFI-RIACT	0.64	0.71	0.65	0.77	0.71	

Accepted manuscript

References

1. Clemmons DR. Value of insulin-like growth factor system markers in the assessment of growth hormone status. Endocrinol Metab Clin North Am 2007; 36:109-129

2. Katznelson L, Laws ER, Jr., Melmed S, Molitch ME, Murad MH, Utz A, Wass JA. Acromegaly: an endocrine society clinical practice guideline. J Clin Endocrinol Metab 2014; 99:3933-3951

3. Giustina A, Chanson P, Kleinberg D, Bronstein MD, Clemmons DR, Klibanski A, van der Lely AJ, Strasburger CJ, Lamberts SW, Ho KK, Casanueva FF, Melmed S. Expert consensus document: A consensus on the medical treatment of acromegaly. Nat Rev Endocrinol 2014; 10:243-248

4. Molitch ME, Clemmons DR, Malozowski S, Merriam GR, Vance ML, Endocrine S. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2011; 96:1587-1609

5. Frystyk J, Freda P, Clemmons DR. The current status of IGF-I assays--a 2009 update. Growth Horm IGF Res 2010; 20:8-18

6. Clemmons DR. Consensus Statement on the Standardization and Evaluation of Growth Hormone and Insulin-like Growth Factor Assays. Clin Chem 2011; 57:555-559

7. Massart C, Poirier JY. Serum insulin-like growth factor-I measurement in the follow-up of treated acromegaly: comparison of four immunoassays. Clin Chim Acta 2006; 373:176-179

Accepted manuscript

8. Pokrajac A, Wark G, Ellis AR, Wear J, Wieringa GE, Trainer PJ. Variation in GH and IGF-I assays limits the applicability of international consensus criteria to local practice. Clin Endocrinol (Oxf) 2007; 67:65-70

9. 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; 46:1776-1783

10. Varewijck AJ, Lamberts SW, van der Lely AJ, Neggers SJ, Hofland LJ, Janssen JA. The introduction of the IDS-iSYS total IGF-1 assay may have far-reaching consequences for diagnosis and treatment of GH deficiency. J Clin Endocrinol Metab 2015; 100:309-316

11. Brabant G, Wallaschofski H. Normal levels of serum IGF-I: determinants and validity of current reference ranges. Pituitary 2007; 10:129-133

12. Junnila RK, Strasburger CJ, Bidlingmaier M. Pitfalls of insulin-like growth factor-I and growth hormone assays. Endocrinol Metab Clin North Am 2015; 44:27-34

13. Chanson P, Arnoux A, Mavromati M, Brailly-Tabard S, Massart C, Young J, Piketty ML, Souberbielle JC, Investigators V. Reference Values for IGF-I Serum Concentrations: Comparison of Six Immunoassays. J Clin Endocrinol Metab 2016; 101:3450-3458

14. Melmed S. Acromegaly. In: Melmed S, ed. The Pituitary. Malden, Mas, USA: Blackwell Science Inc.; 2011:463-474.

15. Chanson P, Salenave S, Kamenicky P. Acromegaly. Handb Clin Neurol 2014; 124:197-219

Accepted manuscript

16. Fermanian J. [Measuring agreement between 2 observers: a quantitative case]. Rev Epidemiol Sante Publique 1984; 32:408-413

17. Fermanian J. [Measurement of agreement between 2 judges. Qualitative cases]. Rev Epidemiol Sante Publique 1984; 32:140-147

18. Bidlingmaier M, Friedrich N, Emeny RT, Spranger J, Wolthers OD, Roswall J, Korner A, Obermayer-Pietsch B, Hubener C, Dahlgren J, Frystyk J, Pfeiffer AF, Doering A, Bielohuby M, Wallaschofski H, Arafat AM. Reference intervals for insulin-like growth factor-1 (IGF-I) from birth to senescence: results from a multicenter study using a new automated chemiluminescence IGF-I immunoassay conforming to recent international recommendations. J Clin Endocrinol Metab 2014; 99:1712-1721

19. Clemmons DR. IGF-I assays: current assay methodologies and their limitations. Pituitary 2007; 10:121-128

20. Kay R, Halsall DJ, Annamalai AK, Kandasamy N, Taylor K, Fenwick S, Webb A, Wark G, Pleasance S, Gurnell M. A novel mass spectrometry-based method for determining insulin-like growth factor 1: assessment in a cohort of subjects with newly diagnosed acromegaly. Clin Endocrinol (Oxf) 2013; 78:424-430

21. Ketha H, Singh RJ. Clinical assays for quantitation of insulin-like-growth-factor-1 (IGF1). Methods 2015; 81:93-98

22. 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

23. Hoofnagle AN, Whiteaker JR, Carr SA, Kuhn E, Liu T, Massoni SA, Thomas SN, Townsend RR, Zimmerman LJ, Boja E, Chen J, Crimmins DL, Davies

Accepted manuscript

SR, Gao Y, Hiltke TR, Ketchum KA, Kinsinger CR, Mesri M, Meyer MR, Qian WJ, Schoenherr RM, Scott MG, Shi T, Whiteley GR, Wrobel JA, Wu C, Ackermann BL, Aebersold R, Barnidge DR, Bunk DM, Clarke N, Fishman JB, Grant RP, Kusebauch U, Kushnir MM, Lowenthal MS, Moritz RL, Neubert H, Patterson SD, Rockwood AL, Rogers J, Singh RJ, Van Eyk JE, Wong SH, Zhang S, Chan DW, Chen X, Ellis MJ, Liebler DC, Rodland KD, Rodriguez H, Smith RD, Zhang Z, Zhang H, Paulovich AG. Recommendations for the Generation, Quantification, Storage, and Handling of Peptides Used for Mass Spectrometry-Based Assays. Clin Chem 2016; 62:48-69

Accepted manuscript