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

Maria Mavromati,[1] Emmanuelle Kuhn,[1,2] Hélène Agostini,[3] Sylvie Brailly-Tabard,[2,4] Catherine Massart,[5] Marie-Liesse Piketty,[6] Armelle Arnoux,[3] Jacques Young,[1,2] Jean-Claude Souberbielle,[6] and Philippe Chanson[1,2]

[1]Service d'Endocrinologie et des Maladies de la Reproduction and Centre de Référence des Maladies Rares de l'Hypophyse, F94275 Le Kremlin-Bicêtre, France; [2]Inserm 1185, Fac Med Paris Sud, Université Paris-Saclay, F94276 Le Kremlin-Bicêtre, France; [3]Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires Paris-Sud, Hôpital de Bicêtre, Unité de Recherche Clinique, F94275 Le Kremlin-Bicêtre, France; [4]Service de Génétique Moléculaire, Pharmacogénétique et Hormonologie, F94275 Le Kremlin-Bicêtre, France; [5]Laboratoire d'Hormonologie, Centre Hospitalier Universitaire de Rennes, Centre d'Investigation Clinique Plurithématique, Inserm 1414, Hôpital Pontchaillou, F29000 Rennes, France; and [6]Service des Explorations Fonctionnelles, Assistance Publique-Hôpitaux de Paris, Hôpital Necker-Enfants Malades, F75015 Paris, France

Context: Insulinlike growth factor I (IGF-I) measurement is essential for the diagnosis and management of growth hormone (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.

Design: In this cross-sectional study, we measured IGF-I with six immunoassays 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 (SDSs). Pairwise concordance between assays was assessed with Bland-Altman plots and with the percentage of observed agreement and the weighted κ 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 κ 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 SDSs, 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. (***J Clin Endocrinol Metab*** **102: 2844–2852, 2017)**

Insulinlike 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 growth hormone (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).

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA

Received 21 January 2017. Accepted 8 May 2017.
First Published Online 12 May 2017

Abbreviations: GH, growth hormone; GHD, growth hormone deficiency; IGF-I, insulinlike growth factor I; LC-MS, tandem liquid chromatography and mass spectrometry; SDS, standard deviation score.

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 World Health Organization reference standard, a >97%-pure recombinant standard that has been well characterized by the National Institute for Biological Standards and Control (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 a healthy reference population (6, 11, 12). Although IGF-I values depend on many factors, such as sex, age, nutritional status, treatments (especially hormonal medications), diabetes, and 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 (SDSs). This prompted us to conduct the VARIETE study (VAleurs de Référence de l'IGF-I Et Transformation En z score) to establish normative reference values for six IGF-I immunoassays in the same healthy adult population, using the same statistical method to calculate SDSs (13). We postulated that this would help to longitudinally assess disease control in a given patient using the IGF-I SDSs, even if IGF-I was measured with more than one assay during follow-up.

In the current study, we measured IGF-I with the same six kits in 102 patients with acromegaly or GH deficiency, and used the age- and sex-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. Thirty-two patients had a blood sample taken at diagnosis (n = 11) or after incomplete surgery and before initiation of medical treatment (n = 21), and 24 patients were sampled during follow-up on medical treatment (cabergoline alone, n = 1; somatostatin analog alone, n = 10; pegvisomant alone, n = 9; somatostatin analog and cabergoline, n = 3; and somatostatin analog and pegvisomant, n = 1) but with variable disease control (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 underwent a clinical examination, had a personal medical history obtained and was sampled at 8:00 AM after an overnight fast. Six patients had serial IGF-I measurements with three to six IGF-I assays (at diagnosis, after pituitary surgery, and during medical treatment with somatostatin analogs). 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 later) used in the recently published VARIETE study (13). The main characteristics of the assays are shown in Supplemental 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), nationwide French cohort study (ClinicalTrials.gov no. NCT01831648) designed to develop reference normative sex- and age-adjusted IGF-I data for the adult general population for each of the different assay techniques widely used in everyday clinical practice in France. The objective of this study was also to propose formulas for calculating IGF-I SDSs, 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- to 20-, 21- to 23-, 24- to 26-, 27- to 29-, 30- to 39-, 40- to 49-, 50- to 59-, 60- to 69-, and 70- to 89-year age groups, respectively. Serum IGF-I was measured with the following six assay kits: iSYS (ImmunoDiagnostic Systems, Boldon, United Kingdom), Liaison XL (Diasorin, Saluggia, Italy), Immulite 2000 (Siemens, Erlangen, Germany), IGF-I 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 2.5th and 97.5th percentiles were calculated. For each sex and age category, the distribution of measurements was normalized by means of sex- and age-specific

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, y	47.1 (19–72)	43 (24–78)
Acromegaly (n = 56)		
Treated , n	11	13
Untreated, n	17	15
GHD (n = 14)		
GH-treated, n	0	1
Untreated, n	11	2
Suspicion of GH disorder (N = 32), n	18	14

Cox-Box power transformation to calculate SDSs. 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 SDSs. 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 the Apple Store (free of charge) yields individual IGF-I SDSs 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

Data were analyzed with Statistical Analysis System software (version 9.4; SAS Institute, Cary, NC). We used scatter plots and Bland-Altman plots 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 >+1.96), normal (SDS between –1.96 and +1.96), and low (SDS <–2), and evaluated pairwise agreement by means of the linearly weighted κ coefficient.

To interpret the κ coefficient, we used the Fermanian scale (16, 17), with κ 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-I SDS values according to the IGF-I assay

Variability between each individual's IGF-I SDS obtained with each of the six immunoassays is illustrated in Fig. 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 (Fig. 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 SDSs were generally concordant in the elevated levels. In three out of six patients with borderline IGF-I SDS after surgery, IGF-I SDS was either normal using some assays, suggesting that the patient was cured, or moderately elevated using other assays, suggesting that the patient had persistent active disease. At 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 (>+1.96), normal (between –1.96 and +1.96), and low (<–1.96) SDS values are shown in Fig. 3. The iSYS and Mediagnost RIA kits classified fewer patients as having "normal" levels (33% and 35%, respectively, *vs* 46% to 49% for the other assays) and, on the contrary, more patients as having "high" IGF-I values (54% and 51%, respectively, *vs* 30% to 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, *vs* 13% to 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 SDSs for each pair of assays are shown in Supplemental Fig. 1.

Two examples of interassay comparisons are shown in Fig. 4. The results obtained with iSYS and Mediagnost RIA were in good overall agreement, with no significant bias on Bland-Altman plots [Fig. 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 effect that was accentuated as the average value increased, especially for raw data [Fig. 4(e–h)].

Pairwise assay concordances (weighted κ coefficient) for categorized IGF-I SDS values are shown in Table 2. The best concordance was found between iSYS and Mediagnost RIA, with a κ coefficient of 0.81. Very good agreement was also observed between Immulite and Mediagnost ELISA (κ coefficient, 0.77), Mediagnost

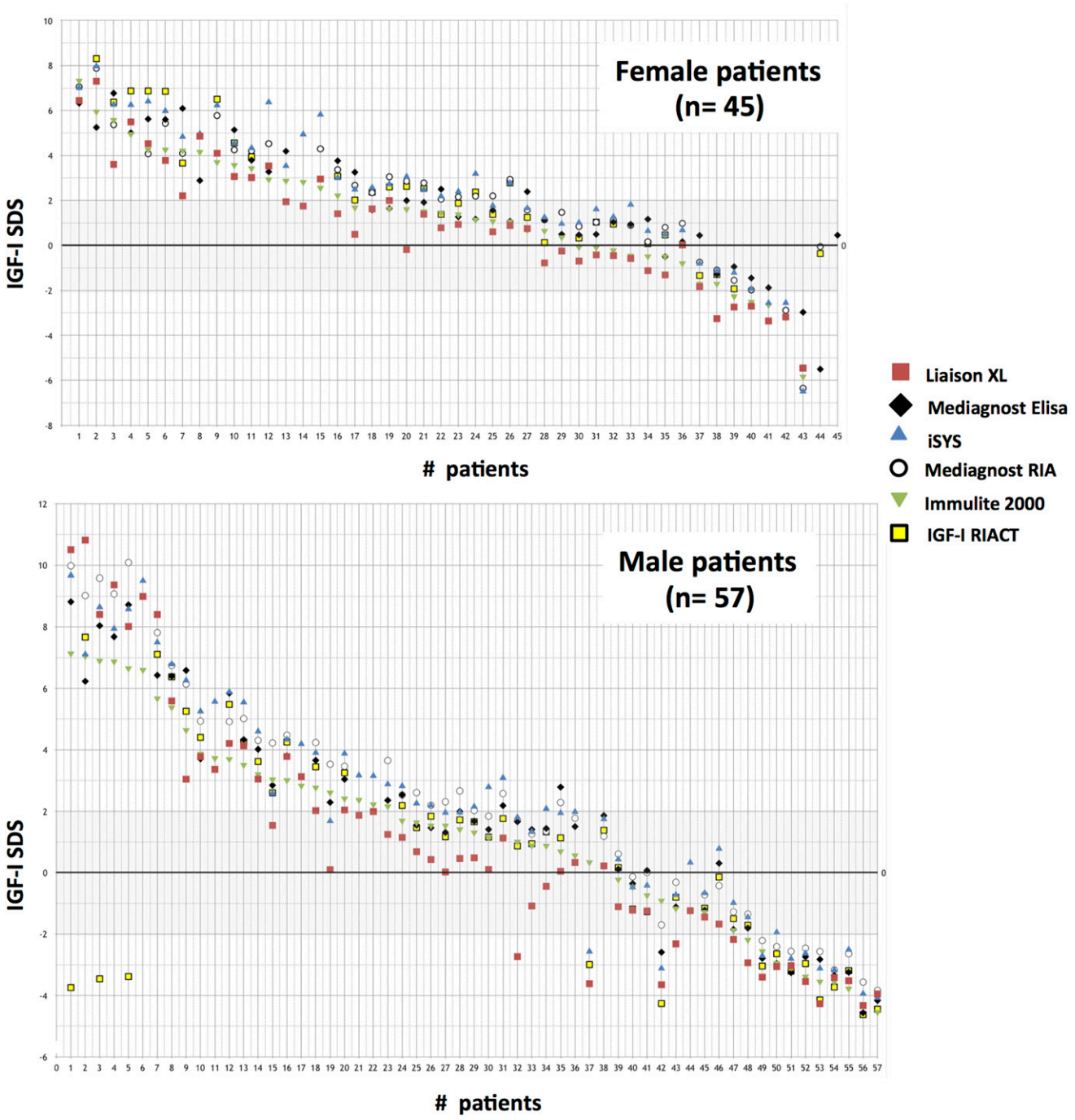

Figure 1. Variability among the six immunoassay SDS values for each of the 45 female patients (upper panel) and the 57 male patients (M; lower panel) with IGF-I disorder ranked by IGF-I Immulite 2000 decreasing value. Each assay is assigned a colored symbol.

ELISA and IGF-I RIACT (κ coefficient, 0.77), Immulite and Liaison XL (κ coefficient, 0.76), and Mediagnost ELISA and RIA (κ coefficient, 0.76), as well as between IGF-I RIACT and iSYS or Mediagnost RIA (κ coefficient, 0.71). The poorest concordance was observed between iSYS and Liaison XL (κ coefficient, 0.50), Mediagnost RIA and Liaison XL (κ coefficient, 0.51), and iSYS and Immulite (κ 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 κ coefficient of 0.81, whereas the worst agreement was between Liaison XL and iSYS and between Immulite and Mediagnost ELISA (κ coefficient, 0.41 for both).

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

We analyzed concordance according to IGF-I SDS classes (high, >1.96; normal, between –1.96 and 1.96;

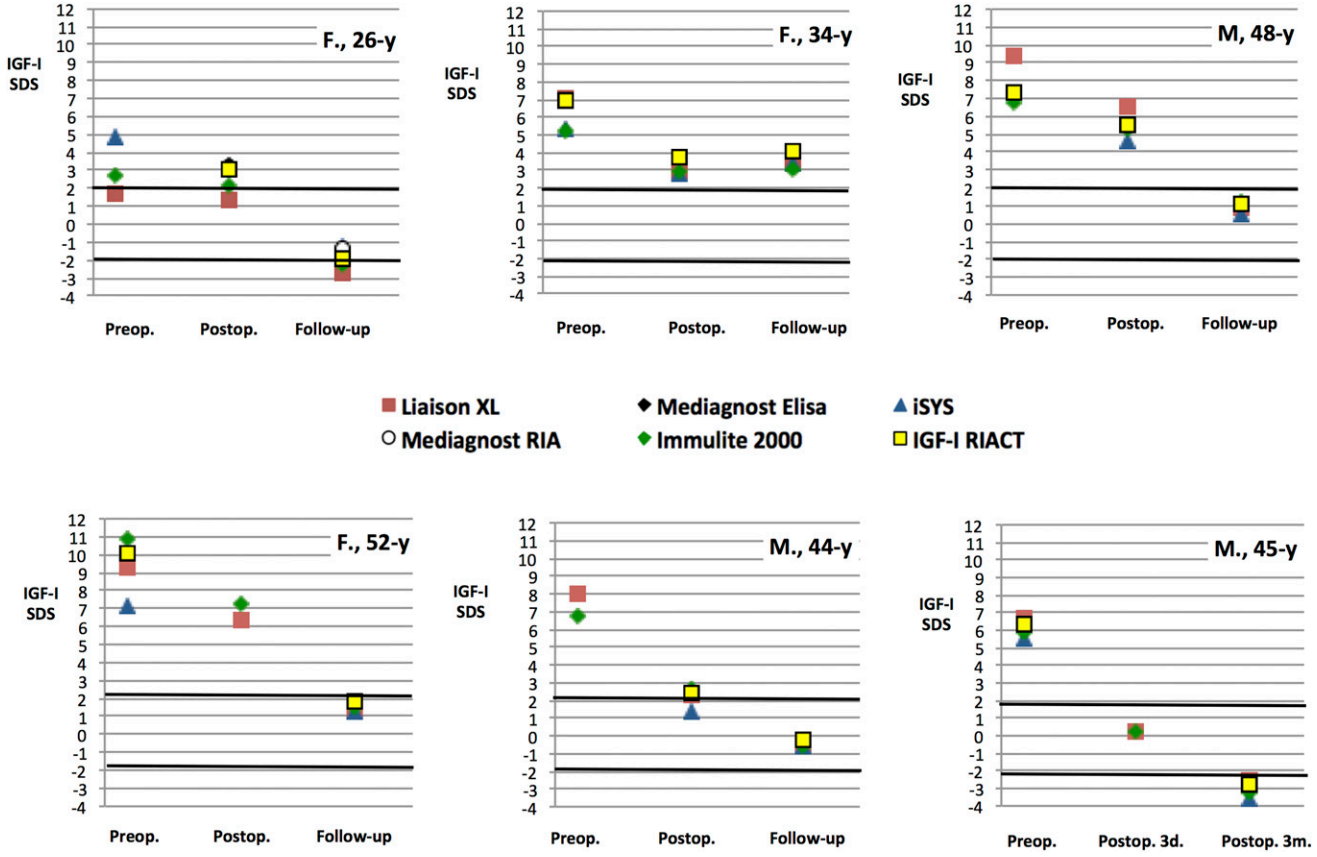

Figure 2. Variability of SDS values obtained with each of the six immunoassays in six patients with acromegaly, at three points of follow-up: diagnosis of acromegaly (Preop.), immediately after surgery (Postop. 3d), and at follow-up, generally under medical treatment. Horizontal lines represent the normal range from +2 to –2 standard deviations. Each assay is assigned a different colored symbol. Postop. 3m., 3 months after surgery.

and 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 κ 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 Supplemental Tables 2 to 7.

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 (Liaison XL and iSYS, iSYS and Immulite, Liaison XL and Mediagnost ELISA, and Immulite and Mediagnost ELISA).

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

For high IGF-I SDSs (>1.96), numbers used for comparisons are variable (n = 16, 24, and 31). In the majority of cases, only one or two assays give different classification. There are more than three misclassified patients when comparing Liaison XL and iSYS, Liaison XL and Immulite, Liaison XL and Mediagnost ELISA, Liaison XL and Mediagnost RIA, and Mediagnost RIA and IGF-I RIACT. 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-I 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 international recombinant IGF-I calibration standard preparation 02/254 and emphasized the importance of

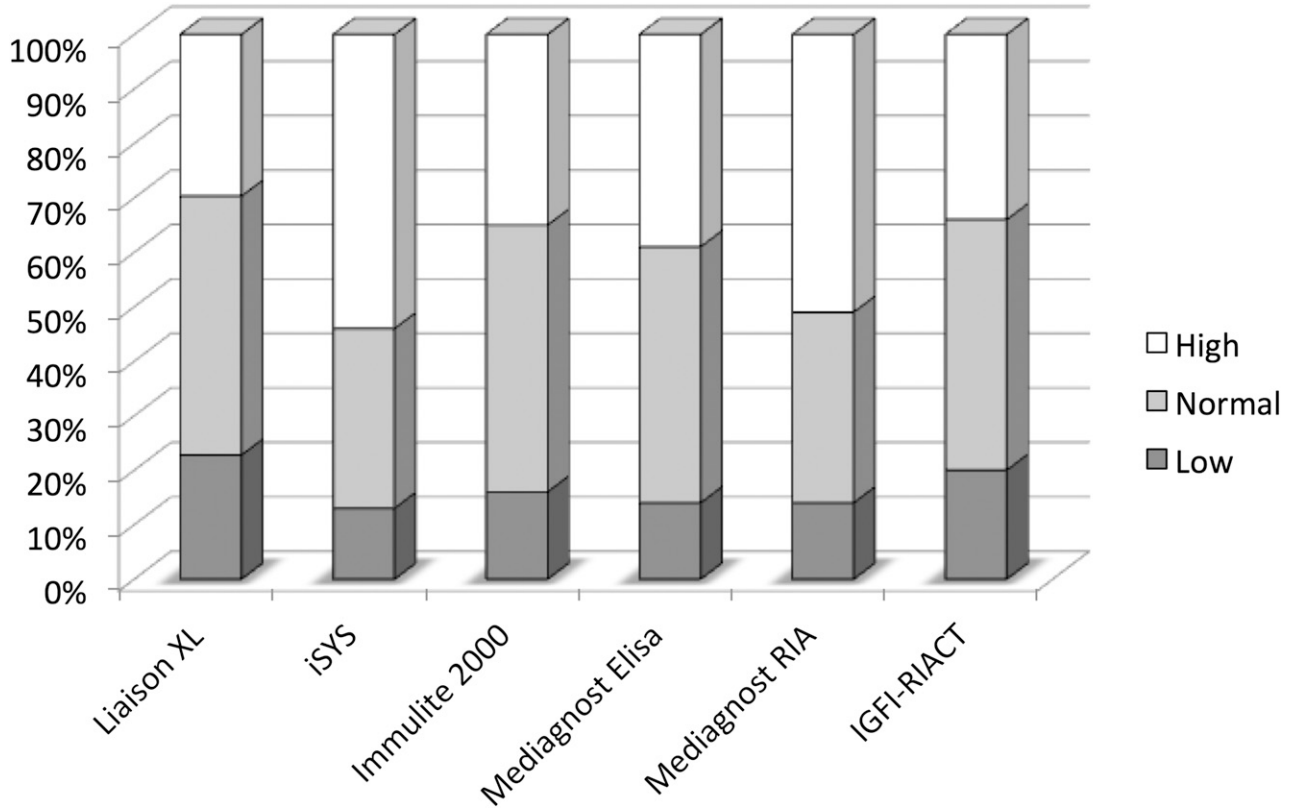

Figure 3. Percentages of patients with normal, low, and high IGF-I levels according to each of the six IGF-I immunoassays.

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 or renal or hepatic failure or taking medications (such as estrogen) that could affect outcome.

Based on this consensus statement, Bidlingmaier *et al.* (18) published normative data for the iSYS IGF-I assay obtained in a cohort of 15,014 healthy subjects, 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 SDSs. Nevertheless, although we ensured the same preanalytical 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 SDSs (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 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 micrograms per liter) were obtained with Immulite and IGF-I RIACT, the two immunoassays calibrated against the old standard International Reference Preparation 87/518 (13). However, when using SDSs 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

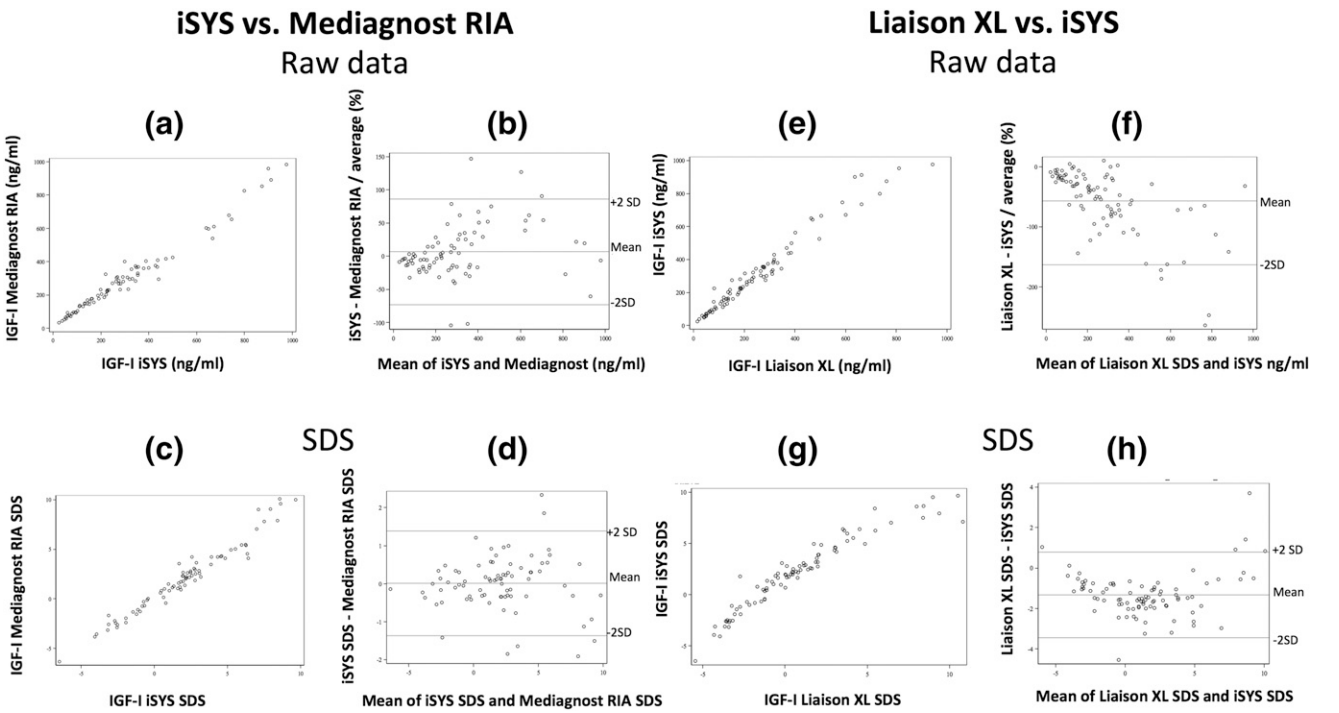

Figure 4. Comparisons between iSYS and Mediagnost RIA expressed as (a) scatter plots and (b) Bland-Altman plots for raw data, or (c) scatter plots and (d) Bland-Altman plots for SDSs, showing excellent overall agreement between the two immunoassays. Comparisons between Liaison XL and iSYS expressed as (e) scatter plots and (f) Bland-Altman plots for raw data, or (g) scatter plots and (h) Bland-Altman plots for SDSs, showing moderate overall agreement between these two immunoassays.

exhibited only moderate agreement (κ coefficients, 0.50 and 0.55, respectively), and only the pair Immulite/Liaison XL showed substantial agreement (κ 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 were 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, with some pairs being clearly more discordant than others.

Another possible explanation for the lack of concordance is a difference in the technical procedure (5, 6, 12, 19). In the current 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 insulinlike growth factor binding proteins 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 insulinlike growth factor binding protein 3.

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 current 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

Table 2. Agreement Between IGF-I Assay Methods, Expressed as Weighted κ Coefficient

κ Coefficient	Liaison XL	iSYS	Immulite 2000	Mediagnost ELISA	Mediagnost RIA	IGF-I 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
IGF-I RIACT	0.64	0.71	0.65	0.77	0.71	

underestimate IGF-I values by comparison with other techniques must be clearly identified. Tandem liquid chromatography and mass spectrometry (LC-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, 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 SDSs, were quite variable in patients 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 that 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 SDSs 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

We thank Céline Piedvache for her help in complementary statistical analysis.

Address all correspondence and requests for reprints to: Philippe Chanson, MD, 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 work was supported by Programme Hospitalier de Recherche Clinique, French Ministry of Health Grant P081216/IDRCB 2009-A00892-55. M.M. received a grant from Fond National Suisse (Grant P1GEP3_155694) for her contribution to this work.

Clinical trial registry: ClinicalTrials.gov no. NCT01831648 (registered 3 April 2013).

Disclosure Summary: The authors have nothing to disclose.

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**(1):109–129.
2. Katznelson L, Laws ER Jr, Melmed S, Molitch ME, Murad MH, Utz A, Wass JA; Endocrine Society. Acromegaly: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2014; **99**(11):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; Acromegaly Consensus Group. Expert consensus document: a consensus on the medical treatment of acromegaly. *Nat Rev Endocrinol*. 2014;**10**(4): 243–248.
4. Molitch ME, Clemmons DR, Malozowski S, Merriam GR, Vance ML, Endocrine S; Endocrine Society. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;**96**(6): 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**(1): 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**(4):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**(1-2): 176–179.
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**(1):65–70.
9. Krebs A, Wallaschofski H, Spilcke-Liss E, Kohlmann T, Brabant G, Völzke 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**(12):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**(1): 309–316.
11. Brabant G, Wallaschofski H. Normal levels of serum IGF-I: determinants and validity of current reference ranges. *Pituitary*. 2007; **10**(2):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**(1):27–34.
13. Chanson P, Arnoux A, Mavromati M, Brailly-Tabard S, Massart C, Young J, Piketty ML, Souberbielle JC, Investigators V; VARIETE Investigators. Reference values for IGF-I serum concentrations: comparison of six immunoassays. *J Clin Endocrinol Metab*. 2016; **101**(9):3450–3458.
14. Melmed S. Acromegaly. In: Melmed S, ed. The Pituitary. Malden, MA: Blackwell Science Inc.; 2011:463–474.
15. Chanson P, Salenave S, Kamenicky P. Acromegaly. *Handb Clin Neurol*. 2014;**124**:197–219.
16. Fermanian J. Measuring agreement between 2 observers: a quantitative case. *Rev Epidemiol Sante Publique*. 1984;**32**(6):408–413.
17. Fermanian J. Measurement of agreement between 2 judges. Qualitative cases. *Rev Epidemiol Sante Publique*. 1984;**32**(2): 140–147.
18. Bidlingmaier M, Friedrich N, Emeny RT, Spranger J, Wolthers OD, Roswall J, Körner A, Obermayer-Pietsch B, Hübener 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**(5):1712–1721.
19. Clemmons DR. IGF-I assays: current assay methodologies and their limitations. *Pituitary*. 2007;**10**(2):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**(3):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**(9):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 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**(1):48–69.