Central laboratory reassessment of IGF-I, IGF-binding protein-3, and GH serum concentrations measured at local treatment centers in growth-impaired children: implications for the agreement between outpatient screening and the results of somatotropic axis functional testing

Berthold P Hauffa1, Nils Lehmann2, Markus Bettendorf3, Otto Mehls3, Helmut-Günther Dörr4, Nikolaus Stahnke5, Heinz Steinkamp6, Elfriede Said6, Michael B Ranke7 and participating members of the German KIGS Board/Medical Outcome Study Group

1Division of Pediatric Endocrinology and Diabetology, University Children’s Hospital, Heidelberg, Germany, 2Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany, 3University Children’s Hospital, Heidelberg, Germany, 4University Hospital for Children and Adolescents, Erlangen, Germany, 5Endokrinologikum, Hamburg, Germany, 6Pfizer Pharma GmbH, Karlsruhe, Germany and 7Pediatric Endocrinology Section, University Children’s Hospital, Tuebingen, Germany

Abstract

Background: Childhood GH deficiency, suspected in the presence of decreased height velocity and short stature, is usually characterized by low IGF-I and IGF-binding protein-3 (IGFBP-3) serum concentrations and is conventionally confirmed by diminished GH peak responses to pharmacological stimuli.

Objective: We evaluated the agreement between different IGF-I (IGFBP-3) assays in predicting GH deficiency and tested whether variability between growth factor screening and pharmacological testing could be diminished by reassessment of growth factor and GH peak concentrations in a single laboratory.

Design: Using the Tuebingen IGF-I (IGFBP-3) RIA, 317 (321) sera from children evaluated for growth disorders in 19 centers were reanalyzed. In 103 children with insulin hypoglycemia and arginine tests, we evaluated how the association between the outcome of growth factor screening and functional testing would change if different assays were employed.

Results: Locally measured IGF-I correlated better than IGFBP-3 with the results of the central laboratory (Tuebingen) assay (slope of the regression curve 1.05; 95% confidence interval (95% CI) 1.01–1.1 versus 1.18; 95% CI 1.09–1.3). Agreement between local and central laboratory assays in predicting GH deficiency was better for IGF-I than for IGFBP-3 assays (κ = 0.59 versus κ = 0.47). The poor agreement between growth factor screening and GH pharmacological testing was not improved when hormone concentrations were remeasured in the central laboratory (κ local = −0.0031, central = 0.12).

Conclusions: In children with impaired growth, growth factor screening reflects different aspects of GH insufficiency than does functional testing. Agreement between these approaches is poor and could not be improved by reduction of assay-related variability.

European Journal of Endocrinology 157 597–603

Introduction

Growth hormone (GH) deficiency is suspected in children with decreased height velocity and short stature, characterized by low insulin-like growth factor (IGF-I) and IGF-binding protein-3 (IGFBP-3) serum concentrations in most patients, and conventionally confirmed by diminished GH peak responses to pharmacological stimuli (1). The agreement between the results of growth factor screening and GH pharmacological testing is highly variable (2–19). Part of this variability is introduced by the use of differently designed hormone assays, using different antibodies and standards, both for measurement of GH and IGF-I or IGFBP-3. We have shown previously that regional use of different GH assays and cutoff values increases the variability not only of hormone measurements, but also of diagnostic and therapeutic decisions regarding growth hormone deficiency (20). Similar restrictions could apply to the measurement of growth factors, often employed in screening for GH deficiency.
Therefore, we wanted to determine the agreement between IGF-I (IGFBP-3) serum concentrations measured in various treatment centers, and measurements performed on the same sera in one central laboratory. We further asked if the agreement of patient diagnostic assignment (‘GH deficient’ versus ‘non-GH deficient’) using growth factor screening alone, or comparing the results of growth factor screening and GH pharmacological testing, would change when different assays, each using its own established reference range, were employed.

**Subjects and methods**

**Samples and subjects**

Twenty-one German secondary and tertiary care centers diagnosing and treating children with growth disorders participated in the study. From January 1, 1998 to December 31, 1999, all consecutive children and adolescents who underwent functional testing of the somatotropic axis or evaluation of spontaneous GH secretion were considered for inclusion. First, IGF-I, IGFBP-3 concentrations (basal samples), and GH concentrations (basal and post-stimulation samples) were measured in each center, using the local assays. If serum was available from the basal sample and the post-stimulation sample containing the peak GH concentration, these were sent to the reference laboratory (University Children’s Hospital, Tuebingen) for re-evaluation.

Here, 317 (321) basal sera whose IGF-I (IGFBP-3) concentrations had been measured locally were available. Measurement results were classified as low, normal or high using the local reference ranges, which differed between assays. Analyte concentrations were then remeasured in the central laboratory using the in-house Tuebingen RIAs for IGF-I and IGFBP-3.

Here, 790 GH peak sera of 434 patients were reassayed centrally. Ninety-one of these results (52 patients, 2 centers) were excluded from the analysis for various reasons (20). Thus, 699 peak GH sera of 382 children and adolescents (258 males and 124 females) from 19 centers were available for analysis. From these, a subgroup of 132 patients with the most abundant test combination (insulin hypoglycemia test and arginine test) were evaluated for changes in the assignment to the diagnostic group of GH deficiency, which occurred with the use of different assays. Patients were diagnosed as GH deficient if the GH concentration remained below the assay-specific cutoff value in both tests. Otherwise, the patients were categorized as non-GH deficient. In 103 of these patients, measurements of IGF-I and IGFBP-3 were available for comparison.

**Ethical approval**

The study protocol was reviewed and approved by the Ethics Committee for Medical Research of the University of Tuebingen.

**Analytical methods**

Assays for IGF-I and IGFBP-3 used in the participating centers (local assays) and their characteristics are detailed in Table 1. GH assays were used as already described (20). Samples were stored frozen for various lengths of time not exceeding 6 months at −20 °C in 300 μl aliquots and sent to the central laboratory at the University Children’s Hospital, Tuebingen (M B R), where IGF-I, IGFBP-3, and GH concentrations were

| Table 1 Characteristics and frequency of use of different insulin-like growth factor-I (IGF-I) and IGF-binding protein-3 (IGFBP-3) immunoassay methods employed by the participating centers. |

<table>
<thead>
<tr>
<th>Assay no.</th>
<th>Assay name (extraction, releasing, blocking agents)</th>
<th>Number of patients</th>
<th>Assay type</th>
<th>Standard calibrated against WHO</th>
<th>Antibodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF-I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Nichols (acid–ethanol, C18)</td>
<td>7</td>
<td>RIA</td>
<td>87/518</td>
<td>Double Ab</td>
</tr>
<tr>
<td>2</td>
<td>DSL (acid–ethanol)</td>
<td>27</td>
<td>IRMA</td>
<td>87/518</td>
<td>Double Ab</td>
</tr>
<tr>
<td>3</td>
<td>DSL (no extraction)</td>
<td>7</td>
<td>IRMA</td>
<td>87/518</td>
<td>Double Ab</td>
</tr>
<tr>
<td>4</td>
<td>Medignost (acid–ethanol, IGF-II)</td>
<td>87</td>
<td>RIA</td>
<td>87/518</td>
<td>Double Ab</td>
</tr>
<tr>
<td>5</td>
<td>Nichols advantage (acid–ethanol, IGF-II)</td>
<td>26</td>
<td>CLIA</td>
<td>87/518</td>
<td>Double Ab</td>
</tr>
<tr>
<td>6</td>
<td>SERIA biochem (acid–ethanol)</td>
<td>9</td>
<td>RIA</td>
<td>87/518</td>
<td>Single Ab</td>
</tr>
<tr>
<td>7</td>
<td>In-house assay Giessen (variant of 4)</td>
<td>18</td>
<td>RIA</td>
<td>87/518</td>
<td>Double Ab</td>
</tr>
<tr>
<td>8</td>
<td>OCTEIA (releasing reagent)</td>
<td>52</td>
<td>EMA</td>
<td>87/518</td>
<td>Two-site Ab</td>
</tr>
<tr>
<td>9</td>
<td>In-house assay Tuebingen (variant of 4)</td>
<td>84</td>
<td>RIA</td>
<td>87/518</td>
<td>Double Ab</td>
</tr>
<tr>
<td>IGFBP-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>DSL</td>
<td>33</td>
<td>IRMA</td>
<td>Recombinant</td>
<td>Double Ab</td>
</tr>
<tr>
<td>2</td>
<td>Medignost</td>
<td>90</td>
<td>RIA</td>
<td>Extractive</td>
<td>Double Ab</td>
</tr>
<tr>
<td>3</td>
<td>Biochem</td>
<td>15</td>
<td>RIA</td>
<td>Extractive</td>
<td>Double Ab</td>
</tr>
<tr>
<td>4</td>
<td>Nichols</td>
<td>80</td>
<td>RIA</td>
<td>Extractive</td>
<td>Double Ab</td>
</tr>
<tr>
<td>5</td>
<td>In-house assay Giessen (variant of 2)</td>
<td>15</td>
<td>RIA</td>
<td>Extractive</td>
<td>Double Ab</td>
</tr>
<tr>
<td>6</td>
<td>In-house assay Tuebingen (variant of 2)</td>
<td>88</td>
<td>RIA</td>
<td>Extractive</td>
<td>Double Ab</td>
</tr>
</tbody>
</table>

IRMA, immunoradiometric assay; CLIA, chemiluminescence immunoassay; EMA, enzyme immunoassay; RIA, radioimmunoassay.
Statistical methods

Analytical results obtained at the different local centers were combined and compared pairwise with the results of the central laboratory by the Bland–Altman plot (21) and the Passing–Bablok regression (22), using MedCalc software, Maria Verke, Belgium (23). These methods are nonparametric and do not require one assay to be assigned the role of an analytical gold standard. Whether the local assays and each of the assays used by the central laboratory agreed in the assignment of a patient to a diagnostic group (GH deficient versus non-GH deficient) was assessed using the McNemar test and the κ statistic (24). A κ value between 0.4 and 0.6 indicates moderate agreement, κ between 0.6 and 0.8 indicates clear agreement, and κ between 0.8 and 1.0 indicates high agreement. The latter analyses were performed using SAS software version 8.2, Cary, NC, USA (25).

Results

In general, central laboratory measurements of serum IGF-I concentrations showed a good agreement with the measurements obtained at the local laboratories (regression equation of central laboratory measurements (x) with pooled local measurements (y): \( y = -6.8 + 1.05x \); Fig. 1A and C). There was no significant deviation of the regression line from linearity. Bland–Altman analysis showed a negligible bias of \(-1.5 \text{ ng/ml}\). The ±1.96 s.d. range of the measurement differences, however, was comparably large. It spanned 170.2 ng/ml and approached one-third of the observed concentration range, with greater differences between results at higher analyte concentrations. A slightly lesser agreement between local and central laboratory measurements was found for IGFBP-3 (regression equation: \( y = -0.9 + 1.18x \); Fig. 1B and D). Bias was 0.36 mg/l, and the ±1.96 s.d. range was 3.0 mg/l, approaching about 45% of the IGFBP-3 concentration range of the samples. The regression line did not deviate from linearity. Regression equations of measurements obtained with individual local assays and the central laboratory assays are shown in Table 2.

When either IGF-I or IGFBP-3 results were used separately for predicting GH deficiency (analyte decreased, GH deficiency likely; analyte normal or elevated, GH deficiency unlikely), agreement between predictions obtained with local and central laboratory assays was better for the IGF-I assay than for the IGFBP-3 assay (κ = 0.59 versus κ = 0.47). When IGF-I and IGFBP-3 measurements were combined to sort the outcomes of the growth factor screening into three categories (IGF-I and IGFBP-3 low, IGF-I or IGFBP-3 low, and IGF-I and IGFBP-3 normal or high), the distribution of the patients between these categories obtained with the use of local assays (Table 3) agreed moderately (κ = 0.46, 95% CI 0.37–0.58) with the distribution seen with the use of the central laboratory assays.

The agreement between growth factor/binding protein screening (suspected GH deficiency: both analytes below reference range) and GH pharmacological testing (GH deficiency: GH below cutoff in two tests) was poor for a group of 103 children and adolescents, whether the comparison was based on the data obtained with the local or central laboratory assays. Agreement did not improve when the hormone concentrations remeasured in the central laboratory were used for comparison (κ local = -0.0031; κ central = 0.12; Table 4). When less stringent criteria were employed for suspected GH deficiency (either IGF-I or IGFBP-3 below reference range), agreement between results of growth factor screening and GH functional testing was still low, when local measurements were considered (IGF-I low: κ local = 0.035; IGFBP-3 low: κ local = -0.0008). There was only a minor improvement of this agreement, when the analysis was based on the central laboratory measurements (IGF-I low: κ central = 0.17; IGFBP-3 low: κ central = 0.14).

Discussion

We have shown that IGF-I and IGFBP-3 concentrations measured locally in 19 German treatment centers correlated well with the assays used in a central laboratory (Tuebingen assay). The correlation was better for IGF-I than for IGFBP-3. When children with growth disturbances were categorized according to their IGF-I and IGFBP-3 concentrations, a moderate agreement was found between local and central assays with regard to assigning individual patients to these categories. There was, however, no agreement concerning the assignment of patients to the diagnostic categories ‘growth hormone deficient’ or ‘non-growth hormone deficient’, between the results of growth factor screening and GH functional testing. The agreement did not improve when only hormone concentrations measured in the central laboratory were used for analysis.

Correlations between locally and centrally measured IGF-I and IGFBP-3 concentrations were better than those reported by us for GH, obtained from the same study population (20). This is due in part to a...
greater procedural standardization of commercially available IGF-I and IGFBP-3 assays. All commercially available IGF-I assays in this study used a human recombinant IGF-I standard calibrated against the International Reference Reagent WHO 87/518. Although this procedure has been criticized as a least common denominator needing improvement (26), it constitutes a more desirable situation than for GH, where different and outdated standards are still used for assay calibration. In growth factor measurement, major differences exist with regard to separation techniques. In our study, however, these differences did not prohibit a rather good agreement between assays, as judged by the similarity of the regression equations. In a previous study, almost identical regression curves were obtained, when IGF-I concentrations in the sera of 427 adult blood donors, measured with four of the assays used in this study (Tuebingen RIA, Mediagnost RIA, DSL IRMA, Nichols Advantage IRMA), were compared (27). Not surprisingly, this agreement between results obtained with different IGF-I assays was reflected in our study with a moderate agreement between the outcomes of diagnostic categorization, not only when IGF-I but also when IGFBP-3 measurements were considered.

IGF-I and IGFBP-3 measurements are commonly employed in the diagnosis of GH deficiency. In a 1995 survey, 88% of pediatric endocrinologists in the USA used IGF-I measurements in the diagnostic work up of children with suspected GH deficiency, at least occasionally or on a regular basis (28). Recent editions of major textbooks of endocrinology include the measurement of IGF-I and IGFBP-3, alone or in combination, in their algorithms for work up of children with short stature (29, 30), mostly for screening.

Figure 1 Comparison of IGF-I (A and C) and IGFBP-3 (B and D) concentrations measured in 19 local centers and remeasured in a central laboratory (Tuebingen). In A and B, for each sample analyte, concentrations measured in the central laboratory (x-axis) are plotted against the concentrations measured in the local laboratories (y-axis). Solid lines represent the regression lines according to Passing and Bablok, with the 95% confidence intervals indicated by dotted lines. C and D display the Bland–Altman plots, where the mean of the local and central measurements (x-axis) is plotted against the difference of these values (y-axis). Horizontal lines indicate the mean and ± 1.96 s.d. range of these differences.
purposes. The appropriate interpretation of the results of this growth factor screening, however, is subject to considerable debate. If GH deficiency is defined by the failure of GH to rise above a predefined value in two independent pharmacological or physiological tests (1), then reported cutoff-dependent sensitivities of the IGF-I (IGFBP-3) measurements to predict GH deficiency range from 34 to 96% (14.9 to 97%) with a specificity ranging from 32 to 97.9% (50 to 97.9%) (5–7, 9, 11, 13, 15, 31–33). The ability of these measurements to serve as a surrogate parameter for GH functional testing in the diagnosis of GH deficiency has been estimated from being high (34) to nonexistent (15). Our data seem to support the latter notion, since there was no agreement between the results of growth factor screening and GH testing. However, we had used the most restrictive criteria for prediction of GH deficiency, requiring that both IGF-I and IGFBP-3 be low. But even if less stringent criteria were employed, agreement improved only slightly with all \( \kappa < 0.2 \) and significant variability remained. This could be explained by the fact that some local investigators had used information for their decision to proceed with GH functional testing not contained in the IGF-I and IGFBP-3 data and unknown to us, e.g. a very poor growth rate with IGF-I and IGFBP-3 concentrations just above the lower limit of normal. This would generate a higher proportion than expected of patients discordantly combining IGF-I and IGFBP-3 values not indicative of GH deficiency with the results of GH functional testing compatible with GH deficiency. This observation implies that in many situations the information provided by IGF-I and IGFBP-3 may be canceled out by the clinical findings during the process of diagnostic decision making.

Further, biological and analytical factors have been noted to explain the poor agreement. Circulating concentrations of IGF-I and IGFBP-3 depend not only on GH secretion, but also on GH-independent causes like genetic factors (35, 36), nutritional status (37, 38), hepatic and renal function (39–41), or underlying diseases presenting with varying degrees of GH resistance. Some of these factors have no or little effect on GH concentration. This would explain the discordant combination of low IGF-I and IGFBP-3 concentrations and normal results of GH functional testing in many patients.

Thus, it cannot be expected that IGF-I or IGFBP-3 concentrations represent a direct mirror image of the

---

**Table 2** Regression equations and 95% confidence intervals (CI) of intercept and slope of the pairwise method comparisons of the local insulin-like growth factor-I (IGF-I) and IGF-binding protein-3 (IGFBP-3) measurements (y) versus the central laboratory (Tuebingen) assay results (x), calculated with the Passing–Bablok method (22). Because of similarities in assay design and lack of difference with regard to the regression equation, the data of both Nichols assays (and the data of the Mediagnost assay and the in-house assay Gießen) were pooled.

<table>
<thead>
<tr>
<th>Local assays (y)</th>
<th>Regression equation with central laboratory (Tuebingen) assay results (x)</th>
<th>Intercept (95% CI)</th>
<th>Slope* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF-I Nichols</td>
<td>( y = -14.8 + 1.12x )</td>
<td>-34.6 to +0.5</td>
<td>0.97–1.38</td>
</tr>
<tr>
<td>DSL (acid–ethanol)( ^a )</td>
<td>( y = -22.6 + 1.37x )</td>
<td>-41.4 to -2.2</td>
<td>1.20–1.63</td>
</tr>
<tr>
<td>Mediagnost/in-house assay Gießen</td>
<td>( y = -5.4 + 1.03x )</td>
<td>-9.0 to -1.9</td>
<td>0.99–1.08</td>
</tr>
<tr>
<td>SERIA Biochem</td>
<td>( y = -48.6 + 1.12x )</td>
<td>-143.1 to -1.0</td>
<td>0.70–1.93</td>
</tr>
<tr>
<td>OCTEIA</td>
<td>( y = -0.2 + 0.85x )</td>
<td>-6.8 to +5.9</td>
<td>0.76–0.96</td>
</tr>
<tr>
<td>IGFBP-3 DSL</td>
<td>( y = 1.0 + 1.16x )</td>
<td>-2.1 to +0.2</td>
<td>0.83–1.61</td>
</tr>
<tr>
<td>Mediagnost/in-house assay Gießen</td>
<td>( y = -0.8 + 1.14x )</td>
<td>-1.2 to -0.5</td>
<td>1.04–1.27</td>
</tr>
<tr>
<td>Biochem</td>
<td>( y = -0.6 + 0.92x )</td>
<td>-3.2 to +0.7</td>
<td>0.42–1.83</td>
</tr>
<tr>
<td>Nichols</td>
<td>( y = -1.0 + 1.30x )</td>
<td>-1.6 to -0.4</td>
<td>1.11–1.52</td>
</tr>
</tbody>
</table>

\( \kappa \), confidence interval. *Significant nonlinearity of the regression lines was not observed in any of the comparisons \( (P>0.1) \).

\( ^a \)For both DSL IGF-I assays (with acid–ethanol extraction and without extraction), regression equations differed significantly and, therefore, could not be pooled. Patient numbers for the assay without extraction were too low to permit calculation of 95% confidence intervals. Therefore, only the data after acid–ethanol extraction are reported.

---

**Table 3** Comparison between the categorization of patients \( (n=312) \) according to their insulin-like growth factor-I (IGF-I) and IGF-binding protein-3 (IGFBP-3) concentrations measured in the local centers and remeasured in a central laboratory (Tuebingen), showing moderate agreement. *\( \kappa \) = kappa, 95% CI = 95% confidence interval for \( \kappa \).

<table>
<thead>
<tr>
<th>Local laboratories (local assays)</th>
<th>IGF-I and IGFBP-3 low</th>
<th>IGF-I or IGFBP-3 low</th>
<th>IGF-I and IGFBP-3 normal or high</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF-I and IGFBP-3 low</td>
<td>40</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>IGF-I or IGFBP-3 low</td>
<td>25</td>
<td>70</td>
<td>24</td>
</tr>
<tr>
<td>IGF-I and IGFBP-3 normal or high</td>
<td>9</td>
<td>29</td>
<td>92</td>
</tr>
</tbody>
</table>

\( *\kappa = 0.46, 95\% CI = 0.37–0.58. \)
GH secretion status. It has been hypothesized that in case analytical sources of IGF-I (IGFBP-3) variation could be minimized, the prediction of GH secretory status would improve (42). Our data do not support this hypothesis. When central laboratory assays were used instead of local assays, a numerical shift occurred between the patients allocated to discordant categories, but prediction of GH deficiency did not improve. This may be explained in part by the fact that intraindividual variation over time is greater for the results of GH functional testing than for IGF-I (IGFBP-3) concentrations (43). This would imply that overall agreement could rather be improved by standardization of the GH assays.

We conclude that the assay-related variability of results between local centers and a reference laboratory is lower for IGF-I than for IGFBP-3. In children with impaired growth, growth factor screening reflects aspects of GH insufficiency different from those reflected by functional testing. Agreement between these diagnostic approaches is poor and could not be improved by reduction of assay-related variability.

Acknowledgements

The participating investigators of the German KIGS Board/Medical Outcome Study Group are as follows: W Andler (Datteln), R Arndt (Neubrandenburg), M Bettendorf (Heidelberg), H J Böhles (Frankfurt/Main), H G Dörr (Erlangen), S Eisberg (Minden), B P Hauffa (Essen), O Hiort (Lübeck), E Korsch (Köln), M Mix (Rostock), K Mohnike (Magdeburg), M Morlot (Hannover), R Mühlenberg (Krefeld), C-J Partsch (Kiel), M B Ranke (Tuebingen), E Schönau (Köln), H-P Schwarz (München), G Simic-Schleicher (Bremen), K-P Ullrich (Gotha), T van Meegen-Freund (Aachen), and Ch Vogel/A Klinghammer (Chemnitz). This work was supported by a grant from Pharmacia Corporation, Erlangen, Germany to B P H and M B R. The authors thank Mrs K Weber for her expert technical assistance and R D Oades, Department of Child and Adolescent Psychiatry, University of Essen for helpful review of the manuscript.

References


Received 21 May 2007
Accepted 8 August 2007