Evaluation of growth hormone assays using ratio plots

MARianne ANdersen,1* PEr HyLtoft PeTersen,2 Ole BlaaBJerg,2 JOrgen HangaARD,1 and ClauS Hagen1

To permit comparison between growth hormone (GH) results obtained using the Pharmacia polyclonal assay and the Delfia monoclonal assay, we used both methods to measure GH concentrations in peak GH responses to the pyridostigmine-growth-hormone-releasing-hormone (PD-GHRH) test and in unstimulated samples from 40 healthy adults and 31 patients with suspected GH deficiency. Ratio plots were used for the comparison, and acceptability criteria were based on inherent analytical imprecision and on analytical quality specifications. The mean ratio (r; Pharmacia/Delfia) for the peak GH responses in 40 healthy adults was calculated to be 1.59, and the 95% prediction interval for ratios fulfilling the imprecision criterion was applied. For GH values > 1 mIU/L, the peak and unstimulated GH ratios in healthy adults and patients were within the 95% prediction interval, and fulfilled the biological criterion as well. Therefore, the conversion factor of 1.59 is applicable for the evaluation of GH-stimulation tests.

The diagnosis of growth hormone deficiency (GHD)3 in adults is a clinical challenge. It requires clinical experience and demands reliable diagnostic tools – in particular a reliable method for measuring serum growth hormone (GH).

Plasma contains >100 forms of GH, and what is commonly understood as “plasma GH,” i.e., the free monomeric 22-kDa form, represents only 21% of the total immunoreactivity in plasma (1). However, the 22-kDa form of GH is the most prevalent form and is the prototype of pituitary GH. This form is a single-chain, 191-amino acid protein. The 20-kDa form is the second most abundant in the circulation. During peak secretion, the relative amounts of the 22-kDa and 20-kDa forms are fairly stable and independent of secretagogues (2), the relative mean percentage concentrations being 76.4% and 15.8%, respectively. This is in contrast to the basal state, where only a fraction of immunoreactive GH in the blood can be attributed to known GH forms (3). From the above description of the nonhomogeneity of GH, it is clear that the antibodies used in an assay would be of great importance. The use of polyclonal antibodies would yield higher results, because epitopes on different GH forms could bind the antibodies, in contrast to monoclonal antibodies directed only against the 22-kDa form of GH. The degree of cross-reactivity for the monoclonal and polyclonal assays has been quantified and specific differences between the two methods shown (4).

Because we previously had evaluated a new stimulation test, the pyridostigmine-growth-hormone-releasing-hormone (PD-GHRH) test, including a reliable cutoff limit for the test (5), it was with some doubt we started comparisons between the well-established Pharmacia assay and the new Delfia assay.

The problems with hormones in general, e.g., deterioration and lack of transferability, and with GH in particular suggested that it would be difficult to compare results obtained from the Pharmacia method based on polyclonal antibodies and those obtained from the Delfia method using monoclonal antibodies specific to the 22-kDa GH form of the hormone. The difficulties with comparisons between assays were demonstrated in a recent paper (4); these difficulties may contribute to the lack of confidence in GH stimulation tests (6).

To permit comparison between results obtained with the Pharmacia and the Delfia methods, we used both methods to measure GH concentrations in samples from 71 individuals: 40 healthy adults and 31 patients with suspected GHD. Both basal samples and peak GH responses after stimulation with pyridostigmine (PD) in combination with GHRH were studied. We chose to compare the results using ratio plots, a modification of the
Altman-Bland difference plot (7), with the application of a 95% prediction interval and with acceptability criteria based on inherent analytical imprecision and on analytical quality specifications (8, 9). We also chose to perform a step-by-step evaluation using well-defined serum samples from groups of healthy adults and patients before and during a stimulation test.

**Materials and Methods**

Seventy-one GH stimulation tests (PD-GHRH tests) were performed on 40 healthy adults [median age, (minimum-maximum), 36 years (22–58)] and on 31 patients with suspected GHD [49 years (18–60)]. The 40 healthy adults were selected into five groups of eight according to age, sex, and the use/nonuse of oral contraceptives (5).

The PD-GHRH test. The participants attended the clinic after fasting overnight; during the experiment they were not permitted to sleep, smoke, eat, or drink anything but tap water. A cannula was inserted into an antecubital vein for blood sampling and the administration of GHRH. At 0900, (time 0), 120 mg of PD (Mestinon, Hoffmann-La Roche) was administered orally; 60 min later, at 1000 (time 0), 1 µg/kg body weight of GHRH (Groliberin, Pharmacia & Upjohn), was administered intravenously as a bolus. Blood samples for the measurement of GH were taken at −60, 0, 20, 30, 60, and 90 min. The GH concentrations corresponding to basal and peak GH responses are presented here.

The samples used in this study were obtained in connection with other investigations (5). The Declaration of Helsinki(II) was observed and the local Ethical Committee has approved the study. All subjects were volunteers, and they signed an informed consent document before taking part in the study.

**METHODS**

The two assays are immunometric: Pharmacia hGH RIA (Pharmacia & Upjohn) and Delfia 22-kDa hGH (Wallac). The Pharmacia hGH assay is a 2-site immunoradiometric method using polyclonal rabbit/sheep antibodies with 

antibodies are directed against two separate antigenic determinants on the 22-kDa hGH molecule with europium as the reporter molecule. All assays were performed according to the manufacturers’ protocols. The standards provided in the kits were calibrated against the 1st International Standard 80/505 from WHO in which 1 mg = 2.6 IU of GH. The estimated precision profile for the expected CVs is shown in Table 1. The in-house cross-reactivity of the Pharmacia and Delfia assays with biosynthetic recombinant hGH (Genotropin, KabiVitrum AB) were 134% and 114%, respectively (ratio 1.18). The quality assessments carried out externally with Ringversuche (Deutsche Gesellschaft für Klinische Chemie E.V.) showed that the median ratio (minimum-maximum) between the means of the participants was 1.10 (1.03–1.14) and close to the ratio estimated in our laboratory on a single occasion (1.18). In both assays, the GH standard was calibrated against the 1st International Standard 80/505 from WHO. This 80/505 standard is pituitary-derived and therefore contains a mixture of GH forms (dimerized and oligomerized) and GH fragments. The detection limit for the Pharmacia assay was 0.4 mIU/L; the intraassay and total CVs for the assay at 0.5 mIU/L were 14% and 17%, respectively, and at 22 mIU/L, the intraassay and total CVs were 4% and 6%, respectively. The detection limit for the Delfia assay was 0.03 mIU/L. The intraassay and total CVs for the assay at 0.5 mIU/L were 5% and 8%, respectively, and at 18 mIU/L, the CVs were 2% and 4%, respectively.

**DATA ANALYSIS**

The mean of ratios [individual GH results determined by the Pharmacia assay divided by the results obtained in the Delfia assay, (GHPharmacia/GHPharmacia)] and the expected distribution of ratios were calculated.

The paired results were evaluated using a modification of the Altman-Bland difference plot (7). A 95% prediction interval was calculated to express the interval within which we would expect 95% of the data points to be found (8, 9). A 95% prediction interval of this kind is in contrast to the 95% limits of agreement described for difference plots by Altman-Bland (7), which are based on calculations of the measured difference between the two methods, and thereby describes the 95% interval for measured differences. During GH stimulation in healthy adults, the percentage of the major GH forms in blood is assumed to be...

---

**Table 1. Estimation of CVtotal for the 95% prediction interval, based on duplicates of samples.**

<table>
<thead>
<tr>
<th>Concentration mIU/L</th>
<th>Delfia</th>
<th>Pharmacia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measured S_w</td>
<td>Measured CV_w</td>
</tr>
<tr>
<td>0.6–5.0</td>
<td>0.10</td>
<td>9%</td>
</tr>
<tr>
<td>5.1–9.0</td>
<td>0.22</td>
<td>4%</td>
</tr>
<tr>
<td>9.1–20.0</td>
<td>0.39</td>
<td>4%</td>
</tr>
</tbody>
</table>

*a From the control, it is known that CV_{between} \sim CV_{within}, or CV_{w+b} \sim 2^{1/2} \times CV_{within}. Using this information, we have estimated CV_{w+b} for each method by extrapolation from the calculated CV_{w} by this formula.

b CV_{total} is estimated, CV_{w+b}^2 + CV_{w+b}^{1/2} = CV_{total}.
be reasonably stable (2). Therefore, we have used the mean of ratios of the peak GH responses to the PD-GHRH test in healthy adults and the known total analytical CVs for the two assays to estimate the 95% prediction interval. The 95% prediction interval is the mean ratio ± 1.96 (CV²\text{Pharmacia} + CV²\text{Delfia})^{1/2} × \text{mean ratio}. With this information, it is possible to set up a prediction interval within which a certain fraction (~95%) of the ratio points are expected to be distributed and to test whether the prediction interval can be applied to related sets of less homogeneous data. This is in accordance with the graphical interpretation of analytical data using difference plots (8, 9). Visual inspection of the scatter of points in relation to the prediction interval can be supplemented by calculation of the mean of ratios and CVs for the different subgroups.

A more accurate 95% prediction interval could be based on the formula (which can be evaluated from Geary (10) and Hinkley (11)):

\[
\frac{r}{(1 - Z^2\text{CV}^2_{\text{denominator}})} \times \\
\left[1 + Z\left(\text{CV}^2_{\text{numerator}} + \text{CV}^2_{\text{denominator}}ight)
- Z^2\text{CV}^2_{\text{numerator}}\text{CV}^2_{\text{denominator}}
\right]^{1/2}
\]

where \( r \) is the mean ratio; \( Z \) is the standard deviation of the gaussian distribution (here 1.96); and the CV values denote the CVs of the numerator and the denominator, respectively (valid for \( Z < 1/\text{CV}^2_{\text{denominator}} \) and \( Z^2 < 1/\text{CV}^2_{\text{numerator}} + 1/\text{CV}^2_{\text{denominator}} \)). When the two CV values are the estimated values for analytical imprecisions of the two methods, Pharmacia and Delfia respectively, this interval describes more correctly the expected distribution of 95% of the measured ratios, if analytical imprecision is the only source of variation. For CV values of the denominator <0.10 (10%), the simplified prediction interval of the mean ratio ± 1.96 (CV²\text{Pharmacia} + CV²\text{Delfia})^{1/2} × \text{mean ratio} is close to the correct prediction interval with limits slightly below the correct limits (Fig. 1). Fig. 1 shows that the simplified method may be valid even for CV values up to 0.15 (15%). The two CV\text{denominator} values of 4% and 8% are indicated in Fig. 1. In European recommendations (12, 13), the analytical quality specification for imprecision is CV\text{analytical} ≤ 0.5CV\text{within-subject} where (CV\text{within-subject}) represented the CV of within-subject variation in accordance with the Cotlove–Harris concept (14). This criterion was also applied to the ratio plots (mean ratio ± 1.96 × \( \frac{1}{2}\text{CV\text{within-subject}} \times \text{mean ratio} \)) as an alternative acceptability criterion setting, the CV\text{within-subject}, equal to 36% (5).

**Results**

Evaluation of the acceptability of using a common factor for the conversion of values between the two analytical methods was performed step-by-step with different but homogeneous groups of sera.
The mean ratio of 1.59 is shown as a solid line. The solid lines with dotted interpolation indicate the 95% prediction interval, an interval within which 95% of the data points are expected to be found, based on the known analytical CVs of the two methods. The quality goal (analytic quality specification) for imprecision based on the biological concept by Cotlove–Harris is $CV_A \leq \frac{1}{2}CV_{within-subject}$, where the within-subject biological variation ($CV_{within-subject}$) for GH (36%) is applied as the two dotted lines. The line for the ratio of the measurement of biosynthetic recombinant hGH by Pharmacia/Delfia (1.18) is shown as well. (A) The peak GH ratios in 40 healthy adults are plotted against peak GH responses measured by Pharmacia. The 40 adults were selected into five groups: ( ), men ≤40 years; ( ), men >40 years; ( ), premenopausal women taking no oral contraceptives; ( ), premenopausal women taking oral contraceptives; and ( ), postmenopausal women. (B) The peak GH ratios in 28 patients with suspected GHD are plotted against peak GH responses (Pharmacia). The data for the three patients with no detectable GH are not shown. The five patients with a GH value of 1 mIU/L are shown as one point, and only one point is visible for the data of the two patients with peak GH responses measured by Pharmacia (6 mIU/L; ratio = 1.5).

**Fig. 2.** Ratio plots comparing the results of peak GH responses with the PD-GHRH test using two different methods, the polyclonal assay (Pharmacia) and the monoclonal assay (Delfia).

SERA FROM PEAK GH RESPONSES IN PATIENTS

Fig. 2B shows the peak GH responses in patients together with the 95% prediction interval, based on the conversion factor 1.59, the analytical goal, and biosynthetic recombinant hGH lines from Figure 2A. For values >1 mIU/L, the ratio points are distributed as ratio points for healthy individuals (mean ± S, 1.50 ± 0.19; n = 23). Below this point, there is a slight tendency to decreasing ratios for decreasing values. Of the eight patients with peak GH responses ≤1 mIU/L, three produced no detectable GH response to either method, and the other five produced a GH response of 1 mIU/L to both methods. However, these five all met the analytical criterion and remained close to the biological criterion for using a common conversion factor of 1.59. From a clinical point of view, these low values are far below the cutoff limit for GHD in adults. The data from the three patients with undetectable GH responses in both assays are not included.
SERA FROM UNSTIMULATED HEALTHY ADULTS
The ratios from measurements of unstimulated healthy individuals are shown in Fig. 3A. For concentration values >1 mIU/L, the distribution of ratios is close to the distribution of stimulated healthy adults (1.73 ± 0.23; n = 24), but premenopausal women taking oral contraceptives have a slightly higher mean ratio (1.85 ± 0.20; n = 8). However, all these points fulfulled the biological criterion, and only one point was outside the analytical limits. The results from the other groups as well as the total distribution clearly satisfied the criterion for values >1 mIU/L.

SERA FROM UNSTIMULATED PATIENTS
Ratios from measurements of unstimulated patients are shown in Fig. 3B. For concentration values >1 mIU/L, the distribution of these ratios is close to the distribution of ratios in stimulated healthy adults, 1.60 ± 0.29 (n = 11). Of the 20 patients with peak GH responses ≥1 mIU/L, 15 patients had no detectable GH. Data from these 15 patients are not included.

Discussion
It was possible to establish a conversion factor that fulfilled the acceptability criteria based on inherent analytical imprecision. Our evaluation indicates that the two methods can be used as substitutes for each other (with the factor) even for values <1 mIU/L.

Fig. 3. Ratio plots comparing the results of basal GH concentrations using two different methods, the polyclonal assay (Pharmacia) and the monoclonal assay (Delfia).

The lines represent the same factors as those in Fig. 2. (A) The basal GH ratios in 40 healthy adults are plotted against basal GH concentrations measured by Pharmacia. The 40 adults were selected into five groups: (■), men ≤40 years; (□), men >40 years; (●), premenopausal women taking no oral contraceptives; (▲), premenopausal women taking oral contraceptives; and (●), postmenopausal women. (B) The basal GH ratios in 16 patients with suspected GHD are plotted against basal GH concentrations (Pharmacia). The data for the 15 patients with no measurable GH concentrations are not shown in the figure.
lytical imprecision and on analytical quality specifications (8,9) for the relationship between the peak GH responses obtained using the Pharmacia and Delfia assays in healthy adults. In patients with GHD, who have few responsive somatotroph cells, the conversion factor was also applicable, suggesting that the relative concentration of the 22-kDa form during peak GH responses to the PD-GHRH test remained the same regardless of the number of somatotroph cells. The PD-GHRH test was able to elicit peak GH responses in all healthy adults (5), and the composition of stimulated GH is reasonably constant (2) compared with the basal state of the somatotroph cell (3). In agreement with the higher biological variation in basal samples, the ratio points in healthy adults were less evenly distributed around the calculated mean for the ratios, 1.59. However, the ratios were within the 95% prediction interval. Visual inspection of the plot revealed slightly higher ratios in premenopausal women who were taking oral contraceptives compared with the other four groups. In women on oral estrogen therapy, the GH-binding protein (GHBP) concentrations may be higher (15,16). Higher GHBP concentrations in vitro (4,17) would probably not change the ratios between the monoclonal and polyclonal assays. However, in vivo the net effect of changes in high-affinity GHBP concentrations is difficult to assess because an increase of the corresponding GH receptor would increase clearance of the 22-kDa form, in contrast to the prolongation of t1/2 by higher circulating GHBP concentrations (18,19). Although the conversion factor was applicable for stimulated GH responses and for unstimulated GH concentrations >1 mIU/L in adult patients with suspected GHD, it is necessary to emphasize that we have not studied children with suspected GHD and patients with adenomatous GH secretion. The comparison of samples with peak GH responses as well as with basal values reveals an unexpected consistency covering both intra- and interindividual biological variation under extreme conditions, showing a total CV of ratios, which to a large extent can be explained by simple inherent analytical imprecision.

In addition to our work, others (4,20) have shown that the use of the 22-kDa form as a standard cannot alone explain the difference between measured GH concentrations using monoclonal and polyclonal assays. However, it is always preferable to have a standard that is as pure and reproducible as possible. Recently, the WHO Expert Committee on Biological Standardization established the preparation coded 88/624 as the first International Standard for Somatropin (Recombinant DNA-derived Human Growth Hormone) (21).

Some previous studies also examined the relationship between GH measurements in monoclonal and polyclonal assays, because marked discrepancies were noted only after monoclonal assays came into widespread use (4,22,23). However, in method comparison, neither the correlation coefficient nor techniques such as regression analysis is appropriate (7). The Pharmacia (polyclonal) and Delfia (monoclonal) assays were compared in a paper by Albertsson-Wikland et al. (22). A high coefficient of correlation was almost guaranteed (7) because the entire range of basal (24-h profiles) and stimulated GH concentrations from young females and males was investigated. They used the same calibrator that we used.

In the same laboratory, Jansson et al. (4) used this standard (WHO 80/505) and reported the comparison of the same two assays using basal samples (24-h profiles) from young females. Their results may be from the same study (22). From the reported regression coefficient of 0.78 (4), we have calculated the mean of ratios, 1/0.78 = 1.28 (their Fig. 2D). This ratio is considerably lower than our estimation of 1.59, and the scatter of points about the regression line is much wider than the scatter of our ratios about the mean ratio. These differences are more distinct than expected from otherwise comparable studies using patient samples to compare two analytical methods for measuring the same component. However, our results are in agreement with the comparison in the same paper (4) of another monoclonal assay (Pharmacia mono) with the polyclonal assay (Pharmacia; their Fig. 2C).

In a study of GH assays by Barth et al. (23), no comparisons were made between Pharmacia and Delfia. However, both of these assays were compared with IDS Gamma BCT, which is essentially a monoclonal assay that does not recognize the 20-kDa GH (24). They used difference plots to present data and found that the difference between Pharmacia and IDS Gamma BCT was greater at stimulated GH concentrations >20 mIU/L. They did not characterize this further, but their results do not conflict with ours because we found that the Pharmacia/Delfia ratio was constant throughout the measured concentration range. In their study, the difference plot is not so informative. More information would probably have been obtained using a ratio plot. Moreover, they used very weak stimulation tests, which, in contrast to the PD-GHRH test (5), will not stimulate GH secretion in all healthy individuals (25).

The advantage of using ratio plots rather than difference plots is evident in situations where the analytical CV values are (or are assumed to be) constant over large concentration ranges and when a proportional relationship is expected. When compared with difference plots of ln-values, the advantage is that the ratios allow immediate interpretation of the data (for expected CV values of ratios up to ~0.15 (15%), whereas the information from a ln-difference plot must first be anti-ln-transformed to be interpreted. An additional advantage of the ratio plot is the clear illustration of the absence of concentration dependency. The lack of concentration dependency supports the hypothesis that, except for the conversion factor (1.59), the GH concentration measurements obtained using the Delfia and Pharmacia methods largely remain within the inherent imprecision of the two methods.

For small values of analytical imprecision (CV<0.1 ~10% for both the numerator and the denominator) the use of simple calculations for the limits of the 95% prediction
interval is justified by the relevant comparison of the calculated \( CV \) of measured ratios with the predicted \( CV \) based on imprecision alone (\( CV_{\text{Pharmacia}}^2 + CV_{\text{Delfia}}^2 \)), which makes interpretation easy. For values >20 mIU/L, the effect of the simple calculation method compared with the correct method (10, 11) is that the ratio for the lower limit is 1.365 instead of 1.374 and that 2.0% of the values will be below this limit instead of the ideal of 2.5%. The corresponding results for the upper limit were 1.815 instead of 1.825 and 3% variance instead of 2.5%. For concentrations <1 mIU/L, the percentages for the lower and upper limits were 1.9% and 3.4%, respectively, compared with the correct value, 2.5%.

This study has demonstrated the use of ratio plots in methods comparison. The importance of a step-by-step evaluation was demonstrated by the use of well-characterized samples from healthy volunteers and patients, according to the expected distribution of ratios based on analytical imprecision and on analytical quality specifications. We have performed an appropriate data analysis for the evaluation of comparability between two GH assays: a polyclonal assay (Pharmacia) and a monoclonal assay (Delfia). We found that it was possible to use a conversion factor to alternate between the two methods without losing diagnostic power from the GH measurements.

We thank Ringversuche (Deutsche Gesellschaft für Klinische Chemie E.V.) for their excellent cooperation. We thank Hanne Peter and Kirsten Pedersen for excellent technical support. This work was supported by the Clinical Institute, Odense University.

References