Problems with GH assays and strategies toward standardization

Martin Bidlingmaier
Endocrine Research Laboratories, Medizinische Klinik, Innenstadt, Ludwig-Maximilians University, Ziemssenstreet 1, 80336 Munich, Germany

Abstract
Disorders affecting GH secretion – either GH deficiency or GH excess (acromegaly) – are biochemically defined through peak or nadir concentrations of human GH in response to dynamic tests. Immunoassays employing polyclonal or monoclonal antibodies are routinely used for the analysis of GH concentrations, and many different assays are available on the market today. Unfortunately, the actual value reported for the GH concentration in a specific patient’s sample to a large extent depends on the assay method used by the respective laboratory. Variability between assay results exceeds 200%, limiting the applicability of consensus guidelines in clinical practice. Reasons for the heterogeneity in GH assay results include the heterogeneity of the analyte itself, the availability of different preparations for calibration, and the interference from matrix components such as GH-binding protein. Furthermore, the reporting of results in mass units or international units together with the application of variable conversion factors led to confusion.

International collaborations proposed measures to improve the comparability of assay results, recommending the use of a single, recombinant calibrator for all assays and reporting only in mass units as first steps. However, because of the differences in epitope specificity of antibodies used in different assays, method-specific cut-off levels for dynamic tests might remain necessary to correctly interpret and compare results from different laboratories.

It is widely accepted that the diagnosis of a disease as well as the respective therapeutic decisions should be based on well-defined criteria, ideally based on evidence of best practice reflecting outcome of international efforts to establish guidelines and consensus documents. Accordingly, several consensus statements on diagnosis and treatment of GH-related disorders have been published (1–3). All these documents recommend the use of a biochemical definition within an appropriate clinical context for both GH deficiency and GH excess, with the biochemical definition being based on the measurements of GH concentrations during dynamic tests (stimulation or suppression) rather than in random samples. Interestingly, the consensus documents refer to published cut-off values for the interpretation of test outcome, but at the same time express concerns regarding the comparability of GH concentrations as measured by different assays. This might sound inconsistent, but reflects an existing dilemma: the weakness of the evidence bases behind many cut-off values to a large degree can be attributed to the discrepancy between the GH concentrations obtained from different assays (4–6). External quality assessment schemes introduced in many countries to monitor the quality of laboratory testing unfortunately mainly document the continuation of disparity and the lack of standardization. Great variability in GH results from different assays has been reported from several countries by different laboratories participating in national external quality assessments (7–9). It is noteworthy that the disparity between GH assay results apparently worsened, especially with the introduction of newer assays based on monoclonal antibodies (10). The method-dependent variability exceeds 100% in many cases, making it impossible to compare results from one study with another or – relevant in clinical practice – from one hospital to another when different laboratories are used.

Of course, the problem of standardization of immunoassays is not restricted to GH (11), but the measurement of GH involves a combination of classical problems: first, the analyte GH is not a homogenous molecule. On the contrary, GH in circulation consists of a wide variety of molecular isoforms (12, 13). Dimers of GH molecules occur, both as hetero- and homodimers, and also multimers are present in human serum (14). Consequently, as specific antibodies used in immunoassay recognize a specific epitope present on the surface of the antigen, each antibody will pick a selected spectrum of isoforms only. This spectrum is most likely broader for
polyclonal antisera (representing a mixture of different antibodies) than for very specific monoclonal antibodies — which might explain the observation that discrepancy between assay results worsened with the introduction of assays based on monoclonal antibodies (10). Also in line with this, the best overall agreement between GH assays was found when the less specific, polyclonal antisera were used (15, 16). There is ongoing debate whether a more specific or a more permissive recognition of GH isoforms is preferable from a clinical point of view. Although a very specific assay based on monoclonal antibodies provides a better definition of what exactly is measured in a sample, different isoforms are biologically active (17). Therefore, we are potentially missing information with assays measuring only one isoform (5).

A second factor interfering with GH assays and contributing to discrepancy between results is the high-affinity GH-binding protein (GHBP), which circulates in human serum (18). It corresponds to the extracellular domain of the GH receptor, and its concentration varies with nutritional and metabolic conditions (19). In circulation, up to 50% of GH is complexed with GHBP; and therefore epitopes might not be accessible for certain antibodies due to steric hindrance. Potentially, this could lead to underestimation of GH concentrations. Especially, assays using monoclonal antibodies in conjunction with a very short incubation time (as in modern automated analyzers) are susceptible to interference from GHBP. The potential negative bias is considerable, approaching 50% in some assays for GHBP concentrations still in the physiological range (20–22).

A third important factor confounding the comparability of assay results is the standard preparation used to calibrate the assay. All immunoassays read the concentration of a hormone in the sample by comparing the signal generated in the sample with that generated by known amounts of the same hormone in the standard curve (calibrators). For GH assays, standard preparations of pituitary origin are widely used. These preparations (International Standard (IS) 66/217 introduced in 1969 and IS 80/505 introduced in 1982 therefore resemble a variety of GH isoforms, although the 22 000 kDa isoform is somewhat enriched. Because of the unknown ‘true’ GH content of the pituitary extracts, the concentrations were arbitrarily chosen at 2 and 2.6 U/mg respectively. More recently, the preparation IS 88/624, which is recombinant 22 000 kDa GH only, was introduced and assigned to a biopotency of 3 U/mg. IS 88/624 was finally replaced by recombinant IS 98/574. The latter preparation today is the reference material of choice to ensure traceability of GH assay results. Obviously, the change in the standard preparation has major influence on the absolute concentrations reported by different assays (9), with results getting generally lower compared with the previous – pituitary-derived – calibration. However, the conversion between the different units provides an additional source of problems in comparing GH assay results; depending on regional or national preferences, GH concentrations have been reported in units/l by some laboratories, but in µg/l by others, with unclear conversion factors in many cases. The variety of standard preparations and units apparently led to considerable confusion among clinicians and scientists when reporting data from clinical studies in international journals. Pokrajac and colleagues (23) found five different factors applied for conversion of GH data from mU/l to µg/l in four major articles in a leading endocrine journal, all published within 5 years.

The heterogeneity of GH assay results is not only an academic problem for scientists, laboratories or clinical chemists. Several studies have shown that difficulties in comparing GH concentrations obtained by different assays have major impact upon clinical decisions, and therefore are relevant confounders in optimizing patient care (24–26). When directly comparing results from different assays in the same clinical samples in one laboratory, disagreement between GH assay methods was confirmed. Furthermore, it was demonstrated that the application of cut-off values from consensus statements led to different clinical decisions than the application of method-specific cut-off values. For example, Markkanen and colleagues analyzed 71 GH nadir samples from oral glucose tolerance tests performed in patients with acromegaly by two different immunoassays (26). For interpretation, they used the ‘consensus criterion’ of GH suppression <1 µg/l for either assays, or they used method-specific cut-off values derived from tests in a cohort of healthy volunteers. Depending on the cut-off values used, the percentage of patients considered ‘in remission’ varied from 48.1 to 85.2% in males and 56.8 to 68.2% in females. The higher discrepancy in males is a consequence of the lower nadir values in healthy males compared with females. In another study in children with GH deficiency (27), peak samples from GH stimulation tests, which had been measured by local laboratories, were reassessed centrally. Based on the central laboratories GH results obtained by one reference assay, 36 out of 132 patients were categorized differently, with 35 patients changing into the GH-deficient group. The situation is of concern especially because the awareness of the problem is still limited, and through the introduction of large central laboratories many clinicians have little impact on where and with which assay samples are measured. A survey among the members of the European Society for Pediatric Endocrinology (28) revealed that only 63% of the participants knew their local GH assay, and even less against which standard their local assay was calibrated. It is noteworthy that the most frequently quoted cut-off values used for the interpretation of stimulation test results in the diagnosis of GH deficiency were 10 µg/l.
and 20 mU/l – regardless of which calibrator the local laboratory used!

In order to improve the comparability of study results from several centers, some authors recommend the central reassessment of samples by a defined GH assay method (27) – an approach that is obviously impractical for patient care on a daily basis. Several guidelines emphasize the need for ‘method-specific cut-off values’ (1, 29–31), but in real life for most of the assays presently used, there are no published studies or reference data available. In the context of differences in epitope specificity of the antibodies used in different assays, the concept of ‘commutability’ deserves attention (32). A practical application of this concept on GH assays has been reported recently (33): laboratories obtained aliquots from the so-called ‘harmonization samples’ (representing ‘average patient samples’), which were measured by all laboratories and used to correct their respective GH assay results from other samples. Using this approach, between-laboratory variability for GH assay results for those samples could be reduced from 15% to <7%. For this concept, the availability of sufficient amounts of adequate harmonization samples is required, and the applicability in a larger international setting should be investigated.

An international collaborative published a summary of the discussions and identified possible first steps toward assay standardization more likely to be implemented (34, 35). The use of a single, recombinant calibrator for all GH assays (IS 98/574) and the reporting of all GH assay results in mass units (µg/l) of the recombinant calibrator are foreseen to be the first important steps toward standardization. Changing to a uniform standard preparation has been shown to lead to a significant reduction in between-laboratory variability in GH assay results from above 35% to below 20% (9). Support of the initiative came from some important journals in the field, which have decided to publish papers on GH data only if these are expressed in mass units of the most recent IS 98/574 (34, 35). Meanwhile, several manufacturers of GH assays announced adoption of the new recombinant calibrator, at least for the European market. Unfortunately, several manufacturers – including the market leader – did not implement the change in the calibrator in all countries at the same time. Thus, GH assays from the same manufacturer now are available with different calibrations in different European countries. If and when the corresponding change will be implemented for assays available on the US market is still unclear. Therefore, researchers must be careful in reporting the correct calibration for GH assays used at the time the samples from their studies have been measured. Furthermore, it remains a primary task for editors and reviewers of scientific journals to ensure that manuscripts reporting GH assay data contain a clear statement regarding the assay method, and also refer to appropriate cut-off values for dynamic tests. Meanwhile, also the International Federation of Clinical Chemistry and Laboratory Medicine has put the issue of GH assays on the agenda by implementing a working group on GH assay standardization. The collaboration with clinical chemists is of importance because – as mentioned above – in many cases the hormone analyses are in some type of large centralized laboratory far away from the clinician.

As outlined above, there are problems with GH assays, which cannot be solved solely by using a single calibrator for all assays. It remains to be investigated to which degree the ‘standardization of the standard’ alone leads to a better comparability of GH assay results, and what the impact upon cut-off values and clinical decisions will be. It is very likely that future activities must address the issue of differences in epitope specificity of different antibodies used to measure a heterogenous substance like GH. From a practical point of view, presently there is no alternative to increasing the awareness of the problem. Knowledge about the GH assays used in each local laboratory and its properties with respect to the application of cut-off values remains mandatory for clinicians to avoid misinterpretation of GH assay results.

Disclosure
This paper forms part of a European Journal of Endocrinology supplement, supported by Ferring. The authors disclose: no potential conflicting relationship with Ferring. This article was subject to rigorous peer review before acceptance and publication.

References


33 Ross HA. Reporting growth hormone assay results in terms of one consensus recombinant standard preparation offers less than optimal reduction of between-method variation. Clinical Chemistry and Laboratory Medicine 2008 46 1334–1335.
