Clinica Chimica Acta 432 (2014) 72–76

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Harmonization of growth hormone measurement results: The empirical approach

H.A. Ross [a,*], E.W.G.M. Lentjes [b], P.M.M. Menheere [c], C.G.J. Sweep [d], on behalf of the Endocrinology Section and project group "Calibration 2000" of the SKML (Dutch Foundation for Quality Assessment in Clinical Laboratories)

[a] Department of Laboratory Medicine, Laboratory of Genetic, Endocrine and Metabolic Diseases, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[b] Department of Clinical Chemistry and Hematology, University Medical Centre Utrecht, Utrecht, The Netherlands
[c] Department of Clinical Chemistry, University Medical Centre Maastricht, Maastricht, The Netherlands
[d] Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 2 March 2013
Received in revised form 28 December 2013
Accepted 5 January 2014
Available online 6 February 2014

Keywords:
Growth hormone
Commutability
Harmonization

ABSTRACT

Growth hormone (hGH) is a measurand belonging to ISO category 4, indicating intrinsic unavailability of a reference measurement procedure and primary standard material. Large between-method differences have been raising confusion, especially in the interpretation of results of stimulation tests for exclusion of juvenile growth hormone deficiency.

Within the framework of the external quality assessment scheme (EQAS) of the SKML (Dutch Foundation for Quality Assessment in Clinical Laboratories), attempts to reduce between-method variation of hGH measurements have been made, starting in 1994 with an inter-laboratory comparison of 9 different immunoassays by using a panel of sera and standard materials available at that time. Methods appeared to differ from each other largely in a systematic, sample-independent manner. These systematic differences are reflected in the hGH measurement results obtained in commutable sera. A commutable serum pool was introduced as a consensus reference material, permitting correction of each method's results to a common scale. Pair wise comparisons ("twin studies") were carried out to investigate and corroborate the effectiveness of this material for harmonization. A significant reduction of the between-laboratory coefficient (CV) of variation from 22 to 9.0% was attained.

1. Introduction

Before recombinant hGH preparations became available on a larger scale, treatment of GH deficiency depended on expensive pituitary hGH preparations of strictly limitedly availability. In The Netherlands, access to treatment was subject to regulation by a semi-governmental institution (*Nederlandse Groeistichting*, Dutch Growth Foundation, since 2007, Dutch Child and Growth Foundation). The outcome of GH stimulation tests was crucial in this respect: after some years of trial a serum GH concentration exceeding 20 mIU/L was considered as excluding GH deficiency. Since initially, clinical laboratories employed radioimmunoassays based on the same anti-hGH antiserum [1] and the same reference standard (known as 1st IRP 66/217), this decision point was more or less unambiguous. With the advent of immunometric techniques and automation, GH assays diversified whereas the 20 mIU/L limit still kept its pivotal function. Inevitably this must have led to unequal chances for patients to be assigned to treatment, depending on the method used by the laboratory that performed the GH assay. Moreover, after recombinant preparations for treatment of GH deficiency had become widely available, this situation was generally considered unacceptable. The use of common reference preparations was found to be ineffective in reducing the differences in GH results between methods. The Dutch Growth Foundation sought advice from the Endocrinology Section of the SKML (Dutch Foundation for Quality Assessment in Clinical Laboratories). This group initiated a study into the possibilities for reduction of between-laboratory variability of GH results, especially at concentrations close to the decision point. It was considered beforehand that this reduction can be achieved only if the between-method differences are largely of a systematic nature. Therefore the primary goal was to ascertain whether this indeed was the case. Once the systematic nature of differences had been established, harmonization in principle may be achieved both by establishment of method-specific factors and by common use of a commutable reference material, enabling laboratories to express their results on a common scale. Moreover, a harmonization procedure should be simple to implement. Adjustment

Abbreviations: GH, growth hormone; hGH, human growth hormone; IEMA, immunoenzymometric assay; IFMA, immunofluorometric assay; ILMA, immunoluminometric assay; RIA, radioimmunoassay; IS, International Standard; EQA, External Quality Assessment; PT, proficiency testing; RMP, reference measurement procedure.

* Corresponding author at: Dept. of Laboratory Medicine, 479 Laboratory of Genetic, Endocrine and Metabolic diseases, Radboud University Nijmegen Medical Centre P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. Tel.: +31 24 3614276.
E-mail address: alec.ross@radboudumc.nl (H.A. Ross).

0009-8981/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cca.2014.01.008

of a laboratory's result to the common scale should not involve more than a single multiplication factor. The first study aimed at identifying a suitable standard preparation among a number of candidates available at that time (1994) for construction of calibration curves in all assay methods. Simultaneously it was studied to what extent the between-method differences are of a systematic nature and how individual sample properties contribute to overall variability. It was found that none of the preparations tested (whether recombinant or of pituitary origin) led to improvement of between-method variability. By contrast, harmonization factors derived from between-method regression coefficients resulted in a reduction of the between-method CV by roughly one-third, indicating that systematic differences contribute significantly to between-method variation. Repetition of the experiment disclosed that method-specific factors are not constant in time, probably due to assay drift. However, it could also be demonstrated that up-to-date estimates of regression coefficients can be obtained from parallel measurements in serum pools. This indicates at least partial commutability of these pools. A special pool serum was composed from donations of healthy individuals during exercise, in order to be able to attain a concentration close to the decision limit of 20 mIU/Ll. To this candidate harmonization material, the mean result (5.84 μg or 17.5 mU per liter) of 7 different assay methods, using the recombinant 1st I.S. 88/624 for constructing standard curves, was assigned as a consensus value. Several tests followed to ascertain commutability of this material and its effectiveness in reducing between-method variation.

A final "twin" study concluded the test phase of the actual implementation of the harmonization process in 2004.

2. Methods

2.1. Assays and samples

Participants in the SKML EQAS for Endocrinology who routinely measured hGH each received 5 aliquots of the lyophilized harmonization serum per year, allowing for inclusion in 10 assay runs. The harmonization factor is obtained by dividing the consensus value of 17.5 mU/L by the concentration measured. A moving average (avg) is obtained as follows:

$$\text{new avg} = 0.8 \times \text{old avg} + 0.2 \times \text{new value}$$

Ten pairs of labs each exchanged 10 patient samples with concentrations ranging from about 5–30 mIU/L. In 9 pairs, LIEMA Immulite 2000 was present. The other assay methods were: LIEMA Immulite1 (5×); IFMA Wallac (3×); RIA In-house (2×) and ILMA Nichols (1×). One in-house RIA was combined with Immulite 1. The effect of harmonization on the between-method CV was evaluated using the harmonization factor reported by each participant.

2.2. Calculations

2.2.1. Regression analysis

Regression coefficients for each couple of methods were obtained by consecutively leaving one sample out before calculating a regression coefficient, so that for each method or method pair a number (equal to the number of samples in the study) of slightly differing regression coefficients was obtained. Thus, for each particular sample a corresponding regression coefficient was calculated from the data of the other samples. This was to prevent the regression coefficient being biased, which would result in an overly favorable estimate of the minimum between-method variation. This consideration is especially important with small numbers of samples. Initially, Deming's regression analysis, adapted for a single parameter model was performed. To obtain regression coefficients that minimize relative rather than absolute distances to the regression line, the antilogarithm of the average difference between log(y) and log(x) was taken. The scatter of points about the regression line was derived from the perpendicular distances of data points to the line (see Section 2.2.2) and compared with the expected scatter due to state-of-the-art imprecision only. The concentration-dependent state-of-the-art CV was calculated according to Steigstra et al. [2].

2.2.2. Commutability

The perpendicular distance of the data point of sample *i*. to the regression line is given by:

$$D_{ij} = \sqrt{\left(y_{ij} - a_{j(-i)} \cdot x_{ij}\right)^2 / \left(1 + a_{j(-i)}{}^2\right)}$$

Here, $a_{j(-i)}$ represents the regression coefficient obtained for method pair *j* excluding sample *i*. Distances were calculated for all data points, both absolute and relative to the level of serum *i*. For the harmonization serum a virtual data point was created by division of the consensus value by the reported harmonization factors for that method pair, so that distances also could be calculated for the harmonization serum.

According to Baadenhuijsen et al. [3], a material is commutable if this distance (expressed as a percentage of the measured value) does not exceed the state-of-the-art CV by more than a factor of 3. Following CLSI Guideline 53a [4], the distance should fall within the confidence interval for the distances of the other data points. Both criteria are evaluated.

2.3. Between-method CV

The between-method CV is obtained from the expression for between-method variance in a twin design:

$$V_{betw(j)} = \frac{\sum_{i=1}^{n} \left(y_{ij} - a_j \cdot x_{ij}\right)^2 / 2}{n}$$

after taking the square root of *V* and dividing by the corresponding mean concentration. In the case of raw, unharmonized data $a_j = 1$ for all *i* and *j*, for calculation of the optimal CV, a_j is replaced by $a_{j(-i)}$, the regression coefficient for method pair *j* with sample *i* left out, and for harmonized CV it is the quotient of harmonization factors for the two methods. Accordingly, the corresponding mean concentrations necessary for calculating the relative distances are the mean of y_{ij} and $a_j \cdot x_{ij}$. Not only CV's for method couples were evaluated. After sorting the data according to GH concentration, three equally sized groups ("low", "medium" and "high") were created and mean optimal CV's and harmonized CV's for these groups were calculated.

3. Results

3.1. Regression analysis

With respect to Immulite 2000 assays, regression coefficients (relative distances minimized) ranged from 0.63 to 1.02. One sample out of 100 exceeded by far the confidence interval of perpendicular distances of its method couple and therefore had to be excluded as an outlier. The scatter about the regression lines in a perpendicular direction did not exceed the state-of-the-art CV, except in one of the pairs (Immulite 1 vs RIA2), indicating that the single parameter regression model was adequate at least in all 9 other cases. Closer inspection of the outlier couple shows that a two parameter model is inappropriate: slope and intercept do not differ from zero, whereas the one-parameter model gives a significant slope. Therefore this couple was not excluded from further analysis.