

# Within-Subject Variability and Analytic Imprecision of Insulinlike Growth Factor Axis and Collagen Markers: Implications for Clinical Diagnosis and Doping Tests

Tuan V. Nguyen,<sup>1</sup> Anne E. Nelson,<sup>1</sup> Christopher J. Howe,<sup>2</sup> Markus J. Seibel,<sup>3</sup> Robert C. Baxter,<sup>4</sup> David J. Handelsman,<sup>3</sup> Ray Kazlauskas,<sup>2</sup> and Ken K. Ho<sup>1\*</sup>

**BACKGROUND:** The utility of insulinlike growth factor (IGF) axis and collagen markers for a growth hormone (GH) doping test in sport depends on their stability and reproducibility. We sought to determine short-term within-subject variability of these markers in a large cohort of healthy individuals.

**METHODS:** We measured IGF-I, IGF binding protein 3 (IGFBP-3), acid labile subunit (ALS), and the collagen markers N-terminal propeptide of type I procollagen (PINP), C-terminal telopeptide of type I collagen (ICTP), and N-terminal propeptide of type III procollagen (PIIINP) in serum samples obtained on multiple occasions (median 3 per participant) over a 2- to 3-week period from 1103 elite athletes (699 men, 404 women) ages 22.2 (5.2) years [mean (SD)]. We estimated between-subject and within-subject variances by mixed-effects ANOVA.

**RESULTS:** Within-subject variance accounted for 32% to 36% and 4% to 13% of the total variance in IGF markers and collagen markers, respectively. The within-subject CV ranged from 11% to 21% for the IGF axis markers and from 13% to 15% for the collagen markers. The index of individuality for the IGF axis markers was 0.66–0.76, and for the collagen markers, 0.26–0.45. For each marker, individuals with initial extreme measured values tended to regress toward the population mean in subsequent repeated measurements. We developed a Bayesian model to estimate the long-term probable value for each marker.

**CONCLUSIONS:** These results indicate that in healthy individuals the within-subject variability was greater for IGF-I than for the collagen markers, and that where a single measurement is available, it is possible to estimate the long-term probable value of each of the markers by applying the Bayesian approach. Such an application can increase the reliability and decrease the cost of detecting GH doping.

© 2008 American Association for Clinical Chemistry

Serum concentrations of insulinlike growth factor I (IGF-I),<sup>5</sup> which together with IGF binding protein 3 (IGFBP-3) and the acid labile subunit (ALS) forms the circulating IGF-I ternary complex, are clinically useful markers for the management of growth disorders and metabolic diseases (1–3) and for cancer development and progression (4–6). Specific peptides related to collagen synthesis and degradation, such as the N-terminal propeptide of type I procollagen (PINP) and the C-terminal telopeptide of type I collagen (ICTP), are used clinically in the management of bone disease (7). The N-terminal propeptide of type III procollagen (PIIINP) is a marker of type III collagen turnover in connective tissue (8) that is used clinically to monitor growth (9) or clinical conditions such as liver fibrosis (10) and prognostically in atherosclerotic disease (11). The serum concentrations of these IGF axis and collagen markers also increase in response to growth hormone (GH) (12–15), indicating the potential of these

<sup>1</sup> Garvan Institute of Medical Research and Department of Endocrinology, Sydney, Australia; <sup>2</sup> Australian Sports Drug Testing Laboratory, National Measurement Institute, Sydney, Australia; <sup>3</sup> ANZAC Research Institute, Concord Hospital, University of Sydney, Sydney, Australia; <sup>4</sup> Kolling Institute of Medical Research, University of Sydney, Royal North Shore Hospital, Sydney, Australia.

\* Address correspondence to this author at: Pituitary Research Unit, Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst 2010 Australia. Fax 0612 9295 8481; e-mail K.Ho@garvan.org.au.

Received February 24, 2008; accepted May 17, 2008.

Previously published online at DOI: 10.1373/clinchem.2008.105726

<sup>5</sup> Nonstandard abbreviations: IGF-I, insulinlike growth factor I; IGFBP-3, IGF binding protein 3; ALS, acid labile subunit; PINP, N-terminal propeptide of type I procollagen; ICTP, C-terminal telopeptide of type I collagen; PIIINP, N-terminal propeptide of type III procollagen; GH, growth hormone; RTM, regression to the mean.

GH-responsive proteins in detecting exogenous GH administration (16).

The utility of a marker, both clinically and for a GH doping test, depends on its sensitivity and ability to discriminate a normal value from an abnormal value. The sensitivity and discriminatory values are, in turn, a function of the stability of these markers over time, which can be assessed by within-subject variability and analytic imprecision. Conceptually, the variation of a biochemical marker can be partitioned into 2 major sources, between-subject and within-subject variation, and the latter further partitioned into normal biological variation and analytical (random) imprecision of the method of measurement.

Despite the importance of IGF axis markers in clinical and scientific practice, the sources and magnitude of these markers' variability have not been well documented. One recent study in normal subjects reported considerable short-term within-subject differences in IGF-I and indicated the need for caution in the use of risk ratios based on single measurements in the clinical setting (17). Most studies to date on short- to long-term within-individual variability of bone turnover markers were based on older women (18–21). No data are available, however, on the stability and reproducibility of the IGF axis and collagen markers in young, healthy athletes; such data are required for the application of these markers to a GH doping test.

In this study, we sought to determine the short-term within-subject variability of the markers IGF-I, IGFBP-3, ALS, PINP, ITCP, and PIIINP in a large cohort of >1000 elite athletes. The within-subject variability was partitioned into biologic and analytic variability and was compared to the between-subject variability to assess the implications of this variability in the clinical setting and the use of these markers to detect GH doping in sport.

## Materials and Methods

### PARTICIPANTS

The participants [ $n = 1103$ , ages 22.2 (5.2) years, mean (SD)] were elite athletes, defined as having competed at state or regional level or higher during the previous 12 months. The athletes, who were at least 14 years old, were required to declare by questionnaire that they had not taken GH, IGF-I, erythropoietin, or any substance known to enhance red blood cell formation during the previous 2 months. The samples were originally collected as part of a study to determine reference ranges for markers of altered erythropoiesis in elite athletes (22), and we have previously defined the influence of demographic factors and sporting type on GH-responsive markers in this cohort (23). The participants represented 10 major sporting categories

and were recruited from 12 countries representing 4 major ethnic groups: Caucasian, Asian, African, and Oceanian and others, which included those who reported mixed ethnic origin and those who declined to report ethnic origin. The study was approved by the Ethics Committee of the Australian Institute of Sport and written informed consent was obtained; blood samples and demographic data provided for this study were coded and not personally identifiable.

### MEASUREMENTS AND LABORATORY ANALYSIS

Blood samples were collected from volunteers on a casual basis, that is, at random with regard to the time of day, food intake, exercise, and competition, as described (22), representing the out-of-competition setting. The day and time of collection and the time since the athlete had last exercised or competed were recorded. We collected 3 venous samples on average from each athlete over a 2- to 3-week period. The median number of samples per subject was 3 (range 1–10).

We stored serum samples at  $-80^{\circ}\text{C}$  before analysis. The stability of these analytes after storage at  $-80^{\circ}\text{C}$  has been indicated by previous studies (24, 25). We measured IGF-I by RIA after acid-ethanol extraction (26) and IGFBP-3 and ALS using polyclonal antibodies (27, 28). The intraassay CVs were IGF-I 5.9%, IGFBP-3 4.6%, and ALS 5.4%. The markers ICTP (intra- and interassay CVs <10%), PINP (intra- and interassay CVs <9% and <12%, respectively), PIIINP (intra- and interassay CVs <7% and <12%) were measured in serum by use of competitive RIAs (Orion Diagnostica) in the same assay batch.

### DATA ANALYSIS

The study design—collection of multiple measurements per subject over a relatively short time—allows comparison of the magnitude of variation that exists between individuals relative to the magnitude of variation within each individual. Accordingly, the total variance of each marker was partitioned into 2 broad components, between-subject and within-subject variance, with the latter being further partitioned into 2 subcomponents, 1 due to biological variation and 1 due to analytic variation. The within-subject variability was subsequently transformed into the original unit of measurement to yield the SD and then expressed as the CV relative to the mean. Estimates of variance components were obtained using a compound symmetric covariance structure in the mixed-effects ANOVA, using the nlme package within the R language (29). Full details of the analysis are shown in the Data Supplement (see Data Analysis) that accompanies the online version of this article at <http://www.clinchem.org/content/vol54/issue8>.

Baseline variable	Men	Women	All
n	699	404	1103
Age, years	22.6 (5.4)	21.4 (4.4)	22.2 (5.1)
Weight, kg	75.0 (14.6)	60.3 (8.5)	69.6 (14.5)
Height, cm	178.0 (9.2)	167.1 (8.2)	174.0 (10.2)
Body mass index, kg/m <sup>2</sup>	23.5 (3.3)	21.6 (2.2)	22.8 (3.1)
Ethnicity, n (%)			
Caucasian	354 (50.6)	236 (58.4)	590 (53.5)
Asian	223 (31.9)	126 (31.2)	349 (31.6)
African	82 (11.7)	27 (6.7)	109 (9.9)
Oceanian and others	40 (5.7)	15 (3.7)	55 (4.5)
IGF-I, $\mu\text{g/L}$	154.8 (45.6) [4.98 (0.31)]	161.9 (52.4) [5.02 (0.33)]	157.4 (48.3) [5.00 (0.32)]
IGFBP-3, mg/L	3.6 (0.60) [1.28 (0.18)]	3.9 (0.61) [1.34 (0.18)]	3.7 (0.62) [1.30 (0.18)]
ALS, nmol/L	283.7 (47.9) [5.63 (0.19)]	324.7 (58.8) [5.76 (0.20)]	298.8 (55.8) [5.68 (0.20)]
PINP, $\mu\text{g/L}$	121.0 (90.0) [4.60 (0.56)]	85.2 (54.9) [4.28 (0.51)]	107.8 (80.7) [4.48 (0.56)]
ICTP, $\mu\text{g/L}$	6.5 (3.0) [1.78 (0.40)]	5.6 (1.6) [1.68 (0.30)]	6.2 (2.6) [1.75 (0.37)]
PIIINP, $\mu\text{g/L}$	5.4 (2.3) [1.63 (0.34)]	5.1 (1.5) [1.58 (0.29)]	5.3 (2.0) [1.61 (0.33)]

<sup>a</sup> Values are mean (SD) unless noted otherwise. Figures in square brackets were calculated from logarithmic transformed data. Parts of this table have been published (23).

In addition, we computed the coefficient of reliability (R), which is defined as the ratio of the between-subject variance to the total variance (the sum of between- and within-subject variances) (30), for each marker. This coefficient can be viewed as the degree to which individuals' marker values remain relatively consistent over repeated measurements, or as a measure of the correlation between probable values and measured values. We also computed the index of individuality as described in the online Data Supplement (see Data Analysis). The presence of within-subject variability raises the issue of regression to the mean (RTM) (31) and the probable value for an individual. Therefore, in a further analysis, we assessed the RTM effect by comparing the change between consecutive measurements, and we estimated the probable value as described (32) using a Bayesian model, further detailed in the online Data Supplement (see Data Analysis).

## Results

Among the 1103 elite athletes in the study, approximately 63% were men and the majority (54%) were Caucasian, followed by Asian (32%), African (10%), and Oceanian and others (4.5%) (Table 1). Although the within-subject SDs were correlated with the means, there were no appreciable differences in the within-subject SDs across sexes or age groups. Also, because the analyses did not reveal any systematic variation be-

tween ethnicities or sport type with respect to between- and within-subject variances, the results are presented for the entire sample.

In all markers, the within-subject SDs were positively correlated with their means ( $r = 0.35\text{--}0.57$ ), which violated the assumption of homogeneity of variances. Therefore, subsequent analyses were performed on natural logarithmic-transformed values. All log-transformed IGF axis and collagen markers were normally distributed, with no appreciable difference in the between-subject variances across age groups.

### VARIANCE COMPONENTS

Between-subject variance formed a major component of the variance of the markers, accounting for 64% to 68% of the total variance of IGF axis markers and 87% to 96% of total variance of collagen markers (Table 2). Thus, within-subject variance accounted for between 4% and 36% of the total variance of the markers and was higher for the IGF axis markers (32%–36%) than for the collagen markers (4%–13%). Most of the within-subject variance was attributable to biological variability (80%–95% of the within-subject variance).

The short-term within-subject SD was relatively greater for IGF-I (CV 20.7%) than for IGFBP-3 and ALS (CV 11.3% and 11.6%, respectively) and ranged from 13% to 15% for the 3 bone turnover markers (Table 3). The CV due to analytic variability, which ranged from 3% to 6.1%, was smaller than the CV for

**Table 2. Estimates of variance components due to within-subject and between-subject sources (analysis based on original nontransformed data).<sup>a</sup>**

	Between-subject variance (% total variability)	Within-subject variance		
		Total within-subject (% total variability)	Biological variance (% within-subject variance)	Analytic variance (% within-subject variance)
IGF-I	1864.70 (63.7)	1063.20 (36.3)	970.39 (92.0)	92.80 (8.0)
IGFBP-3	0.31 (63.9)	0.18 (36.1)	0.14 (83.4)	0.035 (16.6)
ALS	2594.78 (68.4)	1200.90 (31.6)	905.03 (80.2)	295.87 (19.8)
PINP	6389.22 (96.2)	255.41 (3.8)	235.61 (92.8)	19.80 (7.2)
ICTP	6.65 (90.0)	0.66 (10.0)	0.63 (95.0)	0.035 (5.0)
PIIINP	3.93 (87.2)	0.58 (12.8)	0.52 (90.9)	0.058 (9.1)

<sup>a</sup> Components of variance due to biological and analytic variability were calculated as percent of within-subject variance.

biological variability (10%–20%). The coefficient of reliability was higher for the collagen markers (R 0.87–0.96) than for IGF axis markers (R 0.64–0.68) (Table 4), indicating that the within-subject variability was smaller relative to between-subject variability for the collagen markers. The index of individuality for all IGF axis markers was >0.60, whereas the index of individuality for all collagen markers was <0.6. None of the markers had an index of individuality >1.4 (Table 4).

#### REGRESSION TO THE MEAN EFFECTS

To assess the effect of RTM, we conducted an analysis of changes for subjects who had exactly 3 measurements. In this analysis, for each marker we categorized

the first measurement into 3 groups by the top 10% and bottom 10% percentiles; then, in each category, we calculated the mean of absolute change in each marker's values between the 3 measurements (Table 5). As expected from the regression-to-the-mean effect, individuals with initial measurement in the bottom 10% percentile for the marker subsequently increased on average; on the other hand, those in the top 10% percentile subsequently decreased. For example, among individuals with initial IGF-I <107  $\mu\text{g/L}$  ( $n = 76$ ), there was a mean increase of 10.5 (SD 26.9)  $\mu\text{g/L}$  between measurements 2 and 1, which then regressed toward the mean with a mean change of  $-0.8 \mu\text{g/L}$  between measurements 3 and 2. On the other hand, among individuals with initial IGF-I >230  $\mu\text{g/L}$  ( $n = 74$ ), a decrease of 65.3 (SD 62.6)  $\mu\text{g/L}$  between measurements 2 and 1 was observed, and this decrease was less in the subsequent measurements.

In an additional analysis, changes between measurement 2 and measurement 1 ( $\Delta_{2-1}$ ) were correlated with changes between measurement 3 and measurement 2 ( $\Delta_{3-2}$ ). In each subject,  $\Delta_{2-1}$  was classified as decreased if the percent decrease was more than within-subject 1SD, increased if the percent change was more than 1SD, or unchanged if the change was within 1SD (Supplemental Table 6). The common trend emerging from this analysis was that many subjects whose markers' values exhibited a decrease during the first 2 measurements (negative  $\Delta_{2-1}$ ), actually increased in subsequent measurements (positive  $\Delta_{3-2}$ ), and vice versa. For example, for IGF-I, 22% (168/751) of subjects experienced decreased  $\Delta_{2-1}$ , and among these subjects, 18% (31/168) subsequently experienced increased IGF-I values  $\Delta_{3-2}$ . On the other hand, 13% (96/751) of subjects experienced increased  $\Delta_{2-1}$  IGF-I, and of these 55% (53/96) subsequently decreased between measurements 3 and 2 [ $\Delta_{3-2} - 12.4\%$  (21%)].

**Table 3. Within-subject SDs and CVs (analysis based on original nontransformed data).**

	Within-subject variability	Biological variability	Analytic variability
Within-subject SD			
IGF-I, $\mu\text{g/L}$	32.6	31.2	3.10
IGFBP-3, $\text{mg/L}$	0.42	0.37	0.43
ALS, $\text{nmol/L}$	34.6	30.1	4.14
PINP, $\mu\text{g/L}$	16.0	15.3	2.11
ICTP, $\mu\text{g/L}$	0.81	0.79	0.43
PIIINP, $\mu\text{g/L}$	0.76	0.72	0.49
CV			
IGF-I	20.7	19.8	6.1
IGFBP-3	11.3	10.1	5.0
ALS	11.6	10.1	5.8
PINP	14.8	14.2	4.1
ICTP	13.2	12.8	3.0
PIIINP	14.3	13.6	4.5

**Table 4. Coefficient of reliability and index of individuality.**

Marker	Analysis based on original data		Analysis based on log-transformed data	
	Coefficient of reliability	Index of individuality	Coefficient of reliability	Index of individuality
IGF-I	0.64	0.76	0.64	0.76
IGFBP-3	0.64	0.75	0.68	0.69
ALS	0.68	0.68	0.70	0.66
PINP	0.96	0.20	0.94	0.26
ICTP	0.91	0.32	0.87	0.39
PIIINP	0.87	0.38	0.83	0.45

#### ESTIMATES OF PROBABLE VALUES

Using the Bayesian model as described in the online Data Supplement, we estimated probable values of each marker (Fig. 1). As expected from the model, individuals with observed marker values higher than the population mean had lower probable values than those actually measured, whereas individuals with measured values lower than the population mean had higher probable values. For example, an individual with measured IGF-I of 60  $\mu\text{g/L}$  (about 2SD less than population mean) can be expected to have a probable value of 84  $\mu\text{g/L}$ , whereas for an individual with measured IGF-I 250  $\mu\text{g/L}$  (about 2SD greater than population mean), the expected probable value is 200  $\mu\text{g/L}$ . Individuals with measured values closer to the population mean had estimated probable values closer to the measured values. As expected, for markers with higher within-subject variability (such as IGF-I), the observed and estimated probable measurements were further apart than for those markers with lower within-subject variability (such as PINP). The application of the concept of probable value in the detection of abnormal values is further illustrated in the online Data Supplement.

#### Discussion

All biological measurements are subject to within-subject random variation, and estimation of this variation has important clinical and scientific applications. In this study of short-term variability in 1103 healthy adults, within-subject variance ranged from 32% to 36% (for IGF axis markers) and from 4% to 13% (for collagen markers), and when expressed in the original unit of measurement (i.e., CV), the variation was between 11% and 21% of the mean. The coefficient of reliability was higher for the collagen markers than for the IGF axis markers. Regression to the mean effects was observed, and probable values were estimated for individuals based on a Bayesian model.

Although markers of the IGF system have been used extensively in the clinical setting, few studies have examined their reliability. In a study where IGF-I measurements were repeated after 2 weeks in 84 normal subjects, substantial within-subject variation between measurements was observed, resulting in changes from one quartile to another for 40% of the subjects, although a formal estimate of variability was not provided (17). In the current study, the within-subject variability for IGF-I was also high (within-subject CV 21%), whereas the within-subject CVs of IGFBP-3 and ALS were lower (11% and 12%, respectively). The underlying reasons for the relatively high within-subject variation in IGF-I in this study are unknown, but it is interesting that most of the within-subject variation was due to normal biological variation, and the imprecision of measurement accounted for only 6%. The estimates of the short-term within-subject variability in serum bone turnover markers in this study (within-subject CV 13%–15%) are similar to previous estimates in older women and compare favorably to the greater variability observed in other markers such as the C- and N-terminal telopeptides of type I collagen, CTx and NTx (20, 21, 33).

Increased serum concentrations of IGF axis and collagen markers in response to GH administered to young recreational athletes (12–15) have indicated the potential for these markers in detecting GH abuse in sport. Use of these markers in a GH-doping test requires estimation of between-subject variability in elite athletes to define demographically relevant reference ranges, and we have previously demonstrated the influence of demographic factors and sport type on between-subject variation in the same cohort of elite athletes (23). There has been no previous formal evaluation of within-subject variability of the GH-responsive markers, although others have suggested that a single measurement of serum concentrations of IGF-I and IGFBP-3 may not be sufficiently reliable for

**Table 5. Change between measurements 2 and 1 ( $\Delta_{2-1}$ ) and between measurements 3 and 2 ( $\Delta_{3-2}$ ) classified by initial levels of each marker.**

Marker	Initial measurement category <sup>a</sup>	n	Mean (SD) of change	
			$\Delta_{2-1}$	$\Delta_{3-2}$
IGF-I	<107	76	10.5 (26.9)	-0.8 (33.9)
	107-230	601	-3.7 (37.9)	-2.2 (42.6)
	>230	74	-65.3 (62.6)	-23.7 (43.6)
IGFBP-3	<3.0	77	0.31 (0.39)	0.02 (0.37)
	3.0-4.6	575	-0.07 (0.46)	0.05 (0.48)
	>4.6	86	-0.79 (0.82)	-0.11 (0.63)
ALS	<228	70	17.6 (35.8)	4.3 (32.4)
	228-376	598	-0.67 (43.9)	7.4 (44.1)
	>376	70	-46.7 (59.5)	1.4 (60.9)
PINP	<47	75	5.1 (19.1)	-1.6 (18.9)
	47-181	604	2.3 (17.6)	-2.6 (17.9)
	>181	85	-21.0 (41.4)	11.2 (50.3)
ICTP	<3.8	65	0.39 (1.00)	-0.04 (1.08)
	3.8-8.9	621	0.09 (0.92)	-0.16 (1.00)
	>8.9	84	0.05 (1.86)	-0.20 (1.84)
PIIINP	<3.5	72	0.38 (0.78)	-0.09 (0.92)
	3.5-7.4	602	0.00 (0.83)	-0.03 (0.95)
	>7.4	87	-1.05 (1.38)	0.15 (1.77)

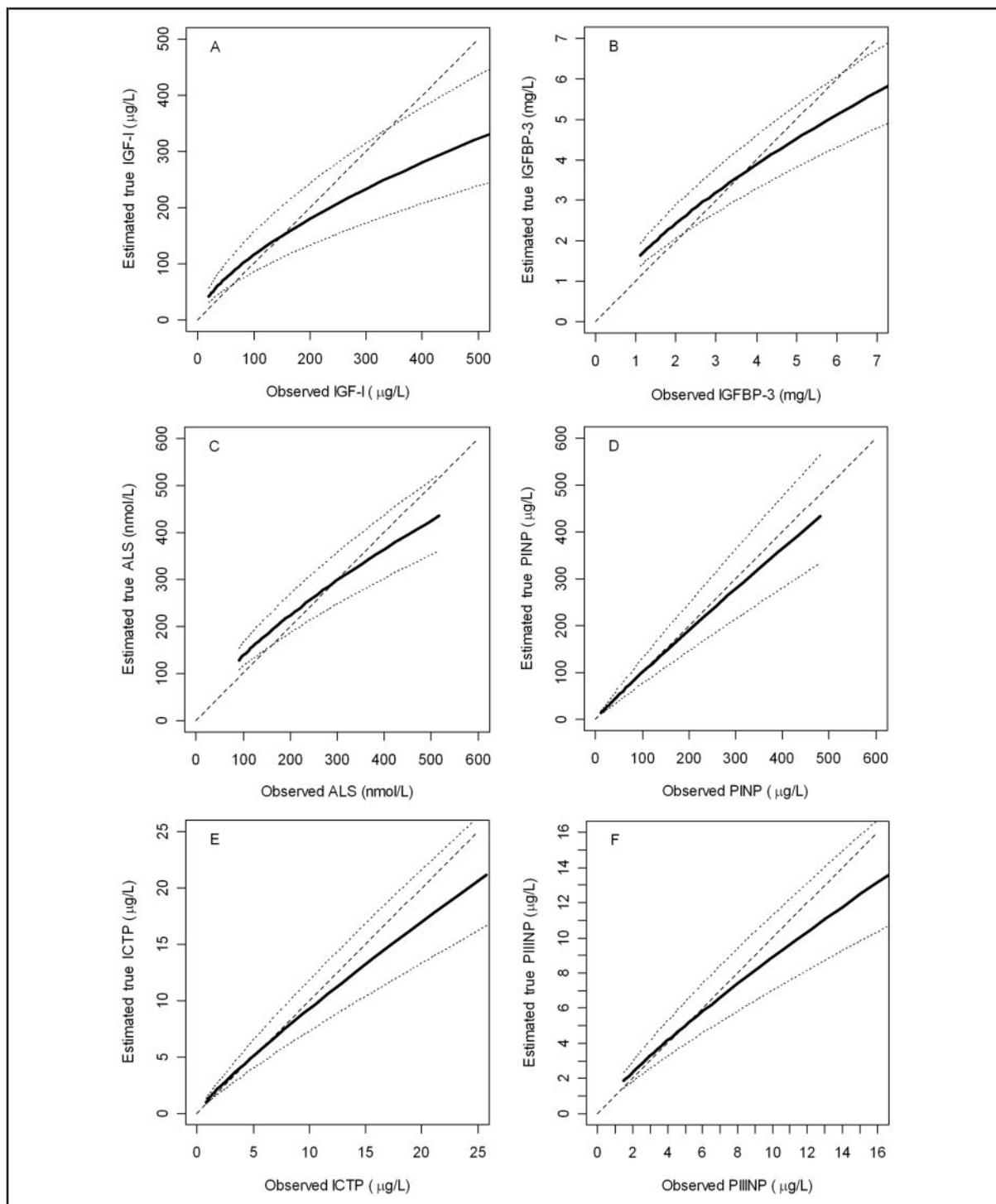
<sup>a</sup> The 3 categories were based on the 10th, 11th to 89th, and 90th percentile values of the first measurement.

an antidoping test (34). Concentrations of the markers IGF-I, ICTP, and PIIINP in 47 athletes measured over a 6-month period, entailing periods of training, competition, recovery, and rest, have been reported to be fairly stable; however, within-subject variability was not formally assessed (35).

Although the index of individuality has classically been interpreted in the context of assessing the utility of the population-based reference interval in relation to the interpretation of an individual's measurements (36), we consider that it can also be used in the reverse sense to judge how closely a single measurement is likely to estimate the mean (true or set point) value of an analyte in the individual. Thus, an index of individuality <0.6 indicates that the population-based reference range is of limited value for the interpretation of measurements in an individual, because test results that are abnormal for the individual can be undetected by the population-based reference range. In such a situation, however, a single measurement will be a much better estimate of the true value for the individual than the mean for the population at large. As the index of individuality increases >0.6, the probability that an abnormal result in an individual will fall outside the

population-based reference range progressively increases until at index values above 1.4 it increases to >95%. Obviously, for an analyte with high index of individuality (e.g., where the variability of measurements within an individual becomes just as large as or larger than the variability between individuals), the utility of a single measurement in an individual is more limited, and in this case the Bayesian approach presented here is useful in estimating the individual's true value because it combines the mean and variance of the individual and the population from which the individual was drawn.

The phenomenon of regression to the mean was observed in this study, in which individuals with extreme measured values at the first measurement on average tended to regress toward the mean in subsequent measurements, such that lower/higher measured values at measurement 1 tend to be higher/lower, respectively, at subsequent measurements in the absence of any biological effect. This phenomenon was largely attributable to the within-subject variability of each marker, which has important implications in clinical practice, because measurements of IGF axis and bone turnover markers can potentially be used to make a



**Fig. 1.** Plots of observed values (x axis) against estimated probable values (y axis) for each marker. IGF-I (A), IGFBP-3 (B), ALS (C), PINP (D), ICTP (E), and PIIINP (F). For each marker, the probable value estimated by the Bayesian model for any observed value is indicated by the solid line  $\pm$  95% CIs (dotted lines). The dashed line indicates the line of identity where the observed and probable values are equal, which would be the case where the marker has no within-subject variability.

diagnosis of GH deficiency or detect GH doping in sport.

The RTM phenomenon has practical importance for the application of a GH diagnostic test. First, a single measurement may result in false-positive or false-negative classification for an individual. Second, the substantial within-subject variability observed in the IGF axis markers suggests that any diagnosis or anti-doping test based on these markers should be based on multiple measurements rather than a single measurement. In practice, however, multiple measurements are not always possible, and the Bayesian approach described in this report can be a useful means to approximate the probable value of a marker for an individual.

The present results should be interpreted within the context of a number of strengths and weaknesses. The study was based on a large sample size with multiple measurements, which allows more accurate and reliable delineation of between- and within-subject variability than studies with small sample sizes. The study participants were all healthy, without any diseases that are likely to affect bone metabolism, and were drawn from all ethnicities. Because there were no significant differences in the estimates of within-subject variability between sexes, age groups, ethnicities, or sport types, the present results could be applied to all young, healthy adults, provided that they are subject to the same method of analysis. Nevertheless, the current study did not attempt to control for other variables that are known to affect IGF axis and collagen markers such as exercise and training (37, 38), food intake (39), and diurnal variation (3, 40). These factors may have contributed to greater within-subject variability in our study. In the clinical setting, controlling for such factors and thus reducing the variability, for example by measuring samples collected in the fasting state or at the same time of day, may be feasible, but in the doping situation would be impractical. Our estimates of within-subject variability using samples collected at random with regard to variables such as time of day and exercise are more likely to reflect the variability in the doping assessment context. Furthermore, in this study, participants were required to declare that they had not taken GH; however, there was no independent verification of the truthfulness of their statement. Assuming that a small minority of subjects may have taken growth hormone, we would expect their markers to be increased, but this would have a negligible effect on the reliability or intrasubject variability of these markers.

This study addresses the challenge of detecting abnormal values with certainty, posed by the presence of within-subject variability in the IGF axis and collagen markers. Even if an individual's marker value is slightly higher than a threshold of abnormality, it may not be possible to ascertain whether the individual's value is abnormal because of within-subject variability. However, the Bayesian approach presented in this report can be used to estimate the probability of abnormality for a given measured value. In essence, the Bayesian approach addresses the following question: "Given a measured marker value for an individual, what is the individual's probable value?" The RTM effect predicts that an individual's extreme value will regress toward the population average; therefore, the probable value can be estimated from the weighted average of population mean and the individual's value, with the weight being the coefficient of reliability. However, this probable value is also subject to sampling variation within the individual. Assuming that the distribution of this further variation is normal, it is possible to estimate the probability that a measured value is above a certain threshold of abnormality. The present study provides all essential parameters required for the estimation.

In conclusion, these results indicate that the within-subject variability in healthy individuals was greater for IGF-I than for the collagen markers. Where a single measurement is available, our findings demonstrate that it is possible to estimate the long-term probable value of each of the markers by applying the Bayesian approach. This modeling strategy not only enhances the reliability and reduces the cost of GH doping tests based on the use of these markers, but has diagnostic applicability beyond doping in sports.

**Grant/Funding Support:** This work was supported by the World Anti-Doping Agency and by the Australian Government through the Anti-Doping Research Program (ADRP) of the Department of Communications, Information Technology and the Arts.

**Financial Disclosures:** None declared.

**Acknowledgments:** We thank Drs. Kin-Chuen Leung and Graham Trout for their contributions to the project and Kevin Hardman, Sri Meka, and James Modzelewski for their technical contribution. We thank Nguyen D. Nguyen for his help in data analysis and graphical presentation and James Boyd for helpful editorial input in the finalization of the manuscript.

## References

1. Leung KC, Ho KK. Measurement of growth hormone, insulin-like growth factor I and their binding proteins: the clinical aspects. *Clin Chim Acta* 2001;313:119–23.
2. Strasburger CJ, Bidlingmaier M, Wu Z, Morrison KM. Normal values of insulin-like growth factor I and their clinical utility in adults. *Horm Res* 2001;55 Suppl 2:100–5.
3. Juul A. Serum levels of insulin-like growth factor I

- and its binding proteins in health and disease. *Growth Horm IGF Res* 2003;13:113–70.
4. Samani AA, Yakar S, Leroith D, Brodt P. The role of the IGF system in cancer growth and metastasis: overview and recent insights. *Endocr Rev* 2006.
  5. Meinbach DS, Lokeshwar BL. Insulin-like growth factors and their binding proteins in prostate cancer: cause or consequence? *Urol Oncol* 2006; 24:294–306.
  6. Renehan AG, Harvie M, Howell A. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and breast cancer risk: eight years on. *Endocr Relat Cancer* 2006;13:273–8.
  7. Seibel MJ. Molecular markers of bone turnover: biochemical, technical and analytical aspects. *Osteoporos Int* 2000;11 Suppl 6:S18–29.
  8. Jensen LT. The aminoterminal propeptide of type III procollagen. Studies on physiology and pathophysiology. *Dan Med Bull* 1997;44:70–8.
  9. Crofton PM, Wade JC, Taylor MR, Holland CV. Serum concentrations of carboxyl-terminal propeptide of type I procollagen, amino-terminal propeptide of type III procollagen, cross-linked carboxyl-terminal telopeptide of type I collagen, and their interrelationships in schoolchildren. *Clin Chem* 1997;43:1577–81.
  10. Zachariae H, Heickendorff L, Sogaard H. The value of amino-terminal propeptide of type III procollagen in routine screening for methotrexate-induced liver fibrosis: a 10-year follow-up. *Br J Dermatol* 2001;144:100–3.
  11. Sundstrom J, Vasan RS. Circulating biomarkers of extracellular matrix remodeling and risk of atherosclerotic events. *Curr Opin Lipidol* 2006;17: 45–53.
  12. Dall R, Longobardi S, Ehrnborg C, Keay N, Rosen T, Jorgensen JO, et al. The effect of four weeks of supraphysiological growth hormone administration on the insulin-like growth factor axis in women and men. GH-2000 Study Group. *J Clin Endocrinol Metab* 2000;85:4193–200.
  13. Longobardi S, Keay N, Ehrnborg C, Cittadini A, Rosen T, Dall R, et al. Growth hormone (GH) effects on bone and collagen turnover in healthy adults and its potential as a marker of GH abuse in sports: a double blind, placebo-controlled study. The GH-2000 Study Group. *J Clin Endocrinol Metab* 2000;85:1505–12.
  14. Kniess A, Ziegler E, Kratzsch J, Thieme D, Muller RK. Potential parameters for the detection of hGH doping. *Anal Bioanal Chem* 2003;376: 696–700.
  15. Sartorio A, Agosti F, Marazzi N, Maffiuletti NA, Cella SG, Rigamonti AE, et al. Combined evaluation of resting IGF-I, N-terminal propeptide of type III procollagen (PIIINP) and C-terminal cross-linked telopeptide of type I collagen (ICTP) levels might be useful for detecting inappropriate GH administration in athletes: a preliminary report. *Clin Endocrinol (Oxf)* 2004;61:487–93.
  16. Nelson AE, Ho KK. Abuse of growth hormone by athletes. *Nat Clin Pract Endocrinol Metab* 2007; 3:198–9.
  17. Milani D, Carmichael JD, Welkowitz J, Ferris S, Reitz RE, Danoff A, Kleinberg DL. Variability and reliability of single serum IGF-I measurements: impact on determining predictability of risk ratios in disease development. *J Clin Endocrinol Metab* 2004;89:2271–4.
  18. Garnero P, Mulleman D, Munoz F, Sornay-Rendu E, Delmas PD. Long-term variability of markers of bone turnover in postmenopausal women and implications for their clinical use: the OFELY study. *J Bone Miner Res* 2003;18:1789–94.
  19. Eastell R, Mallinak N, Weiss S, Ettinger M, Pettinger M, Cain D, et al. Biological variability of serum and urinary N-telopeptides of type I collagen in postmenopausal women. *J Bone Miner Res* 2000;15:594–8.
  20. Scario JK, Garry PJ, Montoya GD, Wilson JM, Baumgartner RN. Critical differences in the serial measurement of three biochemical markers of bone turnover in the sera of pre- and postmenopausal women. *Clin Biochem* 2001;34:639–44.
  21. Hannon R, Blumsohn A, Naylor K, Eastell R. Response of biochemical markers of bone turnover to hormone replacement therapy: impact of biological variability. *J Bone Miner Res* 1998;13: 1124–33.
  22. Sharpe K, Hopkins W, Emslie KR, Howe C, Trout GJ, Kazlauskas R, et al. Development of reference ranges in elite athletes for markers of altered erythropoiesis. *Haematologica* 2002;87:1248–57.
  23. Nelson AE, Howe CJ, Nguyen TV, Leung KC, Trout GJ, Seibel MJ, et al. Influence of demographic factors and sport type on growth hormone-responsive markers in elite athletes. *J Clin Endocrinol Metab* 2006;91:4424–32.
  24. Lomeo A, Bolner A. Stability of several biochemical markers of bone metabolism. *Clin Chem* 2000;46:1200–2.
  25. Ito Y, Nakachi K, Imai K, Hashimoto S, Watanabe Y, Inaba Y, et al. Stability of frozen serum levels of insulin-like growth factor-I, insulin-like growth factor-II, insulin-like growth factor binding protein-3, transforming growth factor beta, soluble Fas, and superoxide dismutase activity for the JACC study. *J Epidemiol* 2005;15 Suppl 1:S67–73.
  26. Baxter RC, Brown AS, Turtle JR. Radioimmunoassay for somatomedin C: comparison with radio-receptor assay in patients with growth-hormone disorders, hypothyroidism, and renal failure. *Clin Chem* 1982;28:488–95.
  27. Baxter RC, Martin JL. Radioimmunoassay of growth hormone-dependent insulinlike growth factor binding protein in human plasma. *J Clin Invest* 1986;78:1504–12.
  28. Baxter RC. Circulating levels and molecular distribution of the acid-labile (alpha) subunit of the high molecular weight insulin-like growth factor-binding protein complex. *J Clin Endocrinol Metab* 1990;70:1347–53.
  29. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. SAS System for Mixed Models. Cary, NC: SAS Institute Inc. 1996;633 p.
  30. Fleiss JL, Shrout PE. The effects of measurement errors on some multivariate procedures. *Am J Public Health* 1977;67:1188–91.
  31. Chapurlat RD, Blackwell T, Bauer DC, Cummings SR. Changes in biochemical markers of bone turnover in women treated with raloxifene: influence of regression to the mean. *Osteoporos Int* 2001; 12:1006–14.
  32. Nguyen TV, Pocock N, Eisman JA. Interpretation of bone mineral density measurement and its change. *J Clin Densitom* 2000;3:107–19.
  33. Alvarez L, Ricós C, Peris P, Guañabens N, Monnegal A, Pons F, Ballesta AM. Components of biological variation of biochemical markers of bone turnover in Paget's bone disease. *Bone* 2000;26:571–6.
  34. Di Luigi L, Guidetti L. IGF-I, IGFBP-2, and -3: do they have a role in detecting rhGH abuse in trained men? *Med Sci Sports Exerc* 2002;34: 1270–8.
  35. Sartorio A, Jubeau M, Agosti F, Marazzi N, Rigamonti A, Muller EE, Maffiuletti NA. A follow-up of GH-dependent biomarkers during a 6-month period of the sporting season of male and female athletes. *J Endocrinol Invest* 2006;29:237–43.
  36. Harris EK. Effects of intra- and interindividual variation on the appropriate use of normal ranges. *Clin Chem* 1974;20:1535–42.
  37. Wallace JD, Cuneo RC, Baxter R, Orskov H, Keay N, Pentecost C, et al. Responses of the growth hormone (GH) and insulin-like growth factor axis to exercise, GH administration, and GH withdrawal in trained adult males: a potential test for GH abuse in sport. *J Clin Endocrinol Metab* 1999; 84:3591–601.
  38. Wallace JD, Cuneo RC, Lundberg PA, Rosen T, Jorgensen JO, Longobardi S, et al. Responses of markers of bone and collagen turnover to exercise, growth hormone (GH) administration, and GH withdrawal in trained adult males. *J Clin Endocrinol Metab* 2000;85:124–33.
  39. Clowes JA, Hannon RA, Yap TS, Hoyle NR, Blumsohn A, Eastell R. Effect of feeding on bone turnover markers and its impact on biological variability of measurements. *Bone* 2002;30: 886–90.
  40. Wolthers OD, Heuck C, Heickendorff L. Diurnal variations in serum and urine markers of type I and type III collagen turnover in children. *Clin Chem* 2001;47:1721–2.