Application of the Athlete Biological Passport approach to the detection of growth hormone doping

Tristan Equey[1], Antoni Pastor[2], Rafael de la Torre Fornell[2], Andreas Thomas[3], Sylvain Giraud[4], Mario Thevis[3], Tiia Kuuranne[4], Norbert Baume[1], Osquel Barroso[1], Reid Aikin[1]

[1]World Anti-Doping Agency (WADA), Montreal, Canada

[2]Integrative Pharmacology and Systems Neurosciences Research Group, Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain; Spanish Biomedical Research Centre in Physiopathology of Obesity and Nutrition (CIBEROBN), Madrid, Spain; University Pompeu Fabra (CEXS-UPF) Barcelona, Spain.

[3]Institute of Biochemistry, German Sport University Cologne, Cologne, Germany

[4]Swiss Laboratory for Doping Analyses, University Center of Legal Medicine, Genève and Lausanne, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Epalinges, Switzerland

Address for Correspondence:

Reid Aikin, World Anti-Doping Agency

800 Rue du Square-Victoria Suite 1700

Montreal, Quebec H4Z 1B7, Canada

Email: reid.aikin@wada-ama.org

Financial Support: This work was supported by funds from the World Anti-Doping Agency.

Disclosure Statement: The authors have nothing to disclose.

Clinical Trial Registration: EudraCT number 2014-000563-41

ABSTRACT

Context: Because of its anabolic and lipolytic properties, growth hormone (GH) use is prohibited in sport. Two methods based on population derived decision limits are currently used to detect human GH (hGH) abuse: the hGH Biomarkers Test and the Isoforms Differential Immunoassay.

Objective: Test the hypothesis that longitudinal profiling of hGH biomarkers through application of the Athlete Biological Passport (ABP) has the potential to flag hGH abuse.

Design: IGF-1 and P-III-NP distributions were obtained from 7 years of anti-doping data and applied as priors to analyse individual profiles from an hGH administration study in recreational athletes.

Setting: Academic and anti-doping laboratories. Elite (n=11,455) and recreational athletes (n=35).

Intervention(s): An open-label, randomized, single site, placebo-controlled administration study was carried out with individuals randomly assigned to 4 arms: placebo, or 3 different doses of recombinant hGH.

Main Outcome Measure(s): Serum samples were analyzed for IGF-1, P-III-NP, and hGH isoforms and the performance of a longitudinal, ABP-based approach was evaluated.

Results: An ABP-based approach set at a 99% specificity level flagged 20/27 individuals receiving hGH treatment, including 17/27 individuals after cessation of the treatment. ABP sensitivity ranged from 12.5-71.4 % across the hGH concentrations tested following 7 days of treatment, peaking at 57.1-100 % after 21 days of treatment, and was maintained between 37.5-71.4 % for the low and high dose groups one week after cessation of treatment.

Conclusions: These findings demonstrate that longitudinal profiling of hGH biomarkers can provide suitable performance characteristics for use in anti-doping programs.

Keywords: growth hormone, anti-doping, biomarkers, athlete biological passport

INTRODUCTION

Growth hormone elicits anabolic and lipolytic properties and is therefore prohibited in sport by the World Anti-Doping Agency (WADA) (1,2). Two methods are currently used for the detection of hGH abuse in sport: 1) the Isoforms Differential Immunoassay based on the ratio of recombinant hGH to endogenous, pituitary hGH (3), and 2) the hGH Biomarkers Test, based on the measurement of two hGH-responsive biomarkers, namely insulin-like growth factor-I (IGF-1) and N-terminal pro-peptide of type III collagen (P-III-NP) (4). Since both approaches utilize population-based thresholds to uncover doping, it is hypothesized that the use of personalized thresholds through the application of the Athlete Biological Passport (ABP) approach may increase the sensitivity to detect hGH abuse.

The ABP is based on the application of adaptive, personalized thresholds to specific biomarkers of doping in order to flag profiles for closer examination. The calculation of such personalized thresholds, which correspond to a critical range defined by a given specificity (ex. 99%) assuming a normal physiological condition, requires an understanding of the population distribution and sources of variation for each biomarker (5,6). In contrast to population-based decision limits for endogenous threshold substances, which are typically set at 99.99% specificity in anti-doping, the ABP uses a lower initial specificity (i.e. 99%) for sensitive flagging of atypical passports for closer examination and to drive anti-doping strategies such as the collection of additional samples, the further analysis of existing samples, carrying out investigations, or placing samples into long-term storage for future analysis. When used to directly sanction an athlete, increased specificity is then brought through a rigorous passport review process (7,8). The ABP is presently applied to biomarkers of blood doping measured in blood samples and to markers of steroid doping measured in urine samples.

Administration studies have established that both IGF-1 and P-III-NP respond in a dose-dependent manner to hGH treatment (9–12), and a discriminant function was developed utilizing both markers (via the GH-2000 score) which improved the sensitivity and specificity of the detection of hGH administration compared to either marker alone (13).The presently employed hGH Biomarkers Test is based on sex-specific population thresholds, at a specificity of 99.99%, for the GH-2000 score measured in a single sample. Previous studies have suggested that longitudinal profiling may improve the ability to detect GH use (14). Interestingly, these studies indicated a significant inter-subject variance for IGF-1 and P-III-NP, suggesting that the use of personalized thresholds through the ABP approach, which removes much of the inter-subject variance, could significantly improve the sensitivity of the detection of GH use (15–17).

The goal of the present study was to develop and validate an ABP-based, longitudinal approach for the use of IGF-1, P-III-NP and the GH-2000 score for the detection of GH use. First, data from authentic anti-doping samples collected over a seven-year period were used to determine the distribution of IGF-1, P-III-NP and GH-2000 scores and to estimate intra- and inter-subject variation in an elite athlete population. These results were then used to develop an adaptive model for the longitudinal monitoring of IGF-1, P-III-NP and GH-2000 score and the performance of this approach was then tested on samples collected during an hGH administration study.

MATERIAL AND METHODS

The first data set of IGF-1 and P-III-NP concentrations is based on values measured by 19 WADA-accredited laboratories between October 2012 and July 2019 where serum samples were collected in accordance with the World Anti-Doping Code, the WADA International Standard for the Protection of Privacy and Personal Information (ISPPPI), and prevailing WADA Technical Documents and Guidelines. In order to estimate priors reflecting a normal physiological condition, all samples from athletes with at least one adverse analytical finding (AAF) reported for a prohibited substance included in the WADA Prohibited List were excluded from the dataset, regardless of the substance or sample matrix, as well as a small number of data entry errors (a total of 1608 samples from 953 individuals were removed). The raw IGF-1 and P-III-NP concentrations, the sample sequence order, the laboratory name, the method used, and the athlete age, gender and sport were compiled into an anonymized dataset. The final dataset includes 15,975 samples collected from 11,455 athletes. See Supplementary Table 1 for a summary of relevant descriptive statistics (18).

Details on the procedure used for the collection, transport, and analysis of the serum samples are available in dedicated WADA Guidelines (4,18). Briefly, serum samples were collected (BD Vacutainer® SSTTM-II Plus tubes; BD Vacutainer® SSTTM-II Plus Advance tubes) and transported to the analyzing laboratory under refrigerated conditions. IGF-1 was quantified by either a bottom-up liquid chromatography-tandem mass spectrometry (LC-MS/MS) method (19), a immunoradiometric assay available from Beckman Coulter Immunotech (Cat# A15729, RRID:AB_2893421, Marseille, France) or a chemiluminescent immunoassay from Immunodiagnostics Systems Limited (IDS, Cat# IS-3900, RRID:AB_2861357, Boldon, UK). The quantification of P-III-NP was performed using a two-site sandwich, chemiluminescent immunoassay on a Siemens ADVIA Centaur platform (Cat# 10492440, RRID:AB_2893415, Siemens Healthcare Laboratory Diagnostics, Camberley, UK) (20), or the competitive radioimmunoassay from Orion Diagnostica (Cat# 68570, RRID:AB_2893420; now Aidian; Espoo, Finland).

The second dataset comes from an open-label, randomized, single site, placebo-controlled administration study with recombinant hGH (Nutropin AQ) in healthy volunteers performed at the Clinical Trials Research Unit (CTRU) of the IMIM (Hospital del Mar Medical Research Institute, Barcelona, Spain). The study (IMIMFTCL/GH4) was approved by the local ethics committee (CEIm-PSMAR) and the Spanish Agency of Medicines and Medical Devices (AEMPS) and a written informed consent was obtained from all participating subjects. The study was registered in the European Union Drug Regulating Authorities Clinical Trials Database (EudraCT number: 2014-000563-41). Briefly, 35 healthy amateur athletes (25 males and 10 females; average age 31.5) performing at least 5 hours per week of moderate to intense physical activity were randomly assigned to one of 4 arms: placebo (6 males, 2 females), Very Low Dose (VL, 0.016 mg/kg; 7 males, 3 females), Low Dose (L, 0.033 mg/kg; 7 males, 3 females), and High Dose (H, 0.066 mg/kg; 5 males, 2 females). The first day of hGH administration was performed in the CTRU and subjects were trained to administer hGH by themselves (auto-administration) daily for the duration of the 3-week treatment period. Subjects were scheduled to be tested 14 times over 3 months. A total of three sample collections were missed by three subjects. Serum samples were collected in accordance to WADA Guidelines (4) and analyzed for IGF-1 by LC-MS/MS and P-III-NP using the Siemens ADVIA Centaur assay. After application of WADA criteria for the measurement of IGF-1, where the absolute difference between measurements made by LC-MS/MS of the T1 and T2 fragments of IGF-1 should not differ by more than 20%, the dataset was reduced to 393 samples, with an average of 11.2 tests per subjects (see Supplementary Table 2 for the relevant descriptive statistics) (18). Serum samples were also analyzed using “Kit 1” (RRID:AB_2893416, CMZAssay GmbH, Germany) of the Isoforms Differential Immunoassay in accordance to the applicable WADA Technical Document (3).

All the statistical analyses have been performed with the R software version 3.6. A significance level of $p < 0.05$ was considered for all hypothesis tests. The Athlete Biological Passport simulations were carried out using Matlab version 9.6 with the Statistics and Machine Learning Toolbox. As established for other modules of the ABP, a standard Bayesian network model is used to 1) detect abnormal samples and 2) detect abnormal sequences of growth hormone (GH) biomarkers in longitudinal data (5,6). In such a model, the latest test result is considered as atypical if its value falls outside the critical range defined by the set specificity $(1-\alpha)\%$, where α is the set acceptable proportion of false positives. Similarly, a sequence is abnormal if it displays an abnormally high variance (6). As in Sottas et al. (2007), the estimated intra- and inter-subject coefficient of variation of the specific biomarker (for a determined assay) was used in addition to its population mean prior to establishing the joint prior distribution (5). Here we choose to model $p(\mu,\sigma) = p(\mu) \cdot p(CV) \cdot \mu$, where $p(CV)\ is$ the intra-subject coefficient of variation probability distribution. No correlation was found between μ and CV for pairs with 6 samples or more (R=-0.16, p=0.32 [IGF-1], R=-0.02, p=0.88 [P-III-NP] and -0.21, p=0.18 [GH-2000], N=40), suggesting that the CV is indeed independent of the mean while a correlation was found between μ and σ, with the exception of GH-2000 score (R=0.56, p<0.01 [IGF-1], R=0.53, p<0.01 [P-III-NP] and R=0.002, p=0.89 [GH-2000], N=40).

RESULTS

Estimation of population mean priors

Figure 1 represents the proposed Bayesian network for the application of the ABP approach to biomarkers of hGH abuse. To calibrate the model, a dataset containing 15,975 serum samples collected over a 7-year period from 11,455 elite athletes was used. The samples were collected from athletes with an average age of 26 years [95% range: 18-37], predominantly male (75.2%), from 132 different nationalities, mainly collected out-of-competition (83.5%), across 78 different sports (21.4% from endurance sports, see Supplementary Tables 1 and 3) (18). Most samples were from athletes tested only once (56.1%), but 931 athletes were tested 3 times or more. A sub-dataset including only samples analyzed by LC-MS/MS (IGF-1) and the Siemens ADVIA Centaur (P-III-NP) was also created (Supplementary Table 1b and 4) (18), as these methods represent a potentially useful assay pairing for routine implementation for the ABP.

Using the elite athlete dataset, median biomarker reference values were determined as a function of age by applying an additive quantile regression model (21). Figure 2 shows biomarker values as a function of athlete age for IGF-1 (LC-MS/MS), P-III-NP (Centaur) and GH-2000 score for each gender as well as the fitted percentile. Supplementary Tables 5-7 (18) report the age reference median value (with standard error; SE) between 15 and 40 years old for both biomarkers, GH-2000 score and genders, and are consistent with other published studies (20,22–24)'. As observed previously in males (26), we also observed a small but significant relationship between age and the GH-2000 score for the pairing involving IGF-1 measurement by LC-MS/MS combined with P-III-NP measured by the Centaur assay in males and also in females. While a correction has been recently applied to the GH-2000 score in males, which is generally suitable for all assay pairings (4,26), for the purposes of the ABP where only one assay pairing will be used, it was preferable to model the age relationship specifically for the LC-MS/MS (IGF-1)-Centaur (P-III-NP) assay pairing, according to Supplementary Table 7 (18).

Estimation of variance components

Using the elite athlete dataset stratified by gender, an estimation of the intra- and inter- subject variance was first performed using a linear mixed effect model (*lme* R package). Due to the skewness of their distribution, IGF-1 and P-III-NP were log-transformed before the estimation. The estimated model includes age, assay, and laboratory as fixed effects and a subject-specific random effect. The resulting estimated variance-covariance structure allows the computation of inter- and intra- subject variance for each specific assay at the exception of GH-2000 score combinations, where considering all possible GH-2000 score combination leads to an over-specified covariance structure. GH-2000 score variance was therefore estimated on the sub-dataset consisting of samples analyzed with LC-MS/MS (IGF-1) and Centaur (P-III-NP) only. The estimated inter-subject variance and intra-subject variance for all the assays are summarized in Table 1. As a robustness test, an expectation-maximization (EM) algorithm for mixtures of normal distributions was run on the empirical distribution of athlete intra-subject coefficients of variation to validate the estimated intra-subject priors (Supplementary Table 8) (18). The results were slightly lower but close to estimates from the mixed model approach. As a higher intra-subject CV results in a more conservative ABP approach, the mixed model estimates were chosen as priors.

Treatment effects

In order to test the performance of an ABP approach for the detection of hGH abuse, serum samples originating from an open-label, randomized, placebo-controlled administration trial with recombinant hGH in healthy recreational athletes (see Figure 3A for study design) were analyzed for IGF-1 by LC-MS/MS and P-III-NP by Siemens ADVIA Centaur. Three doses were included in the study design, where the high (H, 0.066 mg/kg) and low doses (L, 0.033 mg/kg) correspond to those used in the previous GH-2000 studies (9,10), and the very low dose group (VL, 0.016 mg/kg) was chosen to be slightly below the range used in other previous studies (11,26).

Consistent with previous studies (12,27), IGF-1 and P-III-NP demonstrated a dose-dependent response to hGH treatment with IGF-1 levels increasing more rapidly than P-III-NP but with the increased P-III-NP levels persisting longer than IGF-1 following cessation of hGH treatment (Supplementary Figure 1) (18). In males, GH-2000 score levels also showed a dose-dependent increase, where all three hGH doses resulted in a significant increase in GH-2000 score after 7 days of treatment (Figure 3B). The same pattern is observed for female athletes, however the statistical power for such group level analyses in both males and females is limited due to the relatively small sample size.

During the treatment period, the average GH-2000 score for males for the very low (VL), low (L) and high (H) dose group athletes were 8.21 (SD ±1.78), 10.32 (SD ±2.52) and 10.20 (SD ±2.48), respectively, compared to 6.69 (SD ±0.94) for control subjects. When considering the treatment days and the wash-out period, the averages from the VL group were never statistically different from the control group except at day 21 and 63 (p-value = 0.026 and 0.004). The group averages for days 7 to 28 were statistically different from the control group for both L and H dose group. The treatment effect on the GH-2000 score was never statistically different between L and H dose group. The high heterogeneity in the response to the treatment might explain the lack of statistical difference between the two groups (Supplementary Figure 1) (18).

Because the passport approach is able to flag abnormal increases in intra-subject variances, the effects of hGH treatment on intra-subject variance was also examined. A dose-dependent increase in intra-subject variance was observed (Supplementary Figure 2) (18), supporting the applicability of the passport approach to improve the detection of abnormal variations in biomarkers as a response to hGH abuse.

ABP Performance

The performance of a calibrated adaptive model was then assessed on the longitudinal biomarker profiles from individuals treated with recombinant hGH. In order to detect outliers, the specificity of the adaptive model was set at 99% and a universal intra-subject CV is assumed to avoid a strong contraction of the critical range for individuals with very low variation between samples.

An example of the model's performance on a profile from a 44-year-old male from the "very low dose" group is shown in Figure 4. The first sample is evaluated according to population-based priors and with each ensuing baseline sample the thresholds progressively narrow as the model adapts to the athlete's normal biomarker values. After 7 days of hGH treatment, increased IGF-1 and GH-2000 score was observed, with IGF-1 exceeding the upper threshold during the treatment period on days 7; 7.5 and 14. In this example, IGF-1 and GH-2000 score levels in all samples taken during the treatment period exceed the calculated baseline thresholds determined at day -1 (last day before treatment), and IGF-1 continues to be out of this critical range on the day after the cessation of treatment (day 22).

In order to assess the sensitivity of the adaptive model at different time points during and following hGH treatment, each "treatment" or "wash-out" sample was examined separately using all baseline samples from the same individual as prior information. Figure 5 illustrates the sensitivity across groups for each sample during treatment and wash-out periods. As expected, IGF-1 flags outliers quickly after the start of hGH treatment, where after 7 days 50%, 42.9%, and 85.7% of treated samples were flagged for the VL, L, and H doses of hGH, respectively. Three days after the cessation of the treatment (day 24), the IGF-1-based sensitivity was 0% for the VL group, 25% for the L group, and 83.3% for the H group.

The P-III-NP marker was slower to respond to the start of hGH treatment, with the sensitivity ranging between 0-57.1% across hGH doses after 7 days of treatment. However, the P-III-NP signal lasted longer following cessation of hGH treatment; thus, the sensitivity for the detection of individuals receiving high hGH dose is still at 57.1% two weeks after the cessation of treatment (day 35).

The GH-2000 score sensitivity ranged from 12.5-71.4% at day 7, peaking at 57.1-100% on the last day of treatment, and was maintained between 37.5% and 71.4% for the L and H doses one week after the cessation of treatment.

By comparison, the sensitivity of the hGH Biomarkers Test, based on population thresholds for GH-2000 applied to a single sample, is lower than the passport approach at all time points, which is in line with the difference in targeted specificity of both approaches (99% for ABP and 99.99% for Biomarker Test and Isoforms Differential Immunoassay).

In contrast, the Isoforms Differential Immunoassay had sensitivity range of 57.1-100% on the first day of treatment, only hours after hGH administration, and maintained a sensitivity in the range of 42.9-100% across all doses and time points during treatment. The sensitivity of the Isoforms Differential Immunoassay rapidly decreases after cessation of treatment, with 0-60% of samples being flagged on the day after the end of the treatment period and further decreasing to 0% for the remainder of the wash-out period.

At the passport level, 20 of the 27 treated athletes were flagged for the GH-2000 score at least once during the treatment period and 17/27 during the wash-out period. With the GH-2000 ABP approach, 13/27 of the treated athletes were still flagged beyond day 22. Table 2 summarizes the sensitivity at the individual level.

The ability of the calibrated adaptive model to detect abnormal sequences of GH biomarkers in longitudinal data was evaluated using a targeted specificity rate of 99.9%, consistent with other ABP modules. A dose-dependent increase in sensitivity was observed for the sequence-based approach for all three markers (Table 3) with a maximal sensitivity of 86% for the H group (6/7). For the VL dose group, IGF-1 showed the highest sensitivity at 30%, while GH-2000 score had a sensitivity of 10% and no sequence abnormalities were observed for P-III-NP.

Finally, in order to assess the specificity of the ABP approach, profiles were examined for outliers in untreated samples. From the 168 valid baseline and placebo treated samples from all 35 athletes, none were flagged as outliers by the adaptive model for P-III-NP and GH-2000 score at a theoretical specificity of 99% (Table 4). Three samples belonging to two individuals were flagged as outliers for IGF-I (specificity of 98.2%). With regards to the specificity of the model to detect abnormal sequences of biomarkers, none of the placebo group profiles were flagged for an abnormal sequence.

DISCUSSION

The present work describes an adaptive model for the detection of hGH doping in the context of the ABP. This model is calibrated based on population-derived priors estimated from elite athlete samples that can be assumed to capture variations related to factors such as ethnicity, age, training/competition, injury, and inter-laboratory analysis. Although a direct comparison with published population data is confounded by factors such as differences in the population studied, the assays used, and the duration of the study, the intra- and inter-subject coefficient of variation for IGF-1 is in line with the current literature, whereas a larger intra-subject coefficient of variation for P-III-NP and the GH-2000 score were estimated (15–17,28,29). The real anti-doping nature of the dataset, with potentially a non-zero prevalence of injured and doped athletes might explain this result. Given the sources of variation included in the present estimates, coupled with the theoretically improved specificity when using an elevated intra-subject CV, the present model arguably provides more conservative results that are in favor of the athlete. With time, these model parameters may be further refined in light of more harmonized pre-analytical and analytical conditions, which would be expected to reduce analytical uncertainty and further improve the sensitivity of such an ABP approach.

When considering marker performance characteristics, the GH-2000 score provided the best balance of sensitivity and specificity, suggesting it would be an ideal primary biomarker for the ABP that would trigger additional actions on the part of the anti-doping organizations. This finding is not unexpected as the GH-2000 score is based on two orthogonal markers of hGH abuse, linked to different biological pathways not likely to be affected by the same confounding factors. On the other hand, IGF-1 and P-III-NP would arguably be valuable as secondary markers, which could support an atypical passport finding based on the GH-2000 score but would likely not be sufficient to advance a passport case on their own merits. Indeed, when advancing an ABP-based sanction, the profile is reviewed by experts who must weigh the likelihood that the profile is the result of doping against the likelihood that it could be due to any other cause, such as normal variation, injury, disease or analytical issues. The weight of evidence in favor of doping is increased when multiple markers, across multiple samples, all point towards a specific scenario of doping. Thus, a response to an outlier for the GH-2000 score may be to collect additional samples in order to follow the expected decrease in IGF-1 followed by P-III-NP over time.

When applied to the clinical trial dataset, the specificity of the ABP approach for the GH-2000 score and P-III-NP performed in the expected range; however, we did note a slightly lower specificity for IGF-1 than anticipated. Importantly, it is noteworthy that none of these samples flagged for atypical IGF-1 values presented outliers for P-III-NP or the GH-2000 score, and in the absence of additional information from other samples would not provide sufficient evidence of doping to outweigh other possible explanations. Nevertheless, as IGF-1 responds to the beginning and during GH administration, an outlier for IGF-1 may still trigger further analysis of the same sample by the GH Isoforms Differential Immunoassay and/or the collection of an additional sample to examine a potential increase of P-III-NP levels. Additionally, other performance enhancing substances, such as Growth Hormone Releasing Factors, might also be the source of abnormally high IGF-1 levels in serum. Considering that WADA-accredited Laboratories have the analytical capacity to detect these compounds, such additional analyses could also be requested based on passport interpretation.

In order to mimic current practices where samples may be collected before or after exercise, the present clinical trial included 3 samples taken 2 hours after exercise. In all baseline or placebo control samples, exercise did not generate any outliers, confirming previous findings that any potential effects of exercise on IGF-1 or P-III-NP levels subside within 30 minutes following cessation of intense exercise (30–32).

When comparing the ABP approach with currently used population-based thresholds used to establish adverse analytical findings, it is important to acknowledge the difference in the specificity applied for each approach. As a result, a GH-2000-based ABP approach has better sensitivity during the post-treatment phase. Even in situations where such passport evidence would not be sufficient to directly sanction an athlete, the endocrine passport data can also be integrated with data from other sources in order to improve the planning of future tests. Interestingly, the sensitivity of the Isoforms Differential Immunoassay during the treatment period (42.9-100%) suggests that a strategy of performing the Isoforms Differential Immunoassay on relevant atypical samples flagged in the passport may be a viable approach to uncover adverse analytical findings related to hGH abuse. Future studies examining the potential benefits of longitudinal profiling of the Isoforms Differential Immunoassay may also improve the ability to flag hGH use.

When considering the analytical approaches for the ABP, mass spectrometry-based detection methods offer several advantages including the ability to multiplex, improved inter-laboratory reproducibility, and increased stability of the method over time because of the lack of reliance on batches of affinity-based reagents (e.g. inter-batch variability of antibodies or changes of assay platform by manufacturers). Within the past few years, several methods were published to measure either the trypsin digested (bottom-up) or the intact (top-down) IGF-1 protein. While the bottom-up approach was developed and validated first and is applied in routine in some WADA accredited laboratories (19,33–35), the top-down methodology, avoiding the digestion step during the sample preparation, has also been recently validated through an inter-laboratory assessment (36), and offers the potential for a more rapid and cost effective analysis.

Taken together, these findings support the implementation of a module of the ABP aimed at detecting hGH use based on longitudinal profiling of IGF-1, P-III-NP, and the GH-2000 score. Additional markers uncovered through biomarker discovery efforts and additional control of confounding factors can then be layered into this module over time, to progressively improve the performance characteristics of this module.

ACKNOWLEDGEMENTS

The authors would like to thank Pierre-Edouard Sottas for reviewing the manuscript.

DATA AVAILABILITY

The data that support the findings of this study are subject to contractual and/or privacy restrictions. A redacted/anonymized version of the data may be available from the corresponding author upon reasonable request and subject to confidentiality commitments.

REFERENCES

1. World Anti-Doping Agency. *World Anti-Doping Code*.; 2018. https://www.wada-ama.org/en/resources/the-code/world-anti-doping-code

2. Holt RIG, Ho KKY. The Use and Abuse of Growth Hormone in Sports. *Endocr Rev*. 2019;40(4):1163-1185.

3. TD2019GH. World Anti-Doping Agency. Published 2019. Accessed December 9, 2020. https://www.wada-ama.org/en/resources/science-medicine/td2019gh

4. Guidelines - Human Growth Hormone (hGH) Biomarkers Test. World Anti-Doping Agency. Accessed May 3, 2021. https://www.wada-ama.org/en/resources/laboratories/guidelines-human-growth-hormone-hgh-biomarkers-test

5. Sottas P-E, Baume N, Saudan C, Schweizer C, Kamber M, Saugy M. Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. *Biostatistics*. 2007;8(2):285-296.

6. Sottas P-E, Robinson N. A forensic approach to the interpretation of blood doping markers. *Law, Probability and Risk*. 2008;7:191-210.

7. Schumacher YO, D'Onofrio G. Scientific expertise and the Athlete Biological Passport: 3 years of experience. *Clinical chemistry*. 2012;58(6):979-985.

8. World Anti-Doping Agency. *Athlete Biological Passport Operating Guidelines*.; 2018. https://www.wada-ama.org/sites/default/files/resources/files/guidelines_abp_v6_2017_jan_en_final.pdf

9. Dall R, Longobardi S, Ehrnborg C, et al. The Effect of Four Weeks of Supraphysiological Growth Hormone Administration on the Insulin-Like Growth Factor Axis in Women and Men. *J Clin Endocrinol Metab*. 2000;85(11):4193-4200.

10. Longobardi S, Keay N, Ehrnborg C, 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. *J Clin Endocrinol Metab*. 2000;85(4):1505-1512.

11. Kniess A, Ziegler E, Kratzsch J, Thieme D, Müller RK. Potential parameters for the detection of hGH doping. *Anal Bioanal Chem*. 2003;376(5):696-700.

12. Holt RIG, Erotokritou-Mulligan I, McHugh C, et al. The GH-2004 project: the response of IGF1 and type III pro-collagen to the administration of exogenous GH in non-Caucasian amateur athletes. *European journal of endocrinology*. 2010;163(1):45-54.

13. Erotokritou-Mulligan I, Bassett EE, Kniess A, Sönksen PH, Holt RIG. Validation of the growth hormone (GH)-dependent marker method of detecting GH abuse in sport through the use of independent data sets. *Growth Hormone & IGF Research*. 2007;17(5):416-423.

14. Lehtihet M, Bhuiyan H, Dalby A, Ericsson M, Ekström L. Longitudinally monitoring of P-III-NP, IGF-I, and GH-2000 score increases the probability of detecting two weeks' administration of low-

dose recombinant growth hormone compared to GH-2000 decision limit and GH isoform test and micro RNA markers. *Drug Test Anal*. 2019;11(3):411-421.

15. Nguyen TV, Nelson AE, Howe CJ, et al. Within-Subject Variability and Analytic Imprecision of Insulinlike Growth Factor Axis and Collagen Markers: Implications for Clinical Diagnosis and Doping Tests. *Clin Chem*. 2008;54(8):1268-1276.

16. Erotokritou-Mulligan I, Bassett EE, Cowan DA, et al. The use of growth hormone (GH)-dependent markers in the detection of GH abuse in sport: Physiological intra-individual variation of IGF-I, type 3 pro-collagen (P-III-P) and the GH-2000 detection score. *Clinical Endocrinology*. 2010;72(4):520-526.

17. Kniess A, Ziegler E, Thieme D, Müller RK. Intra-individual variation of GH-dependent markers in athletes: Comparison of population based and individual thresholds for detection of GH abuse in sports. *Journal of Pharmaceutical and Biomedical Analysis*. 2013;84:201-208.

18. Equey T, Pastor A, Torre R de la, et al. Supplementary Materials. Figshare. Deposited 14 October 2021. https://doi.org/10.6084/m9.figshare.16814851.v2

19. Guidelines - Blood Sample Collection. World Anti-Doping Agency. Published July 22, 2014. Accessed August 13, 2020. https://www.wada-ama.org/en/resources/world-anti-doping-program/guidelines-blood-sample-collection

20. Cox HD, Lopes F, Woldemariam GA, et al. Interlaboratory Agreement of Insulin-like Growth Factor 1 Concentrations Measured by Mass Spectrometry. *Clinical Chemistry*. 2014;60(3):541-548.

21. Knudsen CS, Heickendorff L, Nexo E. Measurement of amino terminal propeptide of type III procollagen (PIIINP) employing the ADVIA Centaur platform. Validation, reference interval and comparison to UniQ RIA. *Clinical Chemistry and Laboratory Medicine (CCLM)*. 2014;52(2):237-241.

22. Fasiolo M, Wood SN, Zaffran M, Nedellec R, Goude Y. Fast calibrated additive quantile regression. *Journal of the American Statistical Association*. Published online March 11, 2020:1-11.

23. Bidlingmaier M, Friedrich N, Emeny RT, et al. Reference intervals for insulin-like growth factor-1 (igf-i) from birth to senescence: results from a multicenter study using a new automated chemiluminescence IGF-I immunoassay conforming to recent international recommendations. *J Clin Endocrinol Metab*. 2014;99(5):1712-1721.

24. Healy M-L, Dall R, Gibney J, et al. Toward the Development of a Test for Growth Hormone (GH) Abuse: A Study of Extreme Physiological Ranges of GH-Dependent Markers in 813 Elite Athletes in the Postcompetition Setting. *The Journal of Clinical Endocrinology & Metabolism*. 2005;90(2):641-649.

25. Nelson AE, Howe CJ, Nguyen TV, et al. Influence of Demographic Factors and Sport Type on Growth Hormone-Responsive Markers in Elite Athletes. *The Journal of Clinical Endocrinology & Metabolism*. 2006;91(11):4424-4432.

26. Böhning D, Böhning W, Guha N, et al. A correction to the age-adjustment of the GH-2000 score used in the detection of growth hormone misuse. *BMC Research Notes*. 2018;11(1):650.

27. Hermansen K, Bengtsen M, Kjær M, Vestergaard P, Jørgensen JOL. Impact of GH administration on athletic performance in healthy young adults: A systematic review and meta-analysis of placebo-controlled trials. *Growth Hormone & IGF Research*. 2017;34:38-44.

28. Powrie JK, Bassett EE, Rosen T, et al. Detection of growth hormone abuse in sport. *Growth Hormone & IGF Research*. 2007;17(3):220-226.

29. Ankrah-Tetteh T, Wijeratne S, Swaminathan R. Intraindividual variation in serum thyroid hormones, parathyroid hormone and insulin-like growth factor-1. *Ann Clin Biochem*. 2008;45(Pt 2):167-169.

30. Abellan R, Ventura R, Pichini S, et al. Effect of Physical Fitness and Endurance Exercise on Indirect Biomarkers of Recombinant Erythropoietin Misuse. *Int J Sports Med*. 2007;28(1):9-15.

31. Ehrnborg C, Lange KHW, Dall R, et al. The Growth Hormone/Insulin-Like Growth Factor-I Axis Hormones and Bone Markers in Elite Athletes in Response to a Maximum Exercise Test. *The Journal of Clinical Endocrinology & Metabolism*. 2003;88(1):394-401.

32. Wallace JD, Cuneo RC, Lundberg PA. Responses of Markers of Bone and Collagen Turnover to Exercise, Growth Hormone (GH) Administration, and GH Withdrawal in Trained Adult Males. 2000;85(1):10.

33. Wallace JD, Cuneo RC, Baxter R, 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. 1999;84(10):11.

34. Bredehöft M, Schänzer W, Thevis M. Quantification of human insulin-like growth factor-1 and qualitative detection of its analogues in plasma using liquid chromatography/electrospray ionisation tandem mass spectrometry. *Rapid Commun Mass Spectrom*. 2008;22(4):477-485.

35. Lopes F, Cowan DA, Thevis M, Thomas A, Parkin MC. Quantification of intact human insulin-like growth factor-I in serum by nano-ultrahigh-performance liquid chromatography/tandem mass spectrometry: Quantification of insulin-like growth factor-I by nanoUHPLC-MS/MS. *Rapid Commun Mass Spectrom*. 2014;28(13):1426-1432.

36. Kam RKT, Ho CS, Chan MHM. Serum Insulin-like Growth Factor I Quantitation by Mass Spectrometry: Insights for Protein Quantitation with this Technology. *EJIFCC*. 2016;27(4):318-330.

37. Moncrieffe D, Cox HD, Carletta S, et al. Inter-Laboratory Agreement of Insulin-like Growth Factor 1 Concentrations Measured Intact by Mass Spectrometry. *Clinical Chemistry*. 2020;66(4):579-586.

FIGURE LEGENDS

Figure 1. Bayesian network (BN) for the ABP endocrine module. Each node represents a variable and each edge that connects the nodes represents a causal relationship. The solid rectangles represent heterogenous factors controlled for by assessing their impact on the biomarkers of interest (mean and/or coefficient of variation). The dashed rectangle is a dummy variable with two possible states: doped and non-doped. The first line of circles is the mean and coefficient of variation of a longitudinal sequence of a set of endocrine biomarkers. The bottom circle is the set of endocrine biomarker variables. As in Sottas et al. (6), the BN is implemented as a hierarchical model with two levels and returns the probability of doping for an individual athlete.

Figure 2. Individual sample values and fitted percentiles for IGF-1 measured by LC-MS/MS **(A-B)**, P-III-NP measured by Siemens ADVIA Centaur (**C-D)** and the corresponding GH-2000 scores **(E-F)** for male (1,584 samples) and female (1,162 samples) athletes between 15 and 40 years old. The solid red line represents the median, the dashed blue line the 25th and 75th percentile and the black dotted line the 2.5th and 97.5th percentiles.

Figure 3. (A) Study design and timing of sample collection. Serum samples were collected during the three phases of the protocol either in the morning (light grey droplet) or in the afternoon (black droplets). Serum samples were withdrawn either before or after training sessions (🏃) and hGH injection (💉). The droplet is on the left side of the symbols when serum samples were collected before the training/injection, while sample collection after training/injection is depicted with the droplet on the right side. **(B)** Boxplot of GH-2000 score distribution by day for male athletes for each group. The black dashed line represents the applicable GH-2000 population-based decision limits (IGF-1 measured by LC-MS/MS and P-III-NP measured by Siemens ADVIA Centaur).

Figure 4. Passport of a 44 year old male recreational athlete treated with very low dose hGH and analyzed for IGF-1 by LC-MS/MS and P-III-NP by Siemens ADVIA Centaur. IGF-1 generates outliers on 5 occasions (days 7, 7.5, 14, 21 and 22) **(A)** and three outliers were observed for GH-2000 score on days 7.5, 14 and 21 **(C)**. In each graph, the blue line represents the longitudinal marker values and the red lines represent the calculated thresholds from the adaptive model at a 99% specificity. The light red shading indicates the hGH treatment period. In order to compare the sensitivity across different durations of

treatment, the adaptive model is only applied to baseline samples and the limit calculated after the last baseline sample is then applied to all ensuing samples.

Figure 5. Sensitivity during treatment and wash-out periods across treatment groups. The considered “treatment” or “wash-out” sample is evaluated by the adaptive model for IGF-1 **(A)**, P-III-NP **(B)** or GH-2000 score **(C)** considering all available baseline samples from the same individual (IGF-1 measured by LC-MS/MS and P-III-NP measured by Siemens ADVIA Centaur). Sensitivity rate for the Biomarkers Test **(D)** and Isoforms Differential Immunoassay **(E)** are based on the population thresholds defined in the applicable WADA Guidelines (3,4). The treatment period is indicated by grey shading.

			Inter-Subject CV			Intra-Subject CV		
IGF-1			Coef.	Lower	Upper	Coef.	Lower	Upper
	Male	LC-MS/MS	20.2%	18.9%	20.7%	18.6%	17.5%	19.7%
		Immunotech	22.5%	23.3%	24.2%	20.0%	18.8%	21.2%
		IDS	20.2%	18.6%	22.0%	13.1%	12.4%	13.9%
	Female	LC-MS/MS	19.0%	17.1%	21.1%	20.1%	18.3%	22.1%
		Immunotech	23.0%	21.1%	25.0%	22.6%	20.6%	24.8%
		IDS	20.3%	18.1%	22.8%	18.0%	16.4%	19.7%
P-III-NP								
	Male	Centaur	20.2%	19.0%	21.5%	24.5%	23.7%	25.4%
		Orion	22.6%	21.7%	23.6%	21.2%	20.5%	21.9%
	Female	Centaur	22.0%	19.7%	24.5%	28.5%	26.9%	30.2%
		Orion	23.6%	21.5%	25.8%	25.7%	24.3%	27.2%
GH-2000	**Male**		10.6%	9.6%	11.8%	11.8%	11.0%	12.6%
	Female		11.9%	10.6%	13.4%	13.7%	12.8%	14.6%

Table 1. Computed coefficient of variation (CV) from mixed model estimated standard deviations. Lower and upper bounds represent the 95% confidence intervals for each CV. IGF-1 and P-III-NP sample values were log-transformed before estimation of their geometric coefficient of variation. Missing and negative sample values (following log-transformation) were excluded (28 samples from 13 athletes). For IGF-1 and P-III-NP variance estimates, N=11,994 samples corresponding to 8,829 male athletes and N=3,953 samples corresponding to 2,613 female athletes were considered. For estimates of GH-2000 score variance, N=2,749 samples corresponding to 1,787 athletes were analyzed, where only the assay pairing of IGF-1 measured by LC-MS/MS and P-III-NP measured by Siemens ADVIA Centaur was considered.

	GH-2000			Biomarkers Tests			Isoforms Differential Immunoassay		
Dose	All	T	W	All	T	F	All	T	W
VL	40%	40%	20%	10%	10%	0%	90%	90%	0%
L	90%	90%	80%	70%	60%	60%	100%	100%	30%
H	100%	100%	100%	71%	57%	71%	100%	100%	43%

Table 2. Sensitivity across dose groups during the entire administration study (All), the treatment period only (T), or the wash-out period (W). For the application of the ABP approach to the GH-2000 score, each sample is evaluated by the adaptive model considering all available baseline samples for that individual based on the assay pairing of IGF-1 measured by LC-MS/MS and P-III-NP measured by Siemens ADVIA Centaur. Athletes with at least one sample flagged during the period of interest were counted. N=10 for both the very low (VL) and low (L) dose groups, and N=7 for the high (H) dose group.

	Sequence > 99.9%		
Dose	**IGF-1**	**P-III-NP**	**GH-2000**
VL	30%	0%	10%
L	60%	70%	70%
H	86%	86%	86%

Table 3. Sensitivity of the sequence-based ABP approach applied to IGF-1, P-III-NP and the GH-2000 score, where IGF-1 was measured by LC-MS/MS and P-III-NP measured by Siemens ADVIA Centaur. All valid samples from each individual treated with hGH were considered together as one passport. Profiles were flagged as atypical if if the probability of an atypical sequence was outside the 99.9% specificity range. N=10 for both the very low (VL) and low (L) dose groups, and N=7 for the high (H) dose group.

	Specificity		
	All	**M**	**F**
IGF-1	98.2%	97.6%	100%
P-III-NP	100%	100%	100%
GH-2000	100%	100%	100%

Table 4. Assessment of specificity of the ABP approach applied to IGF-1, P-III-NP and the GH-2000 score, where IGF-1 was measured by LC-MS/MS and P-III-NP measured by Siemens ADVIA Centaur. Thirty-five passports corresponding to 168 samples (125 for males and 43 for females) from either the control group or baseline period were analyzed by the adaptive model. Results represent the percentage of unflagged samples.

Figure 1

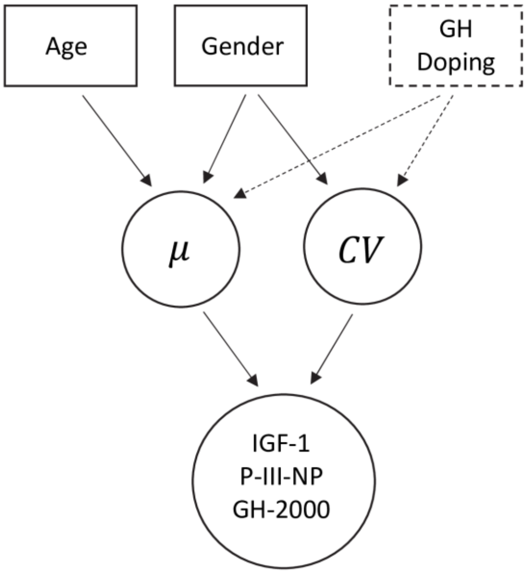


Figure 2

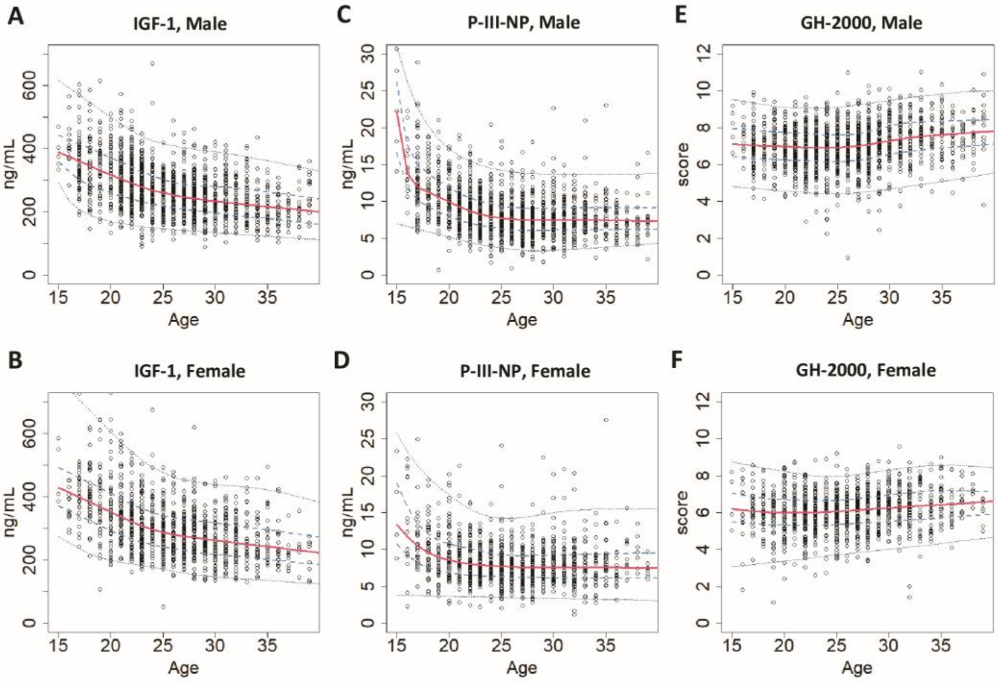

A
IGF-1, Male
ng/mL
0
200
400
600
15
20
25
30
35
Age
C
P-III-NP, Male
ng/mL
0
5
10
15
20
25
30
Age
E
GH-2000, Male
score
0
2
4
6
8
10
12
Age
B
IGF-1, Female
ng/mL
Age
D
P-III-NP, Female
ng/mL
Age
F
GH-2000, Female
score
Age

Figure 3

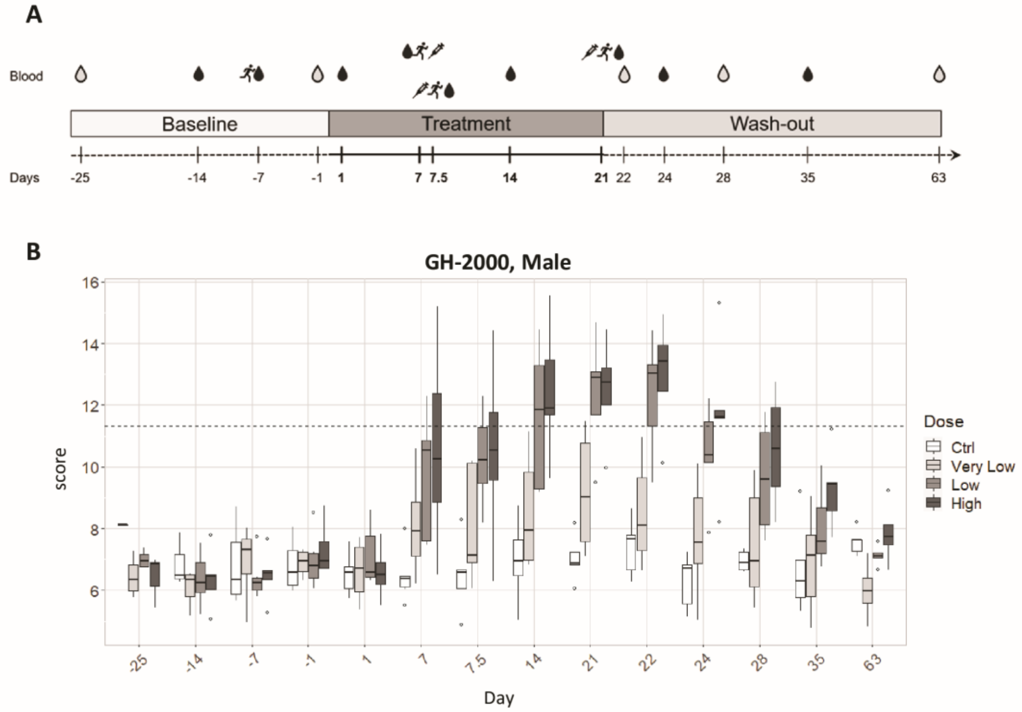

A
Blood
Baseline
Treatment
Wash-out
Days
-25
-14
-7
-1
1
7
7.5
14
21
22
24
28
35
63
B
GH-2000, Male
score
16
14
12
10
8
6
-25
-14
-7
-1
1
7
7.5
14
21
22
24
28
35
63
Day
Dose
Ctrl
Very Low
Low
High

Figure 4

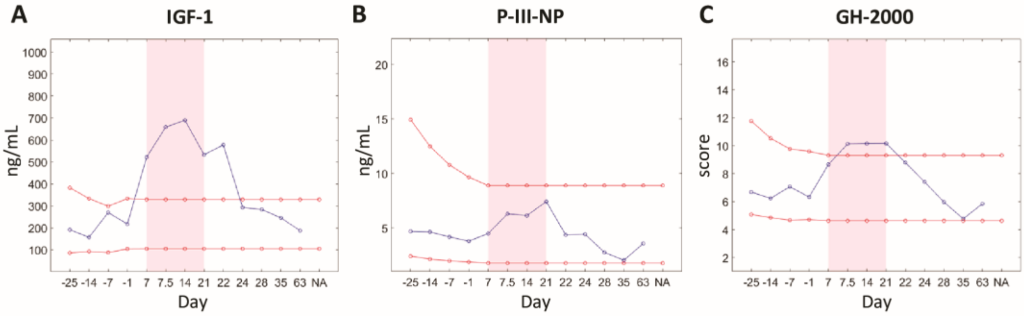
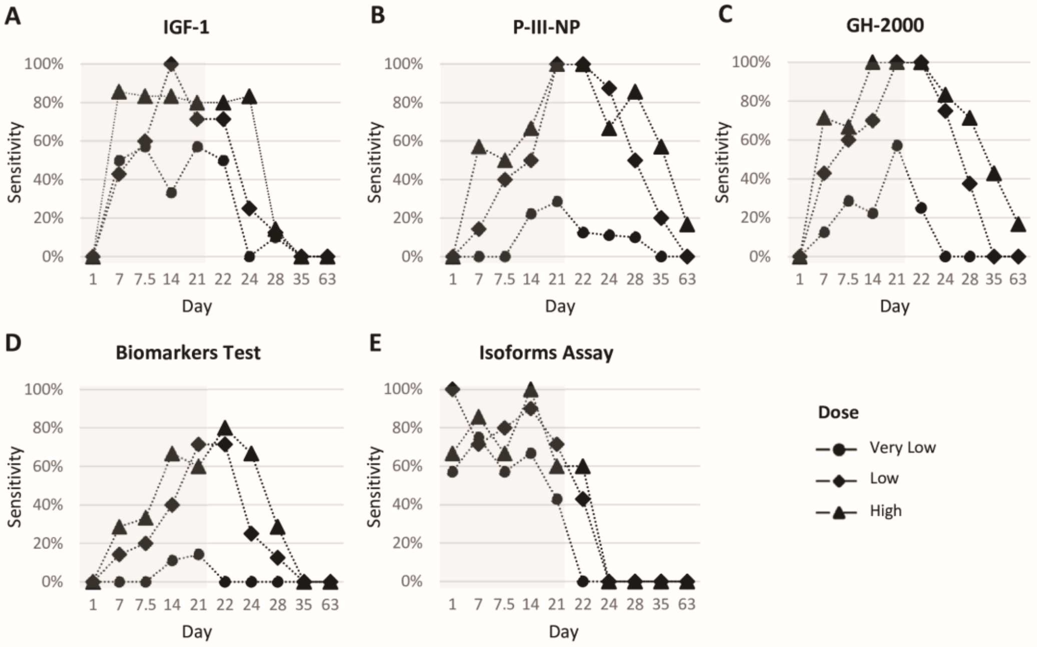

A
IGF-1
1000
900
800
700
600
500
400
300
200
100
ng/mL
-25 -14 -7 -1 7 7.5 14 21 22 24 28 35 63 NA
Day
B
P-III-NP
20
15
10
5
ng/mL
-25 -14 -7 -1 7 7.5 14 21 22 24 28 35 63 NA
Day
C
GH-2000
16
14
12
10
8
6
4
2
score
-25 -14 -7 -1 7 7.5 14 21 22 24 28 35 63 NA
Day

Figure 5


A
IGF-1
B
P-III-NP
C
GH-2000
D
Biomarkers Test
E
Isoforms Assay
Sensitivity
100%
80%
60%
40%
20%
0%
1
7
7.5
14
21
22
24
28
35
63
Day
Dose
Very Low
Low
High