*The Journal of Clinical Endocrinology & Metabolism*, 2022, **XX**, 1–12
https://doi.org/10.1210/clinem/dgac155
Advance access publication 17 March 2022
**Clinical Research Article**

ENDOCRINE SOCIETY

OXFORD

# Reference Curves for Pediatric Endocrinology: Leveraging Biomarker Z-Scores for Clinical Classifications

**Andre Madsen,[1] Bjørg Almås,[1] Ingvild S. Bruserud,[2,3] Ninnie Helen Bakken Oehme,[3] Christopher Sivert Nielsen,[4,5] Mathieu Roelants,[6] Thomas Hundhausen,[7,8] Marie Lindhardt Ljubicic,[9] Robert Bjerknes,[3,10] Gunnar Mellgren,[1,10,11] Jørn V. Sagen,[1,10] Pétur B. Juliusson[3,10,12,*] and Kristin Viste[1,*]**

[1]Hormone Laboratory, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, N-5021 Bergen, Norway
[2]Faculty of Health, VID Specialized University, N-5020 Bergen, Norway
[3]Department of Pediatrics, Haukeland University Hospital, N-5021 Bergen, Norway
[4]Department of Chronic Diseases and Ageing, Norwegian Institute of Public Health, N-0404 Oslo, Norway
[5]Department of Pain Management and Research, Oslo University Hospital, N-0424 Oslo, Norway
[6]Environment and Health, Department of Public Health and Primary Care, KU Leuven, University of Leuven, B-3000 Leuven, Belgium
[7]Department of Medical Biochemistry, Southern Norway Hospital Trust, N-4604 Kristiansand, Norway
[8]Department of Natural Sciences, University of Agder, N-4604 Kristiansand, Norway
[9]Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen, and International Center for Research and Research Training in Endocrine Disruption of Male Reproduction and Child Health (EDMaRC), DK-2100 Copenhagen, Denmark
[10]Department of Clinical Science, University of Bergen, N-5021 Bergen, Norway
[11]Mohn Nutrition Research Laboratory, University of Bergen, N-5021 Bergen, Norway
[12]Department of Health Registries, Norwegian Institute of Public Health, N-5020 Bergen, Norway

**Correspondence:** André Madsen, PhD, Hormone Laboratory, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, N-5021 Bergen, Norway. Email: andre.madsen@uib.no.
*These authors contributed equally to this work

## Abstract

**Context:** Hormone reference intervals in pediatric endocrinology are traditionally partitioned by age and lack the framework for benchmarking individual blood test results as normalized z-scores and plotting sequential measurements onto a chart. Reference curve modeling is applicable to endocrine variables and represents a standardized method to account for variation with gender and age.

**Objective:** We aimed to establish gender-specific biomarker reference curves for clinical use and benchmark associations between hormones, pubertal phenotype, and body mass index (BMI).

**Methods:** Using cross-sectional population sample data from 2139 healthy Norwegian children and adolescents, we analyzed the pubertal status, ultrasound measures of glandular breast tissue (girls) and testicular volume (boys), BMI, and laboratory measurements of 17 clinical biomarkers modeled using the established "LMS" growth chart algorithm in R.

**Results:** Reference curves for puberty hormones and pertinent biomarkers were modeled to adjust for age and gender. Z-score equivalents of biomarker levels and anthropometric measurements were compiled in a comprehensive beta coefficient matrix for each gender. Excerpted from this analysis and independently of age, BMI was positively associated with female glandular breast volume ($\beta = 0.5$, $P < 0.001$) and leptin ($\beta = 0.6$, $P < 0.001$), and inversely correlated with serum levels of sex hormone-binding globulin (SHBG) ($\beta = -0.4$, $P < 0.001$). Biomarker z-score profiles differed significantly between cohort subgroups stratified by puberty phenotype and BMI weight class.

**Conclusion:** Biomarker reference curves and corresponding z-scores provide an intuitive framework for clinical implementation in pediatric endocrinology and facilitate the application of machine learning classification and covariate precision medicine for pediatric patients.

**Key Words:** pediatric endocrinology, biomarker, references, machine learning

**Abbreviations:** AI, artificial intelligence; BGS2, Bergen Growth Study 2; BMI, body mass index; CLSI, Clinical and Laboratory Standards Institute; CV, coefficient of variation; $E_2$, estradiol; FSH, follicle-stimulating hormone; IGF1, insulin-like growth factor 1; LC-MS/MS, liquid chromatography–tandem mass spectrometry; LH, luteinizing hormone; LLOQ, lower limit of quantitation; ML, machine learning; PCA, principal component analysis; RCV, reference change value; ROC, receiver operating characteristics curve; SHBG, sex hormone-binding globulin.

Reproductive hormone references for evaluating blood test results in pediatric patients are essential during clinical investigations of a wide range of conditions including hypo/hypergonadism, differences of sex development (DSDs), and neoplastic and autoimmune conditions that compromise endocrine function. Such pathologies may be associated with abnormal somatic development and altered puberty timing. On the population level, female onset of puberty has decreased by 3 months per decade since the 1970s and has appeared to continue to decline (1). With the secular trend of

**Received:** 17 December 2021. **Editorial Decision:** 10 March 2022. **Corrected and Typeset:** 7 April 2022

earlier puberty timing, particularly observed in girls, pertinent references applied in pediatric endocrinology should periodically be updated. Further, the association between childhood obesity and earlier puberty timing warrants quantitative benchmarking and further attention (2, 3).

During childhood and puberty, circulating levels of hormones and biochemical markers frequently vary considerably with both gender and age. Typically, awakening of the adrenal cortex (adrenarche) precedes attainment of pubic hair (pubarche) and gonadal function (gonadarche) (4). Biochemical reference intervals are fundamental tools to evaluate samples and secure correct diagnosis and treatment. In pediatric endocrinology this necessitates appropriate adjustment or stratification by the major covariates of age, gender, and puberty stage. The well-established and widely applied nonparametric method imposes arbitrary partitioning of age groups to define a series of central 95% CIs in table format (5). However, when assigning a pediatric patient to such predetermined age partitions, the corresponding reference interval will not account for the fact that biochemical observations within most such partitions are likely to exhibit age-dependent skewness, conforming to a non-Gaussian distribution. Further, the Clinical and Laboratory Standards Institute (CLSI) C28-A3c standard upholds that clinically valid reference intervals should ideally be sourced from at least 120 observations (6). In this regard, ethical limitations make it notoriously challenging to recruit cohorts of healthy children to establish sufficiently powered and comprehensive pediatric references (7-9). Notably, the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) and Nordic Reference Interval Project (NORIP) have previously established comprehensive ranges by nonparametric partitioning (10, 11) and quantile regression (12).

Conventional growth charts are ubiquitously used in clinical practice to benchmark the gross anthropometric status of pediatric patients. At the heart of this framework is the LMS method, originally described by Cole and Green (13, 14). Notably, the LMS framework is adopted in most contemporary national growth references, including the growth charts provided by the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) (15, 16). Briefly, the LMS algorithm applies a Box-Cox data transformation and uses 3 parameters to account for the skewness (L), mean (M) and coefficient of variation (S) for each local distribution, effectively providing nonlinear adjustment of age while negating heteroscedasticity (eg, increasing variance with age). Final LMS models enable calculation of standard deviation scores (z-scores) that are typically adjusted for the main covariates age and gender. Such z-scores are also centered at zero (ie, mean for age) and uniformly scaled in terms of SDs and normally distributed, that is, properties that make for ideal input variables in statistical modeling and machine learning (ML). Briefly, supervised ML is an artificial intelligence (AI) method by which an algorithm captures the configuration of several independent feature variables (eg, a biochemical profile) in relation to one dependent variable (eg, a known dichotomy of “disease” or “healthy”) in order to make new and robust predictions (17).

The Bergen Growth Study 2 (BGS2) was conducted in 2016 and has provided new anthropometric puberty references for the Norwegian pediatric population (18, 19) and simple nonparametric hormone references (20, 21). In the current study, we aimed to provide a comprehensive set of LMS gender-specific references curves for 17 pediatric biomarkers. We hypothesized that biomarker z-scores may be useful to quantify associations between hormone levels, pubertal status, and weight class, and thus enable clinical classifications irrespective of patient gender or age. Furthermore, we hypothesized that pediatric overweight may be associated with an altered endocrine profile that, in particular, may be characterized by increased levels of estrogens due to increased adipose tissue aromatase activity. The current reference curves have not been published previously but were recently used to benchmark biomarker z-scores in an unrelated cohort of children exposed to metformin *in utero* due to maternal polycystic ovary syndrome (22).

## Materials and Methods

### Cohort Description

The Bergen Growth Study 2 (BGS2) was conducted in 2016 and comprised a population sample of Norwegian children that was representative of the general Norwegian demographic composition, consisting of approximately 90% Caucasians as described previously (20, 21). Exclusion criteria in the BGS2 cohort included self-reported chronic disease or a medical history of cancer or epilepsy. Data describing puberty status, anthropometric profile, and biochemical data were available for 650 healthy girls and 465 healthy boys in the age interval from 6 to 16 years. The BGS2 total participation rate, that is, ratio of children invited and children enrolled in the study was 43%.

The Norwegian Fit Futures 1 youth study was conducted in 2010 and 2011 to benchmark public health parameters pertaining to lifestyle choices, bone health, and inflammation as described previously (23, 24). All 1117 first year high-school students in Tromsø and Balsfjord municipalities were invited and 1038 participated, yielding a response rate of 93%. From this cohort, previously unpublished data pertaining to steroid hormone levels in girls and boys aged 15 to 18 years was used in the current study. An overview of the current sample sizes and applied exclusion criteria is provided (Table 1). Data regarding puberty status in the Fit Futures 1 cohort was self-reported and did not comply with Tanner staging performed in the BGS2 cohort; these observations were accordingly not included in the current references that were arranged by puberty stages.

Both written parental consent and child assent was required for any examination, and sourcing biochemical data from the biobanks was approved by Norwegian Regional Committees for Medical and Health Research Ethics (approval references REK-2015/128 for BGS2 and REK-2017/1976 for Fit Futures, respectively).

## Methods

### Puberty Evaluation Protocols

For girls in the BGS2 cohort, the depth and diameter of the fibroglandular area was systematically measured with ultrasound in each girl’s left breast in a sagittal plane, unless the right breast visually appeared more mature. Methodological documentation of the ultrasonographic measurement of glandular depth and diameter was described previously (25). Briefly, breasts were palpated and evaluated according to

**Table 1. Cohort sample overview**

| Sample description | BGS2 cohort (6 to <16 y) | | Fit Futures cohort (15-18 y) | |
|---|---|---|---|---|
| Gender | Boys | Girls | Boys | Girls |
| Unique blood samples, n | 451 | 650 | 509 | 486 |
| Excluded due to chronic disease, n | 25 | 27 | 8 | 0 |
| Excluded due to oral contraceptives, n | 0 | 12 | 0 | 167 |
| Excluded due to corticosteroid use, n | 7 | 10 | 10 | 6 |
| Viable blood samples for references, n | 414 | 601 | 491 | 319 |

The current study included observations sourced from the 2 population-based samples of Norwegian children and adolescents enrolled in the Bergen Growth Study 2 (BGS2) and Fit Futures 1 cohorts. Indicated numbers of girls and boys were enrolled and the following exclusion criteria were applied: self-reported history of chronic disease or cancer (excluded from all biomarker references); use of oral contraceptives (excluded from all female biomarker references); use of corticosteroid medication (excluded from cortisol references). The number of viable blood samples used in the current references excludes serum samples that were discarded due to hemolysis or insufficient blood draw volume.

Tanner's classification (26). Ultrasound evaluation of the largest breast was performed with SonoSite Edge (FUJIFILM SonoSite, USA) device with a 15-16 MHz (5 cm) linear transducer. Two consecutive scans were performed and merged when the diameter was 5 to 10 cm, and the measurements were summed. Diameters above 10 cm were not measured or included in analyses. The more mature breast, according to Tanner B or ultrasound breast staging, was used for the analyses. To calculate glandular volume, the formula for a conical shape was applied: volume = (π/3) × radius × 2 × depth. This mathematical formula for a conical shape has also been used by others (27, 28). Using Tanner stages as the gold standard marker of thelarche (Tanner 1 vs Tanner B2+), the optimal cutoff to classify puberty onset in girls corresponded to 0.5 mL of glandular breast tissue volume, and this threshold exhibited a positive predictive value of 60.4%, negative predictive value of 98.6%, and accuracy of 85.2%.

Methodological documentation for ultrasound evaluation of male testicular volume and its mathematical relation to conventional orchidometer milliliters was detailed previously (29). Briefly, the dimensions of the biggest testicle were recorded, and the ellipsoid volume was calculated using Lambert equation (length × width × depth × 0.71). The Norwegian growth chart describing male testicular volume-for-age was sourced from the BGS2 cohort and recently published (19). Anthropometric body mass index (BMI), waist circumference, and subscapular triceps z-scores assigned to participants in the current study were interpolated from the Norwegian national growth charts according to gender and age (30).

## Blood Sample Analyses

Venous blood samples were collected with both parental and child consent if the child was younger than 16 years of age, and with consent of the person if he/she was older than 16 years. Blood samples were collected between 8 am and 2 pm in both studies. Isolated serum was stored in registered biobanks at −80 °C prior to analyses. All biomarkers were analyzed in the standard international (SI) unit framework at the Hormone Laboratory, Department of Medical Biochemistry and Pharmacology, Haukeland University Hospital, accredited in compliance with ISO 15189:2012. Androgens and corticosteroids were analyzed by liquid chromatography–tandem mass spectrometry (LC-MS/MS) multiplex method as described previously (31). For testosterone, the analytical inter-assay coefficient of variation ($CV_A$) was 4% in the range 1.5 to 37 nmol/L and the lower limit of quantification (LLOQ) was 0.02 nmol/L. For BGS2 samples, serum levels of estradiol ($E_2$) were quantified using an ultrasensitive LC-MS/MS method documented previously (32). Here, the $E_2$ analytical inter-assay $CV_A$ was 9.1% in the range 1.7 to 153.3 pmol/L and the LLOQ was 0.58 pmol/L. In the Fit Futures cohort, $E_2$ levels were determined by an LC-MS/MS method with intermediate sensitivity ($CV_A$ 13% at 57 pmol/L; range, 13-2508 pmol/L; LLOQ 13 pmol/L). The 2 methods of $E_2$ determination are traceable to the CRM BCR-576, and no significant bias was detected between the 2 methods when biological samples were run in parallel ($R^2 = 0.96$; average difference = 1.7 % and *t* test *P* = 0.053). Estrogen level data from the 2 cohorts were hence merged without mathematical adjustments. Follicle-stimulating hormone (FSH; $CV_A$ 5% at 5 IU/L; LLOQ 0.1 IU/L), luteinizing hormone (LH; $CV_A$ 7% at 10 IU/L; LLOQ 0.1 IU/L), sex hormone-binding globulin (SHBG; $CV_A$ 6% at 60 nmol/L; LLOQ 2 nmol/L) and insulin-like growth factor 1 (IGF1; $CV_A$ 7% at 18 nmol/L; LLOQ 4 nmol/L) were quantified in BGS2 serum samples using Siemens Immulite 2000 XPi. Enzyme-linked immunosorbent assay kits were used to quantify serum leptin (Mediagnost Cat# E07, RRID: AB_2813737) and adiponectin (Mediagnost Cat# E09, RRID: AB_2813736) in serum samples from the BGS2 cohort. Inter-assay $CV_A$ was determined to 5% at 8.3 µg/L leptin (LOQ 1-100 µg/L) and 8% at 14 µg/mL adiponectin (LOQ 0.6-31 µg/mL). Levels of HDL cholesterol ($CV_A$ 3% at 1.9 mmol/L), LDL cholesterol ($CV_A$ 2.5% at 3.4 mmol/L), total cholesterol ($CV_A$ 3% at 4.4 mmol/L) and triglycerides ($CV_A$ 3% at 1.5 mmol/L) in BGS2 serum samples were quantified by Cobas 8000.

## Hormone Reference Intervals

Biomarker reference curves were modeled using the LMS method provided in the "gamlss" package in R (33). No outliers were removed outside cohort exclusion criteria. The combined triplet of values assigned for L, M, and S enables calculation of z-scores adjusted for gender and age by the following formula: (((X/M)^L)-1)/(L*S) where X is the relevant blood test result in SI units. All LMS models in the current study are provided in Supplemental Table 1 (34). Quality assurance and satisfactory residual distribution of LMS references was assured by QQ-plots, worm-plots, and Q-tests for normality of each reference model. The computational script used to perform the above LMS operations is available in R code (35). Traditionally partitioned and nonparametric

95% reference intervals for all biomarkers were established by bootstrapping and Dixon's outlier removal using the "referenceIntervals" package in R (36) and are provided in table format in Supplemental Table 2 (37). Partitioning of the reference ranges was determined according to CLSI guidelines (6).

Girls using oral contraceptives were not included in any reference intervals, and children using glucocorticoid medication were not included in cortisol and 11-deoxycortisol references, specifically.

### Statistical Analyses

Correlation matrices were computed using the Pearson method with the "reshape" and "ggplot2" packages in R. The $P$ values for the correlation matrices are provided in Supplemental Table 3 (38). Total variance in the biomarker z-score dataset was explored by principal component analysis (PCA), using the "prcomp" and "ggbiplot" functions in R as described previously (21). Supervised machine learning (ML) to predict weight class (BMI-SDS ≥ 1 or BMI-SDS ≤ −1) from all featured biomarker variables was performed by establishing a "randomForest" model and evaluating the resulting confusion matrix using the "caret" package in R. The pipeline script used to perform the above operations is available in R code (39). Complete observations for all biomarker z-scores were available for 122 boys and 172 girls from the BGS2 cohort and combined to one data frame, from which the random forest model was trained using 75% of the data and tested using 25% of the remaining and unseen data with 10-fold cross-validation.

Receiver operating characteristics (ROC) curves were constructed using the "pROC" package in R to evaluate the ability of single biomarkers to distinguish between the weight classes specified above. ROC accuracy was calculated as (true negatives + true positives)/all classification outcomes.

## Results

### Continuous Hormone References

We combined data from the 2 Norwegian cohorts of healthy children and adolescents and modeled circulating steroid hormone levels in girls and boys in relation to chronological age using the LMS growth chart algorithm (Fig. 1). Additional biomarkers analyzed exclusively in the BGS2 cohort included peptide hormones and lipids (Fig. 2). The reference curves showed in Fig. 1 and Fig. 2 are provided as supplementary information and enable anyone to calculate biomarker z-scores adjusted for gender and age. Notably, observations located on the mean-for-age centile have a z-score of 0, corresponding to the 50th percentile.

### Age-adjusted Associations Between Endocrine, Pubertal, and Anthropometric Variables

From the current biomarker reference curves (Fig. 1 and Fig. 2), age-adjusted z-scores were calculated for each cohort participant. Combining these biomarker z-scores with conventional anthropometric z-scores, we next calculated the Pearson correlation between all variables, according to gender (Fig. 3). Hence, the provided correlation coefficients are standardized beta coefficients that specify relationships between all variables in terms of SD and irrespective of age. Sample sizes for these analyses were 552 to 995 girls and 419 to 910 boys, since puberty endpoints were not included in the Fit Futures dataset. No correlation was observed between z-scores and chronological age, indicating successful adjustments for age. In boys, both total testosterone and LH z-scores were positively associated with testicular volume-for-age ($\beta = 0.4$ and $P < 0.001$ for both). In girls, LH, FSH, and IGF1 z-scores were positively associated with $E_2$ ($\beta = 0.5$ to 0.6, respectively; $P < 0.001$ for both). BMI z-scores associated positively with male testicular volume-for-age ($\beta = 0.2$ and $P < 0.001$) and female glandular breast tissue volume-for-age ($\beta = 0.5$ and $P < 0.001$) but negatively with SHBG in both genders ($\beta = -0.4$ and $P < 0.001$ for both). Strong β coefficients were observed between obesometric variables and the same was also true for related biomarkers in the steroid hormone synthesis pathway, eg, between estrone ($E_1$) and estradiol ($E_2$).

### Endocrine Features of Pubertal Phenotypes

A refined analysis of hormone profile in relation to pubertal phenotypes was achieved by stratifying the BGS2 cohort according to attainment of pubarche or central puberty onset at the time of examination (Table 2). Specifically, participants were grouped according to attainment of pubic hair (Tanner pubic hair stage PH2) and/or canonical markers of pubertal onset for boys (testicular volume ≥ 4 mL) and girls (Tanner breast stage B2+ ie, thelarche). The earliest and latest occurrences of puberty onset were between 10 and 13 years for boys and between 8 and 12 years for girls. Boys exhibiting central pubertal onset without pubarche had significantly higher z-scores of LH, total testosterone, and testicular volume than prepubertal peers with or without pubarche. Girls exhibiting gonadarche but no pubarche had significantly higher z-scores of FSH, $E_2$, and glandular breast tissue volume than prepubertal peers without pubarche. Cohort participants presenting with both pubarche and gonadarche exhibited pubertal and endocrine z-scores markedly above the mean for age.

### Leveraging Biomarker Z-Scores for Clinical Classifications

We initially hypothesized that weight classes would associate with differential biomarker profiles, and we pursued this hypothesis as a classification problem. In order to characterize the biomarker profile and explore the utility of z-scores in clinical classifications, we applied a principal component analysis (PCA) to "fingerprint" the biomarker profiles associated with BMI weight class. This analysis included the biomarker z-score profile for 154 "underweight" (BMI z-score ≤ −1.0) and all 140 "overweight" (BMI z-score ≥ 1.0) girls and boys in the BGS2 study. In addition to biomarker z-scores, female glandular tissue-for-age and male testicular volume-for-age were included to evaluate associations between gonadal development and weight class. The resulting PCA biplot showed a partially distinct clustering of underweight and overweight biomarker z-score profiles (Fig. 4). From the previous beta coefficients matrices (Fig. 2), we observed a positive and association between BMI and circulating levels of leptin for both genders ($\beta = 0.6$ and $P < 0.001$). In line with this finding, the PCA factor analysis (arrows) indicated that leptin, along with SHBG and IGF1, were important biomarkers of weight class. Subsequent ROC analyses verified that classification of overweight was achieved by leptin (88.8% accuracy), SHBG

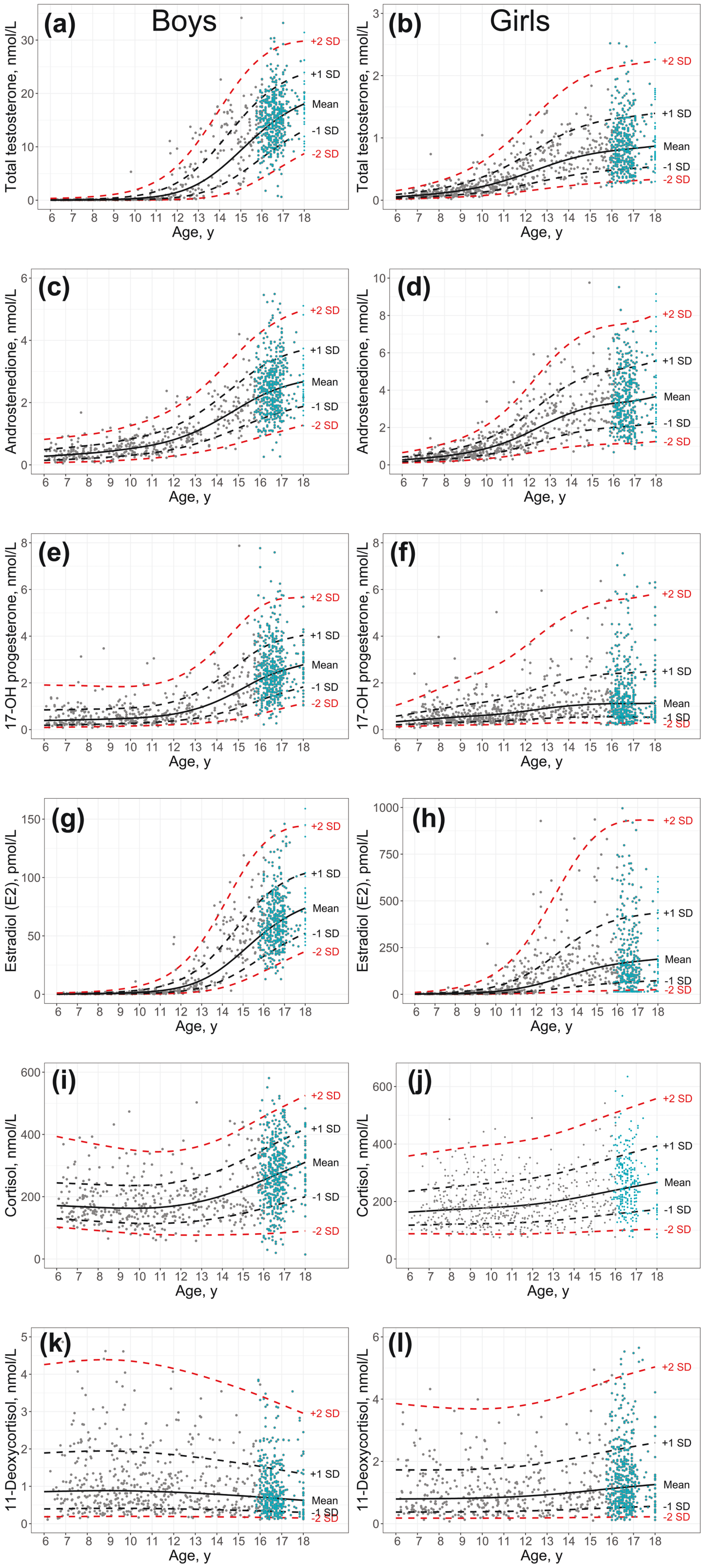


**Figure 1.** Continuous steroid hormone reference curves. Steroid hormone levels in individuals enrolled in the Bergen Growth Study 2 (black dots) and Fit Futures cohort (green dots) were quantified by LC-MS/MS. Male (left column panels a, c, e, g, i, k) and female (right column panels b, d, f, h, j, l) references were modeled separately for indicated hormones. Continuous centiles indicating the mean for age (p50, solid lines) and discrete SDs from the mean (dashed lines) were fitted using the LMS algorithm. The −2 and + 2 SD curves correspond to percentiles p2.2 and p97.8, respectively, and the vertical range between these centiles approximate the 95% CI at any age. Abbreviations: 11-DOC, 11-deoxycortisol; 17-OHP, 17-hydroxyprogesterone; y, chronological age in years.

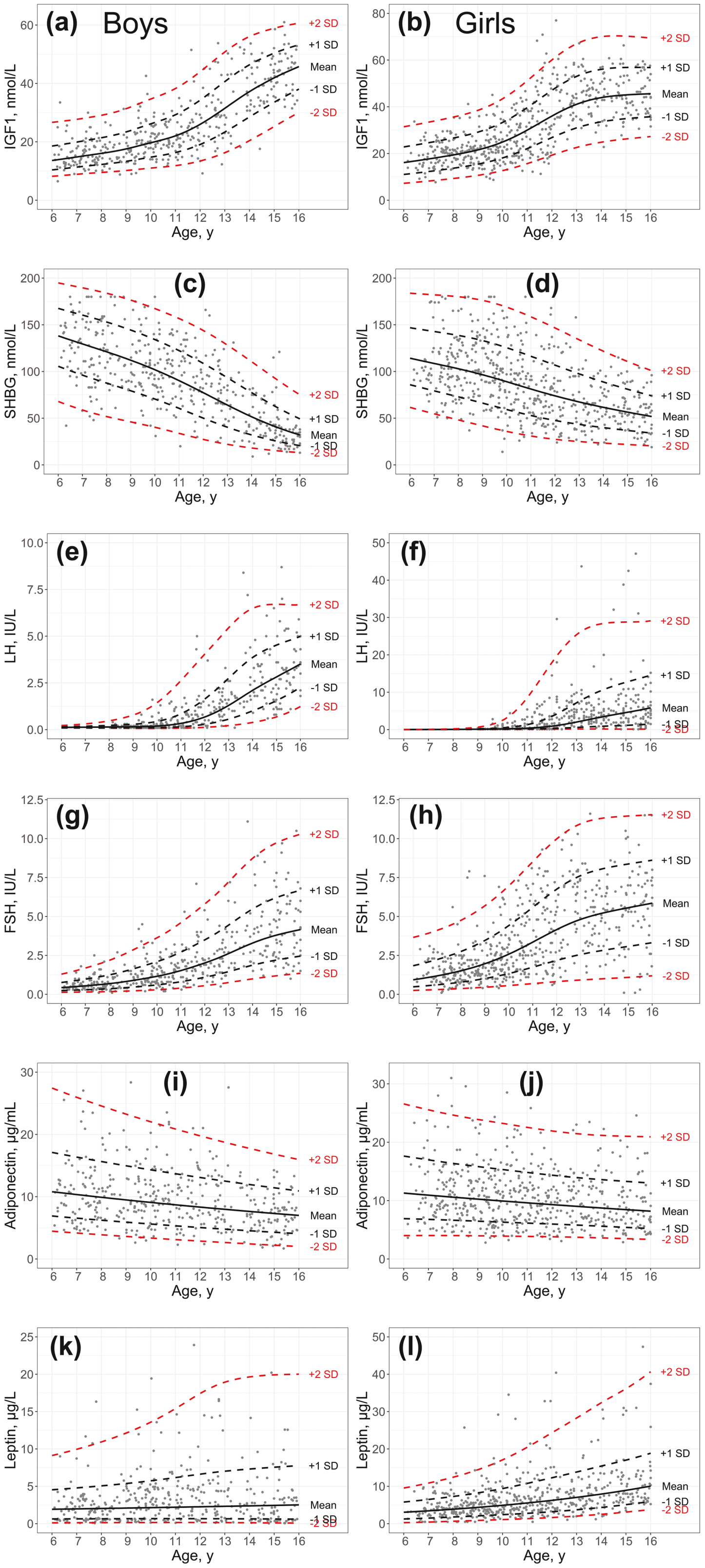
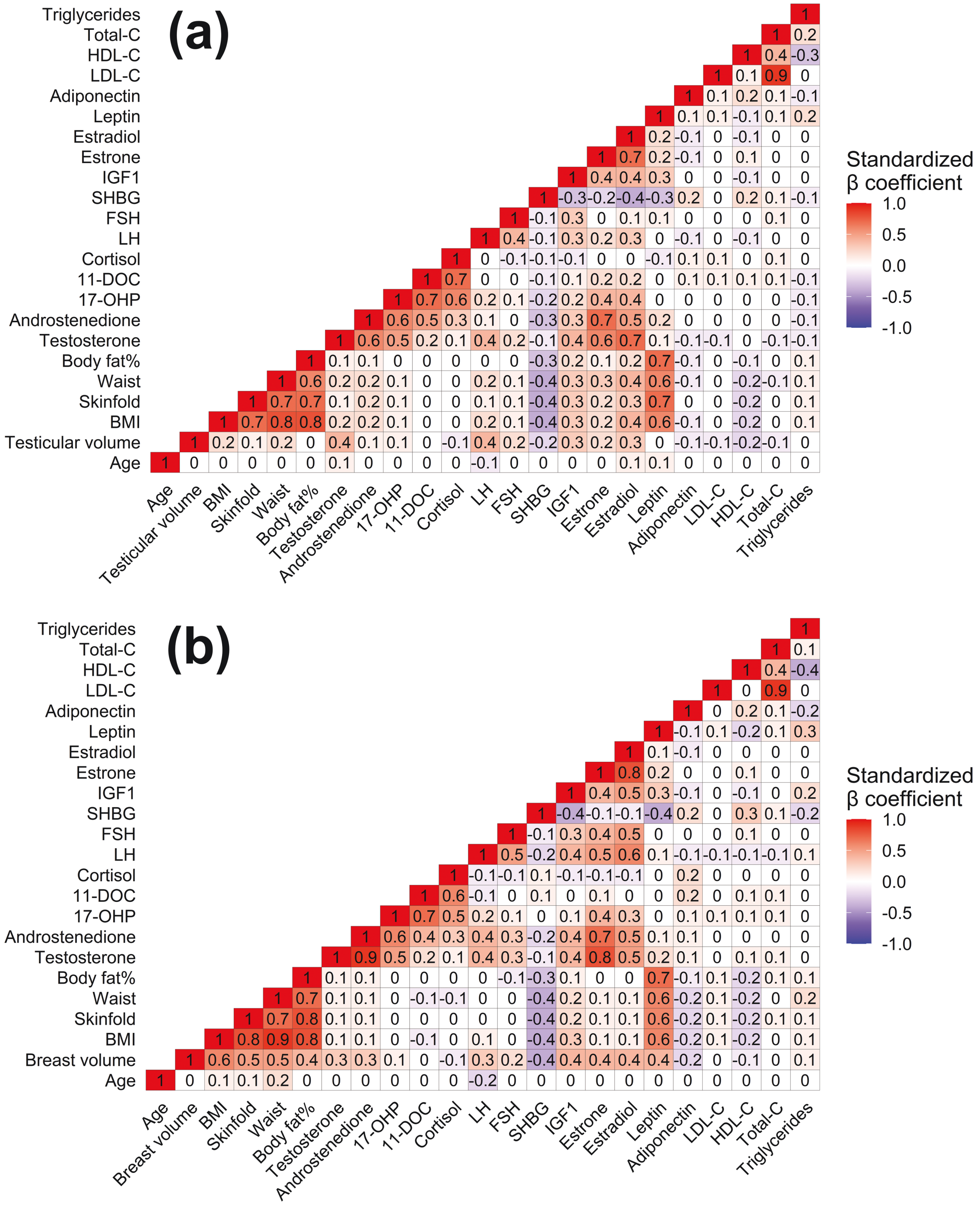


**Figure 2.** Continuous biomarker reference curves. Biomarker levels quantified in serum samples from the BGS2 cohort were modeled as reference curves using the LMS algorithm. Male (left column panels a, c, e, g, i, k) and female (right column panels b, d, f, h, j, l) references were modeled separately. Abbreviations: FSH, follicle-stimulating hormone; IGF1, insulin-like growth factor 1; LH, luteinizing hormone; SHBG, sex hormone-binding globulin; y, chronological age in years.

(75.2% accuracy), and IGF1 (69.4% accuracy), respectively. As a proof-of-concept for weight class prediction using the entire biomarker z-score profile, we trained a supervised machine learning (ML) model using the decision tree–based "random forest" algorithm. By evaluating only biomarker z-scores, this ML classification model was able to predict BMI



(a)

| | Age | Testicular volume | BMI | Skinfold | Waist | Body fat% | Testosterone | Androstenedione | 17-OHP | 11-DOC | Cortisol | LH | FSH | SHBG | IGF1 | Estrone | Estradiol | Leptin | Adiponectin | LDL-C | HDL-C | Total-C | Triglycerides |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Triglycerides | | | | | | | | | | | | | | | | | | | | | | | 1 |
| Total-C | | | | | | | | | | | | | | | | | | | | | | 1 | 0.2 |
| HDL-C | | | | | | | | | | | | | | | | | | | | | 1 | 0.4 | -0.3 |
| LDL-C | | | | | | | | | | | | | | | | | | | | 1 | 0.1 | 0.9 | 0 |
| Adiponectin | | | | | | | | | | | | | | | | | | | 1 | 0.1 | 0.2 | 0.1 | -0.1 |
| Leptin | | | | | | | | | | | | | | | | | | 1 | 0.1 | 0.1 | -0.1 | 0.1 | 0.2 |
| Estradiol | | | | | | | | | | | | | | | | | 1 | 0.2 | -0.1 | 0 | -0.1 | 0 | 0 |
| Estrone | | | | | | | | | | | | | | | | 1 | 0.7 | 0.2 | -0.1 | 0 | 0.1 | 0 | 0 |
| IGF1 | | | | | | | | | | | | | | | 1 | 0.4 | 0.4 | 0.3 | 0 | 0 | -0.1 | 0 | 0 |
| SHBG | | | | | | | | | | | | | | 1 | -0.3 | -0.2 | -0.4 | -0.3 | 0.2 | 0 | 0.2 | 0.1 | -0.1 |
| FSH | | | | | | | | | | | | | 1 | -0.1 | 0.3 | 0 | 0.1 | 0.1 | 0 | 0 | 0 | 0.1 | 0 |
| LH | | | | | | | | | | | | 1 | 0.4 | -0.1 | 0.3 | 0.2 | 0.3 | 0 | -0.1 | 0 | -0.1 | 0 | 0 |
| Cortisol | | | | | | | | | | | 1 | 0 | -0.1 | -0.1 | -0.1 | 0 | 0 | -0.1 | 0.1 | 0.1 | 0 | 0.1 | 0 |
| 11-DOC | | | | | | | | | | 1 | 0.7 | 0 | 0 | -0.1 | 0.1 | 0.2 | 0.2 | 0 | 0.1 | 0.1 | 0.1 | 0.1 | -0.1 |
| 17-OHP | | | | | | | | | 1 | 0.7 | 0.6 | 0.2 | 0.1 | -0.2 | 0.2 | 0.4 | 0.4 | 0 | 0 | 0 | 0 | 0 | -0.1 |
| Androstenedione | | | | | | | | 1 | 0.6 | 0.5 | 0.3 | 0.1 | 0 | -0.3 | 0.3 | 0.7 | 0.5 | 0.2 | 0 | 0 | 0 | 0 | -0.1 |
| Testosterone | | | | | | | 1 | 0.6 | 0.5 | 0.2 | 0.1 | 0.4 | 0.2 | -0.1 | 0.4 | 0.6 | 0.7 | 0.1 | -0.1 | -0.1 | 0 | -0.1 | -0.1 |
| Body fat% | | | | | | 1 | 0.1 | 0.1 | 0 | 0 | 0 | 0 | 0 | -0.3 | 0.2 | 0.1 | 0.2 | 0.7 | -0.1 | 0 | -0.1 | 0 | 0.1 |
| Waist | | | | | 1 | 0.6 | 0.2 | 0.2 | 0.1 | 0 | 0 | 0.2 | 0.1 | -0.4 | 0.3 | 0.3 | 0.4 | 0.6 | -0.1 | 0 | -0.2 | -0.1 | 0.1 |
| Skinfold | | | | 1 | 0.7 | 0.7 | 0.1 | 0.2 | 0.1 | 0 | 0 | 0.1 | 0.1 | -0.4 | 0.3 | 0.2 | 0.3 | 0.7 | 0 | 0 | -0.2 | 0 | 0.1 |
| BMI | | | 1 | 0.7 | 0.8 | 0.8 | 0.2 | 0.2 | 0.1 | 0 | 0 | 0.2 | 0.1 | -0.4 | 0.3 | 0.2 | 0.4 | 0.6 | -0.1 | 0 | -0.2 | 0 | 0.1 |
| Testicular volume | | 1 | 0.2 | 0.1 | 0.2 | 0 | 0.4 | 0.1 | 0.1 | 0 | -0.1 | 0.4 | 0.2 | -0.2 | 0.3 | 0.2 | 0.3 | 0 | -0.1 | -0.1 | -0.2 | -0.1 | 0 |
| Age | 1 | 0 | 0 | 0 | 0 | 0 | 0.1 | 0 | 0 | 0 | 0 | -0.1 | 0 | 0 | 0 | 0 | 0.1 | 0.1 | 0 | 0 | 0 | 0 | 0 |

(b)

| | Age | Breast volume | BMI | Skinfold | Waist | Body fat% | Testosterone | Androstenedione | 17-OHP | 11-DOC | Cortisol | LH | FSH | SHBG | IGF1 | Estrone | Estradiol | Leptin | Adiponectin | LDL-C | HDL-C | Total-C | Triglycerides |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Triglycerides | | | | | | | | | | | | | | | | | | | | | | | 1 |
| Total-C | | | | | | | | | | | | | | | | | | | | | | 1 | 0.1 |
| HDL-C | | | | | | | | | | | | | | | | | | | | | 1 | 0.4 | -0.4 |
| LDL-C | | | | | | | | | | | | | | | | | | | | 1 | 0 | 0.9 | 0 |
| Adiponectin | | | | | | | | | | | | | | | | | | | 1 | 0 | 0.2 | 0.1 | -0.2 |
| Leptin | | | | | | | | | | | | | | | | | | 1 | -0.1 | 0.1 | -0.2 | 0.1 | 0.3 |
| Estradiol | | | | | | | | | | | | | | | | | 1 | 0.1 | -0.1 | 0 | 0 | 0 | 0 |
| Estrone | | | | | | | | | | | | | | | | 1 | 0.8 | 0.2 | 0 | 0 | 0.1 | 0 | 0 |
| IGF1 | | | | | | | | | | | | | | | 1 | 0.4 | 0.5 | 0.3 | -0.1 | 0 | -0.1 | 0 | 0.2 |
| SHBG | | | | | | | | | | | | | | 1 | -0.4 | -0.1 | -0.1 | -0.4 | 0.2 | 0 | 0.3 | 0.1 | -0.2 |
| FSH | | | | | | | | | | | | | 1 | -0.1 | 0.3 | 0.4 | 0.5 | 0 | 0 | 0 | 0.1 | 0 | 0 |
| LH | | | | | | | | | | | | 1 | 0.5 | -0.2 | 0.4 | 0.5 | 0.6 | 0.1 | -0.1 | -0.1 | -0.1 | -0.1 | 0.1 |
| Cortisol | | | | | | | | | | | 1 | -0.1 | -0.1 | 0.1 | -0.1 | -0.1 | -0.1 | 0 | 0.2 | 0 | 0 | 0 | 0 |
| 11-DOC | | | | | | | | | | 1 | 0.6 | -0.1 | 0 | 0.1 | 0 | 0.1 | 0 | 0 | 0.2 | 0 | 0.1 | 0.1 | 0 |
| 17-OHP | | | | | | | | | 1 | 0.7 | 0.5 | 0.2 | 0.1 | 0 | 0.1 | 0.4 | 0.3 | 0 | 0.1 | 0.1 | 0.1 | 0.1 | 0 |
| Androstenedione | | | | | | | | 1 | 0.6 | 0.4 | 0.3 | 0.4 | 0.3 | -0.2 | 0.4 | 0.7 | 0.5 | 0.1 | 0.1 | 0 | 0 | 0 | 0 |
| Testosterone | | | | | | | 1 | 0.9 | 0.5 | 0.2 | 0.1 | 0.4 | 0.3 | -0.1 | 0.4 | 0.8 | 0.5 | 0.2 | 0.1 | 0 | 0.1 | 0.1 | 0 |
| Body fat% | | | | | | 1 | 0.1 | 0.1 | 0 | 0 | 0 | 0 | -0.1 | -0.3 | 0.1 | 0 | 0 | 0.7 | -0.1 | 0.1 | -0.2 | 0.1 | 0.1 |
| Waist | | | | | 1 | 0.7 | 0.1 | 0.1 | 0 | -0.1 | -0.1 | 0 | 0 | -0.4 | 0.2 | 0.1 | 0.1 | 0.6 | -0.2 | 0.1 | -0.2 | 0 | 0.2 |
| Skinfold | | | | 1 | 0.7 | 0.8 | 0.1 | 0.1 | 0 | 0 | 0 | 0 | 0 | -0.4 | 0.2 | 0.1 | 0.1 | 0.6 | -0.2 | 0.1 | -0.2 | 0.1 | 0.1 |
| BMI | | | 1 | 0.8 | 0.9 | 0.8 | 0.1 | 0.1 | 0 | -0.1 | 0 | 0.1 | 0 | -0.4 | 0.3 | 0.1 | 0.1 | 0.6 | -0.2 | 0.1 | -0.2 | 0 | 0.1 |
| Breast volume | | 1 | 0.6 | 0.5 | 0.5 | 0.4 | 0.3 | 0.3 | 0.1 | 0 | -0.1 | 0.3 | 0.2 | -0.4 | 0.4 | 0.4 | 0.4 | 0.4 | -0.2 | 0 | -0.1 | 0 | 0.1 |
| Age | 1 | 0 | 0.1 | 0.1 | 0.2 | 0 | 0 | 0 | 0 | 0 | 0 | -0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 3.** Standardized β coefficient matrices for puberty development, anthropometry, and hormone profile. Age-adjusted z-scores derived from anthropometric LMS growth charts and the current biomarker LMS reference curves were correlated to obtain standardized beta coefficients that describe relationships between all variables. To exemplify the readout, 1 SD score increase in BMI incurs a 0.6 SD score increase in circulating levels of leptin, regardless of age. Testicular volume-for-age z-scores were included in the top (a) male matrix and corresponding z-scores for female glandular tissue volume-for-age were included in the bottom (b) female matrix. Standardized β coefficients were calculated as the linear regression (Pearson r) between pairwise z-scores and colored according to the indicated heatmap scale. Complete statistical analyses including β coefficient *P* values are available in Supplemental Table 3. (38).

weight class with an accuracy of 94.5% (95% CI, 86.6% to 98.5%), as shown in the classification table (Table 3).

## Discussion

A critical function of clinical laboratories is to construct and maintain updated biochemical references to guide medical decision making. Establishing suitable references for the pediatric population is especially challenging due to ethical, practical, and regulatory impediments to the recruitment of healthy blood donors while also keeping up to date with the secular trend of earlier pubertal onset. The current and widely adopted nonparametric method to construct reference intervals for arbitrarily partitioned age groups may not

**Table 2. Baseline characteristics of male and female puberty phenotypes**

| **Male baseline characteristics (puberty onset age range, 10-13 years)** | | | | |
|---|---|---|---|---|
| **Boys, ages 10-13** | **TV < 4 mL; PH1** | **TV ≥ 4 mL; PH1** | **TV < 4 mL; PH2+** | **TV ≥ 4 mL; PH2+** |
| Sample size, n | 69 | 23 | 20 | 37 |
| Attained testicular vol. ≥ 4 mL, % | 0% | 100% | 0% | 100% |
| Attained pubic hair ≥ PH2, % | 0% | 0% | 100% | 100% |
| Age, y | 10.74 (10.07 to 12.54) | 11.82 (10.37 to 12.92) | 11.64 (10.36 to 12.78) | 12.47 (11.01 to 12.93) |
| Testicular volume, z-score | −0.55 (−1.84 to 1.01) | 0.79 (−0.69 to 3.13) | −0.72 (−1.82 to 0.71) | 0.34 (−0.95 to 2.30) |
| LH, z-score | −0.75 (−1.97 to 1.44) | 0.81 (−0.93 to 1.99) | −0.30 (−1.95 to 1.92) | 0.54 (−0.55 to 2.39) |
| FSH, z-score | −0.25 (−2.39 to 1.35) | 0.07 (−0.86 to 1.93) | 0.16 (−1.66 to 1.42) | 0.27 (−1.43 to 1.77) |
| Testosterone, z-score | −0.37 (−1.59 to 0.85) | 0.19 (−1.10 to 2.92) | −0.46 (−1.43 to 1.34) | 0.64 (−0.66 to 2.29) |
| **Female baseline characteristics (puberty onset age range, 8-12 years)** | | | | |
| **Girls, ages 8-12** | **No thelarche; PH1** | **Thelarche; PH1** | **No thelarche; PH2+** | **Thelarche; PH2+** |
| Sample size, n | 92 | 28 | 11 | 35 |
| Attained breasts ≥ Tanner B2, % | 0% | 100% | 0% | 100% |
| Attained pubic hair ≥ PH2, % | 0% | 0% | 100% | 100% |
| Age, y | 9.23 (8.08 to 11.67) | 10.46 (8.58 to 12.30) | 9.93 (8.15 to 11.26) | 11.31 (9.90 to 11.95) |
| Glandular tissue volume, z-score | −0.49 (−2.09 to 1.24) | 0.68 (−1.01 to 1.82) | −0.13 (−1.61 to 1.44) | 1.16 (−0.86 to 1.96) |
| LH, z-score | 0.20 (−1.95 to 1.90) | 0.06 (−1.54 to 2.34) | −0.41 (−1.39 to 1.85) | 1.05 (−2.05 to 1.85) |
| FSH, z-score | −0.15 (−2.06 to 1.59) | 0.53 (−0.72 to 2.41) | −0.05 (−1.44 to 1.14) | 0.53 (−1.49 to 1.84) |
| $E_2$, z-score | −0.40 (−2.55 to 1.21) | 0.42 (−1.26 to 2.64) | −0.12 (−1.61 to 1.24) | 0.88 (−1.73 to 2.03) |

Participants in the BGS2 cohort were stratified by differential puberty phenotypes at the time of examination, and the resulting sample sizes and baseline characteristics are presented as median (p2.5 to p97.5). The earliest and latest occurrences of puberty onset in the dataset, defined by attainment of 4 mL orchidometer testicular volume (boys) or Tanner stage B2 (girls), were set as respective age limits for this stratification analysis. Abbreviations: $E_2$, estradiol; FSH, follicle-stimulating hormone; LH, luteinizing hormone; PH, Tanner pubic hair stage; SDS, z-score measured in SD from the mean for age; US, ultrasound; y, years.

satisfactorily capture the age-dependent trends that are continuous in nature. Compared with the continuous distributions and centiles obtained by the LMS method, nonparametric reference intervals would necessarily require partitioning of age groups (eg, 6 to < 9 years; 9 to < 13 years, and so on) into discrete distributions. Furthermore, there is a demand to device new reference frameworks that enable quantitative precision medicine and integrate with AI approaches to improve clinical investigations and patient treatment strategies. Although growth curves were implemented decades ago and have become standardized tools in pediatrics, the potential for more advanced clinical utilization of this framework remains unexplored, particularly with respect to nonanthropometric variables. Although there are several examples of AI tools implemented to enhance radiologic predictions of disease and bone age in pediatrics (40, 41) and biochemistry profiles for hematologic disease in adults (42), equivalent progress has not been made to enable computer-aided diagnosis in pediatric endocrinology.

The current study applied the conventional "LMS" growth chart framework to model reference curves for 17 biomarkers in the pediatric population. Serum levels of several of these biomarkers increase by powers of 10 throughout puberty and are challenging to resolve in age-partitioned reference intervals. Precedence for using the semiparametric LMS algorithm to model anthropometric parameters of pediatric development is provided in the official WHO, CDC, and national growth charts worldwide. Furthermore, some previous publications demonstrate the successful application of this framework applied to steroid sex hormones (43, 44) and the biomarker IGF1 (45, 46). Importantly, disclosing the L, M, and S parameters enables health personnel elsewhere to implement the relevant reference curves and calculate z-scores according to their patients' gender and age. Reference curves in the current study describe the gender-specific and age-dependent variation of 17 different biomarkers, and these have been made available in Supplemental Table 1 (34).

The references presented in the current manuscript were generated from a population sample representative of the general Norwegian demography, corresponding to approximately 89% Caucasian, 6% Asian, 3% African, and 0.5% Hispanic according to the latest census update (47). Although our use of LC-MS/MS and common mainstream commercial instruments to quantify biomarkers may provide more robust generalization to other laboratories, the current references should be interpreted with caution and validated prior to clinical implementation elsewhere, and also with respect to other ethnicities.

Diurnal variation was not accounted for in the current reference curves but decreasing in hormone levels throughout the day should be considered in clinical practice, particularly with respect to cortisol and testosterone where we observed significantly higher hormone levels in morning samples (before 10:00) than afternoon samples (after 10:00) in pubertal children. Statistical tests were performed according to CLSI guidelines to determine whether stratification of reference ranges according to time of blood draw was warranted, and appropriately stratified nonparametric reference intervals accounting for sample time of day in the current study are provided in the Supplemental Table 2 (37). Further, cyclical hormone variation should be considered when sampling gonadotropins, estradiol, 17-hydroxyprogesterone, and

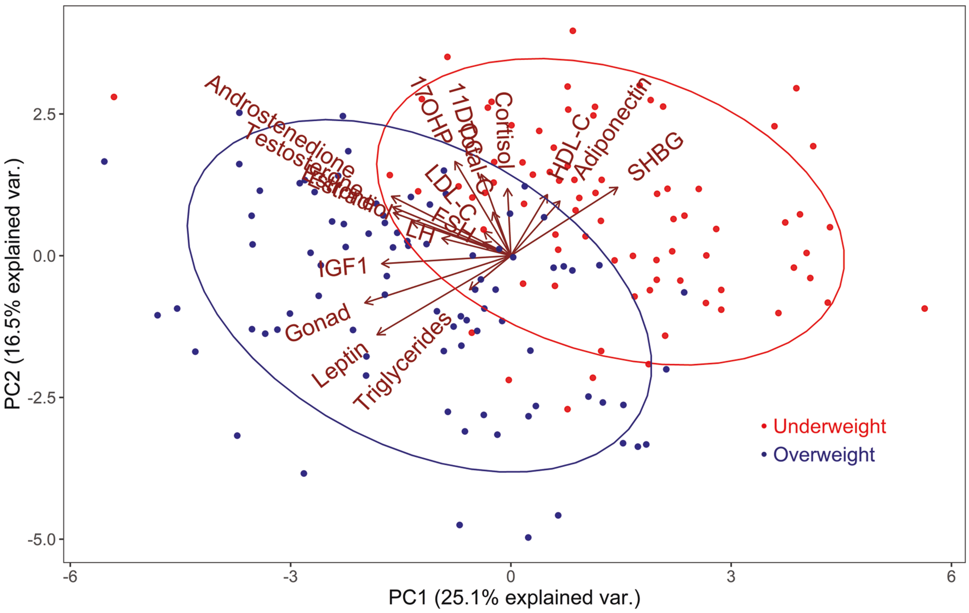


**Figure 4.** Association between biomarker levels and weight class. Dimension reduction by principal component analysis (PCA) was applied to 17 biomarkers and puberty status in terms of testicular volume or glandular tissue volume in 154 underweight (BMI-SDS ≤ –1.0) and 140 overweight (BMI-SDS ≥ 1.0) boys and girls. Directional contribution of individual variables to dataset variance is shown in the biplot in relation to clusters for underweight (red dots) and overweight (blue dots) BMI weight classes. The 1.5 SD confidence ellipses define each weight class cluster in terms of the dataset variance.

**Table 3. Classification of BMI weight class by applying machine learning to the biomarker profile**

| | | Reference | |
|---|---|---|---|
| | | Underweight | Overweight |
| **Prediction** | Underweight | 37 | 3 |
| | Overweight | 1 | 32 |

Biomarker z-scores from 294 underweight (BMI z-score ≤ –1.0) and 'overweight' (BMI z-score ≥ 1.0) children were included in the analysis, and the random forest decision tree classification model was trained using 75% of the data prior to prediction of BMI weight class in the remaining 25% unseen data shown in the current confusion matrix. Classification performance of the ML model exceeded that of any individual biomarkers of BMI weight class. A satisfactory measure of classification agreement was estimated for the ML model: Cohen's kappa of 0.89 (95% CI, 0.74-0.89).

4-androstenedione in girls. For girls exhibiting a regular menstruation cycle in the Fit Futures cohort, we were able to partition reference intervals by both menstrual cycle week and use of oral contraceptives, and for menarcheal patients we therefore recommend consulting the reference limits provided as supplemental information.

The clinical utility of anthropometric growth charts is integral in pediatric practice, and we propose that this statistical framework is also applicable to biomarkers in pediatric endocrinology. While nonparametric hormone references represent clinical cutoff values and primarily provide a qualitative indication as to whether the patient is within or outside the reference interval, LMS models enable quantitative benchmarking and longitudinal tracking of patients' biomarker levels in terms of z-scores. This feature was recently leveraged to quantify endocrine abnormalities in a pediatric cohort that were exposed to metformin in utero due to treatment of maternal polycystic ovary syndrome (22). Applying reference curves to monitor individual patients over time may also be useful for evaluating pediatric endocrinopathies, differences of sex development, and general follow-up. In this respect, the steroid hormone 11-deoxycortisol is an integral biomarker of congenital adrenal hyperplasia due to 11-hydroxylase deficiency, and a general biomarker of virilization, hirsutism, and further used in other diagnostic contexts of suspected Cushing disease or adrenal insufficiency (48). We were unfortunately not able to quantify levels of dehydroepiandrosterone (DHEA) or DHEA sulfate (DHEAS) in the current study.

Biomarker z-scores are subject to the same considerations of biological ($CV_I$) and analytical ($CV_A$) variation as results denoted in absolute concentration units. The "critical difference" threshold obtained by calculating reference change value (RCV) by the classical formula [$RCV = 2^{1/2} \times Z \times (CV_A^2 + CV_I^2)^{1/2}$] defines whether or not a new sample result (eg, during longitudinal tracking) should be regarded as a significant patient change. The "critical difference" obtained by multiplying the absolute value of the previous sample with the RCV percentage can be readily converted to equivalent z-scores using the LMS formula outlined in the "Methods" section.

By application of the current reference curves and calculation of multiple z-scores for each cohort participant, we were able to parameterize a quantitative profile of biochemical and anthropometric measures that may enable personalized medicine. The standardized correlations between biomarker and anthropometric z-scores in Fig. 3 are novel and meaningful, because it is otherwise not feasible to obtain the absolute unit correlation between 2 variables (eg, testicular volume measured in mL, and serum LH measured in IU/L) that are both confounded by age-dependent and nonlinear variation. Similarly, with regard to Table 2, the observed differences in biomarker z-scores according to pubertal phenotypes are not attributable to variation in age. Results from Table 2 show that boys and girls exhibiting pubarche without testicular maturation or breast development had significantly lower LH-for-age compared to children with established pubertal onset. Surprisingly, no significant differences in adrenal steroid hormones were observed between any of the groups. The gender-specific beta correlation matrices in Fig. 3 provide quantitative adjustments to be applied to patient endpoints during a clinical investigation. The standardized beta coefficients featured in Fig. 3 warrant some further discussion. First, the association between weight class and early puberty timing is well established, particularly with respect to female development (2). The results in Fig. 3 provide a quantitative measure of the association between BMI and female glandular tissue volume (β = 0.5; $P < 0.001$). With regard to boys, we observed only a minor standardized association between BMI and male testicular volume (β = 0.2; $P < 0.001$). Irrespective of age and gender, BMI was negatively associated with SHBG (β = −0.4; $P < 0.001$). In clinical practice, the results shown in Fig. 3 can be used as follows: since a 1 SD increase in BMI z-score associates with a 0.4 decline in SHBG z-score, blood sample results of underweight or overweight persons can be calibrated according to the patient's weight class. Conversely, weight class and adiposity may be regarded as significant covariates of SHBG levels in children. Our reference beta coefficient matrices should enable clinicians to implement adjustments for gender, age, BMI, and other clinically relevant features when evaluating patient biomarker levels. We propose that applying such quantitative adjustments for key clinical covariates is an effective practice of personalized precision medicine in pediatric endocrinology.

Lastly, we visualized systemic differences in the biomarker profile according to BMI weight classes and demonstrated a use-case for leveraging the biomarker z-scores interpolated from the current reference curves in ML classification. This analysis sought to examine whether the endocrine profiles (comprising all biomarkers featured in Figs. 1 and 2, total cholesterol, HDL cholesterol, LDL cholesterol, and triglycerides) associated with "overweight" and "underweight" BMI weight classes could be resolved by a machine learning model. Importantly, using normalized biomarker z-scores adjusted for age and gender, it was feasible to combine girls and boys of all ages in this analysis, whereas this would not be the case with biomarkers denoted in absolute concentration units. Notably, adiponectin and leptin are adipocyte-derived adipokines that modulate whole-body energy balance and exhibit profoundly dysfunctional signaling in obese and insulin resistant individuals (49). Peripheral insulin resistance and hyperinsulinemia further dysregulate circulating cholesterol composition and hepatic synthesis of SHBG and IGF1 (50, 51). Results from the current PCA demonstrated that children with mild overweight (BMI z-score ≥ 1) exhibit a materially altered biomarker profile compared with underweight (BMI z-score ≤ −1) peers. Moreover, the corresponding PCA biplot demonstrated that more advanced pubertal characteristics for age was a defining feature of overweight children. Supervised machine learning is an AI method to produce an inferred function from labeled data (eg, using several feature variables to explain a known phenotype dichotomy) in order to computerize classification of new cases. The current "random forest" classification model to infer BMI weight class by biomarker profile outperformed the predictive values of any individual biomarkers. Although the population samples in the current study included only healthy children, we propose that supervised training of such classification models may provide useful clinical tools to diagnose and manage pediatric diseases, unfavorable metabolic profiles, and endocrinopathies.

In conclusion, the LMS framework was used to configure reference charts for 17 circulating biomarkers, by which patients may be benchmarked in terms of age- and sex-adjusted equivalent z-scores. Differential attainment of pubic hair and/or gonadarche during the puberty onset age window was associated with distinct differences for both anthropometric and biomarker z-scores. Finally, we compiled a comprehensive association map of clinical variables and demonstrate high-accuracy machine-aided classification of a clinical dichotomy only by evaluating the biomarker z-score profile.

## Acknowledgments

The authors wish to thank the parents and participants of the BGS2 and Fit Futures cohorts. The study was funded by the Western Norway Regional Health Authority (grant no. 911975, recipient: P. Juliusson and grant no. 912221, recipient: Andre Madsen) and the Laboratory Medicine and Pathology, Haukeland University Hospital (recipient: Hormone Laboratory by Bjørg Almås (Fit Futures) and Jørn V. Sagen (BGS2)).

## Grants

This study was funded by the Western Norway Regional Health Authority (grants no. 911975 and 912221) and internal funding from Laboratory Medicine and Pathology, Haukeland University Hospital.

## Disclosures

The authors have nothing to disclose.

## Data Availability

Restrictions apply to the availability of data generated or analyzed during this study to preserve patient confidentiality or because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided.

## References

1. Eckert-Lind C, Busch AS, Petersen JH, *et al*. Worldwide secular trends in age at pubertal onset assessed by breast development among girls: A systematic review and meta-analysis. *JAMA Pediatr*. 2020;174(4):e195881.
2. Li W, Liu Q, Deng X, Chen Y, Liu S, Story M. Association between Obesity and Puberty Timing: A Systematic Review and Meta-Analysis. *Int J Environ Res Public Health*. 2017;14(10):1266. doi:10.3390/ijerph14101266
3. Brix N, Ernst A, Lauridsen LLB, *et al*. Childhood overweight and obesity and timing of puberty in boys and girls: cohort and sibling-matched analyses. *Int J Epidemiol*. 2020;49(3):834-844.
4. Rosenfield RL. Normal and premature adrenarche. *Endocr Rev*. 2021;42(6):783-814.
5. Rustad P, Felding P, Lahti A, Hyltoft Petersen P. Descriptive analytical data and consequences for calculation of common reference intervals in the Nordic Reference Interval Project 2000. *Scand J Clin Lab Invest*. 2004;64(4):343-370.
6. Wayne, PA. *Ep28-a3c - Defining, establishing, and verifying reference intervals in the clinical laboratory*.

7. Loh TP, Antoniou G, Baghurst P, Metz MP. Development of paediatric biochemistry centile charts as a complement to laboratory reference intervals. *Pathology.* 2014;46(4):336-343.
8. Zierk J, Arzideh F, Rechenauer T, *et al*. Age- and sex-specific dynamics in 22 hematologic and biochemical analytes from birth to adolescence. *Clin Chem.* 2015;61(7):964-973.
9. Hoq M, Matthews S, Karlaftis V, *et al*. Reference values for 30 common biochemistry analytes across 5 different analyzers in neonates and children 30 days to 18 years of age. *Clin Chem.* 2019;65(10):1317-1326.
10. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. *Crit Rev Clin Lab Sci.* 2017;54(6):358-413.
11. Hilsted L, Rustad P, Aksglaede L, Sorensen K, Juul A. Recommended Nordic paediatric reference intervals for 21 common biochemical properties. *Scand J Clin Lab Invest.* 2013;73(1):1-9.
12. Asgari S, Higgins V, McCudden C, Adeli K. Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: comparisons to traditionally partitioned reference intervals. *Clin Biochem.* 2019;73:82-89.
13. Cole TJ. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr.* 1990;44(1):45-60.
14. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med.* 1992;11(10):1305-1319.
15. de Onis M, Onyango AW, Borghi E, Siyam A, Nishida C, Siekmann J. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ.* 2007;85(9):660-667.
16. Flegal KM, Cole TJ. Construction of LMS parameters for the Centers for Disease Control and Prevention 2000 growth charts. *Natl Health Stat Report.* 2013(63):1-3.
17. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak.* 2019;19(1):281.
18. Bruserud IS, Roelants M, Oehme NHB, *et al*. References for ultrasound staging of breast maturation, Tanner breast staging, pubic hair, and menarche in Norwegian girls. *J Clin Endocrinol Metab.* 2020;105(5):1599-1607.
19. Oehme NHB, Roelants M, Saervold Bruserud I, *et al*. Reference data for testicular volume measured with ultrasound and pubic hair in Norwegian boys are comparable with Northern European populations. *Acta Paediatr.* 2020;109(8):1612-1619.
20. Madsen A, Oehme NB, Roelants M, *et al*. Testicular ultrasound to stratify hormone references in a cross-sectional Norwegian study of male puberty. *J Clin Endocrinol Metab.* 2020;105(6):1888-1898.
21. Madsen A, Bruserud IS, Bertelsen BE, *et al*. Hormone References for Ultrasound Breast Staging and Endocrine Profiling to Detect Female Onset of Puberty. *J Clin Endocrinol Metab.* 2020;105(12):e4886-e4895.
22. Hanem LGE, Salvesen O, Madsen A, *et al*. Maternal PCOS status and metformin in pregnancy: Steroid hormones in 5-10 years old children from the PregMet randomized controlled study. *PLoS One.* 2021;16(9):e0257186.
23. Winther A, Dennison E, Ahmed LA, *et al*. The Tromso Study: Fit Futures: a study of Norwegian adolescents' lifestyle and bone health. *Arch Osteoporos.* 2014;9:185. doi:10.1007/s11657-014-0185-0
24. Iordanova Schistad E, Kong XY, Furberg AS, *et al*. A population-based study of inflammatory mechanisms and pain sensitivity. *Pain.* 2020;161(2):338-350.
25. Bruserud IS, Roelants M, Oehme NHB, *et al*. Ultrasound assessment of pubertal breast development in girls: intra- and interobserver agreement. *Pediatr Radiol.* 2018;48(11):1576-1583.
26. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child.* 1969;44(235):291-303.
27. Fugl L, Hagen CP, Mieritz MG, *et al*. Glandular breast tissue volume by magnetic resonance imaging in 100 healthy peripubertal girls: evaluation of clinical Tanner staging. *Pediatr Res.* 2016;80(4):526-530.
28. Calcaterra V, Sampaolo P, Klersy C, *et al*. Utility of breast ultrasonography in the diagnostic work-up of precocious puberty and proposal of a prognostic index for identifying girls with rapidly progressive central precocious puberty. *Ultrasound Obstet Gynecol.* 2009;33(1):85-91.
29. Oehme NHB, Roelants M, Bruserud IS, *et al*. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys - intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatr Radiol.* 2018;48(12):1771-1778.
30. Juliusson PB, Roelants M, Nordal E, *et al*. Growth references for 0-19 year-old Norwegian children for length/height, weight, body mass index and head circumference. *Ann Hum Biol.* 2013;40(3):220-227.
31. Methlie P, Hustad SS, Kellmann R, *et al*. Multisteroid LC-MS/MS assay for glucocorticoids and androgens, and its application in Addison's disease. *Endocr Connect.* 2013;2(3):125-136.
32. Bertelsen BE, Kellmann R, Viste K, *et al*. An ultrasensitive routine LC-MS/MS method for estradiol and estrone in the clinically relevant sub-picomolar range. *J Endocr Soc.* 2020;4(6):bvaa047.
33. Stasinopoulos M. Package 'gamlss'. https://cran.r-project.org/web/packages/gamlss/gamlss.pdf
34. Madsen A, Almås B, Bruserud IS, *et al*. Supplemental Table 1. *Figshare Digital Repository*. Deposited December 9, 2021. https://doi.org/10.6084/m9.figshare.17153336.v1
35. Madsen A. Application of the LMS algorithm in R. Deposited 8 December 2021. https://github.com/andremadsen/LMS-reference-curve
36. Finnegan D. Package 'referenceIntervals'. https://cran.r-project.org/web/packages/referenceIntervals/referenceIntervals.pdf. Accessed September 5, 2021.
37. Madsen A, Almås B, Bruserud IS, *et al*. Supplemental Table 2. *Figshare Digital Repository*. Deposited 9 December 2021. https://doi.org/10.6084/m9.figshare.17153369.v4
38. Madsen A, Almås B, Bruserud IS, *et al*. Supplemental table 3. *Figshare Digital Repository*. Deposited 9 December 2021. https://doi.org/10.6084/m9.figshare.17153378.v1
39. Madsen A. Random forest ML classification model in R. https://github.com/andremadsen/Random-Forest-ML
40. Offiah AC. Current and emerging artificial intelligence applications for pediatric musculoskeletal radiology. *Pediatr Radiol.* 2021.
41. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging.* 2009;28(1):52-66.
42. Guncar G, Kukar M, Notar M, *et al*. An application of machine learning to haematological diagnosis. *Sci Rep.* 2018;8(1):411.
43. Frederiksen H, Johannsen TH, Andersen SE, *et al*. Sex-specific Estrogen Levels and Reference Intervals from Infancy to Late Adulthood Determined by LC-MS/MS. *J Clin Endocrinol Metab.* 2020;105(3):754-768.
44. Soeborg T, Frederiksen H, Mouritsen A, *et al*. Sex, age, pubertal development and use of oral contraceptives in relation to serum concentrations of DHEA, DHEAS, 17alpha-hydroxyprogesterone, Delta4-androstenedione, testosterone and their ratios in children, adolescents and young adults. *Clin Chim Acta.* 2014;437:6-13. doi:10.1016/j.cca.2014.06.018
45. 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.
46. Isojima T, Shimatsu A, Yokoya S, *et al*. Standardized centile curves and reference intervals of serum insulin-like growth factor-I (IGF-I) levels in a normal Japanese population using the LMS method. *Endocr J.* 2012;59(9):771-780.
47. Statistics Norway. Immigrants and Norwegian-born to immigrant parents. 2021. www.ssb.no/en/befolkning/innvandrere/statistikk/innvandrere-og-norskfodte-med-innvandrerforeldre. Accessed October 20, 2021.
48. Tonetto-Fernandes V, Lemos-Marini SH, Kuperman H, Ribeiro-Neto LM, Verreschi IT, Kater CE. Serum 21-deoxycortisol,

17-hydroxyprogesterone, and 11-deoxycortisol in classic congenital adrenal hyperplasia: clinical and hormonal correlations and identification of patients with 11beta-hydroxylase deficiency among a large group with alleged 21-hydroxylase deficiency. *J Clin Endocrinol Metab*. 2006;91(6):2179-2184.

49. Landecho MF, Tuero C, Valenti V, Bilbao I, de la Higuera M, Fruhbeck G. Relevance of Leptin and Other Adipokines in Obesity-Associated Cardiovascular Risk. *Nutrients*. 2019;11(11):2664. doi:10.3390/nu11112664
50. Wallace IR, McKinley MC, Bell PM, Hunter SJ. Sex hormone binding globulin and insulin resistance. *Clin Endocrinol (Oxf)*. 2013;78(3):321-329.
51. Ormazabal V, Nair S, Elfeky O, Aguayo C, Salomon C, Zuniga FA. Association between insulin resistance and the development of cardiovascular disease. *Cardiovasc Diabetol*. 2018;17(1):122.