UZ LEUVEN | LABORATORIUMGENEESKUNDE

CAT

Critically Appraised Topic

Title: Review of pediatric reference intervals for testosterone, dihydrotestosterone, and testosterone/dihydrotestosterone ratio by mass spectrometry: need for harmonization

Author: Kobe Truijens

Supervisors: Prof. James C. Doery, Prof Zhong X. Lu

Search/methodology verified by: Prof. James C. Doery, Prof Zhong X. Lu

Date: December 2024

Introduction

Accurate measurement of testosterone (T), dihydrotestosterone (DHT), and T/DHT ratio is important for diagnosing multiple pediatric endocrine disorders. Mass spectrometry (MS) provides superior sensitivity and potential standardization for steroid analysis, yet harmonized age-specific reference intervals remain lacking, complicating clinical decision-making.

Methods

We conducted a systematic review of studies (published 1980–2024) on reference intervals for T, DHT, and T/DHT ratio by MS in individuals aged 0–18 years. An overview of the published RIs, including corresponding partitions, was compiled, and graphical representations were made to assess potential differences and heterogeneity. To evaluate limitations of the RIs, all data relevant to the establishment of the RIs was extracted from the selected articles. Additionally, a 10-year retrospective analysis of T and DHT data measured by LC-MS/MS was performed, comparing results with published RIs to evaluate abnormal flagging rates and variability in clinical interpretation.

Results

Based on the selection criteria, 14 articles were included. RIs for T were reported in all studies, while only three provided RIs for DHT, and none reported RIs for T/DHT ratio. Graphical overviews of the published RIs revealed substantial variability in age and pubertal stage partitions, as well as heterogeneity in the applied age-partitions. Application of real-world data on the published reference intervals showed considerable variability in interpretation depending on the source of the RIs, resulting in inconsistencies in clinical interpretation.

Conclusion

This review highlights limitations in current RIs for T and DHT, including variability in lower and upper limits and inconsistent age partitions. These issues stem from factors such as small sample sizes, opportunistic sampling, and inadequate consideration of developmental changes like mini-puberty and adrenarche. Establishing RIs that accurately reflect the physiological and biochemical changes during childhood development is essential to improve diagnostic accuracy.

CLINICAL/DIAGNOSTIC SCENARIO

Accurate measurement and interpretation of testosterone (T), dihydrotestosterone (DHT) and other serum androgens are essential in pediatric endocrinology for diagnosing various endocrine disorders, such as congenital adrenal hyperplasia, 5α-reductase deficiency, hypogonadism and delayed or premature puberty (1–3).

The reliable analysis of these steroids has historically been a significant challenge for clinical laboratories (4,5). Fifty years ago, radioimmunoassays were introduced for T analysis (6). Despite their acceptable analytical performance, these assays required time-consuming purification steps. Therefore, automated direct (electro-)chemiluminescent immunoassays have been developed and still are the primary method for analyzing serum steroids due to their practicality and high sample throughput (4,5,7). However, immunoassays often lack specificity and the necessary sensitivity, especially in pediatric and female populations, where T and DHT concentrations are typically very low (4,5,7,8).

Due to the analytical limitations of immunoassays, mass spectrometry (MS) has become the gold standard for steroid analysis, particularly liquid chromatography-tandem mass spectrometry (LC-MS/MS), which offers superior sensitivity, accuracy, and precision compared to other methods (4,7,9). Recent advancements in MS technology have expanded its use from reference laboratories to numerous clinical laboratories worldwide. These improvements enable laboratories to analyze a wide range of steroids with high accuracy and precision across varied concentrations and in large sample volumes (4,7). Unlike immunoassays, LC-MS/MS can detect very low levels of steroids, making it particularly useful for measuring T and DHT in the female and pediatric population (8–10). Additionally, unlike immunoassays which have substantial interassay variability, MS-based assays can be traced to an international standard, enhancing standardization. While MS-based assays have improved the comparability of the interlaboratory results, there is still a need for harmonized age-related reference intervals to ensure accurate interpretation of results (9,11–18). Currently, there are significant discrepancies in the pediatric population, with inconsistencies in age-partitions and variations in reported reference intervals, even when pubertal stages are used for partitioning (18).

While MS-based assays have improved the comparability of the interlaboratory results, there is still a need for harmonized age-related reference intervals to ensure accurate interpretation of results (9,11–18). Currently, T reference intervals are inconsistent due to the use of different assays and methods (9). Establishing reliable pediatric reference intervals is essential to accurately reflect the physiological and biochemical changes that occur during a child's development (15,19). Serum levels of T and DHT vary significantly in the pediatric population. In boys, concentrations are high at birth but rapidly decrease within the first week, becoming nearly undetectable. Around the second to fourth week, concentrations begin to rise again, peaking between one and three months of age. This is followed by a gradual decline to the age of six months, this is called the 'mini-puberty' (3,19,20). In contrast, girls maintain low T levels during infancy (19). Besides the possible effect of adrenarche, these concentrations tend to stay low in both boys and girls until the start of puberty (3,19,20,34).

QUESTION(S)

1. *What reference intervals for testosterone, dihydrotestosterone, and T/DHT ratio, analyzed using mass-spectrometry, are currently available in the literature?*

2. *What are the limitations of the currently published reference intervals for these androgens?*

3. *Do published reference intervals still show heterogeneity and variability when using standardized MS-based assays?*

SEARCH TERMS

PubMed (Medline) database: 1980 – 2024. Search conducted at 29/05/2024.

Search Strategy	
Testosterone	"Testosterone"[Mesh] OR "Testosterone"[tiab]
Dihydrotestosterone	"Dihydrotestosterone"[Mesh] OR "Dihydrotestosterone"[tiab] OR "DHT"[tiab]
Reference Intervals	"Reference Values"[Mesh] OR "Reference Values"[tiab] OR "Reference Value"[tiab] OR "Reference Ranges"[tiab] OR "Reference Range"[tiab] OR "Reference Interval"[tiab] OR "Reference Intervals"[tiab] OR "Reference Limit"[tiab] OR "Limit of Normal"[tiab]
Mass Spectrometry	"Mass Spectrometry"[Mesh] OR "Mass Spectrometry"[tiab] OR "MS"[tiab] OR "liquid chromatography/mass spectrometry"[tiab] OR "LC/MS"[tiab] OR "LC-MS"[tiab]
Final search combination	(Testosterone OR Dihydrotestosterone) AND Reference Intervals AND Mass Spectrometry ("Testosterone"[Mesh] OR "Testosterone"[tiab] OR "Dihydrotestosterone"[Mesh] OR "Dihydrotestosterone"[tiab] OR "DHT"[tiab]) AND ("Reference Values"[Mesh] OR "Reference Values"[tiab] OR "Reference Value"[tiab] OR "Reference Ranges"[tiab] OR "Reference Range"[tiab] OR "Reference Interval"[tiab] OR "Reference Intervals"[tiab] OR "Reference Limit"[tiab] OR "Limit of Normal"[tiab]) AND ("Mass Spectrometry"[Mesh] OR "Mass Spectrometry"[tiab] OR "MS"[tiab] OR "liquid chromatography/mass spectrometry"[tiab] OR "LC/MS"[tiab] OR "LC-MS"[tiab])

Methods

Search Strategy

We conducted a systematic review of the literature for all the publication from 1980 to 2024 on pediatric reference intervals for testosterone, dihydrotestosterone and T/DHT ratio measured via mass spectrometry. The key concepts for our search were "testosterone," "dihydrotestosterone," "reference intervals," and "mass spectrometry." The search was executed in the PubMed database utilizing Medical Subject Headings (MeSH) to ensure the inclusion of relevant studies. Additionally, we supplemented the MeSH terms with relevant synonyms using free text terms to identify articles not yet indexed in PubMed or not covered by our MeSH terms.

Screening and Selection

Articles were screened through a decision cascade based on the relevance of the title followed by the abstract. The inclusion criteria were: (1) reporting of reference intervals, (2) utilization of mass spectrometry-based assays, (3) involvement of T, DHT, or T/DHT ratio, and (4) analysis of plasma and/or serum samples. Articles meeting these criteria proceeded to full manuscript screening. During this phase, we restricted our selection to reference intervals in individuals aged 0-18 years old. This specific criterion was omitted during the title/abstract screening to prevent the exclusion of studies providing comprehensive reference intervals that include but do not explicitly mention pediatric reference intervals. We also examined the reference lists of the studies included to identify additional studies that were not included in our literature search. The article search was limited to articles written in English. The process of study screening and selection was performed by one reviewer only.

Data Extraction

All articles included underwent an in-depth analysis. To create an overview of the available pediatric reference intervals for T, DHT, and T/DHT ratio, all reference intervals were extracted from the selected articles, along with their corresponding partitions and the sample size of each partition. All reference interval data was converted into SI units (i.e. nmol/L) for uniformity. The intervals were categorized by source and organized according to the provided partitions (age groups, pubertal stages, and sex). We extracted all relevant data for establishment of the reference intervals. This included analytical characteristics (e.g. method of analysis/manufacturer, calibrator traceability), characteristics of the reference population (e.g. age and sex composition, in- and exclusion criteria, ethnicity), methodological characteristics (partitions, partitioning criteria, size of the confidence interval used to establish RIs), and cohort size.

Data Analysis and Synthesis

The results of our systematic review are presented descriptively, featuring a summary of the study selection process and characteristics, as well as a tabular overview of the reported reference intervals. To highlight potential differences and heterogeneity, the lower-limit and upper-limit of the extracted reference intervals were presented as graphical representations sorted by age, sex and Tanner stage. All analyses were performed in Microsoft Excel (Microsoft, Redmond, WA, USA) and R (http://www.r-project.org/).

Real World Data Comparison

To compare the extracted reference intervals with real-world data, we analyzed 10 years of retrospective T and DHT data by LC-MS/MS from children aged 0-18 years. This retrospective dataset was obtained at a tertiary clinical laboratory (Monash Health Pathology, Melbourne, Australia) and covered the entire period where their analytical method remained stable. The data extraction included age, sex and test results for T and DHT. This retrospective data analysis adhered to the hospital's ethical guidelines and was approved by the ethical board.

To ensure a reliable analysis, we minimized the inclusion of pathological data by removing outliers and irrelevant points through a data-cleaning process. As the dataset was anonymized, we had limited clinical information to exclude samples based on clinical criteria. Therefore, we restricted the analysis to 'solo' samples to exclude potentially diseased individuals who are more likely to undergo retesting. Additionally, we used Tukey's far-outlier detection method to eliminate overtly abnormal results.

To assess the published reference intervals, we applied the lower and upper limits obtained from the different articles against our retrospective dataset. This approach allowed us to determine abnormal flagging rates to evaluate the reference intervals from the studies included.

Results

Study Selection

Our search strategy initially identified 201 articles. Screening by title and abstract reduced the selection to 53 articles. The full manuscripts of these 53 articles were assessed for eligibility, leading to the exclusion of 40 articles for various reasons including no reference intervals were available or no reference intervals were available for children (n = 33), analysis focused on specific age groups (adolescents partitioned by menstrual cycle, premature neonates) (n = 2), and the analyses were on dried blood spots (n=5), not serum or plasma. This resulted in a final selection of 13 articles (1,21–32). Additionally, screening the reference lists of these articles and related sources yielded one more relevant article (33), resulting in a final selection of 14 articles included in the literature review. A study selection flowchart of our literature search is available in Figure 1.

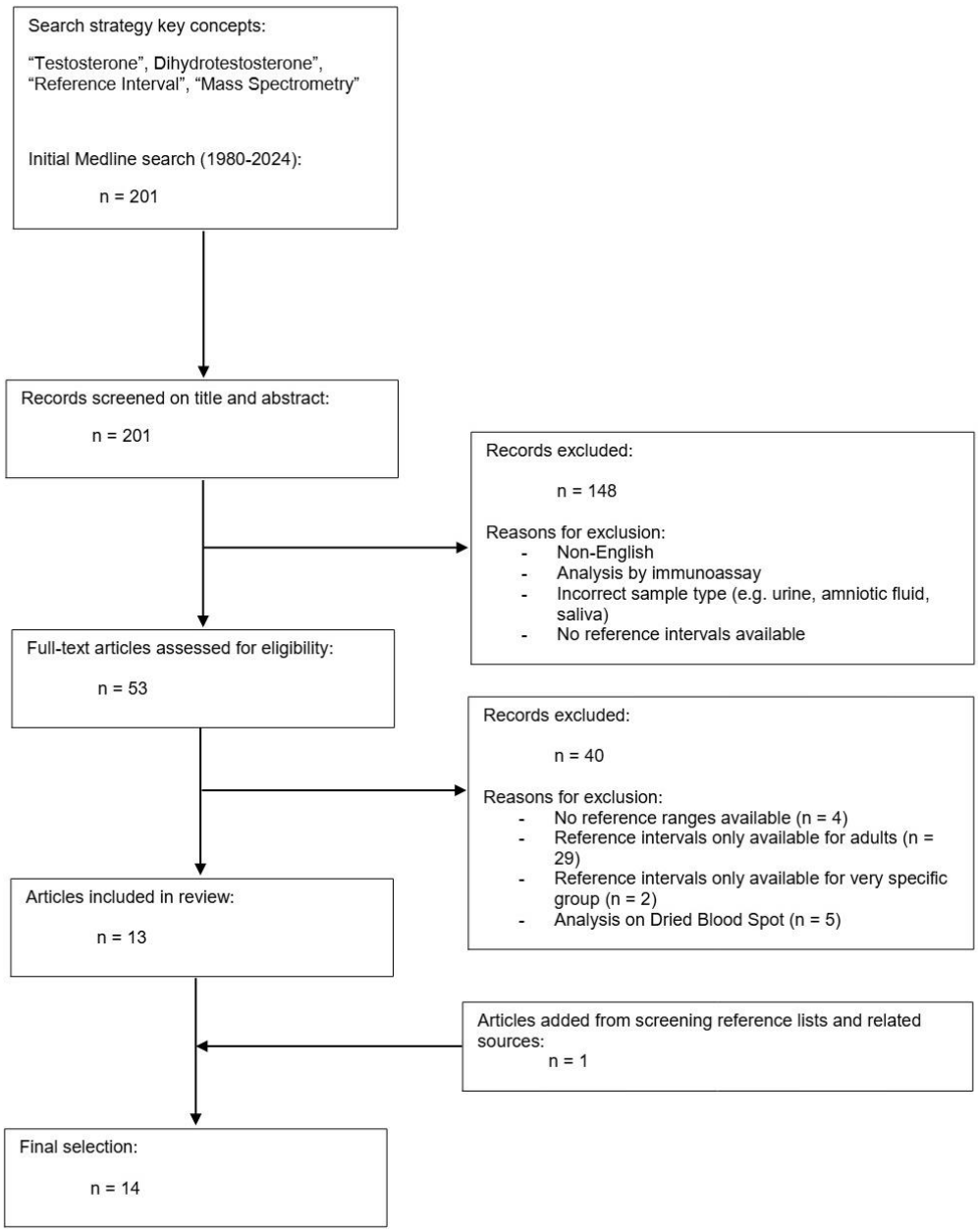

Figure 1. Study selection flowchart

Study Characteristics

An overview of the final selected studies and their characteristics are listed in Supplemental Table 1. Testosterone reference intervals were reported in all 14 articles reviewed, while DHT reference intervals were documented in only three studies (21,25,33). None of the articles provided reference intervals for T/DHT ratio. Of the 14 included studies, 12 utilized LC-MS/MS for their analyses, while the remaining two used GC-MS/MS (25,30). Serum samples were used in 13 studies, with only one study using plasma samples (33). Thirteen of the 14 articles reported reference intervals for both boys and girls. The remaining article focused solely on boys (23). Age was used as a partitioning criterion in 12 articles (1,21–24,27–33), nine articles reported intervals based on pubertal stage (21,23–28,32,33). Pubertal stages were evaluated using various methods, including Tanner pubic hair stages (23,26), Tanner breast stages (25,26), Tanner genital stages (26) or unspecified Tanner stages (24,27,28,32,33). Some studies assessed pubertal stages through testicular volume (23,25,26,33), and one study used age-partitioned pubertal stages (21). While all studies partitioned reference intervals by sex, some combined intervals due to small sample sizes or a lack of statistically significant gender differences in prepubertal children (21,30). Only two of the 14 studies mentioned that their calibrator could be traced to an international standard (NIST 971) (22,24), while the rest only specified the supplier of their calibrator materials. Ethnicity of the study populations was reported in only 4 studies (1,21,23,27).

Reference Intervals

A comprehensive overview of all extracted reference intervals sorted by author and subgroup can be found in Supplemental Table 2-5. We included a total of 252 reference intervals for testosterone and 52 for DHT. The lower limit of testosterone ranged from the limit of quantification (LOQ) up to 11.1 nmol/L for men and up to 1 nmol/L for women, while the upper limit ranged from 0.2 nmol/L to 44.7 nmol/L for men and 0.24 nmol/L to 2.7 nmol/L for women. For DHT, the lower limit ranged from <0.027 nmol/L to 0.646 nmol/L for men and <0.027 nmol/L to 0.1 nmol/L for women, while the upper limit ranged from 0.227 nmol/L to 3.4 nmol/L for men and 0.1 nmol/L to 1 nmol/L for women. The age partitions used in the studies varied widely, ranging from intervals as short as 6 days (9-15 days old) and as broad as 10 years (0-10 years old) (21,30). While most studies established reference intervals using the 95% confidence interval, some reported only the observed range of testosterone or DHT as the reference interval (21,31,33). One study applied the p5-p95 interval (25). Focusing on a specific age and sex, e.g. a 12-year-old boy, the upper-limit for testosterone ranges from 1.12 nmol/L to 25.9 nmol/L (1,32). As previously mentioned, reference intervals partitioned by pubertal stage were mainly evaluated by Tanner stage. In most cases, pubertal stages were evaluated by a qualified physician, in one study Tanner stage classification was done by self-reporting (28).

A graphical representation of the lower- and upper-limit of all age-partitioned reference intervals can be found in Figures 2-4. These figures illustrate the well-established evolution of T and DHT during childhood development, with low levels before puberty and increasing androgen concentrations during puberty. Additionally, some studies include the phenomenon of mini-puberty, visible in the figures as elevated upper limits during the first few months of life. Although a distinct pattern can be observed, there is a relatively wide spread in the lower and upper limits across different age groups. This variability is most pronounced during infancy and puberty. Figures 5–6 provide a graphical representation of the reference intervals categorized by Tanner stage. Due to heterogeneity in cut-offs in testicular volume cut-offs, reference intervals partitioned by testicular volume were not included for comparison. These figures display the published lower and upper limits from all selected articles for each Tanner stage; if only

the upper limit was reported, it is represented as a dot. Compared to the age-partitioned reference intervals, these figures show a less distinct pattern of androgen concentration evolution in relation to pubertal development.

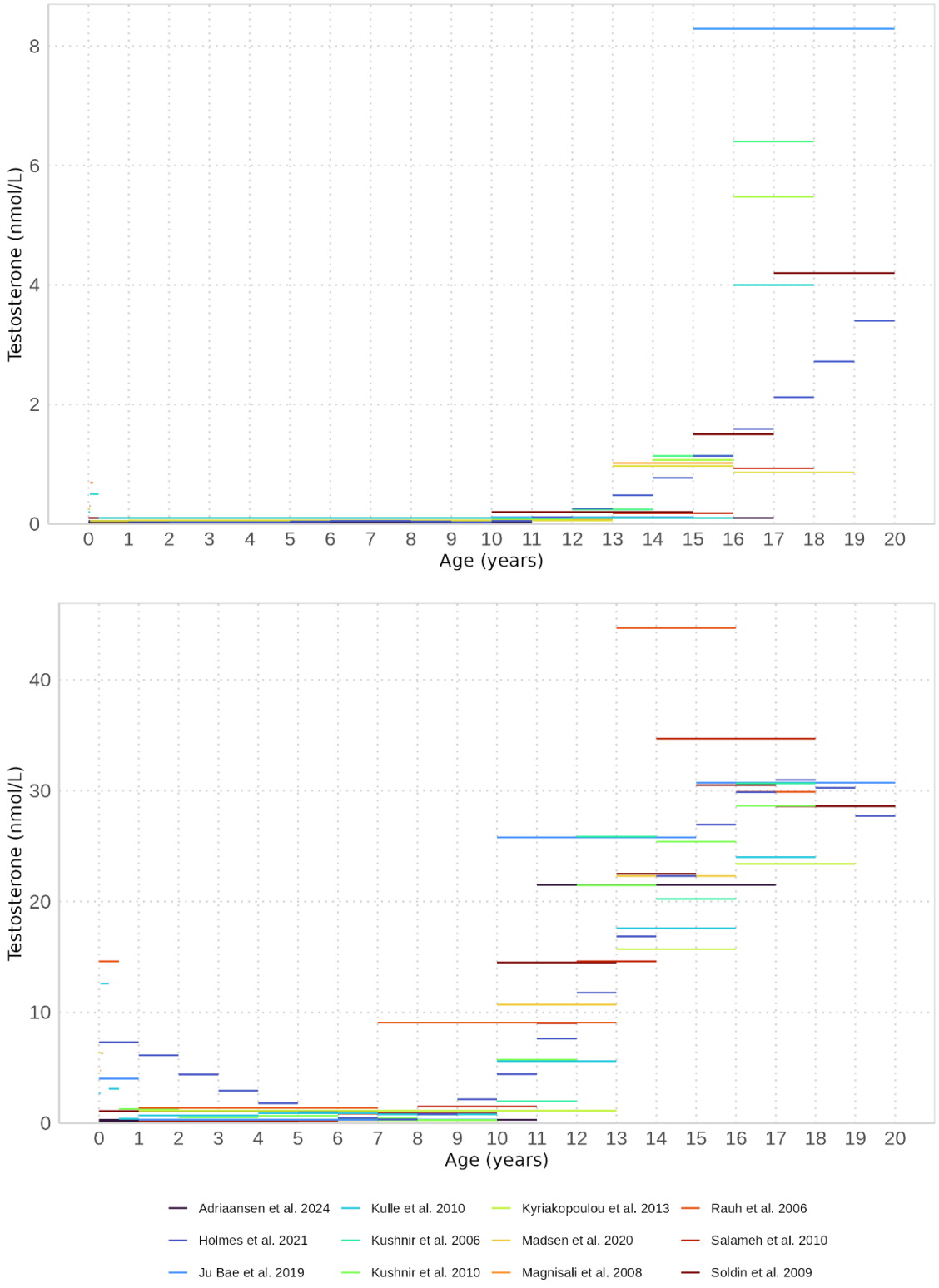

Figure 2. Summary of Pediatric Published Age-Partitioned Testosterone Reference Intervals for males (A: lower-limit; B: upper-limit)

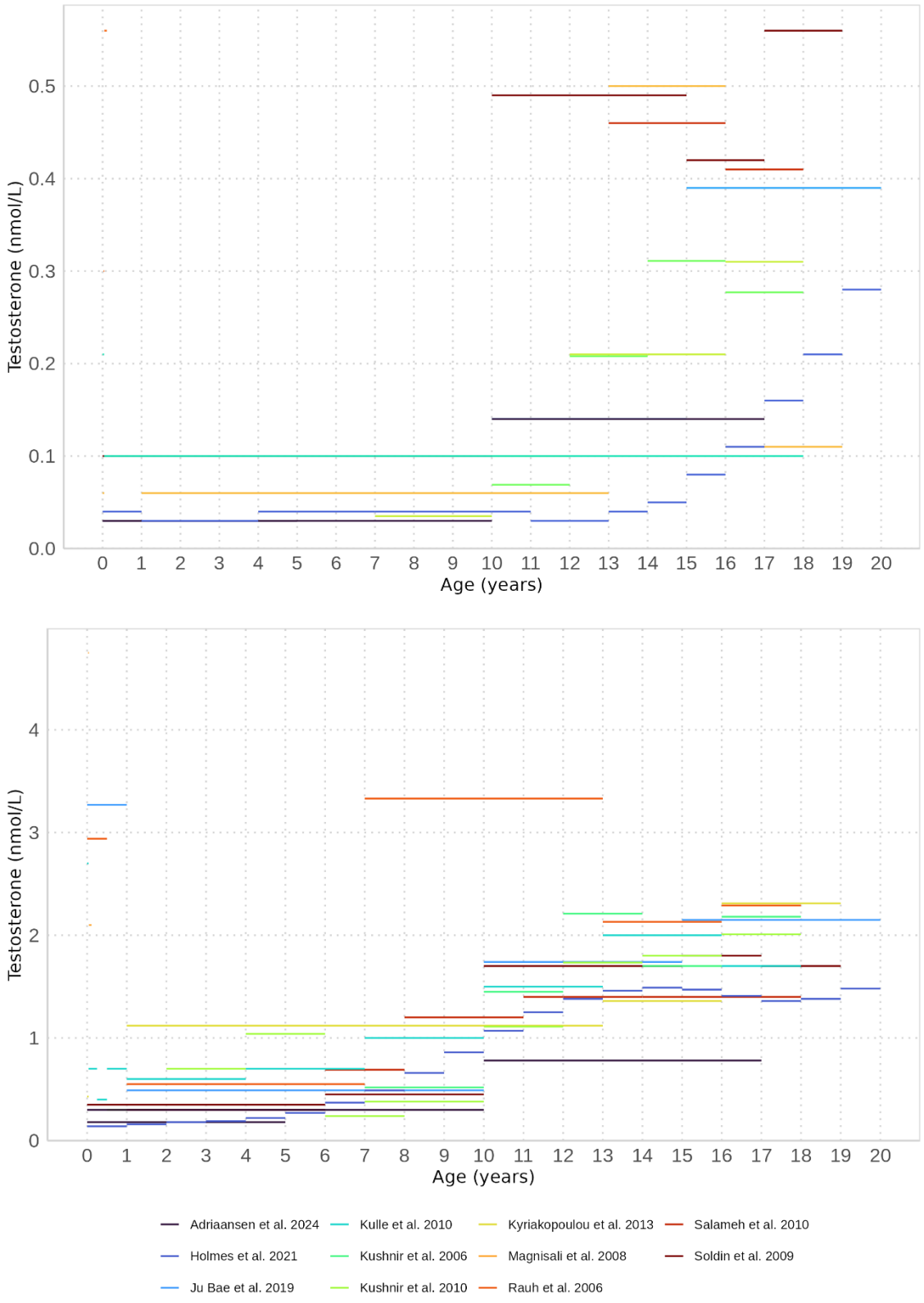

Figure 3. Summary of Pediatric Published Age-Partitioned Testosterone Reference Intervals for females (A: lower-limit; B: upper-limit)

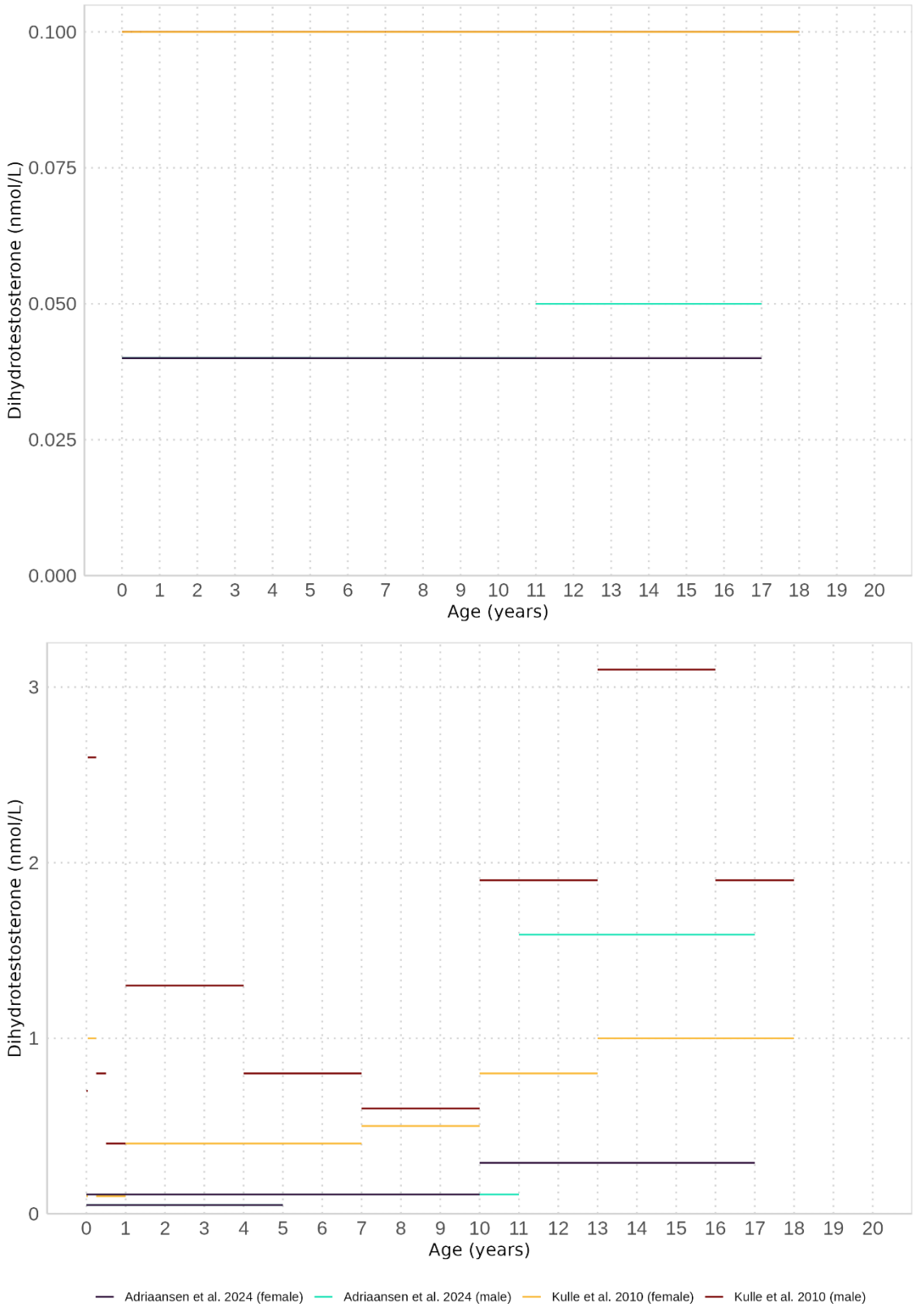

Figure 4. Summary of Pediatric Published Age-Partitioned Dihydrotestosterone Reference Intervals for males and females (A: lower-limit; B: upper-limit)

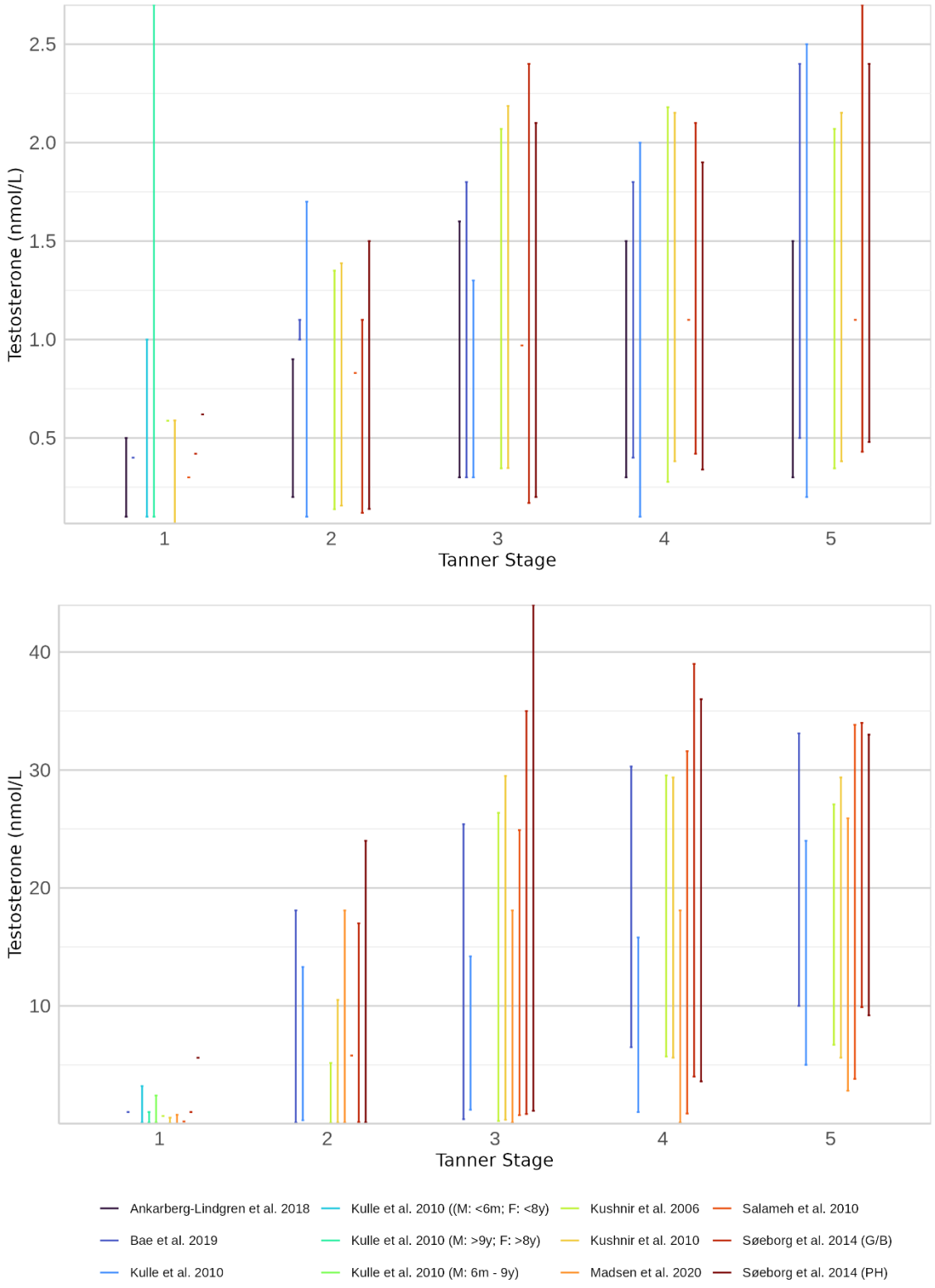

Figure 5. Summary of Published Testosterone Reference Intervals Partitioned by Tanner stage (A: female; B: male) (If only an upper-limit was available for a certain reference it is visualized as a dot)

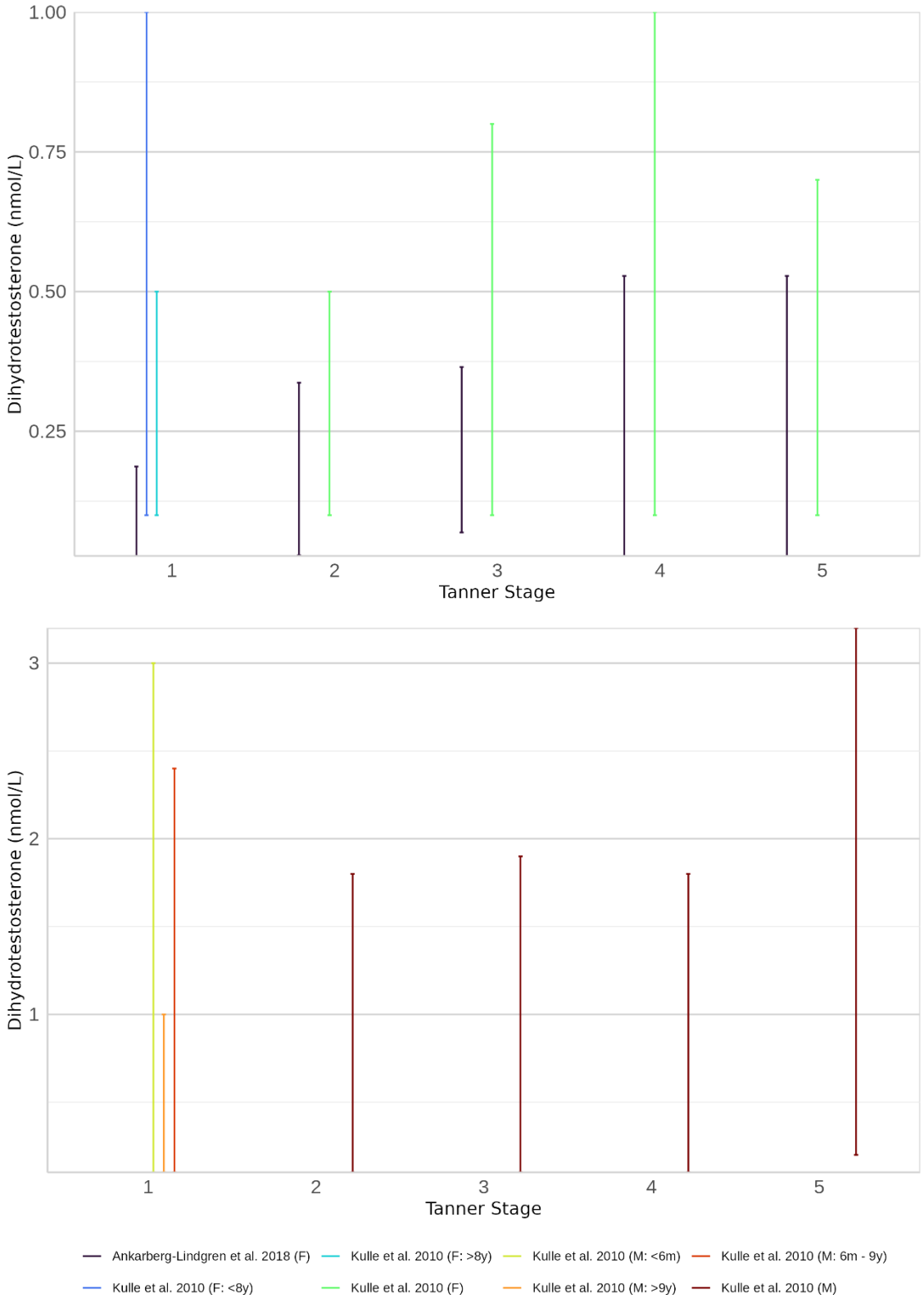

Figure 6. Summary of Published Dihydrotestosterone Reference Intervals Partitioned by Tanner stage (A: female; B: male)

Application of Reference Intervals on Real World Data

To verify the published reference intervals, we conducted a 10-year retrospective data analysis at our institute (Monash Health Pathology). The results of this analysis are presented in Figures 7 and 8. These figures plot testosterone concentrations for male and female children aged 0–18 years against a selection of the published reference intervals, highlighting variability in interpretation.

These figures illustrate that the interpretation of testosterone results can differ significantly depending on the reference intervals used. For instance, a 12-year-old boy with a testosterone level of 13 nmol/L would be considered within the normal range by three of the selected articles but pathological by two. Across all articles publishing age-partitioned reference intervals for boys, the abnormal flagging rate for this value is 33%, meaning four out of 12 articles would assess this result as abnormal, while eight would consider it normal.

Clinically, the most significant abnormality in pubertal boys is the interpretation of decreased testosterone levels in the diagnostic work-up of hypogonadism. Based on our figures we can observe that the heterogeneity in lower limit thresholds is also apparent, leading to variability in interpretation. For instance, a 16-year-old boy with a testosterone level of 3.5 nmol/L would be flagged as too low by 44% of the published reference intervals, while 56% would interpret it as normal. In contrast, for young girls, the primary concern is pathologically elevated testosterone levels. A 15-year-old girl with a testosterone level of 1.5 nmol/L would be flagged as abnormally high by 40% of the published reference intervals but regarded as normal by 60%. These examples emphasize how variability in published reference intervals can impact clinical interpretation.

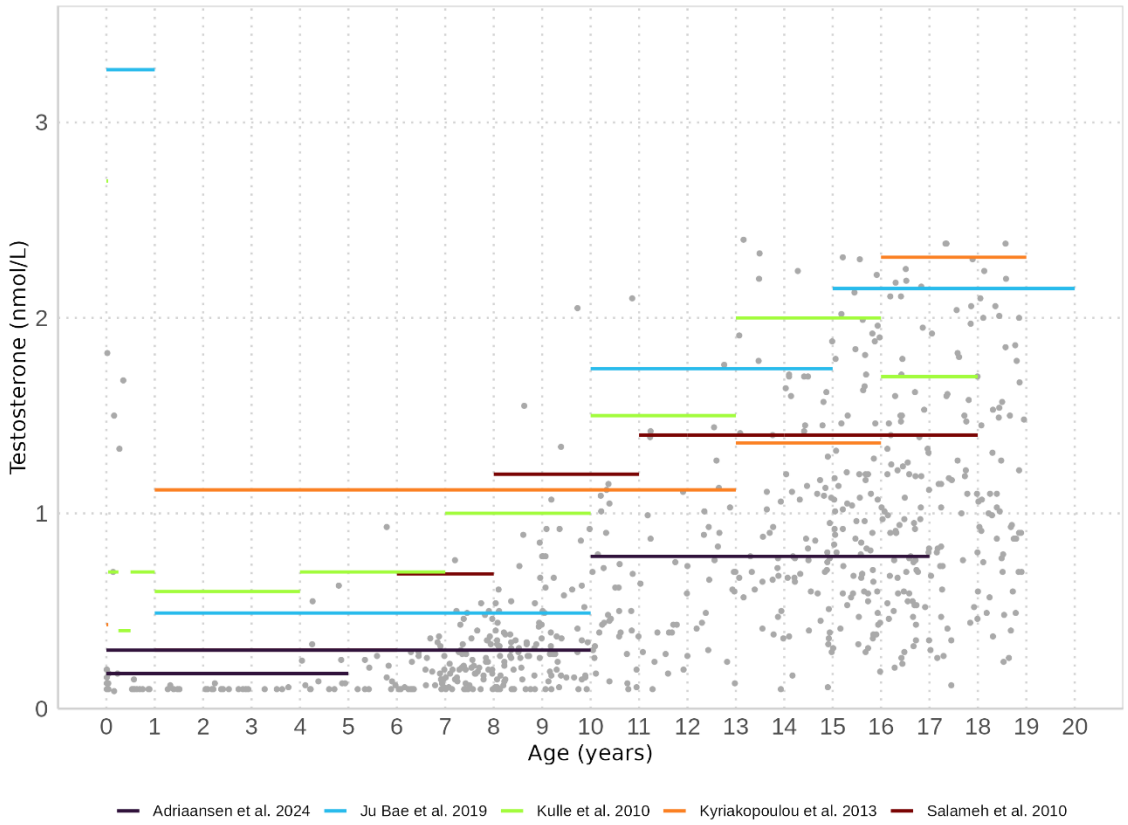

Figure 7. Application of Selected Age-Partitioned Upper Limits on Real-World Data (female)

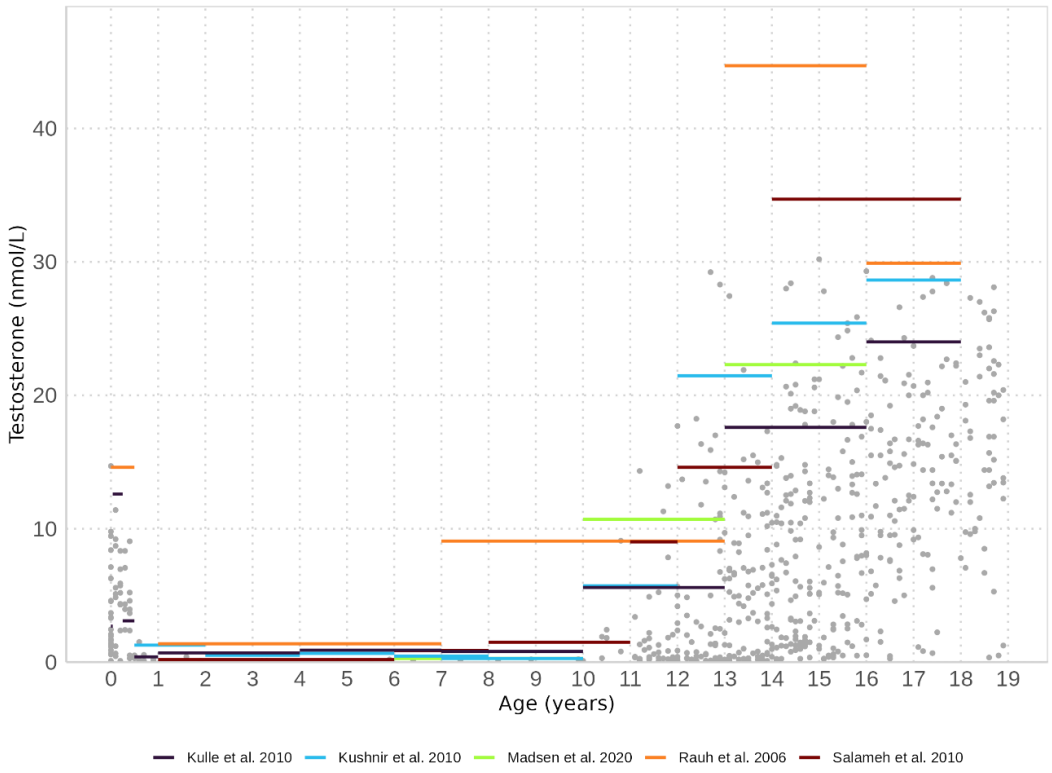

Figure 8. Application of Selected Age-Partitioned Upper Limits on Real-World Data (male)

Discussion

This literature review identified several shortcomings in the published reference intervals for testosterone and dihydrotestosterone. A graphical overview of these intervals revealed wide variation in both the upper and lower limits, as well as inconsistencies in the age partitions, with some subgroups showing widely differing cut-offs. This variability was further emphasized by applying the RIs to real-world data, which demonstrated significant interpretative differences, leading to misclassification of results as normal or abnormal depending on the source of the reference interval. While the heterogeneity observed in the Tanner stage-partitioned reference intervals is less pronounced than that seen in the age-partitioned intervals, it remains notable, especially given that these intervals were derived from homogenous subgroups using a standardized mass spectrometry method. Analysis of the reference intervals and the study characteristics of the selected articles revealed several weaknesses and limitations contributing to these shortcomings.

Age Partitions and Physiological Variability

A key limitation contributing to the variability in published RIs is the inconsistency in the age partitions applied across different studies. The rationale for age partitioning was not provided in the majority of articles (7 out of 12). When it was mentioned, it included Clinical Laboratory Standards Institute (CLSI) recommendations (24), existing literature (21), and 'standard' age partitions (23). Only three articles used statistical methods to establish age partitions (1,22,24). This inconsistency results in significant variations and abrupt changes when transitioning from one age category to another, as illustrated in our graphical representation. Even when RIs are established within a homogenous group such as pubertal stages (based on Tanner stage as well as testicular volume), there remains notable variability. This variance can be attributed to several factors such as the subjective nature of assessing pubertal stages and the different methods of assessing pubertal stage (e.g. orchidometer versus ultrasound for testicular volume, evaluation by physician versus self-reporting, Tanner genital stage versus Tanner pubic hair stage etc.). Furthermore, it has been previously stated that the use of Tanner stages to characterize steroid concentrations in prepubertal children is not always appropriate (19).

Inconsistently applied partitions may place children in different developmental stages within the same subgroup, which fail to consider physiological changes that occur during child development such as 'mini-puberty', adrenarche and gonadarche (22). The increase in androgen levels during 'mini-puberty' should result in a higher upper limit of normal for T and DHT in infant boys under the age of 6 months. This is not adequately addressed in many studies, which fail to differentiate infants undergoing 'mini-puberty' from other prepubertal children. Consequently, this can result in an underestimation of the upper limit of normal in infant boys. This is supported by the reference intervals for T and DHT established by Kulle et al. (33). They provided separate reference intervals for boys aged 0-6 months and 6 months to 9 years within Tanner stage 1. The younger group, experiencing 'mini-puberty', had a higher upper limit of normal for both T and DHT. Although 4 studies specifically established reference intervals in the age group beneath 6 months to take these endocrine fluctuations into account, none have subgroups with a sample size above 25, limiting their generalizability (1,30,31,33).

Similarly, adrenarche, which begins around age 5 and marks the onset of adrenal androgen production in both boys and girls, significantly influences T and DHT levels, but is not translated into separate age-partitions by most studies (34). This may result in an overestimation of the upper limit for boys aged 6 months to 5 years and for girls aged 0 to 5 years. Adriaansen et al. observed that children aged 0-5 years had a lower upper limit of normal compared

to the broader group of prepubertal children (boys aged 0-9 years and girls aged 0-10 years), illustrating the impact of adrenarche (21).

Our retrospective data analysis confirms these hormonal fluctuations (Figure 7-8). We observed high T concentrations in the first months of life, even in young girls (though less pronounced). In girls, this is followed by a gradual rise around age 5, consistent with the onset of adrenarche, which then moderately increases until and during puberty. In boys, the increase in T during adrenarche is less apparent due to a more pronounced elevation around age 10, marking the onset of puberty, which reaches much higher levels than those seen in female children.

Challenges in Sample Size and Data Representation

The limited cohort size in multiple studies leads to small sample sizes in different subgroups. In total, reference intervals for T and DHT were provided for 304 subgroups, including partitions by age, sex and pubertal stage. A sample size of the subgroup was provided in 240 subgroups (T: 188; DHT: 52). For T, 120 subgroups did not meet the acceptable threshold of 120 samples required to establish reference intervals of the CLSI EP28-A3C, with only 1 subgroup of 52 meeting this threshold for DHT (35). The primary issue when establishing RIs in the pediatric population is that obtaining blood samples from healthy children is considered an invasive procedure. Consequently, obtaining a sufficient sample size raises numerous practical, logistical and ethical difficulties (15,19). This challenge is further complicated by the rapidly changing physiology of children. Due to physiological changes during child development, numerous subgroups are needed, which further increases the required sample size. As result only a limited number of studies reach a sufficiently large cohort to establish meaningful and accurate RIs for each subgroup. This issue is most pronounced in very young children. In the age group under six months, no article reported a cohort larger than 25 samples (1,30,31,33). A consequence of the small sample sizes is that some articles established RIs based on the range of all results in the studied cohort instead of the 95% CI as recommended by CLSI EP28-A3C guideline (21,31,33). The small sample sizes used to establish many of the published reference intervals raise concerns regarding their generalizability, as smaller cohorts may include outliers that can significantly affect the results.

Inaccuracies due to Opportunistic Sampling

Because of the difficulties obtaining a sufficiently large cohort, many RIs many studies relied on opportunistic samples from routine clinical care instead of healthy volunteers (21,22,25,27,29–31,33). This approach involves samples collected from hospitalized children, patients undergoing elective surgeries, and leftover specimens intended for unrelated analyses). Since these samples are taken for unrelated medical reasons, it is assumed there is no impact on the selected parameters. Nevertheless, we noted that in some cases samples were also collected for steroid analysis or to exclude endocrine diseases, raising the likelihood of including pathological results (29,33). Opportunistic samples lack standardized preanalytical conditions, resulting in further inaccuracies (15). This was also noted by several authors, who stated that their reference intervals could be affected by the fact that samples were taken from children awaiting surgery, which may lead to elevated levels of adrenal androgens (21,27). The use of opportunistic samples can introduce bias and variability in the reference intervals, as the health status of these patients may not accurately reflect that of a typical healthy population.

Variability due to Ethnic Differences

The included RI studies are primarily based on Caucasian populations, with all studies conducted in European and North American cohorts, thereby overlooking potential ethnic differences in T and DHT levels. It has been stated in the literature that average T levels are higher in African populations compared to Caucasian populations (36,37). However, other studies could not support these findings and even report lower T concentrations in non-Hispanic Black adolescents (12-15 years old) compared to non-Hispanic whites (38,39). These ethnic variations are currently not considered in the published reference intervals.

Conclusion

Our systematic review assessed the current evidence on reference intervals for testosterone, dihydrotestosterone and T/DHT ratio analyzed using mass spectrometry. We found only a limited number of RIs published for DHT and none for T/DHT ratio. Despite numerous published RIs for T and the potential standardization offered by MS, our review highlights substantial variability. This variability can be attributed to several key issues, including age partitions that do not account for physiological variability, difficulties in obtaining sufficient sample sizes, and inaccuracies arising from opportunistic sampling. These limitations result in RIs that may not accurately reflect the dynamic endocrine changes in pediatric populations.

Based on these observations we conclude that more studies are needed to establish reliable RIs for DHT and T/DHT ratio. Additionally, efforts should focus on harmonizing testosterone RIs, which are critical in diagnosing and managing various pediatric endocrine disorders. Future studies should involve large, well-defined cohorts to statistically evaluate the physiological changes associated with child development. This would enable the establishment of age partitions that accurately represent the biochemistry of children and are more practical than those based on pubertal stages.

Alternatively, age partitions could be assessed by collecting a substantial amount of data from routine clinical care across multiple clinical centers, enabled by the standardization of MS-based assays. This approach offers a more accessible, less invasive, and cost-effective method to establish robust, age-partitioned reference intervals that accurately reflect physiological changes. Combining data from multiple laboratories would enhance the generalizability and harmonization of these reference intervals, ensuring their broader applicability in clinical practice.

TO DO/ACTIONS

In this critically appraised topic, we aim to raise awareness about the existing variation in reference intervals for testosterone and DHT. Through this effort, we intend to initiate a "call for data." We invite clinical laboratories to share routine clinical results of testosterone and DHT analyzed by mass spectrometry. By compiling a sufficiently large dataset, we seek to statistically evaluate age partitions that accurately reflect the evolving physiology of developing children, utilizing indirect methods for establishing reference intervals and applying statistical partitioning techniques.

REFERENCES

1. Kyriakopoulou L, Yazdanpanah M, Colantonio DA, Chan MK, Daly CH, Adeli K. A sensitive and rapid mass spectrometric method for the simultaneous measurement of eight steroid hormones and CALIPER pediatric reference intervals. Clin Biochem. 2013 May 46(7–8):642–51.
2. King B, Natale C, Hellstrom WJG. Testosterone Assays. Urologic Clinics of North America. 2022 Nov 1;49(4):665–77.
3. Ahmed SF, Achermann JC, Arlt W, Balen AH, Conway G, Edwards ZL, et al. UK guidance on the initial evaluation of an infant or an adolescent with a suspected disorder of sex development. Clin Endocrinol (Oxf) . 2011 Jul 10;75(1):12–26.
4. Fanelli F, Belluomo I, Di Lallo VD, Cuomo G, De Iasio R, Baccini M, et al. Serum steroid profiling by isotopic dilution-liquid chromatography–mass spectrometry: Comparison with current immunoassays and reference intervals in healthy adults. Steroids . 2011 Feb 1;76(3):244–53.
5. Tavita N, Greaves RF. Systematic review of serum steroid reference intervals developed using mass spectrometry. Clin Biochem . 2017 Dec 1;50(18):1260–74.
6. Furuyama S, Mayes DM, Nugent CA. A radioimmunoassay for plasma testosterone. Steroids . 1970 Jul 1;16(C):415–28.
7. Stanczyk FZ, Clarke NJ. Advantages and challenges of mass spectrometry assays for steroid hormones. J Steroid Biochem Mol Biol . 2010 Aug;121(3–5):491–5.
8. Matsumoto AM, Bremner WJ. Serum Testosterone Assays—Accuracy Matters. J Clin Endocrinol Metab . 2004 Feb;89(2):520–4.
9. Li JR, Goodman X, Cho J, Holditch-Davis D. The Variability and Determinants of Testosterone Measurements in Children: A Critical Review. Biol Res Nurs . 2021 Oct 18;23(4):646–57.
10. Hamer HM, Finken MJJ, van Herwaarden AE, du Toit T, Swart AC, Heijboer AC. Falsely elevated plasma testosterone concentrations in neonates: importance of LC-MS/MS measurements. Clinical Chemistry and Laboratory Medicine (CCLM) . 2018 May 24;56(6):e141–3.
11. Plebani M. Harmonization in laboratory medicine: the complete picture. Clinical Chemistry and Laboratory Medicine (CCLM) . 2013 Apr 1;51(4):741–51.
12. Berg J, Lane V. Pathology Harmony; a pragmatic and scientific approach to unfounded variation in the clinical laboratory. Ann Clin Biochem . 2011 May 9;48(Pt 3):195–7.

13. Greaves RF, Pitkin J, Ho CS, Baglin J, Hunt RW, Zacharin MR. Hormone Modeling in Preterm Neonates: Establishment of Pituitary and Steroid Hormone Reference Intervals. J Clin Endocrinol Metab . 2015 Mar;100(3):1097–103.
14. Koerbin G, Sikaris K, Jones GRD, Flatman R, Tate JR. An update report on the harmonization of adult reference intervals in Australasia. Clinical Chemistry and Laboratory Medicine (CCLM) . 2018 Dec 19;57(1):38–41.
15. Kohse KP. National and international initiatives and approaches for the establishment of reference intervals in pediatric laboratory medicine. LaboratoriumsMedizin . 2015 Jul 1;39(4):197–212.
16. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Mårtensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. Scand J Clin Lab Invest . 2004;64(4):271–84.
17. Travison TG, Vesper HW, Orwoll E, Wu F, Kaufman JM, Wang Y, et al. Harmonized Reference Ranges for Circulating Testosterone Levels in Men of Four Cohort Studies in the United States and Europe. J Clin Endocrinol Metab . 2017 Apr 1;102(4):1161–73.
18. Tavita N, Greaves RF. Systematic review of serum steroid reference intervals developed using mass spectrometry. Clin Biochem . 2017 Dec 1;50(18):1260–74.
19. Lyle AN, Pokuah F, Dietzen DJ, Wong ECC, Pyle-Eilola AL, Fuqua JS, et al. Current State of Pediatric Reference Intervals and the Importance of Correctly Describing the Biochemistry of Child Development. JAMA Pediatr . 2022 Jul 1;176(7):699.
20. Winter JSD, Hughes IA, Reyes FI, Faiman C. Pituitary-Gonadal Relations in Infancy: 2. Patterns of Serum Gonadal Steroid Concentrations in Man from Birth to Two Years of Age. J Clin Endocrinol Metab . 1976 Apr;42(4):679–86.
21. Adriaansen BPH, Oude Alink SE, Swinkels DW, Schröder MAM, Span PN, Sweep FCGJ, et al. Reference intervals for serum 11-oxygenated androgens in children. Eur J Endocrinol . 2024 Jan 3;190(1):96–103.
22. Holmes DT, van der Gugten JG, Jung B, McCudden CR. Continuous reference intervals for pediatric testosterone, sex hormone binding globulin and free testosterone using quantile regression. Journal of Mass Spectrometry and Advances in the Clinical Lab . 2021 Nov;22:64–70.
23. Madsen A, Oehme NB, Roelants M, Bruserud IS, Eide GE, Viste K, et al. Testicular Ultrasound to Stratify Hormone References in a Cross-Sectional Norwegian Study of Male Puberty. J Clin Endocrinol Metab . 2020 Jun 1;105(6):1888–98.
24. Bae YJ, Zeidler R, Baber R, Vogel M, Wirkner K, Loeffler M, et al. Reference intervals of nine steroid hormones over the life-span analyzed by LC-MS/MS: Effect of age, gender, puberty, and oral contraceptives. 2019;
25. Ankarberg-Lindgren C, Dahlgren J, Andersson MX. High-sensitivity quantification of serum androstenedione, testosterone, dihydrotestosterone, estrone and estradiol by gas chromatography–tandem mass spectrometry with sex- and puberty-specific reference intervals. J Steroid Biochem Mol Biol . 2018 Oct;183:116–24.
26. Søeborg T, Frederiksen H, Mouritsen A, Johannsen TH, Main KM, Jørgensen N, et al. Sex, age, pubertal development and use of oral contraceptives in relation to serum concentrations of DHEA, DHEAS, 17α-hydroxyprogesterone, Δ4-androstenedione, testosterone and their ratios in children, adolescents and young adults. Clinica Chimica Acta . 2014 Nov 1;437:6–13.

27. Kushnir MM, Blamires T, Rockwood AL, Roberts WL, Yue B, Erdogan E, et al. Liquid Chromatography–Tandem Mass Spectrometry Assay for Androstenedione, Dehydroepiandrosterone, and Testosterone with Pediatric and Adult Reference Intervals. Clin Chem . 2010 Jul 1;56(7):1138–47.
28. Salameh WA, Redor-Goldman MM, Clarke NJ, Reitz RE, Caulfield MP. Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: Total and free testosterone reference ranges. Steroids . 2010 Feb;75(2):169–75.
29. Soldin OP, Sharma H, Husted L, Soldin SJ. Pediatric reference intervals for aldosterone, 17α-hydroxyprogesterone, dehydroepiandrosterone, testosterone and 25-hydroxy vitamin D3 using tandem mass spectrometry. Clin Biochem . 2009 Jun;42(9):823–7.
30. Magnisali P, Dracopoulou M, Mataragas M, Dacou-Voutetakis A, Moutsatsou P. Routine method for the simultaneous quantification of 17α-hydroxyprogesterone, testosterone, dehydroepiandrosterone, androstenedione, cortisol, and pregnenolone in human serum of neonates using gas chromatography–mass spectrometry. J Chromatogr A . 2008 Oct;1206(2):166–77.
31. Rauh M, Gröschl M, Rascher W, Dörr HG. Automated, fast and sensitive quantification of 17α-hydroxy-progesterone, androstenedione and testosterone by tandem mass spectrometry with on-line extraction. Steroids . 2006 Jun;71(6):450–8.
32. Kushnir MM, Rockwood AL, Roberts WL, Pattison EG, Bunker AM, Fitzgerald RL, et al. Performance Characteristics of a Novel Tandem Mass Spectrometry Assay For Serum Testosterone. Clin Chem . 2006 Jan 1;52(1):120–8.
33. Kulle AE, Riepe FG, Melchior D, Hiort O, Holterhus PM. A Novel Ultrapressure Liquid Chromatography Tandem Mass Spectrometry Method for the Simultaneous Determination of Androstenedione, Testosterone, and Dihydrotestosterone in Pediatric Blood Samples: Age- and Sex-Specific Reference Data. J Clin Endocrinol Metab . 2010 May;95(5):2399–409.
34. Rege J, Karashima S, Lerario AM, Smith JM, Auchus RJ, Kasa-Vubu JZ, et al. Age-dependent Increases in Adrenal Cytochrome b5 and Serum 5-Androstenediol-3-sulfate. J Clin Endocrinol Metab . 2016 Dec 1;101(12):4585–93.
35. CLSI, EP28-A3C: documents, defining, establishing and verifying reference intervals in the clinical laboratory. 2010;Third ed.
36. Winters, SJ, Brufsky A, Weissfeld J, Trump DL, Dyky MA, Hadeed V. Testosterone, sex hormone-binding globulin, and body composition in young adult African American and Caucasian men. Metabolism . 2001 Oct;50(10):1242–7.
37. Ross R, Bernstein L, Judd H, Hanisch R, Pike M, Henderson B. Serum testosterone levels in healthy young black and white men. J Natl Cancer Inst. 1986 Jan;76(1):45–8.
38. Rohrmann S, Nelson WG, Rifai N, Brown TR, Dobs A, Kanarek N, et al. Serum Estrogen, But Not Testosterone, Levels Differ between Black and White Men in a Nationally Representative Sample of Americans. J Clin Endocrinol Metab . 2007 Jul 1;92(7):2519–25.
39. Lopez DS, Peskoe SB, Joshu CE, Dobs A, Feinleib M, Kanarek N, et al. Racial/ethnic differences in serum sex steroid hormone concentrations in US adolescent males. Cancer Causes & Control . 2013 Apr 25;24(4):817–26.

Supplemental Table 2. Published Age-Partitioned Reference Intervals for Testosterone by MS methods (part 1 of 2)

(Abbreviations: y: year; m: month, w: week, d: day, M: male, F: female, TV: testicular volume, PH: pubic hair, B: breast, G: genital)

First Author	Year	Reference Interval Definition	Subgroup	Testosterone Reference Intervals (nmol/L)								
				Mixed Sex			Male			Female		
				n	LL	UL	n	LL	UL	n	LL	UL
Adriaansen	2024	2.5-97.5th	0-5y	62	<0,03	0,18						
			0-10y (M); 0-9y old (F)		<0,03	0,3	94			39		
		Range	11-17y (M); 10-17y (F)				12	0,1	21,5	16	0,14	0,78
Holmes	2021	2.5-97.5th	0-<1y					0,05	7,31		0,04	0,14
			1-<2y					0,04	6,13		0,03	0,16
			2-<3y					0,03	4,39		0,03	0,18
			3-<4y					0,03	2,93		0,03	0,19
			4-<5y					0,03	1,79		0,04	0,22
			5-<6y					0,04	0,97		0,04	0,27
			6-<7y					0,05	0,47		0,04	0,37
			7-<8y					0,05	0,34		0,04	0,49
			8-<9y					0,04	0,81		0,04	0,66
			9-<10y					0,03	2,15		0,04	0,86
			10-<11y					0,05	4,42		0,04	1,07
			11-<12y					0,11	7,63		0,03	1,25
			12-<13y					0,26	11,77		0,03	1,38
			13-<14y					0,48	16,85		0,04	1,46
			14-<15y					0,77	22,3		0,05	1,49
			15-<16y					1,14	26,94		0,08	1,47
			16-<17y					1,59	29,87		0,11	1,41
			17-<18y					2,12	30,97		0,16	1,36
			18-<19y					2,72	30,26		0,21	1,38
			19-<20y					3,4	27,72		0,28	1,48
Madsen	2020	2.5-97.5th	6-<10y				169	0,02	0,28			
			10-<13y				137	0,07	10,7			
			13-<16y				108	1,02	22,3			
Bae	2019	2.5-97.5th	0-<1y				124		≤ 4,01	156		≤ 3,27
			1-10y				855		≤ 0,32	602		≤ 0,49
			10-15y				730	0,11	25,78	622	0,14	1,74
			15-20y				336	8,29	30,72	319	0,39	2,15
Kyriakopoulou	2013	2.5-97.5th	0–14d				18	0,25	6,36	13	0,06	0,43
			15d–<1y				34	0,05	8,9	28	0,02	0,22
			1–<13y				138	0,06	1,12	138	0,06	1,12
			13–<16y				18	0,97	15,7	17	0,5	1,36
			16–<19y				30	0,86	23,4	31	0,11	2,31
Kushnir	2010	2.5-97.5th	6–24m				123		<1.28	92		**<0.3**
			2–3y				125		<0.520	126		**<0.7**
			4–5y				125		<0.659	127		**<1.04**
			6–7y				125		<0.451	131		<0.24
			7–9y				206	0,059	0,28	206	0,035	0,381
			10–11y				140	0,080	5,721	148	0,10	1,11
			12–13y				143	0,10	21,46	142	0,21	1,7335
			14–15y				141	1,07	25,41	143	0,21	1,80
			16–17y				136	5,478	28,64	138	0,31	2,01

Supplemental Table 2. Published Age-Partitioned Reference Intervals for Testosterone by MS methods (part 2 of 2)

(Abbreviations: y: year; m: month, w: week, d: day, M: male, F: female, TV: testicular volume, PH: pubic hair, B: breast, G: genital)

Salameh	2010	2.5-97.5th	1–5.9y						<0.2			<0.3
			6–7.9y						<0.87			<0.69
			8–10.9y						<1.5			<1.2
			11–11.9y						<9.02			<1.4
			12–13.9y						<14.6			<1.4
			14–17.9y						<34.7			<1.4
Soldin	2009	2.5-97.5th	0 – 6y				158	0,1	1,1			
			7 – 9y				148	0,1	0,87			
			10 – 12y				188	0,2	14,5			
			13 – 14y				245	0,2	22,5			
			15 – 16y				209	1,5	30,5			
			17 – 19y				123	4,20	28,6			
			0 – 5y							127	0,1	0,35
			6 – 9y							171	0,1	0,45
			10 – 14y							190	0,49	1,7
			15 – 16y							175	0,42	1,8
			17 – 18y							123	0,56	1,7
Magnisali	2008	not specified	9-15d	7	0,3	4,75						
			16-40d				9	0,69	6,32	10	0,56	2,1
Rauh	2006	Range (UL of mixed group is p97.5)	0-0.5y	40			25		<14,6	15		<2,94
			1-6y	70		<0,77	41		<1,38	29		<0,55
			7-12y	62		<3,33	29		<9,08	33		<3,33
			13-15y	28			18	0,18	44,7	10	0,46	2,13
			16-18y	23			12	0,93	29,9	11	0,41	2,29
Kushnir	2006	2.5-97.5th	7–9y				105		<0,311	103		<0,519
			10–11y				68	0,069	1,97	69	0,069	1,45
			12–13y				69	0,242	25,85	69	0,208	2,21
			14–15y				70	1,14	20,24	70	0,311	1,7
			16–17y				70	6,4	30,66	70	0,277	2,18
Kulle	2010	Range	<1w				8	0,2	2,7	6	0,21	2,7
			2w -2m				19	0,5	12,6	23	<0,1	0,7
			3-5m				12	0,1	3,1	10	<0,1	0,4
			6-11m				4	0,1	0,4	10	0,1	0,7
			1-3y				9	0,1	0,7	9	<0,1	0,6
			4-6y				10	0,1	0,9	13	<0,1	0,7
			7-9y				10	0,1	0,8	14	0,1	1
			10-12y				18	0,1	5,6	17	0,1	1,5
			13-15y				26	0,1	17,6	19	0,1	2
			16-18y				21	4	24	10	0,1	1,7

Supplemental Table 3. Published Reference Intervals for Testosterone by MS methods, all subgroups other than age-partitions (part 1 of 2)

(Abbreviations: y: year; m: month, w: week, d: day, M: male, F: female, TV: testicular volume, PH: pubic hair, B: breast, G: genital)

First author	Year	Reference Interval Definition	Subgroup	Testosterone Reference intervals (nmol/L)					
				Male			Female		
				n	LL	UL	n	LL	UL
Madsen	2020	2.5-97.5th	10-< 13y old (TV < 2.7 mL)	84	0,06	3,42			
			10-< 13y old (TV ≥ 2.7 mL)	52	0,16	12,8			
			Prepubertal (TV < 2.7 mL)	247	0,02	0,77			
			Pubertal (TV: 2.7-12.8 mL)	115	0,24	17,5			
			Tanner PH stage 1	249	0,02	2,17			
			Tanner PH stage 2/3/4	90	0,11	18,1			
			Tanner PH stage 5	64	2,81	25,9			
Bae	2019	2.5-97.5th	Tanner stage 1	266		≤ 1.0	224		≤ 0.4
			Tanner stage 2	75	0,1	15,5	99	1	1,1
			Tanner stage 3	63	0,4	25,4	74	0,3	1,8
			Tanner stage 4	73	6,5	30,3	68 (not on OCP)	0,4	1,8
			Tanner stage 5	114	10	33,1	90 (not on OCP)	0,5	2,4
Ankarberg-Lindgren	2018	5-95th	TV: 1-2 mL	9	0,1	0,4			
			TV: 3-6 mL	9	0,2	2,6			
			TV: 8-12 mL	14	2,2	22,9			
			TV: 15-25 mL	9	11,1	23,6			
			Tanner B stage 1				16	0,1	0,5
			Tanner B stage 2				15	0,2	0,9
			Tanner B stage 3				13	0,3	1,6
			Tanner B stage 4/5				16	0,3	1,5
Søeborg	2014	Mean ±2SD	Tanner PH stage 1	272	<LOQ	5,6	134	<LOQ	0,62
			Tanner PH stage 2	74	0,12	24	32	0,14	1,5
			Tanner PH stage 3	29	1,1	44	32	0,2	2,1
			Tanner PH stage 4	46	3,6	36	33	0,34	1,9
			Tanner PH stage 5	69	9,2	33	160	0,48	2,4
			Tanner PH stage 6	428	11	33			
			Tanner G stage 1 (M)	215	<LOQ	1			
			Tanner G stage 2 (M)	109	0,14	17			
			Tanner G stage 3 (M)	52	0,84	35			
			Tanner G stage 4 (M)	37	4	39			
			Tanner G stage 5 (M)	504	9,9	34			
			Tanner B stage 1 (F)				122	<LOQ	0,42
			Tanner B stage 2 (F)				32	0,12	1,1
			Tanner B stage 3 (F)				32	0,17	2,4
			Tanner B stage 4 (F)				132	0,42	2,1
			Tanner B stage 5 (F)				79	0,43	2,7
			TV: < 3mL	81	<LOQ	0,69			
			TV: 3 - < 4mL	70	<LOQ	1,4			
			TV: 4 - < 11 mL	141	0,19	22			
			TV: 11 - < 16 mL	54	3,3	37			
			TV: > 16 mL	502	10	34			

Supplemental Table 3. Published Reference Intervals for Testosterone by MS methods, all subgroups other than age-partitions (part 2 of 2)

(Abbreviations: y: year; m: month, w: week, d: day, M: male, F: female, TV: testicular volume, PH: pubic hair, B: breast, G: genital)

Kushnir	2010	2.5-97.5th	Tanner stage 1	278	0,055	0,520	296	0,066	0,589	
			Tanner stage 2	131	0,11	10,51	120	0,156	1,388	
			Tanner stage 3	140	0,3467	29,50	135	0,347	2,186	
			Tanner stage 4/5	204	5,617	29,37	205	0,382	2,151	
			Before menarche				413	0,066	1,21	
			After menarche, ≤18 years				323	0,347	2,18	
Salameh	2010	2.5-97.5th	Tanner stage 1			<0,2			<0,3	
			Tanner stage 2			<5,79			<0,83	
			Tanner stage 3		0,73	24,9			<0,97	
			Tanner stage 4		0,87	31,6			<1,1	
			Tanner stage 5		3,817	33,833			<1,1	
Kushnir	2006	2.5-97.5th	Tanner stage 1	143		<0,657	147		<0,588	
			Tanner stage 2	66	0,069	5,16	66	0,138	1,35	
			Tanner stage 3	65	0,242	26,37	85	0,346	2,07	
			Tanner stage 4	98	5,71	29,54	77	0,277	2,18	
			Tanner stage 5	42	6,71	27,09	123	0,346	2,07	
Kulle	2010	Range	Tanner stage 1 (M: <6m; F: <8y)	41	0,1	3,2	71	<0,1	1	
			Tanner stage 1 (M: 6m-9y)	35	0,1	2,4				
			Tanner stage 1 (M: >9y; F: >8y)	13	0,1	1	14	0,1	2,7	
			Tanner stage 2	11	0,3	13,3	10	0,1	1,7	
			Tanner stage 3	11	1,2	14,2	10	0,3	1,3	
			Tanner stage 4	10	1	15,8	11	0,1	2	
			Tanner stage 5	18	5	24	12	0,2	2,5	
			TV: 1-2 mL (<6m)	40	0,1	15				
			TV: 1-2 mL (6m-9y)	26	0,1	1,5				
			TV: 1-2 mL (>9y)	7	0,11	1				
			TV: 3-4mL	7	0,1	10				
			TV: 5-10mL	9	1,2	15				
			TV: 11-15mL	15	1,7	16,4				
			TV: >15mL	10	5	24				

Supplemental Table 4. Published Age-Partitioned Reference Intervals for Dihydrotestosterone by MS methods.

(Abbreviations: y: year, m: month, w: week, M: male, F: female, TV: testicular volume, B: breast)

First author	Year	Reference Interval Definition	Subgroup	Dihydrotestosterone Reference Intervals (nmol/L)								
				Mixed Sex			Male			Female		
				n	LL	UL	n	LL	UL	n	LL	UL
Adriaansen	2024	2.5-97.5th	0-5y	62	<0,04	0,05						
			0-10y (M); 0-9y old (F)		<0,04	0,11	94			39		
		Range	11-17y (M); 10-17y (F)				12	0,05	1,59	16	0,04	0,29
Kulle	2010	Range	<1w				8	<0,1	0,7	6	<0,1	0,1
			2w -2m				19	<0,1	2,6	23	<0,1	1
			3-5m				12	<0,1	0,8	10	<0,1	0,1
			6-11m				4	<0,1	0,4	10	<0,1	0,1
			1-3y				9	<0,1	1,3	9	<0,1	0,4
			4-6y				10	0,1	0,8	13	<0,1	0,4
			7-9y				10	0,1	0,6	14	<0,1	0,5
			10-12y				18	0,1	1,9	17	<0,1	0,8
			13-15y				26	0,1	3,1	19	0,1	1
			16-18y				21	0,1	1,9	10	0,1	1

Supplemental Table 5. Published Reference Intervals for Dihydrotestosterone by MS methods, all subgroups other than age-partitions

(Abbreviations: y: year, m: month, w: week, M: male, F: female, TV: testicular volume, B: breast)

First author	Year	Reference Interval Definition	Subgroup	Dihydrotestosterone Reference					
				Male			Female		
				n	LL	UL	n	LL	UL
Ankarberg-Lindgren	2018	5-95th	TV: 1-2 mL	9	<0,027	0,227			
			TV: 3-6 mL	9	0,08	0,31			
			TV: 8-12 mL	14	0,18	1,242			
			TV: 15-25 mL	9	0,65	1,363			
			Tanner B stage 1				16	<0,027	0,187
			Tanner B stage 2				15	0,028	0,337
			Tanner B stage 3				13	0,069	0,365
			Tanner B stage 4/5				16	0,027	0,528
Kulle	2010	Range	Tanner stage 1 (M: <6m; F: <8y)	41	0,1	3	71	0,1	1
			Tanner stage 1 (M: 6m-9y)	35	0,1	2,4			
			Tanner stage 1 (M: >9y; F: >8y)	13	0,1	1	14	0,1	0,5
			Tanner stage 2	11	0,1	1,8	10	0,1	0,5
			Tanner stage 3	12	0,1	1,9	10	0,1	0,8
			Tanner stage 4	10	0,1	1,8	11	0,1	1
			Tanner stage 5	18	0,2	3,2	12	0,1	0,7
			TV: 1-2 mL (<6m)	40	<0,1	3,4			
			TV: 1-2 mL (6m-9y)	26	<0,1	1,3			
			TV: 1-2 mL (>9y)	7	0,1	0,7			
			TV: 3-4mL	7	0,1	1,7			
			TV: 5-10mL	9	0,1	1,7			
			TV: 11-15mL	15	0,3	1,9			
			TV: >15mL	10	0,3	3,2			

UZ LEUVEN | LABORATORIUMGENEESKUNDE

Supplemental Table 1: Study Characteristics (part 1 of 2)

(Abbreviations: T: testosterone, DHT: dihydrotestosterone, TV: testicular volume, M: male, F: female)

First author	Year	Available parameters	Method	Analyser/Manufacturer	Calibrator traceability	Matrix	Partitioning	criteria for partitions	Country	Ethnicity	Age	Cohort size	Inclusion criteria	Exclusion criteria
Adriaansen	2024	T, DHT	LC-MS/MS	Xevo TQ-XS StepWave XSTM (Waters)	Sigma-Aldrich	Serum	Age (converted to pubertal stage), sex	Literature	The Netherlands	Western-European (mostly) + Mediterranean	0-17y	256 (146 M, 110 F)	- Age 0 to <18 years - Need for a venipuncture or intravenous drip placement for the purpose of general anesthesia or sedation	- Use oral contraceptives/glucocorticoids - Iron deficiency anemia - Any systemic underlying disease (malignancy, asthma, diabetes, congenital heart disease, kidney failure, congenital immunodeficiency, etc.) - Acute infection - Trauma or operation 48 h ago - Inflammation, infection (CRP >5 mg/L) - ALT >40 IU/L - Treatment with iron preparations - Children admitted for eartube insertion, adenotomy, or adenotonsillectomy because of recurrent infections of the upper airways
Holmes	2021	T	LC-MS/MS	20AC liquid chromatography system (Shimadzu), API 5000 triple quadrupole mass spectrometer (SCIEX)	NIST 971	Serum	Age, sex	Yearly reference intervals	Canada	NA	6m-19y (F); 1m-19y (M)	421 (195 M, 226 F)	- Serum samples for routine analysis for reactivity to allergens	- No clinical exclusion criteria
Madsen	2020	T	LC-MS/MS	1290 UPLC system (Agilent), API 5500 triple quadrupole mass spectrometer (SCIEX)	Steraloids, Newport, RI	Serum	Age, pubertal stage (TV, Tanner pubic hair stage)	Standard age-partitioned references confirmed by Harris and Boyd method	Norway	Majority of included children had two Norwegian parents*	6-16y	491 (M only)	- Healthy individuals from 6 schools in Norway	- Chronic diseases that could affect growth - Nonthreatening scrotal pathologies including microcalcifications, hydrocele, or cryptorchidism
Bae	2019	T	LC-MS/MS	Prominence UFLC systemn (Shimadzu), QTRAP 6500 (SCIEX)	6PLUS1® multilevel serum calibrators (Chromsystems Instruments & Chemicals GmbH, Munich,Germany) (NIST 971).	Serum	Age, sex, pubertal stage (Tanner stage)	Principles of partitioning from CLSI, calculated by GAMLSS for fitted model	Germany	NA	0.33-79y	4678 (2512 M, 2166 F)	- All children, parents and pregnant women that are interested in the study and not suffering from any chronic, chromosomal, and syndromal diseases	- Concomitant endocrine diseases - Glucocorticoid therapy - Hepatitis - Pregnancy - Body mass index (BMI) ≥ 33 in adults or BMI standard de- viation score (SDS) below -3 or over +3 in children and adolescents
Ankarberg-Lindgren	2018	T, DHT	GC-MS/MS	7890B GC (Agilent), 7000 triple quadrupole mass spectrometer (Agilent)	Ceriliant	Serum	Sex, pubertal stage (TV, Tanner breast stage)	Only pubertal stage partitions	Sweden	NA	5-18y	101 (41 M, 60 F)	- Leftover routine samples	- No subject of hormone replacement therapy or pubertal blockers

Supplemental Table 1: Study Characteristics (part 2 of 2)

(Abbreviations: T: testosterone, DHT: dihydrotestosterone, TV: testicular volume, M: male, F: female)

Søeborg	2014	T	LC-MS/MS	Accela LC system, TSQ Vantage triple stage quadropole mass spectrometer (Thermo Fisher)	Sigma-Aldrich	Serum	Sex, pubertal stage (Tanner pubic hair stage (M/F), Tanner genital stage (M), Tanner breast stage, TV)	Only pubertal stage partitions	Denmark	NA	0.2-27.6y	1798 (1133 M, 665 F)	- Healthy volunteers	NA
Kyriakopoulou	2013	T	LC-MS/MS	1200 series HPLC (Agilent) + ZPI 4000 QTRAP mass spectrometer (SCIEX)	Sigma-Aldrich	Serum	Age, sex	Harris and Boyd method	Canada	Multi-ethnic	0-18y	465 (238 M, 227 F)	- Healthy children	- History of chronic illness or metabolic disease - Acute illness within the previous month - Use of prescribed medication over the previous 2 weeks
Kushnir	2010	T	LC-MS/MS	1200 series pump (Agilent), API 4000 triple- quadrupole mass spectrometer (SCIEX)	Sigma-Aldrich	Serum	Age, sex, pubertal stage (Tanner stage), menstrual status	Not specified	USA	>90% of participants were white	6m-17y	2517 children (1264 M, 1253 F)	- Apparently healthy volunteers (7-17y: research setting; 6m-6y; before elective surgical procedure)	- Use of prescription medications - Known medical conditions
Salameh	2010	T	LC-MS/MS	Aria TX-4 HTLC, Finnigan TSQ Quantum Ultra tandem mass spectrometer (ThermoFisher)	Sigma–Aldrich	Serum	Age, sex, pubertal stage (Tanner stage)	Not specified	USA	NA	8-90y	499 (children + adults; 264 M, 235 F)	- Apparently normal volunteers (the population studied is "normal" by self-report)	- Abnormal AST, ALT, and WBC - Use of medications - History of steroid use or endocrine related disease
Soldin	2009	T	LC-MS/MS	HPLC (Shimadzu), API-5000 triple-quadrupole mass spectrometer (SCIEX)	Sigma-Aldrich	Serum	Age, sex	Not specified	USA	NA	0-18y	1857 (1071 M, 786 F)	- Blood specimens on which one or more steroids had been ordered were obtained	- 'Diseased population' (by application of Hofman method)
Magnisali	2008	T	GC-MS/MS	Finnigan TRACE 2000 series GC (Thermo Fisher), Finnigan TraceMS (Thermo Fisher)	Sigma-Aldrich	Serum	Age, sex	Not specified	Greece	NA	9-40d	26 (12 M, 14 F)	- Samples screened for glucose-6-phosphate dehydrogenase deficiency - Healthy full term infants	NA
Rauh	2006	T	LC-MS/MS	1100 series binary pump (Agilent), API 4000 tandem mass spectrometer (SCIEX)	Sigma–Aldrich	Serum	Age, sex	Not specified	Germany	NA	1d-18y	223 (125 M, 98 F)	- Children who were admitted to the University Children's Hospital in Erlangen	- Endocrine, cardiac, renal and hepatic diseases - Water and electrolyte disorders
Kushnir	2006	T	LC-MS/MS	1100 series HPLC pump (Agilent), API 4000 triple-quadrupole mass spectrometer (SCIEX)	Sigma-Aldrich	Serum	Age, sex, pubertal stage (Tanner stage)	Not specified	USA	NA	7-17y	763 (382 M, 381 F)	- Apparently healthy volunteers	- Women on oral contraceptives/female hormone replacement therapy
Kulle	2010	T, DHT	LC-MS/MS	Acquity UPLC system (Waters), Quattro Premier/XE triple quad mass spectrometer (Waters)	Sigma-Aldrich	Plasma	Age, sex, pubertal stage (Tanner stage, TV)	Not specified	Germany	NA	0-18y	269 (138 M, 131 F)	- Leftover samples from blood checks before minor surgery (e.g. inguinal hernia repair, circumcision, and tonsillectomy) or for exclusion of endocrine diseases (e.g. presumed growth disorder or presumed hypothyroidism)	- Active signs of endocrine or systemic disease - Use of steroid medication

UZ LEUVEN | LABORATORIUMGENEESKUNDE