Johannes Böhm, Boris Rolinski, Sven Schneider and Julian E. Gebauer*

Automated age and sex partitioning of reference intervals based on routine laboratory data

https://doi.org/10.1515/labmed-2025-0210
Received August 15, 2025; accepted October 11, 2025;
published online November 25, 2025

Abstract

Objectives: In this study we evaluated the performance of the decision and regression tree–based algorithm `rpart` for age- and sex-partitioning of reference intervals (RIs) in an indirect setting, and compared its results with those derived from the CALIPER study.

Methods: Routine laboratory data for γ-glutamyltransferase (GGT), creatinine (CREA), and alkaline phosphatase (AP) were analyzed. Indirect RIs were estimated using the reflimR package, with partitions defined either by CALIPER or suggested by `rpart`. Corresponding partitions were compared using the Harris and Boyd method and assessed for clinical plausibility.

Results: `rpart` produced clinically and statistically meaningful partitioning schemes in mixed populations with unknown proportions of pathological values. Compared with CALIPER, `rpart` often yielded comparable RIs but avoided non-significant splits. Notable differences were observed in specific age groups, particularly in early childhood partitions for GGT and neonatal partitions for AP. Validation against original CALIPER RIs showed good agreement in most cases, with some broader intervals in the indirect approach.

Conclusions: These findings support the use of `rpart` not only in carefully preselected reference cohorts but also in heterogeneous routine laboratory datasets. Combined with indirect RI estimation via reflimR, automated partitioning offers a practical and reproducible approach to establishing clinically relevant RIs. Further multicenter studies should be conducted to confirm these results.

***Corresponding author: Dr. med. Julian E. Gebauer**, Department of Laboratory Medicine, Elblandkliniken, Nassauweg 7, Meißen, Germany, E-mail: julian-gebauer@gmx.de. https://orcid.org/0000-0003-0216-3607
Johannes Böhm and Sven Schneider, Institute for Laboratory and Transfusion Medicine, Klinikum Passau – Medizin Campus Niederbayern, Passau, Germany
Boris Rolinski, Department of Laboratory Medicine, Elblandkliniken, Nassauweg 7 Meißen, Germany

Keywords: reference intervals; indirect methods; machine learning; age and sex partitioning; regression tree model

Introduction

For the clinical interpretation of analysis results in laboratory medicine, method-specific reference intervals are required to classify a measurement as increased or decreased. Biological factors such as age, sex, ethnicity or environmental conditions as well as the analysis method have a significant influence on reference intervals. According to the CLSI/IFCC standard [1], reference intervals are defined as the central 95 % range of a healthy population. The estimation of reference intervals from healthy subjects specifically acquired for this purpose is called the direct approach. To ensure sufficient statistical significance, a minimum number of measurements per each subpopulation is required [2]. The resulting necessary number of healthy subjects for the direct estimation of RI leads to a considerable effort.

An alternative approach of establishing RIs is the use of indirect methods. In contrast to the direct approach, indirect estimated RI are calculated on the basis of retrospectively collected routine data. However, these contain not only healthy but also pathological results. Therefore, the presumably healthy population must be extracted to estimate RI. There are various statistical approaches for this [3–6]. Out of these, we decided to use `reflimR` [6] as it yields robust results with little computational effort.

The indirect approach may require complex data analysis to assess clinical validity. However, RI intervals from routine data may reflect the actual patient population better than the direct method applied to highly selected reference cohorts. In addition, indirect methods can be particularly helpful for population groups where it is not possible to obtain sufficient data from healthy volunteers e.g. for children. For more detailed information on the advantages and disadvantages of indirect RI methods the reader is referred to references [7, 8].

In our previous work, we addressed the importance of RI partitioning [9]. The partitioning of RI is typically based on biological background, using predefined sex and age groups. We previously demonstrated that mathematical partitioning

using regression tree algorithms yields comparable results. For this, we utilized data from the CALIPER study, which was collected from a healthy study population. Now, we aim to show that this methodology can also be applied to the indirect determination of RI. A key aspect of this approach is addressing potential bias in the partitioning results caused by pathological subpopulations within the data.

Materials and methods

Based on our previous work, we decided to use the analytes γ-glutamyl transferase, creatinine, and alkaline phosphatase.

Data collection was performed retrospectively from the database of the laboratory information system of Klinikum Passau for the period from the beginning of 2020 to the end of 2024. The measurements of the respective analytes were performed on ADVIA-analyzers (Siemens Healthineers AG, Forchheim, Germany) using the ADVIA Chemistry XPT assays γ-glutamyltransferase (GGT), creatinine_2 (Crea_2), and alkaline phosphatase_2, concentrated (ALP_2c). The data was preprocessed, whereby only the first measurement value per patient ID was used. All subsequent measurements were discarded. For analysis, the dataset was anonymized so that only the measured analyte, sex and age in days remained. Informed consent was not required due to the anonymized, retrospective data analysis according to the statement of our Ethics Committee (Ethikkomission Uni Regensburg, 25-4055-104).

Data analysis was performed in `R` (version 4.4.2) [10]. Data processing and visualization was performed using the tidyverse-packages `dplyr`, `tidyr` and `ggplot2` [11]. Partitioning was performed using `rpart` [12] and `rpart.plot` [13]. The `rpart` algorithm was applied with the default control parameters. Key parameters include the complexity parameter (cp=0.01), which sets the minimum improvement in model fit required for a split, and the minimum number of observations in a terminal node (minbucket=7), which ensures that each resulting group contains enough data for stable estimation. Other parameters, such as minsplit (the minimum number of observations required to attempt a split) and maxdepth (the maximum number of consecutive splits), also influence the granularity and depth of the tree. Adjusting these parameters allows researchers to control how finely or broadly the data are partitioned. Higher values for cp or minbucket produce fewer, broader partitions, while lower values allow more detailed subgrouping. In our dataset, which was sufficiently large, the default settings generated robust and clinically interpretable age- and sex-partitions without producing excessively small or spurious subgroups. It should be noted that `rpart` can generate either decision trees for categorical outcomes or regression trees for continuous outcomes. In this study, regression trees and therefore the ANOVA-method were used, allowing the algorithm to model continuous laboratory values and determine optimal age- and sex-based splits in a data-driven manner. Estimation of reference intervals was performed using `reflimR` [6]. For statistical comparison of the reference intervals, the Harris & Boyd method was used, as described by Lahti 2004 [14].

Results

Figures 1–3 provide a visual summary of the analysis results. As in Klawitter et al. [9], Subfigure A shows the indirectly estimated reference intervals (RIs) based on the raw data, using the age and sex partitioning according to CALIPER. Subfigure B then displays the indirect RIs using partitioning derived from the `rpart` algorithm. For improved graphical presentation, extreme outliers were excluded from the visualizations of subfigures A and B. This concerned three GGT values>600 U/l, 18 creatinine values>2 mg/dL, and six alkaline phosphatase values>800 U/L. Subfigure C displays the segmentation results generated by the `rpart` algorithm in the form of a decision tree. Each split (node) and terminal group (leaf) is shown together with the number of observations it contains as well as the corresponding percentage of the total dataset. This representation illustrates how the algorithm recursively divides the population according to age and sex until no further statistically significant partitioning can be achieved. Additionally, Subfigure D presents a bar chart comparing the individual RIs side by side. The original reference intervals from Colantonio et al. [15] are also shown here. Mean±1 standard deviation is indicated with vertical lines and whiskers, while the permissible uncertainty is visualized as an extension on either side of the bars.

An insufficient sample size was encountered only for the RI of alkaline phosphatase (AP) in newborns. Therefore, the reference interval was estimated using the central 95 % percentiles, which is indicated by a dashed line in Figure 3A and B.

The combined RIs for males and females are shown in black, the RIs for males in blue, and those for females in red. For better visual presentability some extreme values were truncated in the plots for better resolution of the reference intervals.

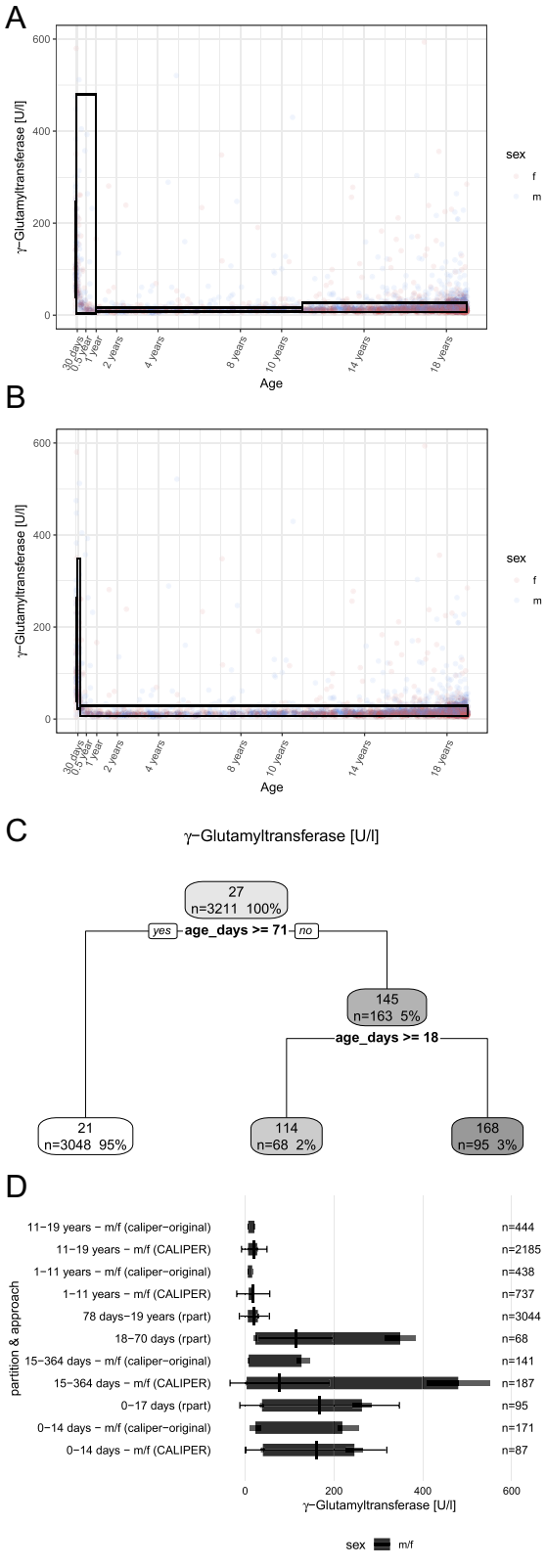

Figure 1: Reference intervals for γ-glutamyltransferase. (A) RI of partitions suggested by CALIPER for GGT. (B) RI of partitions suggested by rpart for GGT. (C) Flow chart for age- and sex-dependent reference intervals for GGT. (D) RI chart for age- and sex-dependent reference intervals for GGT.

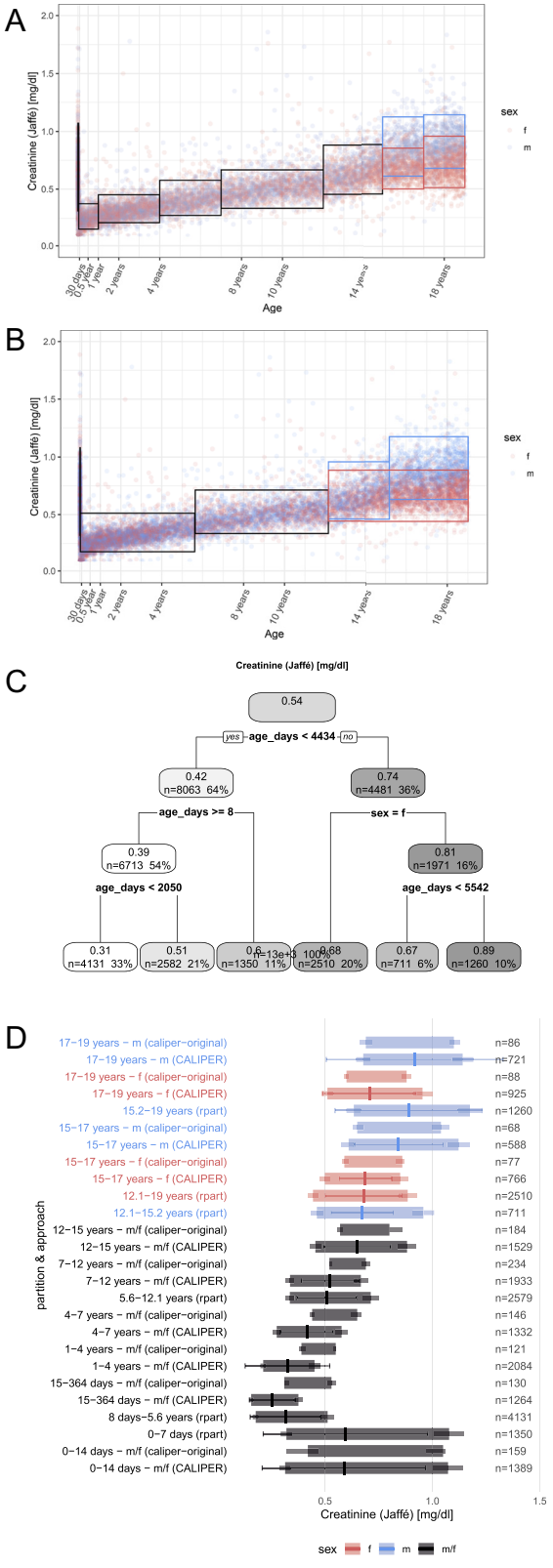

Figure 2: Reference intervals for creatinine. (A) RI of partitions suggested by CALIPER for creatinine. (B) RI of partitions suggested by rpart for creatinine. (C) Flow chart for age- and sex-dependent reference intervals for creatinine. (D) RI chart for age- and sex-dependent reference intervals for creatinine.

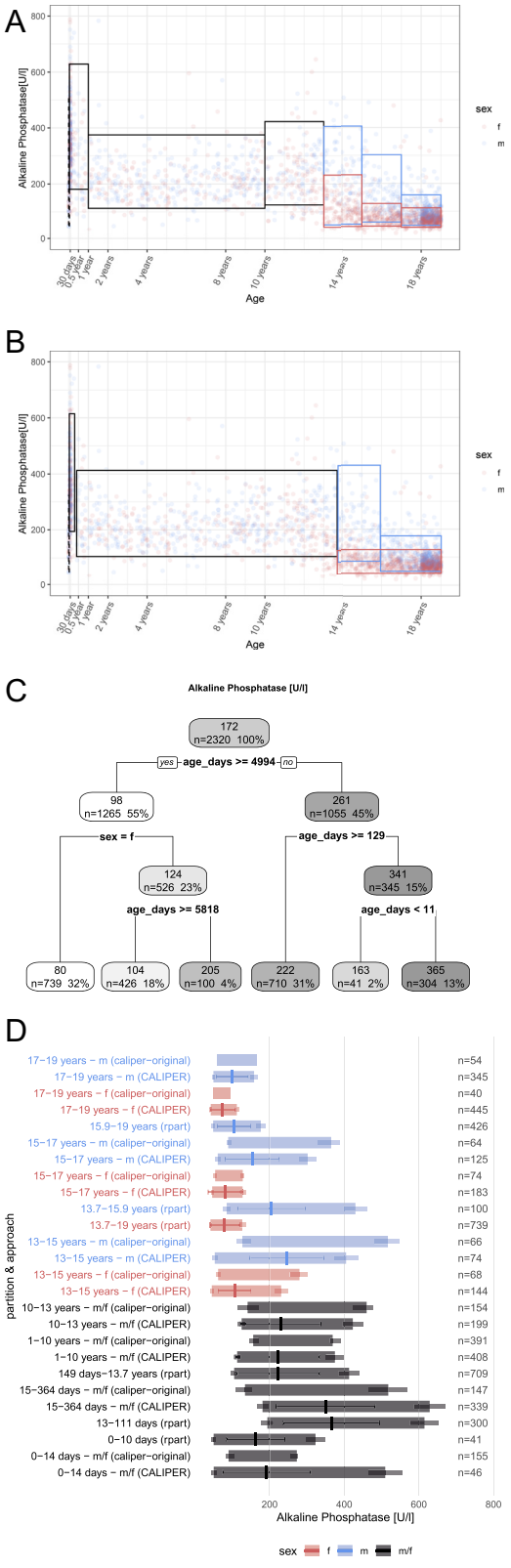

Figure 3: Reference intervals for alkaline phosphatase. (A) RI of partitions suggested by CALIPER for alkaline phosphatase. (B) RI of partitions suggested by rpart for alkaline phosphatase. (C) Flow chart for age- and sex-dependent reference intervals for alkaline phosphatase. (D) RI chart for age- and sex-dependent reference intervals for alkaline phosphatase.

The corresponding numerical values underlying the plots can again be found in the appendix. Supplementary Table 1 contains the reference interval data. Supplementary Table 2 includes the Harris & Boyd comparison results.

For GGT, the CALIPER method proposed four age-based partitions without sex differentiation, while the `rpart` algorithm suggested three age splits, likewise without distinguishing between sexes. CALIPER defined age thresholds at 14 days, 1 year, and 11 years. In contrast, `rpart` yielded a broader age partition ranging from 78 days to 18 years, with an additional early split at 17 days for neonates.

According to the Harris and Boyd statistical comparison, no significant differences were found between the partitioning schemes derived from CALIPER and `rpart`. Within the `rpart` segmentation, only the first split at 18 days reached statistical significance. For CALIPER, the first two age splits showed statistically significant differences.

For CREA, CALIPER suggests sex-independent reference intervals with age-based splits at 14 days, 1 year, 4 years, 7 years, 12 years, 15 years, and 17 years. From 15 to 18 years, CALIPER introduces a sex-based separation, providing distinct reference intervals for males and females. In contrast, `rpart` separates neonates after just 8 days. It then consolidates the four CALIPER age intervals into two partitions with only one split at approximately 5.6 years. From around 12 years onward, `rpart` suggest sex-specific partitioning. For females, a single uniform interval is suggested up to 18 years, whereas for males, an additional subdivision occurs at approximately 15 years.

Within each of the two approaches, all consecutive unisex partitions were found to differ significantly from one another according to the Harris and Boyd method. In contrast, non-significant splits were also found between the corresponding sex-specific partitions of `rpart` and CALIPER. No statistically significant differences were detected when comparing corresponding male or female partitions between CALIPER and `rpart`.

For AP, both approaches yield broadly similar partitioning schemes at first glance. Within the first year of life, however, `rpart` proposes narrower partitions with splits at 11 and 129 days, compared to CALIPER, which defines age thresholds at 14 days and 1 year. Following this, `rpart` introduces a wide age partition from 129 days to 13.7 years, thereby avoiding a statistically unnecessary split suggested by CALIPER at 10 years. Thereafter, `rpart` suggests a uniform interval for females from 13.7 to 18 years, while for males, an additional split is introduced at approximately 15.9 years.

From a statistical perspective, `rpart` avoids non-significant splits in contrast to CALIPER. The sex-independent partitions of both methods show no significant differences. Interestingly, however, the male-specific partitions differ significantly between the two approaches.

Discussion

Building on our previous work [9], in which we demonstrated that the decision tree-based algorithm rpart can generate valid partitions for the estimation of reference intervals using the direct method based on data from presumably healthy individuals, the present study extends the applicability of this approach to an indirect setting using routine laboratory data. Our results show that rpart is equally capable of producing clinically and statistically meaningful partitioning schemes when applied to mixed populations, where the presence and proportion of pathological values are unknown. Compared to the CALIPER study, rpart was able to propose a medically and statistically reasonable age- and sex-based partitioning of routine laboratory data.

The examination of rpart partitioning for individual analytes again revealed the typical characteristics of the rpart algorithm, as we previously described in Klawitter et al. [9]. When considering the first interval corresponding to neonates, CALIPER established 14 days as the age cutoff for all analytes used in this study. However, rpart deviated from this threshold in all three cases (17 days for GGT, 7 days for creatinine, and 10 days for AP). CALIPER's selection of a round age boundary (2 weeks) appears intuitive from a human perspective. In contrast, rpart determines age boundaries purely mathematically based on the available data. The lack of interpolation in rpart can lead to gaps in reference intervals, as observed for GGT between the second (10–70 days) and third (78–6939 days) rpart intervals. Therefore, age partitions cannot be directly transferred from rpart for application to other datasets without careful consideration. To evaluate rpart's partitioning suggestions, the age distribution of routine data should be examined for potential gaps or temporal clustering. From a medical perspective, when considering neonatal intervals, the transition between the first and second intervals should always be evaluated in its temporal context and which might be more relevant than changes between hard cut-offs at 7, 10 or 14 days.

For GGT, rpart proposes two significantly narrower partitions for neonates and infants compared to CALIPER. The non-significant CALIPER split at 11 years is avoided by rpart. Otherwise, the respective partitions are very similar when considering the upper and lower limits (see Figure 1D).

For creatinine, it is notable that four unisex CALIPER intervals were combined into two by rpart. In compensation, rpart introduces sex-specific partitioning earlier. Additionally, rpart suggests a broader female partition from 4434 days onward, while males are further subdivided in this age range. This conventionally atypical division into sex-dependent partitions with different age thresholds avoids a non-significant split between the last CALIPER unisex partition and the CALIPER female partition. From the perspective of creatinine's biological background, the use of continuous reference ranges has already been proposed [16]. It is therefore not surprising that individual partitions between rpart and CALIPER can be offset relative to each other in a meaningful way. Comparison of the reference limits of individual partitions between CALIPER and rpart again shows minimal differences.

When comparing the age and sex partitions for AP between CALIPER and rpart, rpart initially appears to be a consolidation of CALIPER with a reduction in the number of partitions. The significant differences between some sex-dependent partitions are equally unsurprising from the graphical comparison.

To assess the validity of the derived reference intervals, we incorporated the corresponding intervals from the original CALIPER study into Subfigures D for direct comparison. In most cases, the indirectly estimated reference intervals, based on rpart or CALIPER partitions, closely matched those reported in the original study. Nonetheless, a few notable discrepancies were observed. Specifically, for GGT in the CALIPER partition covering 15 days to 1 year, and for AP in the partition from birth to 14 days, the indirectly estimated intervals were substantially wider than those established in the healthy CALIPER study population.

Conclusions

These findings support the conclusion that both using reflimR for indirect estimation and applying automated partitioning with rpart can produce clinically meaningful reference intervals. However, the present analysis is limited by its reliance on routine data from a single laboratory.

In addition, this study focused on analytes with relatively straightforward partitioning patterns. Future studies should therefore extend this approach to more challenging analytes, where age- and sex-related stratification is less obvious, to further assess the robustness of the method. To strengthen these conclusions and evaluate their broader applicability, multicenter comparisons across diverse populations and analytical settings will also be required.

Research ethics: Ethical approval was granted by the Ethics Committee of the University of Regensburg under an expedited procedure in accordance with § 15 of the Professional Code of Conduct for Physicians in Bavaria. Following review of the application form and brief project description

(submitted 21 January 2025, reference number: 25-4055-104), the committee found no professional-ethical or legal objections to the conduct of the study.

Informed consent: Informed consent was not necessary due to the retrospective and anonymized data analysis. Not applicable.

Author contributions: J. Böhm and J. Gebauer conceived the study, analysed the data and took the lead in writing the manuscript. B. Rolinski and S. Schneider contributed to the interpretation and provided critical feedback. The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning Tools: ChatGPT and Claude.ai were used for translations, language editing and code refactoring.

Conflict of interest: The authors state no conflict of interest.

Research funding: None declared.

Data availability: Upon request.

References

1. Horowitz GL, Altaie S, Boyd JC, Ceriotti F, Garg G, Horn P, et al. C28-A3c: Defining, establishing, and verifying reference intervals in the Clinical Laboratory; approved guideline – third edition; 28th series, 3rd ed. Wayne, Pennsylvania, USA: Clinical; Laboratory Standards Institute; 2008:30 p.
2. Ichihara K, Boyd JC. An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48: 1537–51.
3. Bohn MK, Adeli K. Application of the TML method to big data analytics and reference interval harmonization. J Lab Med 2021;45: 79–85.
4. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. Clin Chem Lab Med 2019;57:1933–47.
5. Ammer T, Rank CM, Schuetzenmeister A. refineR: reference interval estimation using real-world data; 2023. Available from: https://CRAN.R-project.org/package=refineR [Accessed 29 Apr 2024].
6. Hoffmann G, Klawitter S, Klawonn F. reflimR: reference limit estimation using routine laboratory data; 2024. Available from: https://CRAN.R-project.org/package=reflimR [Accessed 29 Apr 2024].
7. Jones GRD, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, et al. Indirect methods for reference interval determination – review and recommendations. Clin Chem Lab Med 2018;57:20–9.
8. Yang D, Su Z, Zhao M. Big data and reference intervals. Clin Chim Acta 2022;527:23–32.
9. Klawitter S, Böhm J, Tolios A, Gebauer JE. Automated sex and age partitioning for the estimation of reference intervals using a regression tree model. J Lab Med 2024;48:223–37.
10. R Core Team. R: a language and environment for statistical computing; 2023. Available from: https://www.R-project.org/ [Accessed 29 Apr 2024].
11. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. J Open Source Softw 2019;4:1686.
12. Therneau T, Atkinson B. Rpart: recursive partitioning and regression trees; 2023. Available from: https://CRAN.R-project.org/package=rpart [Accessed 29 Apr 2024].
13. Milborrow S. Rpart.plot: plot 'rpart' Models: an Enhanced Version of 'plot.rpart'; 2022. Available from: https://CRAN.R-project.org/package=rpart.plot [Accessed 29 Apr 2024].
14. Lahti A. Partitioning biochemical reference data intosubgroups: Comparison of existing methods. Clin Chem Lab Med 2004;42: 725–33.
15. Colantonio DA, Kyriakopoulou L, Chan MK, Daly CH, Brinc D, Venner AA, et al. Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. Clin Chem 2012;58:854–68.
16. Li K, Hu L, Peng Y, Yan R, Li Q, Peng X, et al. Comparison of four algorithms on establishing continuous reference intervals for pediatric analytes with age-dependent trend. BMC Med Res Methodol 2020;20: 136.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/labmed-2025-0210).