Sandra Klawitter, Johannes Böhm, Alexander Tolios and Julian E. Gebauer*

Automated sex and age partitioning for the estimation of reference intervals using a regression tree model

Abstract

Objectives: Reference intervals (RI) play a decisive role in the interpretation of medical laboratory results. An important step in the determination of RI is age- and sex specific partitioning, which is usually based on an empirical approach by graphical representation. In this study, we evaluate an automated machine learning approach.

Methods: This study uses pediatric data from the CALIPER RI (Canadian Laboratory Initiative on Pediatric Reference Intervals) study. The calculation of potential partitions is carried out using a regression tree model included in the rpart package of the statistical programming language R. The Harris & Boyd method is used to compare the corresponding partitions suggested by rpart and CALIPER. For better comparability, the reference ranges of the partitions of both approaches are then calculated using reflimR.

Results: Most of the partitions suggested by rpart or CALIPER show sufficient heterogeneity among themselves to justify age- and/or sex-specific RI partitioning. With only few individual exceptions, both methods yield comparable results. The partitions of both approaches for albumin and γ-glutamyl-transferase are very similar to each other. For creatinine rpart suggests a slightly earlier distinction between the sexes. Alkaline phosphatase shows the most pronounced differences. In addition to a considerable earlier sex split, rpart suggests different age intervals for both sexes, resulting in three partitions for females and four partitions for males.

Conclusions: Our findings indicate that the automated analysis provided by rpart yields results that comparable to traditional methods. Nevertheless, the medical plausibility of the automatic suggestions needs to be validated by human experts.

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

Introduction

Reference intervals (RI) are an important information for the interpretation of laboratory results. It is the responsibility of medical laboratories to validate the existing reference intervals or to establish their own. According to the CLSI/IFCC standard [1] RI are usually estimated from the central 95 % interval of a healthy population, a method known as the direct RI estimation. Values outside of this interval are classified as decreased or elevated.

An essential issue in determining RI is understanding the factors that influence variations in analyte concentrations [2]. Age, sex and ethnicity have a major impact on biology and are therefore important factors influencing the RI of laboratory parameters. In addition, different analytical methods can lead to deviating RI. For the estimation of RI, a minimum of 120 reference subjects is recommended by the CLSI guidline [1]. However, this was later considered as too low [3]. In principle, the number of individuals is required for each subgroup like sex or age cohort. This results in a considerable number of healthy individuals who must be included in the process to establish direct RI.

One of the major challenges in establishing RI is the partitioning in subgroups. While this is relatively straightforward for sex, it becomes more complex for ethnicity and particularly challenging when dealing with the categorization/binning of numeric variables such as age. Partitioning by age groups is a crucial method for enhancing the diagnostic value of RI [4]. Currently, there are three basic principles for partitioning:

(1) Partitioning by social categories (e.g., newborn, infant, adolescence, adult) [5].

---

*Corresponding author: Dr. med. Julian E. Gebauer, MVZ Labor Krone GbR, Siemensstr. 40, Bad Salzuflen, Germany, E-mail: julian-gebauer@gmx.de. https://orcid.org/0000-0003-0216-3607
Sandra Klawitter, Ostfalia University of Applied Sciences, Wolfenbüttel, Germany
Johannes Böhm, Klinikum Passau, Institut für Labor- und Transfusionsmedizin, Passau, Germany
Alexander Tolios, Department of Transfusion Medicine and Cell Therapy (UTZ 4i), Medical University of Vienna, Vienna, Austria

(2) Partitioning based on data and its biochemical and medical context. This is the most commonly used approach to date [1, 4, 5].

(3) Newer and sophisticated approaches with continuous RI [6–9].

While partitioning by social categories simplifies the classification of values for laypeople, it poorly represents physiological processes, especially in childhood [5]. On the other side, continuous RI have the greatest potential to accurately reflect biological backgrounds. However, the validation and use of continuous RI are currently not covered by normative documents [7]. Therefore the partitioning based on the data with its biological background can be considered as the current standard, which is also used in this work and the CALIPER (Canadian Laboratory Initiative on Pediatric Reference Intervals) study [10].

The aim of this study is to evaluate automated and therefore potentially unbiased tools for age- and sex-dependent segmentation. To that end, the authors chose to utilize decision tree algorithms since they display several properties which are useful when using machine learning-tools in a medical context:

– they are robust to outliers, which makes the results more reliable,

– they are easy to visualize and understand even without any specialized knowledge besides the ability to read flowcharts, and

– they are by default non-opaque, which helps to interpret the results on the basis of medical knowledge (so-called explainable AI).

Theoretical background of the decision tree model

The decision tree algorithm creates a model to assign classes or target values to data. The rpart-package utilize so-called "Classification and Regression Tree (CART)" models, which can be applied for both categorical or numerical targets [11]. The model is computed using training data to learn relationships between the target value (measured value, e.g. albumin concentration) and the data features (age and sex). This is considered "supervised" machine learning, since the algorithm is trained using a separate dataset (the *training set*) with an outcome variable (the measured analyte), and tries to re-create that mapping on a separate dataset (the *test set*) without that variable.

For its decision, the tree utilizes univariate splits by assessing the ANOVA method. ANOVA, short for Analysis of Variance, is used to determine if there are significant differences between the means of multiple groups by comparing the variance within groups to the variance between groups. In regression trees, ANOVA helps to evaluate potential splits for each feature (age and sex) at each node to improve the homogeneity of the resulting groups. It assesses the quality of a split by measuring how much it reduces the variance between the subsets. The best split is the one that significantly reduces the variance, leading to the most homogeneous subsets. This process is repeated recursively for each node in the tree until a stopping criterion e.g. maximum tree size or minimum target group size is reached.

Materials and methods

Selection of analytes

We chose to utilize a set of laboratory values which are measured commonly in a large number of individuals with varying clinical conditions. Those analytes include albumin, γ-glutamyl transferase, creatinine, and alkaline phosphatase.

Albumin (ALB) was chosen due to its relatively constant synthesis rate in the liver, without a known age dependence. The main source of pre-analytical error described is orthostatic redistribution [12]. Therefore we consider it as the "negative control" of our study and do not expect significant age or sex clusters.

The other three analytes were selected specifically for their age-dependent physiological changes. We anticipated finding corresponding physiological processes through a good partitioning procedure.

γ-Glutamyl transferase (GGT) is a marker for hepatobiliary diseases. It mainly undergoes a significant postnatal decline, and although the exact cause is to the best of our knowledge unknown, it is associated with childbirth [13]. Additionally, in adults the upper reference limit of GGT is considered too high due to the extent of socially accepted alcohol consumption. Subsequently, textbooks [12, 14] describe initially a narrowing of the range of the reference interval in childhood, followed by widening during adolescence. This makes it an interesting case for the automated evaluation of reference values.

Creatinine (CREA) is a key parameter for assessing kidney function. It is characterized by low intraindividual with high interindividual variance. This is due to the dependence of serum creatinine on muscle mass as well as meat intake. Thus, we expect age- and sex-dependent partitioning of reference intervals related to the onset of puberty.

Alkaline phosphatase (AP) is divided into at least 15 isoforms and is involved in a variety of metabolic processes [12]. In routine measurements, physiological variants are mainly the liver, bone, and small intestine isoforms. Due to the use of total AP in this study, the small intestine isoform primarily represents a pre-analytical error source related to postprandial increases. However, the bone isoform is particularly relevant because it strongly linked to growth during childhood. Specifically, relevant literature [12] describes a median constant AP until the age of 10, followed by an increase up to the age of 14 (factor 2–3), and then a subsequent decrease. The liver isoform, as a marker for liver and bile duct disease, should not be relevant in this study of a healthy population.

Selection of data

ALB (albumin G for the Bromocresol Green assay), GGT, CREA (enzymatic assay), and AP data were retrieved from Colantonio et al. (2012) [10]. We decided to use these data from the CALIPER study data primarily for the following reasons:

- CALIPER is considered as an established standard concerning pediatric reference intervals [10]. Other sources like textbooks, often inadequately represent reference ranges for children.
- the data were collected from reference persons who were proven to be healthy. Otherwise, the partitioning would be biased by potentially non-healthy measurement results.
- the data from the study are freely available and have been frequently both viewed and reviewed. This makes it easier for others to reproduce and verify our results.

Data analysis

Data analysis was conducted in R (version 4.3.2) [15]. We utilized the R-packages `dplyr` and `tidyr` [16] for data processing, `ggplot2` [17] for data visualization, rpart [18] for partitioning, rpart.plot [19] for visualization of the regression trees, and `reflimR` [20] for the estimation of reference intervals. The differences between partitioning suggested by rpart ord CALIPER were assessed using the Harris & Boyd method as detailed in Lahti 2004 [4]. Among other methods, some of which are considered better for comparing RI partitions [3], we chose to use the Harris and Boyd method to achieve better comparability with the CALIPER study. Mean, standard deviation, and sample size

for Harris & Boyd method were computed from their respective partitions. In those cases where the sample size of the subgroup was too small for a valid RI estimation with `reflimR`, the central 95 % interval was used as a proxy. In contrast to the CALIPER study, we opted for the consistent calculation of RI using `reflimR` to enhance comparability. The 'reflimR'-package provides a method for estimating reference intervals using truncated quantile-quantile plots. Although the method is intended for the use as an indirect method on routine laboratory data, it also provides very robust results for data from a direct appraoch. The R code used for the data analysis is available on github (https://github.com/gebauerj/ri_partitioning_rpart). In addition, an interactive application using the R-package `shiny`, which allows the reader to re-perform our analyses with their own data, is also available on github (https://github.com/SandraKla/AdRI_rpart).

Results

Figures 1 to 4 present a visual summary of the results obtained for the four analytes: ALB, GGT, CREA, and AP, respectively. Each Figure is subdivided into three parts. Subfigure A displays the age and sex partitions recommended by CALIPER alongside the RI estimated using `reflimR`. Subfigure B shows the same data utilizing the partitioning suggested by rpart. The RI are shown as boxes for girls in red, boys in blue, or for the combined sexes in black. RI calculated as percentiles (in the case of small sample sizes) are indicated by dashed lines. Subfigure C illustrates the outcomes of the rpart algorithm in the form of a decision tree. For each split, the response variable value of the model for respective node or leaf is displayed together with the number of observations and the percentage of the total observations. The underlying absolute numbers for reference limits and age partitions can be found in Tables 1 and 2. The results of the comparisons of specific partitions using the Harris & Boyd method are shown in Table 3.

For ALB, rpart suggests two very simple partitions, which only consider age (see Figure 2). These the splits occur at approx. one and ten years. In contrast, CALIPER suggests a sex-dependent subdivision from 15 years onwards, with male individuals having a much broader scattering than female individuals. As a result, in older age groups, the CALIPER reference intervals cover more data points than rpart, although when comparing the intervals, the Harris & Boyd method again do not indicate a significant difference

(C)
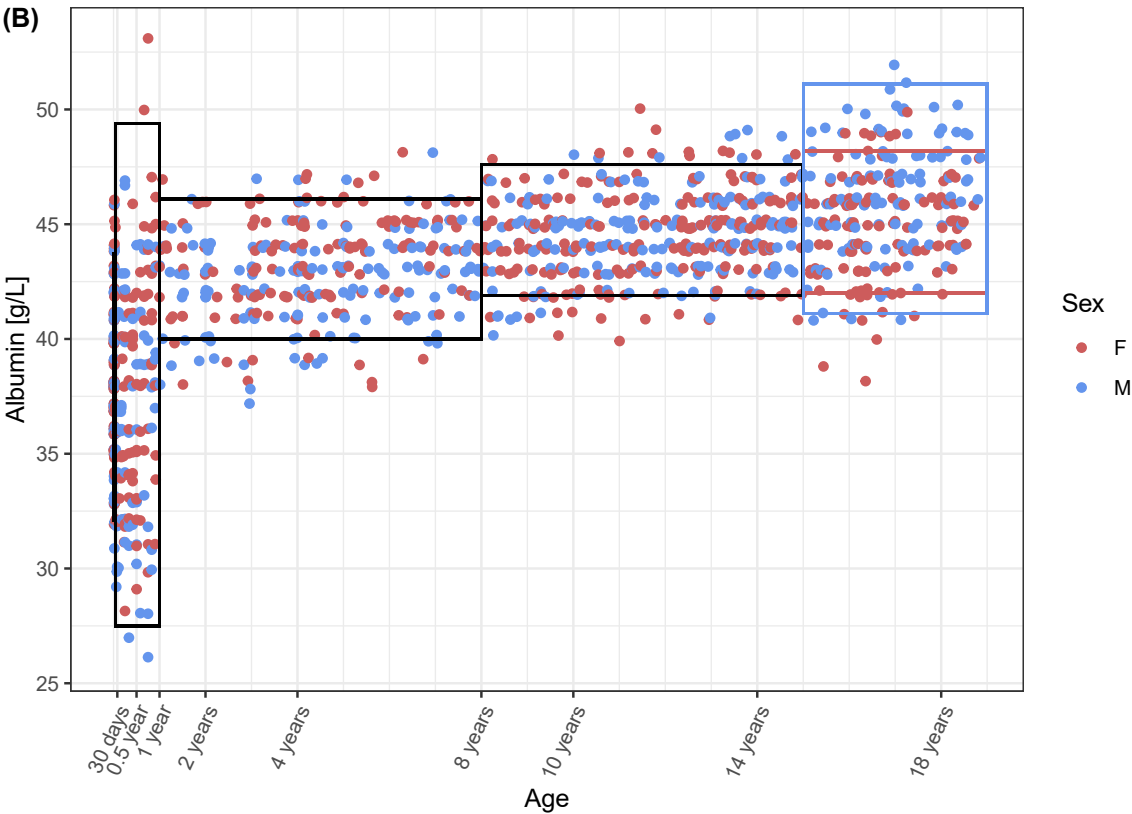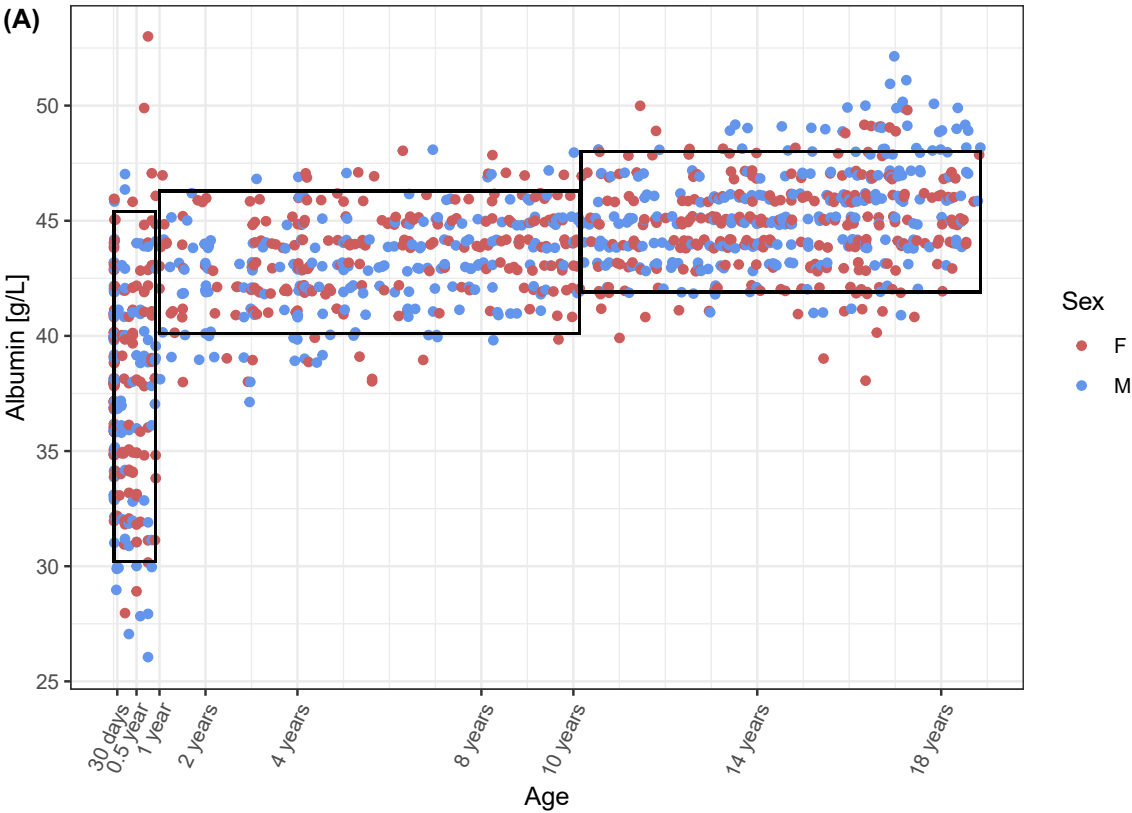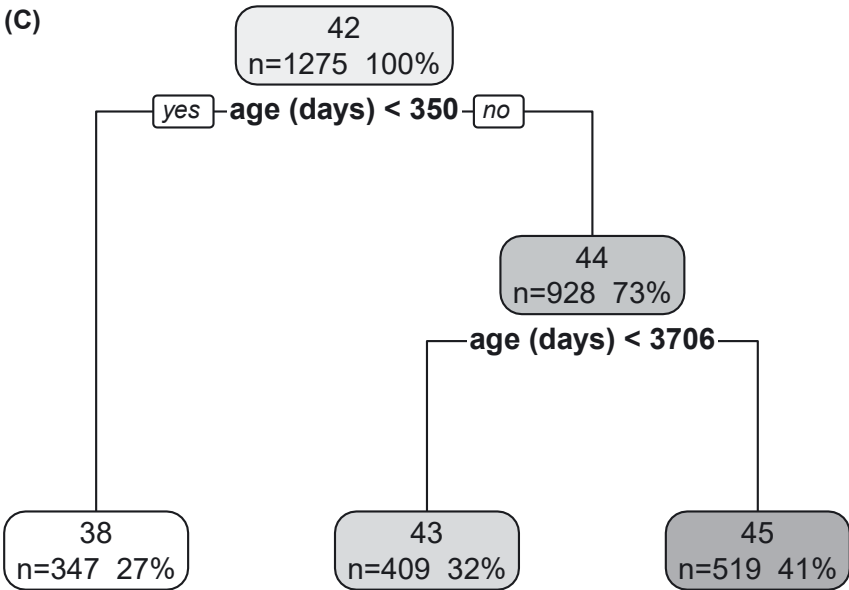
Figure 1: Analysis of albumin (ALB). (A) RI of partitions suggested by CALIPER for ALB. (B) RI for partitions suggested by rpart for ALB. (C) Flow chart for age- and sex-dependent reference intervals for ALB.

between the corresponding rpart partition and the two CALIPER partitions for boys and girls.

The analysis of GGT also reveals very similar results with both approaches. Significant age splits are seen at about one month and one year, and again CALIPER suggests an additional split at an age of 11 years, which is missing in rpart. According to the Harris & Boyd method, there is no statistical need for this subdivision.

(A)

(B)

(C)

Figure 2: Analysis of γ-glutamyltransferase (GGT). (A) RI of partitions suggested by CALIPER for GGT. (B) RI for partitions suggested by rpart for GGT. (C) Flow chart for age- and sex-dependent reference intervals for GGT.

For the remaining age progression, rpart assumes a constant reference interval for both sexes, while CALIPER suggests an additional split at 11 years with a particular effect on young males. According to the Harris & Boyd method, there is no significant difference between the partitions suggested by rpart or CALIPER, even if the

(A)

(B)

(C)

Figure 3: Analysis of creatinine (CREA). (A) RI of partitions suggested by CALIPER for CREA. (B) RI for partitions suggested by rpart for CREA. (C) Flow chart for age- and sex-dependent reference intervals for CREA.

CALIPER proposal looks somewhat closer to the blue dots on the right-hand side of the graph.

Both approaches suggest quite similar partitions for CREA, albeit with greater complexity of age partitioning.

This is due to the fact that creatinine levels (in contrast to ALB and GGT) increase more or less continuously from infancy to adulthood, which is difficult to depict using constant age groupings. In view of this challenge, the position of the

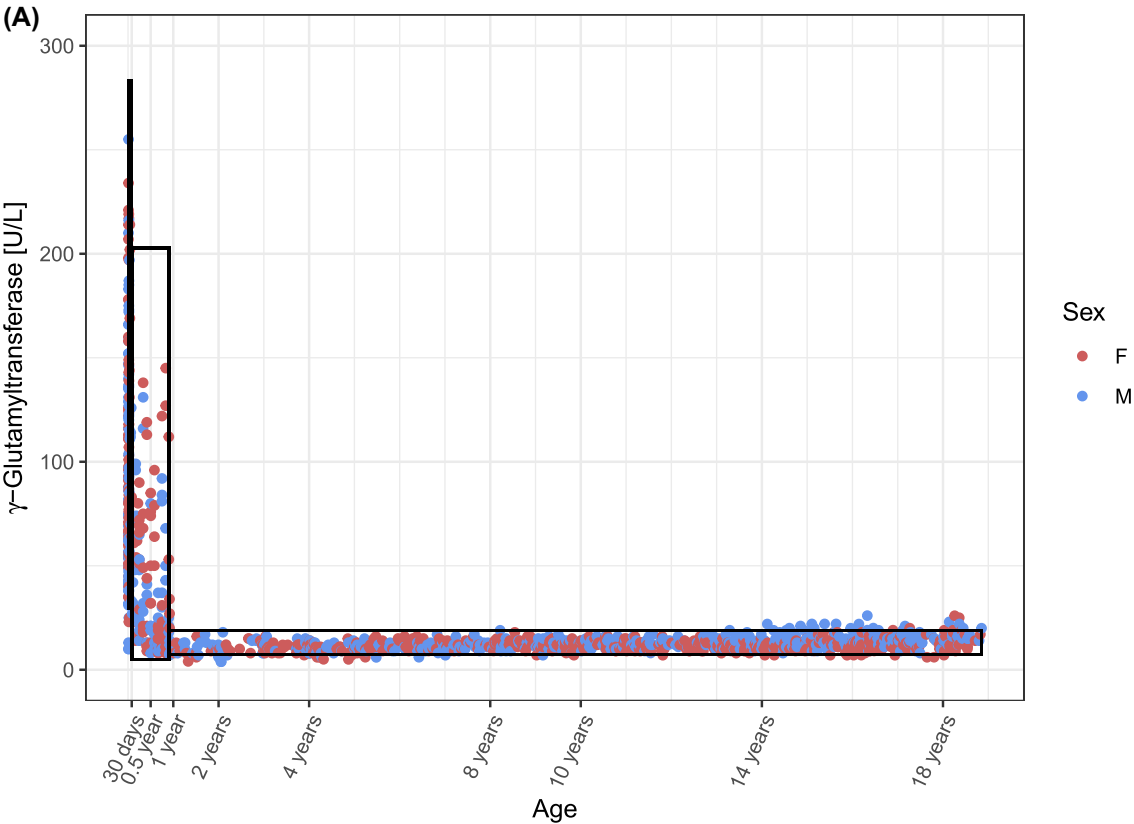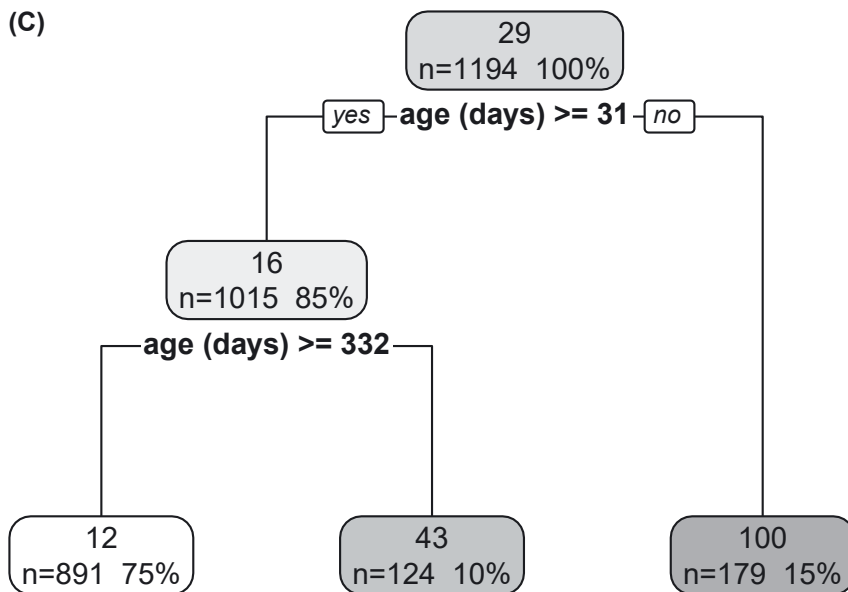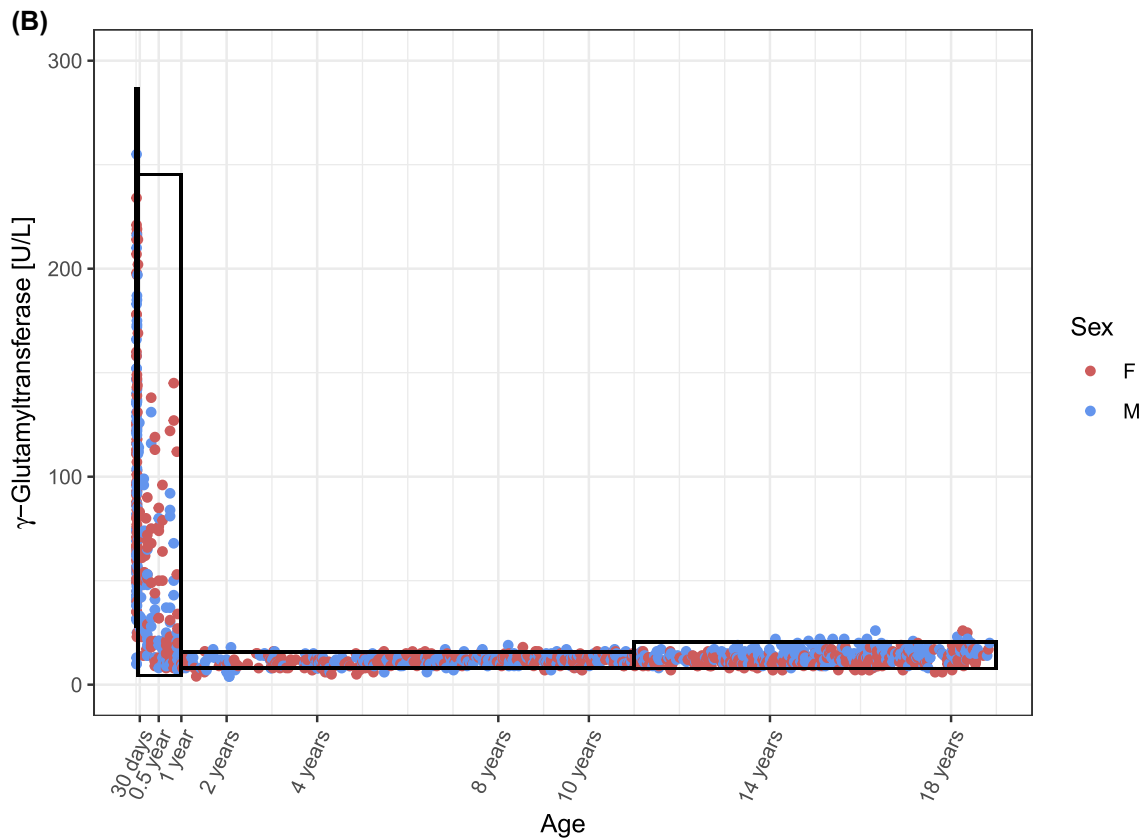**(A)**

(B)

(C)
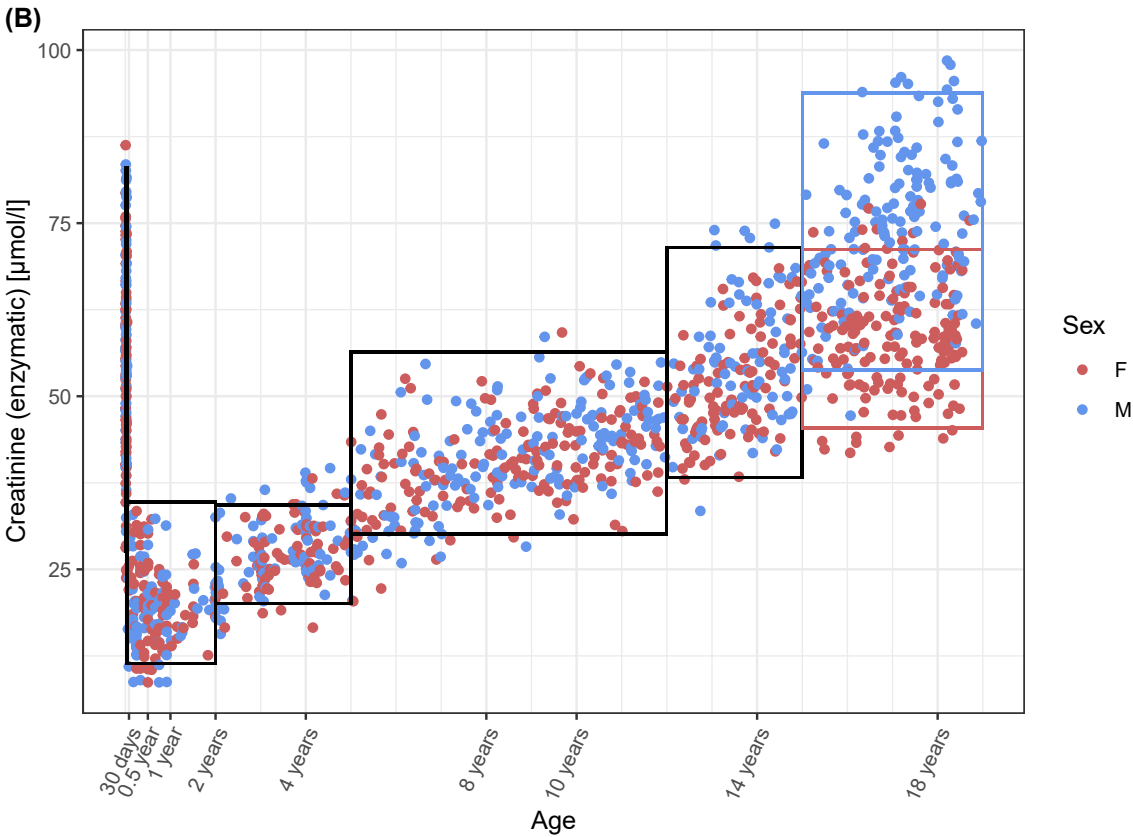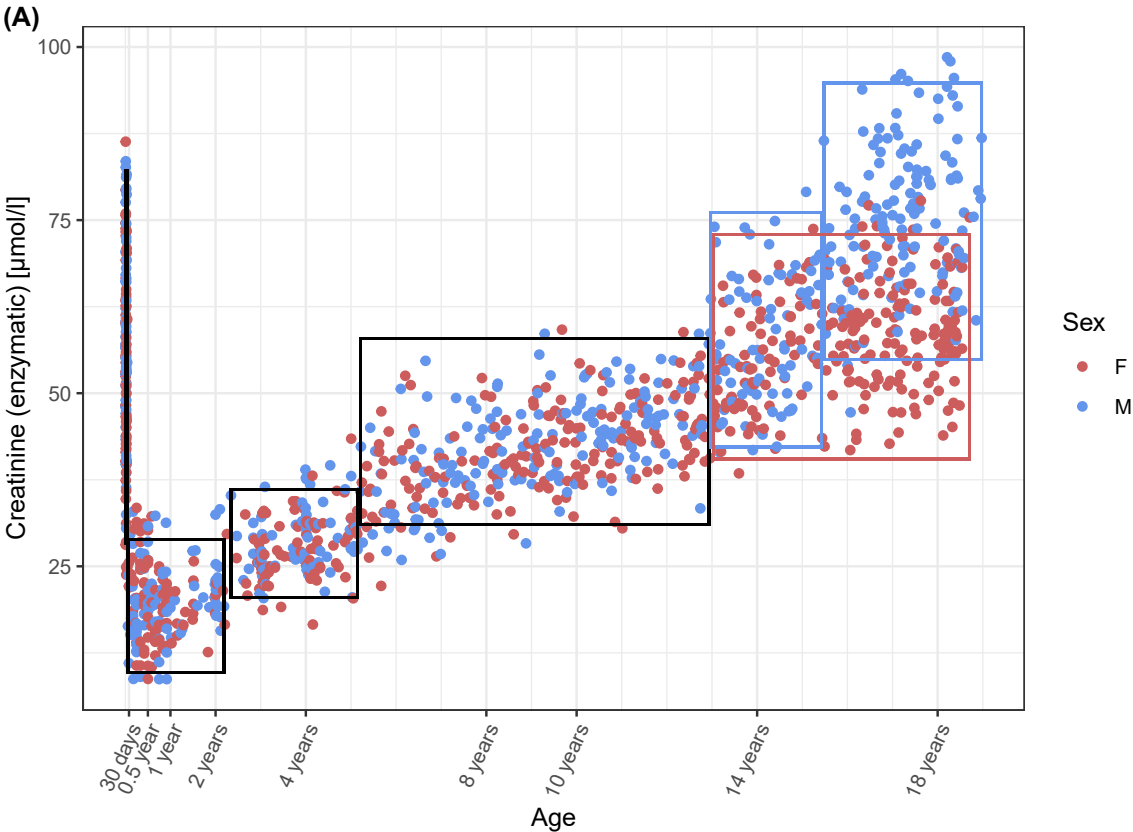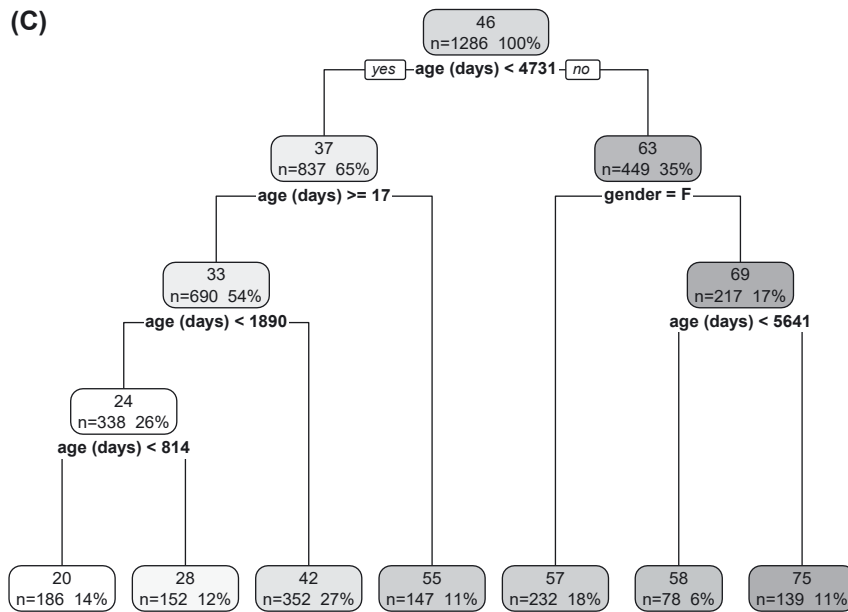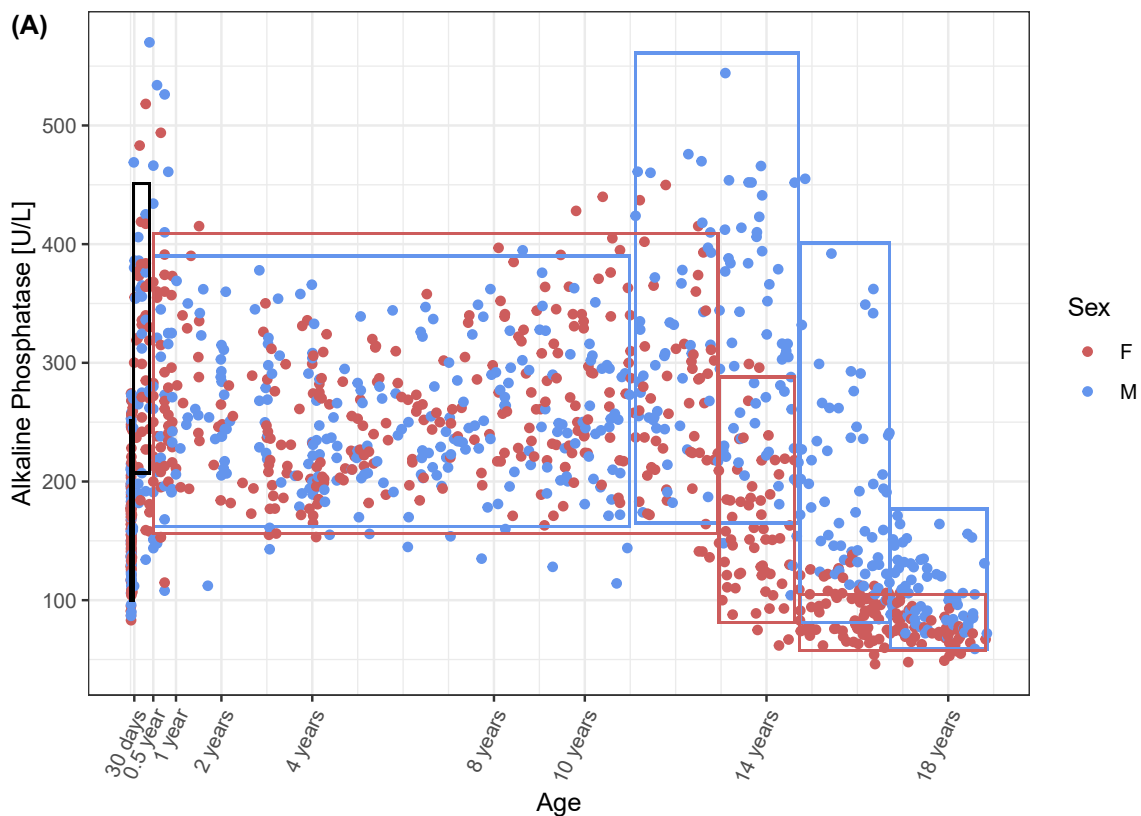
Figure 4: Analysis of alkaline phosphatase (AP). (A) RI of partitions suggested by CALIPER for AP. (B) RI for partitions suggested by rpart for AP. (C) Flow chart for age- and sex-dependent reference intervals for AP.
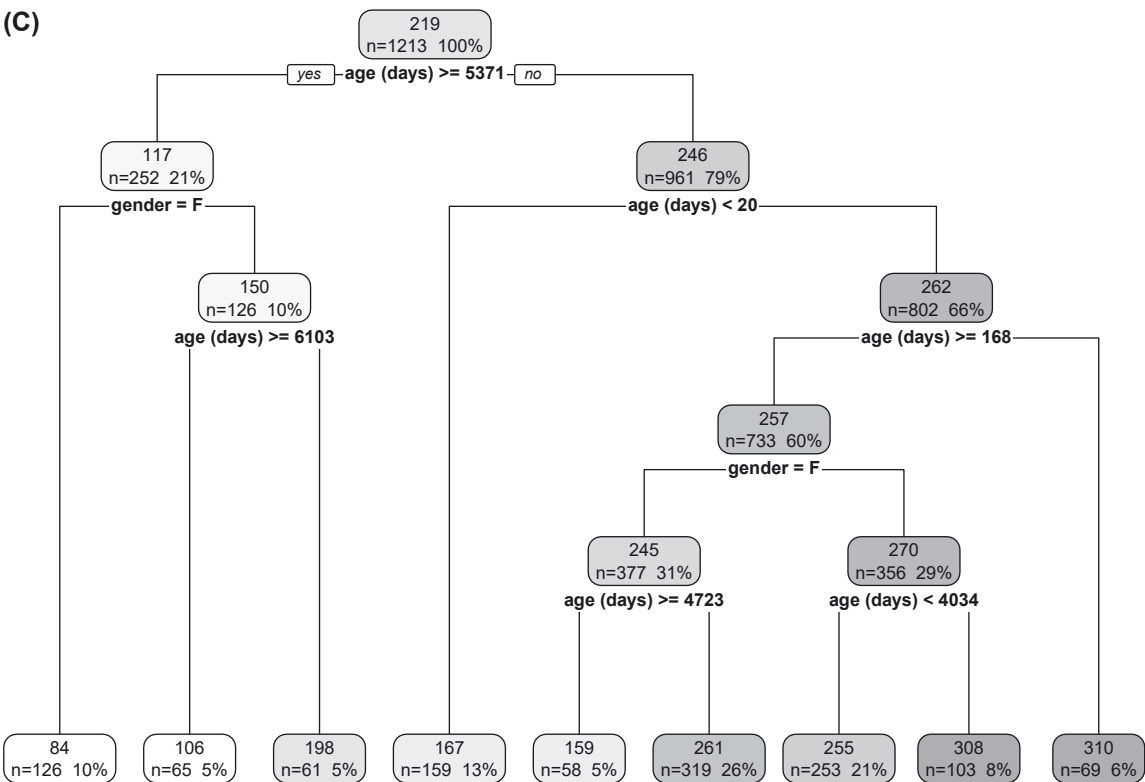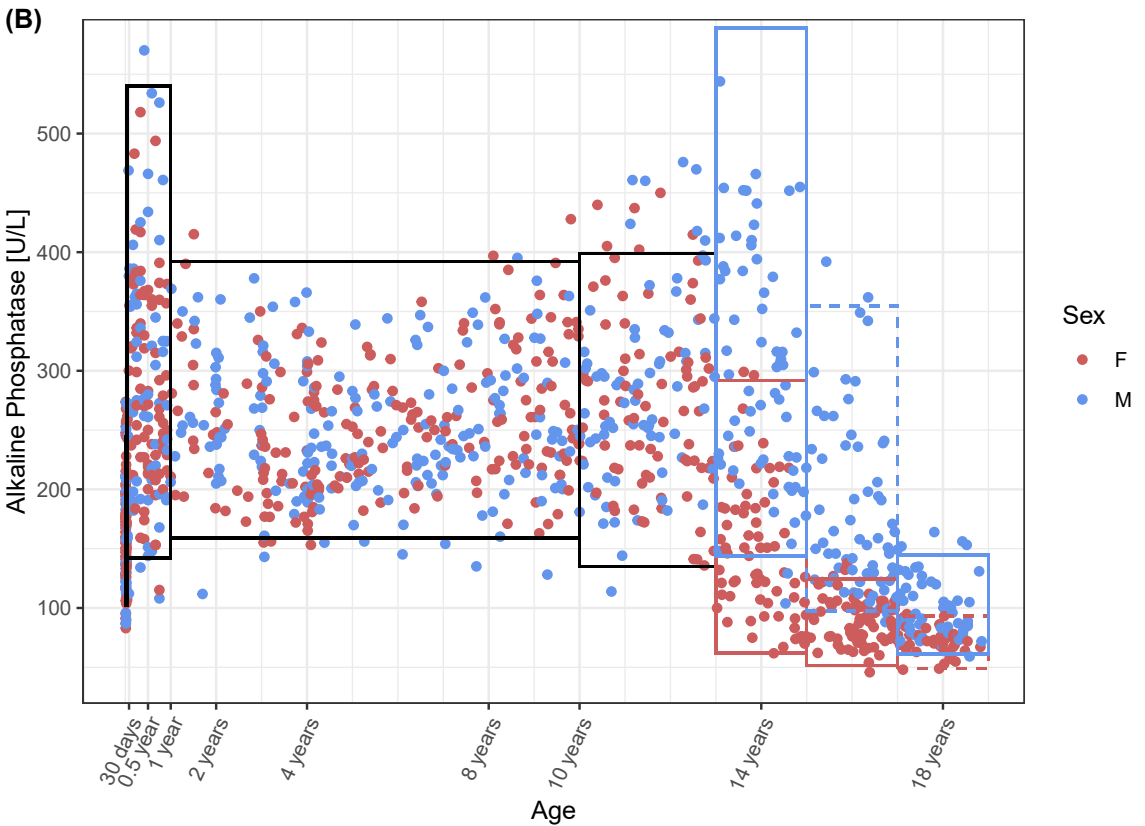
age splits proposed by rpart matches that of CALIPER remarkably well. For children older than 13 years, rpart suggests only one partition for girls but two for boys with a further subdivision at 15.5 years. CALIPER, on the other hand, suggests an additional sex-independent partition and then only divides into separate partitions for boys and girls from 15 years onwards. Again, significant differences can be observed between the individual segments within both approaches, with no significant differences between the two approaches in the Harris & Boyd analysis.

AP is the most complex analyte examined here. It is therefore not surprising that the clearest differences between the two approaches can be seen here. The complexity is primarily due to the strong age and sex-dependent kinetics of AP. In the first year of life, there is initially a significant increase in AP with considerable variability, followed by a rapid decline to a relatively stable level, which persists until around the age of 10. Subsequently, there is a noticeable sex differentiation associated with puberty. Both approaches attempt to map this complex process as effectively as possible. In this regard, the solution from CALIPER appears clearer due to largely sex-independent partitions, while the rpart RI cover more data points.

When comparing the partitions within the approaches, we found that there are no significant differences between the successive CALIPER segments from 15 days up to an age of 13 years. Similarly, there are no significant differences in the transition to the first CALIPER segment for boys aged 13 years and older. In contrast, all subdivisions within rpart are significant. When comparing both approaches, there is initially no significant difference between the sex-dependent rpart segments and the unisex CALIPER segments up to the age of 13 years. As expected, the comparison of the male CALIPER and rpart segments shows no significant differences. However, when comparing the female segments, a significant difference is found between the last two segments of CALIPER and rpart.

Discussion

This proof-of-principle study supports our hypothesis that automated tools like the decision tree algorithm implemented in rpart can be used in the generation of reference intervals. The partitioning suggestions of the machine learning algorithm matches the judgment of human expertise surprisingly well, even for complex age progressions, and therefore appear suitable to estimate age ranges for calculating RI.

Instead of visually inspecting the data as in CALIPER [10], rpart relies on a purely mathematical basis to calculate the age and sex partitions. According to the used outcome metric (the Harris & Boyd's test), all partitions suggested by rpart differed significantly from each other. This is an improvement over the visual and thus more subjective CALIPER results, which according to the Harris & Boyd's test was sometimes prone to oversegmentation.

For the analytes GGT, ALB and CREA, the automatic suggestions are in good agreement with their biological background and generally agree with the assessments made by human experts. For AP, an increase in the variance of the measured values between 10 and 14 years is also described in the literature [12]. It is noticeable that the rpart partitions reflect this kinetics better than the CALIPER partitions (see Figure 4, fewer data points outside the RI of rpart compared to CALIPER). In addition, there are problems with small sample sizes for the reflimR method within the two dashed CALIPER partitions and the non-significant (non-)differences between several subsequent CALIPER partitions (as described in the results section). Therefore, the partitioning performed by rpart seems to be more stringent than the one done by CALIPER.

A notable difference between the suggested partitions from rpart and CALIPER is that rpart sometimes suggests gender-specific partitions with clearly differing age ranges. This is particularly noticeable for creatinine (CREA) (Figure 3B) and alkaline phosphatase (AP) (Figure 4B). From a mathematical perspective, rpart aims to minimize the number of partitions. Conversely, humans might tend to create sex-specific partitions with identical age intervals during visual partitioning. We observed that some consecutive partitions in CALIPER do not show significant differences. Therefore, we consider partitioning suggested by rpart to be mathematically superior. However, for clinical use, this must be evaluated within the biological and clinical context.

The Harris & Boyd method is currently regarded as an established method for the comparison of reference intervals [1], although some methodological problems have been described [4]. One of them is that Harris & Boyd's test assumes an approximate normal distribution of the data. The problem with skewed data is briefly mentioned in the original paper, but not discussed further [21]. A critical evaluation of the precise statistical problems of the methodology is beyond the scope of this publication. For several reasons, Haeckel and Wosniok [22] argue that there are almost no normally distributed analytes in clinical chemistry and propose that all quantities should be considered lognormally distributed. To briefly address the main features of the problem, we used the lognormally

distributed analyte GGT to check whether the logarithmization of the measured values in our analysis leads to a different interpretation. In that case rpart suggested one additional age segment without any medical relevance, which is why we refrained from further consideration of the problem (data not shown).

With regression tree algorithms, however, there are other parameters influencing the results. Firstly, the sample size can have a major impact. Decision trees are known to easily overfit to their training set [23, 24]. In this case, tuning of the hyperparameters is required to ensure that the individual partitions reach an appropriate size. This process is also known as pruning. In particular, parameters such as the minimum number of data points per partition or the maximum number of branches need to be adapted so that non-relevant and redundant parts of the tree are removed.

It should be noted that the actual boundaries of the partitions in the decision trees of rpart are calculated from the mean value of the limit values between the two partitions. While this leads to more stringent reference interval estimations, the algorithm doesn't interpolate in case of missing values, which might lead to gaps in the RI range. To illustrate this, we intentionally plotted the rpart boxes in the scatter plots with the minimum and maximum age values of each age partition to reveal any gaps in the calculation. In our case, the data were relatively evenly distributed across age, but when interpreting with rpart, one should be aware of possible biases due to underrepresented age intervals.

In addition, we want to stress that without careful preparation of the real-world data, analyses like this one cannot successfully be performed [2]. After all, the data from the routine laboratory requires a clean structure comparable to that of the CALIPER data in order to obtain meaningful results. On one hand, there are technical challenges such as the extraction of data from the laboratory information system (LIS) with potential subsequent anonymization. On the other hand, there are also medical issues to consider, such as how to deal with metadata like a "sample is hemolytic"-flag or the "measured value is below the detection limit"-information.

Our study demonstrates that rpart can be used for establishing reference intervals in an unbiased and mathematically rigorous fashion. This approach offers a valuable tool and can provide some guidance when trying to establish partitions in reference intervals. Thus, we underscore the importance of understanding the biochemical and statistical context. Furthermore, all derived conclusions should be carefully validated within their specific context.

Appendix

Table 1: Age limits (in days) and reference intervals from the CALIPER study.

Analyte	Unit	Sex	Min age	Max age	Lower limit	Upper limit
ALB	g/L	M/F	0	14	33	45
ALB	g/L	M/F	15	364	28	47
ALB	g/L	M/F	365	2919	38	47
ALB	g/L	M/F	2920	5474	41	48
ALB	g/L	M	5475	6934	41	51
ALB	g/L	F	5475	6934	40	49
GGT	U/L	M/F	0	14	23	219
GGT	U/L	M/F	15	364	8	127
GGT	U/L	M/F	365	4014	6	16
GGT	U/L	M/F	4015	6934	7	21
CREA	µmol/L	M/F	0	14	28	81
CREA	µmol/L	M/F	15	729	9	32
CREA	µmol/L	M/F	730	1824	18	38
CREA	µmol/L	M/F	1825	4379	27	54
CREA	µmol/L	M/F	4380	5474	40	72
CREA	µmol/L	M	5475	6934	55	95
CREA	µmol/L	F	5475	6934	43	74
AP	U/L	M/F	0	14	90	273
AP	U/L	M/F	15	364	134	518
AP	U/L	M/F	365	3649	156	369
AP	U/L	M/F	3650	4744	141	460
AP	U/L	M	4745	5474	127	517
AP	U/L	M	5475	6204	89	365
AP	U/L	M	6205	6934	59	164
AP	U/L	F	4745	5474	62	280
AP	U/L	F	5475	6204	54	128
AP	U/L	F	6205	6934	48	95

Table 2: Age limits (in days), reference intervals with permissible uncertainty (PU) and descriptive statistics of each partition.

	Unit	Approach	Sex	Min age	Max age	Lower limit	Upper limit	Lower limit PU	Upper limit PU	Method	Mean	SD	Truncated n
ALB	g/L	rpart	M/F	4	335	30.20	45.40	28.9–31.5	43.6–47.2	reflimR	37.70	4.22	338
ALB	g/L	rpart	M/F	365	3699	40.10	46.30	39.1–41.1	45.2–47.4	reflimR	43.24	2.05	408
ALB	g/L	rpart	M/F	3711	6881	41.90	48.00	40.9–42.9	46.9–49.1	reflimR	45.09	2.12	519
ALB	g/L	CALIPER	M/F	0	14	32.10	43.70	30.9–33.3	42.2–45.2	reflimR	38.18	3.10	188
ALB	g/L	CALIPER	M/F	15	364	27.50	49.40	26.1–28.9	47.1–51.7	reflimR	37.14	5.13	159
ALB	g/L	CALIPER	M/F	365	2919	40.00	46.10	39–41	45–47.2	reflimR	42.95	2.07	298
ALB	g/L	CALIPER	M/F	2920	5474	41.90	47.60	41–42.8	46.6–48.6	reflimR	44.53	1.82	388
ALB	g/L	CALIPER	M	5475	6934	41.10	51.10	39.9–42.3	49.6–52.6	reflimR	46.11	2.37	123
ALB	g/L	CALIPER	F	5475	6934	42.00	48.20	41–43	47.1–49.3	reflimR	44.83	2.23	119
GGT	U/L	rpart	M/F	3	29	29.70	283.40	25.4–34	258.6–308.2	reflimR	99.68	52.64	179
GGT	U/L	rpart	M/F	33	329	4.90	202.80	3.46–6.34	179.09–226.51	reflimR	43.06	33.84	124
GGT	U/L	rpart	M/F	335	6881	7.40	18.90	6.88–7.92	17.8–20	reflimR	12.25	3.29	891
GGT	U/L	CALIPER	M/F	0	14	28.00	286.50	23.8–32.2	261.1–311.9	reflimR	98.40	52.43	167
GGT	U/L	CALIPER	M/F	15	364	4.40	245.40	2.89–5.91	215.08–275.72	reflimR	47.52	41.23	145
GGT	U/L	CALIPER	M/F	365	4014	8.00	15.60	7.55–8.45	14.83–16.37	reflimR	11.25	2.57	438
GGT	U/L	CALIPER	M/F	4015	6934	7.90	20.60	7.34–8.46	19.39–21.81	reflimR	13.17	3.41	444
CREA	µmol/L	rpart	M/F	3	15	28.30	82.20	26.1–30.5	77.1–87.3	reflimR	55.34	13.53	147
CREA	µmol/L	rpart	M/F	22	799	9.70	28.90	8.95–10.45	27.1–30.7	reflimR	20.02	5.60	183
CREA	µmol/L	rpart	M/F	855	1878	20.50	36.10	19.4–21.6	34.4–37.8	reflimR	27.92	4.70	151
CREA	µmol/L	rpart	M/F	1902	4720	31.00	57.90	29.3–32.7	55.1–60.7	reflimR	42.16	6.79	350
CREA	µmol/L	CALIPER	M/F	0	14	27.60	83.00	25.4–29.8	77.8–88.2	reflimR	55.31	13.61	145
CREA	µmol/L	CALIPER	M/F	15	729	11.40	34.70	10.5–12.3	32.5–36.9	reflimR	20.26	6.93	170
CREA	µmol/L	CALIPER	M/F	730	1824	20.10	34.30	19.1–21.1	32.8–35.8	reflimR	27.07	4.85	155
CREA	µmol/L	CALIPER	M/F	1825	4379	30.10	56.40	28.5–31.7	53.7–59.1	reflimR	40.84	6.93	321
CREA	µmol/L	CALIPER	M/F	4380	5474	38.30	71.50	36.2–40.4	68.1–74.9	reflimR	53.22	8.32	183
CREA	µmol/L	rpart	M	4733	5632	42.20	76.10	40–44.4	72.5–79.7	reflimR	57.64	8.95	78
CREA	µmol/L	rpart	M	5650	6926	54.90	94.80	52.1–57.7	90.5–99.1	reflimR	75.48	10.26	139
CREA	µmol/L	CALIPER	M	5475	6934	53.80	93.80	51.1–56.5	89.5–98.1	reflimR	74.54	10.56	151
CREA	µmol/L	rpart	F	4755	6829	40.50	72.90	38.4–42.6	69.5–76.3	reflimR	57.17	7.93	232
CREA	µmol/L	CALIPER	F	5475	6934	45.40	71.20	43.4–47.4	68.3–74.1	reflimR	58.66	7.72	161
AP	U/L	rpart	M/F	4	18	100.00	251.00	93.1–106.9	236.6–265.4	reflimR	168.05	45.32	152
AP	U/L	rpart	M/F	22	152	207.00	451.00	194–220	427–475	reflimR	310.08	92.72	69
AP	U/L	CALIPER	M/F	0	14	102.00	244.00	95.2–108.8	230.3–257.7	reflimR	165.90	43.83	153
AP	U/L	CALIPER	M/F	15	364	142.00	540.00	129–155	503–577	reflimR	287.15	94.87	149
AP	U/L	CALIPER	M/F	365	3649	159.00	392.00	148–170	370–414	reflimR	253.82	56.71	391
AP	U/L	CALIPER	M/F	3650	4744	135.00	399.00	125–145	374–424	reflimR	277.56	77.60	154
AP	U/L	rpart	M	183	4010	162.00	390.00	151–173	368–412	reflimR	255.34	66.69	253
AP	U/L	rpart	M	4058	5367	165.00	561.00	151–179	524–598	reflimR	307.73	90.30	103
AP	U/L	rpart	M	5380	6098	81.00	401.00	72.6–89.4	371.3–430.7	reflimR	197.72	80.05	61
AP	U/L	rpart	M	6108	6881	59.00	177.00	54.4–63.6	166–188	reflimR	105.92	28.01	65
AP	U/L	CALIPER	M	4745	5474	144.00	589.00	131–157	548–630	reflimR	298.27	98.87	66
AP	U/L	CALIPER	M	5475	6204	97.30	354.52	88.8–105.8	330.7–378.3	quantile	179.47	73.27	64
AP	U/L	CALIPER	M	6205	6934	61.20	144.80	57.1–65.3	136.7–152.9	reflimR	101.44	26.01	54
AP	U/L	rpart	F	183	4719	156.00	409.00	145–167	385–433	reflimR	260.93	64.79	319
AP	U/L	rpart	F	4727	5333	81.00	288.00	74–88	268.8–307.2	reflimR	158.67	54.32	58
AP	U/L	rpart	F	5374	6871	57.60	104.70	54.5–60.7	99.8–109.6	reflimR	83.75	18.72	126
AP	U/L	CALIPER	F	4745	5474	62.00	292.00	55.7–68.3	270.7–313.3	reflimR	146.26	56.40	68
AP	U/L	CALIPER	F	5475	6204	51.70	124.60	48.2–55.2	117.6–131.6	reflimR	88.97	18.88	74
AP	U/L	CALIPER	F	6205	6934	48.98	93.05	46.3–51.7	88.5–97.6	quantile	71.72	11.38	40

Table 3: Results of the comparisons of different partitions (age in days, mean and SD of the analyte) using the Harris & Boyd method. A z-score higher than the critical value (crit.val) indicates significant differences between two partitions.

Analyte	Partition 1							Partition 2							Harris & Boyd		
	Approach	Min age	Max age	Sex	Mean	SD	Trunc n	Approach	min age	Max age	Sex	Mean	SD	Trunc n	z.score	crit.val	Signif
ALB	CALIPER	0	14	M/F	38.18	3.10	188	CALIPER	15	364	M/F	37.14	5.13	159	2.23	3.61	
ALB	CALIPER	15	364	M/F	37.14	5.13	159	CALIPER	365	2919	M/F	42.95	2.07	298	13.71	4.14	a
ALB	CALIPER	365	2919	M/F	42.95	2.07	298	CALIPER	2920	5474	M	44.53	1.82	388	10.39	5.07	a
ALB	CALIPER	2920	5474	M/F	44.53	1.82	388	CALIPER	5475	6934	M	46.11	2.37	123	6.82	4.38	a
ALB	CALIPER	2920	5474	M/F	44.53	1.82	388	CALIPER	5475	6934	F	44.83	2.23	119	1.37	4.36	
ALB	rpart	4	335	M/F	37.70	4.22	338	rpart	365	3699	M/F	43.24	2.05	408	22.09	5.29	a
ALB	rpart	365	3699	M/F	43.24	2.05	408	rpart	3711	6881	M/F	45.09	2.12	519	13.43	5.90	a
ALB	rpart	4	335	M/F	37.70	4.22	338	CALIPER	0	14	M/F	38.18	3.10	188	1.50	4.44	
ALB	rpart	4	335	M/F	37.70	4.22	338	CALIPER	15	364	M/F	37.14	5.13	159	1.19	4.32	
ALB	rpart	365	3699	M/F	43.24	2.05	408	CALIPER	365	2919	M/F	42.95	2.07	298	1.81	5.15	
ALB	rpart	3711	6881	M/F	45.09	2.12	519	CALIPER	2920	5474	M/F	44.53	1.82	388	4.28	5.83	
ALB	CALIPER	5475	6934	M	46.11	2.37	123	CALIPER	5475	6934	F	44.83	2.23	119	4.34	3.01	a
ALB	rpart	3711	6881	M/F	45.09	2.12	519	CALIPER	5475	6934	M	46.11	2.37	123	4.40	4.91	a
ALB	rpart	3711	6881	M/F	45.09	2.12	519	CALIPER	5475	6934	F	44.83	2.23	119	1.14	4.89	
GGT	CALIPER	0	14	M/F	98.40	52.43	167	CALIPER	15	364	M/F	47.52	41.23	145	9.59	3.42	a
GGT	CALIPER	15	364	M/F	47.52	41.23	145	CALIPER	365	4014	M/F	11.25	2.57	438	10.59	4.68	a
GGT	CALIPER	365	4014	M/F	11.25	2.57	438	CALIPER	4015	6934	M/F	13.17	3.41	444	9.47	5.75	a
GGT	rpart	3	29	M/F	99.68	52.64	179	rpart	33	329	M/F	43.06	33.84	124	11.39	3.37	a
GGT	rpart	33	329	M/F	43.06	33.84	124	rpart	335	6881	M/F	12.25	3.29	891	10.13	6.17	a
GGT	rpart	3	29	M/F	99.68	52.64	179	CALIPER	0	14	M/F	98.40	52.43	167	0.23	3.60	
GGT	rpart	3	29	M/F	99.68	52.64	179	CALIPER	15	364	M/F	47.52	41.23	145	10.00	3.49	a
GGT	rpart	33	329	M/F	43.06	33.84	124	CALIPER	15	364	M/F	47.52	41.23	145	0.97	3.18	
GGT	rpart	335	6881	M/F	12.25	3.29	891	CALIPER	365	4014	M/F	11.25	2.57	438	6.06	7.06	
GGT	rpart	335	6881	M/F	12.25	3.29	891	CALIPER	4015	6934	M/F	13.17	3.41	444	4.70	7.08	
CREA	CALIPER	0	14	M/F	55.31	13.61	145	CALIPER	15	729	M/F	20.26	6.93	170	28.06	3.44	a
CREA	CALIPER	15	729	M/F	20.26	6.93	170	CALIPER	730	1824	M/F	27.07	4.85	155	10.33	3.49	a
CREA	CALIPER	730	1824	M/F	27.07	4.85	155	CALIPER	1825	4379	M/F	40.84	6.93	321	25.07	4.22	a
CREA	CALIPER	1825	4379	M/F	40.84	6.93	321	CALIPER	4380	5474	M/F	53.22	8.32	183	17.05	4.35	a
CREA	CALIPER	5475	6934	M/F	74.54	10.56	151	CALIPER	5475	6934	F	58.66	7.72	161	15.08	3.42	a
CREA	rpart	3	15	M/F	55.34	13.53	147	rpart	22	799	M/F	20.02	5.60	183	29.69	3.52	a
CREA	rpart	22	799	M/F	20.02	5.60	183	rpart	855	1878	M/F	27.92	4.70	151	14.02	3.54	a
CREA	rpart	855	1878	M/F	27.92	4.70	151	rpart	1902	4720	M/F	42.16	6.79	350	27.02	4.33	a
CREA	rpart	3	15	M/F	55.34	13.53	147	CALIPER	0	14	M/F	55.31	13.61	145	0.02	3.31	
CREA	rpart	22	799	M/F	20.02	5.60	183	CALIPER	15	729	M/F	20.26	6.93	170	0.36	3.64	
CREA	rpart	855	1878	M/F	27.92	4.70	151	CALIPER	730	1824	M/F	27.07	4.85	155	1.56	3.39	
CREA	rpart	1902	4720	M/F	42.16	6.79	350	CALIPER	1825	4379	M/F	40.84	6.93	321	2.50	5.02	
CREA	rpart	4755	6829	F	57.17	7.93	232	CALIPER	4380	5474	M/F	53.22	8.32	183	4.90	3.94	a
CREA	rpart	4733	5632	M	57.64	8.95	78	CALIPER	4380	5474	M/F	53.22	8.32	183	3.73	3.13	a

Table 3: (continued)

Analyte	Partition 1							Partition 2							Harris & Boyd		
	Approach	Min age	Max age	Sex	Mean	SD	Trunc n	Approach	min age	Max age	Sex	Mean	SD	Trunc n	z.score	crit.val	Signif
CREA	rpart	1902	4720	M/F	42.16	6.79	350	rpart	4733	5632	M	57.64	8.95	78	14.39	4.01	[a]
CREA	rpart	4733	5632	M	57.64	8.95	78	rpart	5650	6926	M	75.48	10.26	139	13.35	2.85	[a]
CREA	rpart	5650	6926	M	75.48	10.26	139	CALIPER	5475	6934	M	74.54	10.56	151	0.77	3.30	
CREA	rpart	1902	4720	M/F	42.16	6.79	350	rpart	4755	6829	F	57.17	7.93	232	23.65	4.67	[a]
CREA	CALIPER	4755	6829	F	57.17	7.93	232	CALIPER	5475	6934	F	58.66	7.72	161	1.86	3.84	
AP	CALIPER	0	14	M/F	165.90	43.83	153	CALIPER	15	364	M/F	287.15	94.87	149	14.20	3.37	[a]
AP	CALIPER	15	364	M/F	287.15	94.87	149	CALIPER	365	3649	M/F	253.82	56.71	391	4.02	4.50	
AP	CALIPER	365	3649	M/F	253.82	56.71	391	CALIPER	3650	4744	M/F	277.56	77.60	154	3.45	4.52	
AP	CALIPER	3650	4744	M/F	277.56	77.60	154	CALIPER	4745	5474	M	298.27	98.87	66	1.51	2.87	
AP	CALIPER	3650	4744	M/F	277.56	77.60	154	CALIPER	4745	5474	F	146.26	56.40	68	14.17	2.89	[a]
AP	CALIPER	4745	5474	M	298.27	98.87	66	CALIPER	5475	6204	F	146.26	56.40	68	10.89	2.24	[a]
AP	CALIPER	5475	6204	M	179.47	73.27	64	CALIPER	5475	6204	F	88.97	18.88	74	9.61	2.27	[a]
AP	CALIPER	6205	6934	M	101.44	26.01	54	CALIPER	6205	6934	F	71.72	11.38	40	7.48	1.88	[a]
AP	CALIPER	4745	5474	M	298.27	98.87	66	CALIPER	5475	6204	M	179.47	73.27	64	7.80	2.21	[a]
AP	CALIPER	5475	6204	M	179.47	73.27	64	CALIPER	6205	6934	M	101.44	26.01	54	7.95	2.10	[a]
AP	CALIPER	4745	5474	F	146.26	56.40	68	CALIPER	5475	6204	F	88.97	18.88	74	7.98	2.31	[a]
AP	CALIPER	5475	6204	F	88.97	18.88	74	CALIPER	6205	6934	F	71.72	11.38	40	6.08	2.07	[a]
AP	rpart	4	18	M/F	168.05	45.32	152	rpart	22	152	M/F	310.08	92.72	69	12.09	2.88	[a]
AP	rpart	22	152	M/F	310.08	92.72	69	rpart	183	4010	M	255.34	66.69	253	4.59	3.47	[a]
AP	rpart	22	152	M/F	310.08	92.72	69	rpart	183	4719	F	260.93	64.79	319	4.19	3.81	[a]
AP	rpart	183	4010	M	255.34	66.69	253	rpart	4058	5367	M	307.73	90.30	103	5.33	3.65	[a]
AP	rpart	4058	5367	M	307.73	90.30	103	rpart	5380	6098	M	197.72	80.05	61	8.10	2.48	[a]
AP	rpart	5380	6098	M	197.72	80.05	61	rpart	6108	6881	M	105.92	28.01	65	8.48	2.17	[a]
AP	rpart	183	4719	F	260.93	64.79	319	rpart	4727	5333	F	158.67	54.32	58	12.78	3.76	[a]
AP	rpart	4727	5333	F	158.67	54.32	58	rpart	5374	6871	F	83.75	18.72	126	10.23	2.63	[a]
AP	rpart	4	18	M/F	168.05	45.32	152	CALIPER	0	14	M/F	165.90	43.83	153	0.42	3.38	
AP	rpart	22	152	M/F	310.08	92.72	69	CALIPER	15	364	M/F	287.15	94.87	149	1.69	2.86	
AP	rpart	183	4010	M	255.34	66.69	253	CALIPER	15	364	M/F	287.15	94.87	149	3.60	3.88	
AP	rpart	183	4719	F	260.93	64.79	319	CALIPER	15	364	M/F	287.15	94.87	149	3.06	4.19	
AP	rpart	4058	5367	M	307.73	90.30	103	CALIPER	3650	4744	M/F	277.56	77.60	154	2.77	3.10	
AP	rpart	4058	5367	M	307.73	90.30	103	CALIPER	4745	5474	M	298.27	98.87	66	0.63	2.52	
AP	rpart	5380	6098	M	197.72	80.05	61	CALIPER	5475	6204	M	179.47	73.27	64	1.33	2.17	
AP	rpart	6108	6881	M	105.92	28.01	65	CALIPER	6205	6934	M	101.44	26.01	54	0.90	2.11	
AP	rpart	183	4719	F	260.93	64.79	319	CALIPER	3650	4744	M/F	277.56	77.60	154	2.30	4.21	
AP	rpart	183	4719	F	260.93	64.79	319	CALIPER	4745	5474	F	146.26	56.40	68	14.81	3.81	[a]
AP	rpart	183	4719	F	260.93	64.79	319	CALIPER	5475	6204	F	88.97	18.88	74	40.55	3.84	[a]
AP	rpart	4727	5333	F	158.67	54.32	58	CALIPER	4745	5474	F	146.26	56.40	68	1.26	2.17	
AP	rpart	5374	6871	F	83.75	18.72	126	CALIPER	5475	6204	F	88.97	18.88	74	1.89	2.74	
AP	rpart	5374	6871	F	83.75	18.72	126	CALIPER	6205	6934	F	71.72	11.38	40	4.90	2.49	[a]

[a]Significant z.score.

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, 3rd ed. Wayne: Clinical and Laboratory Standards Institute; 2008. (28th series; vol. 30).

2. 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.

3. Ichihara K, Boyd JC. An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48:1537–51.

4. Lahti A. Partitioning biochemical reference data into subgroups: comparison of existing methods. Clin Chem Lab Med 2004;42:725–33.

5. Sikaris KA. Physiology and its importance for reference intervals. Clin Biochem Rev 2014;35:3–14.

6. 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.

7. Ma C, Yu Z, Qiu L. Development of next-generation reference interval models to establish reference intervals based on medical data: current status, algorithms and future consideration. Crit Rev Clin Lab Sci 2024;61:298–316.

8. Zierk J, Baum H, Bertram A, Boeker M, Buchwald A, Cario H, et al. High-resolution pediatric reference intervals for 15 biochemical analytes described using fractional polynomials. Clin Chem Lab Med 2021;59:1267–78.

9. Klawitter S, Kacprowski T. A visualization tool for continuous reference intervals based on GAMLSS. J Lab Med 2023;47:165–70.

10. 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.

11. Breiman L, Friedman J, Charles C, Olshen R. Classification and regression trees. New York: Chapman; Hall/CRC; 2017.

12. Thomas L. Labor und diagnose; 2024. Available from: https://www.labor-und-diagnose.de/ [Accessed 29 Apr 2024].

13. Hirfanoglu IM, Unal S, Onal EE, Beken S, Turkyilmaz C, Pasaoglu H, et al. Analysis of serum gamma-glutamyl transferase levels in neonatal intensive care unit patients. J Pediatr Gastroenterol Nutr 2014;58:99–101.

14. Gortner L, Meyer S, editors. Pädiatrie. 5., vollständig überarbeitete Auflage. Stuttgart; New York: Georg Thieme Verlag; 2018. (Thieme eRef).

15. R Core Team. R: a language and environment for statistical computing; 2023. Available from: https://www.R-project.org/ [Accessed 29 Apr 2024].

16. 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.

17. Wickham H. ggplot2: elegant graphics for data analysis; 2016. Available from: https://ggplot2.tidyverse.org [Accessed 29 Apr 2024].

18. Therneau T, Atkinson B. Rpart: recursive partitioning and regression trees; 2023. Available from: https://CRAN.R-project.org/package=rpart [Accessed 29 Apr 2024].

19. 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].

20. 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].

21. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. Clin Chem 1990;36:265–70.

22. Haeckel R, Wosniok W. Observed, unknown distributions of clinical chemical quantities should be considered to be log-normal: a proposal. Clin Chem Lab Med 2010;48:1393–6.

23. Lantz B. Machine learning with R: expert techniques for predictive modeling, 3rd ed. Birmingham, UK: Packt; 2019. (Expert insight).

24. Bramer M, editor. Avoiding overfitting of decision trees. In: Principles of data mining. London: Springer; 2007:119–34 pp.