



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)
**Journal of Mass Spectrometry and
 Advances in the Clinical Lab**

journal homepage: www.sciencedirect.com/journal/journal-of-mass-spectrometry-and-advances-in-the-clinical-lab



Indirect reference intervals using an R pipeline

Dustin R. Bunch^{a,b,*}

^a Nationwide Children's Hospital, Department of Pathology and Laboratory Medicine, 700 Children's Dr, Columbus, OH 43205, United States

^b The Ohio State University, Department of Pathology, 410 West 10th Ave, Columbus, OH 43210, United States

ARTICLE INFO

Keywords:

Reference interval
 R markdown tutorial
 Mixtools

ABSTRACT

Background: Indirect reference intervals require robust statistical approaches to separate the pathological and healthy values. This can be achieved with a data pipeline created in R, a freely available statistical programming language.

Methods: A data pipeline was created to ingest, partition, normalize, remove outliers, and identify reference intervals for testosterone (Testo; $n = 7,207$) and aspartate aminotransferase (AST; $n = 5,882$) using data sets from NHANES.

Results: The estimates for AST and Testo determined by this pipeline approximated current RIs. Care should be taken when using this pipeline as there are limitations that depend on the pathology of the analyte and the data set being used for RI estimation.

Conclusions: R can be used to create a robust statistical reference interval pipeline.

1. Introduction

Reference intervals (RI) are an integral component of the clinical service that a laboratory offers. For many analytes, this is the only interpretative information given with a patient's result. The clinical laboratory is charged with determining RIs that are specific to the population served and to the specific assays in use. Several organizations, including the Clinical Laboratory Standards Institute give guidance and describe methodologies to determine RIs (CLSI C28-A3c) [1]. The direct method is performed using a "healthy" patient population either prospectively recruited or through banked and/or residual patient samples. The health status is confirmed by direct questionnaires in the case of prospective recruitment and/or through chart review. The second methodology is termed indirect because it employs patient information from a database to determine RI. A comparison of the pros and cons of the direct and indirect methods is available in Jones et al. [2]. However, this publication will focus on creating a pipeline for indirect RI determination.

Indirect RI determinations typically use data sets from a laboratory informatics system (LIS) or the electronic health record (EHR). These

sources inherently contain data intermixed with pathological and normal values. To address this, indirect methods require robust statistical methods to separate the normal from the pathological results. These robust statistical methods can be applied with consultation of biostatisticians, specialized statistical software, and/or statistical programming languages, such as R [3,4].

R is an extremely flexible statistical programming language for creating reports and for statistical packages that allow for the extension of the base language [5,6]. These packages allow the end-user to focus on the application of complex statistically approaches without requiring coding of statistical concepts themselves. In addition, many packages have extensive documentation with online tutorials. R can be used to create a data pipeline for routine and *ad hoc* analysis of data. The typical statistical procedures required to determine RI, described in Ichihara and Boyd's International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) recommendations [7], are easily employed by a R data pipeline. Herein is described a R based simplified data pipeline for verification of RIs with the underlying code supplied for the user.

Abbreviations: ANOVA, Analysis of variance; AST, aspartate aminotransferase; CLSI, Clinical Laboratory Standards Institute; EHR, electronic health record; LC-MS/MS, Liquid chromatography tandem mass spectrometry; LIS, Laboratory informatics system; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine; RI, reference interval; SDI, Standard deviation index; SDR, Standard deviation ratio; Testo, Testosterone; TukeyHSD, Tukey multiple pairwise-comparisons; z_5 , Critical z-score.

* Address: Nationwide Children's Hospital, Department of Pathology and Laboratory Medicine, 700 Children's Dr, Columbus, OH 43205, United States.

E-mail address: dustin.bunch@nationwidechildrens.org.

<https://doi.org/10.1016/j.jmsacl.2022.02.004>

Available online 23 February 2022

2667-145X/© 2022 THE AUTHORS. Publishing services by ELSEVIER B.V. on behalf of MSACL This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2. Methods

An R data pipeline requires many statistical packages. The following packages are critical to the pipelines workflow (tidyverse [8], mixtools [9], janitor [10], here [11], readxl, fitdistrplus [12], mosaic [13], lubridate [14], eeptools [15], haven [16], and bestNormalize [17]) and were used to create a RI determination pipeline (Fig. 1).

A subset of the 2017–2018 NHANES will be used for demonstration purposes [18]. The following files will be used: demographics (DEMO_J.XPT), standard biochemistry profile (BIOPRO_J.XPT), alcohol use (ALQ_J.XPT), and medical conditions (MCQ_J.XPT). Each data set comes with documentation available at <https://wwwn.cdc.gov/nchs/nhanes/>. A second data set was extracted from the 2015–2016 NHANES for sex steroid hormones (TST_I.XPT) and the data demographics (DEMO_I.XPT) to demonstrate a sex separation.

The full code for this tutorial is available at https://github.com/dustinrbunch/jmsacl_reference_interval and includes the custom functions used for this tutorial.

2.1. Importing Data

During data import, column names are standardized using the janitor package. The standardization removes special characters (e.g., %, +, -) that are important to the R language. This can cause a loss of information and needs to be addressed with additional logic in the code or within the data set before import. Result data is converted to a numeric data type as part of the data type conversion. This causes non-numeric data, such as less than and greater than results, to be converted to 'NA'. The 2017–2018 data set was filtered on data import to contain the following columns: seqn, lboxsasi (Aspartate Aminotransferase [AST; IU/L]) performed on the Roche Cobas 6000 (c501 module), riagendr (gender), ridageyr (age in years), alq121 (Alcohol consumption in past year), mcq170l (Do you currently have a liver condition), mcq510a (fatty liver), mcq510b (liver fibrosis), mcq510c (liver cirrhosis), mcq510d (viral hepatitis), mcq510e (autoimmune hepatitis), and mcq510f (other liver disease). The 2015–2016 data set was filtered on data import to contain seqn, riagendr (gender), ridageyr (age in years), and lbxtst (testosterone [Test; ng/dL]) performed by liquid chromatography

tandem mass spectrometry (LC-MS/MS).

2.2. Partitions

The full data set was first separated based on analyte then on sex and/or age for the examples. For this example, AST had no partitions ($n = 5882$), while testosterone was partitioned based on sex and into age groups (F: 2776, M:2595). Analysis of variance testing (ANOVA) can be performed to determine partitions if the following assumptions are met; the residuals are normally distributed (approximately), the population variances are equal, and the observations are independent. When variance is unknown the 2-sample Welch's t-test can be used. The Kruskal–Wallis test is a non-parametric method similar to 1-way ANOVA. Tukey multiple pairwise-comparisons (TukeyHSD) is used for determining which group means differ from others in the group. For a tutorial on one-way ANOVA testing, an available resource is <http://www.sthda.com/english/wiki/one-way-anova-test-in-r>.

2.3. Normalization

The fitdistrplus package was used to perform the statistical summary of the data set including the kurtosis, skewness, and to plot a histogram of the data set. All data transformations were performed using the Yeo-Johnson method as deployed by the bestNormalize package.

2.4. Outliers

The detection of outliers or extreme values can be done with either a univariate approach, such as the Tukey method (Horn's Algorithm [19]) for parametric data or the Dixon method for non-parametric data or a multivariate approach. The Tukey method for outlier detection was applied to these data sets.

2.5. Identifying Reference Intervals

A resampling bootstrap method using the mosaic package, set at 500 iterations, in conjunction with a modern maximum likelihood method from the mixtools package [20], was used to identify RI. Multiple

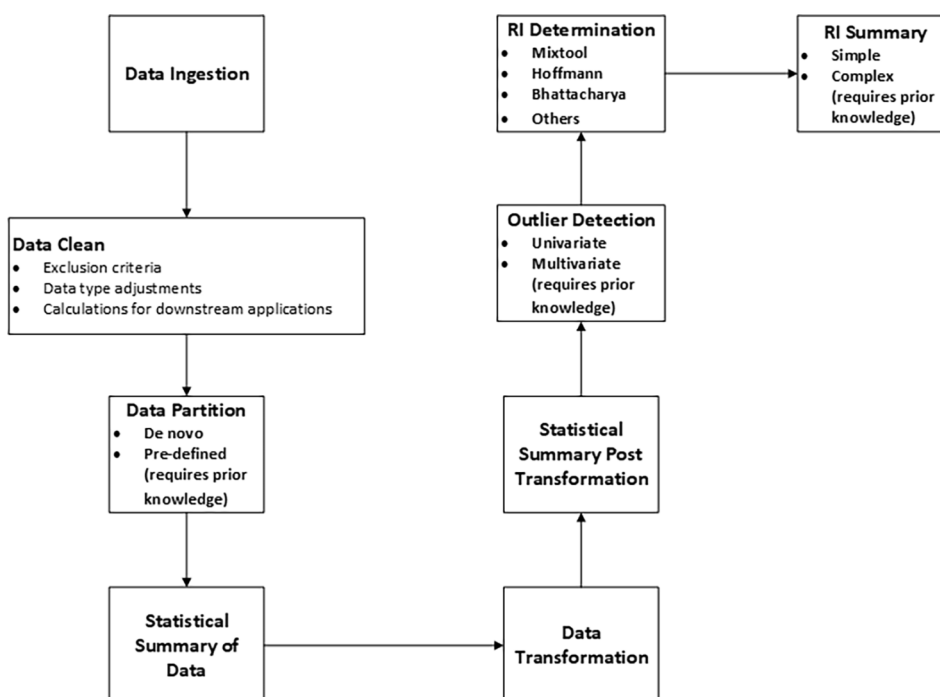


Fig. 1. .

permutations of the data set were applied for RI identification. For comparison purposes, RI were estimated with transformed and non-transformed data with and without outlier removal. The RI are summarized and saved to csv file for review.

3. Results and Discussion

3.1. Data Import [Code Block 1]

Data is paramount to indirect RI. Ideally, indirect RI are determined using all the data available. However, there are parameters within the data that should be addressed and/or recognized before import. The first parameter to understand is data stability or data history. Have the data sets changed such that they will require adjustments? Common changes to consider are instrument or assay changes, which often occur in data collected over long periods. A second parameter to understand is the changing of patient demographics and/or populations over time, including the dynamics around in- and out-patients. Generally, in-patients are admitted due to a pathological or surgical event, which often shifts assay results. Indirect data sets can give incorrect RI results when more data is pathological than normal or if the assay is used in a select population with a high probability of disease. For purposes of this example, the data set did not require separation of out- and in-patient information. However, typical hospital acquired data sets would require in-patient data exclusion, which can be performed by a combination of dplyr from tidyverse and stringr. Stringr is a package that employs regex, a language for working with text data.

```
## Code Block 1:
## Data import ##
# read in data
demo_data_2015 <- haven::read_xpt(here::here
  ("data",
   "DEMO_I.XPT")) %>%
  janitor::clean_names() %>%
  dplyr::select(seqn, riagendr, ridageyr)

testo_data <- haven::read_xpt(here::here
  ("data",
   "TST_I.XPT")) %>%
  janitor::clean_names() %>%
  dplyr::select(seqn, lbxtst)

# join data from two sources
testo_ri <- full_join(testo_data,
  demo_data_2015,
  by = "seqn")
```

3.2. Data Wrangling [Code Block 2 & 3]

Data type checks were performed and adjusted as needed with either R base functions (e.g., as.numeric and as.factor) or date functions (e.g., convert_to_datetime) from the janitor package. Following data type adjustment, data cleaning should be performed, which can include various processes that are dependent on the data set. When applying functions, non-conforming data is often converted to 'NA' values. These

can be removed through the drop_na function from the tidyr package (tidyverse) before proceeding. For data sets that are severely left-shifted a significant portion of the data could be removed due to being less-than the AMR yielding a non-numeric result. For these data sets, this pipeline is not ideal and other methods would need to be used.

The most common partitioning requirements are sex and age. Sex usually has three possible entries of female, male and unknown. Unknown sex often has limited data that makes it unsuitable for downstream processing, so should be removed for sex partitioning. For data extracted from a LIS or EHR, the age in the system can come in different formats, which include day, month, or year. To simplify downstream processes including age partitioning, age calculated in days is the most useful in a pediatric setting and years in an adult population. The eep-tools package provides useful tools for calculating age through the age_calc function. AST was chosen as an example due to the availability of the self reported alcohol consumption and liver disease status, which is known to impact AST result. AST data was filtered if the patient was known to have a liver condition.

```
## Code Block 2:
## Data Wrangling ##
# check data types
#glimpse(testo_ri) #Remove comment '#' to use.

# adjust datatypes
testo_ri$riagendr <-
  testo_ri$riagendr %>%
  dplyr::recode(`1` = "M", `2` = "F",
               .default = NA_character_) %>%
  as.factor()

testo_ri$seqn <- as.character(testo_ri$seqn)

# remove NAs from results
testo_ri <- testo_ri %>%
  drop_na(lbxtst)

testo_ri <- testo_ri %>%
  filter(ridageyr > 17)

# Remove those with known liver conditions
ast_ri <- ast_ri %>%
```

For the alcohol consumption category (alq121) zero codes for no alcohol consumption while 1–10 code for 1 having the most frequent consumption and 10 being the lowest frequency. A histogram or scatterplot of the condition being tested vs the analytes concentration is easy way to quickly visualize possible breakpoints in the data. As the frequency of alcohol increases there is a broadening of the standard deviation and a slight increase in the median and extreme values (Fig. 2).

Based on Fig. 2, a break should occur either at group 3 or group 4 with group 4 corresponding to having an alcoholic beverage more than once a month. From the three statistical methods (ANOVA, Welch's t-test, and Kruskal–Wallis test), The Kruskal–Wallis test is the most appropriate test for this data set comparison and was found to be significant. The ANOVA, Welch's t-test, and Kruskal–Wallis tests will determine if there are significant differences, but these test are unable to identify which are

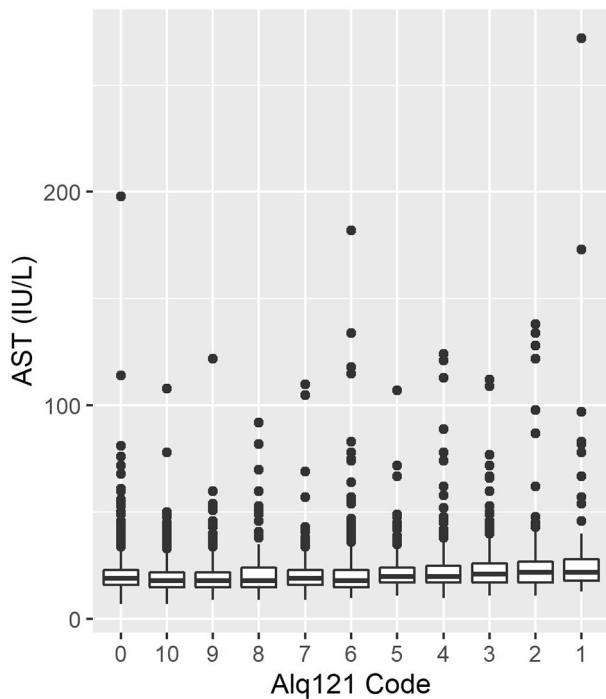


Fig. 2. .

different. The TukeyHSD statistical test is able to identify which are different. When the TukeyHST test was performed on the AST data versus the alcohol consumption there were 22 significant differences based on a p-value of ≤ 0.05 . These differences were mainly between groups 1–4 and all other groups. When groups 1–4 were removed and

the TukeyHSD test performed, there were no significant differences between the groups. A similar procedure can be used for the *de novo* determination of age and sex partitions.

```
## Code Block 3:
## Data Wrangling/Partitioning Statistics ##
ast <- ast_ri %>%
  dplyr::filter(alq121 != 77, alq121 != 99)
ast_anova.liver <- anova_test(
  ast, ast$lbxsassi, ast$alq121)

ast.anova.df1 <- as.data.frame(
  as.matrix(ast_anova.liver[[3]]$c2))

ast_sig1 <- ast.anova.df1 %>%
  dplyr::filter(`p adj` < 0.05)

ast_ri <- ast %>%
  dplyr::filter(alq121 != 4, alq121 != 3,
               alq121 != 2, alq121 != 1)
ast <- ast_ri
ast_anova.liver.etho <- anova_test(
  ast, ast$lbxsassi, ast$alq121)

ast.anova.df1 <- as.data.frame(
  as.matrix(ast_anova.liver.etho[[3]]$c2))

ast_sig2 <- ast.anova.df1 %>%
  dplyr::filter(`p adj` < 0.05)
```

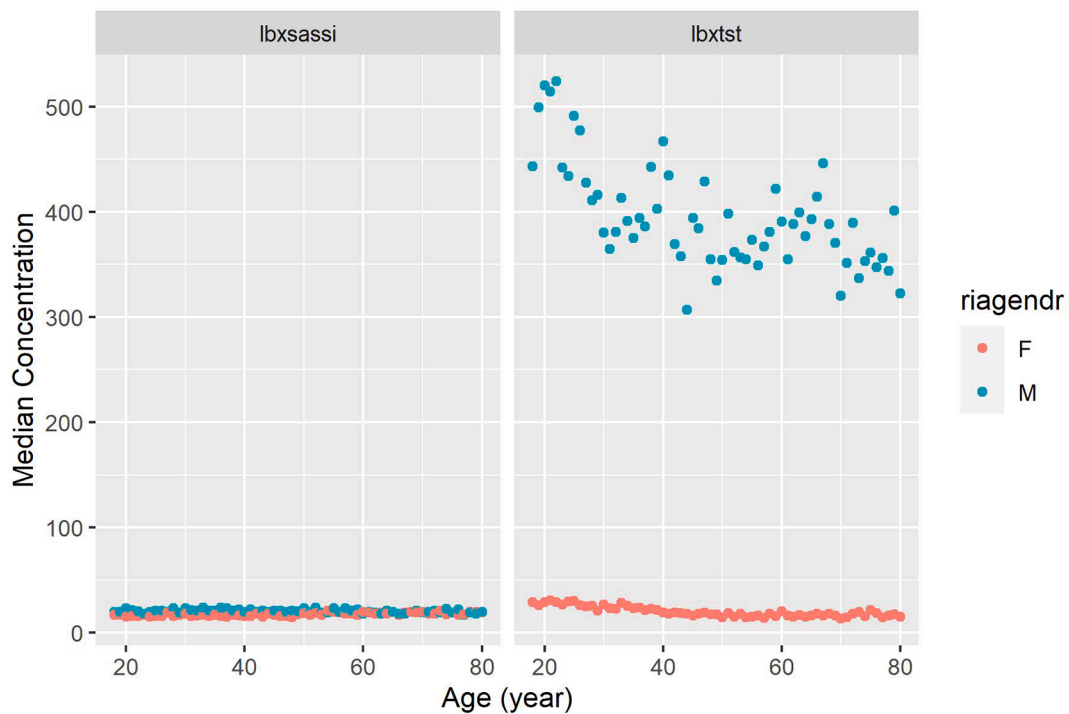


Fig. 3. .

3.3. Partitions [Code Block 4]

Partitioning statistics can be broken into two different use cases. The first use case is for the *de novo* determination of partitions. There are multiple statistical techniques that are useful for identifying partitions, such as those used above for data wrangling/partitioning ANOVA, Welch's t-test, and Kruskal–Wallis. Fig. 3 plots the age versus the concentration of AST and Testosterone with gender identified. From Fig. 3, a slight gender separation can be seen for AST, but when statistical tests are applied it is considered insignificant. There is a major split among age and gender for Testosterone, which is verified by statistical test. The code is available for *de novo* partitions in the full code (code block test_deno_partition1); however, as an example, partitions will be made based on existing reference intervals from a reference laboratory. For other statistical methods for determining partitions, see Ichihara et al., which details current recommended statistical methods (univariate and multivariate) for identifying sources of variation in the data set [7]. The caveat of multivariate techniques is that they require prior knowledge of the analyte in question. The second use case for partitioning statistics is to determine if an analyte data set recapitulates the partitions recommended elsewhere, such as published or in use RI. Frequently, the laboratory wants to compare newly derived RI with existing RI. There are a few common partitioning statistics to determine whether there is a significant difference between RI. When comparing to an existing RI without the original data set available for analysis, the Harris-Boyd method can be used even with the general downsides outlined by Lahti (Table 2) [21]. The mean and standard deviation can be ascertained from a RI based on the mean equals [(Upper Limit – Lower Limit)/2 + Lower Limit] and the standard deviation equals (Upper Limit – Lower Limit)/4 assuming the RI is based on a central 95%. The number of patient samples after exclusion, cleaning, and outlier removal included in each partition are in Table 1 and ranged from 63 to 380. All partitions had greater than 120 samples, which is the minimum recommended for establishing a RI in a direct data collection scheme [1], except the male 14 year partition.

```
## Code Block 4:
## Partitioning with Known Ranges ##
test_age <- testo_ri %>%
  dplyr::select(lbxtst, riagendr, ridageyr) %>%
  filter(!is.na(lbxtst)) %>%
  filter(!is.na(ridageyr)) %>%
  filter(!is.na(riagendr))

#### Males ####
test_age_m <- test_age %>%
  filter(riagendr == "M")

# Using an existing RI from a reference
# lab the split will be >= 19 for males and
# for female >= 19.

testo_19_80_r <- test_age_m %>%
  filter(as.numeric(as.character(ridageyr)) > 18
        & as.numeric(as.character(ridageyr)) < 81)

#### Females ####

test_age_f <- test_age %>%
  filter(riagendr == "F")

testo_f19_45_r <- test_age_f %>%
  filter(as.numeric(as.character(ridageyr)) > 18
        & as.numeric(as.character(ridageyr)) < 46)

testo_f46_80_r <- test_age_f %>%
  filter(as.numeric(as.character(ridageyr)) > 45
        & as.numeric(as.character(ridageyr)) < 81)

indirect_ri_list <- list(
  # No partitions needed.
  "AST_Both_All" = ast_ri %>%
    dplyr::select(lbxsassi),
  "Testo_M_19-80" = testo_19_80_r %>%
    dplyr::select(lbxtst),
  "Testo_F_20-45" = testo_f19_45_r %>%
    dplyr::select(lbxtst),
  "Testo_F_45-80" = testo_f46_80_r %>%
    dplyr::select(lbxtst)
)
```

3.4. Data Transformation [Code Block 5]

The distribution of laboratory data is often non-Gaussian. However, many of the statistical techniques used for RI determination assume a Gaussian distribution of the data. Typically, skewness and kurtosis are used to determine the shape of the distribution. Ideally, a skewness of 0 ± 0.5 is approximately symmetric and a kurtosis of 3 ± 2 has insignificant tailing. For most clinical laboratory data, a transformation needs to be performed if an approximation to a Gaussian curve is needed. Within a data pipeline where all data is treated in a similar fashion, the Yeo-Johnson transformation, a modification of the Box-Cox transformation, a historic normalizing transformation, is a universal transformation [22]. Table 1 shows the statistical summary for each partition before and after transformation. Once Yeo-Johnson transformation is performed all groups approximate a Gaussian curve with skewness from -0.1 to 0.2 and kurtosis from 3.9 to 5.6 on the high end.

Table 1
Partition Group Summary Statistics.

Analyte	Gender	Age Range (yr)	n	Pre-transformation						Post-transformation							
				Min	Max	Mean	SD	Skewness	Kurtosis	Outliers (%)	Min	Max	Mean	SD	Skewness	Kurtosis	Outliers (%)
AST	Both	All	5611	6	272	21.4	12.5	7.3	90.3	380 (6.8)	-6.1	3.4	0	1	-0.1	3.9	135 (2.4)
Testo	M	19–80	2534	1.72	2000	422.4	200.4	1.4	7.5	83 (3.3)	-3.8	5.1	0	1	0.2	4.6	63 (2.5)
Testo	F	20–45	1248	1.61	583	28.3	26.7	10.1	172.6	63 (5)	-5.9	4.3	0	1	-0.1	4.4	20 (1.6)
Testo	F	45–80	1467	0.53	444	21.1	25	8.8	110.6	83 (5.7)	-4.8	4.4	0	1	-0.1	5.6	39 (2.7)

```
## Code Block 5:
## Data Transformation ##
# data.test is a custom function that
# uses the fitdistrplus package. For
# documentation purposes I produce the
# histograms and Cullen & Frey graphs.
# Stats are saved for the analytes pre
# and post transformation & outlier
# removal.

# This code automatically prints histogram,
# qq-plot, and Cullen & Frey graph
# pretransformation
# indirect_ri_list %>%
  # map_depth(data.test, .depth = 2)

indirect_ri_list_trans <-
  indirect_ri_list %>%
  map_depth(data.tran, .depth = 2)
# Extract the transformed data
indirect_ri_list_tran <-
  lapply(indirect_ri_list_trans,
         '['[ , 1)
indirect_ri_list_tran <-
  lapply(indirect_ri_list_tran,
         '['[ , 1)

# These list are used to do the
# backtransformations downstream
indirect_ri_list_tran1 <-
  lapply(indirect_ri_list_trans,
         '['[ , 1)
indirect_ri_list_tran1 <-
  unlist(indirect_ri_list_tran1,
         recursive = F)

# This code automatically prints histogram,
# qq-plot, and Cullen & Frey graph
# posttransformation
# map(map(indirect_ri_list_tran1, "x.t"),
# data.test)
```

3.5. Outlier Detection [Code Block6]

Theoretically, using the Tukey method on a Gaussian data set would remove 0.7% of the data from the tail [7]. Outlier removal before transformation was 3.3% to 6.8% and after transformation was 1.6 to 2.7%. This is not a drastic difference for those close to Gaussian, but can be significant for the groups that were least Gaussian (Table 1). To determine if the outlier removal changes the outcome of determining RI, the data set was subjected to four different workflows (Table 2): 1. transformed with outliers removed 2. transformed without outliers removed 3. No transformation with outliers removed 4. No transformation without outliers removed. Transformation had a noticeable impact on the RI outcome with ~3% difference in the outliers removed. As an example the female Testo groups has an upper limit mean RI estimate of 49 ng/dL while the non-transformed estimates were 22.0 ng/dL and 33.8 ng/dL. Sometime negative estimates can be obtained and are often due to left shifted data which is usually resolved through transformation.

Table 2
RI Data Pretreatment Summary.

Analyte	Gender	Age Range (yr)	Reference Lab RI LL	Reference Lab RI UL	Transformed-Outliers Removed			Transformed-No Outlier Removed			No Transformation-Outlier Removed			No Transformation-No Outlier Removed		
					LL Mean (95% CI)	UL Mean (95% CI)	UL Mean (95% CI)	LL Mean (95% CI)	UL Mean (95% CI)	UL Mean (95% CI)	LL Mean (95% CI)	UL Mean (95% CI)	UL Mean (95% CI)	LL Mean (95% CI)	UL Mean (95% CI)	
AST	Both	All	10	35	12.2 (12.2–12.2)	37.6 (37.5–37.7)	12.5 (12.5–12.5)	35.5 (35.4–35.5)	10.8 (10.8–10.8)	23.2 (23.1–23.2)	9.6 (9.6–9.6)	28.7 (28.7–28.7)				
Testo	M	19–80	264	916	147.1 (147.1–147.2)	707.9 (707.2–708.5)	153.2 (152.8–153.5)	727.1 (726.2–728)	115.1 (114.5–115.7)	705.7 (697.9–713.4)	108.6 (108.6–108.6)	617.6 (617.6–617.6)				
Testo	F	20–45	10	55	10.2 (10.2–10.3)	65.5 (65.3–65.6)	9.6 (9.6–9.6)	65.7 (65.7–65.8)	6.7 (6.7–6.7)	31 (30.9–31.1)	3.9 (3.9–3.9)	44.5 (44.5–44.5)				
Testo	F	45–80	7	40	6.4 (6.4–6.4)	49 (48.9–49)	6.1 (6.1–6.1)	46.2 (46.1–46.4)	3.4 (3.4–3.4)	22 (21.9–22)	0.8 (0.8–0.8)	33.8 (33.8–33.8)				

```
## Code Block 6:
## Outlier detection and removal ##
# Remove outliers using the custom
# remove.outlier function which uses
# the Tukey Method.
indirect_ri_list_out <-
  indirect_ri_list_tran %>%
  map_depth(remove.outlier, .depth = 2)
```

3.6. Reference Intervals Determination [Code Bloc 7]

RI were estimated with and without transformation, and with and without outliers removed. The results of the methods and expected RIs are in Table 2. The data pipeline can be utilized for Gaussian data without affecting the final RI determination (unpublished data). The following statistical calculations are performed to improve confidence in the estimated RI and allow one to compare the different estimate made in this tutorial. However, sometimes investigation into the reference RI and clinical discussion is warranted. The comparability of the estimated RI to established RIs are defined in Table 3 with z-scores, standard deviation ratios (SDR), and standard deviation index (SDI) calculations. Z-scores were compared to the critical z value (z5) defined by Harris & Boyd [21], SDR is expected to be < 1.6, and SDI is a measure of the number of SDs a mean deviates from a reference mean. The expected value for SDI would be zero, but > 1.25 would be more than 1 SD away from the mean and would be concerning based on method comparison bias, as such the cut-off used was 1.25.

The estimates for the “transformed with outliers removed” best recapitulates the laboratory reference intervals used based on the z-scores and SDRs. When the estimates do not recapitulate the reference used, further investigation into the data analysis, patient demographics and/or diagnostic information, in search of unaccounted data variation, is needed. First steps could include: (i) investigating the providence of the reference RI and assay used to generate the data, (ii) examining whether the assays used to measure the data set were different from those used to create the reference RI. For the Testo data set, the reference laboratory used a LC-MS/MS assay. For AST, the reference RI and data set were performed on the same assay.

An inherit limitations of this pipeline concerns the assumption of one healthy range and one pathological range yielding a bimodal distribution for the data set. This is coded into the custom mixtool function and the selection of the correct limits. The mixtools k-value, which is the number of components or modes, is set to two in this example. Since the k-value is at 2, the lambda value is set to >0.5 for the RI selection. The lambda value is a reflection of the amount of data under the RI estimate. With these settings in place, there are two issues that can occur. The first issue is a data set that has a majority of pathological data, resulting in a lambda value >0.5, thereby causing the selection of an incorrect RI estimate. The second issue is similar to the first, but is a case where analytes have two pathological intervals. In such cases, with an example being thyroid stimulating hormone (TSH), the k-value should be set to 3 and the selection of the lambda will require deeper investigation and possibly manual interruption. Additionally, this pipeline is not optimal for identifying RIs for an analyte with no expected healthy and pathological ranges. To guard against these issues, sample verification of any RI with a minimum of 20 patients samples prior to implementation is recommended.

Table 3
Z-score Statistic with Standard Deviation Index.

Analyte	Gender	Age Range (yr)	Reference Lab RI Mean				Transformed-Outliers Removed				Transformed-No Outlier Removed				No Transformation-Outlier Removed				No Transformation-No Outlier Removed									
			Mean	SD	SDI	SDI	Mean	SD	z	z5	SDR	SDI	Mean	SD	z	z5	SDR	SDI	Mean	SD	z	z5	SDR	SDI				
AST	Both	All	22.5	6.2	24.9	6.4	20.2	34.2	1.03	0.4	24	5.8	13.2	34.2	1.07	0.2	17	3.1	57.4	33	2	0.9	19.1	4.8	32.5	34.2	1.29	0.5
Testo	M	19–80	590	163	427.5	140.2	38	23	1.16	1.0	440.2	143.5	34.7	23	1.14	0.9	410.4	147.7	40.4	22.6	1.1	1.1	363.1	127.2	55.2	23	1.28	1.4
Testo	F	20–45	32.5	11.2	37.8	13.8	10.5	16.1	1.23	0.5	37.6	14	10	16.1	1.25	0.5	18.9	6.1	36.7	15.7	1.84	1.2	24.2	10.2	19.4	16.1	1.1	0.7
Testo	F	45–80	23.5	8.2	27.7	10.7	11.9	17.5	1.3	0.5	26.1	10	7.7	17.5	1.22	0.3	12.7	4.7	42.5	17	1.74	1.3	17.3	8.2	20.5	17.5	1	0.8

SD, standard deviation; SDI, standard deviation index = $|\text{mean}_1 - \text{mean}_2| / \text{SD}$ [acceptable range ≤ 1.25]; SDR, standard deviation ratio [acceptable ≤ 1.5]; z, z-score = $|\text{mean}_1 - \text{mean}_2| / \sqrt{[(\text{SD}_1/n_1) + (\text{SD}_2/n_2)]}$; z5, critical z value = $5 * \sqrt{(n/120)}$, z-scores that exceed the z5 value are considered statistically different. RI that deviate from current RIs are bolded.

```
## Code Block 7:
## Reference Interval Estimation ##
# The estimation is bootstrapped "n" times
# using resample from the mosaic package
# To speed up the process, capture.output
# and invisible are included which helps
# when running n = 500. Mixtool.drb is a
# custom function that standardizes the
# mixtool parameters and allows mapping
# of the list. Mixtool.limits estimates
# the reference limits using the mixtool
# results.

n <- 500

analyte_name <- names(indirect_ri_list_out)

indirect_ri_list_out <-
  unlist(indirect_ri_list_out,
         recursive = F)

indirect_ref_list_ml <- map(1:n,
  ~ map(map(map(indirect_ri_list_out,
                na.omit), mixtool.drb),
        mixtool.limits))

indirect_ref_list_ml <-
  transpose(indirect_ref_list_ml)
indirect_ref_list_ml <-
  modify_depth(indirect_ref_list_ml, 1,
               transpose)
```

4. Conclusion

While R can be used to create a robust statistical RI determination pipeline, a fully automated data pipeline may be difficult to achieve with all of the clinical discrimination required. However, each segment of the pipeline presented here allows for review and interrogation of the data to help answer the clinical questions that arise during the RI determination process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] CLSI. Defining, establishing, and verifying reference intervals in the clinical laboratory; third edition [C28-A3] 2008.
- [2] G.R.D. Jones, R. Haeckel, T.P. Loh, K. Sikaris, T. Streichert, A. Katayev, et al., Indirect methods for reference interval determination - review and recommendations, Clin. Chem. Lab Med. 57 (2018) 20–29, <https://doi.org/10.1515/cclm-2018-0073>.
- [3] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
- [4] S. Haymond, S. Master, Why clinical laboratorians should embrace the R programming language a case for learning R as a gateway to laboratory medicines digital future, Clin. Lab News 46 (2020) 12–17.
- [5] D.R. Bunch, K. Firmender, R. Harb, J.M. El-Khoury, First- and Second-Trimester Reference Intervals for Thyroid Function Testing in a US Population, Am. J. Clin. Pathol. 155 (2020) 776–780, <https://doi.org/10.1093/ajcp/aqaa165>.
- [6] G.M. Ziegler, J.L. Slaughter, M. Chaudhari, H. Singh, P.J. Sanchez, D.R. Bunch, Preterm to term infant postmenstrual age reference intervals for thyroid-

- stimulating hormone and free thyroxine, *Pediatr. Res.* (2021), <https://doi.org/10.1038/s41390-021-01838-3>.
- [7] K. Ichihara, J.C. Boyd, An appraisal of statistical procedures used in derivation of reference intervals, *Clin. Chem. Lab Med.* 48 (2010) 1537–1551, <https://doi.org/10.1515/ccm.2010.319>.
- [8] H. Wickham, M. Averick, J. Bryan, W. Chang, L.D. McGowan, R. François, et al., Welcome to the tidyverse, *J. Open Source Softw.* 4 (2019) 1686, <https://doi.org/10.21105/joss.01686>.
- [9] T. Benaglia, D. Chauveau, D.R. Hunter, D.S. Young, Mixtools: An r package for analyzing finite mixture models, *J. Stat. Softw.* 32 (2009) 1–29, <https://doi.org/10.18637/jss.v032.i06>.
- [10] Firke S. Janitor: Simple tools for examining and cleaning dirty data 2020.
- [11] Müller K. Here: A simpler way to find your files 2017.
- [12] M.L. Delignette-Muller, C. Dutang, Fitdistrplus: An r package for fitting distributions, *J. Stat. Softw.* 64 (2015) 1–34, <https://doi.org/10.18637/jss.v064.i04>.
- [13] R. Pruim, D.T. Kaplan, N.J. Horton, The mosaic package: Helping students to think with data using r, *R J.* 9 (2017) 77, <https://doi.org/10.32614/RJ-2017-024>.
- [14] G. Grolemund, H. Wickham, Dates and times made easy with lubridate, *J. Stat. Softw.* 40 (2011) 1–25, <https://doi.org/10.18637/jss.v040.i03>.
- [15] Knowles JE. Eeptools: Convenience functions for education data 2020.
- [16] Wickham H, Miller E. Haven: Import and export SPSS, stata and SAS files 2021.
- [17] R.A. Peterson, J.E. Cavanaugh, Ordered quantile normalization: A semiparametric transformation built for the cross-validation era, *J. Appl. Stat.* (2019) 1–6, <https://doi.org/10.1080/02664763.2019.1630372>.
- [18] Disease Control C for, National Center for Health Statistics (NCHS) P (CDC). National health and nutrition examination survey data 2018.
- [19] H.E. Solberg, A. Lahti, Detection of outliers in reference distributions: Performance of horns algorithm, *Clin. Chem.* 51 (2005) 2326–2332, <https://doi.org/10.1373/clinchem.2005.058339>.
- [20] D.T. Holmes, K.A. Buhr, Widespread incorrect implementation of the hoffmann method, the correct approach, and modern alternatives, *Am. J. Clin. Pathol.* 151 (2018) 328–336, <https://doi.org/10.1093/ajcp/aqy149>.
- [21] A. Lahti, Are the common reference intervals truly common? Case studies on stratifying biochemical reference data by countries using two partitioning methods, *Scand. J. Clin. Lab Invest.* 64 (2004) 407–430, <https://doi.org/10.1080/00365510410006027>.
- [22] I. Yeo, R.A. Johnson, A new family of power transformations to improve normality or symmetry, *Biometrika* 87 (2000) 954–959, <https://doi.org/10.1093/biomet/87.4.954>.