

# Estimating Reference Intervals from Routine Laboratory Data Using Indirect Reference Interval Methods

**Kelly Doyle, PhD, DABCC, FADLM**

Associate Professor of Pathology, University of Utah  
Medical Director of Special Chemistry and Endocrinology, ARUP Laboratories

ARUP INSTITUTE FOR LEARNING

AUGUST 2024

# OBJECTIVES

1. Summarize the establishment and use of reference intervals in clinical laboratory practice.
2. Discuss the application of direct and indirect methods to determine population/sex/age-based reference intervals.
3. Describe how different estimation methods can overcome analyte specific challenges including skewed, partial, or overlapping distributions.

# Reference Intervals (RIs)

- *Are:*
  - » a representation of the typical distribution of analyte values observed in a healthy reference population
  - » are reported to clinicians to support the interpretation of clinical pathology results
- *Are not:*
  - » the same as clinical decision limits (CDL)

## Regulatory requirements in US and ISO:

- » **CLIA:** RIs must be provided with lab results when applicable
- » **CAP:** Each laboratory to establish or verify RIs
- » **ISO15189:** Verify RIs
  - When new clinical data is available
  - When a method has been in use for an extended time



October 2010

# EP28-A3c

Defining, Establishing, and Verifying  
Reference Intervals in the Clinical Laboratory;  
Approved Guideline—Third Edition

CLSI Guidance  
document  
Reaffirmed April 2016



# Direct Method

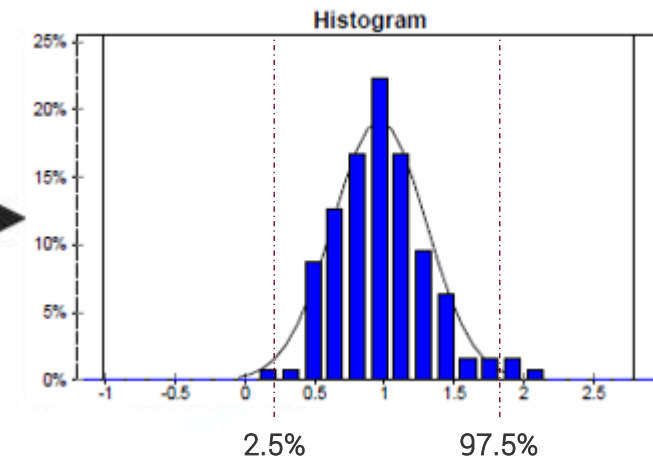
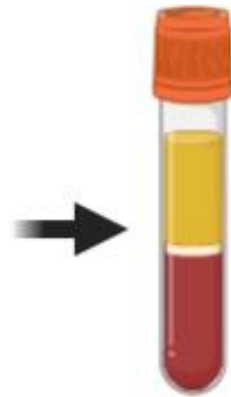
Planning and pre-collection

Patient preparation and collection

Specimen processing

Specimen testing

Data collection and analysis



# Benefits and Limitations

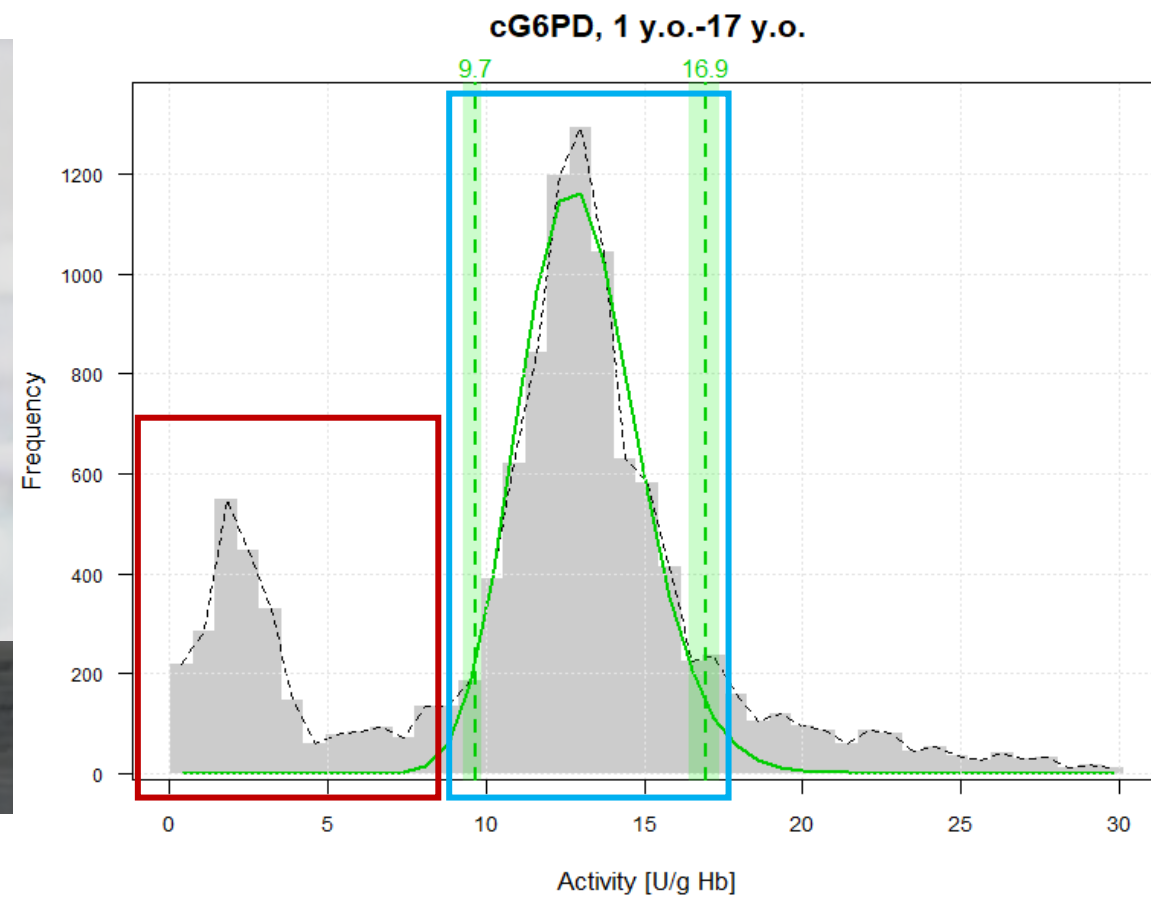
- Benefits

- » Controlled, well-characterized reference group
- » Defined protocols for volunteer preparation, specimen collection, and analysis
- » Simple statistical methods used

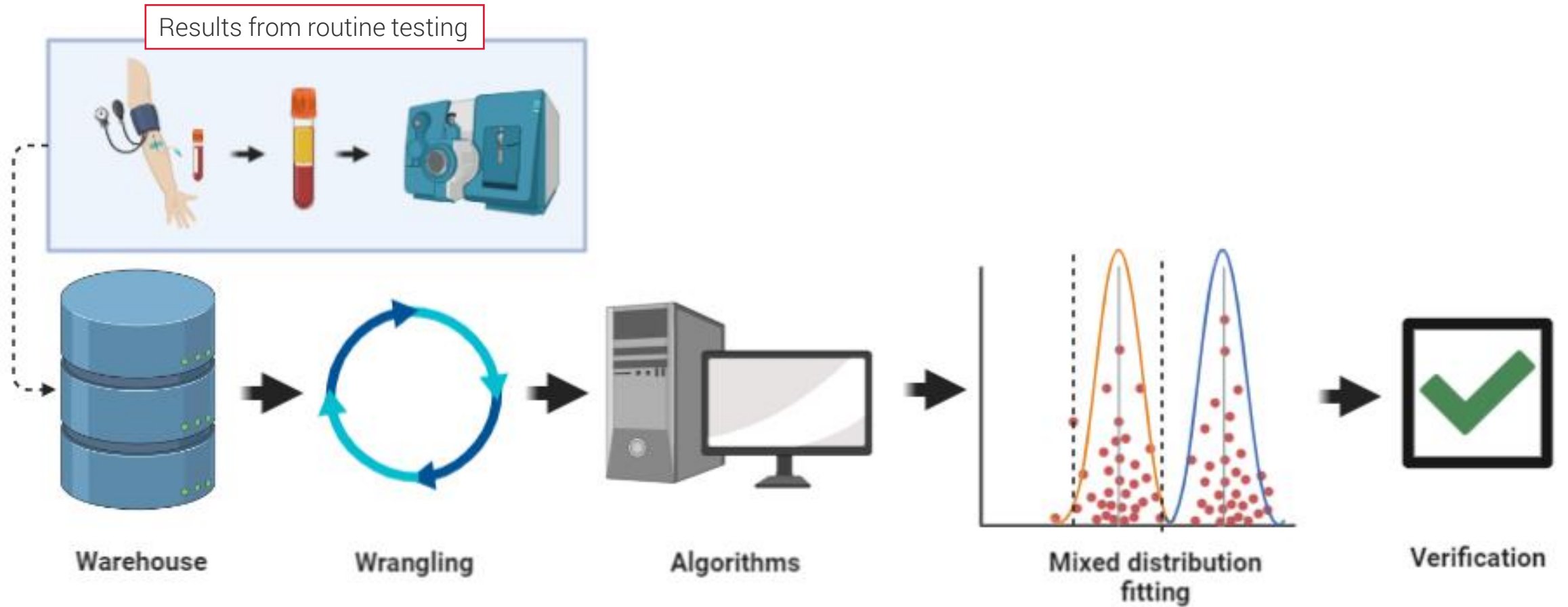
- Limitations

- » Cost
- » Time
- » Ethical issues – e.g., pediatric or pregnant individuals
- » Selection bias – volunteer reference group may not be truly representative
- » Challenge to encompass all groups – pediatrics, pregnant, elderly, ethnic
- » Preanalytical conditions may not fully represent all variables
- » Possible harm to subjects – bruising, time
- » Difficult to define/assure health of subjects

# ■ Indirect Methods



# Indirect Method





# Benefits and Limitations

- Benefits
  - » Cost
  - » Time
  - » Ethical advantages
  - » Mixed datasets – i.e., pathological and non-pathological test results
  - » Data collected under routine lab processes
  - » Difficult to collect scenarios (e.g., 24-hour urine)
  - » When many in the general population may be excluded (e.g., PTH in elderly)
- Limitations
  - » Useful after the testing has commenced - need data!
  - » There may be too little data for some partitions
  - » May be too challenging to minimize the influence of pathological results

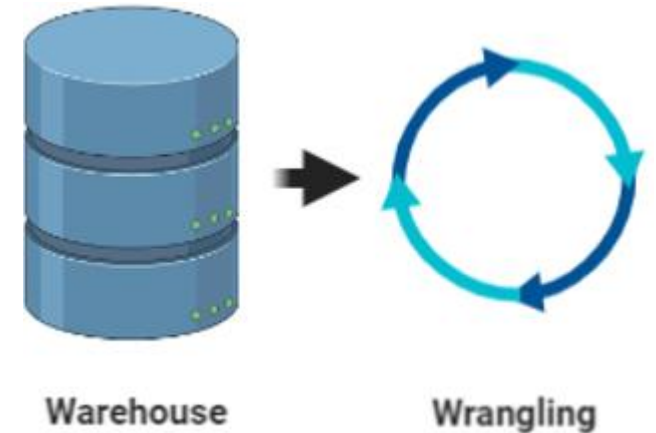


# ■ Conducting an Indirect Reference Interval Study

Overview and Use of the Indirect Methods

# Initial Process Steps

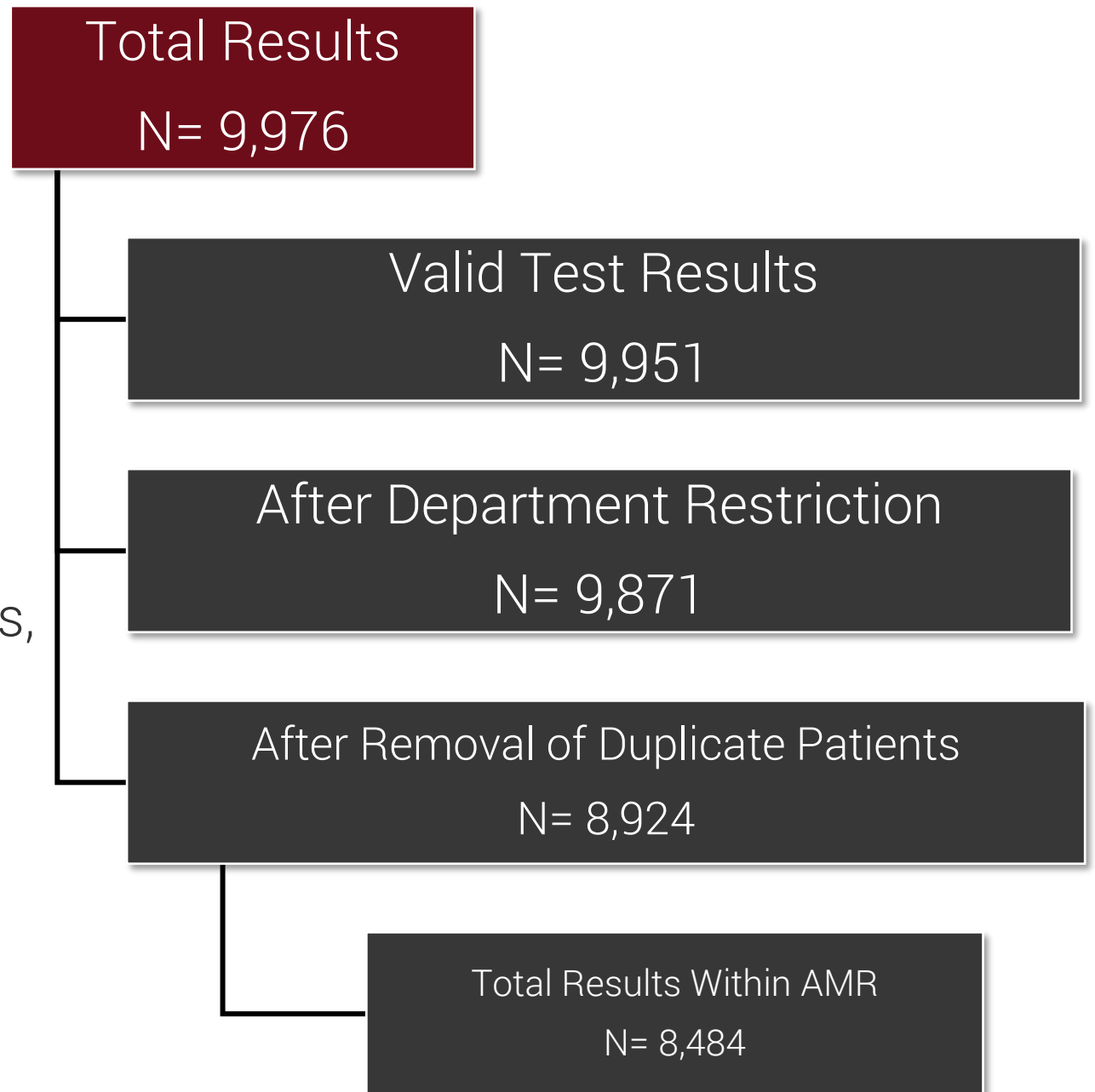
- **Step 1: Feasibility – analytic and population stability**
  - » QC review, moving medians, proficiency testing result review, method history, draw site/patient mix history (pathological fraction <30%)
- **Step 2: Pre-Data Collection**
  - » IRB, data sources, data sets
- **Step 3: Data Collection**
  - » Deidentification
  - » Age (decimal), sex, result time, instrument ID, collection site, result
- **Step 4: Data Cleanup**
  - » Collection sites (e.g., oncology, diabetes clinics), repeat patients (or repeat test results), test specimens, result truncation



# Data Cleanup

## ❖ Document the process!

Exclusion criteria, data cleanup steps, transformations, truncation, etc.



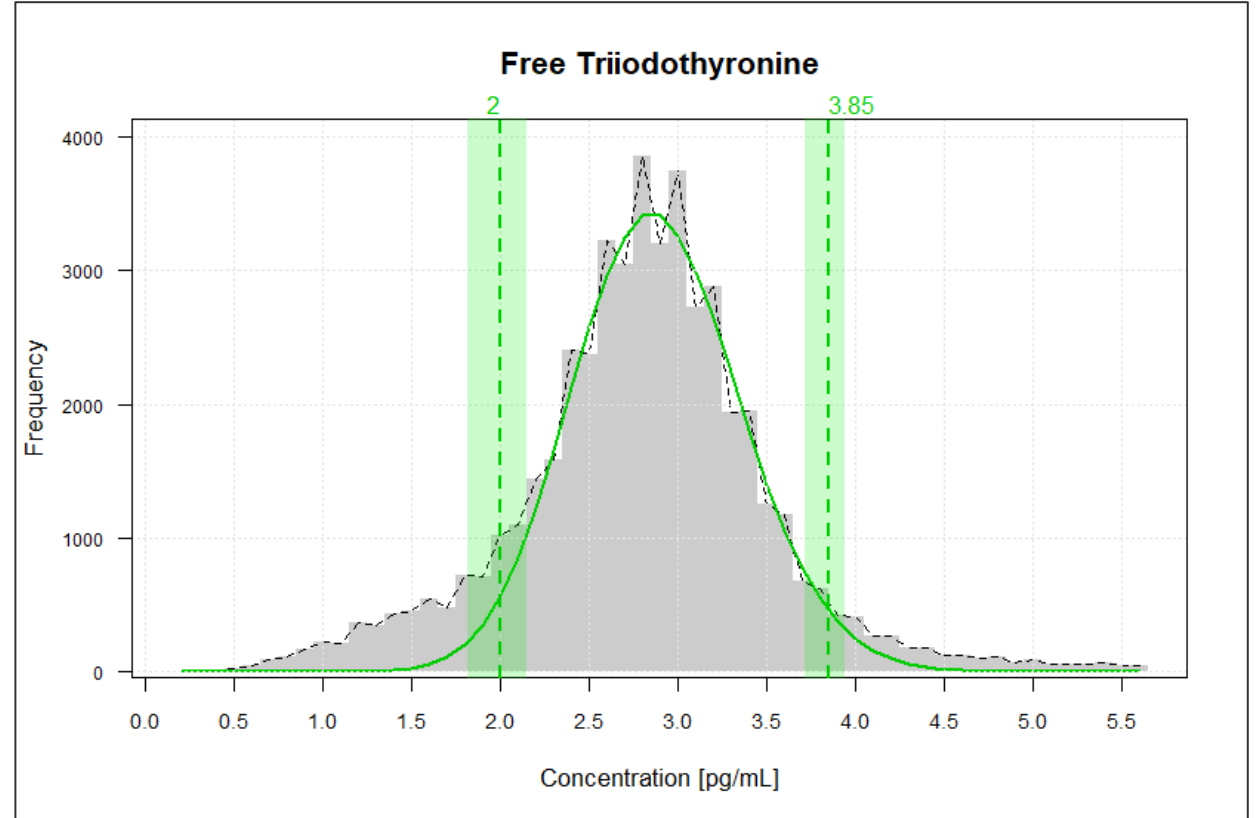
# Dataset Structure

Patient_ID	Age_decimal	Sex	Result	Verified_time	Device_ID
00001	27.29	M	1.34	05/22/2022 12:32:00	5500-6
00002	18.75	F	1.17	05/22/2022 12:33:00	5500-6

Recommendation: keep all data together as a single .csv file, perform filtering in R.

# Step 5 – Data Processing and Distribution Modeling

- Initial review of data set
  - » Distribution
    - Skew
    - < or > impact on histogram
- Choose an appropriate indirect method

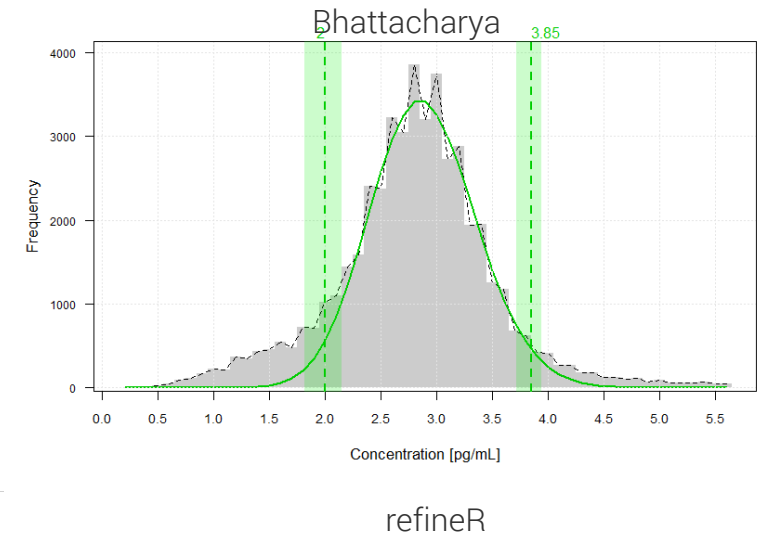
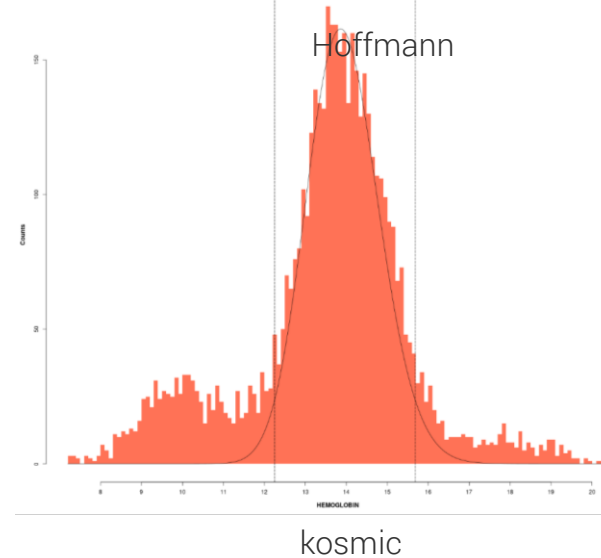
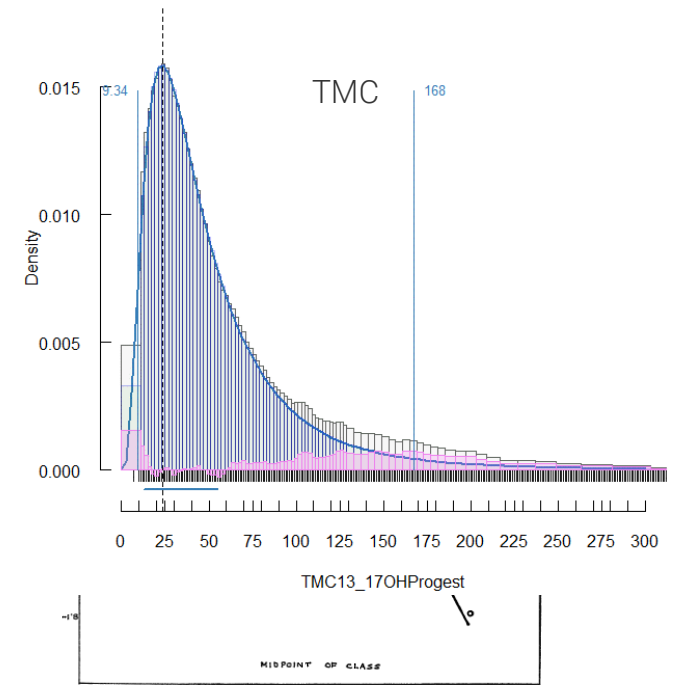
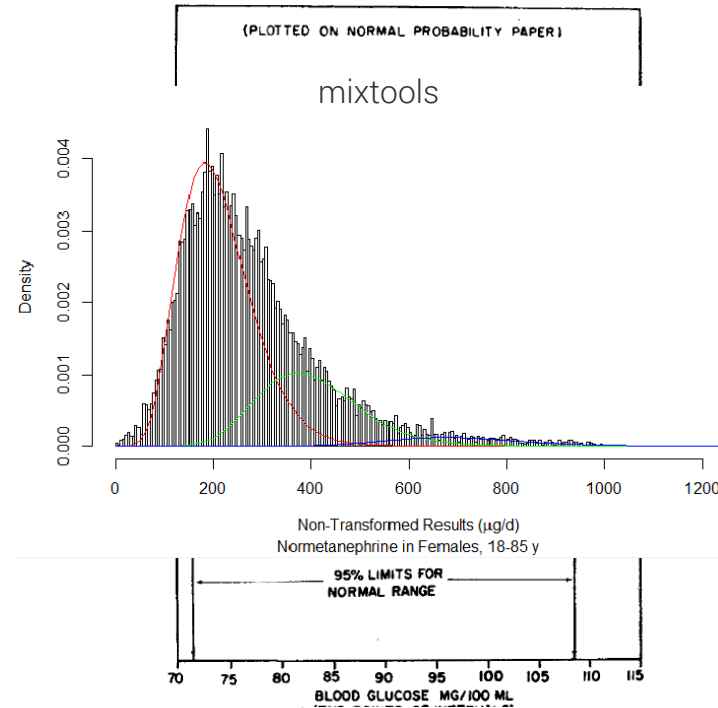


INDIRECT METHOD

# Methods/Algorithms

Hoffmann  
Bhattacharya

Mixtools  
Truncated Minimum Chi-Square (TMC)  
Kosmic  
refineR



```

76 rollmedian_data <- rollmedian(Data.DF$RESULT, k = N+1, fill = NA, align = 'right')
77 Data_median.DF <- Data.DF %>% mutate(rollmedian_data)
78 RM.plot <- ggplot(Data_median.DF, aes(VER.WHEN, rollmedian_data)) +
79   geom_line() +
80   coord_cartesian(ylim = c(0,100)) +
81   scale_x_datetime(date_breaks = "1 year", date_minor_breaks = "1 month", date_labels = ("%b %y")) +
82   labs(title = "Rolling Median", y = "Median", subtitle = ANALYTE, caption = paste("Based on N =", N))
83 RM.plot
84
85 #refiner ("BoxCox", "modBoxCoxFast", "modBoxCox")
86 #is.na(Data.DF$RESULT)
87 Data.DF[is.na(Data.DF$RESULT),]
88 Data.RI <- Data.DF[!is.na(Data.DF$RESULT),]
89
90 fit_OHP_F_16_17 <- findRI(Data = Data.RI$RESULT,
91   model = "BoxCox",
92   Nbootstrap = 1200,
93   seed = 123)
94 print(fit_OHP_F_16_17)
95 ri <- getRI(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS")
96 plot(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS",
97   xlim = c(0,300), xlab = "Concentration [ng/dL]", title = "17-hydroxyprogesterone, Females, 16-17 y.o.",
98   Nhist=50, showValue = FALSE)
99 mtext(text = round(ri$PointEst, 0), at = ri$PointEst, col = "green3", cex = 1.0)
100
101 save(fit_OHP_F_16_17, file = "./refiner_model_OHP_F_16_17.RData")
102
103
104 # to load again into the workspace to use the model for further plots or additional analysis
105 fit_OHP_F_16_17 <- get(load(file = "./refiner_model_OHP_F_16_17.RData"))
106

```

82:62 17-OHP, Females, 16-17 y.o.

```

R 4.2.2 - U:/Stats math/Reference Intervals/Indirect RI/Tools/refiner/refiner_TestScripts/test_data/
sigma: 1.95
shift: 0
cost: -14.2
NP fraction: 0.941
> ri <- getRI(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS")
> plot(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS",
+   xlim = c(0,300), xlab = "Concentration [ng/dL]", title = "17-hydroxyprogesterone, Females, 16-17 y.o.",
+   Nhist=50, showValue = FALSE)
> mtext(text = round(ri$PointEst, 0), at = ri$PointEst, col = "green3", cex = 1.0)
>
> save(fit_OHP_F_16_17, file = "./refiner_model_OHP_F_16_17.RData")
> Data.DF[is.na(Data.DF$RESULT),]
[1] PatID RESULT AGE.DECIMAL SEX VER.WHEN Device
<0 rows> (or 0-length row.names)
> Data.RI <- Data.DF[!is.na(Data.DF$RESULT),]
>
> fit_OHP_F_16_17 <- findRI(Data = Data.RI$RESULT,
+   model = "BoxCox",
+   Nbootstrap = 1200,
+   seed = 123)

```

Environment History Connections

Global Environment

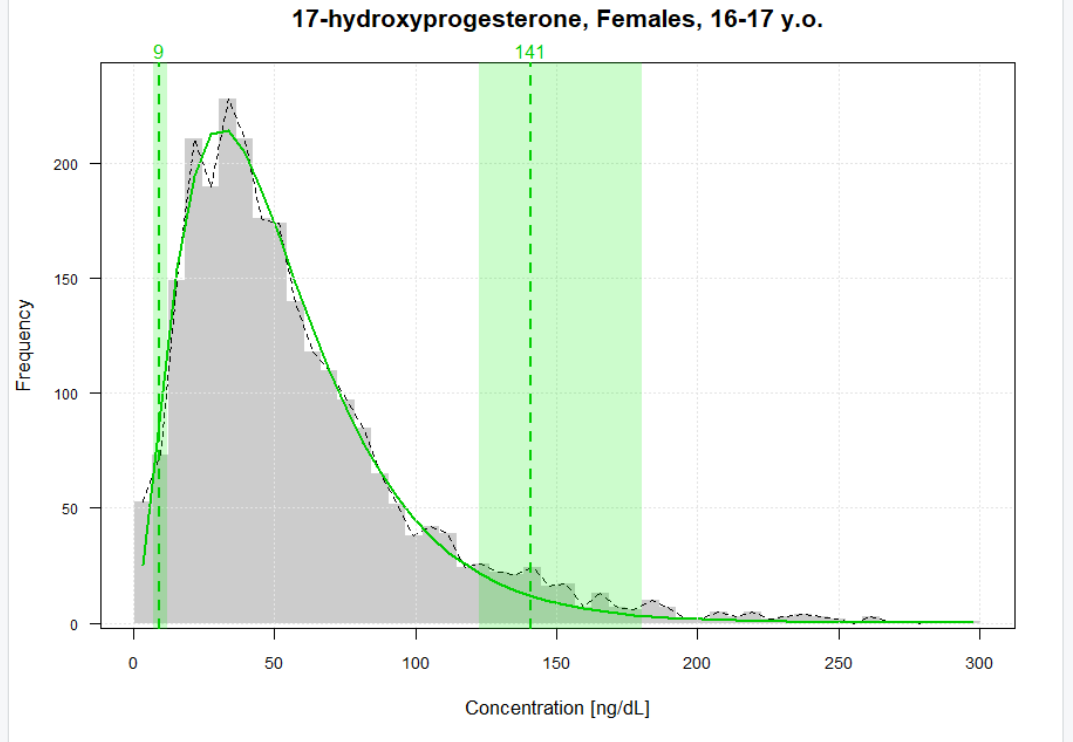
rawData.DF	LIST OF 27
ri	123565 obs. of 6 variables
RM.plot	2 obs. of 4 variables
RM.plot	List of 9

values

ANALYTE	"17-OHP, Females, 16-17 y.o."
myPaths	chr [1:2] "C:/Program Files/R/R-4.2.2/library" "C:/Users/204783/Documents/R/R..."
N	50
rollmedian_data	num [1:2505] NA NA NA NA NA NA NA NA NA ...

Files Plots Packages Help Tutorial Viewer Presentation

Zoom Export Publish





# What if I Don't Know a Coding Language?

kosmic

Upload data file

Advanced settings

Start over

## Reference Interval Estimation from Mixed Distributions using Truncation Points and the Kolmogorov-Smirnov Distance

### About

This is a web application accompanying the publication

[Reference Interval Estimation from Mixed Distributions using Truncation Points and the Kolmogorov-Smirnov Distance \(kosmic\)](#)  
by Jakob Zierk<sup>1, 2</sup>, Farhad Arzideh<sup>3</sup>, Lorenz A. Kapsner<sup>2</sup>, Hans-Ulrich Prokosch<sup>2</sup>, Markus Metzler<sup>1</sup> and Manfred Rauh<sup>1</sup>  
<sup>1</sup>Department of Pediatrics and Adolescent Medicine, University Hospital Erlangen, Erlangen, Germany  
<sup>2</sup>Chair of Medical Informatics, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany  
<sup>3</sup>Institute of Clinical Chemistry, University of Cologne, Cologne, Germany

This web application's aim is to enable evaluation of the reported algorithm for the estimation of reference intervals using in- or outpatient laboratory test results. Please refer to the publication for details regarding the algorithm and see <https://gitlab.miracum.org/kosmic> for the source code, Python bindings, and precompiled windows binaries. Importantly, **this is not a medical device/product** but an ongoing research project.

If you are looking for R bindings to kosmic, Devon Buchanan provides the tidykosmic R package at <https://www.divinenephron.co.uk/tidykosmic/>.

**Upload a text (\*.csv) or Excel file (\*.xlsx)** with at least **one column containing the test results**. Optionally, you can specify columns for **sex** and **age** - these columns can be used to subset your dataset. Additionally, you can specify a **patient ID** column - if your dataset contains multiple test results per patient ID, a single test result per patient (selected randomly) is used. Alternatively, you can use an example dataset of hemoglobin test results.

You can review the uploaded dataset and the selected subset using the left-hand pane. Confirm your selection and press 'Calculate Reference Interval' to see your results.

Pressing the 'Confirm selection'-button creates the filtered subset of the provided

Upload data file

Upload laboratory test results (accepted formats: \*.xlsx, \*.csv)

Browse... No file selected

Alternative: use an example data set:

Load example 'hemoglobin' test results.

***kosmic***

<https://kosmic.diz.uk-erlangen.de/>

Version: 0.0.7

© Universitätsklinikum Erlangen

## Upload data file

Upload laboratory test results (accepted formats: \*.xlsx, \*.csv)

Browse...

No file selected

Alternative: use an example data set:

Load example 'hemoglobin' test results.

## Select Columns

### Decimals

1

Caution: Increasing the number of decimals results in a substantially longer computation time.

Test result (required):

HEMOGLOBIN

Patient ID (optional):

HEMOGLOBIN

Age (optional):

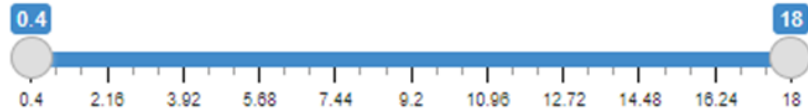
AGE

Sex (optional):

SEX

## Specify age range

Age



## Specify gender

Male:

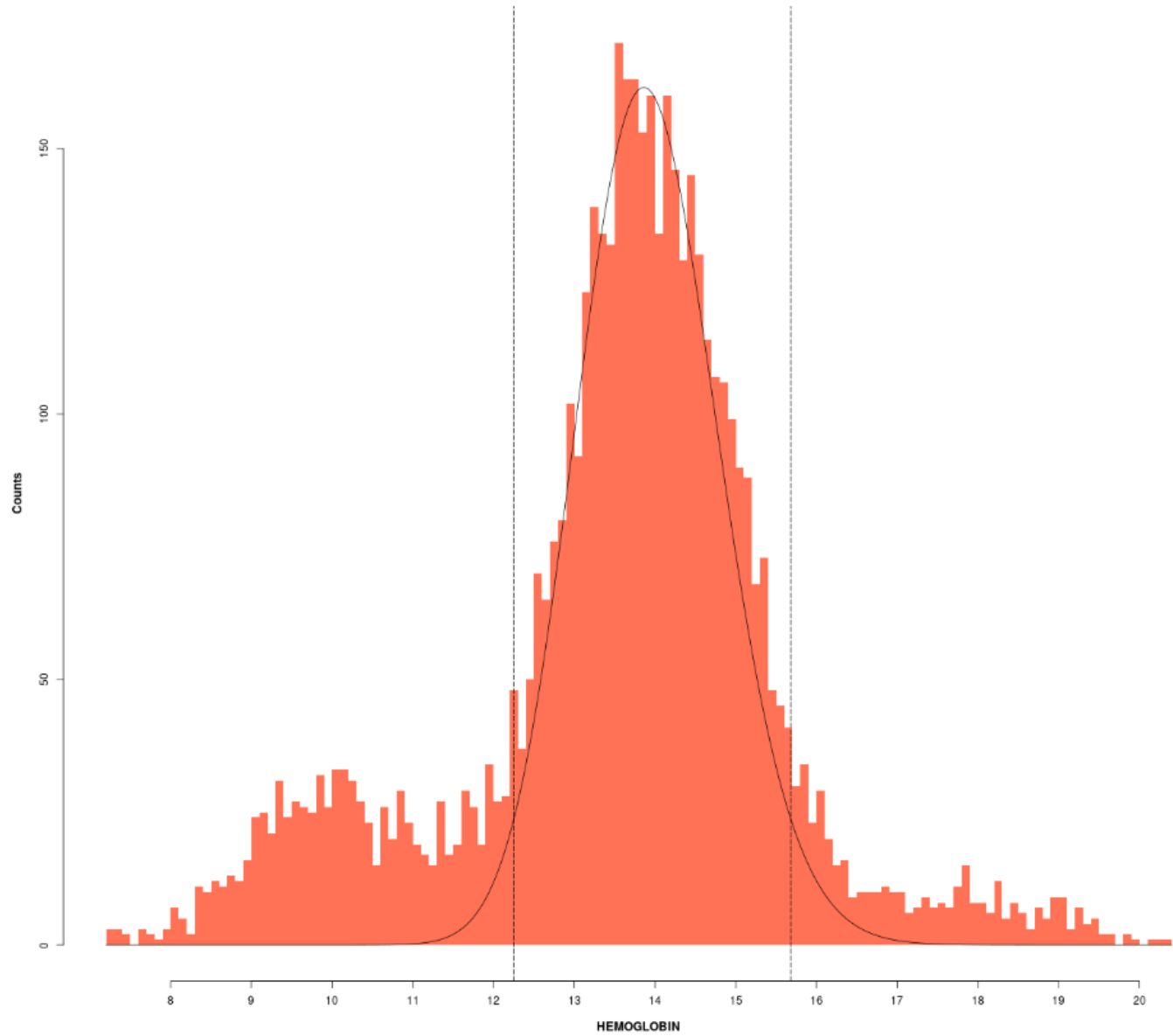
m

Female:

w

Select gender subset:

Female  Male



Download Plot

Download Report

N:	5016
2.5 % Percentile:	12.2496
50 % Percentile:	13.8606
97.5 % Percentile:	15.6834
Lambda:	0
Mu:	2.62905
Sigma:	0.0630387
T1:	12.6
T2:	14.4
Decimals:	1
T1 min:	0.05
T1 max:	0.3
T2 min:	0.7
T2 max:	0.95
SD:	0.8
Tolerance:	1e-07

```

76 rollmedian_data <- rollmedian(Data.DF$RESULT, k = N+1, fill = NA, align = 'right')
77 Data_median.DF <- Data.DF %>% mutate(rollmedian_data)
78 RM.plot <- ggplot(Data_median.DF, aes(VER.WHEN, rollmedian_data)) +
79   geom_line() +
80   coord_cartesian(ylim = c(0,100)) +
81   scale_x_datetime(date_breaks = "1 year", date_minor_breaks = "1 month", date_labels = ("%b %y")) +
82   labs(title = "Rolling Median", y = "Median", subtitle = ANALYTE, caption = paste("Based on N =", N))
83 RM.plot
84
85 #refiner ("BoxCox", "modBoxCoxFast", "modBoxCox")
86 #is.na(Data.DF$RESULT)
87 Data.DF[is.na(Data.DF$RESULT),]
88 Data.RI <- Data.DF[!is.na(Data.DF$RESULT),]
89
90 fit_OHP_F_16_17 <- findRI(Data = Data.RI$RESULT,
91   model = "BoxCox",
92   Nbootstrap = 1200,
93   seed = 123)
94 print(fit_OHP_F_16_17)
95 ri <- getRI(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS")
96 plot(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS",
97   xlim = c(0,300), xlab = "Concentration [ng/dL]", title = "17-hydroxyprogesterone, Females, 16-17 y.o.",
98   Nhist=50, showValue = FALSE)
99 mtext(text = round(ri$PointEst, 0), at = ri$PointEst, col = "green3", cex = 1.0)
100
101 save(fit_OHP_F_16_17, file = "./refiner_model_OHP_F_16_17.RData")
102
103
104 # to load again into the workspace to use the model for further plots or additional analysis
105 fit_OHP_F_16_17 <- get(load(file = "./refiner_model_OHP_F_16_17.RData"))
106

```

Environment History Connections

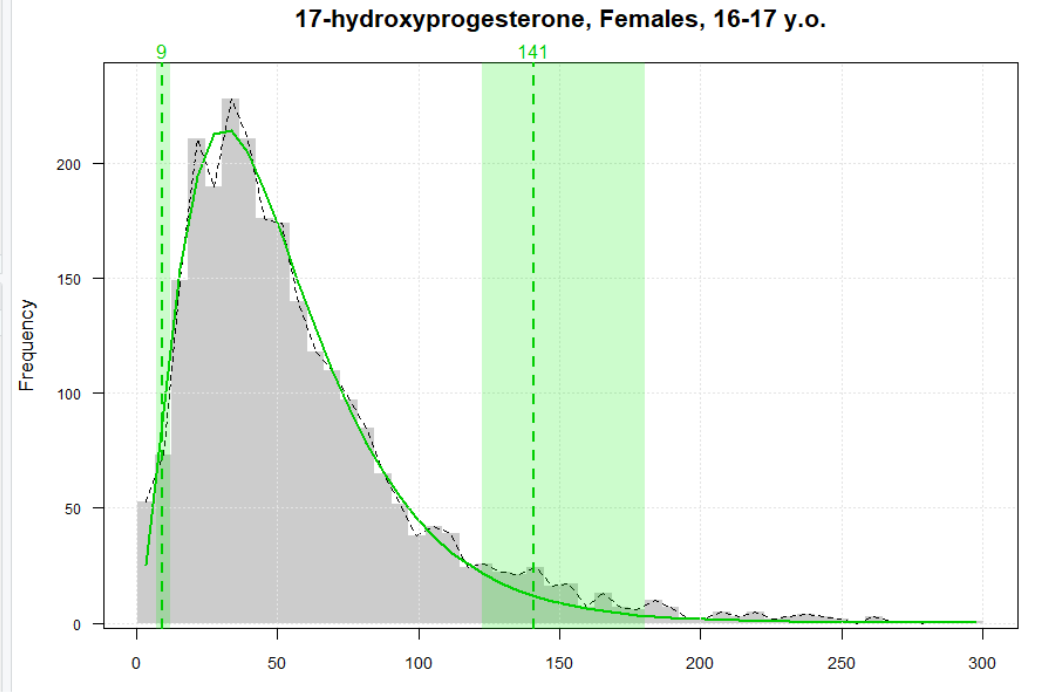
Global Environment

rawData.DF	LIST OF 27
ri	123565 obs. of 6 variables
RM.plot	2 obs. of 4 variables
RM.plot	List of 9

values

ANALYTE	"17-OHP, Females, 16-17 y.o."
myPaths	chr [1:2] "C:/Program Files/R/R-4.2.2/library" "C:/Users/204783/Documents/R/R..."
N	50
rollmedian_data	num [1:2505] NA NA NA NA NA NA NA NA NA ...

Files Plots Packages Help Tutorial Viewer Presentation



```

R 4.2.2 - U:/Stats math/Reference Intervals/Indirect RI/Tools/refiner/refiner_TestScripts/test_data/
sigma: 1.95
shift: 0
cost: -14.2
NP fraction: 0.941
> ri <- getRI(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS")
> plot(fit_OHP_F_16_17, RIperc = c(0.025, 0.975), CIprop = 0.95, pointEst = "medianBS",
+   xlim = c(0,300), xlab = "Concentration [ng/dL]", title = "17-hydroxyprogesterone, Females, 16-17 y.o.",
+   Nhist=50, showValue = FALSE)
> mtext(text = round(ri$PointEst, 0), at = ri$PointEst, col = "green3", cex = 1.0)
>
> save(fit_OHP_F_16_17, file = "./refiner_model_OHP_F_16_17.RData")
> Data.DF[is.na(Data.DF$RESULT),]
[1] PatID RESULT AGE.DECIMAL SEX VER.WHEN Device
<0 rows> (or 0-length row.names)
> Data.RI <- Data.DF[!is.na(Data.DF$RESULT),]
>
> fit_OHP_F_16_17 <- findRI(Data = Data.RI$RESULT,
+   model = "BoxCox",
+   Nbootstrap = 1200,

```

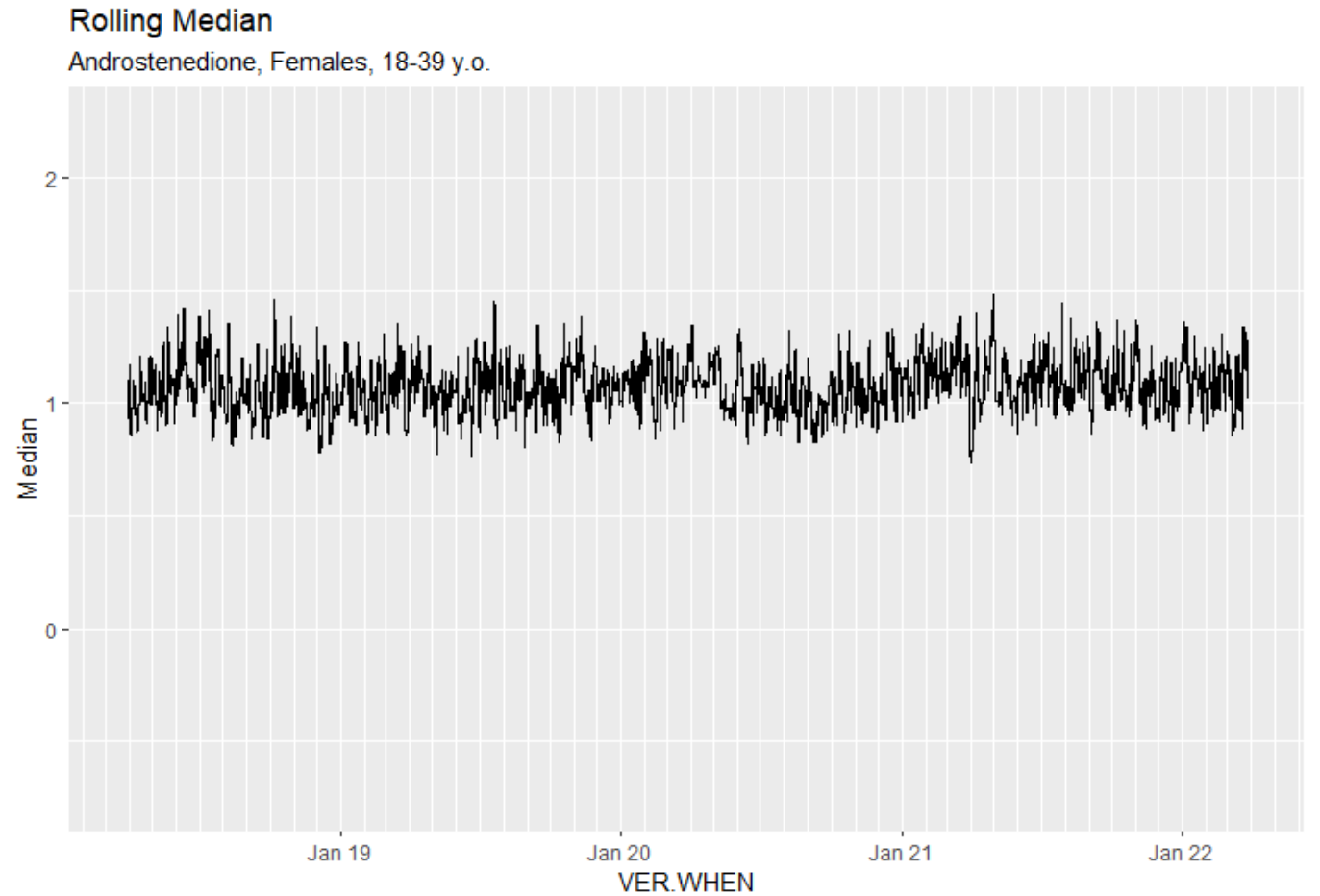
# Data Review: Moving Medians

Highly recommended to review data integrity across data set

Rolling median:

```
library(zoo)  
rollmedian(data, K (block size), align = 'right')
```

```
library(ggplot2)  
ggplot(data, aes(date, results))
```



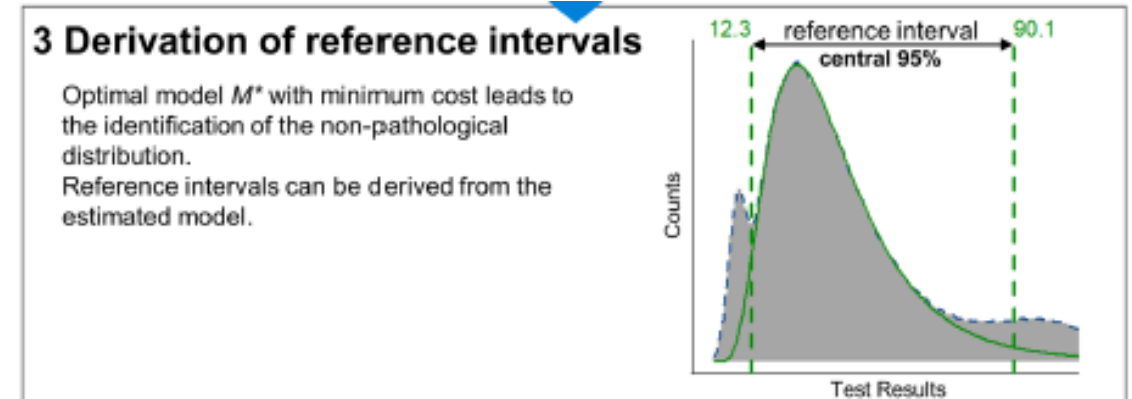
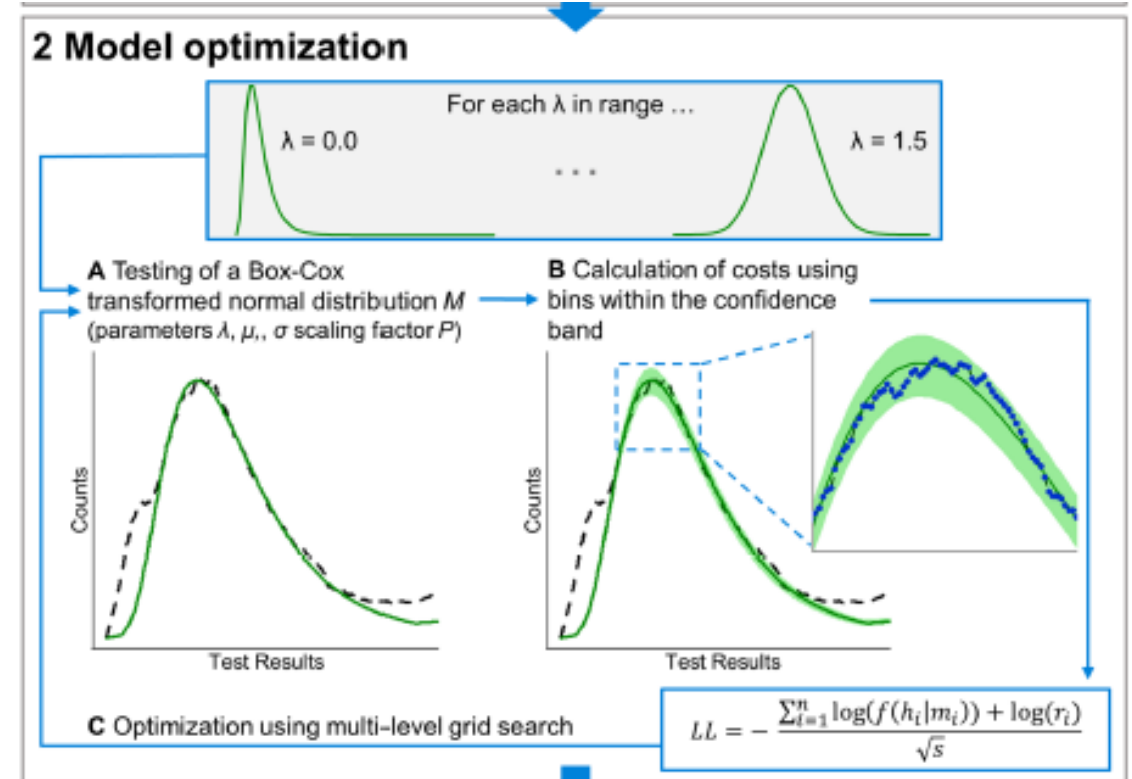
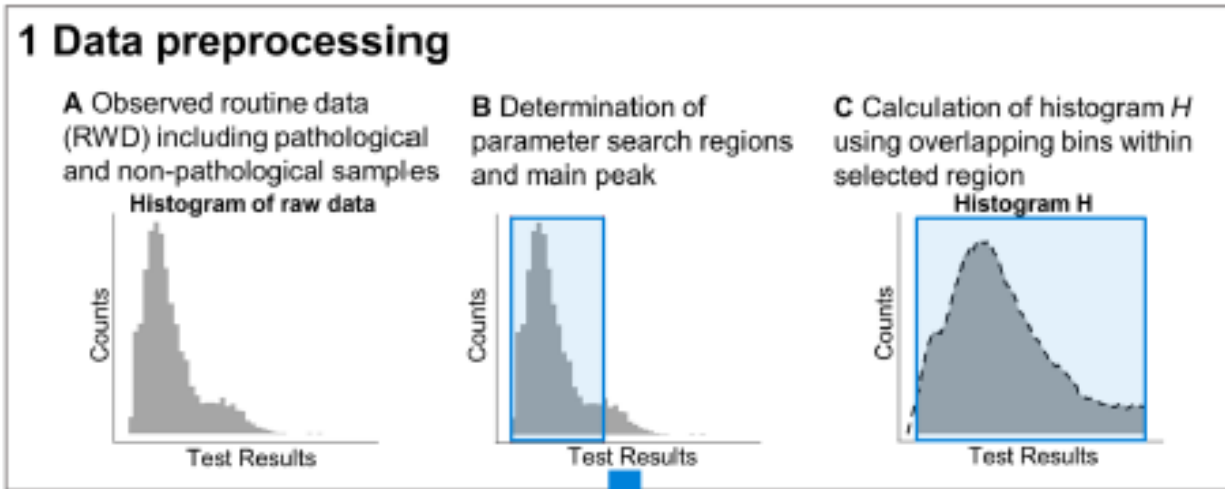
Based on N = 50

# refineR Package - R

## refineR: A Novel Algorithm for Reference Interval Estimation from Real-World Data

Tatjana Ammer<sup>1,2</sup>, André Schützenmeister<sup>2</sup>, Hans-Ulrich Prokosch<sup>1</sup>, Manfred Rauh<sup>3</sup>, Christopher M. Rank<sup>2,5</sup> & Jakob Zierk<sup>3,4,5</sup>

PMID: 34362961



# refineR Package - R

JOURNAL ARTICLE

## Estimation of Reference Intervals from Routine Data Using the refineR Algorithm—A Practical Guide

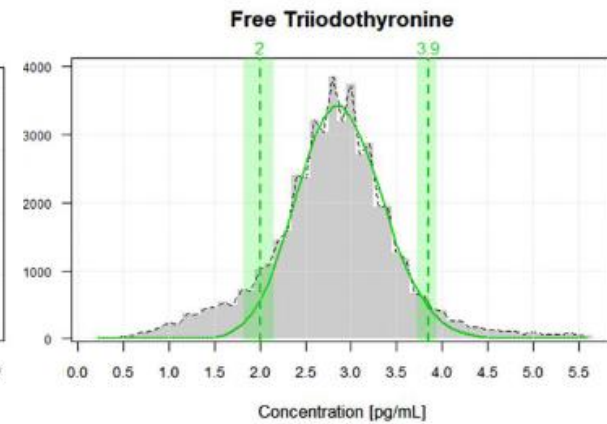
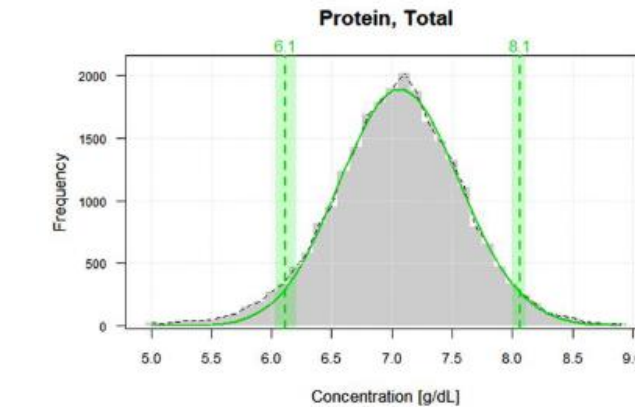
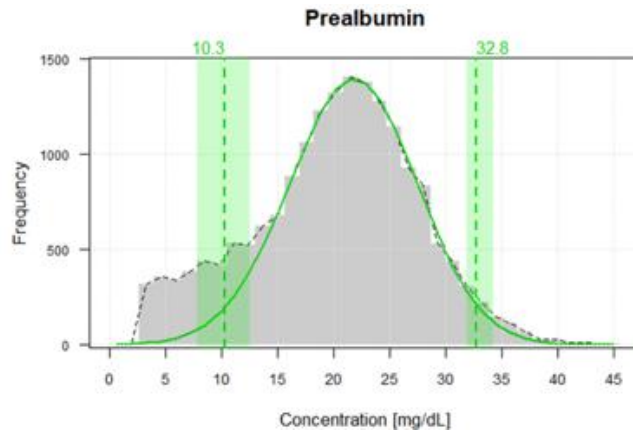
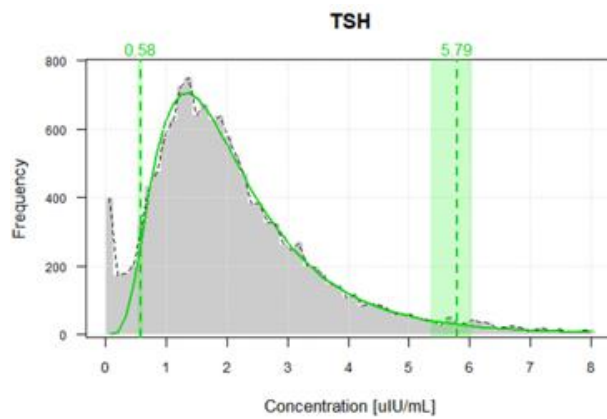
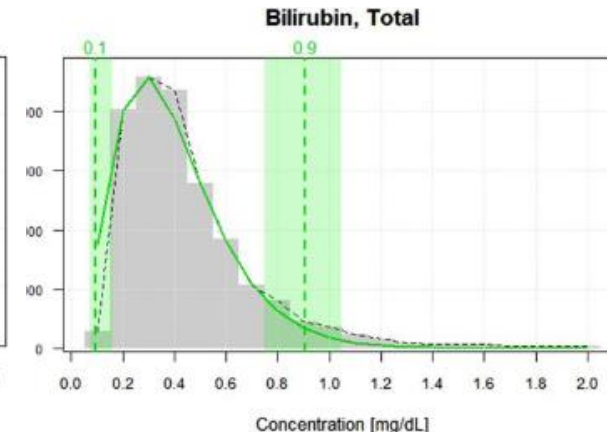
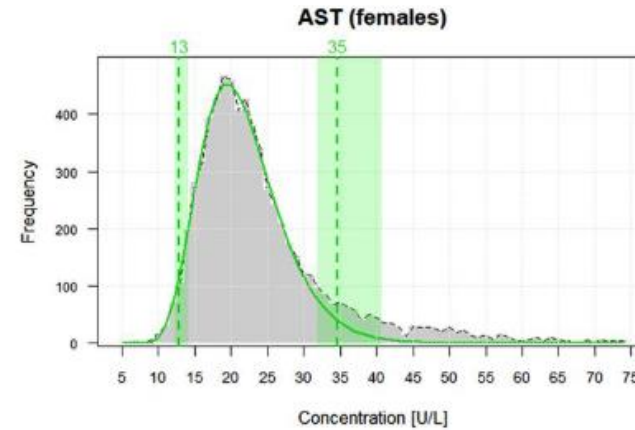
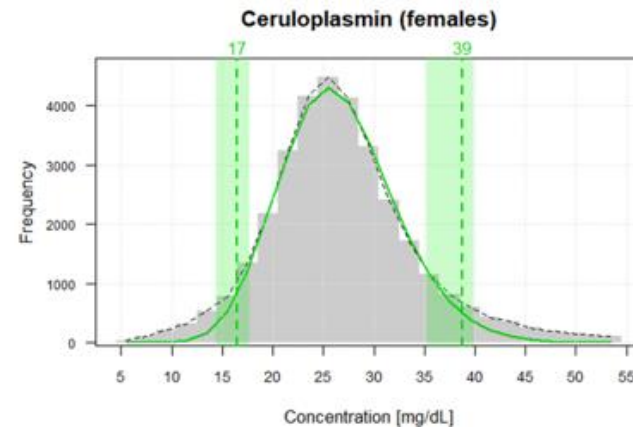
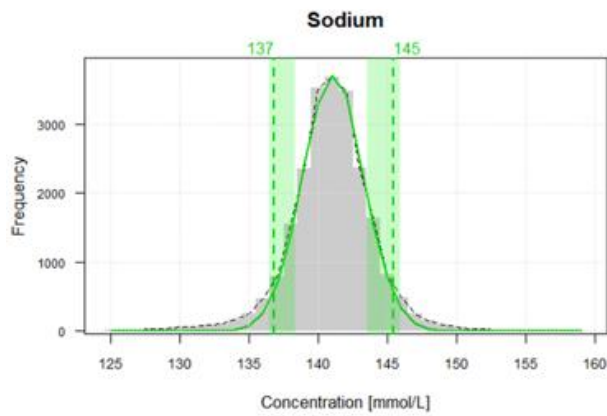
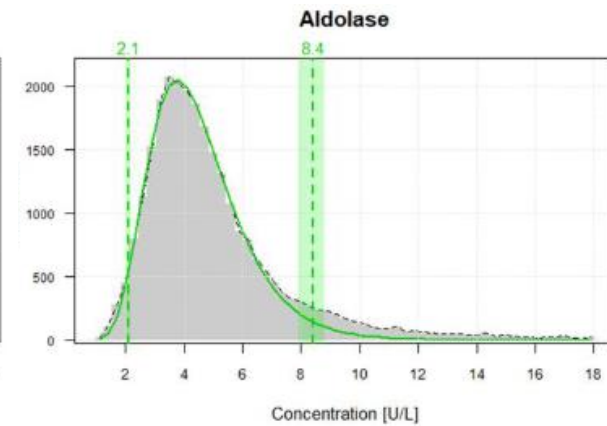
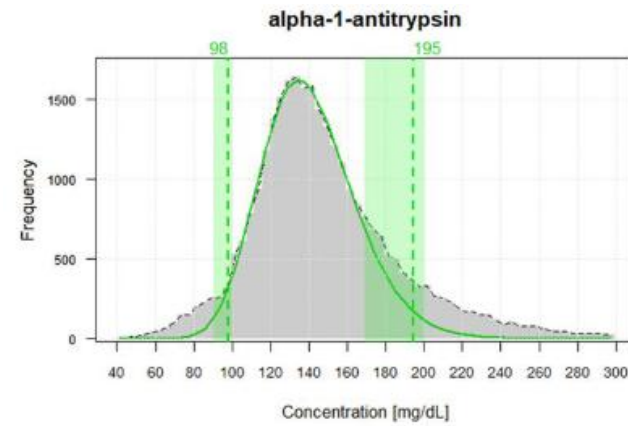
Tatjana Ammer , André Schützenmeister, Christopher M Rank, Kelly Doyle

[Author Notes](#)

*The Journal of Applied Laboratory Medicine*, Volume 8, Issue 1, January 2023, Pages 84–91, <https://doi.org/10.1093/jalm/jfac101>

**Published:** 04 January 2023    **Article history** ▼

# refineR: Various Distributions

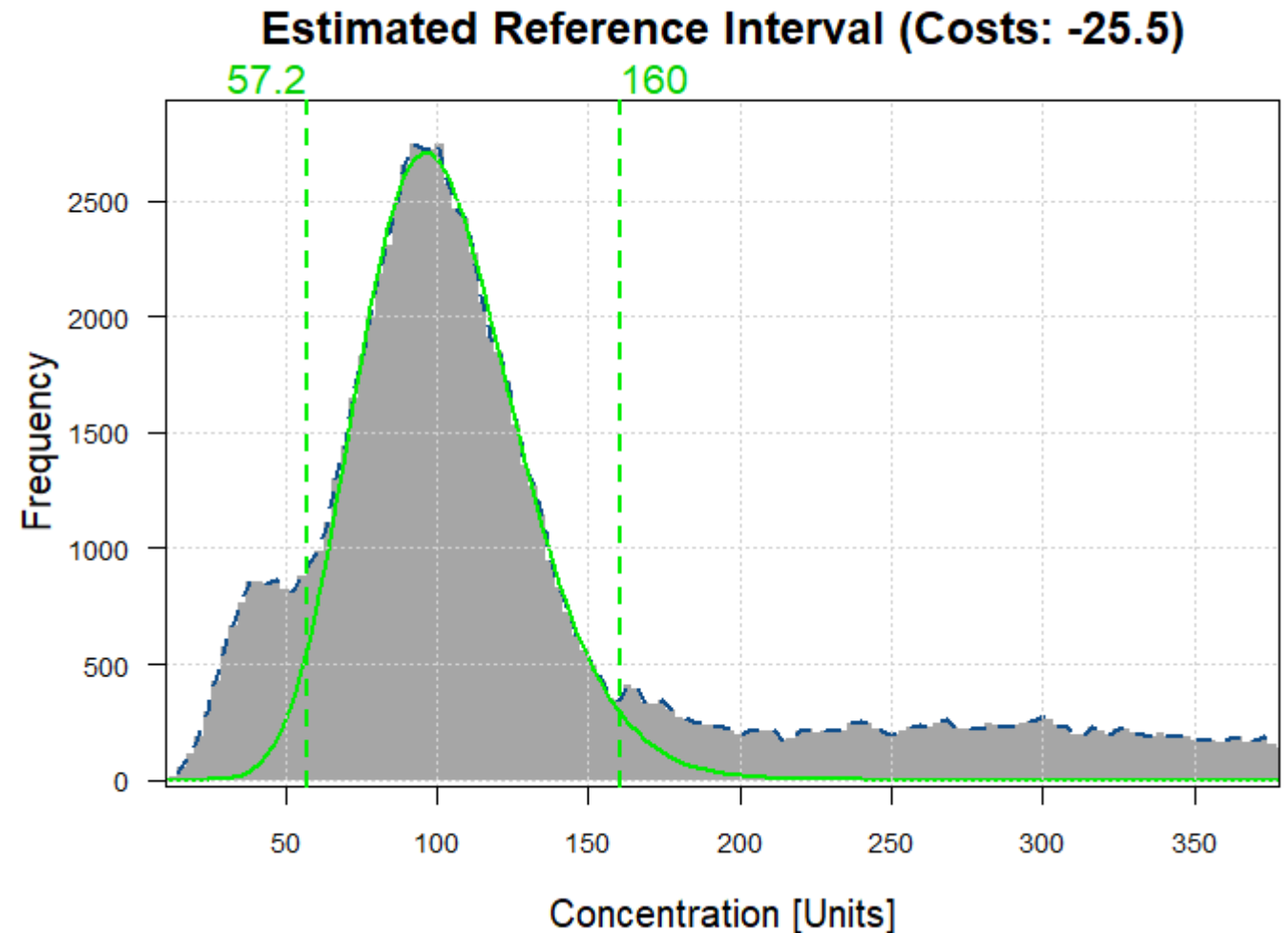




# refineR Package Vignettes

```
library(refineR)  
  
# first example  
data(testcase1)  
resRI <- findRI(Data = testcase1)  
print(resRI)  
plot(resRI, showPathol = FALSE)
```

N = 10,000 simulated measurements  
80%/20% non-pathological to pathological



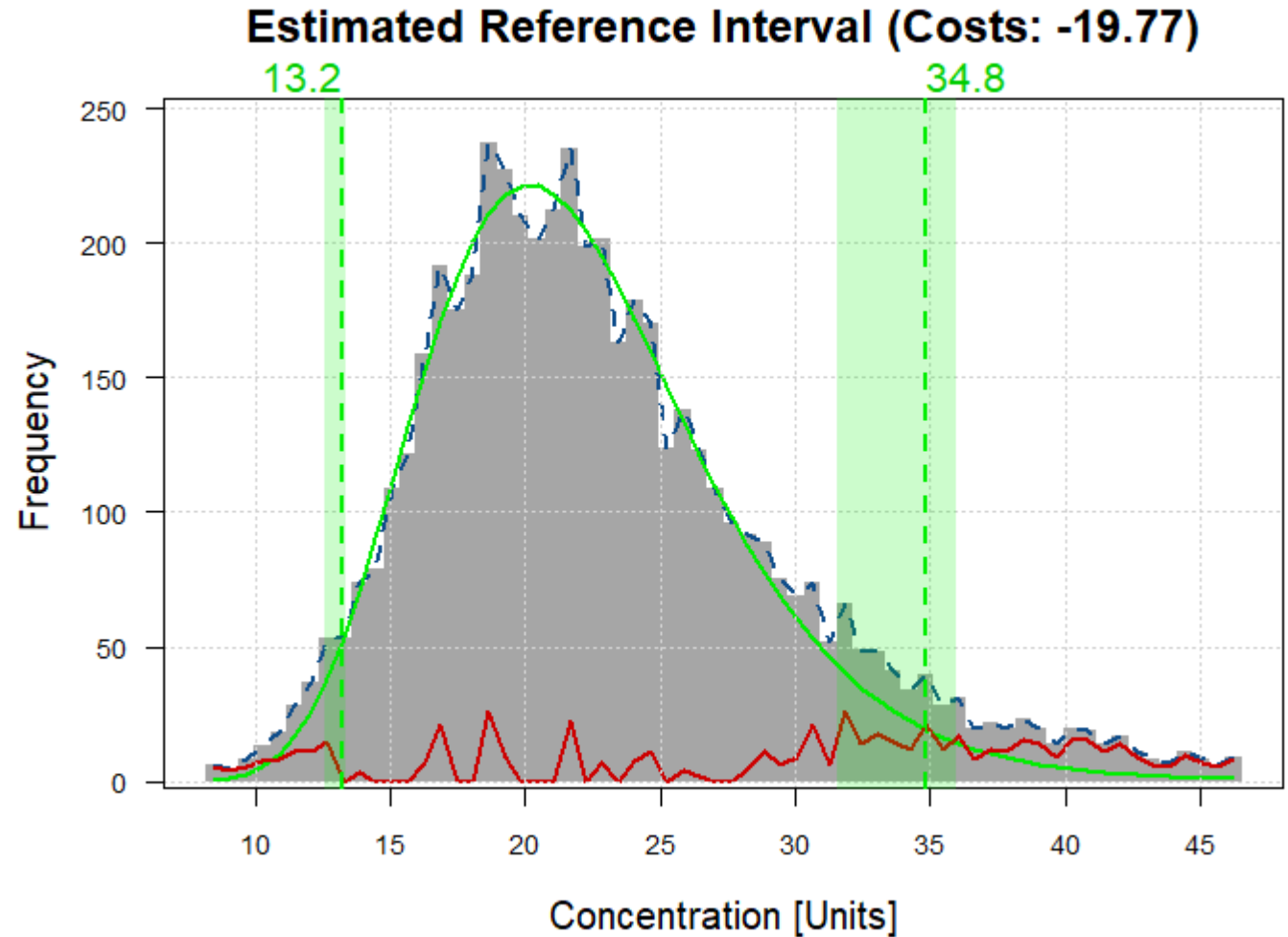
# Example: First Attempt

## Thyroxine Binding Globulin

Package insert claim: 13 – 31  $\mu\text{g}/\text{mL}$  (n = 75)

refineR estimated: 13.2 – 34.8  $\mu\text{g}/\text{mL}$  (N = 5,872)

Literature RI: 15 – 34  $\mu\text{g}/\text{mL}$



# refineR – Getting StaRted

```
# load refineR package
library(refineR)

# To open the help page
?refineR

# Load filtered/cleaned/partitioned data
rawData.DF <- read.csv("andro_all.csv")
Data.DF <- rawData.DF %>% filter(AGE.DECIMAL > 17.99,
                                AGE.DECIMAL < 40.00, SEX == "F")

# Run refineR estimation using findRI
fit <- findRI(Data=Data.DF)

# Print summary of estimated model
Print(fit)

# Run refineR findRI plus bootstrapping
fit.bs <- findRI(Data=Data.DF, NBootstrap=200)

# Print summary of estimated model
print(fit.bs)
```

```
Reference Intervals
-----
lower limit [ 2.5% perc]: 0.366
upper limit [97.5% perc]: 2.82

Model Parameters
-----
method: refineR (v1.6.0)
model: BoxCox
N data: 25101
rounded: yes (base: 0.001)
point est: fullDataEst
lambda: 0.103
mu: 0.0693
sigma: 0.523
shift: 0
cost: -29.9
NP fraction: 0.988
```

```
Reference Intervals
-----
lower limit [ 2.5% perc]: 0.366 (0.335; 0.384)
upper limit [97.5% perc]: 2.82 (2.16; 2.88)

Model Parameters
-----
method: refineR (v1.6.0)
model: BoxCox
N data: 25101
N bootstrap: 200
rounded: yes (base: 0.001)
point est: fullDataEst
lambda: 0.103
mu: 0.0693
sigma: 0.523
shift: 0
cost: -29.9
NP fraction: 0.988
```

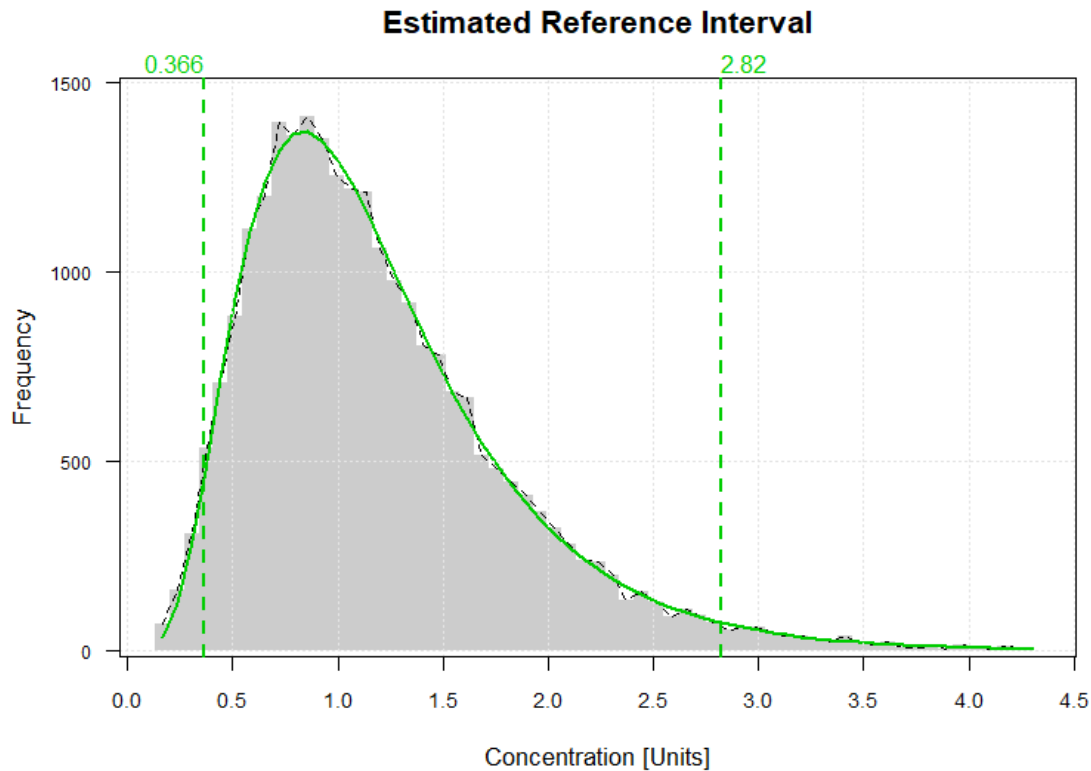
# refineR – Getting StaRted

```
# Run getRI to compute estimates for specified  
percentiles, confidence region, and point estimate  
position  
ri_fit.bs<- getRI(fit.bs, RIperc = c(0.01, 0.025,  
0.975, 0.99), CIprop = 0.95,  
pointEst ="fullDataEst"))
```

	Percentile	PointEst	CILow	CIHigh
1	0.010	0.2953126	0.2585747	0.3165687
2	0.025	0.3660094	0.3350618	0.3841925
3	0.975	2.8189462	2.1591106	2.8829025
4	0.990	3.3434405	2.4504120	3.4453127

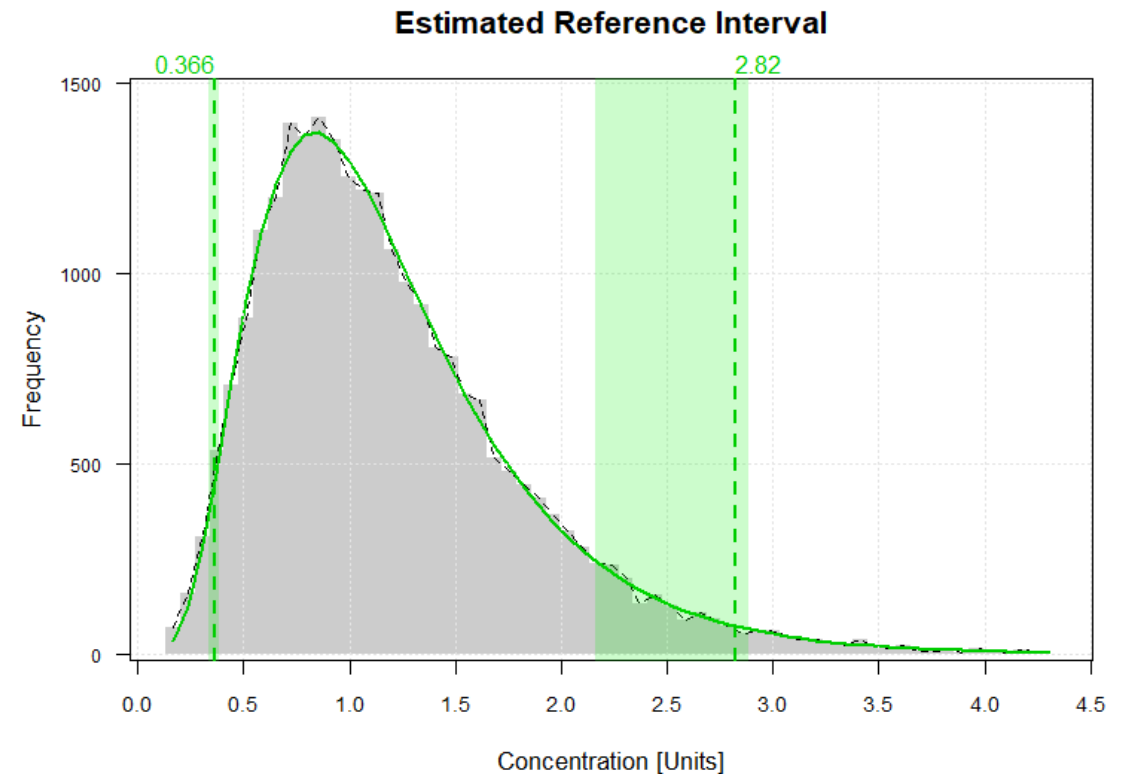
# refineR – Androstenedione Example

```
# Default plot function  
plot(fit)
```



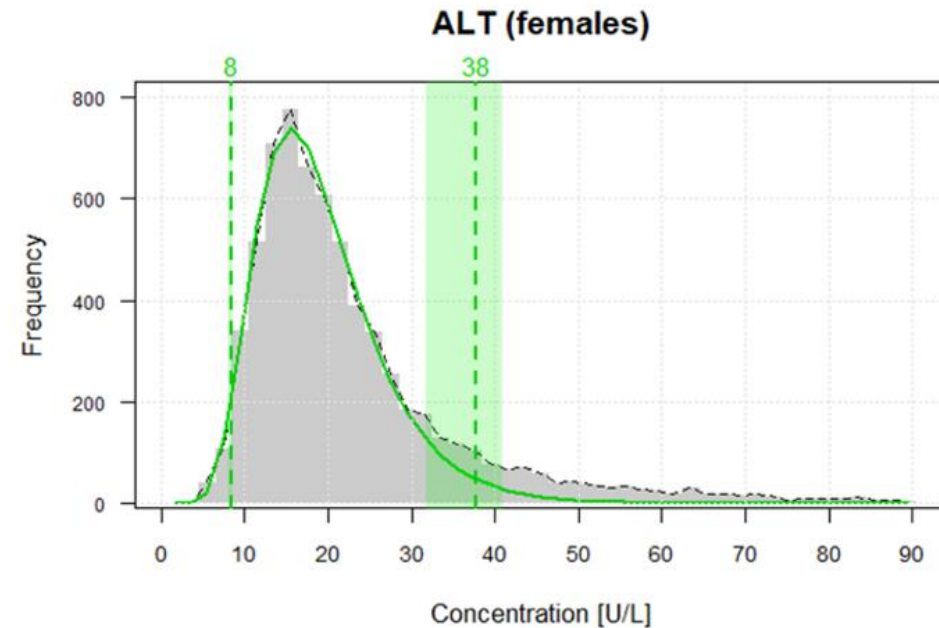
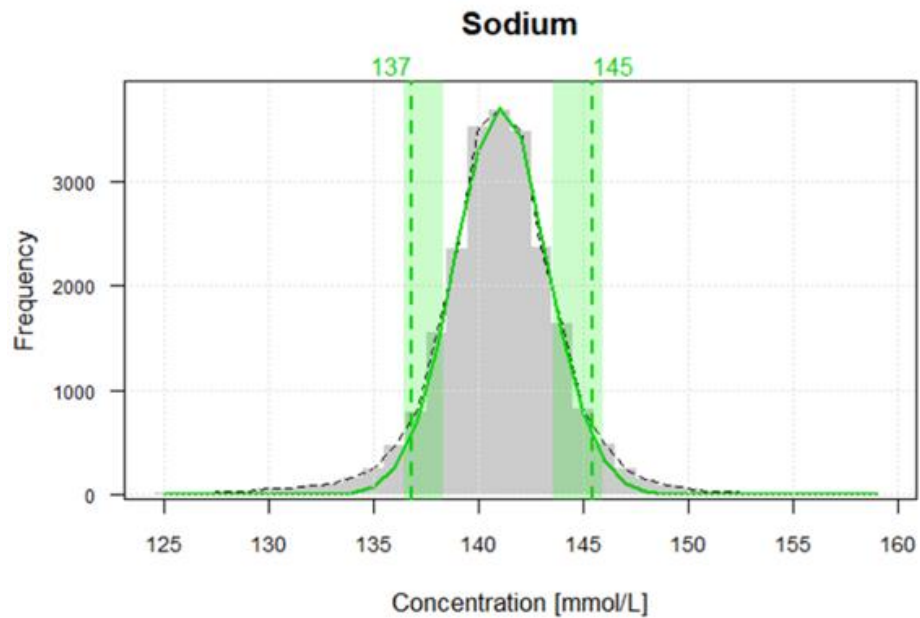
```
Reference Intervals  
-----  
lower limit [ 2.5% perc]: 0.366  
upper limit [97.5% perc]: 2.82
```

```
# Default plot function  
plot(fit.bs)
```



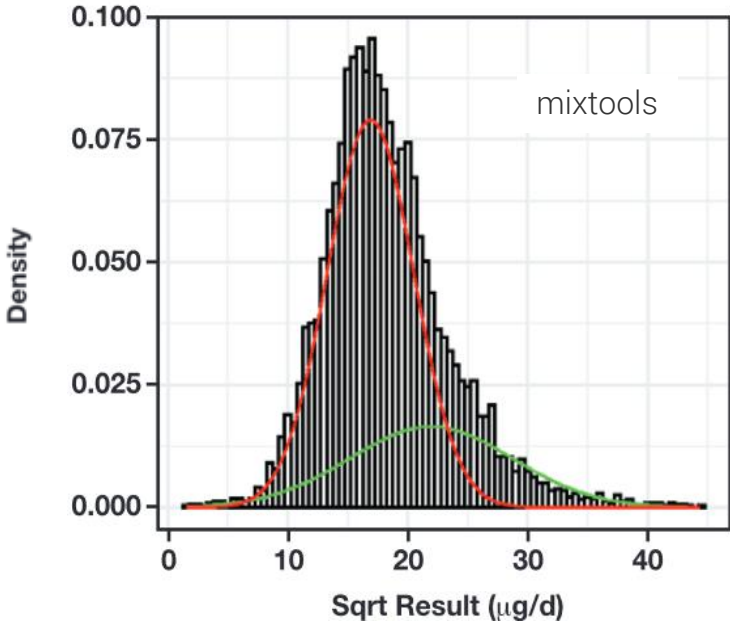
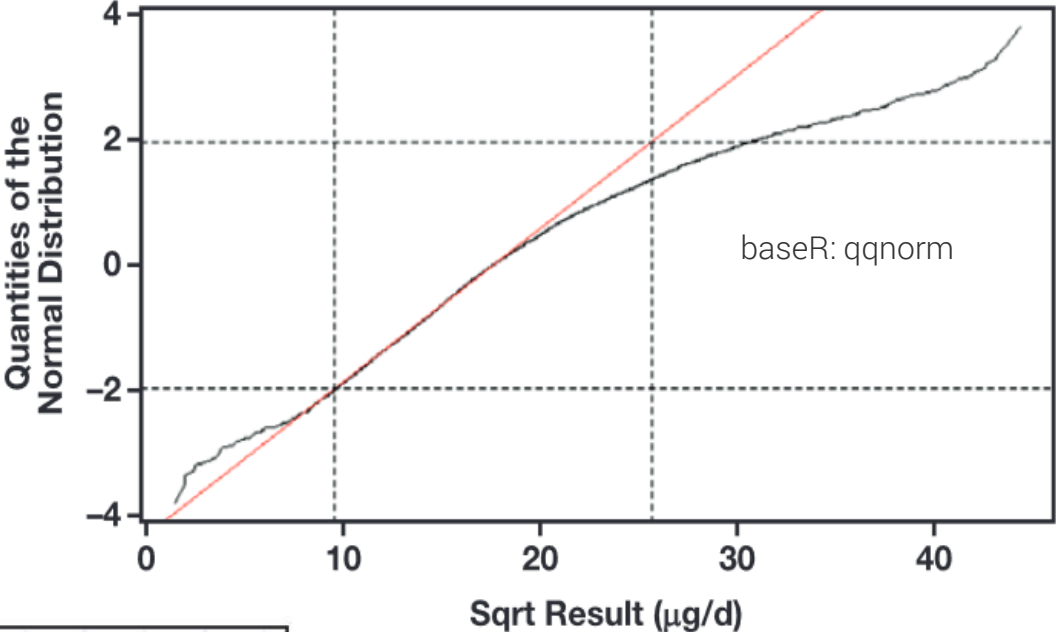
```
Reference Intervals  
-----  
lower limit [ 2.5% perc]: 0.366 (0.335; 0.384)  
upper limit [97.5% perc]: 2.82 (2.16; 2.88)
```

# Indirect RI Estimation Normal and Skewed Data Sets

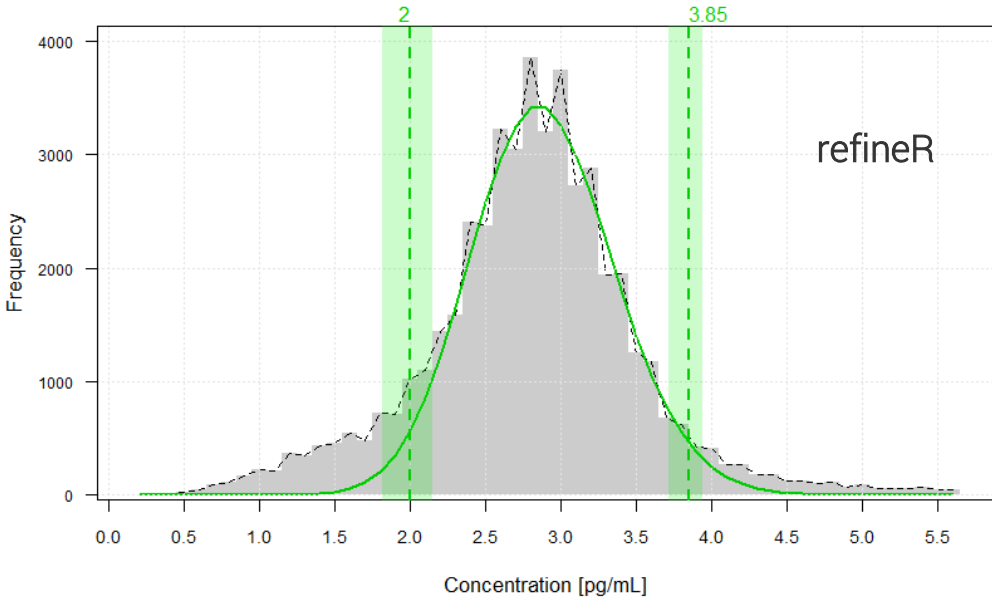


# Working with Skewed Distributions -- Transformed Normal Approach

- Manual transformation
- Evaluate normalcy (e.g., QQ-plot)
- Estimation of ranges (e.g., Hoffman, mixtools)
- Not recommended

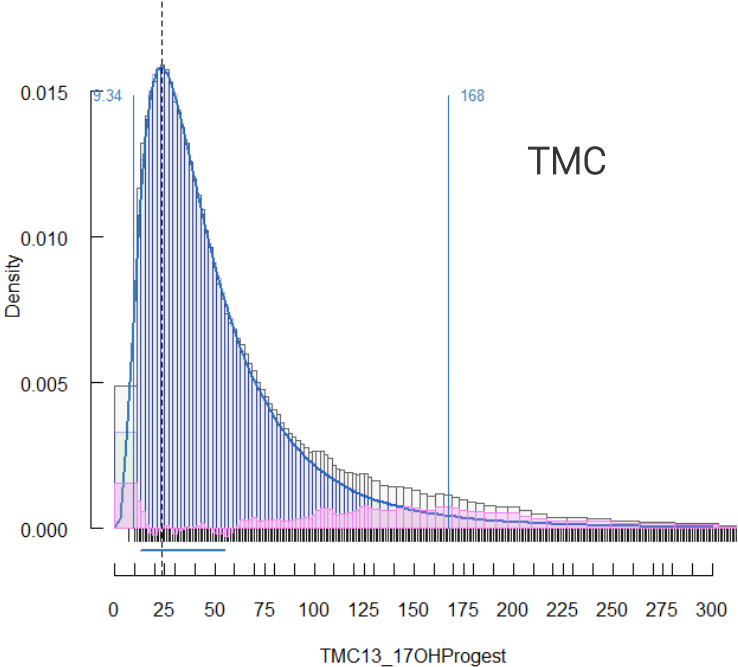
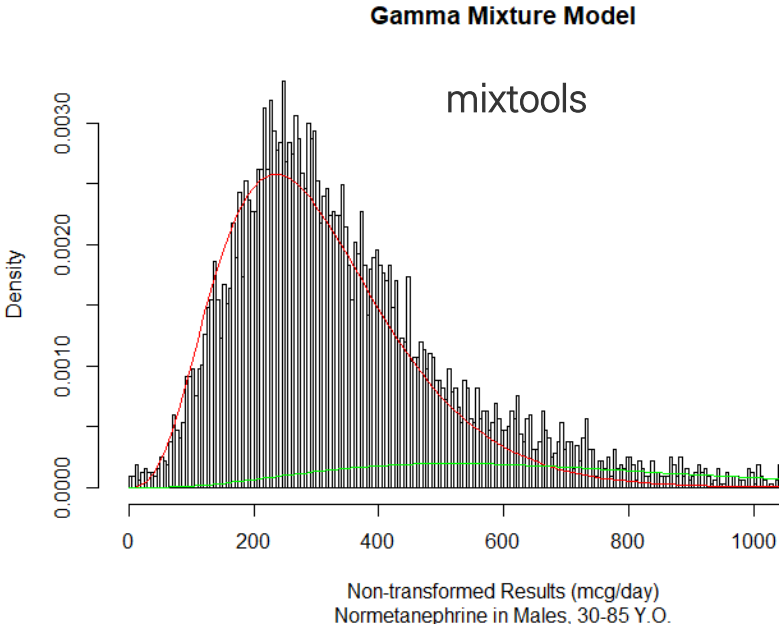


# Skewed Distributions: Gamma Model or Iterative Fitting



`library(mixtools)`  
-non-Gaussian gamma mixture model

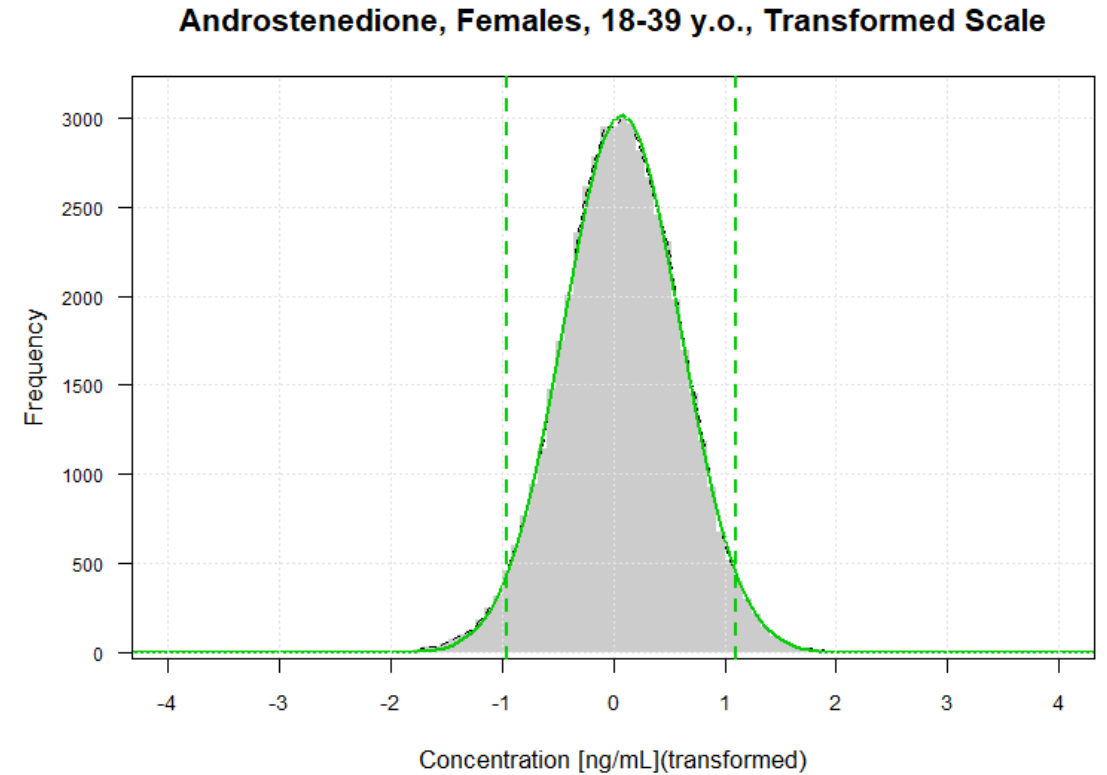
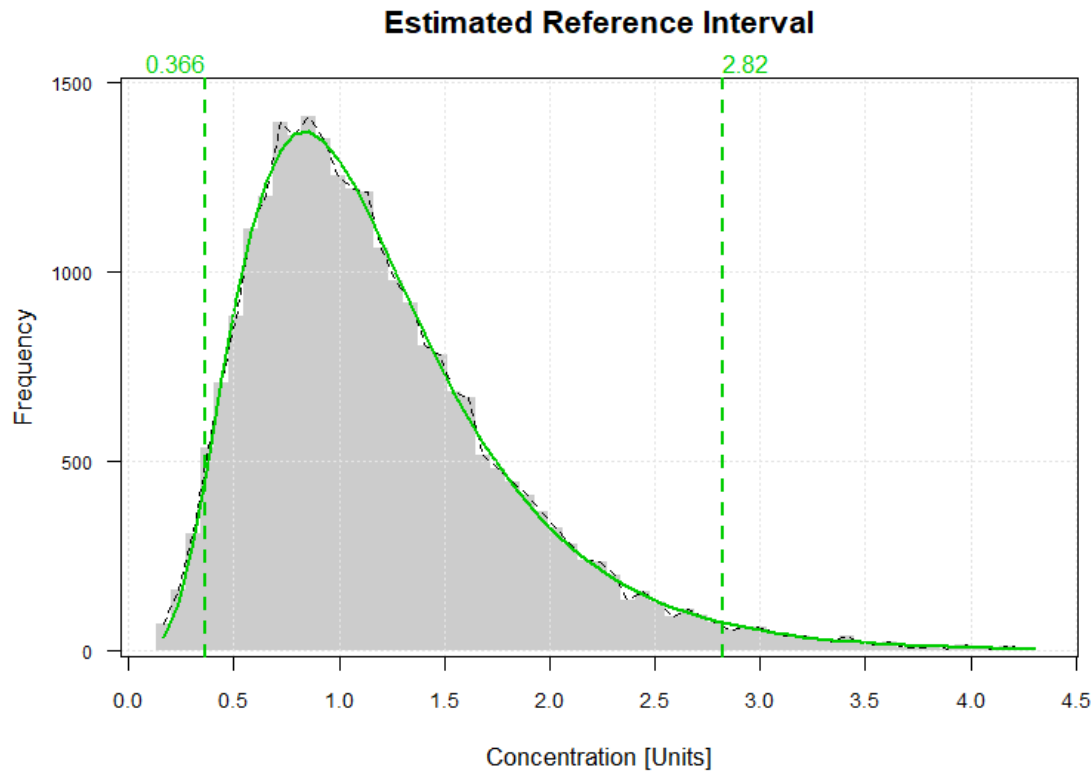
`library(refineR)`  
TMC program (multiple R packages)





# Lightly and Heavily Skewed Data - refineR

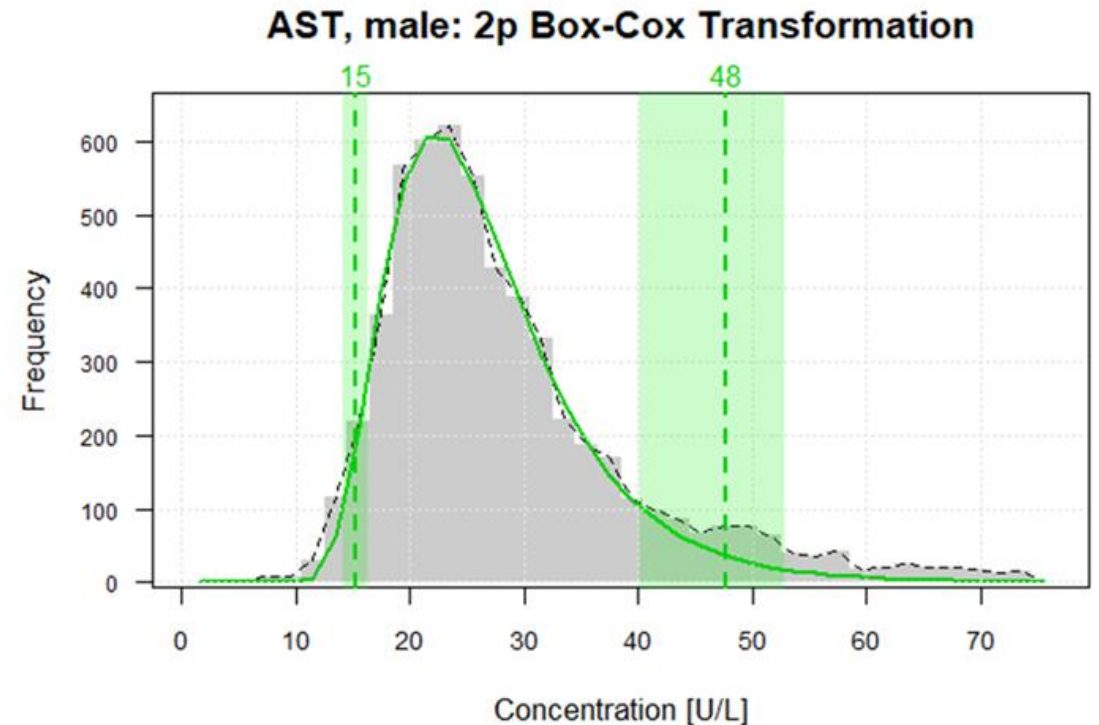
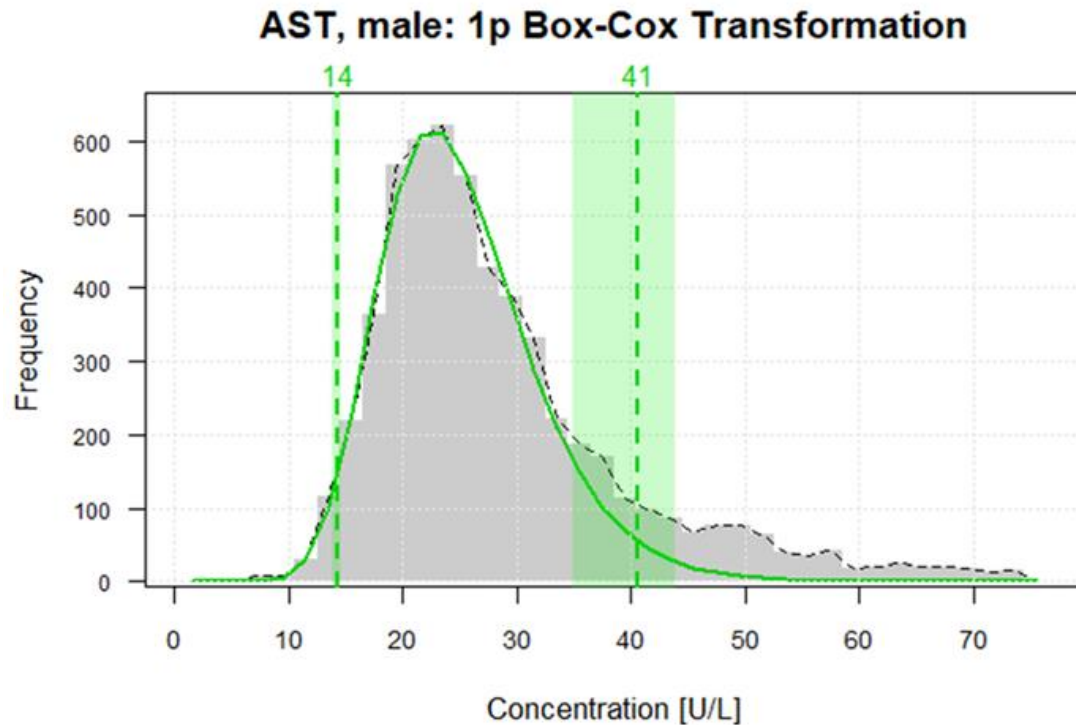
`library(refineR)` – 1 parameter Box-Cox (default)



refineR allows for graphing in transformed space.  
Shift away from 0 suggests trying 2 –parameter transformation.

# refineR (1.5) : Two-Parameter Box-Cox

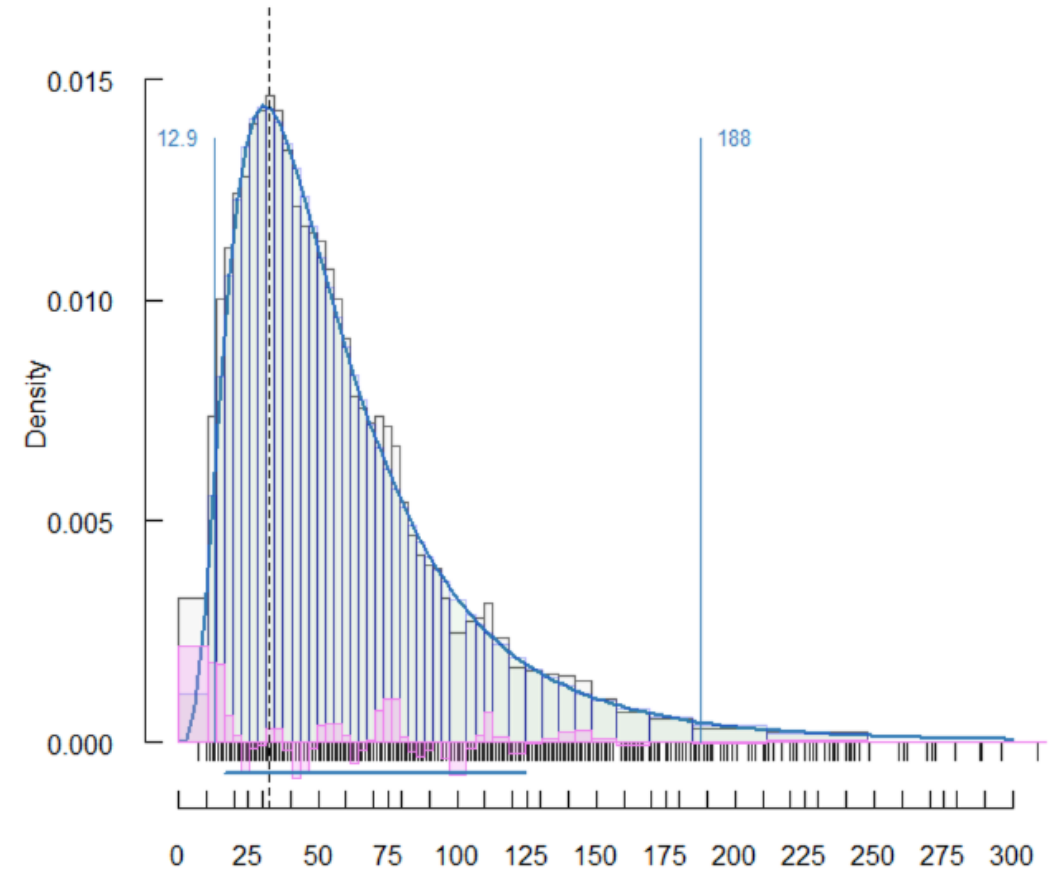
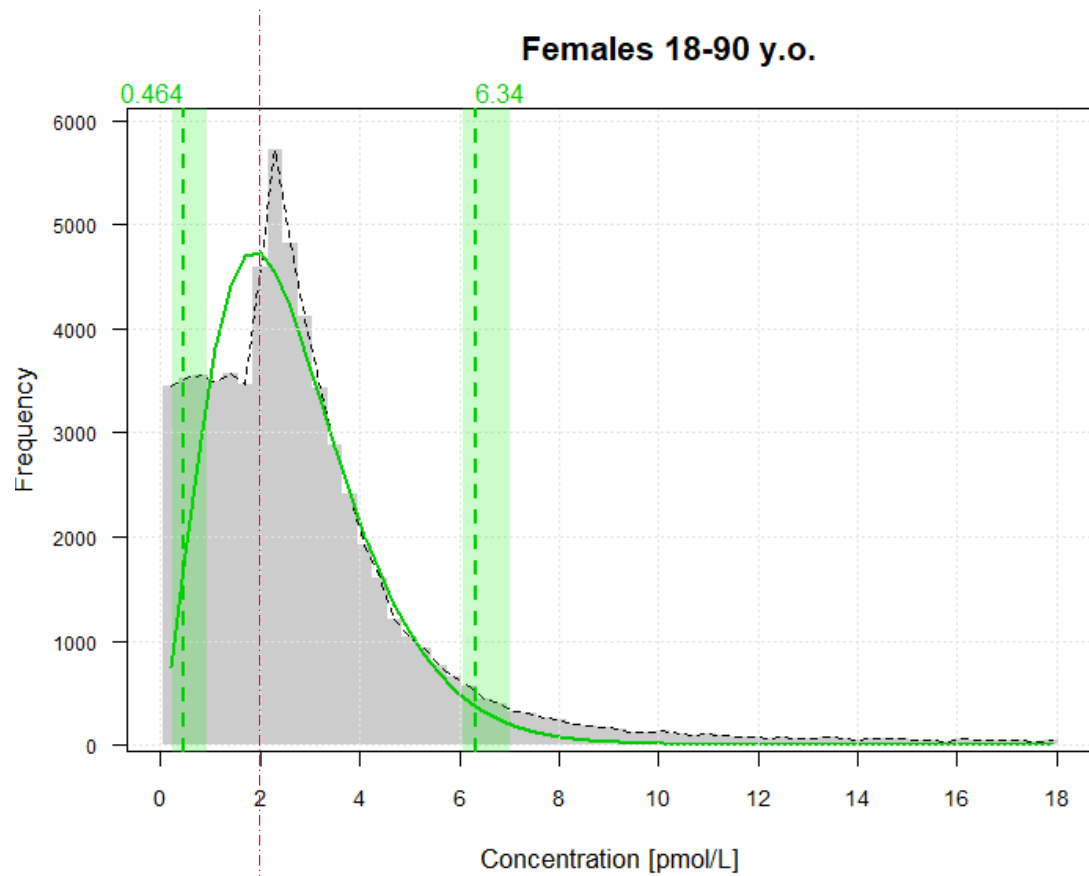
library(refineR) – 2 parameter Box-Cox



```
fit.bs <- findRI(Data=Data.DF, model = "BoxCox",  
NBootstrap=200)
```

```
fit.bs <- findRI(Data=Data.DF, model = "ModBoxCox",  
NBootstrap=20)
```

# Indirect RI Estimation: Partial Distributions (Limited Assay Sensitivity)



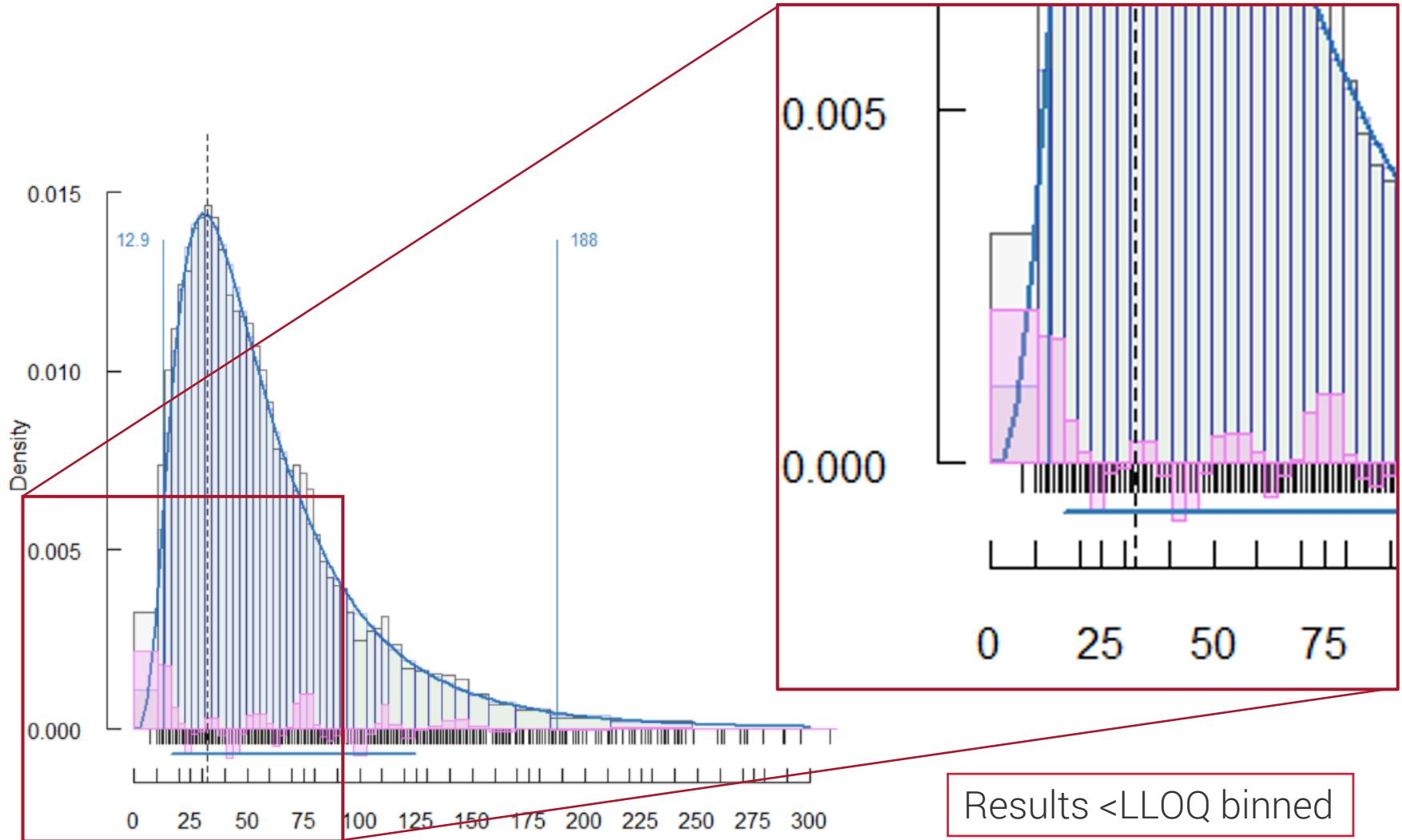
2023-03-13\_090748 / TMC13\_17OHProgesteronedata\_c2\_oAll\_dAll\_F\_16-17 / n = 3158

# TMC Package in R for Partial Distributions

17-OH Progesterone

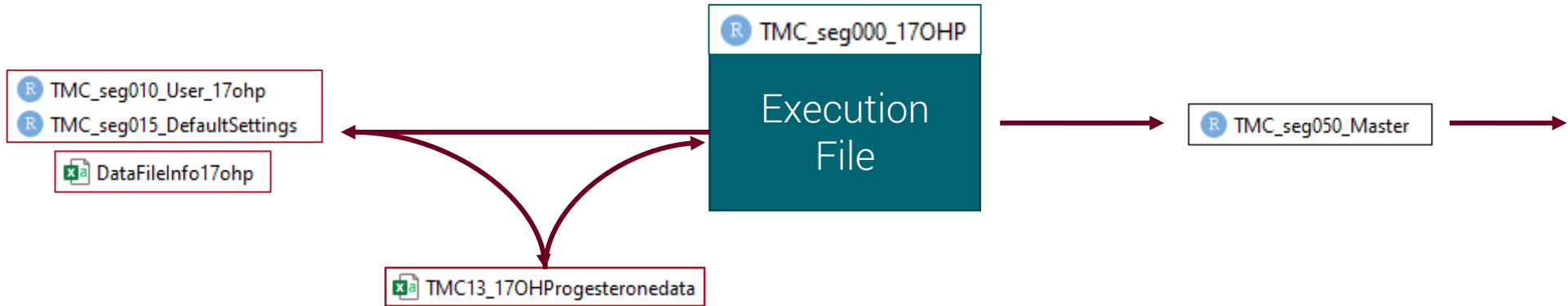
5% of results are <LLOQ

TMC can handle "<" results as its own bin



TMC13\_17OHProgest  
2023-03-13\_090748 / TMC13\_17OHProgesteronedata\_c2\_oAll\_dAll\_F\_16-17 / n = 3158

# TMC Package Workflow in R



- R TMC\_seg051\_AgeClassFix3
- R TMC\_seg051\_AgeClassFix4
- R TMC\_seg051\_DynAgeClass
- R TMC\_seg052\_AgeClassDyn3
- R TMC\_seg052\_AgeClassDyn4
- R TMC\_seg056\_MasterPlots
- R TMC\_seg057\_MasterTables
- R TMC\_seg060\_RemoveOldFiles
- R TMC\_seg070\_ReadData
- R TMC\_seg075\_FinalParSettings
- R TMC\_seg080\_NamesPerFile
- R TMC\_seg089\_OpenWindows
- R TMC\_seg090\_Plot\_Age\_Val
- R TMC\_seg094\_OpenWindows
- R TMC\_seg095\_Plot\_DateTime\_Val
- R TMC\_seg099\_OpenWindows
- R TMC\_seg100\_Analysis
- R TMC\_seg100\_Analysis\_2022-10-06
- R TMC\_seg101\_Analysis\_Ini
- R TMC\_seg101\_Ini\_Analysis\_V2
- R TMC\_seg102\_Analysis\_npa
- R TMC\_seg104\_Analysis\_qqw
- R TMC\_seg105\_Analysis\_tmc
- R TMC\_seg106\_Analysis\_tmu
- R TMC\_seg108\_Analysis\_tuk
- R TMC\_seg109\_Analysis\_qqs
- R TMC\_seg109\_Analysis\_qqw
- R TMC\_seg109\_TMC\_BS
- R TMC\_seg110\_NamesPerStratum
- R TMC\_seg149\_OpenWindows
- R TMC\_seg150\_PlotRL\_vs\_Age\_by\_S
- R TMC\_seg159\_OpenWindows
- R TMC\_seg160\_PlotRL\_vs\_Age
- R TMC\_seg201\_OpenWindows
- R TMC\_seg202\_SimResults
- R TMC\_seg202\_SimResultsA
- R TMC\_seg202\_SimResultsB
- R TMC\_seg202\_SimResultsC
- R TMC\_seg202\_SimResultsD
- R TMC\_seg203\_SimResultsA
- R TMC\_seg211\_NamesPerScena

DataFileInfo17ohp

FileNo;FileName;Path;Label;Value;Rounding;Sex;SexCodeM;SexCodeF;Age;DateTime;Device;DecChar

0001;TMC13\_17OHPProgesteronedata.csv;U:/Stats math/Reference Intervals/Indirect RI/Tools/TMC/TMC13/data/;TMC13\_17OHPProgest;2;1;4;M;F;3;5;6;

	A
1	PatId;Value;Age;Sex;DateTime;Device
2	0001;56;27;F;01.01.2016 03:22;MS1
3	0002;75;15;F;01.01.2016 03:22;MS2

Dataset

# Indirect RI Estimation - Many RI Partitions

Example: Endocrine-related analytes often have sex- and age-specific partitions

- Analysis of each partition can lead to repetitive and time-consuming code writing

Solution: Code functions that loop over parameters from a table/spreadsheet

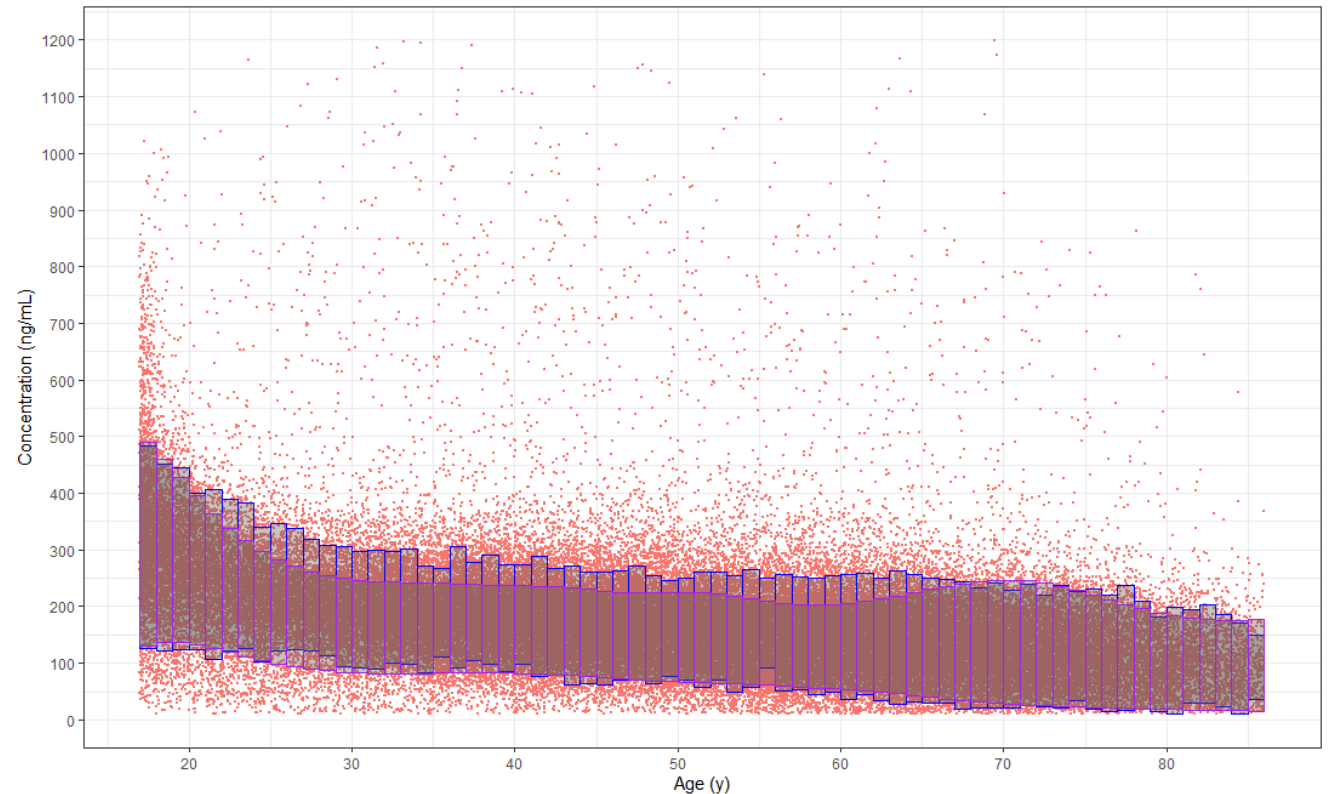
```
Functions
twoparFUN function (df, mu1, mu2, sd1, sd2, co, sex_sel, age_bin_start, age_bin_end)

params
$ .df      :
$ dataset.name :
$ .mu1     :
$ .mu2     :
$ .sd1     :
$ .sd2     :
$ .co...7  :
$ .data    :
$ .mu1_val :
$ .mu2_val :
$ .sd1_val :
$ .sd2_val :
$ .co      :
$ age_bin_start:
$ age_bin_end:
$ sex      :
```

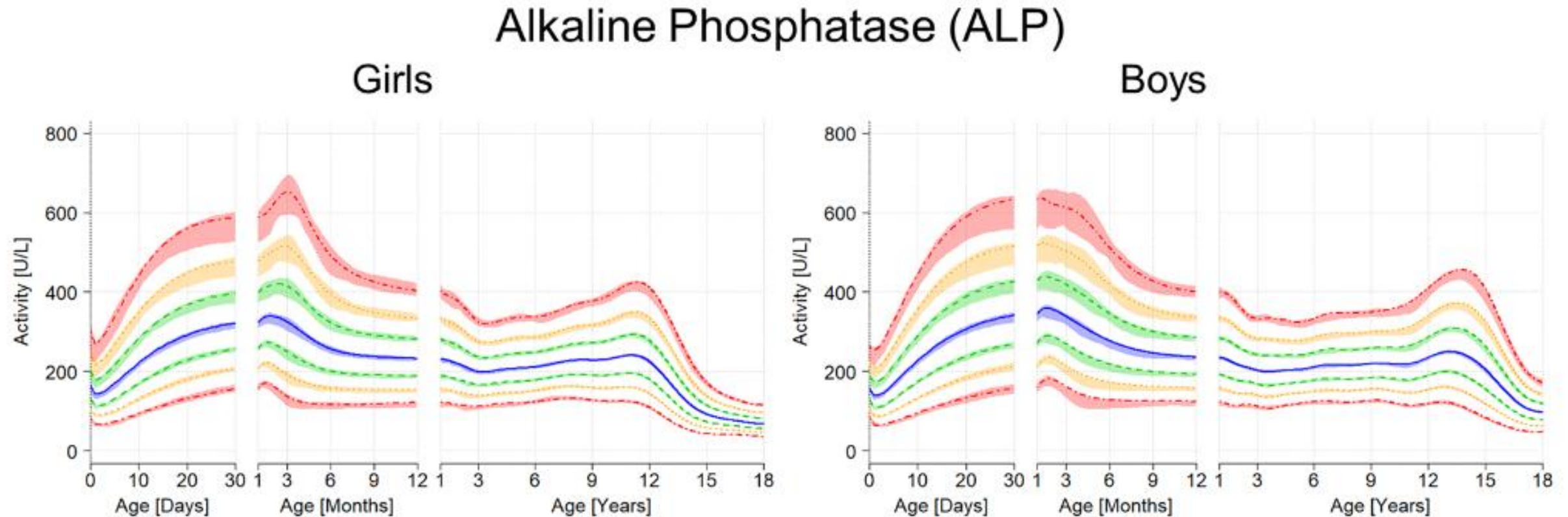
```
df
$ id      :
$ sex     :
$ years_old :
$ _result :
$ collection:
$ verified :
$ _result:
```



Crock pot method  
"Set it and forget it!"



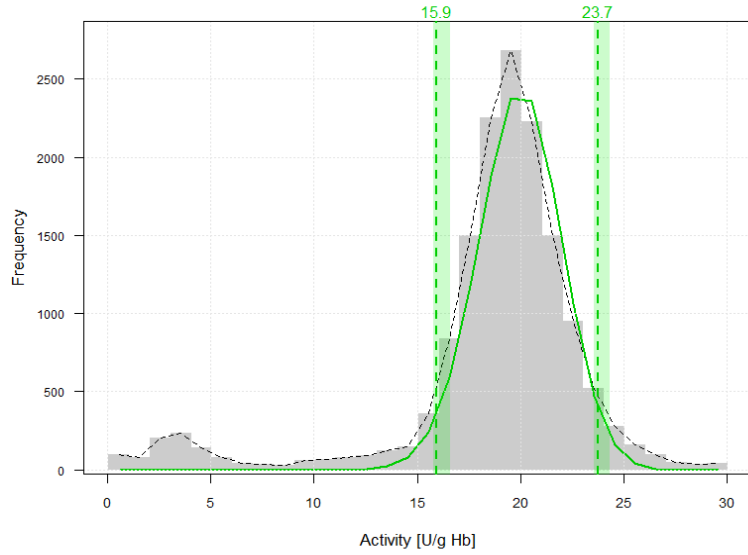
# Continuous Indirect Reference Interval Estimation



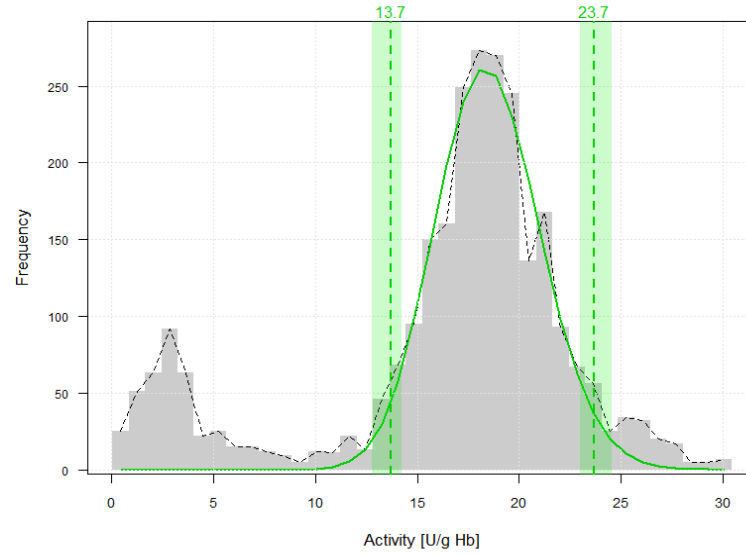
R Packages used: Runner pipeline = gamlss + refineR

# Example of Verifying Reference Intervals – G6PD Activity

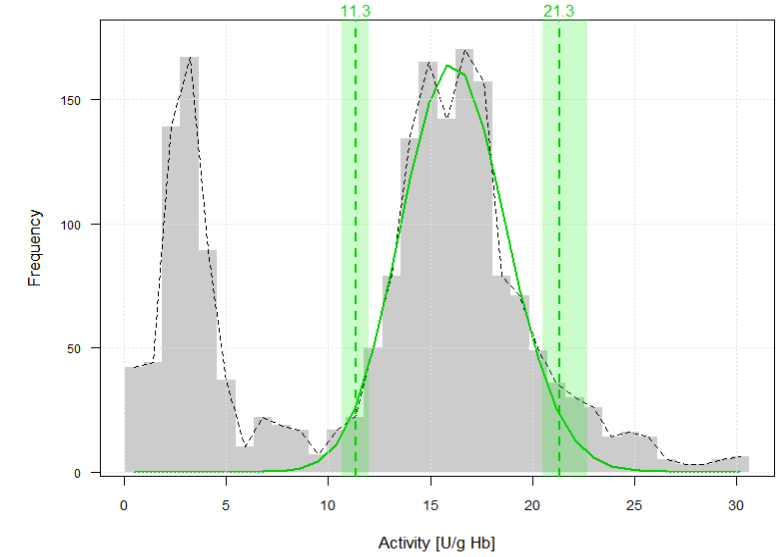
**cG6PD, 0-7 d.o.**



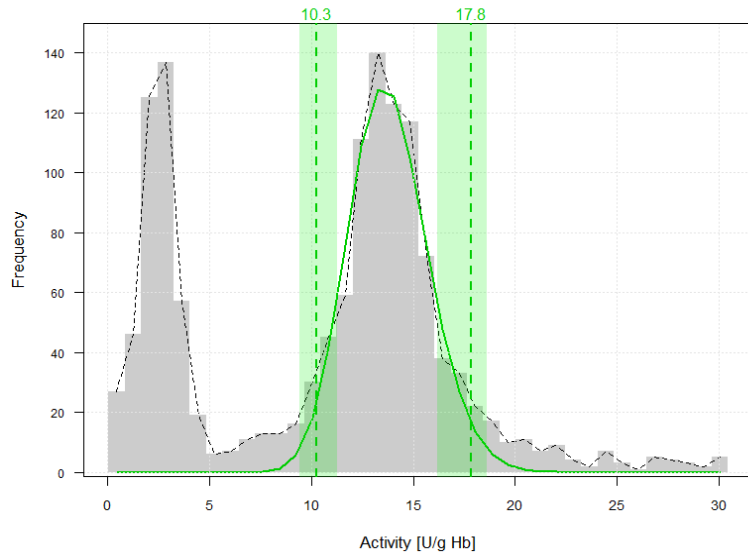
**cG6PD, 8 - 30 d.o.**



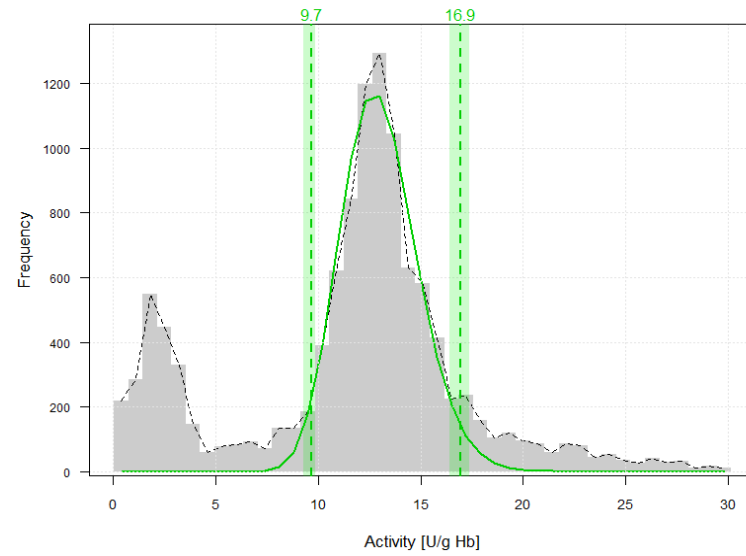
**cG6PD, 1-5 m.o.**



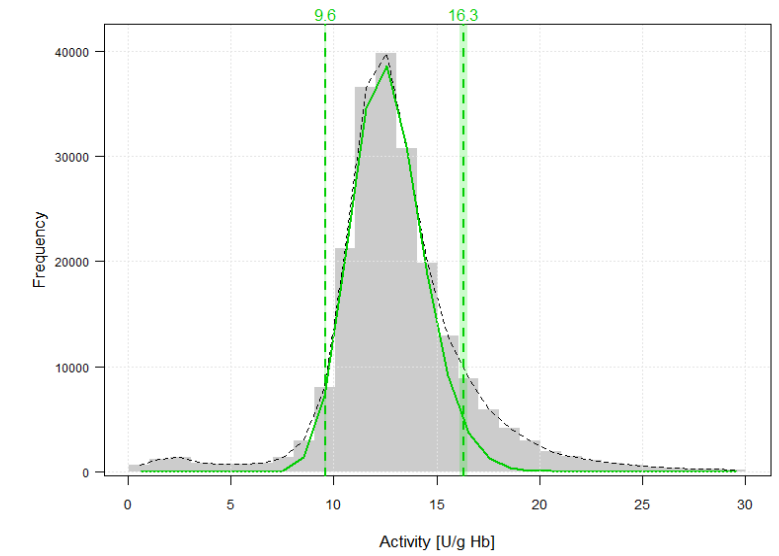
**cG6PD, 6-12 m.o.**



**cG6PD, 1 y.o.-17 y.o.**

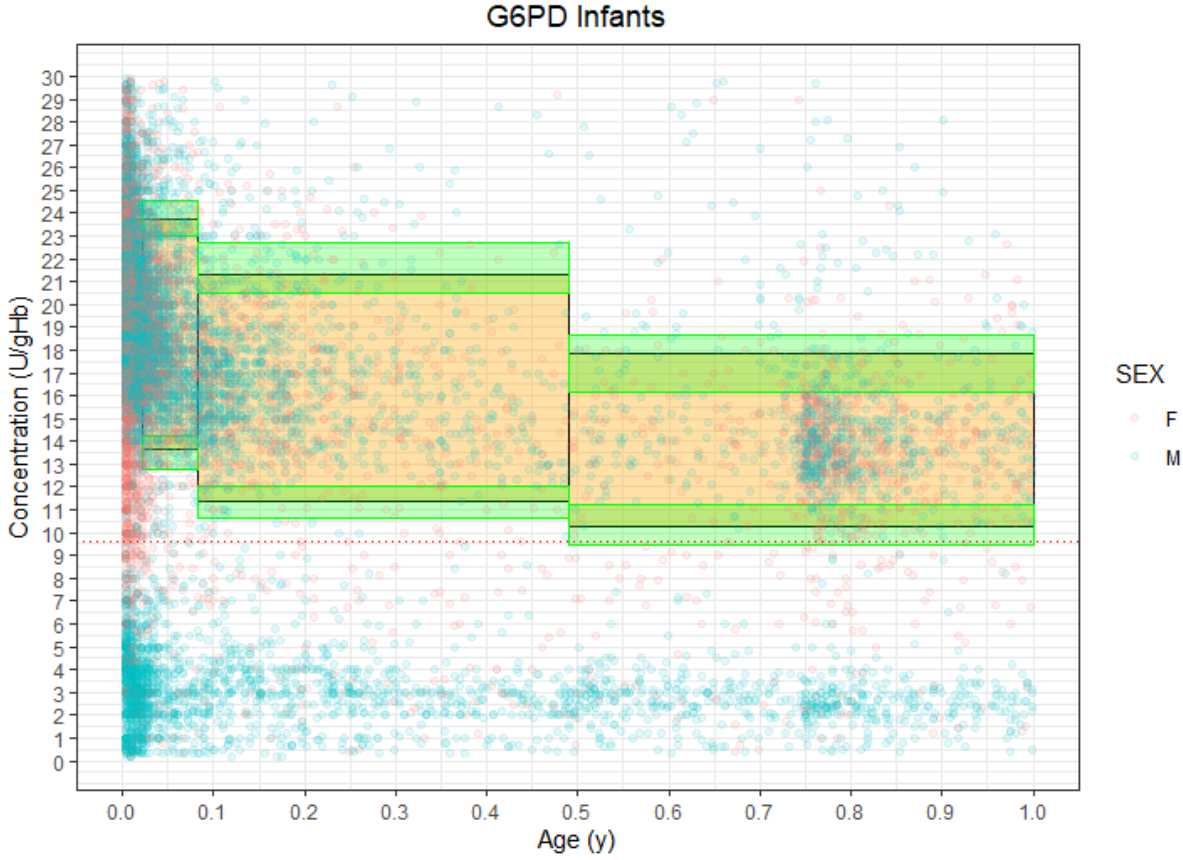
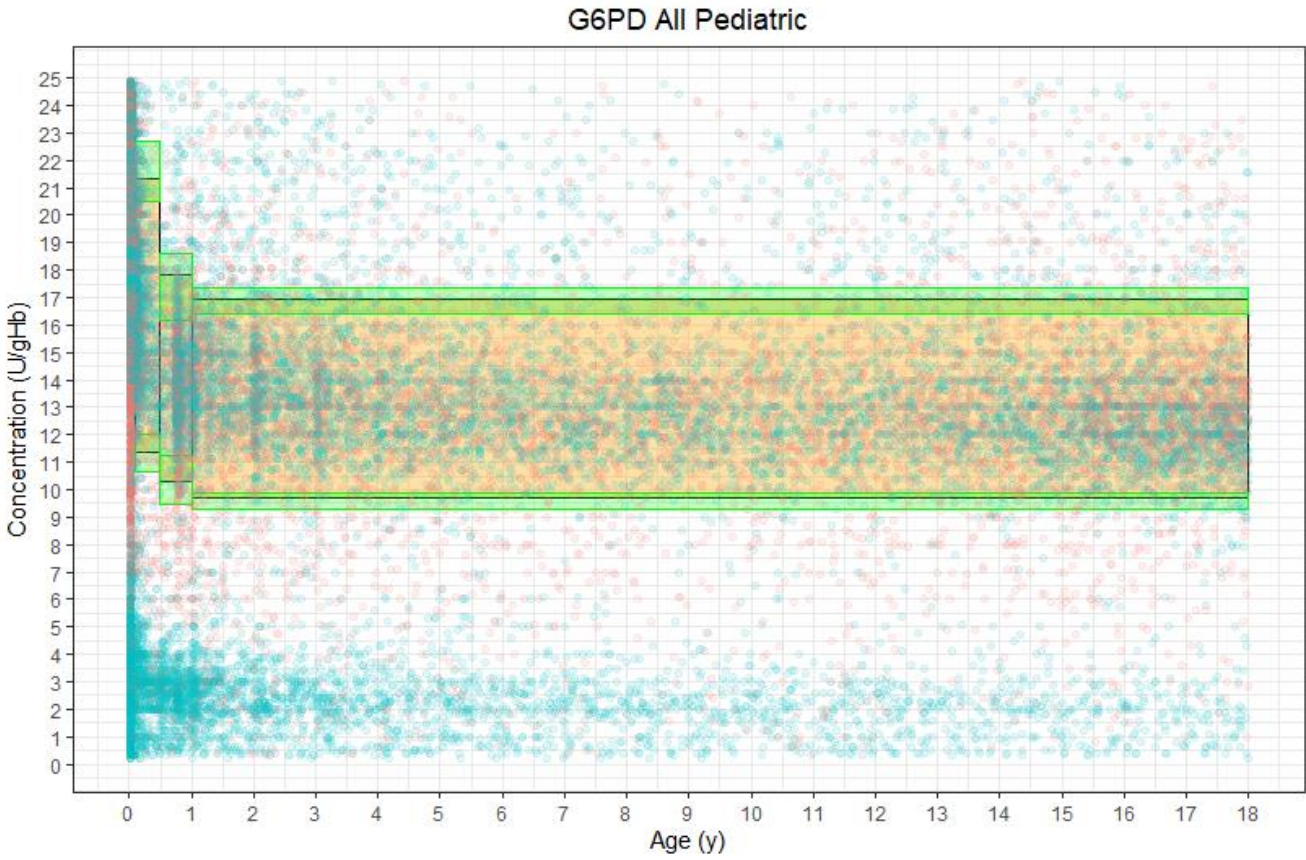


**cG6PD, 18 - 89 y.o.**



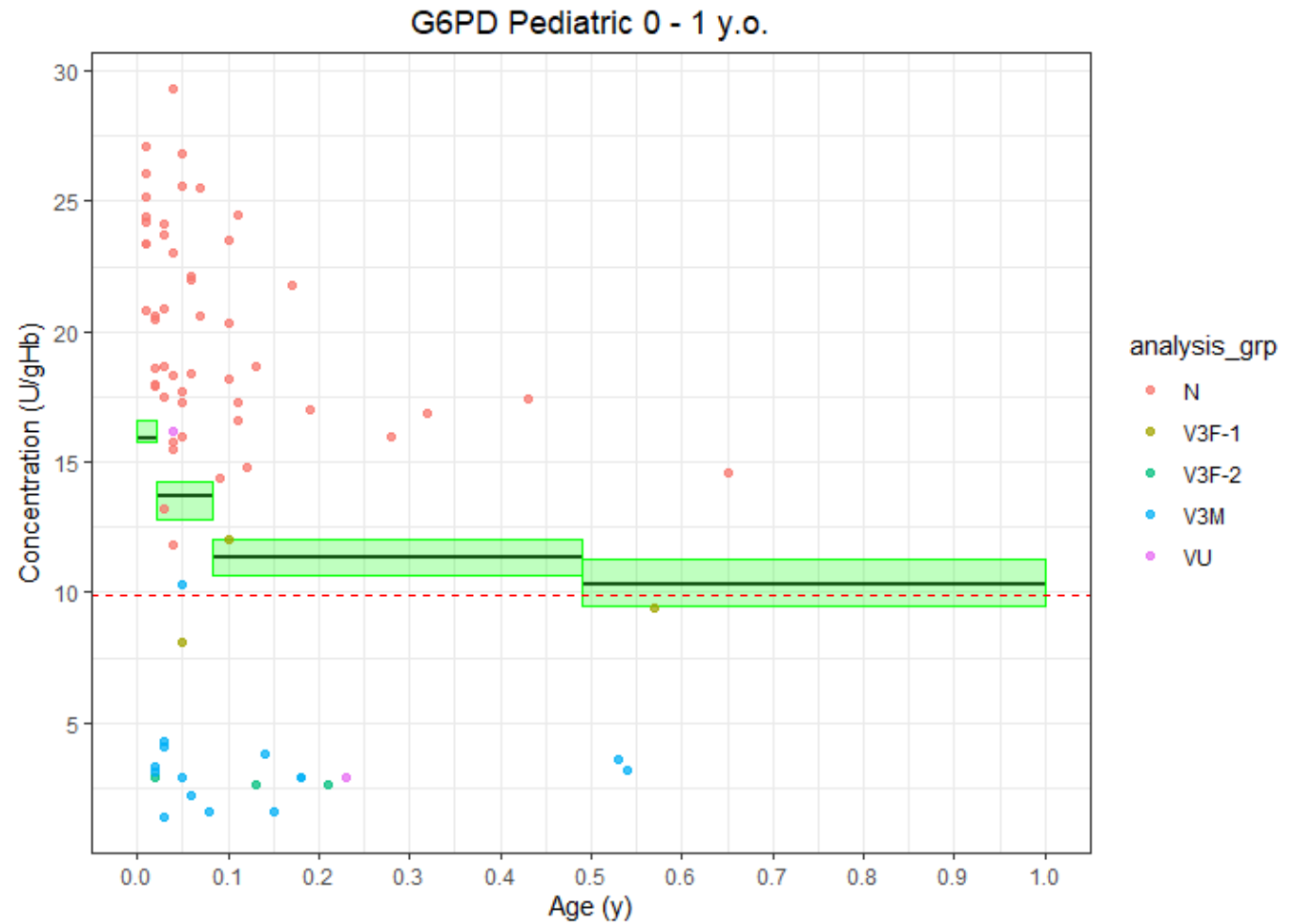
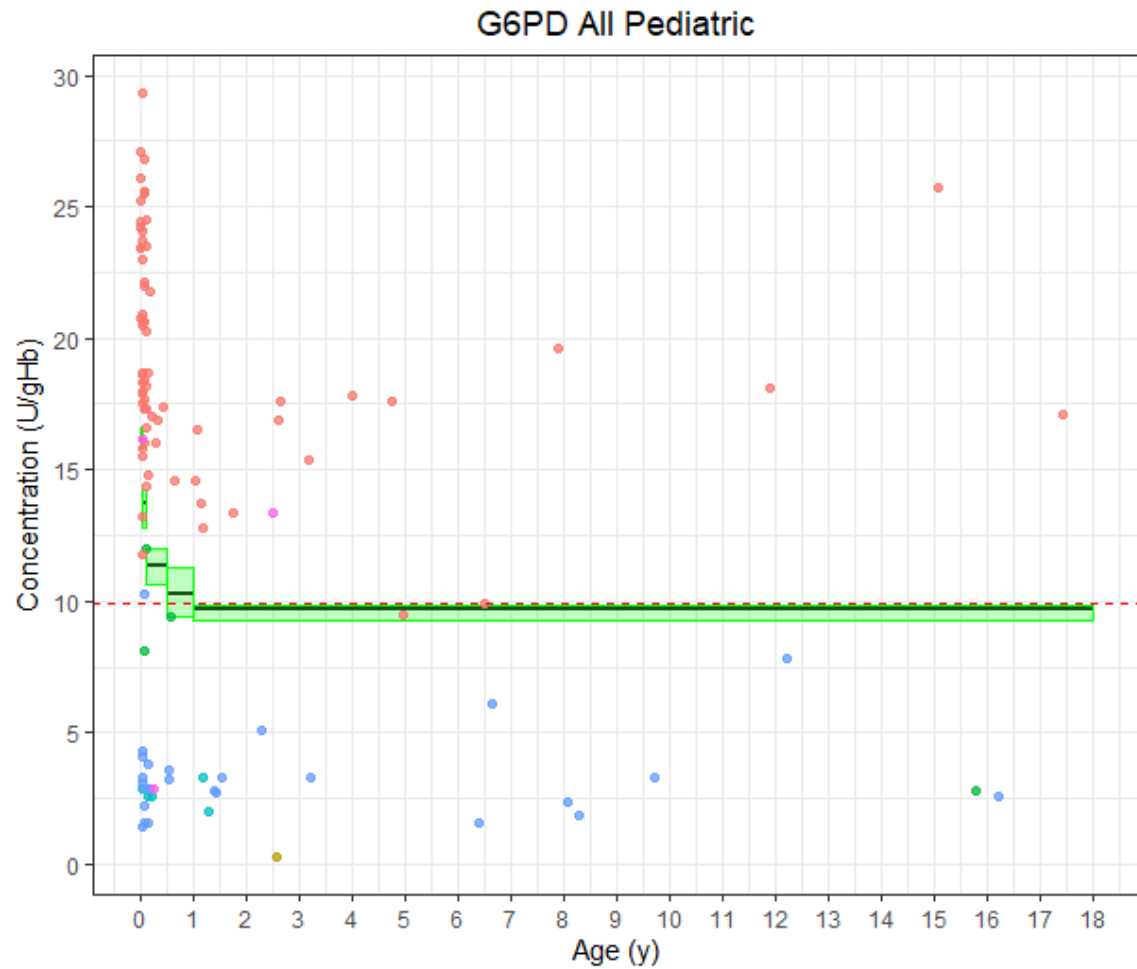


Estimated reference intervals coincide with dot plot trends...



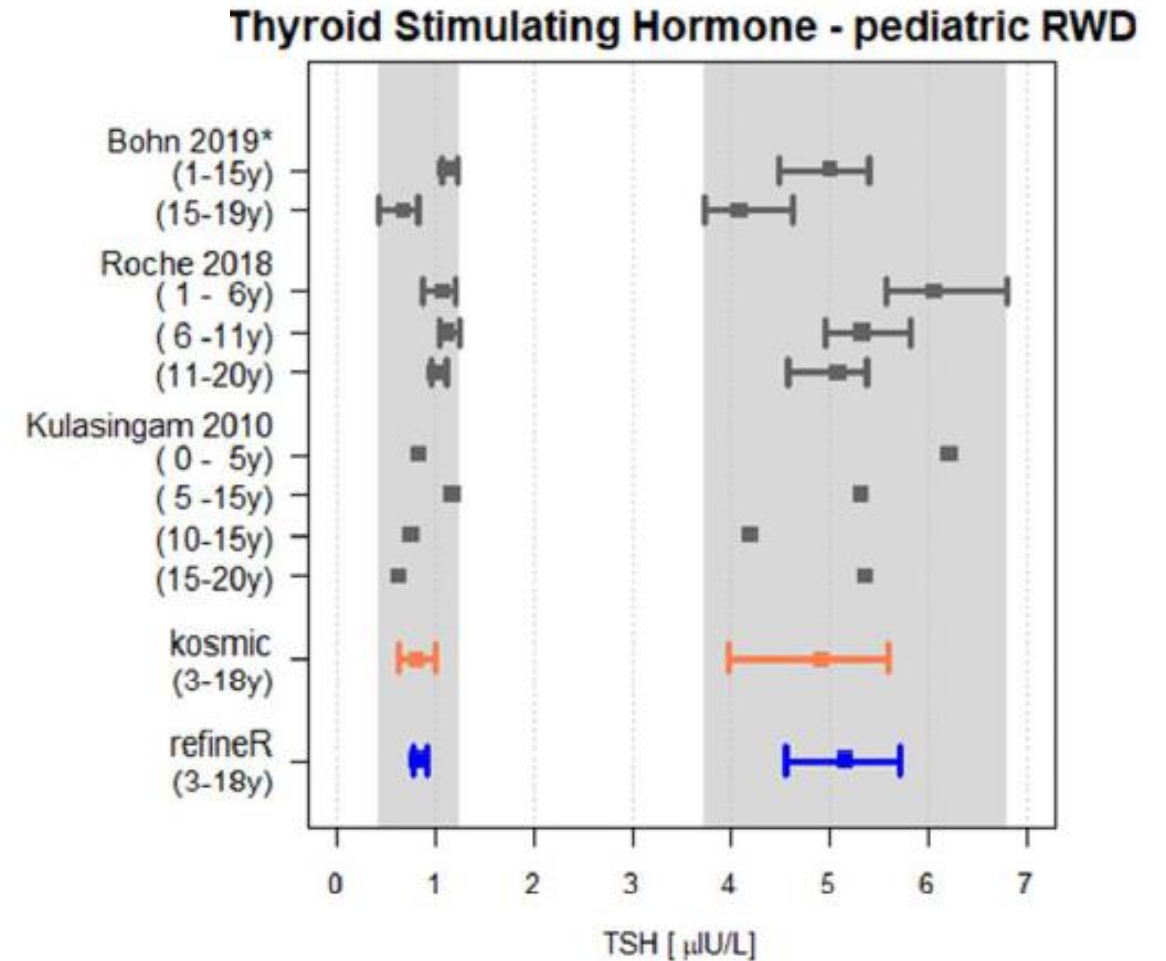
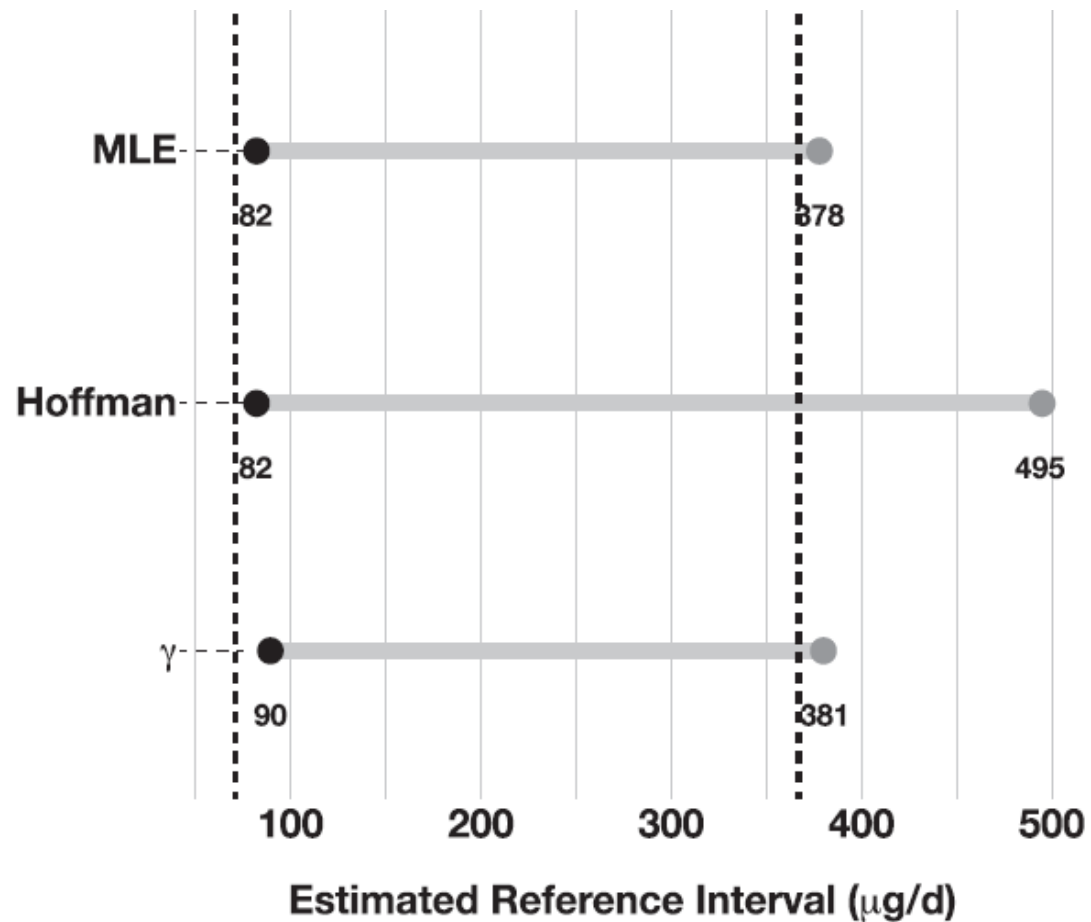
...but is there another way to verify clinical appropriateness?

# Estimated Phenotype RI With Genotype Data



Only the lower estimated RI (CI = 95%) is shown with adult LRI shown as a red dashed line (9.9)

# Compare to Other Methods and Published Studies



# Words of Caution

- Indirect methods do not work for all analytes
  - » Patient demographic/cohort
    - Pathological data
    - Circadian cycles
    - Inpatient/outpatient
    - Collection techniques
    - Tanner stages
- Do not rely too much on too little data

# IFCC Committee Guidance

2018 IFCC C-RIDL recommendations

“encourage the use of indirect methods to establish and verify reference intervals, to promote publication of such intervals with clear explanation of the process used, and also to support the development of improved statistical techniques for these studies.”

» Task Force on Global Reference Interval Database (TF-GRID)

The screenshot shows the IFCC Global Reference Interval Database interface. At the top is the IFCC logo (International Federation of Clinical Chemistry and Laboratory Medicine). Below it is the title "Global Reference Interval Database". A search bar contains the placeholder text "Search for analyte name". Below the search bar are four filter sections: "Filter by age" with a drop-down menu showing "Paediatric (<18)", "Adult (≥18y)", and "Pregnancy"; "Filter by sex" with a drop-down menu showing "Female", "Male", and "All"; "Filter by Country" with a drop-down menu showing "Afghanistan", "Albania", and "Algeria..."; and "Filter by Date" with a drop-down menu showing "Since 2000", "Since 2001", and "Since 2002...". Below the filters are three blue boxes with white text: "200 Studies Included", "40 Analytes Included", and "50 Countries Included". At the bottom, there is a row of 18 "Insert disclaimer" text elements.



## Resources

- CLSI document EP28-A3c. Wayne, PA: Clinical and Laboratory Standards Institute; 2008. Reaffirmed: April 2016
- Jones RD, et al. *Clin Chem Lab Med* 2019;57(1):20-29.
- Ozarda Y. *Biochemia Medica* 2016;26(1):5-16.
- Doyle K, Bunch DR. *Crit Rev Clin Lab Sci.* 2023 Sep;60(6):466-482.
- Holmes DT, Buhr KA. *Am J Clin Pathol.* 2019 Feb 4;151(3):328-336. PMID: 30475946.
- IFCC Committee on Reference Intervals and Decision Limits. *Clin Chem Lab Med.* 2018 Dec 19;57(1):20-29. PMID: 29672266.



*ARUP is a nonprofit enterprise of the University of Utah and its Department of Pathology.*