

# Indirect Estimation of Reference Intervals from Clinical “Big Data”

## Method Requirements for Full Automation

Author: Tobias Ueli BLATTER

Date: July 22, 2023

Version: 1.0.0

Check for Updates [here](#)

## Contents

0.1 Abstract . . . . .	3
0.2 Learning Objectives . . . . .	3
0.3 Contact Information . . . . .	3
0.4 Disclosures . . . . .	3
<b>1 Introduction</b>	<b>4</b>
1.1 Background . . . . .	4
1.2 Advancing Reference Interval Estimation . . . . .	4
1.3 The Clinical Use of Big Data . . . . .	6
<b>2 Case Study: Reference Intervals Estimation</b>	<b>9</b>
2.1 Data Source . . . . .	9
2.2 Data Preparation . . . . .	9

2.3	Data Exploration . . . . .	10
2.4	Data Stratification . . . . .	13
<b>3</b>	<b>Reference Interval Estimation</b>	<b>17</b>
3.1	Traditional Methods . . . . .	17
3.2	Indirect Estimation Approach . . . . .	20
<b>4</b>	<b>Method Requirements for Full Automation</b>	<b>23</b>
4.1	The BioRef Infrastructure: Fully automated RI estimation . . . . .	24
<b>5</b>	<b>Summary</b>	<b>27</b>
<b>6</b>	<b>Copyright</b>	<b>28</b>

## 0.1 Abstract

The evaluation of a patient's laboratory test result by the use of reference intervals (RIs) is an important part of diagnostic medicine. A variety of newer methodologies (indirect methods) have provided new possibilities to establishing RIs directly from laboratory test results collected during routine, a *Big Data* source. These novel methods can provide precise RI estimations, especially in patient groups, such as pediatric or multimorbid populations, where conventional methods are unable to. The implementation of these methods into standardized data analysis pipelines still requires general consensus regarding acceptable veracity of clinical *Big Data*, the relevant patient factors to consider, appropriate stratification procedures.

## 0.2 Learning Objectives

- Name the ethical, legal, and social implications (ELSI) of indirect RI estimation from clinical *Big Data*.
- State the most important hurdles for implementing automated RI estimation methods in clinical data analysis pipelines.

## 0.3 Contact Information

Tobias Ueli Blatter, MSc Bioinformatics  
Department of Clinical Chemistry, University Hospital Bern  
Freiburgstrasse  
CH-3010 Bern  
[tobias.blatter\[at\]extern.insel.ch](mailto:tobias.blatter@extern.insel.ch)  
<http://compmed.ch/>

## 0.4 Disclosures

Nothing to disclose

# 1 Introduction

## 1.1 Background

Reference intervals (RIs) are widely used in various medical fields to aid physicians in identifying potentially pathological states of patient's test results. They refers to a coverage of a specific range (i.e. 95%) of values obtained from a pre-defined population that help establish a baseline for comparison and interpretation of individual test results (Figure 1<sup>1</sup>).

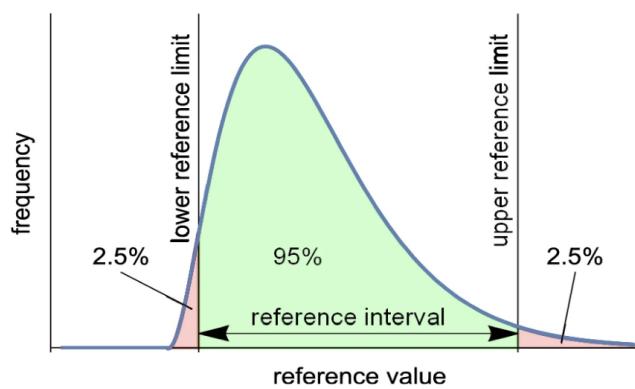


Figure 1: The reference interval, encapsulated by the reference limit, covers a specific range of reference values.

## 1.2 Advancing Reference Interval Estimation

Accredited clinical laboratories need to establish and verify the RIs for the analyses they offer independently and on a regular basis by admissible international guidelines (ISO 15189:2012). The gold standard of inferring RIs has long been **direct methodologies**, where test results are sampled from a healthy reference population<sup>2</sup> (Figure 2, left). To establish said healthy cohort, significant resources

<sup>1</sup>Figure adapted from Davis, C.Q. and Hamilton, R., Reference ranges for clinical electrophysiology of vision. *Doc Ophthalmol* (2021). DOI

<sup>2</sup>CLSI, Guideline EP28-A3c: "Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory: Approved Guideline" (2016). 3rd Edition

are required to assemble healthy patients across both administrative genders and relevant age-ranges. Furthermore, the absence of a comprehensive definition for “health” that encompasses both the normative elements (*well-being* and *functioning*), and the descriptive elements related to health evaluation (*test result assessment*), it is often difficult to establish reference populations across the different patient groups (i.e. pregnant, chronically ill, or older patients) present in a general admission hospital.

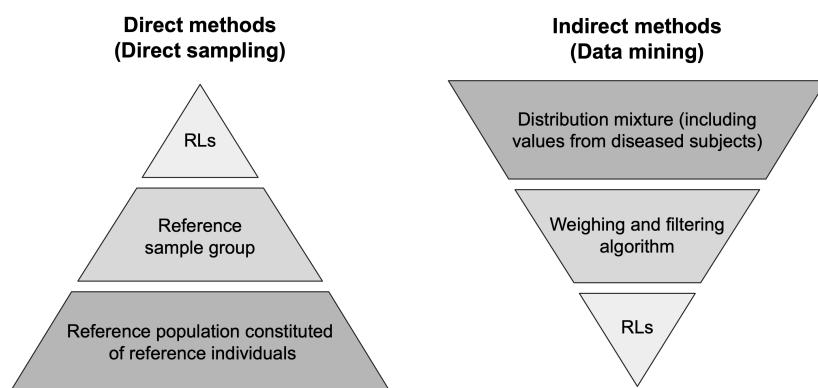


Figure 2: The different approaches that direct and indirect methodologies (methods) take to estimated the reference limits (RLs) from the underlying data.

**Indirect methodologies** of RI estimation offer a way to address the aforementioned shortcomings, as they sample and weight test results directly from a mixed clinical population, which contains both “physiological” (non-pathological) and pathological test results from routine patients (Figure 2, right). As some of these indirect methods have been developed fairly recently, they need to be validated across many different data sets and adapted to be integrated into standard data analysis pipelines. The IFCC Committee for Reference Intervals and Decision Limits (C-RIDL) is driving this effort and guides the ongoing process. This script follows the recommendations of reference interval estimation from clinical routine data laid out by the committee<sup>3</sup>.

---

<sup>3</sup>Jones, G.R.D., C-RIDL, et al., Indirect Methods for Reference Interval Determination - Review and Recommendations, *Clinical Chemistry and Laboratory Medicine* (2018). DOI

### 1.3 The Clinical Use of Big Data

Among medical disciplines, laboratory medicine has consistently embraced a high level of digitization. The data generated during clinical routine, being collected for screening, diagnostic or monitoring purposes, adheres to high-quality specifications and reproducibility standards. This data embodies three important pillars of *Big Data*: **Volume** (Amount of data), **Variety** (Diversity of data) and **Veracity** (Accuracy of data). Laboratory data can be augmented with valuable metadata (i.e. patients' demographic information) from the hospital's IT systems, if this metadata is encoded in adherence to international standards (Figure 3<sup>4</sup>).

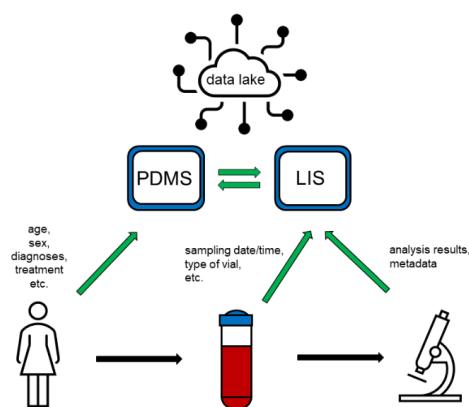


Figure 3: Patients' data is entered into the patient data management system (PDMS), predominantly manually, while information about samples collected as well as about analyses conducted is entered into the laboratory information system (LIS), either manually or automatically. PDMS and LIS should be connected to form “data lakes”, comprising of various types of interlinked data.

Effectively operational clinical data requires clear adherence to international encoding standards, efficient ETL (Extract, Transform, Load) processes, careful data governance, and modern data security solutions<sup>5</sup>. Consistency in data is beneficial for both clinical practice and clinical research. In the age of artificial intelligence (AI) and machine learning, it is important for laboratory medicine as

<sup>4</sup>Blatter, T.U., et al., Big Data in Laboratory Medicine - FAIR Quality for AI?, *Diagnostics* (2022). [DOI](#)

<sup>5</sup>Blatter, T.U., et al., Big Data in Laboratory Medicine - FAIR Quality for AI?, *Diagnostics* (2022). [DOI](#)

a scientific discipline to embrace well-curated and highly-enriched data in order to thrive.

**Encoding** of laboratory (meta-)data ensures consistency and interoperability, enables seamless integration and utilizes the data within a *Big Data* framework. Globally Unique, Persistent and Resolvable Identifiers (GUPRIs) are needed for any variables associated with the measurement collection and testing procedure. A variety of nomenclatures have been established (non-exhaustive list):

1. [LOINC](#) (Logical Observation Identifiers Names and Codes) - A standardized coding system for laboratory and clinical observations, providing unique identifiers for various laboratory tests and measurements.
2. [GUDID](#) (Global Unique Device Identification Database) - A database containing unique identifiers for medical devices.
3. [EUDAMED](#) (European Database on Medical Devices) - A database for the registration and identification of medical devices in the European Union.
4. [GMDN](#) (Global Medical Device Nomenclature) - A nomenclature system used to classify medical devices.
5. [EMDN](#) (European Medical Device Nomenclature) - A nomenclature system specific to Europe for classifying medical devices.

In order to facilitate the accessibility of *Big Data* in laboratory medicine, data must be prepared through **efficient** and **reliable ETL processes**. Here, a few international standards, such as the [FHIR](#) (Fast Healthcare Interoperability Resources) and the Resource Description Framework ([RDF](#)), provide a framework and query language for representing and querying healthcare data in a standardized manner.

**Scalable bioinformatics infrastructures** help to manage observational data from different sources with the help of a common underlying data model and standardized vocabulary for organizing and analyzing electronic health record (EHR) data. Especially noteworthy are the Observational Health Data Sciences and Informatics ([OHDSI](#)) collaborative and the [i2b2](#) (Informatics for Integrating Biology and the Bedside) tranSMART Foundation, which both offer open-source and open-data models designed for organizing healthcare data.

The increasing availability of large volumes of clinical data highlights the importance to address novel ethical concerns related to *Big Data*: In a *Big Data* set, it's detrimental that the individual's privacy is protected and the individual's consent in participation is respected. Not only is careful **data governance** required, but also the hindrance of re-identification of individual patients from high-dimensional data (i.e. by differential privacy<sup>6</sup>) and newer models of consent management (i.e. providing patient-centric consent management<sup>7</sup>) have to be implemented from the beginning of the clinical data management. Furthermore, aggregated data cannot be considered by default anonymized data anymore. Aggregated data has the potential to reveal information about individuals (e.g., membership in a sensitive cohort, undisclosed private/sensitive attributes) through statistical inference even if the data itself does not directly identify specific persons<sup>8</sup>.

Access to laboratory data must be carefully managed and compliance with regulatory requirements and ethical approval is crucial before access is given. Before any use of clinical data for research purposes, the link between the laboratory measurement (with metadata) and the ID of the patient has to be either reversibly removed (de-identification) or fully removed (anonymization). Compliance with (inter-)national data protection laws (see [GDPR and research](#)) is non-negotiable to ensure the protection of individuals' privacy and confidentiality.

---

<sup>6</sup>Ficek, J., et al., Differential Privacy in Health Research: A Scoping Review., *Journal of the American Medical Informatics Association*, (2021), [DOI](#)

<sup>7</sup>Tith, D., et al., Patient Consent Management by a Purpose-Based Consent Model for Electronic Health Record Based on Blockchain Technology, *Healthcare Informatics Research*, (2020), [DOI](#)

<sup>8</sup>Raisaro, J. L., et al., Addressing Beacon Re-Identification Attacks: Quantification and Mitigation of Privacy Risks, *Journal of the American Medical Informatics Association*, (2017), [DOI](#)

## 2 Case Study: Reference Intervals Estimation

### 2.1 Data Source

In this script, we'll be using a modified [HCV](#) data set, provided by the UC Irvine Machine Learning Repository (CC BY 4.0 license). The modified data set contains laboratory test results of 615 potential blood donors. For each patient, the person's age (in years), sex (f,m) and measurements of 6 laboratory analytes are recorded.

### 2.2 Data Preparation

```
# Read the csv
dt <- read.csv(file = "data/hcvdat0.csv", sep = ";")
# Show the head of the data
knitr::kable(head(dt, 2), format = "pipe")
```

Category	Age	Sex	ALT	AST	BIL	CHOL	CREA	GGT
0=Blood Donor	32	m	7.7	22.1	7.5	3.23	106	12.1
0=Blood Donor	32	m	18.0	24.7	3.9	4.80	74	15.6

The following analytes are recorded:

Code	Full name	LOINC Unit
ALT	Alanine Aminotransferase	U/L
AST	Aspartate Aminotransferase	U/L
BIL	Bilirubin	µmol/L
CHOL	Cholesterol	mg/dL
CREA	Creatinine	mg/dL
GGT	Gamma-Glutamyl Transferase	U/L

## 2.3 Data Exploration

The use of the HCV data is beneficial, as for every person (row) it is indicated whether or not the person was deemed eligible for a blood donation and we can therefore assume that they are seemingly “healthy”. In the data there are also some people recorded that were not qualified for blood donation due to an underlying medical condition, which could result in potentially pathological measurement for some analytes.

```
# Shows the amount of patients (female & male) per category  
knitr::kable(table(dt[,c(1,3)]),  
            format = "pipe")
```

	f	m
0=Blood Donor	215	318
0s=suspect Blood Donor	1	6
1=Hepatitis	4	20
2=Fibrosis	8	13
3=Cirrhosis	10	20

The different analytes present the following distributions (see Figure 4). To visually assess if the analytes (grouped by “health status”) follows a normal distribution, a Q-Q-plot is used to comparing the quantiles of the data against the quantiles of a theoretical normal distribution (Figure 5).

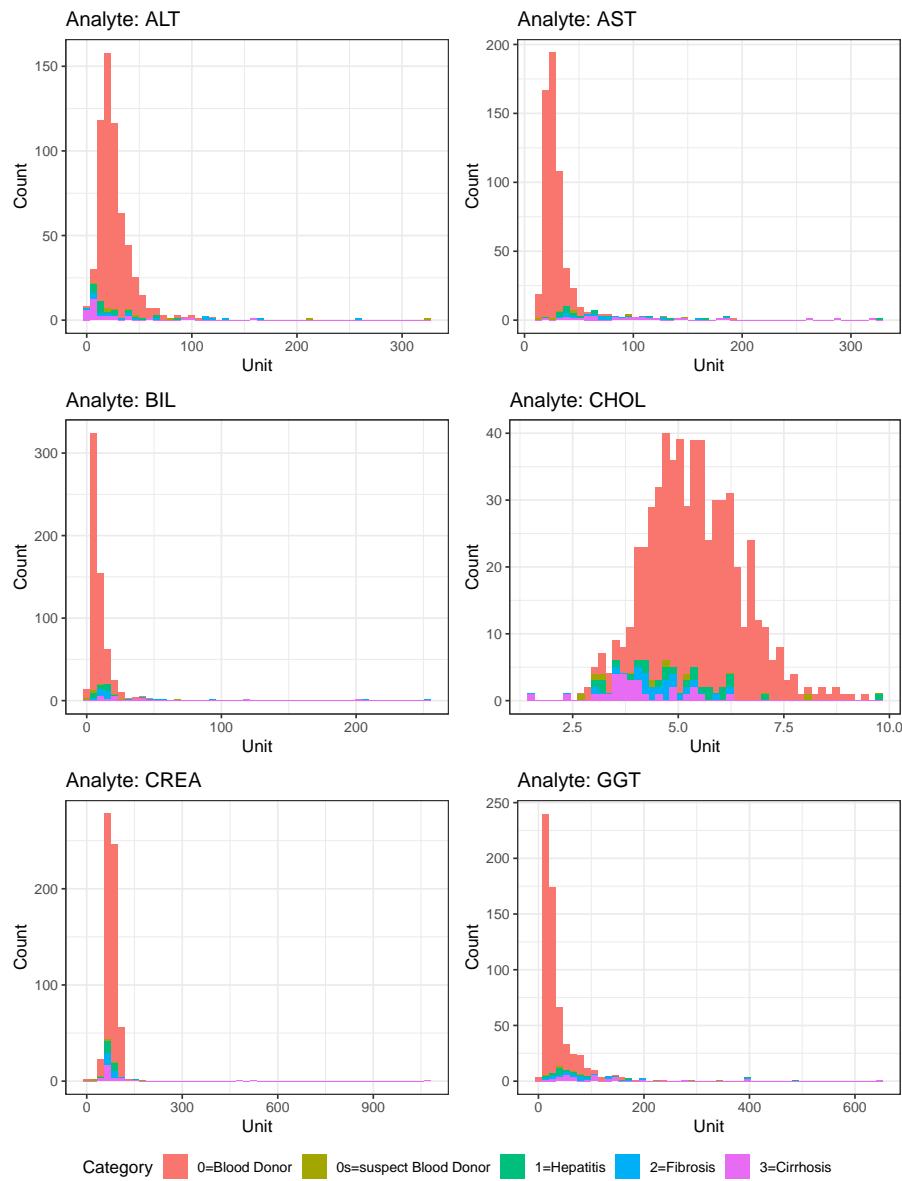


Figure 4: Full histogram of the six analytes from the dataset. There are extreme values present in the reference distribution of some analytes (AST, BIL, CREA).

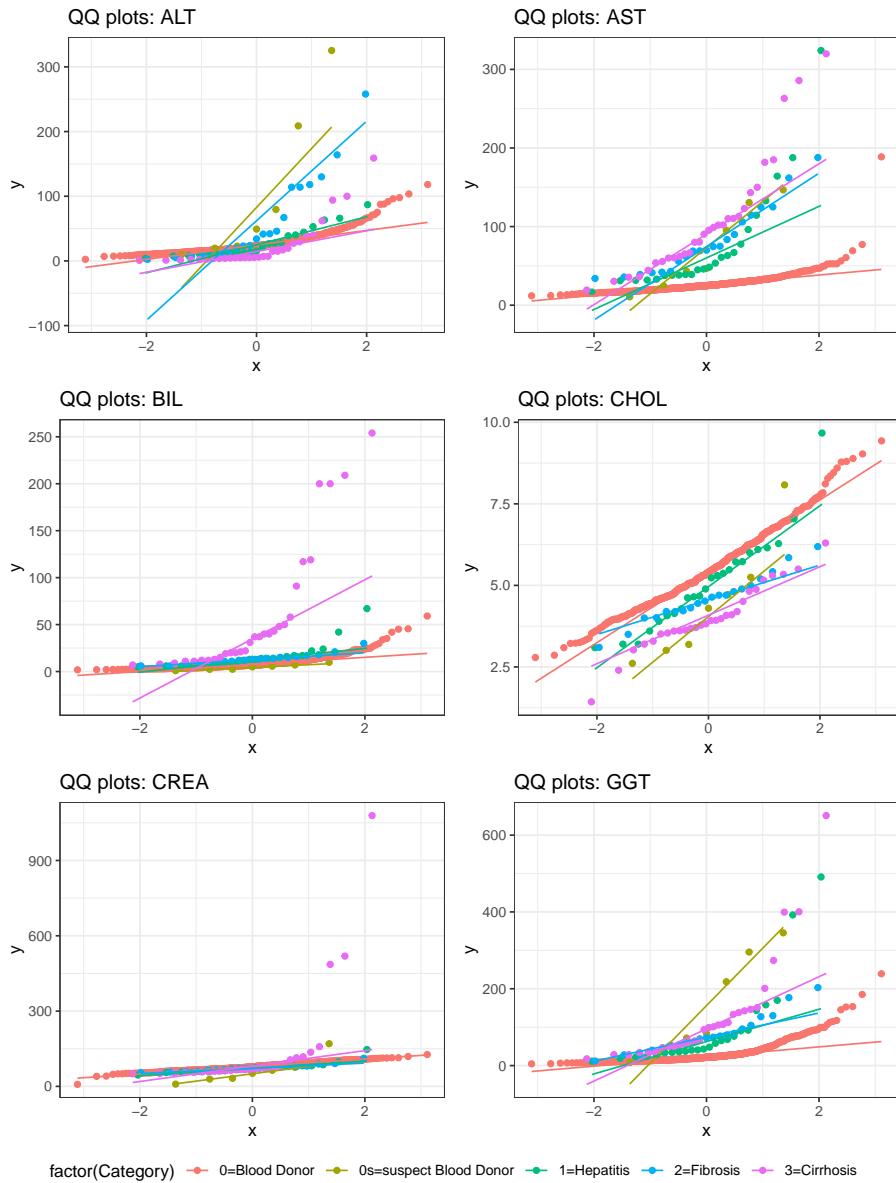


Figure 5: Quantile-Quantile (Q-Q) plots of the six analytes from the dataset. The assumption of a Gaussian distribution for the main mode of the data ("blood donors") can for most analytes be guaranteed. "Outliers" are mostly originating from the diseased populations ("1=Hepatitis", "2=Fibrosis" or "3=Cirrhosis").

## 2.4 Data Stratification

### 2.4.1 By Sex

For the distinct variable “Sex”, we can use a one-way ANOVA to see whether or not the values for each analyte should be stratified. ANOVA (Analysis of Variance) is used to test for significant differences in means among two or more groups. In this case, the groups are defined by the levels of the “Sex” variable (m / f). The table below shows the result of the ANOVA for each analyte:

Analyte	F_value	p_value
CREA	153.245	1.51e-31
ALT	39.592	6.00e-10
GGT	22.367	2.81e-06
BIL	19.431	1.23e-05
AST	14.266	1.74e-04
CHOL	0.308	5.79e-01

The following observations between male and female patients can be made:

1. Significant differences in mean values were observed for CREA, ALT, GGT, BIL, and AST ( $p < 0.001$  for all).
2. No significant difference in mean values was found for CHOL ( $p > 0.05$ ).

The figure on the next page shows the “Sex” stratification (Figure 6). If multiple distinct (non-continuous) variables are considered for stratification the *Ichihara method* can be used. This method utilizes a nested ANOVA (two- or three-level) and separates sources of variations (SD) into three components<sup>9</sup>. The relative magnitude of each SD is expressed as its ratio to between-individual SD. With this, confounding influences of other factors can be handled, allowing judgment on the necessity of partitioning after adjusting for these variables.

---

<sup>9</sup>Ichihara, K. and James, C.B., An Appraisal of Statistical Procedures Used in Derivation of Reference Intervals., *Clinical Chemistry and Laboratory Medicine*, (2010), DOI

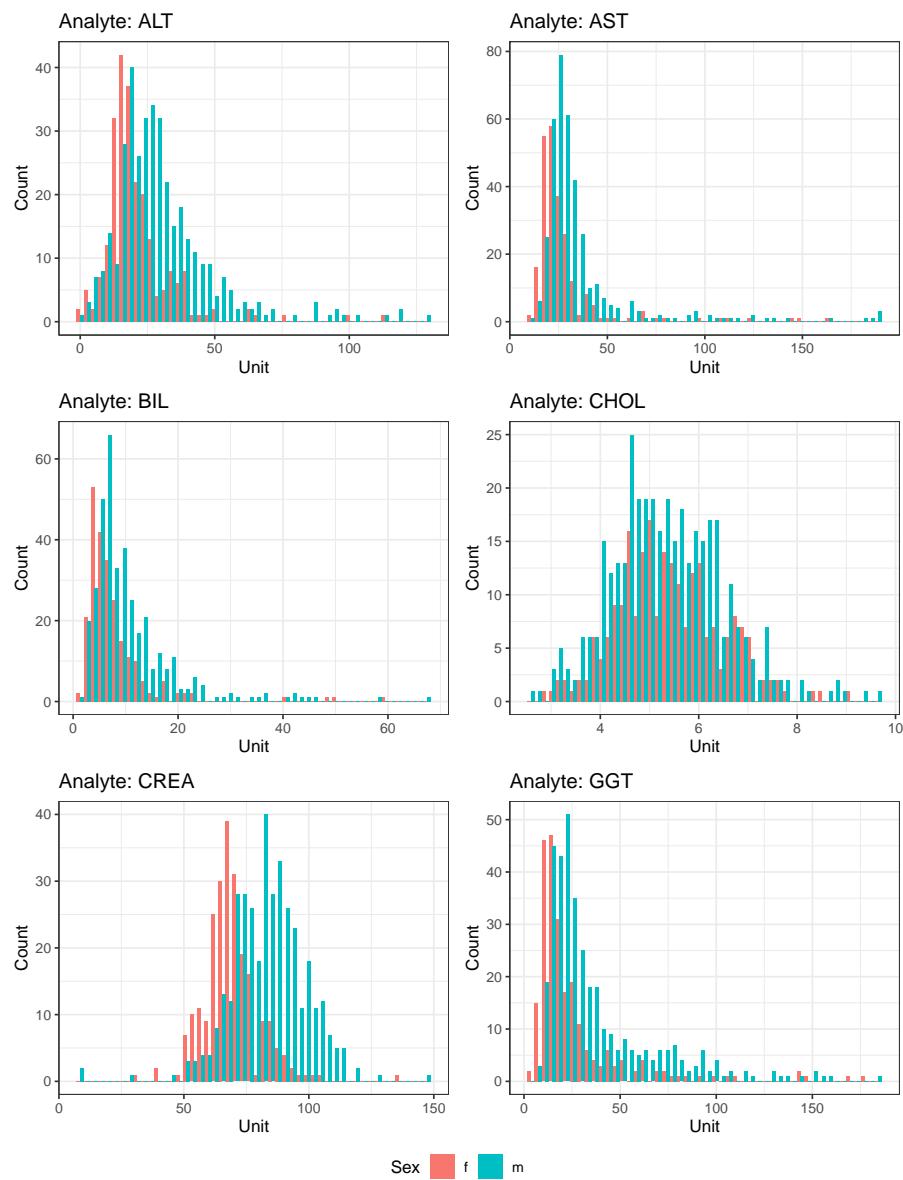


Figure 6: Full histogram of the six analytes from the dataset, stratified by sex (m and f).

## 2.4.2 By Age

For continuous variables such as age, scatter plots can be used to help visualize the relationship between the continuous variable and the test measurement. For analytes showcasing regression between age and measurement, one should consider introducing relevant breakpoints (1 vs. 5 vs. 10 years depending on the available sample size).

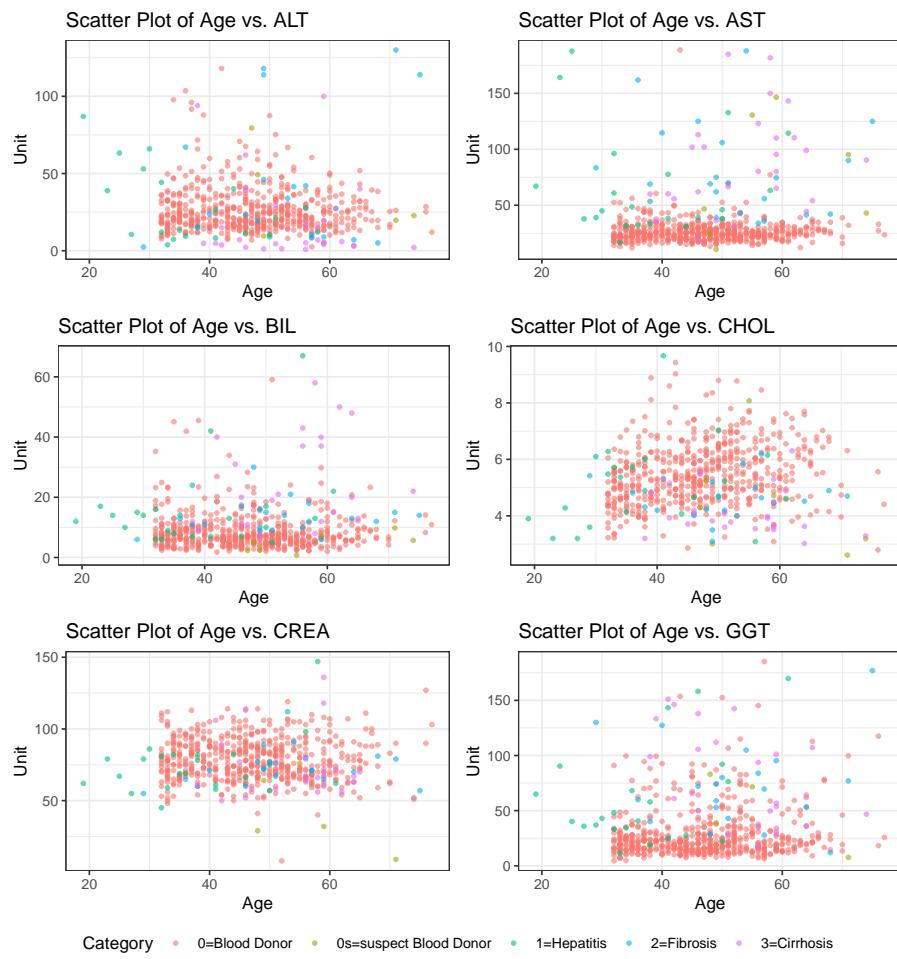
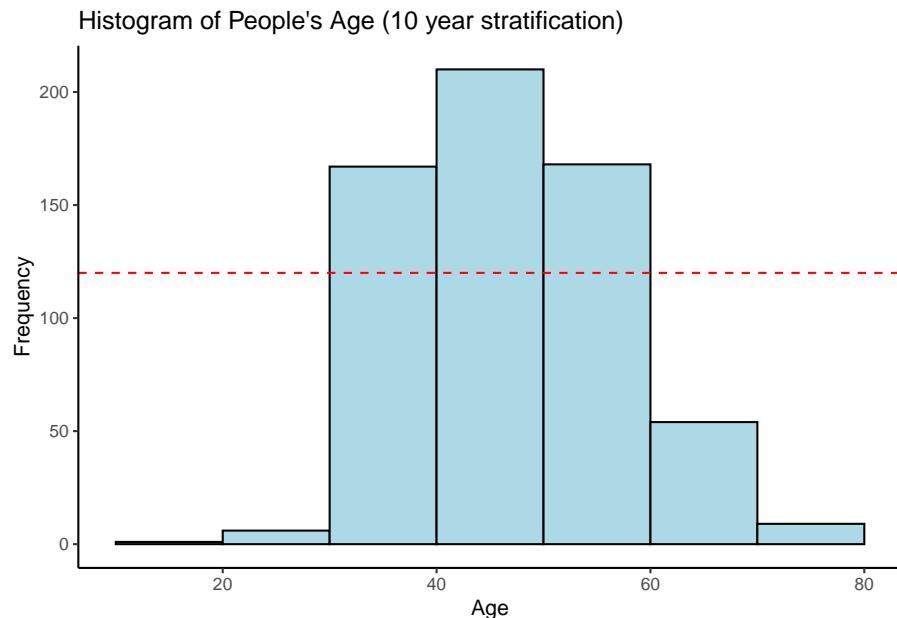


Figure 7: Scatterplots of the six analytes, showcasing the relationship between the patients' age and the measurement

One has to consider that by age stratification, the sample sizes for each stratum might be significantly reduced and fewer reference values are available in the data.



Apart from considering separate partitions, another approach is to create smoothed reference intervals (i.e. *continuous reference intervals*), which can be useful for graphical data analysis. This approach may be particularly beneficial when dealing with data from the pediatric age group (usually very low sample sizes).

## 3 Reference Interval Estimation

### 3.1 Traditional Methods

To calculate RIs by a direct methodologies we use the established *referenceIntervals* package<sup>10</sup>

```
# Calculates the reference and confidence intervals (CIs)
# Methods: "p" (default) for parametric
#           "n" for non-parametric
#           "r" for robust method

ri_direct <- refLimit(data = x,
                       # + e.g. Horn method for outlier removal
                       # --> [Q1 - 1.5 IQR, Q3 + 1.5 IQR]
                       out.method = "horn", out.rm = TRUE,

                       # Parametric method
                       RI = "n", refConf = 0.95,

                       # Bootstrapping CIs
                       CI = "n",
                       limitConf = 0.9)
```

---

<sup>10</sup>Finnegan, D. “referenceIntervals: Reference Intervals.” from CRAN, the Comprehensive R Archive Network, 2022. [CRAN](#)

### 3.1.1 Non-parametric method

For the six analytes (stratified by Sex), this results in the following RIs:

	Male		Female	
Analyte	Lower RL	Higher RL	Lower RL	Higher RL
ALT	3.64	39.63	5.08	57.8
AST	14.47	40.5	16.68	53
BIL	2.26	13.41	2.69	23.13
CHOL	3.32	7.42	3.2	7.43
CREA	52	89.3	58	111.75
GGT	7.14	49.88	10.39	93.82

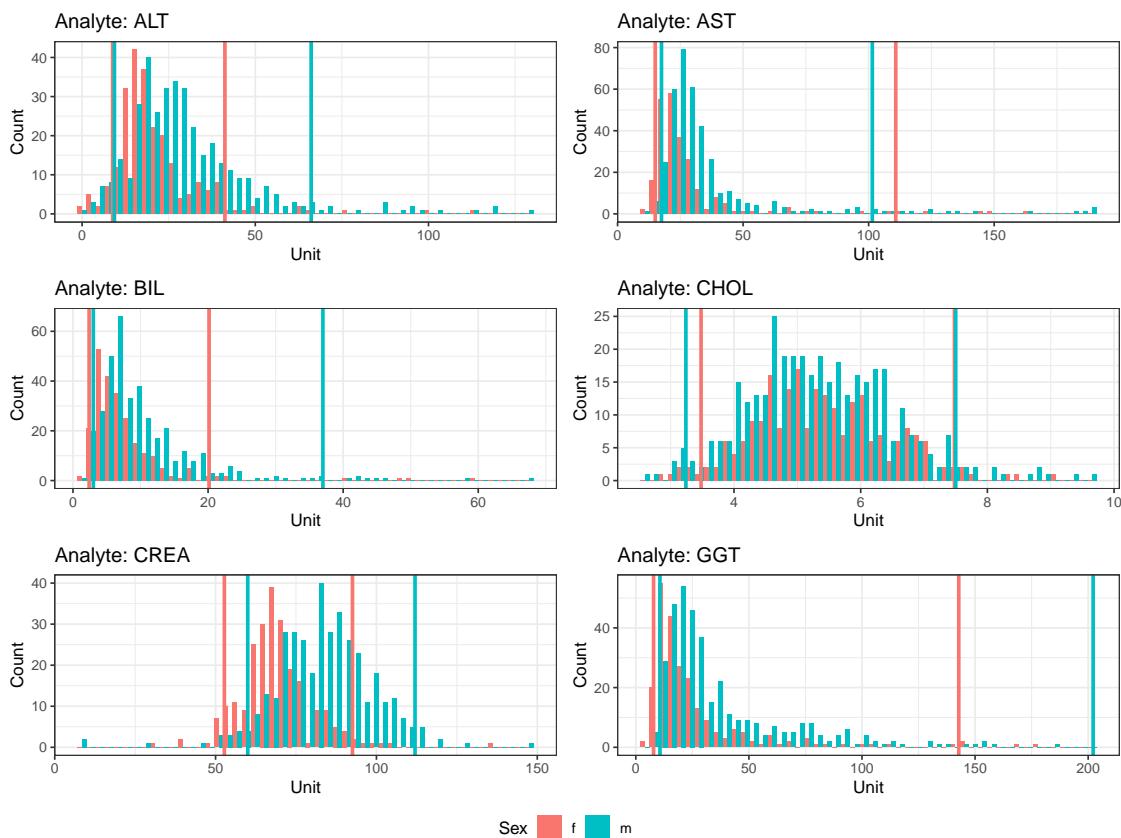


Figure 8: Estimated non-parametric RIs on the global histograms (stratified by Sex)

### 3.1.2 Parametric method

For the six analytes (stratified by Sex), this results in the following RIs:

	Male		Female	
Analyte	Lower RL	Higher RL	Lower RL	Higher RL
ALT	3.13	35.58	2.36	52.26
AST	10.93	36.02	11.95	45.85
BIL	0.77	12.13	0	19.03
CHOL	3.41	7.32	3.2	7.39
CREA	50.47	86.77	57.42	110.24
GGT	0	40.58	0	77.98

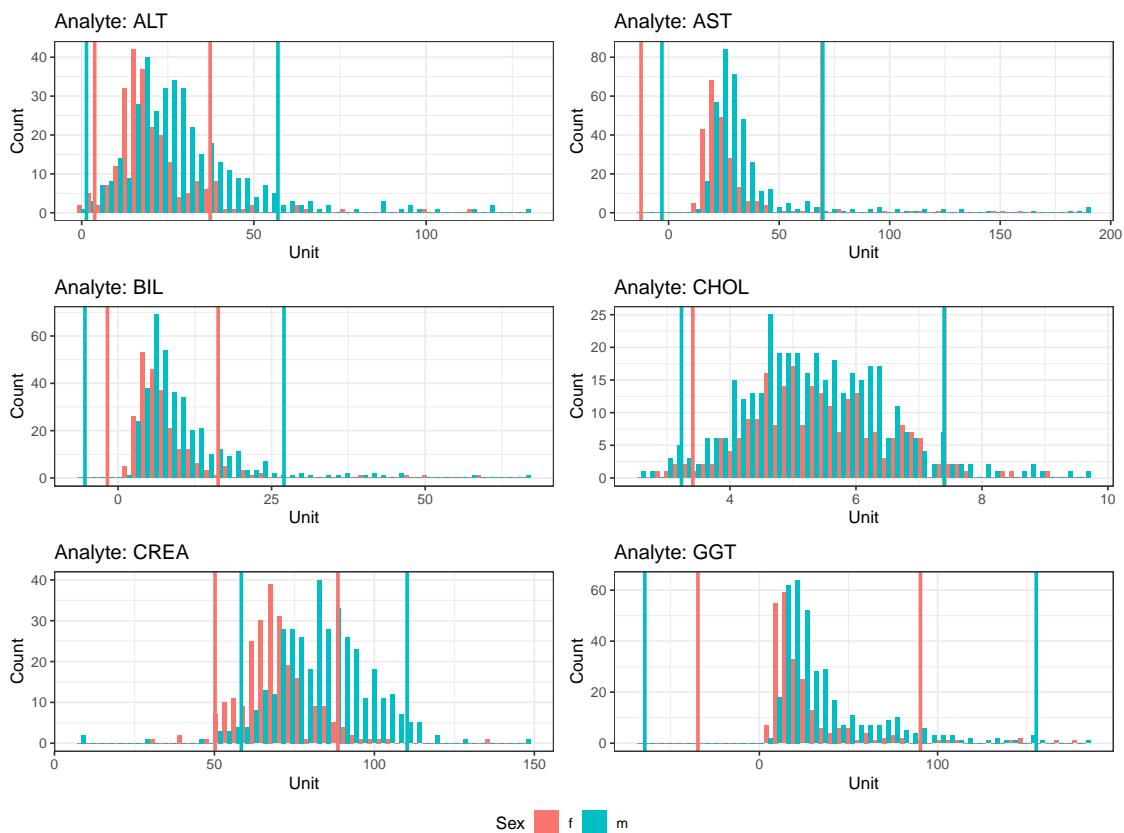


Figure 9: Estimated parametric RIs on the global histograms (stratified by Sex)

## 3.2 Indirect Estimation Approach

For indirect methods, usually 1000 subjects are considered a small sample size and above 10,000 as a large sample size. In populations that are underrepresented in a database (e.g., extremes of age), smaller sample sizes can be considered.

### 3.2.1 Using refineR

The refineR algorithm was recently described and is provided as an open-source R-package (see <https://CRAN.R-project.org/package=refineR>)<sup>11</sup>. The algorithm can estimate RIs from real-world data consisting of a mixed distribution of non-pathological and pathological test results. It is assumed that the majority of test results in the input data set are non-pathological and that their distribution can be described by a Box-Cox transformed normal distribution. Further, it is presumed that a region of test result concentrations exists, where the fraction of pathological test results is negligible. The shape of the distribution of pathological test results can be arbitrary. Two steps are required:

- 1) Define Sex and Age partition (if considering an analyte with sex or age dependency), and ensure population and assay stability by applying and reviewing quality control measures. Arbitrary outlier exclusion or data truncation is not required.
- 2) Apply the refineR algorithm with the default 1-parameter Box-Cox transformation to estimate a model of the non-pathological distribution (for each partition).

---

<sup>11</sup>Ammer, T., et al., refineR: A Novel Algorithm for Reference Interval Estimation from Real-World Data, *Nature Scientific Reports*, (2021), DOI

### 3.2.2 Running refineR

Here, *refineR* finds the RIs from mixed routine data (*x*) using the function `findRI()` (the default function runs with a 1-parameter Box-Cox transformation):

```
# Estimate the model parameters by
fit <- findRI(Data= x )

# Print summary of estimated model
print(fit)

# Or just the estimated RIs
getRI(fit)

# The default plotting function
plot(fit)
```

The following RI can be generated:

Analyte	Male		Female	
	Lower RL	Higher RL	Lower RL	Higher RL
ALT	10.2	65.11	6.08	28.59
AST	14.68	41.52	12.77	38.83
BIL	3.46	14.54	2.18	16.29
CHOL	3.32	7.75	3.71	7.56
CREA	58.95	112.92	51.73	80.69
GGT	9.52	55.23	5.91	47.13

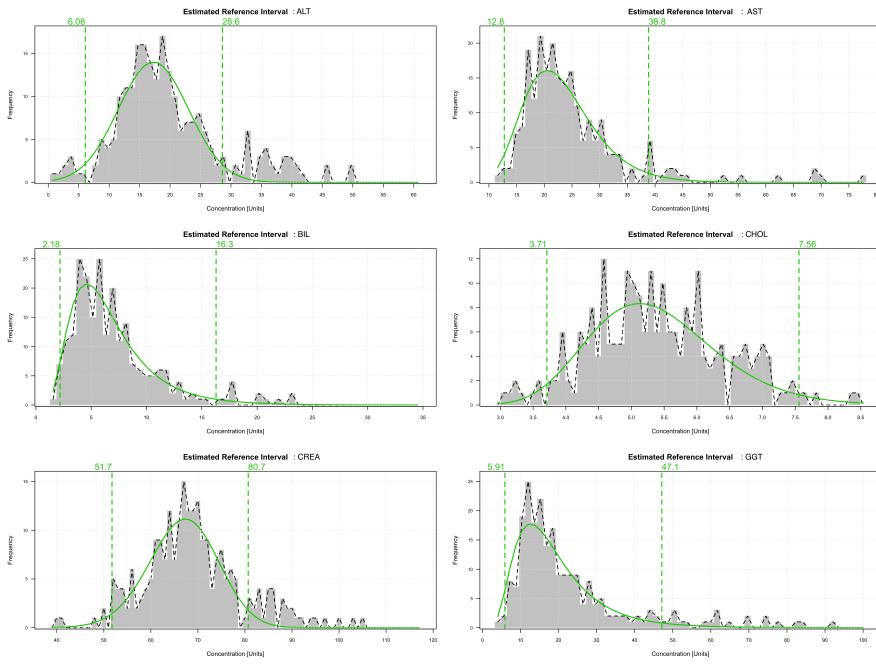


Figure 10: Estimated RIs from refine R for the female population

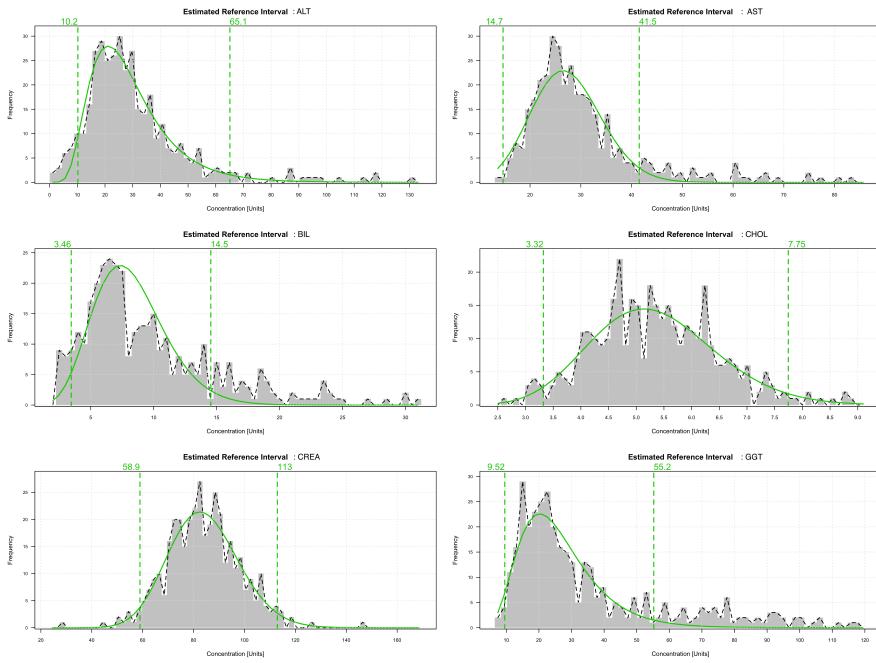


Figure 11: Estimated RIs from refine R for the male population

## **4 Method Requirements for Full Automation**

### **Scalability**

- The method should be scalable to handle large and diverse data sets from different clinical laboratories or institutions without compromising performance: Consider the computational requirements, memory usage, and storage capacity needed to handle increasing data volumes.

### **Efficiency**

- The method should be computationally efficient to analyze the data in a reasonable amount of time: Optimize data processing steps to reduce redundancy and increase efficiency and avoid time-consuming algorithms that hinder quick turnaround.

### **Accuracy**

- Validation of the method can be done against an appropriate gold standard direct methods. The method should provide accurate and reliable estimates of reference intervals: Account for potential sources of bias or confounding in the data and take steps to minimize their impact on the results.

### **Reproducibility**

- The method should yield consistent results when applied to different data sets or subsets of the same data set (also at different times): Provide clear documentation and code for the methodology used to enable reproducibility by others.

### **Interpretability**

- The method should be transparent and interpretable. This can be achieved by providing clear explanations of the algorithm and parameters used in RI estimation: Consider ways to visualize and communicate the estimated reference intervals effectively to healthcare professionals and researchers.

## 4.1 The BioRef Infrastructure: Fully automated RI estimation

**BioRef** is a national infrastructure for generating precise reference intervals for diagnostic medicine

With the BioRef project, we have developed a multi-center computational framework, where specialized web applications estimate and assess patient group-specific reference intervals based on clinical routine data from four Swiss Hospitals. We have established a common legal governance and interoperability framework for our clinical partners to share their data either to a central database via a national and secure data sharing network or providing their data in a decentralized way via “TI4Health”, a secure and encrypted data-accessing system, allowing each data provider to abide to the restrictions laid out by their ethics waivers.

*Figures from the publication, forthcoming in JMIR<sup>12</sup>:*

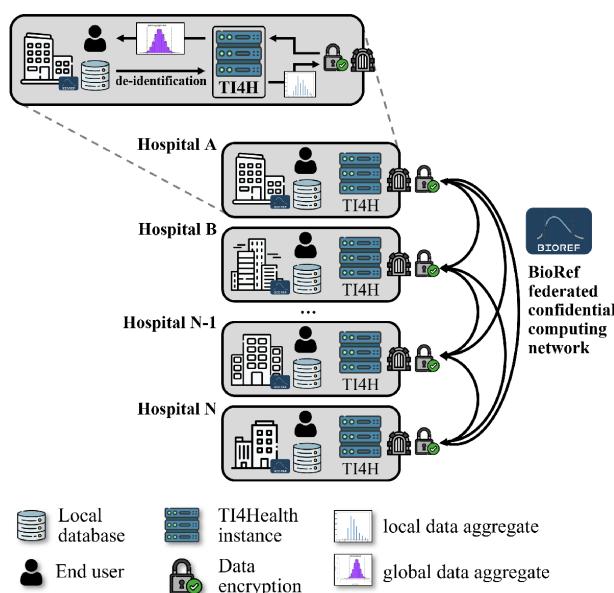


Figure 12: Illustration of the BioRef federated analytics infrastructure. In the decentralized approach, data is de-identified on site by the individual data providers of the consortium (hospital A, hospital B, ...) and uploaded to the on-premise TI4Health instance. Data is analyzed via the federated confidential computing network without any raw data of the consortium members being revealed.

<sup>12</sup>J Med Internet Res. 2023 Jul 14. DOI: <https://doi.org/10.2196/47254>. [Epub ahead of print]

The deployed web applications, which allow intuitive and interactive data stratification by patient factors (such as age, administrative sex and personal medical history) and laboratory analysis features (such as device, analyzer and test kit identifier) are accessible for registered physicians and researchers. As we are evaluating our deployed framework, we are currently establishing the on-boarding of future national and international partners, refining the statistical analysis for multi-cohort patient queries and adjusting the web-interfaces to build clinically viable diagnostic tools.

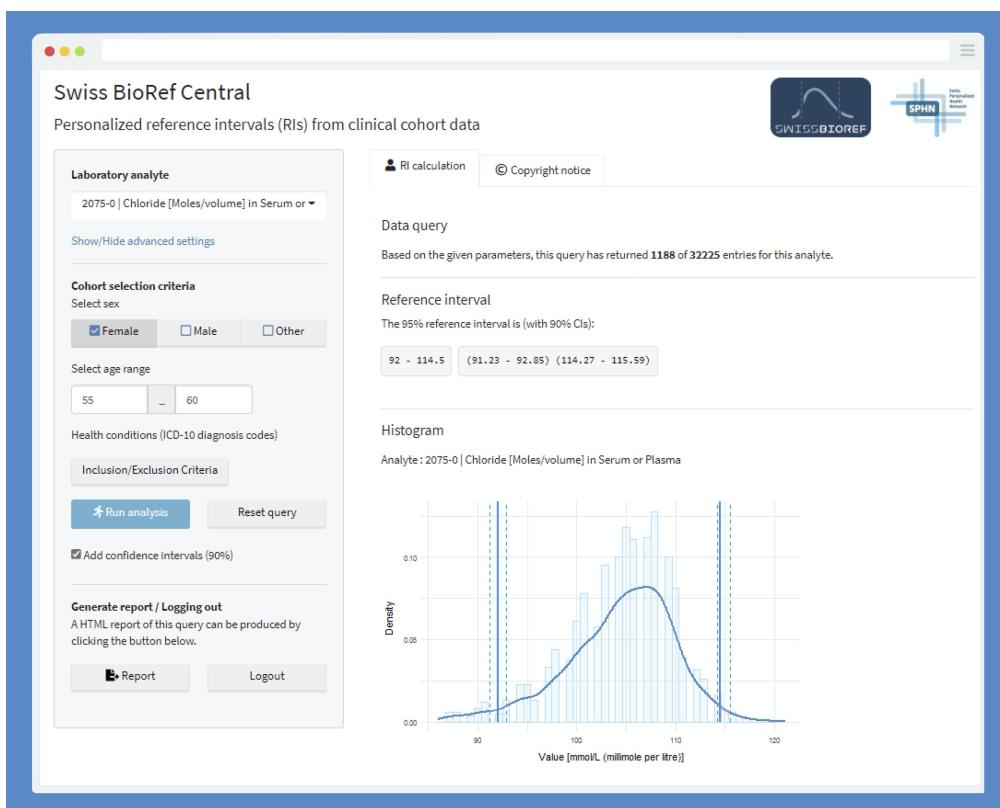


Figure 13: Graphical user interface of the Swiss BioRef Central web application. The web applications show the estimates for reference intervals for “Chloride in Serum or Plasma” (LOINC: 2075-0) for a female patient cohort with age 55-60 years as an exemplary query.

Establishing an opportunity for clinical physicians and researchers to define precise reference intervals in a convenient and reproducible way on-the-fly is a vital part of practicing precision medicine today.

We further suggest that additional patient parameters (in addition to age and administrative gender) such as the specific combinations of diagnoses should be considered while analyzing locally derived reference intervals.

Especially for older patients, distinguishing between “disease” and natural aging processes can be challenging. The functional decline observed in old age can be attributed to either a specific disease or simply the aging process itself. Age-related health concerns become more significant in aging populations, and defining an appropriate reference becomes crucial. These reference intervals should encompass both physiological changes that occur with age and an increasing proportion of values that might be considered abnormal in a younger population but are common in the aging patient population. Rather than attempting to establish reference intervals solely as “normal ranges” for aging populations, a concept of “expectation ranges” is proposed. These expectation ranges aid in evaluating a specific patient’s test results within the context of similar patients, often referred to as “digital twins.” By incorporating specific diagnoses, it is possible to adjust and fine-tune these expected ranges to cater to various multimorbid conditions (e.g., diabetes, hyperlipidemia, coronary heart disease, or renal impairment).

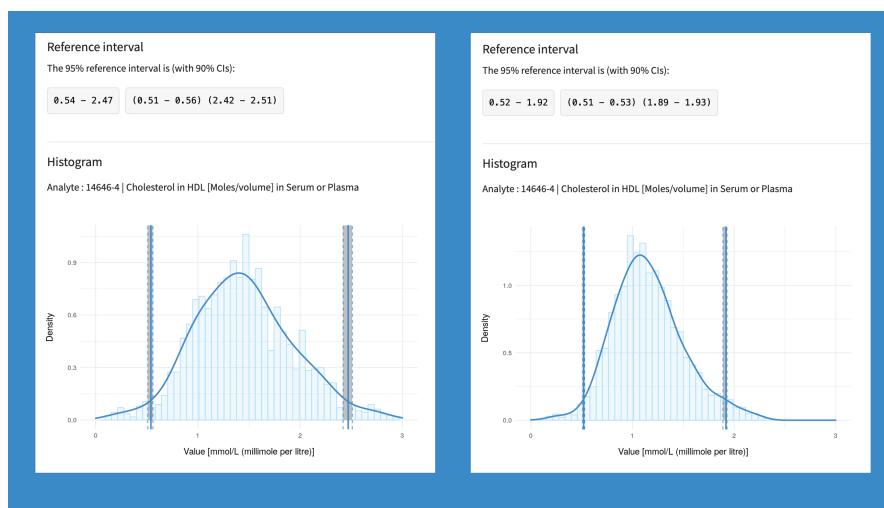


Figure 14: Estimated reference intervals for «Cholesterol in HDL [Moles/volume] in Serum or Plasma» (LOINC 14646-4) for female patients ( $n = 1'848$ , left) and male patients ( $n = 5'026$ , right), 60-65 years old

## 5 Summary

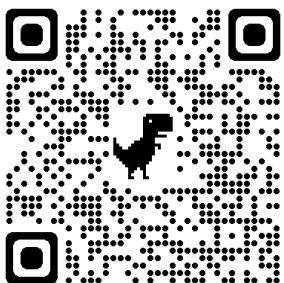
This document discusses the importance of reference intervals (RIs) in diagnostic medicine and showcases a newer methodology for establishing RIs directly from laboratory test results collected during routine with this script, serving as an interactive document. The case study presented showcases a reference intervals estimation approach using a modified HCV data set. Data preparation, exploration, and stratification by sex and age are demonstrated. Traditional and indirect methods for RI estimation are compared, with the refineR algorithm presented as an open-source R-package for indirect RI estimation from real-world data. Indirect methodologies are highlighted as they offer advantages in patient groups where conventional methods are limited, such as pediatric or multimorbid populations. There is a clear need general consensus on the use of clinical Big Data in standardized data analysis pipelines, considering ethical, legal, and social implications (ELSI) associated with indirect RI estimation. Furthermore, there are five method requirements for fully automated RI estimation, including scalability, efficiency, accuracy, reproducibility, external validation, and interpretability.

Overall, with this script, I hoped to provide an insights into the advancements and challenges in reference interval estimation and the integration of clinical Big Data in diagnostic medicine.

## 6 Copyright

Copyright © Tobias Ueli Blatter, Department of Clinical Chemistry, University Hospital Bern, 2023. All rights reserved.

Check for Updates [here](#) or use the QR code below:



**License:** This script is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0) license. This means that you are free to use, share, and adapt the script for non-commercial purposes, provided proper attribution is given and the script is not used for publication without explicit written permission by the author.

**Terms of Use:** By accessing and using this script, you agree to abide by the following terms: You are not allowed to publish or distribute the content of this script, in whole or in part, without prior written consent. This includes any form of publication, including journal articles, conference proceedings, or books. However, you are permitted to use and adapt the script for personal, educational, or non-commercial purposes within the scope of the Creative Commons license stated above.

**Citation Requirement:** When using or referencing this script, proper attribution must be given. Please include the author's name, title of the script, date of creation, and a reference to the original [source](#).