UAB
Universitat Autònoma de Barcelona

COMPARISON OF MULTIPLE (CLASSIC AND NOVEL) INDIRECT METHODS TO ESTABLISH REFERENCE INTERVALS

Luisa María Martínez Sánchez

COMPARISON OF MULTIPLE (CLASSIC AND NOVEL) INDIRECT METHODS TO ESTABLISH REFERENCE INTERVALS

Doctoral Thesis submitted by

Luisa María Martínez Sánchez

For the degree of Doctor

Directors

Dr. Christa Cobbaert

Dr. Francisco Rodríguez Frías

Dr. Wendy den Elzen

Tutor

Dr. Francisco Rodríguez Frías

Doctoral Thesis Program in Biochemistry, Molecular Biology and Biomedicine

Department of Biochemistry and Molecular Biology

Barcelona, 2023

Physiology describes our life's journey and is only when we are familiar with that journey that we can appreciate a pathological departure

Sikaris KA.

Physiology and its importance for reference intervals

Acknowledgements

The development of this doctoral thesis could not have been done without the support, encouragement and help from a lot of people.

I would like to thank my thesis directors for the opportunity of becoming part of such a great team. To Christa for giving me the chance to visit her laboratory in the LUMC and for being an inspirational leader. To Paco for his trust right from the beginning of this journey, for encouraging my interest in Data Science, and for being a great mediator when needed. To Wendy for being my guide and support all these years, for getting the best out of me, and being always available to teach me new things. I have learnt from you not only research skills but life skills, which have shaped me as the person I am now.

A todas mis compañeras del laboratorio que han estado siempre ahí, tanto en mis años de residencia como durante la tesis. Especialmente a aquellas que han escuchado mis problemas unos días y celebrado mis triunfos otros. Al grupo de Bhattacharya, especialmente a Yoli, Pablo y Albert por tantas horas de “brainstorming” sin saber hacia dónde nos llevaría esto.

A mis compañeras y compañeros de vida durante estos años por estar ahí para mí.

Abbreviations

ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
BMI	Body Mass Index
BV	Biological Variation
CK	Creatine Kinase
CLSI	Clinical Laboratory Standards Institute
C-RIDL	Committee on Reference Intervals and Decision Limits
EQA	External Quality Assessment
GDPR	General Data Protection Regulation
GGT	Gamma-glutamyl Transferase
IFCC	International Federation of Clinical Chemistry
ISO	International Organization for Standardization
IVD	In Vitro Diagnostic
IVDR	In Vitro Diagnostic Medical Devices Regulation
LDH	Lactate Dehydrogenase
LIS	Laboratory Information System
MeSH	Medical Subject Headings
NUMBER	Nederlandse UniforMe Beslisgrenzen En Referentie-intervallen
RLE	Reference Limit Estimator
RMS	Reference Measurement Systems
SKML	Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek
TMC	Truncated minimum chi-squared
TML	Truncated maximum likelihood

Table of content

Summary

Reference intervals are essential decision-making tools for results evaluation in laboratory reports. The calculation of reference intervals is traditionally a laborious and time-consuming process requiring significant resources and consisting of recruiting at least 120 individuals for calculating the interval where 95% of them lies (direct methods). In the past years the rapid evolution of clinical laboratories due to its automatization has allowed a big amount of clinical laboratory data to be available. This, together with data science strategies that have also evolved, has allowed alternative approaches for reference intervals calculation to arise (indirect methods). Those approaches use the already available laboratory data to calculate the reference intervals by means of statistical methods. The aim of this thesis is twofold. On the one hand, to explore the differences in the reference interval results obtained by three indirect methods (NUMBER Dutch method, Reference Limit Estimator German method and Bhattacharya traditional method) using a dataset from Vall d'hebron clinical laboratories. On the other hand, to provide easily accessible tools and descriptions that enable laboratory specialists to calculate reference intervals for their own laboratory using indirect methods.

In the first study presented within this thesis, reference intervals were calculated for 16 biochemistry tests using the Dutch indirect method NUMBER. This method considers biochemically related tests for outlier elimination and dataset cleaning and then calculate the reference intervals using the mean and two times the standard deviation. Then, obtained reference intervals were compared with the original NUMBER results obtained in the Dutch population. In addition, results were also compared with reference intervals from the Reference Limit Estimator method using the same dataset from Vall d'Hebron. For tests following a normal distribution, similar reference intervals were found between Vall d'Hebron and the Dutch study. The upper limits of Gamma-glutamyl transferase were markedly higher in the Dutch study compared to Vall d'Hebron results which suggest a lifestyle component. Creatine kinase and uric

acid reference intervals were higher in both populations compared to conventional reference intervals.

In the second study presented in this thesis, 8 biochemistry laboratory tests were analysed for reference intervals calculation by the Bhattacharya method using the Excel Spreadsheet created by St Vincent's hospital. Bhattacharya is a graphical method for identifying a Gaussian distribution (reference population) in the midst of a complete dataset. This method is known to require the subjective input of the user, which results in important between user differences in calculated reference intervals. An important reduction of between users' variability when using the tool was found for most tests after applying the criteria defined as part of the study.

In summary, we found that medical test results following a normal distribution result in comparable and consistent reference intervals between indirect methods. Therefore, a simple indirect method including data cleaning and the calculation of percentiles (or means and standard deviations) is a feasible and cost-efficient approach for calculating reference intervals. For other tests, more statistically complex methods are necessary. Yet, for generating standardized calculated reference intervals that are traceable to higher order materials and methods, efforts should also focus on test standardization and bias assessment using commutable trueness verifiers. Future efforts should focus on creating appropriate and easy to use tools for indirect reference intervals calculation; on performing clinical validation studies, together with clinicians, in which clinical information and patient follow up is available and on investigating the added value of personalized reference intervals.

Resumen

Los intervalos de referencia son herramientas esenciales para la evaluación de los resultados en los informes de laboratorio. El cálculo de intervalos de referencia es tradicionalmente un proceso laborioso y que requiere de abundantes recursos; consiste en reclutar al menos 120 individuos y encontrar el intervalo en el que se encuentra el 95% de sus resultados (métodos directos). En los últimos años, la rápida evolución de los laboratorios clínicos debido a su automatización ha permitido disponer de una gran cantidad de datos. Esto, junto con las estrategias de ciencia de datos que también han evolucionado, ha dado lugar al estudio de aproximaciones alternativas que utilizan los datos de laboratorio ya disponibles para calcular los intervalos de referencia mediante métodos estadísticos (métodos indirectos). El objetivo de esta tesis es, por un lado, explorar las diferencias en los resultados de intervalos de referencia obtenidos por tres métodos indirectos (método holandés *NUMBER*, método alemán *Reference Limit Estimator* y método tradicional *Bhattacharya*) utilizando un conjunto de datos de los laboratorios clínicos de Vall d'Hebron. Por otro lado, proporcionar herramientas y descripciones de fácil acceso que permitan a los especialistas de laboratorio calcular intervalos de referencia propios utilizando métodos indirectos.

En el primer estudio presentado en esta tesis, se calcularon intervalos de referencia para 16 pruebas bioquímicas utilizando el método indirecto *NUMBER*. Este método considera pruebas relacionadas clínicamente para la eliminación de valores atípicos y luego calcula los intervalos de referencia como la media y dos veces la desviación estándar. Posteriormente, se compararon los intervalos de referencia obtenidos en el estudio previo de *NUMBER* para la población holandesa, con los resultados obtenidos para la población de los laboratorios Vall d'Hebron. Además, los resultados del estudio también se compararon con los intervalos de referencia calculados con el método *Reference Limit Estimator,* utilizando el mismo conjunto de datos de Vall d'Hebron. Para las pruebas que siguen una distribución normal, se encontraron intervalos de referencia similares entre Vall d'Hebron con ambos métodos y el estudio holandés. Los

límites superiores de la gamma-glutamil transferasa fueron marcadamente más altos en el estudio con población holandesa, lo cual sugiere una posible influencia del estilo de vida. Los intervalos de referencia de la creatina quinasa y el ácido úrico fueron más altos en ambas poblaciones en comparación con los intervalos de referencia convencionales.

En el segundo estudio presentado en esta tesis, se analizaron 8 pruebas bioquímicas para el cálculo de intervalos de referencia por el método de *Bhattacharya* utilizando la hoja de cálculo de Excel creada por el hospital *St Vincent*. *Bhattacharya* es un método gráfico para identificar una distribución Gaussiana (población de referencia) en medio de un conjunto de datos. Se conoce que este método requiere la entrada de parámetros de forma subjetiva por parte del usuario, lo que resulta en importantes diferencias entre usuarios en los intervalos de referencia calculados. Después de aplicar una serie de criterios definidos como parte del estudio, se encontró una importante reducción de la variabilidad entre usuarios al usar la herramienta.

En resumen, encontramos que los resultados de pruebas médicas que siguen una distribución normal dan lugar a intervalos de referencia comparables y consistentes entre métodos indirectos. Por lo tanto, un método indirecto simple que incluya limpieza de datos y cálculo de percentiles o medias y desviaciones estándar, es un enfoque factible y eficiente en términos de costes para calcular intervalos de referencia mientras que, para otras pruebas, se necesitan métodos estadísticamente más complejos. Además, para la generación de valores de referencia estandarizados que sean trazables a materiales y métodos de orden superior, muchos esfuerzos deberían estar enfocados en la armonización y normalización de las pruebas de laboratorio y en la evaluación de posibles sesgos analíticos mediante controles de calidad conmutables. En un futuro próximo las líneas de investigación deberían orientarse en desarrollar herramientas adecuadas y fáciles de usar para el cálculo indirecto de intervalos de referencia; en la realización de estudios de validación clínica en los que se disponga de información alternativa a los datos de laboratorio para realizar el seguimiento de pacientes y en la investigación del valor añadido de los intervalos de referencia personalizados.

1. INTRODUCTION

1. Introduction

One of the most important roles of specialists in clinical chemistry and laboratory medicine is to design and generate informative laboratory reports to help clinicians in the interpretation of medical test results. Laboratory reports often provide test result and its corresponding reference intervals which are commonly used as a decision-making tool (1).

Reference intervals are delimited by its upper and lower reference limit and are generally defined as the results from the central 95% of a population free from disease (reference population) (2,3). Therefore, by definition, 5% of the results from non-diseased people will fall outside the reference interval. There are certain situations in which the description above does not apply, such as when reference intervals are based on the 99th percentile or when decision limits are used. These exceptions will not be addressed in this thesis (4). The quality of the reference intervals plays an equally important role in result interpretation as the quality of the result itself (5). For laboratory specialists, it is therefore important to know the concept of reference intervals, how to obtain reliable reference intervals, and how these strategies evolved in the past years. A central question is how, as experts, we can improve this tool and allow easier and accurate interpretation of the test results.

The concept 'Reference interval' was used for the first time in 1969 by Saris and Grasbeck in contrast with the hazy concept 'normal' that was used until then (6). This first definition was developed by a specific expert panel created with this purpose. Nevertheless, the first official recommendations about the theory and production of reference values were published by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) nine years later, in 1978 (7). After that, other national scientific societies published their own recommendations based on the international one [French (SFBC) (8), Spanish (SEQCML) (9), Scandinavian societies (10), etc]. From 1987 to 1991, six papers were published by the IFCC with recommendations and procedures to produce reference intervals (11–16).

The following years after these recommendations two initiatives arose that increased the interest in reference interval calculations. On the one hand, the European Directive 98/79 on In Vitro Diagnostic (IVD) for the first time obliged medical devices manufacturers to include appropriate reference intervals into their product inserts (17). On the other hand, the International Organization for Standardization 15.189 standard for clinical laboratory accreditation states the need for each laboratory to review periodically their own reference intervals (18). The Committee on Reference Intervals and Decision Limits (C-RIDL) was established under the umbrella of IFCC in 2005 and a final guideline for the Defining, Establishing and Verifying reference intervals based on the original recommendations was published in 2010 by the Clinical Laboratory Standards Institute (CLSI) (2,3). This document has been widely used and followed to produce reference intervals by what we know now as the "direct method".

Even with all the improvements gained in the last decades, implementation of the theory of reference values to the clinical practice is not straightforward. Laboratories often do not have the resources to calculate reference intervals specific for their method and population following the guidelines (2). The alternative of using reference intervals recommended by manufacturers or any other source is often preferred. Even though, some guidelines exist for reference intervals transferability, application of them is not optimized in practice. Several countries and the C-RIDL have been working during the last decade on standardization and harmonization efforts (19–25) but appropriate approaches that overcome the drawbacks from the current definition of reference intervals are not yet available.

1.1. Direct vs Indirect methods for reference intervals calculation

According to the CLSI recommendation published in 2010 (2,3,26) reference intervals should be calculated using the direct method by selecting a minimum of 120 healthy individuals per partition to be able to calculate 90% confidence intervals (27). Those should be systematically selected after knowing the characteristics of the healthy reference population. After selecting them and verifying they are in good health by means of questionnaires, medical and/or physical tests; phlebotomy is performed, usually at the laboratory site. Then, the samples are analysed and, after all test results are available, reference intervals are calculated using statistical analyses (calculating directly the 95% range either as mean ± 2 times the standard deviation or as the 2.5 and 97.5 percentiles in the non-parametric approaches). The advantages and disadvantages of this method are presented in **Box1**.

Box1. Advantages and disadvantages of direct reference intervals calculation:

Advantages:

1. the reference group is well-characterized and controlled
2. simple statistical methods can be performed to calculate the direct reference intervals
3. the definition of reference values and the protocol are standardized.

Disadvantages:

1. selection bias may occur, due to the complexity to select, contact and enrol 120 healthy random individuals (sampling bias).
2. preanalytical conditions may not reflect usual care, as most primary care samples are subject to transportation
3. it is not feasible to determine age/sex-dependent reference intervals for tests that are age and sex dependent, such as serum creatinine which increases rapidly with age and differs between men and women
4. terms as ”reference population” and ”health” are subjective, and characteristics of a healthy subject are difficult to define
5. bias may occur due to the relatively small sample size
6. it requires many steps and therefore more time, resources and costs
7. it is not feasible for some tests in some matrices, such as cerebrospinal, peritoneal or synovial fluid which are difficult to obtain in healthy individuals, or for some populations, such as children and geriatric individuals.

In some fields as nanotechnology, finance or economics, strategies are classified according to their direction. In that sense, top-down approaches are those going from the general to the specific and bottom-up approaches starts at the specific and moves through the general. According to that, the explained (direct) approach currently recommended for reference interval calculation could be considered a bottom-up strategy. As it starts at analysing in deep the

reference population with their specific characteristics and then uses that knowledge to recruit individuals and infer from them the distribution of the reference population. The alternative approach, known as the indirect method, would then be a top-down strategy. It starts from a general overview of the total population available in the laboratory information system (LIS) and from that, apply mathematical strategies to uncover the distribution of only the reference population. **Figure 1** represent what it was explained above.

The final objective in both approaches is to have a realistic picture of the reference population. One of the almost philosophical questions arising from that is: What type of individual is part of our reference population? Or, similar to that: What does it mean to be free from disease? Our two approaches would answer those two questions differently. With the direct method the "type of individual" selected as part of the reference population is well defined and the answer will have to be given by health professionals that will normally have a bias in their perspective of what "health" is. Using the indirect approach these questions would be answered by the data, which also could be confounded (biased), being the confound or bias decreased as the amount of data increase.

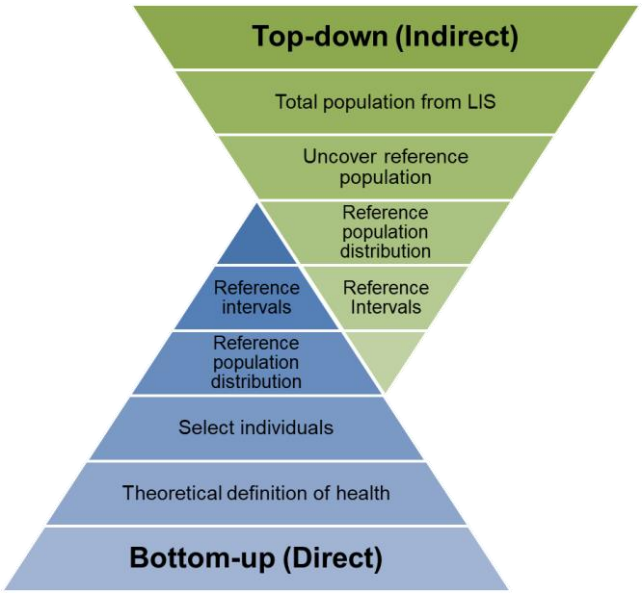


Figure 1. Direct and Indirect strategies for reference intervals calculation defined as bottom-up and top-down approaches.

Automation and informatization has increased in the clinical laboratories since the late fifties. First, analytical assays were automatized, and assay duration was decreased. In 1957, when Leonard T.Skeggs published his work "*An automatic method for colorimetric assays*" (28) it was the start of the automation era in the clinical laboratories. Then, in the seventies, the first automatic analysers controlled by microprocessors appeared which importantly improved sample handling. Finally, the introduction of computers has increased processing capacity, not only to control analysers, but also to introduce management systems for laboratories and LIS that integrate different laboratory disciplines to monitor the total testing process and generate automated reports. Nowadays there is a trend to centralize diagnostic laboratories and to archive medical test results from big geographical areas into a single LIS, ensuring a similar data structure and easy data extraction. In this context, indirect methods for reference intervals calculation are emerging as a suitable alternative since it overpasses many of the drawbacks from direct methods (29) (Box 1).

1.2. Biological, pre-analytical and analytical considerations

Some considerations are needed prior to reference interval calculations. First, some biologically determined issues, such as stratification or data transformations, should be assessed prior to any calculation method. Then, preanalytical issues should be studied as it is known that sample collection, tube type, centrifugation conditions and processing time are conditions that significantly affect test results. Finally, also analytical considerations are needed as reference intervals may be method and/or calibrator specific and could introduce interlaboratory variations that should be known in advance. To avoid preanalytical and analytical problems adequate quality controls are needed. All these considerations will be treated within this section.

1.2.1. Biological – Stratification, Biological variation, and Data transformation

Clinical laboratory results vary between subjects and within subjects for different reasons such as normal physiological processes, (epi-)genetic differences, environmental factors, and pathology (30). Knowledge about all these reasons is important to calculate, interpret and communicate reference intervals.

Stratification and biological variation

Some biological variables may importantly influence test results and will determine a different reference population. If there are statistically significant or clinically relevant differences, reference intervals should be established based on these groups due to the implications they may have for clinical management of patients. Several variables can be considered depending on the analytical measurand: age, sex, ethnicity, body mass index or lifestyle, among others. Age and sex are the two most used partitioning elements. A recent study published by Özcürümez and Haeckel (31) created a list of biological variables influencing reference intervals. When the difference in reference intervals between subpopulations is clinically relevant, stratification is recommended. Some approaches for that were already outlined (32,33).

It is important to note that some of the indirect methods that will be explained in the next section allow the presentation of continuous reference intervals avoiding the need for partitioning, by applying regression and cubic spline techniques (34–36).

Reference intervals calculations in measurands with circadian variation should be correctly designed. It is recommended to collect the data for at least one year to avoid any possible circadian or circa-seasonal effect. Phlebotomy is recommended to be performed always in the same timeframe (normally between 7 and 10 am). In cases where some circadian variation or time of phlebotomy cannot be assured, it is essential to consider the expected degree of confounding (or bias) and state this as a limitation of the reference intervals study. Examples of measurands with circadian rhythms can be found elsewhere (37–39).

Data transformation

The distribution of the test results is an important issue for reference interval calculation. This point must be assessed for each medical test studied. Most methods for reference interval calculation assume a Gaussian or close to Gaussian distributions. Therefore data transformation is needed when this assumption is not accomplished. For large data sets (used for indirect methods) the formal tests of normality are very sensitive to a deviation from normality (40) and other approaches as visual inspection are usually followed to assess normality. The most common data transformation used is box cox (note that log transformation is a type of box cox) or Manly transformation (41).

For the direct methods, having defined a priori the "normal" population, statistical management of the data is oriented to decide which statistical test is more suitable to use. For this, the possible outliers (i.e. Tukey exclusion test), and the normality of the distribution for the selection of parametric (mean ±2 standard deviation) or non-parametric (percentiles) methods are assessed. To check for normality, several tests are available. Due to the small sample sizes in direct methods (120 individuals) the Shapiro-Wilk test would be the preferred option since it provides more power than the Kolmogorov-Smirnov test (42).

1.2.2. Pre-analytical – Phlebotomy and Sample handling

Some preanalytical issues are also known to affect test results and therefore potentially influence the reference intervals calculations. Posture is one of them; phlebotomy in outpatients is supposed to be done always in a seated position, with some exceptions for clinical tests known to be highly affected by the posture (43). Instead, inpatients are usually in a supine position. This issue is important to be considered when using inpatients and/or outpatient data for reference interval calculations by indirect methods.

Other preanalytical variables affecting the test results are tube type, and conditions (of time and temperature) between phlebotomy and centrifugation which will have to be well recorded and

reported when possible. Also, endogenous interferents as bilirubin, hemoglobin or lipemia should ideally be collected with the test results when collecting data for direct or indirect reference interval calculation in order to enable beforehand decision making regarding the need for exclusion based in case of significant interference on the medical test results.

1.2.3. Analytical – Methods and Quality control

Before using any analytical data for reference intervals calculation, IVD test standardization/harmonization and test result validity should be considered (44). Both the European IVDD 98/79/EC (1998) (17) and the IVDR 2017/747 (2017) (45) demand metrological traceability of controls and calibrators to higher order reference measurement procedures and reference materials when available (46) and, on top of that, ISO 17511:2020 demands traceability of patients' test results to higher order Reference Measurement Systems. Thus, it is important that, 1) tests are standardized by the IVD industry and meet predefined analytical performance specifications 2) laboratory specialists are aware of these regulations and implement these standardized tests (47), and 3) targeted commutable materials for trueness verification are used in EQA-programs. In the Netherlands, these latter materials were developed and considered to be the 'Holy Grail' of the Calibration 2.000 program (48). The implementation of the Dutch External Quality Assessment (EQA) Program 'SKML Combi New Style' in 2005, using commutable and targeted sera, has proven to be very effective in reducing median inter-laboratory coefficients of variation for electrolytes, substrates and enzymes in the Netherlands (49). A comparability study between analytical methods in Spain, using also commutable materials from SKML, shows that implementation of internationally endorsed Reference Measurement Services by the IVD-manufacturers -in line with the metrological traceability concept and with measurement uncertainty within limited allowable measurement uncertainty is still inadequate/ insufficient. Ricos et al. have already recommended the change to pyridoxal phosphate methods for alanine aminotransferase (ALT) and aspartate aminotransferase (AST) measurements, the use of enzymatic method for creatinine

measurement, the change to pyruvate-to-lactate methods for lactate dehydrogenase (LDH) measurement and the use of commutable calibrators for electrolytes (50,51). These recommendations are important to recall.

Thus, medical laboratories should preferentially select field methods that are traceable to IFCC recommended methods and use commutable calibration materials and/or value-assigned EQA materials to 1) improve between and within laboratory variation and methods equivalence (49), 2) allow calculation and comparison of reference intervals between laboratories using the direct or indirect method, 3) allow the implementation of national common or even global harmonized reference intervals, and 4) implement a sustainable surveillance system to structurally monitor the established common reference intervals.

In addition, it is important that the stability over time of the medical test used is controlled and monitored by stringent internal and external control samples (using preferably commutable value-assigned external quality assurance programs, if available), reducing possible variability due to changes of lots in either the reagents or the calibration materials. When commutable external quality materials are not available, comparison of daily, weekly and/or monthly averages or medians could be a good method to test for longitudinal stability (33).

1.3. Overview of indirect methods

In the indirect methods for reference interval calculation, the statistical data management plan has the greatest weight to obtain the best possible information from the available data set. In these methods, data generated for the diagnosis/monitoring of individuals are used (re-used) for the identification of new information (in this case obtaining population reference intervals). Having adequate statistical methods is very important to achieve this goal. In the following sections, specific needs, and statistical options for calculating reference intervals using indirect methods are described.

1.3.1. Data reuse

As broadly explained in previous sections, the indirect methods reveal the reference population within the total data already available in the LIS. Therefore, data reuse is one of the key questions when we start using those methods.

Data source

In general, it is recommended for most methods to use data from primary care patients and/or outpatients. In the population of data from primary care (or outpatients), a significant number of individuals will be 'healthy'. Many of the medical test results of these individuals will be derived from regular health checks or to rule out disease (and in general very few test results are likely to reflect pathology). Inpatients have acute pathophysiological conditions, are subjected to shock treatments with an abundant supply of intravenous fluids between other situations that may contribute to the introduction of noise in the data (29). Also, depending on the setting of laboratory, it may be important to eliminate data from patients suffering from a specific disease, some subgroups of disease, patients using certain drugs, or when phlebotomy was performed at home (e.g. when primary care patients could not visit the laboratory due to illness). If information on underlying disease is directly available, this is the preferred way to set inclusion/exclusion criteria if needed. However, when the information about individual pathological conditions is not available in the laboratory information system, other information about the medical test request could be used (e. g. the specialty of the physician requesting the test, a combination of tests requested by specific protocols). E.g. when establishing reference intervals for serum creatinine, exclusion of the test data from patients which were referred to the clinical laboratory by the nephrologist or urologist, could be recommended, as these patients may have underlying kidney pathology. As an alternative, some studies also exclude data from subjects who had repeated serial measurements (52) as this could indicate pathological conditions of patients that require follow up and may introduce confounding (bias)

of the calculated reference intervals. Nevertheless, some indirect methods, such as Bhattacharya or Hoffman, do seem to allow the use of data from inpatients because pathological results can be detected and will be (automatically) statistically deleted.

Sample size

When using the indirect approach, in general, sample size does not imply a limitation due to the large amount of data available. Despite this, it is appropriate to define minima that ensure statistical robustness and low confidence intervals for each calculated reference interval (41). According to IFCC C-RIDL (29) it is recommended to use at least 1,000 data points, with at least 750 data points for each category (usually by sex and age) (53). It is also important to note that rounding might also influence the accuracy of the result, more importantly as sample size decreases.

1.3.2. Statistical analyses

In the different projects that have been described in the literature (25,32,52,54–59) the statistical methods used can be grouped into two main data management strategies based on the assumption of where the data from the reference population is found within the total datasets:

- **Group A:** It is based on the assumption that results from individuals with some pathological condition can be separated from the reference population using pre-cleaning steps of the database. These methods are performed in two phases: i) the pre-cleaning step and ii) the calculation of the reference intervals by means of a simple method (calculating directly the 95% range either as mean ± 2 times the standard deviation or as the 2.5 and 97.5 percentiles in the non-parametric approaches).

- **Group B:** It assumes that in the total dataset there are two or more overlapping distributions or a mixture of distributions where the biggest one contains the reference

population. It applies statistical methods over the entire data collection to unravel the reference population. Some authors (41) divide this group into two, one including statistical techniques considering overlapping distributions and another group considering the mixture decomposition technique. As some methods are a combination of those two, for explanation purposes we included them in the same group.

These two strategies with the most important examples for each one of them are explained below.

1.3.2.1. Group A

There are different methods included in group A where, after a 'cleaning' step of the database, reference intervals are calculated in the same way as in the direct method (calculating directly the 95% range either as mean ± 2 times the standard deviation or as the 2.5 and 97.5 percentiles in the non-parametric approaches). Therefore, these methods only differ in the process of database cleaning.

Some authors exclude results from individuals with clinically relevant information that can be related to pathological conditions, e.g. in case of patients with a history of any chronic disease, abnormal BMI, hypertension, positive for thyroglobulin antibodies or undergoing thyroid ultrasound for calculation of thyroid hormones reference intervals (60). Other authors exclude patients with repeated measurements within a specific time period and with abnormal results in related tests (59). Another method included within this group is the Dutch NUMBER method, where an outlier exclusion method (Tukey method) is used for elimination of results from clinically interrelated tests (25), attempting to further exclude data from potentially diseased populations. For that, the authors created groups of clinically related tests and, if an individual is an outlier for any of the medical tests from a group, their result will not be used to calculate reference intervals for any of the tests within the group, but they can be used to calculate reference intervals for tests in other groups (i.e. if an individual is an outlier for sodium their

results for calcium, potassium and chloride will be excluded while their results for creatinine and urea will be included). The main novelty introduced in the latter method is that they exclude only test results and not individuals. A more detailed explanation about that method is found in section 3 (Indirect determination of biochemistry reference intervals using outpatient data).

These methods have the advantages of being simple to apply and easy to understand by non-statisticians, and to have the capacity to detect individuals or results not pertaining to the reference population anywhere in the distribution (as the clinical rules for the pre-cleaning step are independent of the data distribution). Nevertheless, those clinical rules need to be very well designed by laboratory specialists to mitigate the risk that results from patients with limited clinical information or with conditions not considered by the clinical rules are not removed and might affect the identification of the reference population.

1.3.2.2. Group B

As explained before, the data of healthy and non-healthy individuals show a certain degree of overlap, which depends on the type of test and/or its unequivocal contribution to the definition of health and disease. Based on this premise, some statistical methods have been directly applied to laboratory databases that allow these two populations to be adequately separated. There are two classical methods that are based on using graphic strategies to perform this separation: the Hoffmann method and the Bhattacharya method (29). Both methods try, by different means, to identify a gaussian distribution within the total dataset as the reference population. In databases in which, in addition to the population of healthy individuals, there is another population of individuals (usually non-healthy) with a significant size, this second population negatively influences the determination of the reference intervals using the Hoffmann method. In contrast, the Bhattacharya method is less influenced by the patient population (29). An important limitation of the Bhattacharya method is the subjective influence of the variables defined when applying the method. It is necessary to define the bin size (size of

the ranges of numerical values into which the data are sorted), bin location and number of bins in each data set used. A better overview of these problems and a well-constructed solution is found in section 4 (Harmonization of indirect reference intervals calculation by the Bhattacharya method). In both Hoffmann and Bhattacharya methods, the graphical representation of the data plays a fundamental role in the estimation of the reference intervals, but this is not always necessary in the case of Bhattacharya method. A comparison between the indirect Bhattacharya methods and the - IFCC recommended direct method, published in 1990 (61), showed important differences between calculated reference intervals. It was shown that observed differences were due to the statistical methods and not just to the reference population and that those differences depend on the shape of the distribution.

Modern methods classified within this group are the Truncated maximum likelihood (TML) and the Truncated minimum chi-squared (TMC). TML was developed by Arzideh and colleagues (54) and it is based on the maximum likelihood estimation of the mean, the standard deviation and the variance from a power normal distribution that includes only the subjects from the reference population ('healthy' subjects). This power normal distribution is selected by truncation using an optimization algorithm (54). First, non-parametric density functions are estimated for the distribution of the total sample group (combined non-diseased and diseased) using smoothed kernel density estimation. In the next step, two density functions are obtained: one for the healthy population and another for non-healthy population. The deviation from the normal distribution is detected by a goodness of fit test for identifying the non-healthy population. Finally, the intersection points between the density function (healthy and non-healthy) establish the truncation points for the power normal distribution. Percentiles of this estimated distribution of the reference population are then considered the reference intervals. This method is freely available online in an automated program, known as the Reference Limit Estimator (RLE). Some results from the RLE are presented in section 3 (Indirect determination of biochemistry reference intervals using outpatient data).

The TMC method developed by Wosniok and Haeckel (62) is similar to the TML method in several aspects. It also estimates the mean, the standard deviation and the variance from a power normal distribution that include only the subjects from the reference population ('healthy' subjects). In this method results are included in ranges as not continuous data, plotted into a histogram, and modelled. It identifies the interval including the reference population, determined by truncation points, by fitting a power normal distribution to a series of candidate possible truncation points. An assessment is done per each of the candidate truncation points and the best one is selected according to the minimization of the chi-squared distance. As in TML, percentiles of this estimated distribution of the reference population are considered the reference intervals. These last methods also have the option to generate continuous reference intervals using the technique of "splines" (cubic smoothing spline) and avoiding the need of age partitioning.

The techniques of this group allow the identification and separation of the reference population in a robust way. Nevertheless, they are more difficult to apply and to interpret clinically.

1.3.3. Advantages and Limitations of indirect methods

Indirect methods are arising as a very promising solution for reference intervals calculation by individual laboratories and for having better harmonized reference intervals between geographical areas. Variation in the reference intervals between clinical laboratories affects patients directly, leading to disparity in clinical interpretations from the same results or unnecessary repetition of medical tests (46,63). This reality has become more important nowadays since people are increasingly moving (within the country) and visiting doctors in different healthcare settings. National or regional electronic systems from primary care are receiving results from different laboratories. Harmonization of reference intervals, obtained by an indirect data mining approach, will enable harmonized data exchange between healthcare

systems and help reduce the need for repeated laboratory tests when patients are seen by different doctors in different care settings.

Other advantages, opposed to the disadvantages of the direct method presented in Box 1, include: 1. When using primary care data, a variety of individuals will be included in the dataset avoiding the selection and sampling bias; 2. Sample handling will be the same as in routine analyses; 3. It allows the representation of different ages and even the calculation of continuous reference intervals by age as explained for the TML and TMC methods; 4. A subjective definition of reference population is not needed, avoiding the difficulty to define 'health'; 5. Higher robustness with lower confidence intervals derived from large sample sizes; 6. Time, resources and costs are significantly reduced and 7. It opens the possibility to explore the calculation of reference intervals in special individual cases, such as in children or in matrices that are difficult to obtain such as cerebrospinal fluid.

Nevertheless, some limitations of the indirect approaches must be considered: 1. the possible effect of diseased subpopulations on the derived reference intervals , 2. Not an oficial method is available yet to check if the obtained reference intervals are correct and valid, 3. several (pre-)analytical changes or inconsistencies (i.e. methodology changes, calibrator or reagent lot changes, quality control issues) could lead to potential errors , 4. several statistical methods have been proposed but no consensus or official recommendations about 'which method to use when' are available yet. Thus, indirect approaches need further validation.

2. AIM AND OUTLINE OF THIS THESIS

2. Aim and outline of this thesis

2.1. Aim

The aim of this thesis is twofold. On the one hand, to explore the differences in the reference interval results obtained by three indirect methods (NUMBER Dutch method, Reference Limit Estimator German method and Bhattacharya traditional method) using the same dataset. On the other hand, to provide easily accessible tools and descriptions that enable laboratory specialists to calculate reference intervals for their own laboratory using indirect methods.

2.2. Study population

The studies presented in this thesis were performed with an anonymized dataset extracted from Vall d'Hebron laboratories in Barcelona. Basic biochemistry medical test results analysed from January 1st 2018 until and including 31st of December 2018 in the Clinical Laboratories were used in the first and second study and, results analysed from January 1st 2019 to December 31st 2019, were also included in the second study. The data were included in the studies only when requested by primary care centres, employees analytical control centres, sexual and reproduction centres, and geriatric centres. Test results were excluded when phlebotomy was performed in the hospital (inpatients), drug addiction centres, mental health centres, external emergency centres, the prison women centre, or at home (e.g. when primary care patients could not visit the laboratory due to illness).

Based on this common dataset and as outlined thereafter, for each study presented, other preanalytical or clinical considerations were differently applied for further filtering of the data.

2.3. Outline of this thesis

The aim of section 3 is to explore the Dutch NUMBER method and apply it to a dataset from a single laboratory in Barcelona. The section includes a description of the method, created for

calculating national standardized or harmonized reference intervals for clinical chemistry tests in The Netherlands, applied to the dataset in our study. As the method was already applied to Dutch laboratories, the reference intervals obtained were compared with the results published in the first NUMBER project in the Netherlands. Observed differences were discussed and several hypotheses were made to explain them. Section 3 also contains a comparison between results by NUMBER and the Reference Limit Estimator method using the same dataset previously described. Reference intervals results with Bhattacharya method are presented in Section 4. In that section, Bhattacharya analyses were performed using the St Vincent's hospital Spreadsheet available online (http://www.sydpath.stvincents.com.au/). For the appropriate use of this tool, specific criteria were defined to reduce inherent subjectivity of the method and to reduce between-user variability. In section 5 the results are summarized and discussed. Section 5 includes a description of the possible implications of this thesis and recommendations for further validation studies to guarantee that the calculated reference intervals based on indirect methods are fit-for-clinical purpose. It also contains the next research questions arose during the development of the presented thesis. Section 6 includes the final conclusions derived.

3. INDIRECT DETERMINATION OF BIOCHEMISTRY REFERENCE INTERVALS USING OUTPATIENT DATA

3. Indirect determination of biochemistry reference intervals using outpatient data

3.1. Summary

Objectives: The aims of this study were to determine reference intervals in an outpatient population from Vall d'Hebron laboratory using an indirect approach previously described in a Dutch population (NUMBER project), to compare the reference intervals results between the Vall d'Hebron population and the Dutch population using the same method, and to calculate the reference intervals for the Vall d'Hebron population with another alternative method known as the Reference Limit Estimator.

Material and Methods: We used anonymized test results from individuals visiting general practitioners and analyzed during 2018. Analytical quality was assured by adequate performance on external quality assessment (EQA) programs, daily and monthly average monitoring and by assessing longitudinal accuracy between 2018 and 2020 (using trueness verifiers from Dutch EQA). Per test, outliers by biochemically related tests were excluded, data were transformed to a normal distribution (if necessary) and means and standard deviations were calculated, stratified by age and sex. In addition, the Reference Limit Estimator method was also used to calculate reference intervals using the same dataset. Finally, for standardized tests, the reference intervals obtained were compared with the published NUMBER results calculated in the Dutch population and flagging rates (percentages of measurements below and above the lower and upper reference limits) were calculated with an independent dataset using the reference intervals currently used in Vall d'Hebron (direct reference intervals) and the reference intervals calculated using the NUMBER method (indirect reference intervals).

Results: Reference intervals were calculated for 16 biochemistry medical tests using data from 509,408 clinical requests. For biochemical tests following a normal distribution, similar reference

intervals were found between Vall d'Hebron and the Dutch study. The upper limits of Gamma-glutamyl transferase were markedly higher in the Dutch study compared to Vall d'Hebron results which suggest a lifestyle component supported by two main facts: the differences were more important in adult individuals and further age stratification showed the same pattern. For creatinine and urea, reference intervals increased with age in both populations which support earlier studies on the age-related decline in renal function. Creatine kinase and uric acid reference intervals were higher in both populations compared to conventional reference intervals. In general, the results calculated with the NUMBER method and the Reference Limit Estimator method showed in a great extent similar results, but lower reference intervals were found with the Reference Limit Estimator method for Gamma-glutamyl transferase, creatinine and creatine kinase. Regarding flagging rates, higher flagging rates were noted for the currently used reference intervals compared with the indirectly calculated reference intervals by the NUMBER method.

Conclusions: Medical test results following a normal distribution showed comparable and consistent reference intervals between studies. Therefore, a simple indirect method is a feasible and cost-efficient approach for calculating reference intervals. Yet, for generating standardized calculated reference intervals that are traceable to higher order materials and methods, efforts should also focus on test standardization and bias assessment using commutable trueness verifiers. Adequate implementation of common, metrologically traceable reference intervals is the goal for guaranteeing safe and clinically effective use of medical tests. As a first step, in this study, method and population specific refined reference intervals were derived for biochemistry tests.

3.2. Publication

Luisa Martinez-Sanchez[1,2,3*], Christa M Cobbaert[2], Raymond Noordam[4], Nannette Brouwer[5], Albert Blanco-Grau[1], Yolanda Villena-Ortiz[1,3], Marc Thelen[6,7,8], Roser Ferrer-Costa[1], Ernesto Casis[1] , Francisco Rodríguez-Frias[1,3], Wendy PJ den Elzen[1,2,9,10].

[1] Biochemistry Department, Clinical Laboratories, Vall d'Hebron University Hospital, Barcelona, Spain.

[2] Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Centre, Leiden, The Netherlands.

[3] Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[4] Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands.

[5] Diagnost-IQ, Expert Centre for Clinical Chemistry, Purmerend, The Netherlands.

[6] Laboratory for Clinical Chemistry and Hematology, Amphia, Breda, The Netherlands

[7] Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek, Nijmegen, The Netherlands.

[8] Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, the Netherlands.

[9] Atalmedial Diagnostics Centre, Amsterdam, The Netherlands.

[10] Department of Clinical Chemistry, Amsterdam UMC, Amsterdam, The Netherlands

PLoS One. 2022 May 19;17(5):e0268522.

10.1371/journal.pone.0268522

PLOS ONE

RESEARCH ARTICLE

Indirect determination of biochemistry reference intervals using outpatient data

Luisa Martinez-Sanchez[1,2,3]*, Christa M. Cobbaert[2], Raymond Noordam[4], Nannette Brouwer[5], Albert Blanco-Grau[1], Yolanda Villena-Ortiz[1,3], Marc Thelen[6,7,8], Roser Ferrer-Costa[1], Ernesto Casis[1], Francisco Rodríguez-Frias[1,3☯], Wendy P. J. den Elzen[1,2,9,10☯]

1 Clinical Laboratories, Biochemistry Department, Vall d'Hebron University Hospital, Barcelona, Spain, **2** Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Centre, Leiden, The Netherlands, **3** Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Spain, **4** Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands, **5** Diagnost-IQ, Expert Centre for Clinical Chemistry, Purmerend, The Netherlands, **6** Laboratory for Clinical Chemistry and Hematology, Amphia, Breda, The Netherlands, **7** Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek, Nijmegen, The Netherlands, **8** Department of Laboratory Medicine, Radboud University Medical Centre, Nijmegen, The Netherlands, **9** Atalmedial Diagnostics Centre, Amsterdam, The Netherlands, **10** Department of Clinical Chemistry, Amsterdam Public Health research institute, Amsterdam UMC, Amsterdam, The Netherlands

☯ These authors contributed equally to this work.
* luisa.maria.martinez.lm@gmail.com

Citation: Martinez-Sanchez L, Cobbaert CM, Noordam R, Brouwer N, Blanco-Grau A, Villena-Ortiz Y, et al. (2022) Indirect determination of biochemistry reference intervals using outpatient data. PLoS ONE 17(5): e0268522. https://doi.org/10.1371/journal.pone.0268522

Editor: Michael Nagler, Inselspital, Bern University Hospital, SWITZERLAND

Received: August 11, 2021

Accepted: May 2, 2022

Published: May 19, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The aim of this study was to determine reference intervals in an outpatient population from Vall d'Hebron laboratory using an indirect approach previously described in a Dutch population (NUMBER project). We used anonymized test results from individuals visiting general practitioners and analysed during 2018. Analytical quality was assured by EQA performance, daily average monitoring and by assessing longitudinal accuracy between 2018 and 2020 (using trueness verifiers from Dutch EQA). Per test, outliers by biochemically related tests were excluded, data were transformed to a normal distribution (if necessary) and means and standard deviations were calculated, stratified by age and sex. In addition, the reference limit estimator method was also used to calculate reference intervals using the same dataset. Finally, for standardized tests reference intervals obtained were compared with the published NUMBER results. Reference intervals were calculated using data from 509,408 clinical requests. For biochemical tests following a normal distribution, similar reference intervals were found between Vall d'Hebron and the Dutch study. For creatinine and urea, reference intervals increased with age in both populations. The upper limits of Gamma-glutamyl transferase were markedly higher in the Dutch study compared to Vall d'Hebron results. Creatine kinase and uric acid reference intervals were higher in both populations compared to conventional reference intervals. Medical test results following a normal distribution showed comparable and consistent reference intervals between studies. Therefore a simple indirect method is a feasible and cost-efficient approach for calculating reference intervals. Yet, for generating standardized calculated reference intervals that are traceable to higher order materials and methods, efforts

should also focus on test standardization and bias assessment using commutable trueness verifiers.

Introduction

Specialists in clinical chemistry should provide accurate and useful information into their clinical laboratory reports. Reference intervals are commonly presented together with the actual analytical results. Their correct evaluation is crucial due to their use as a clinical decision-making tool [1]. Most manufacturers provide reference intervals in their technical documentation. According to ISO15189:2012, it is the responsibility of the laboratory to either validate them, find reference intervals from other sources or calculate the appropriate reference intervals for their method and population. Two different approaches to calculate reference intervals could be used: (a) The procedure recommended by the International Federation of Clinical Chemistry (IFCC), known as the direct method and [2,3] (b) an alternative approach, known as the indirect method [4].

The direct approach uses a bottom-up strategy. In this sense, the reference population will be analysed in detail in order to unravel their characteristics and then a realistic "model" will be constructed to derive the distribution of the reference population and the reference intervals. This methodology has been widely used and standardized [2], but it is laborious and expensive. In addition, it struggles with selection bias, in combination with subjective terms as "reference population" and "health" [5]. As an alternative approach, the indirect method uses a top-down approach. It starts by acquiring a general overview of the total population by analysing clinical data from the laboratory information system (LIS) and, from this, filtering to uncover the distribution of the reference population and the reference intervals. This approach has several advantages, since 'big' analytical data is more accessible nowadays [4]. Automation has increased in clinical laboratories. This has resulted in the centralization of medical tests from a big geographical area around Vall d'Hebron into a single LIS, which guarantees a common diagnostic test process and a similar data structure for extraction [6].

As a result of differences between reference intervals provided by different manufacturers and individual efforts to verify or select them from the literature, reference intervals vary per laboratory potentially resulting in unequal treatment and patient harm [7]. Standardization and harmonization efforts, which are currently successfully employed in several countries, are necessary to improve presentation and interpretation of laboratory results [8–14]. In the Netherlands, we previously determined nationally standardized reference intervals for clinical chemistry tests using an indirect "big data" approach [14]. A simple and straightforward workflow using the same approach is presented in this work. First, we determined indirect reference intervals using the NUMBER approach in a dataset of routine clinical chemistry values of the Vall d'Hebron laboratory population in Barcelona. The clinical laboratory Vall d'Hebron is the result of a fusion between three laboratories of the Catalan Institute of Health in Barcelona in 2014. It processes between 15,000 and 18,000 samples a day and covers a population of 1.2 million people, resulting in a very large amount of medical test results a year. This provided us with a unique opportunity to use only the data of a single laboratory using one single method to establish reference intervals, which is very important given the lack of harmonization in Spain [15]. Secondly, for those tests that are internationally standardized and produce test results traceable to standards and/or methods of higher order, we compared the reference intervals obtained from this study with the results published in the first NUMBER project in

the Netherlands [14]. Finally, the reference intervals for creatinine kinase and uric acid were investigated, since no consensus was obtained yet in the NUMBER project [14].

Material and methods

Study design

We extracted anonymized medical test results from individuals visiting general practitioners, analysed from January 1st 2018 until and including 31st of December 2018 in the Clinical Laboratory Vall d'Hebron from the LIS. The presented study was considered suitable from the point of view of ethics and science by the corresponding Clinical Research Ethics Comittee.

We included test results from patients visiting primary care centres, employees analytical control centres, sexual and reproduction centres and geriatric centres. Test results were excluded when phlebotomy was performed in the hospital (inpatients), drug addiction centres, mental health centres, external emergency centres, the prison women centre, or at home (e.g. when primary care patients could not visit the laboratory due to illness) since we expected substantial differences in health status in these settings that can add noise to the data [4]. We performed sensitivity analyses to compare the distribution between all the included centres, showing no signs of sample or sex bias between centres (results not shown).

Pre-analytical and analytical considerations

Samples were collected from 62 blood collection centres and were transported via 8 different routes to the laboratory. Serum tubes for biochemistry tests included separating gel and coagulation activator (BD Vacutainer®). Phlebotomy order of draw was always performed as advised by the EFLM pre-analytical workgroup [16]. The samples were transported to the laboratory in cool boxes with a temperature monitoring system. After arriving in the laboratory, the samples were centrifuged either 12 minutes at 3,500 rpm (2,438 g) when handled manually or 10 min at 3,000 rpm (2,113 g) when on the track.

Eighteen biochemistry tests were measured on three parallel AU5800 chemistry analysers (Beckman Coulter®). Detailed descriptions of the methods and the recommended reference intervals (calculated by direct approaches) according to Beckman's IFU are presented in S1 Table. Tests included: albumin (CRM470 traceable), calcium (NIST-SRM-909bL1 traceable), creatinine (NIST-SRM-967 L1 traceable), lactate dehydrogenase (LDH) (not traceable to higher order reference material (NTRM)), magnesium (NIST-SRM-909bL2 traceable), anorganic phosphate (NTRM), total bilirubin (NIST-SRM-916a traceable), total protein (NIST-SRM-927c traceable), uric acid (traceable to isotope dilution Mass Spectrometry), urea (NIST-SRM-909bL1 traceable), chloride (NIST-SRM-919 traceable), potassium (NIST-SRM-918 traceable), sodium (NIST SRM-919 traceable), alkaline phosphatase (ALP) (NTRM), alanine aminotransferase (ALT) (NTRM), aspartate aminotransferase (AST) (NTRM), gamma glutamyltransferase (GGT) (traceable to IFCC reference method) and creatine kinase (CK) (traceable to IFCC reference method).

Analytical quality assurance

To assure the outpatient data quality, we first examined the monthly results from external quality control scheme from the Spanish Society of Clinical Chemistry (SEQC), basic biochemistry scheme. In this scheme, the results obtained in our laboratory are compared with the average calculated from every laboratory participating in the program using the same analytical method and/or instrument. When our result was within one time the standard deviation from other laboratories participating in the scheme using the same method, data from this

particular month and test were accepted as valid. When our result was above or below three standard deviations, we excluded the data from that particular test and instrument for that month. When the result was between the second and third standard deviation, we analysed the daily average outpatient results for the particular test and month.

Daily averages were investigated to ensure longitudinal accuracy of the results over time. Averages were calculated per batch of 200 results a day and were compared with the average per month and year. Plots were visually inspected in order to decide whether the analytical quality was sufficient using the biological variation of the monthly and yearly mean as a reference and comparing it visually with the daily mean.

Finally, due to the lack of commutable trueness verifiers in 2018, we further validated the quality of the obtained reference intervals by using a new data extraction of test results from 2020. In 2020, our laboratory participated in the fortnightly EQA scheme from the Dutch EQA organizer Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek (SKML) which uses commutable and value-assigned trueness verifiers [17]. In all EQA reports, the Multi sample evaluation (MUSE) scores for all tests were > = 1 (meaning a total allowable error sigma value over 2), indicating adequate performance for all tests [18]. To verify the calculated reference intervals deduced from the 2018 data, outpatient data from July to October 2020 were selected, considering the same analytical and pre-analytical considerations explained previously for the main data. To that end, we designed an algorithm that computed 2,000 random samples of 200 test results each time. Next, for each random sample of 200 test results, we calculated the proportion of cases residing within the reference intervals deduced from the 2018 dataset. Then we calculated the mean of these 2000 proportions for each test. When the mean of the proportions (Prop.2020) was higher than 95 %, we considered the reference interval as valid. This protocol was based on the CLSI EP28-A3C for reference intervals transference modifying the sample number from 20 to 200 and repeating the protocol 2,000 times [2].

Clinical criteria

To avoid pre-analytical issues that could confound the reference intervals, results from hemolyzed, lipemic and icteric samples were excluded when indices were > = 2 on a 0–5 scale (Beckman Coulter® AU5800, S2 Table). In addition, since the icteric index could also be a good indicator for liver dysfunction, samples with icteric indices > = 1 were also excluded for total bilirubin, ALT, AST, ALP and GGT.

For the calculations on CK, individuals with AST results higher than decision limits in Vall d'Hebron laboratory (50 U/L in men and 35 U/L in women) were excluded, in order to exclude patients with skeletal muscle injury [19].

Statistical analyses

Reference intervals were calculated per test using an automatic calculator programmed in R [20] (version 3.6.1), following the workflow presented in Fig 1.

Firstly, we used the Tukey method [21] to identify and discard outliers. The lower and upper cut-offs for outlier exclusion were defined as Q1-(1.5xIQR) and Q3+(1.5xIQR), respectively, being Q1 the lower sample quartile, Q3 the upper sample quartile and IQR the interquartile range (Q3-Q1). The same workflow and outlier exclusion procedures were used as the ones described in the NUMBER project [14], where outliers from biochemically related tests based on defined groups were excluded. Defined groups were:

- Electrolytes: calcium, chloride, potassium, sodium
- Bone: calcium, magnesium, phosphate

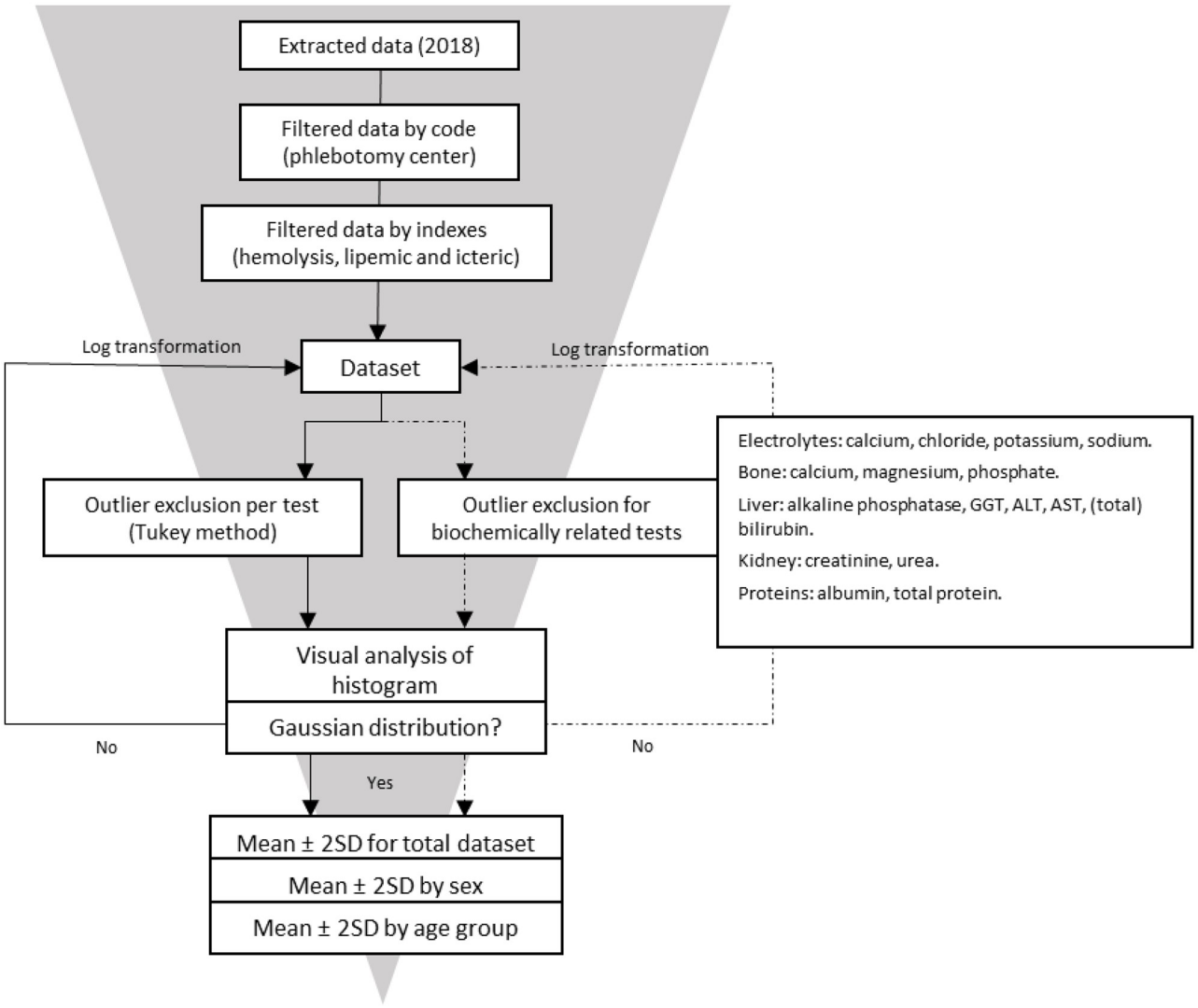


Fig 1. Study workflow. Workflow used for calculating reference intervals in Vall d'Hebron laboratory hospital by an indirect method based on the NUMBER study.

https://doi.org/10.1371/journal.pone.0268522.g001

- Liver: alkaline phosphatase, GGT, ALT, AST, (total) bilirubin
- Kidney: creatinine, urea
- Proteins: albumin, total protein

For calcium, two groups of tests were considered biochemically related. The histograms were visually inspected, and formal tests were performed (Z score for Skewness and Kurtosis) to determine the presence of a normal Gaussian distribution. Given the large numbers of test results, the formal tests of normality were very sensitive to a deviation from normality [22]. In such cases, visual inspection was considered decisive. If a normal distribution was absent, we performed a log transformation on the original data.

The reference intervals were calculated as mean plus/minus two times the standard deviation (mean ± 2SD) both for the total dataset and per subgroup when a minimum of 120 test

results per group were available. Also 90% confidence intervals for the lower and upper limit were calculated. We used pre-defined subgroups analogous to the NUMBER project [14]:

- Sex: Male / Female
- Age:
  - Newborns /infants: <28 days of age (WHO definition), 28 days to <1 year
  - 1–5, 6–12, 13–18, 19–50, 51–65, 66–80, 80+ years

In addition, in order to test a recently published hypothesis [23] stating that certain differences between indirect studies may be due to diverse age representations into the age groups, sensitivity analyses with additional age categories were performed for ALT and GGT.

Per test and per group boxplots were visually inspected after outlier elimination in order to decide whether or not subgroup differentiated reference intervals were necessary. In addition, reference intervals results were compared with the reference limit estimator method employed by the group of Haeckel, Wosniok and Arzideh [24] using the same dataset.

Lastly, flagging rates were calculated to verify the clinical suitability of the reference intervals using an independent dataset (January–June 2019). The percentages of measurements below and above the lower and upper reference limits were calculated per test.

Results

We extracted anonymized test results from a total of 530,778 clinical requests for a period of one year from the laboratory system of the Clinical Laboratory Vall d'Hebron University Hospital. After filtering by phlebotomy centre, 3.01% clinical requests were excluded. We discarded an additional 0.70% of the clinical requests because of hemolysis, 0.02% because of icteria, and 0.35% because of lipemia. The final dataset consisted of 509,408 requests.

Analytical performance, based on monthly external quality controls was adequate for SEQC material for all tests, except for ALP in December 2018. For this period, ALP results were excluded from the analyses. Daily average results showed stable performance over the year for all tests. In the S1 Fig we show an example for calcium.

Outlier exclusion by biochemically related tests ranged from 1.27 to 16.50%. Albumin, total protein, magnesium, phosphate, calcium, sodium, potassium and chloride followed a Gaussian distribution; for all other tests we obtained a Gaussian distribution after log transformation. The calculated reference intervals by the indirect approach are presented in Table 1, stratified for sex and age categories, if necessary. Results from the reference interval quality verification protocol, tested with the new dataset from 2020 (110,237 clinical requests), are also presented in Table 1, showing acceptable results (>95%) for all tests except for some age groups, particularly for creatinine and magnesium. Confidence intervals for the lower and upper limits per test are presented in S3 Table.

In Table 2, the obtained Vall d'Hebron reference intervals from the normally distributed tests are compared with results from the Dutch NUMBER project [14]. The kidney and liver parameters for both studies are graphically displayed in different age categories for men and women in Fig 2. Similar results for GGT were found when we increased the number of age categories (S2 Fig). In addition, results from the calculated reference intervals for creatine kinase and uric acid for the Vall d'Hebron hospital and the Dutch project are presented in Fig 3.

The obtained Vall d'Hebron reference intervals for the normally distributed tests, compared with results from the Dutch NUMBER project, stratified for sex and age categories, if necessary.

Table 1. Obtained Vall d'Hebron reference intervals results using the indirect approach from the NUMBER project, stratified for sex and age categories when necessary.

Test	Unit	Gender	Age, years	n	Calculated reference intervals:		
					Low	High	Prop. 2020*
Albumin	g/dL (g/L)	M	1–5	330	3.8 (38)	4.9 (49)	95.6
			6–18	914	4.1 (41)	5.0 (50)	92.8
			19–50	4281	4.0 (40)	5.1 (51)	89.4
			51–65	3660	3.8 (38)	4.9 (49)	96.2
			66–80	4704	3.5 (35)	4.9 (49)	98.6
			80+	4299	3.2 (32)	4.7 (47)	99.0
		F	1–5	282	3.9 (39)	4.9 (49)	93.2
			6–18	1116	4.0 (40)	5.0 (50)	94.5
			19–50	6622	3.7 (37)	4.9 (49)	94.0
			51–65	5966	3.8 (38)	4.8 (48)	93.4
			66–80	7865	3.6 (36)	4.7 (47)	93.6
			80+	10767	3.2 (32)	4.6 (46)	97.2
ALP	U/L	M	13–18	381	74	218	73.9
			19–50	11577	46	133	89.2
			51–65	9928	45	135	96.2
			66–80	9520	44	137	96.2
			80+	4569	46	155	94.5
		F	13–18	880	50	184	97.9
			19–50	15757	39	130	96.3
			51–65	14428	49	152	93.2
			66–80	15285	47	147	94.5
			80+	10185	46	157	94.4
ALT	U/L	M	1–12	2391	9	32	97.6
			13–18	2978	8	38	92.8
			19–50	45779	10	55	86.3
			51–65	37916	11	51	92.5
			66–80	42390	9	43	95.6
			80+	21014	7	34	92.0
		F	1–12	2308	9	31	97.5
			13–18	4748	7	27	95.4
			19–50	77714	7	35	95.7
			51–65	52710	9	42	97.2
			66–80	60878	8	36	97.2
			80+	4353	6	29	95.5
AST	U/L	M	1–5	557	25	51	99.5
			6–12	921	20	42	92.6
			13–18	1262	15	38	91.6
			19+	50953	13	38	96.1
		F	1–5	407	26	51	99.5
			6–12	1035	18	42	96.7
			13–18	1980	13	30	93.8
			19+	78383	13	36	95.9
Bilirubin (total)	mg/dL (μmol/L)	M	6–12	123	0.23 (4)	0.84 (14)	NA
			13–18	301	0.29 (5)	1.34 (23)	97.3
			19+	16323	0.32 (6)	1.30 (22)	95.1

(*Continued*)

Table 1. (Continued)

Test	Unit	Gender	Age, years	n	Calculated reference intervals:		
					Low	High	Prop. 2020*
		F	6–18	573	0.23 (4)	1.10 (19)	93.9
			19+	24215	0.28 (5)	1.04 (18)	95.5
Calcium	mg/dL (mmol/L)	M + F	1–5	318	9.2 (2.29)	10.7 (2.67)	NA
			6–12	954	9.3 (2.32)	10.5 (2.63)	97.7
			13–18	1358	9.2 (2.29)	10.5 (2.61)	95.4
			19+	46602	8.8 (2.20)	10.3 (2.58)	93.6
Chloride	mmol/L	M + F		784	98	108	91.1
Creatinine	mg/dL (μmol/L)	M	6–12	1317	0.37 (33)	0.61 (54)	68.1
			13–18	3517	0.47 (41)	1.07 (94)	93.0
			19–50	53345	0.65 (57)	1.17 (103)	66.3
			51–65	44666	0.62 (54)	1.23 (109)	78.8
			66–80	48705	0.62 (55)	1.36 (121)	83.1
			80+	22032	0.63 (56)	1.53 (135)	85.5
		F	6–12	1364	0.37 (33)	0.59 (53)	67.5
			13–18	4804	0.45 (40)	0.83 (74)	86.6
			19–50	80957	0.47 (41)	0.90 (79)	82.1
			51–65	56832	0.47 (41)	0.95 (84)	85.5
			66–80	67350	0.46 (41)	1.09 (96)	90.7
			80+	47750	0.48 (42)	1.37 (121)	97.8
GGT	U/L	M	1–5	174	7	20	NA
			6–12	322	9	23	97.2
			13–18	1452	8	36	95.5
			19–50	31582	9	79	96.0
			51–65	26758	12	95	93.9
			66–80	28080	11	84	96.5
			80+	13160	8	79	99.1
		F	1–5	146	8	17	NA
			6–12	341	8	22	95.7
			13–18	2191	7	26	97.2
			19–50	48040	7	48	99.0
			51–65	35997	8	71	99.7
			66–80	40412	8	65	100
			80+	27174	7	66	100
LDH	U/L	M + F	6–12	257	359	643	NA
			13–18	340	274	531	NA
			19–50	2573	256	507	NA
			51–65	1963	274	534	NA
			66–80	2039	270	551	NA
			80+	1539	266	584	NA
Magnesium	mg/dL (mmol/L)	M + F		4571	1.8 (0.72)	2.4 (1.00)	89.5
Phosphate	mg/dL (mmol/L)	M	1–5	147	4.2 (1.34)	5.4 (1.74)	NA
			6–12	405	4.2 (1.33)	5.3 (1.7)	94.8
			13–18	390	3.6 (1.16)	5.4 (1.72)	94.4
			19–50	3132	2.4 (0.77)	4.7 (1.51)	96.8
			51–65	2917	2.2 (0.72)	4.3 (1.38)	90.0
			66+	6640	2.2 (0.71)	4.2 (1.34)	92.7

(*Continued*)

Table 1. (Continued)

Test	Unit	Gender	Age, years	n	Calculated reference intervals:		
					Low	High	Prop. 2020*
		F	1–5	126	4.3 (1.39)	5.4 (1.71)	96.8
			6–12	428	4.1 (1.32)	5.3 (1.71)	96.3
			13–18	729	3.5 (1.11)	5.2 (1.67)	94.5
			19–50	5874	2.6 (0.84)	4.7 (1.51)	93.0
			51–65	7970	2.7 (0.88)	4.7 (1.5)	92.8
			66+	18548	2.6 (0.84)	4.5 (1.43)	93.9
Potassium	mmol/L	M + F		257189	3.60	5.09	95.4
Sodium	mmol/L	M + F		256775	136	144	95.7
Total protein	g/dL (g/L)	M + F		35141	6.1 (61)	8.0 (80)	94.8
Urea	mg/dL (mmol/L)	M	1–5	227	15 (2.5)	45 (7.5)	93.4
			6–12	755	19 (3.1)	47 (7.7)	88.7
			13–18	996	18 (3.0)	47 (7.8)	85.7
			19–50	4709	20 (3.3)	54 (9.0)	86.1
			51–65	4457	21 (3.5)	61 (10.2)	93.9
			66–80	5680	23 (3.9)	75 (12.6)	95.6
			80+	3662	27 (4.5)	93 (15.5)	94.5
		F	1–5	167	16 (2.6)	43 (7.2)	91.0
			6–12	767	17 (2.8)	44 (7.4)	92.9
			13–18	1220	16 (2.7)	42 (7.0)	92.0
			19–50	6850	16 (2.7)	46 (7.7)	93.9
			51–65	5524	20 (3.4)	58 (9.7)	94.8
			66–80	6986	22 (3.7)	72 (12.0)	96.4
			80+	7668	25 (4.1)	97 (16.2)	96.1

Obtained Vall d'Hebron reference intervals using the indirect approach from the NUMBER project stratified by sex and age categories when necessary.

M: Male, F: Female.

[a]The reference intervals obtained from the dataset in 2018 were validated using a new dataset in 2020 when the laboratory participated in a type 1 EQA scheme. Proportion (Prop.) 2020 indicates the proportion of data from 2020 inside the calculated reference intervals. When the mean of the proportions was higher than 95%, we considered the calculated reference intervals verified.

https://doi.org/10.1371/journal.pone.0268522.t001

Results calculated using the reference limit estimator method are presented in S4 Table and S3 Fig.

Flagging rates from an independent dataset, for both the calculated reference intervals in this study and the currently used reference intervals in Vall d'Hebron laboratory are presented in Fig 4.

Discussion

Application of big data to healthcare has been a matter of interest in recent years [25]. Consequently, in laboratory medicine, where quantitative data is generated every day, machine learning, data mining, business intelligence and related concepts are starting to be used for different purposes including analytical and quality management [25]. For the determination of reference intervals, for which classical (direct) recommendations are laborious and expensive, various statistical (indirect) methods have been developed using big data [4]. It is important to remark that some specialists are concerned about the possible bias due to the presence of unhealthy individuals in the dataset. Standard and detailed protocols following this approach

Table 2. Reference intervals results from normally distributed tests.

Test	Unit	Gender	Age, years	Vall d'hebron RI:		NUMBER RI:	
				Low	High	Low	High
Albumin	g/dL (g/L)	M	6–18	4.2 (42)	5.1 (51)	4.0 (40)	5.2 (52)
			19–50	4.0 (40)	5.1 (51)	3.9 (39)	5.1 (51)
			51–65	3.8 (38)	4.9 (49)	3.7 (37)	4.9 (49)
			66–80	3.5 (35)	4.9 (49)	3.6 (36)	4.8 (48)
			80+	3.2 (32)	4.7 (47)	3.6 (36)	4.6 (46)
		F	1–5	4.0 (40)	5.0 (50)	3.9 (39)	5.0 (50)
			6–18	4.0 (40)	5.1 (51)	4.0 (40)	5.1 (51)
			19–50	3.7 (37)	4.9 (49)	3.8 (38)	4.9 (49)
			51–65	3.8 (38)	4.8 (48)	3.8 (38)	4.9 (49)
			66–80	3.6 (36)	4.7 (47)	3.7 (37)	4.8 (48)
			80+	3.2 (32)	4.6 (46)	3.6 (36)	4.7 (47)
Calcium	mg/dL (mmol/L)	M + F	6–12	9.3 (2.32)	10.5 (2.63)	9.2 (2.29)	10.3 (2.56)
			13–18	9.2 (2.29)	10.5 (2.61)	8.9 (2.23)	10.3 (2.57)
			19+	8.8 (2.20)	10.3 (2.58)	8.7 (2.18)	10.2 (2.55)
Chloride	mmol/L	M + F		98	108	97	108
Phosphate	mg/dL (mmol/L)	M	13–18	3.6 (1.16)	5.4 (1.72)	2.9 (0.88)	4.8 (1.53)
			19–50	2.4 (0.77)	4.7 (1.51)	1.9 (0.62)	4.1 (1.32)
			51–65	2.3 (0.72)	4.3 (1.38)	1.9 (0.62)	4.1 (1.32)
			66+	2.2 (0.71)	4.2 (1.34)	1.9 (0.62)	4.1 (1.32)
		F	13–18	3.5 (1.11)	5.2 (1.67)	2.6 (0.82)	4.8 (1.52)
			19–50	2.6 (0.84)	4.7 (1.51)	2.3 (0.73)	4.5 (1.44)
			51–65	2.8 (0.88)	4.7 (1.50)	2.3 (0.73)	4.5 (1.44)
			66+	2.6 (0.84)	4.5 (1.43)	2.3 (0.73)	4.5 (1.44)
Potassium	mmol/L	M + F		3.6	5.1	3.8	5.2
Magnesium	mg/dL (mmol/L)	M + F		1.75 (0.72)	2.43 (1.00)	1.73 (0.71)	2.38 (0.98)
Sodium	mmol/L	M + F		136	144	136	145
Total protein	g/dL (g/L)	M + F		6.1 (61)	8.0 (80)	6.1 (61)	7.9 (79)

https://doi.org/10.1371/journal.pone.0268522.t002

are not available yet. However, the IFCC committee on Reference Intervals and Decision Limits (c-RIDL) recently recommended and promoted the development and assessment of indirect methods, stimulating future consensus for a harmonized indirect approach [26].

In the present study, we calculated reference intervals in an outpatient population from Vall d'Hebron laboratory using the NUMBER approach created for calculating nationally standardized reference intervals for clinical chemistry tests in The Netherlands [14]. The normally distributed tests (Table 2) showed similar reference intervals between both studies and other previous projects such as the Canadian project CALIPER (direct method) [27], the Australian and New Zeeland project ARIA (direct method) [8], or the German projects (indirect methods) [23,24]. This suggests that standardized tests allow global and common use of reference intervals and a straightforward indirect method could be a valuable approach for these normally distributed tests. The comparison of the results from this study with the reference limit estimator method (S4 Table and S3 Fig) support this idea as nearly equal reference interval calculations were obtained with both methods for tests with a normal distribution.

In this project, the upper reference limits for liver enzymes from the Dutch project were always substantially higher than the upper reference limits from Vall d'Hebron laboratory. We previously already hypothesized about potential explanations for the higher upper limits in the

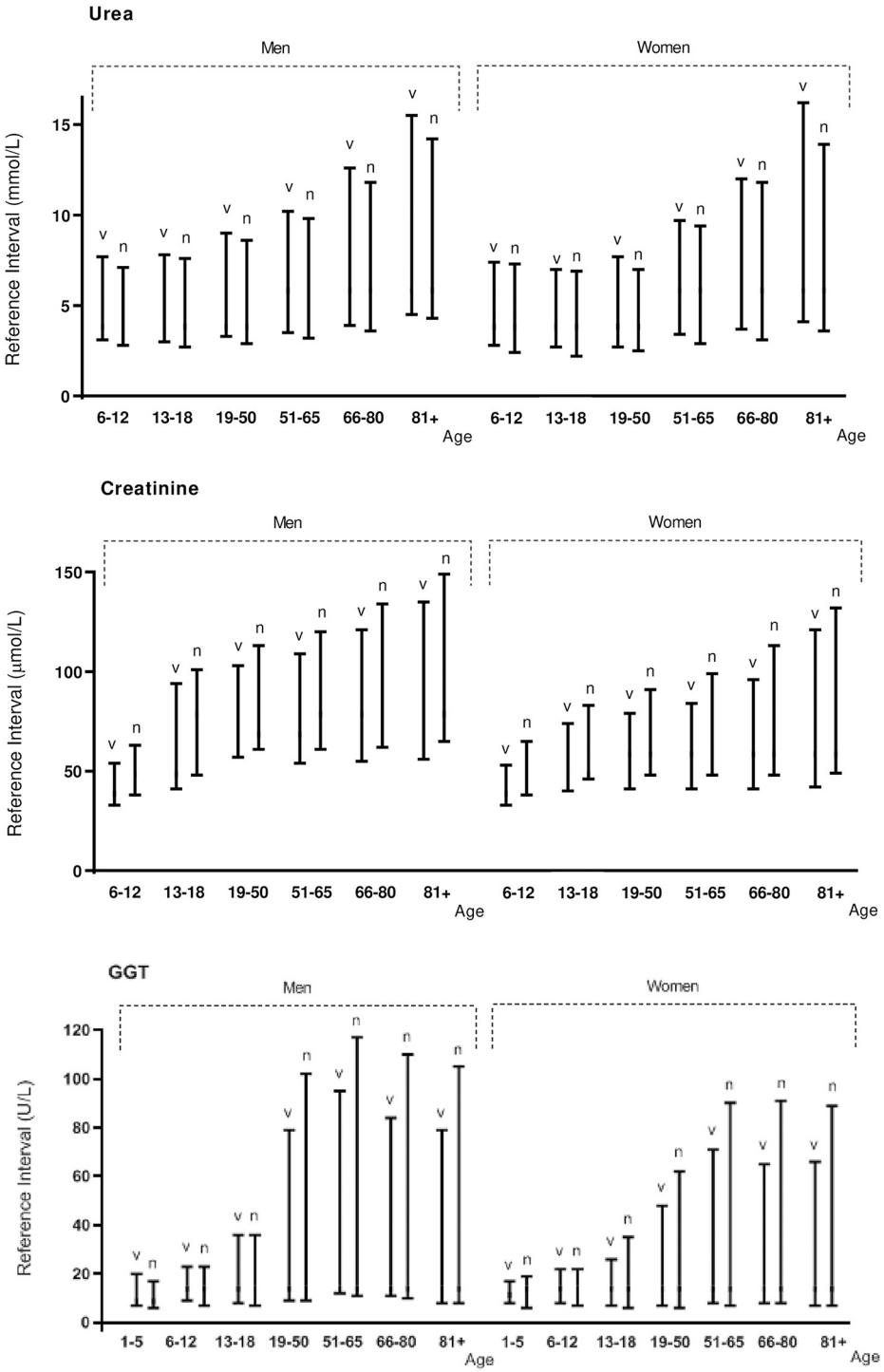


Fig 2. Urea, creatinine and GGT results. Age and sex effects on the reference intervals for creatinine, urea and GGT for Vall d'Hebron (v) and NUMBER (n).

https://doi.org/10.1371/journal.pone.0268522.g002

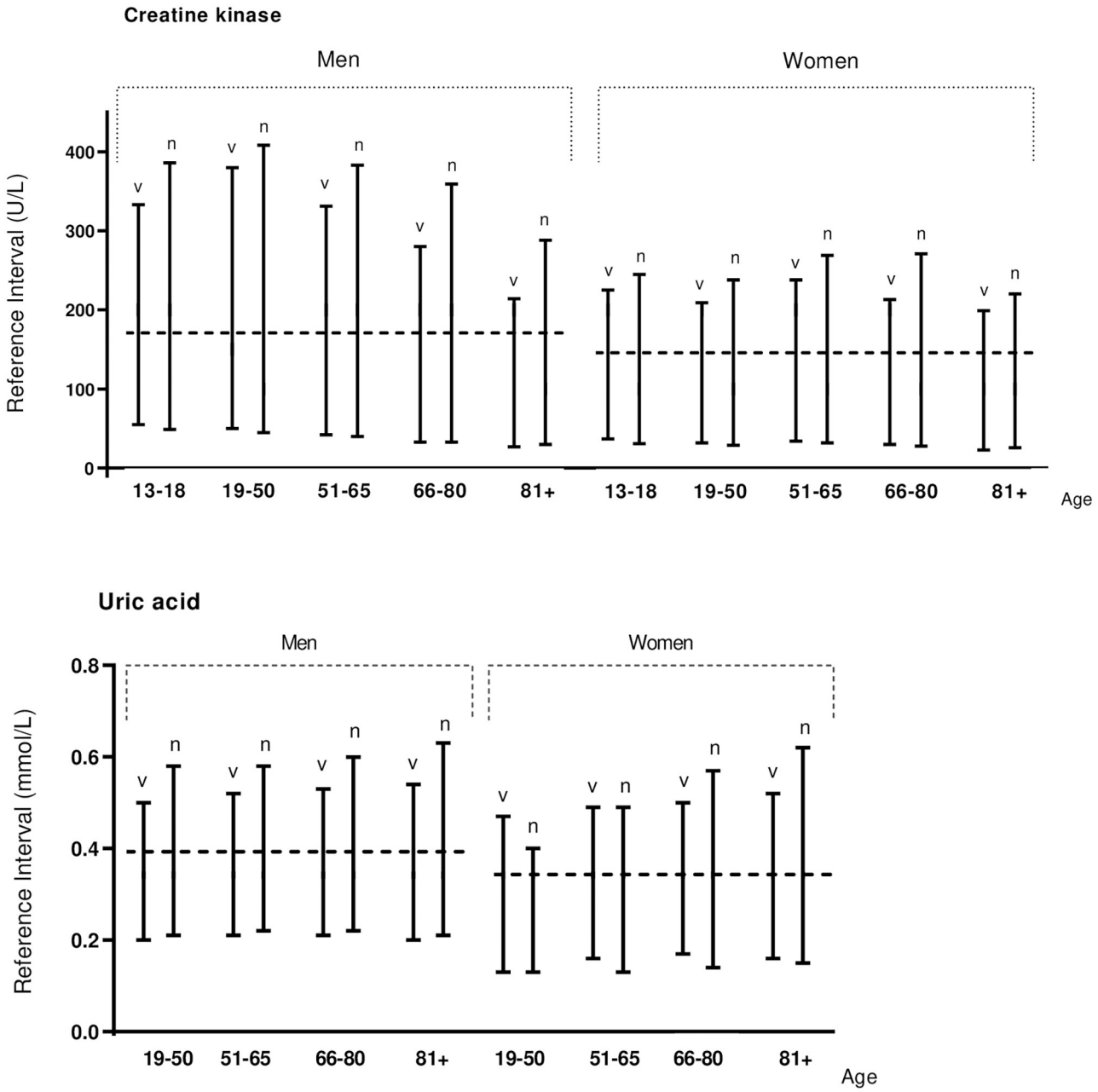


Fig 3. Creatine kinase and uric acid results. Reference intervals for creatine kinase and uric acid for Vall d'Hebron (v) and NUMBER (n), stratified for age and sex. Currently used upper reference interval in Vall d'Hebron are shown as slashed lines.

https://doi.org/10.1371/journal.pone.0268522.g003

Netherlands [14], as a result of the Dutch lifestyle and diet. The only IFCC-standardized method for liver parameters in our study was GGT and the differences for this test between Vall d'Hebron results in Barcelona and the NUMBER project could support this hypothesis.

Alcohol consumption and increased body mass index have been related with higher ALT, GGT and AST results in the population from the Nordic Reference Interval Project (NORIP) [28]. Interestingly, in 2009, Strømme and colleagues, using data from the NORIP project, showed reference intervals results for ALT in northern Europe which are similar to our Dutch

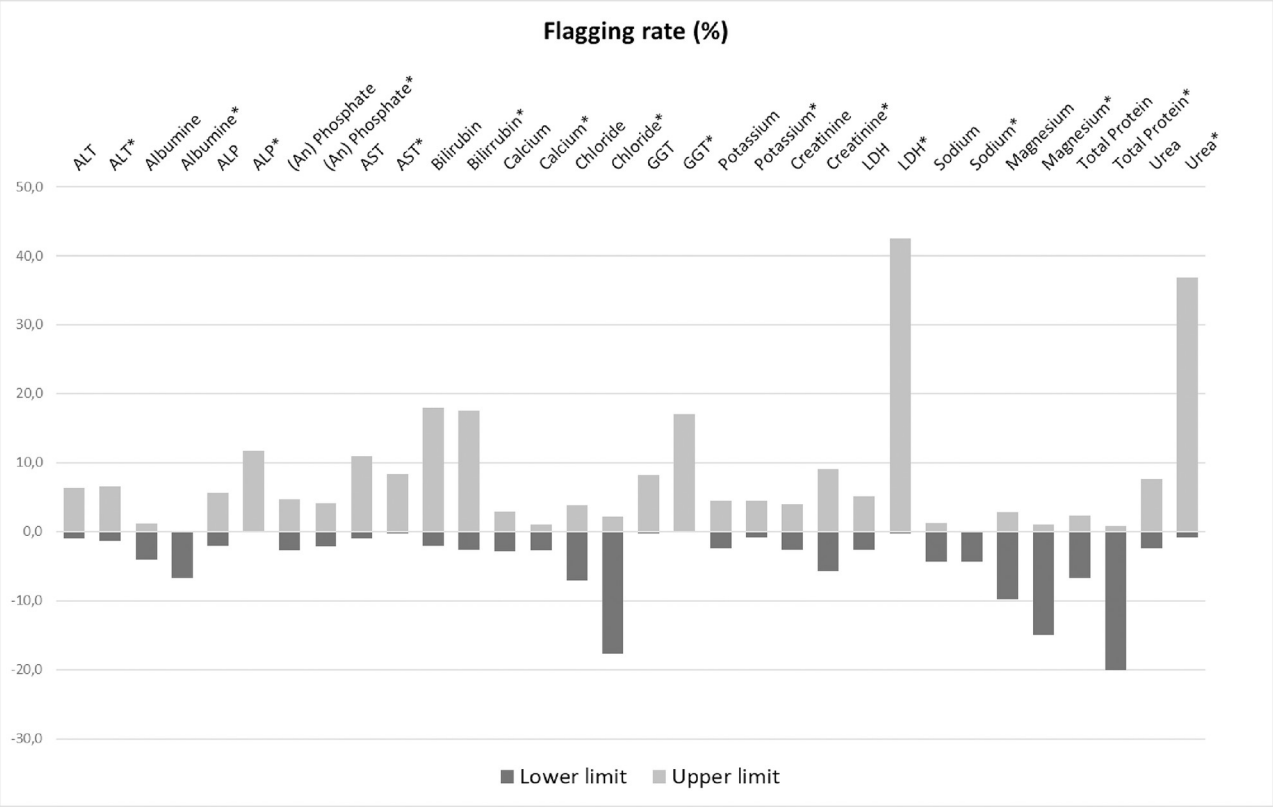


Fig 4. Flagging rates. Percentage of individuals upper or lower (represented as negative) the reference intervals, for an independent dataset (January-June 2019) for both calculated reference intervals and currently used reference intervals in Vall d'Hebron (*).

https://doi.org/10.1371/journal.pone.0268522.g004

results [29]. They already highlighted the differences observed between the Nordic reference intervals and the reference intervals calculated for the Italian population, which in their turn are similar to the calculated reference intervals in our study for the population in Vall d'Hebron [29,30]. In a recent publication, Wosniok et al. addressed these differences in calculated reference intervals from different studies for liver parameters [23]. They proposed it may be due to diverse age representations in the age groups. In order to test this hypothesis, we repeated the analyses for GGT, applying more age categories, in both the Vall d'Hebron and NUMBER datasets (S2 Fig). Since these results showed the same tendency, we consider differences in lifestyle a potential alternative hypothesis. In addition, the reference intervals for GGT are only significantly higher in the adult Dutch population (when diet or alcohol do start to play a role) and not in children, indicating a lifestyle component. The Mediterranean diet has been associated with favourable health outcomes [31], and with decreasing levels of ALT, AST and GGT in patients with non-alcoholic fatty liver disease, supporting this hypothesis [32]. It is important to remark that the reference intervals for the liver parameters that were calculated using the reference limit estimator method (S4 Table and S3 Fig) were not as high in the Vall d'Hebron population as with NUMBER method, but were still higher than the reference intervals that are now commonly applied in clinical laboratories. This supports the idea that, for skewed distributions, it is still necessary to further explore the best indirect method for references interval calculation.

For creatinine and urea, similar age distributions were found in the Vall d'Hebron outpatient sample compared to the Dutch national sample, even though the methodology for creatinine differed (Jaffe vs enzymatic, Fig 2), which support earlier studies on the age related decline in renal function [33].

Interestingly, for reasons yet unclear, in age group 19–50 years, for albumin, ALP, ALT, creatinine and urea, the resulting reference interval is usually smaller in male patients and the Prop. 2020 is always lower (<90%) when comparing to the results in female patients. No explanation was found for the significantly elevated reference intervals for CK and uric acid in the NUMBER project [14], as the calculated reference intervals were substantially higher than those currently applied in the participating laboratories. In the Vall d'Hebron sample, we confirmed the Dutch observations and also found reference intervals higher than currently used and recommended for these tests. Nevertheless, compared with the Duch results, the upper limits of the reference intervals calculated in Vall d'Hebron laboratory were lower for all age groups for both CK and uric acid (Fig 3). For CK, differences between currently used and calculated reference intervals are particularly extreme, which has been already observed in other studies [34,35]. This finding might be explained by the high incidence of some related comorbidities such as metabolic syndrome or high blood pressure [36] together with the use of statins. For uric acid, the obtained higher limits in both studies are also higher than cut-off values associated with worse progression of kidney disease [37] and higher than the cut-off defined by the solubility limit of uric acid [14].

Our analyses show important differences in flagging rates between the currently used reference intervals in Vall d'Hebron and the new calculated reference intervals in an independent dataset. In general terms, too much flagging is noted for currently used reference intervals. This highlights the need for establishing adequate reference intervals, as frequent flagging may distract attention from true pathological results [38]. In addition to that, we found, in general, higher flagging in our study compared to the Dutch NUMBER study which may be explained by the additional pre-analytical and clinical criteria used in the current study.

For some of the calculated reference intervals the confidence intervals for lower and upper limits (S3 Table) included only the reported limit, due to the large sample size, emphasizing the robustness of the presented results.

It is also important to remark that, in general, the results calculated with the NUMBER method and the Reference Limit Estimator method (S4 Table and S3 Fig) show in a great extent similar results across age group, but for a few laboratory tests there are some remarkable differences that deserve further study. Lower reference intervals were found with the Reference Limit Estimator method for GGT, creatinine and CK.

Our study has several strengths. First, compared to the direct method of establishing reference intervals, the applied automatic indirect approach is cost-efficient and avoids collecting and analysing material from healthy control donors. Second, it mimics preanalytical conditions of real samples. In addition, we had the unique opportunity to experiment with the Dutch NUMBER approach and to do head-to-head comparisons between the reference intervals obtained for the Dutch population with the reference intervals calculated in the Vall d'Hebron population for standardized tests. Lastly, results using the NUMBER method were also compared with the reference limit estimator method [24] using the same dataset.

We also acknowledge several limitations. First, since we used anonymous laboratory test results, clinical information was not available. Although we tried to select a healthy population as much as possible, test results from unhealthy persons may have been included in our datasets. Second, because of our completely anonymized databases, we did not exclude individuals visiting practitioners more than once a year leading to a possible bias. Third, structural monitoring with commutable, value-assigned trueness verifiers (type 1 EQA-materials) was not

available in 2018. However, blinded type 1 EQA materials from the Dutch SKML were used in 2020, which is essential for proving metrological traceability of results from standardized test. By using a random sampling method with a dataset from 2020 we confirmed adequate analytical performance and verified the reference intervals calculated in the 2018 dataset. The COVID-19 pandemic and the resulting differences in patient population hampered us in using a dataset from 2020 to calculate the reference intervals. Fourth, we selected one statistical method (NUMBER method) to calculate reference intervals, and compared these with the reference limit estimator method [24]. Several statistical methods have been proposed so far but no consensus or official recommendations about 'which method to use when' are available yet [4]. We recommend that, on an international level, indirect (statistical) reference interval methods are compared, in order to reach consensus on criteria to decide which statistical method should be applied for which test. Given the comparable results between studies applying indirect methods to establish reference intervals, indirect methods are a promising tool for laboratories to develop cheap, specific and updated reference intervals.

In conclusion, using an indirect approach, we determined population-specific reference intervals for 16 biochemistry tests from the Vall d'Hebron region, some being more sex and age specific than in the product inserts. Reference intervals of normally distributed biochemical tests were comparable to those found in a Dutch outpatient sample, indicating that the indirect method is an appropriate approach for deducing reference intervals. In order to verify the applicability of SI-traceable reference intervals obtained by indirect methods across outpatient populations, equivalence of test results from SI-standardizable tests must be verified thoroughly using type 1 EQA-materials. To conclude, adequate implementation of common, metrologically traceable reference intervals is the ultimate goal for guaranteeing safe and clinically effective use of medical tests, as required by the upcoming EU IVD Regulation 2017/746. As a first step, method (Beckman)- and population (Vall d'Hebron region)- specific refined reference intervals were derived for biochemistry tests.

Supporting information

S1 Fig. Daily averages plot for calcium. Daily average is represented as points, monthly average as black lines and the average of the year as red lines. Slashed lines represent biological variation from monthly (black) or yearly (red) average and were used as an indication for person to person variation. Decisions about quality stability were made by visual inspection of the plots.
(PDF)

S2 Fig. GGT reference interval results by age. Different age representation for the calculated reference intervals for ALT and GGT for Vall d'Hebron (V) and NUMBER (N).
(PDF)

S3 Fig. Comparison between indirect reference intervals using two methods. NUMBER method and reference limit estimator (RLE) method. Representation of reference intervals from S4 Table were made just when the number of data per both methods were higher than 500. *Reference interval results calculated with less data than the recommended by the RLE method (4.000).
(PDF)

S1 Table. Methods principles and metrological traceability of general clinical chemistry tests used in Vall d'Hebron for determining reference intervals. LOINC codes for international units are also shown in the table.
(PDF)

S2 Table. Corresponding approximate serum concentrations of intralipid, bilirubin and free hemoglobin for the 0–5 scale for indices.
(PDF)

S3 Table. Calculated reference intervals using NUMBER method presented together with the 90% confidence interval.
(PDF)

S4 Table. Comparison of indirect reference intervals using the NUMBER method and the reference limit estimator (RLE) method. Confidence intervals are presented for the RLE method. Results are presented in calculated and international units.
(PDF)

Acknowledgments

We would like to thank Dr. Farhad Arzideh for his advice and assistance in the use of reference limit estimator.

Author Contributions

Conceptualization: Nannette Brouwer.

Data curation: Albert Blanco-Grau, Yolanda Villena-Ortiz, Wendy P. J. den Elzen.

Formal analysis: Luisa Martinez-Sanchez, Raymond Noordam, Albert Blanco-Grau, Yolanda Villena-Ortiz, Wendy P. J. den Elzen.

Investigation: Luisa Martinez-Sanchez, Yolanda Villena-Ortiz, Wendy P. J. den Elzen.

Methodology: Christa M. Cobbaert, Raymond Noordam, Nannette Brouwer, Marc Thelen, Wendy P. J. den Elzen.

Resources: Christa M. Cobbaert, Marc Thelen, Roser Ferrer-Costa, Ernesto Casis, Francisco Rodríguez-Frias.

Supervision: Christa M. Cobbaert, Nannette Brouwer, Francisco Rodríguez-Frias, Wendy P. J. den Elzen.

Validation: Luisa Martinez-Sanchez, Albert Blanco-Grau.

Writing – original draft: Luisa Martinez-Sanchez, Wendy P. J. den Elzen.

Writing – review & editing: Luisa Martinez-Sanchez, Christa M. Cobbaert, Raymond Noordam, Nannette Brouwer, Marc Thelen, Roser Ferrer-Costa, Ernesto Casis, Francisco Rodríguez-Frias, Wendy P. J. den Elzen.

References

1. Koerbin G, Sikaris KA, Jones GRD, Ryan J, Reed M, Tate J. Evidence-based approach to harmonised reference intervals. Clin Chim Acta 2014; 432:99–107. https://doi.org/10.1016/j.cca.2013.10.021 PMID: 24183842
2. Gary LH, Sousan A, James CBM, Ceriotti F, Garg U, Horn P, et al. EP28-A3c: Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition. Clin Lab Stand Inst 2010; 28:30.
3. Ozarda Y, Ichihara K, Barth JH, Klee G. Protocol and standard operating procedures for common use in a worldwide multicenter study on reference values. Clin Chem Lab Med 2013; 51:1027–40. https://doi.org/10.1515/cclm-2013-0249 PMID: 23633469
4. Martinez-Sanchez L, Marques-Garcia F, Ozarda Y, Blanco A, Brouwer N, Canalias F et al. Big data and reference intervals: rationale, current practices, harmonization and standardization prerequisites and

future perspectives of indirect determination of reference intervals using routine data Adv Lab Med 2021; 2:9–16.

5. Gräsbeck R, Siest G, Wilding P, Williams GZ, Whitehead TP. Provisional recommendation on the theory of reference values (1978). Part 1. The concept of reference values. Clin Chem 1979; 25:1506–8. PMID: 455695
6. Seaberg RS, Stallone RO, Statland BE. The role of total laboratory automation in a consolidated laboratory network. Clin Chem 2000; 46:751–6. PMID: 10794773
7. Berg J, Lane V. Pathology Harmony; a pragmatic and scientific approach to unfounded variation in the clinical laboratory. Ann Clin Biochem 2011; 48:195–7. https://doi.org/10.1258/acb.2011.011078 PMID: 21555538
8. Tate JR, Sikaris KA, Jones GRD, Yen T, Koerbin G, Ryan J, et al. Harmonising adult and paediatric reference intervals in Australia and New Zealand: an evidence-based approach for establishing a first panel of chemistry analytes. Clin Biochem Rev 2014; 354:213–35. PMID: 25678727
9. Flegar-Meštrić Z, Perkov S, Radeljak A. Standardization in laboratory medicine: Adoption of common reference intervals to the Croatian population. World J Methodol 2016; 6:93–100. https://doi.org/10.5662/wjm.v6.i1.93 PMID: 27019800
10. Borai A, Ichihara K, Al Masaud A, Tamimi W, Bahijri S, Armbuster D, et al. Establishment of reference intervals of clinical chemistry analytes for the adult population in Saudi Arabia: a study conducted as a part of the IFCC global study on reference values. Clin Chem Lab Med 2016; 54:843–55. https://doi.org/10.1515/cclm-2015-0490 PMID: 26527074
11. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Mårtensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. Scand J Clin Lab Invest 2004; 64:271–84. https://doi.org/10.1080/00365510410006324 PMID: 15223694
12. Evgina S, Ichihara K, Ruzhanskaya A, Skibo I, Vybornova N, Vasiliev A, et al. Establishing reference intervals for major biochemical analytes for the Russian population: a research conducted as a part of the IFCC global study on reference values. Clin Biochem 2020; 81:41–58.
13. Parker ML, Adeli K. Pediatric and adult reference interval harmonization in Canada: an update. Clin Chem Lab Med 2018; 57:57–60. https://doi.org/10.1515/cclm-2017-0965 PMID: 29303773
14. Den Elzen WPJ, Brouwer N, Thelen MH, Le Cessie S, Haagen IA, Cobbaert CM. NUMBER: Standardized reference intervals in the Netherlands using a "big data" approach. Clin Chem Lab Med 2018; 57:42–56. https://doi.org/10.1515/cclm-2018-0462 PMID: 30218599
15. Ricós C, Fernandez-Calle P, Marqués F, Minchinela J, Salas A, Martínez-Bru C, et al. Impact of implementing a category 1 external quality assurance scheme for monitoring harmonization of clinical laboratories in Spain. Adv Lab Med 2020; 1:20200008.
16. Cornes M, van Dongen-Lases E, Grankvist K, Ibarz M, G, Lippi G, et al. Order of blood draw: Opinion Paper by the European Federation for Clinical Chemistry and Laboratory Medicine (EFLM) Working Group for the Preanalytical Phase (WG-PRE). Clin Chem Lab Med 2017; 55:27–31. https://doi.org/10.1515/cclm-2016-0426 PMID: 27444170
17. Jansen RTP, Cobbaert CM, Weykamp C, Thelen M. The quest for equivalence of test results: the pilgrimage of the Dutch Calibration 2.000 program for metrological traceability. Clin Chem Lab Med 2018; 56:1673–84. https://doi.org/10.1515/cclm-2017-0796 PMID: 29341939
18. Thelen M, Jansen R, Weykamp C, Steigstra H, Meijer R, Cobbaert CM. Expressing analytical performance from multi-sample evaluation in laboratory EQA. Clin Chem Lab Med 2017; 55:1509–16. https://doi.org/10.1515/cclm-2016-0970 PMID: 28182577
19. Noakes TD. Effect of Exercise on Serum Enzyme Activities in Humans. Sports Med 1987; 4:245–67. https://doi.org/10.2165/00007256-198704040-00003 PMID: 3306866
20. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing 2018; Viena, Austria.
21. Ichihara K, Boyd JC on behalf of the IFCC Committee on Reference Intervals and Decision Limits (C-RIDL). An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010; 48:1537–51. https://doi.org/10.1515/CCLM.2010.319 PMID: 21062226
22. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 2012; 10:486–9. https://doi.org/10.5812/ijem.3505 PMID: 23843808
23. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. Clin Chem Lab Med 2019; 57:1933–47. https://doi.org/10.1515/cclm-2018-1341 PMID: 31271548
24. Arzideh F, Wosniok W, Gurr E, Hinsch W, Schumann G, Weinstock N, et al. A plea for intra-laboratory reference limits. Part 2. A bimodal retrospective concept for determining reference limits from intra-

laboratory databases demonstrated by catalytic activity concentrations of enzymes. Clin Chem Lab Med 2007; 45:1043–57. https://doi.org/10.1515/CCLM.2007.250 PMID: 17867994

25. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: a systematic review. Int J Med Inform 2018; 114:57–65. https://doi.org/10.1016/j.ijmedinf.2018.03.013 PMID: 29673604

26. Jones GRD, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, et al. Indirect methods for reference interval determination—review and recommendations. Clin Chem Lab Med 2018; 57:20–29. https://doi.org/10.1515/cclm-2018-0073 PMID: 29672266

27. Adeli K, Higgins V, Nieuwesteeg M, Raizman JE, Chen Y, Wong SL, et al. Biochemical marker reference values across pediatric, adult, and geriatric ages: establishment of robust pediatric and adult reference intervals on the basis of the Canadian Health Measures Survey. Clin Chem 2015; 61:1049–62. https://doi.org/10.1373/clinchem.2015.240515 PMID: 26044506

28. Alatalo P, Koivisto H, Kultti J, Bloigu R, Niemelä O. Evaluation of reference intervals for biomarkers sensitive to alcohol consumption, excess body weight and oxidative stress. Scand J Clin Lab Invest 2010; 70:104–11. https://doi.org/10.3109/00365510903548818 PMID: 20073674

29. Strømme JH, Rustad P, Steensland H, Theodorsen L, Urdal P. Reference intervals for eight enzymes in blood of adult females and males measured in accordance with the International Federation of Clinical Chemistry reference system at 37 degrees C: part of the Nordic Reference Interval Project. Scand J Clin Lab Invest 2004; 64:371–84. https://doi.org/10.1080/00365510410002742 PMID: 15223701

30. Prati D, Taioli E, Zanella A, Della Torre E, Butelli S, Del Vecchio E, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. Ann Intern Med 2002; 137:1–10. https://doi.org/10.7326/0003-4819-137-1-200207020-00006 PMID: 12093239

31. Galbete C, Kröger J, Jannasch F, Iqbal K, Schwingshackl L, Schwedhelm C, et al. Nordic diet, Mediterranean diet, and the risk of chronic diseases: the EPIC-Potsdam study. BMC Med 2018; 16:99. https://doi.org/10.1186/s12916-018-1082-y PMID: 29945632

32. Biolato M, Manca F, Marrone G, Cefalo C, Racco S, Miggiano GA, et al. Intestinal permeability after Mediterranean diet and low-fat diet in non-alcoholic fatty liver disease. World J Gastroenterol 2019; 25:509–20. https://doi.org/10.3748/wjg.v25.i4.509 PMID: 30700946

33. Denic A, Glassock RJ, Rule AD. Structural and functional changes with the aging kidney. Adv Chronic Kidney Dis 2016; 23:19–28. https://doi.org/10.1053/j.ackd.2015.08.004 PMID: 26709059

34. Lilleng H, Johnsen SH, Wilsgaard T, Bekkelund SI. Are the currently used reference intervals for creatine kinase (CK) reflecting the general population? The Tromsø Study. Clin Chem Lab Med 2011; 50:879–84. https://doi.org/10.1515/CCLM.2011.776 PMID: 22070220

35. Capasso M, De Angelis MV, Di Muzio A, Uncini A. Caveats in determining reference intervals for serum creatine kinase. Am Heart J 2008; 155:e5. https://doi.org/10.1016/j.ahj.2007.11.024 PMID: 18215582

36. Brewster LM, Mairuhu G, Bindraban NR, Koopmans RP, Clarck JF, Montfrans GA. Creatine kinase is associated with blood pressure. Circulation 2006; 114:2034–9. https://doi.org/10.1161/CIRCULATIONAHA.105.584490 PMID: 17075013

37. Tsai CW, Lin SY, Kuo CC, Huang CC. Serum uric acid and progression of kidney disease: a longitudinal analysis and mini-review. PLoS One 2017; 12:e0170393. https://doi.org/10.1371/journal.pone.0170393 PMID: 28107415

38. Tate JR, Koerbin G, Adeli K. Opinion paper: deriving harmonized reference intervals–Global activities. EJIFCC 2016; 27:48–65. PMID: 27683506

3.3. Supplementary material

S1 Table. Methods principles and metrological traceability of general clinical chemistry tests used in Vall d'Hebron for determining reference intervals. LOINC codes for international units are also shown in the table.

Test (LOINC*)	Traceability	Method short description	Catalogue Number	Currently used and recommended reference intervals (adults)
Albumin (61151-7)	CRM 470	bromocesol green	OSR6202	35 – 52 g/L
Calcium (2000-8)	NIST SRM 909bL1	reaction with arsenazo III	OSR61117	2.20 – 2.65 mmol/L
Creatinine (14682-9)	NIST SRM 967 L1	Jaffé method IDMS traceable (compensated)	OSR6178	M: 59 – 104 µmol/L F: 45 – 84 µmol/L
Lactate Dehydrogenase (14805-6)	Beckman Master Cal*	pyruvate to lactate method	OSR6126	208 – 378 U/L
Magnesium (2601-3)	NIST SRM 909b L2	direct method with xylidyl blue in basic reaction	OSR6189	M: 0.73 – 1.06 mmol/L F: 0.77 – 1.03 mmol/L
(Anorganic) Phosphate (14879-1)	Beckman Master Cal*	reaction with molybdate	OSR6222	0.81 – 1.45 mmol/L
Total Bilirubin (14631-6)	NIST SRM 916a	reaction with 3,5-tetrafluoroborato de diclorofenilodiazonio stabilized by diazonium salt	OSR6212	5 – 21 µmol/L
Total Protein (2885-2)	NIST SRM 927c	copper reaction in basic solution (Biuret)	OSR6232	66 – 83 g/L
Uric Acid (14933-6)	Isotope Dilution Mass Spectrometry	uricase total reaction to form alantoine and hydrogen peroxide and coupled Trinder reaction for hydrogen peroxide determination	OSR6298	M : 208.3 – 428.4 µmol/L F: 154.7 – 357.0 µmol/L
Urea (22664-7)	NIST SRM 909b L1	indirect method with urease reaction to form ammonium ion and carbonate and coupled reaction with glutamate dehydrogenase for ammonium determination	OSR6234	2.8 – 7.2 mmol/L
Chloride (2075-0)	NIST SRM 919	indirect ion selective electrodes	A28937 & A28945	98 – 107 mmol/L
Potassium (2823-3)	NIST SRM 918	indirect ion selective electrodes	A28937 & A28945	3.5 – 5.1 mmol/L
Sodium (2951-2)	NIST SRM 919	indirect ion selective electrodes	A28937 & A28945	136 – 145 mmol/L

Alkaline Phosphatase (6768-6)	Beckman Master Cal**	IFCC recommended method	OSR6004	M : 43 – 115 U/L F: 33 – 98 U/L
Alanine Aminotransferase (1744-2)	Beckman Master Cal**	IFCC recommended method without pyridoxyl 5 phosphate	OSR6107	M: <50 U/L F: <35 U/L
Aspartate Aminotransferase (88112-8)	Beckman Master Cal**	IFCC recommended method without pyridoxyl 5 phosphate	OSR6209	M: <50 U/L F: <35 U/L
Gamma-Glutamyltransferase (2324-2)	IFCC reference method	IFCC recommended method	OSR6120	M: <55 U/L F: <38 U/L
Creatine Kinase (2157-6)	IFCC reference method	IFCC recommended method	OSR6279	M: <171 U/L F: <145 U/L

* LOINC codes are indicated for the units presented in the same table (international units)

** Beckman Coulter system calibrator catalogue number 66300 with values determined by Beckman Coulter selected measurement procedure.

S2Table. Corresponding approximate serum concentrations of intralipid, bilirubin and free hemoglobin for the 0–5 scale for indices.

Scale	Intralipid mg/dL	Bilirubin mg/ dL	Hemoglobin mg/dL
1	40 – 99	1.4 – 4.9	50-99
2	100-199	5.0 – 9.9	100-199
3	200-299	10.0 – 19.9	200-299
4	300-500	20 - 40	300-500
5	>500	> 40	>500

S3 Table. Calculated reference intervals using NUMBER method presented together with the 90% confidence interval.

Test	Unit	Sex	Age. years	n	Calculated reference intervals:			
					Low	90% CI	High	90% CI
Albumin	g/dL	M	1-5	330	3.8	(3.76-3.84)	4.9	(4.86-4.94)
			6-18	914	4.1	(4.08-4.12)	5.0	(4.98-5.02)
			19-50	4281	4.0	(3.99-4.01)	5.1	(5.09-5.11)
			51-65	3660	3.8	(3.79-3.81)	4.9	(4.89-4.91)
			66-80	4704	3.5	(3.49-3.51)	4.9	(4.89-4.91)
			80+	4299	3.2	(3.18-3.22)	4.7	(4.68-4.72)
		F	1-5	282	3.9	(3.86-3.94)	4.9	(4.86-4.94)
			6-18	1116	4.0	(3.98-4.02)	5.0	(4.98-5.02)
			19-50	6622	3.7	(3.69-3.71)	4.9	(4.89-4.91)
			51-65	5966	3.8	(3.79-3.81)	4.8	(4.79-4.81)

			66-80	7865	3.6	(3.59-3.61)	4.7	(4.69-4.71)
			80+	10767	3.2	(3.19-3.21)	4.6	(4.59-4.61)
ALP	U/L		13-18	381	74	(68.8-79.2)	218	(212.8-223.2)
			19-50	11577	46	(45.4-46.6)	133	(132.4-133.6)
		M	51-65	9928	45	(44.4-45.6)	135	(134.4-135.6)
			66-80	9520	44	(43.3-44.7)	137	(136.3-137.7)
			80+	4569	46	(44.9-47.1)	155	(153.9-156.1)
			13-18	880	50	(46.8-53.2)	184	(180.8-187.2)
			19-50	15757	39	(38.5-39.5)	130	(129.5-130.5)
		F	51-65	14428	49	(48.4-49.6)	152	(151.4-152.6)
			66-80	15285	47	(46.4-47.6)	147	(146.4-147.6)
			80+	10185	46	(45.2-46.8)	157	(156.2-157.8)
ALT	U/L		1-12	2391	9	(8.7-9.3)	32	(31.7-32.3)
			13-18	2978	8	(7.6-8.4)	38	(37.6-38.4)
			19-50	45779	10	(9.9-10.2)	55	(54.9-55.2)
		M	51-65	37916	11	(10.9-11.1)	51	(50.9-51.1)
			66-80	42390	9	(8.9-9.1)	43	(42.9-43.1)
			80+	21014	7	(6.9-7.1)	34	(33.9-34.1)
			1-12	2308	9	(8.7-9.3)	31	(30.7-31.3)
			13-18	4748	7	(6.8-7.2)	27	(26.8-27.2)
			19-50	77714	7	(6.9-7.1)	35	(34.9-35.1)
		F	51-65	52710	9	(8.9-9.1)	42	(41.9-42.1)
			66-80	60878	8	(7.9-8.1)	36	(35.9-36.1)
			80+	4353	6	(5.8-6.2)	29	(28.8-29.2)
AST	U/L		1-5	557	25	(24.2-25.8)	51	(50.2-51.8)
			6-12	921	20	(19.5-20.5)	42	(41.5-42.5)
		M	13-18	1262	15	(14.6-15.5)	38	(37.6-38.5)
			19+	50953	13	(12.9-13.1)	38	(37.9-38.1)
			1-5	407	26	(25.1-26.9)	51	(50.1-51.9)
			6-12	1035	18	(17.5-18.5)	42	(41.5-42.5)
		F	13-18	1980	13	(12.7-13.3)	30	(29.7-30.3)
			19+	78383	13	(12.9-13.1)	36	(35.9-36.1)
Bilirubin (total)	mg/dL (µmol/L)		6-12	123	0.23	(0.19-0.27)	0.84	(0.8-0.88)
		M	13-18	301	0.29	(0.25-0.33)	1.34	(1.30-1.38)
			19+	16323	0.32	(0.31-0.33)	1.3	(1.29-1.31)
			6-12	573	0.23	(0.20-0.26)	1.1	(1.07-1.13)
		F	19+	24215	0.28	(0.28-0.28)	1.04	(1.04-1.04)
Calcium	mg/dL (mmol/L)		1-5	318	9.2	(9.14-9.26)	10.7	(10.64-10.76)
		M + F	6-12	954	9.3	(9.27-9.33)	10.5	(10.47-10.53)
			13-18	1358	9.2	(9.18-9.22)	10.5	(10.48-10.52)

			19+	46602	8.8	(8.80-8.80)	10.3	(10.30-10.30)
Chloride	mmol/L	M + F		784	98	(97.8-98.3)	108	(107.8-108.3)
Creatinine	mg/dL (µmol/L)		6-12	1317	0.37	(0.37-0.37)	0.61	(0.61-0.61)
		M	13-18	3517	0.47	(0.46-0.48)	1.07	(1.06-1.08)
			19-50	53345	0.65	(0.65-0.65)	1.17	(1.17-1.17)
			51-65	44666	0.62	(0.62-0.62)	1.23	(1.23-1.23)
			66-80	48705	0.62	(0.62-0.62)	1.36	(1.36-1.36)
			80+	22032	0.63	(0.63-0.63)	1.53	(1.53-1.53)
		F	6-12	1364	0.37	(0.37-0.37)	0.59	(0.59-0.59)
			13-18	4804	0.45	(0.45-0.45)	0.83	(0.83-0.83)
			19-50	80957	0.47	(0.47-0.47)	0.9	(0.90-0.90)
			51-65	56832	0.47	(0.47-0.47)	0.95	(0.95-0.95)
			66-80	67350	0.46	(0.46-0.46)	1.09	(1.09-1.09)
			80+	47750	0.48	(0.48-0.48)	1.37	(1.37-1.37)
GGT	U/L		1-5	174	7	(6.3-7.7)	20	(19.3-20.7)
			6-12	322	9	(8.5-9.6)	23	(22.5-23.6)
			13-18	1452	8	(7.5-8.5)	36	(35.5-36.5)
		M	19-50	31582	9	(8.7-9.3)	79	(78.7-79.3)
			51-65	26758	12	(11.6-12.4)	95	(94.6-95.4)
			66-80	28080	11	(10.7-11.3)	84	(83.7-84.3)
			80+	13160	8	(7.6-8.4)	79	(78.6-79.4)
			1-5	146	8	(7.5-8.5)	17	(16.5-17.5)
			6-12	341	8	(7.5-8.5)	22	(21.5-22.5)
			13-18	2191	7	(6.7-7.3)	26	(25.7-26.3)
		F	19-50	48040	7	(6.9-7.1)	48	(47.9-48.1)
			51-65	35997	8	(7.8-8.2)	71	(70.8-71.2)
			66-80	40412	8	(7.8-8.2)	65	(64.8-65.2)
			80+	27174	7	(6.8-7.3)	66	(65.8-66.3)
LDH	U/L		6-12	257	359	(346.6-371.5)	643	(630.6-655.5)
			13-18	340	274	(264.2-283.8)	531	(521.2-540.8)
			19-50	2573	256	(252.5-259.5)	507	(503.5-510.5)
		M + F	51-65	1963	274	(269.9-278.1)	534	(529.9-538.1)
			66-80	2039	270	(265.6-274.4)	551	(546.6-555.4)
			80+	1539	266	(260.3-271.7)	584	(578.3-589.7)
Magnesium	mg/dL (mmol/L)	M + F		4571	1.8	(1.79-1.81)	2.4	(2.39-2.41)

Phosphate	mg/dL (mmol/L)		1-5	147	4.2	(4.13-4.27)	5.4	(5.33-5.47)
		M	6-12	405	4.2	(4.16-4.24)	5.3	(5.26-5.34)
			13-18	390	3.6	(3.54-3.66)	5.4	(5.34-5.46)
			19-50	3132	2.4	(2.37-2.43)	4.7	(4.67-4.73)
			51-65	2917	2.2	(2.17-2.23)	4.3	(4.27-4.33)
			66+	6640	2.2	(2.18-2.22)	4.2	(4.18-4.22)
		F	1-5	126	4.3	(4.23-4.37)	5.4	(5.33-5.47)
			6-12	428	4.1	(4.06-4.14)	5.3	(5.26-5.34)
			13-18	729	3.5	(3.46-3.54)	5.2	(5.16-5.24)
			19-50	5874	2.6	(2.58-2.62)	4.7	(4.68-4.72)
			51-65	7970	2.7	(2.68-2.72)	4.7	(4.68-4.72)
			66+	18548	2.6	(2.59-2.61)	4.5	(4.49-4.51)
Potassium	mmol/L	M + F		257189	3.6	(3.60-3.60)	5.1	(5.09-5.09)
Sodium	mmol/L	M + F		256775	136	(136.0-136.0)	144	(144.0-144.0)
Total protein	g/dL (g/L)	M + F		35141	6.1	(6.09-6.11)	8	(7.99-8.01)
Urea	mg/dL (mmol/L)		1-5	227	15	(13.6-16.4)	45	(43.6-46.4)
		M	6-12	755	19	(18.3-19.7)	47	(46.3-47.7)
			13-18	996	18	(17.4-18.6)	47	(46.4-47.6)
			19-50	4709	20	(19.7-20.4)	54	(53.7-54.4)
			51-65	4457	21	(20.6-21.4)	61	(60.6-61.4)
			66-80	5680	23	(22.5-23.5)	75	(74.5-75.5)
			80+	3662	27	(26.2-27.8)	93	(92.2-93.8)
		F	1-5	167	16	(14.5-17.5)	43	(41.5-44.5)
			6-12	767	17	(16.3-17.7)	44	(43.3-44.7)
			13-18	1220	16	(15.5-16.5)	42	(41.5-42.5)
			19-50	6850	16	(15.8-16.3)	46	(45.8-46.3)
			51-65	5524	20	(19.6-20.4)	58	(57.6-58.4)
			66-80	6986	22	(21.6-22.4)	72	(71.6-72.4)
			80+	7668	25	(24.4-25.6)	97	(96.4-97.6)

S4 Table. Comparison of indirect reference intervals using the NUMBER method and the reference limit estimator (RLE) method. Confidence intervals are presented for the RLE method. Results are presented in calculated and international units.

				NUMBER			Reference limit estimator method				
Test	Unit	Sex	Age	n	Low limit	High limit	n	Low limit	90%CI	High limit	90%CI
Albumin	g/dL (g/L)	M	18-50	4281	4.0 (40)	5.1 (51)	4352	4.0 (40)	(3.89-4.13)	5.0 (50)	(4.89-5.19)
			51-65	3660	3.8 (38)	4.9 (49)	3684	3.8 (38)	(3.70-3.96)	5.0 (50)	(4.79-5.11)
			66-80	4704	3.5 (35)	4.9 (49)	4806	3.5 (35)	(3.38-3.66)	5.1 (51)	(4.90-5.28)
			80+	4299	3.2 (32)	4.7 (47)	4721	3.2 (32)	(3.03-3.29)	4.8 (48)	(4.56-4.94)
		F	19-50	6622	3.7 (37)	4.9 (49)	6640	3.8 (38)	(3.66-3.9)	4.8 (48)	(4.69-4.99)
			51-65	5966	3.8 (38)	4.8 (48)	5960	3.8 (38)	(3.70-3.94)	4.8 (48)	(4.61-4.89)
			66-80	7865	3.6 (36)	4.7 (47)	7933	3.7 (37)	(3.56-3.8)	4.8 (48)	(4.60-4.90)
			80+	10767	3.2 (32)	4.6 (46)	11805	3.3 (33)	(3.19-3.45)	4.6 (46)	(4.47-4.81)
ALP	U/L	M	13-18	381	74	218	935	74	(68.3-78.7)	190	(178.5-200.5)
			19-50	11577	46	133	17401	47	(43.3-50.3)	133	(125.0-141.2)
			51-65	9928	45	135	14449	46	(42.9-49.9)	133	(125.0-141.2)
			66-80	9520	44	137	12880	45	(41.7-48.5)	128	(119.7-135.3)
			80+	4569	46	155	6056	46	(42.3-49.7)	147	(137.2-156.0)
		F	13-18	880	50	184	1126	52	(48.8-55.0)	108	(102.3-113.5)
			19-50	15757	39	130	19451	39	(36.1-42.3)	122	(114.0-129.4)
			51-65	14428	49	152	18368	51	(47.1-55.1)	151	(141.9-160.7)
			66-80	15285	47	147	18705	49	(45.0-52.6)	143	(133.9-151.5)
			80+	10185	46	157	12627	47	(42.8-50.4)	148	(138.6-157.4)

ALT	U/L	M	19-50	45779	10	55	54216	10	(9.3-11.3)	47	(43.4-50.2)
			51-65	37916	11	51	43602	11	(10.0-12.2)	50	(46.8-54.0)
			66-80	42390	9	43	46988	10	(8.8-10.6)	39	(36.6-42.0)
			80+	21014	7	34	23185	7	(6.6-8.0)	30	(27.5-31.7)
		F	19-50	77714	7	35	82790	8	(6.9-8.3)	27	(25.2-28.8)
			51-65	52710	9	42	56864	9	(8.6-10.2)	36	(33.1-37.9)
			66-80	60878	8	36	64624	9	(8.4-10.0)	30	(28.0-31.8)
			80+	4353	6	29	46539	7	(6.0-7.2)	24	(22.8-26.0)
AST	U/L	M	1-5	557	25	51	841	24	(22.7-25.9)	56	(52.5-58.5)
			6-12	921	20	42	1401	19	(18.3-20.5)	39	(37.1-41.1)
			13-18	1262	15	38	1770	16	(14.9-16.9)	35	(32.8-36.6)
			19+	50953	13	38	65475	14	(13.0-15.0)	37	(34.8-39.2)
		F	1-5	407	26	51	655	25	(24.1-26.7)	44	(41.8-45.8)
			6-12	1035	18	42	1505	18	(16.4-18.6)	37	(35.2-39.2)
			13-18	1980	13	30	2206	14	(12.7-14.3)	28	(26.6-29.6)
			19+	78383	13	36	89705	13	(12.1-13.9)	33	(30.9-34.7)
Bilirubin (total)	mg/dL (µmol/L)	M	6-12	123	0.23 (4)	0.84 (14)	291	0.23 (4)	(0.21-0.24)	0.44 (7)	(0.41-0.46)
			13-18	301	0.29 (5)	1.34 (23)	452	0.19 (3)	(0.17-0.21)	0.88 (15)	(0.82-0.95)
			19+	16323	0.32 (6)	1.30 (22)	19086	0.31 (5)	(0.28-0.34)	1.26 (22)	(1.17-1.35)
		F	6-18	573	0.23 (4)	1.10 (19)	786	0.21 (4)	(0.19-0.23)	0.93 (16)	(0.86-1.00)
			19+	24215	0.28 (5)	1.04 (18)	26404	0.28 (5)	(0.26-0.30)	0.95 (16)	(0.89-1.01)
Calcium	mg/dL (mmol/L)	M + F	1-5	318	9.2 (2.29)	10.7 (2.67)	951	9.4 (2.33)	(9.12-9.58)	10.9 (2.71)	(10.61-11.13)
			6-12	954	9.3 (2.32)	10.5 (2.63)	1686	9.4 (2.35)	(9.22-9.58)	10.4 (2.59)	(10.18-10.58)
			13-18	1358	9.2 (2.29)	10.5 (2.61)	1629	9.2 (2.31)	(9.03-9.45)	10.5 (2.62)	(10.27-10.73)
			19+	46602	8.8 (2.20)	10.3 (2.58)	60579	8.9 (2.21)	(8.65-9.09)	10.3 (2.57)	(10.05-10.53)

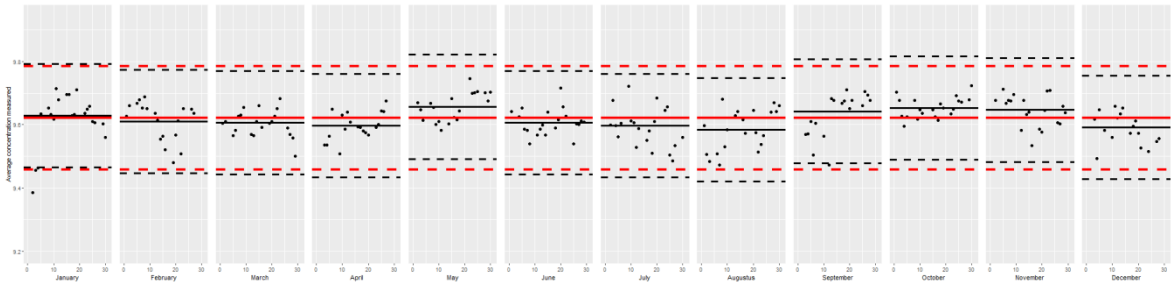
Chloride	mmol/L	M + F		784	98	108	883	99	(97.2-100.8)	108	(106.1 - 109.9)
Creatinine	mg/dL (µmol/L)	M	19-50	53345	0.65 (57)	1.17 (103)	53215	0,65 (57)	(0.61-0.68)	1.12 (99)	(1.07-1.17)
			51-65	44666	0.62 (54)	1.23 (109)	44921	0.63 (56)	(0.60-0.67)	1.17 (103)	(1.11-1.22)
			66-80	48705	0.62 (55)	1.36 (121)	50515	0.64 (57)	(0.60-0.68)	1.28 (114)	(1.22-1.35)
			80+	22032	0.63 (56)	1.53 (135)	25461	0.63 (55)	(0.58-0.67)	1.53 (136)	(1.45-1.62)
		F	19-50	80957	0.47 (41)	0.90 (79)	81069	0.46 (41)	(0.44-0.49)	0.86 (76)	(0.81-0.90)
			51-65	56832	0.47 (41)	0.95 (84)	56770	0.48 (42)	(0.45-0.50)	0.89 (79)	(0.85-0.93)
			66-80	67350	0.46 (41)	1.09 (96)	67772	0.48 (43)	(0.45-0.51)	0.99 (88)	(0.94-1.04)
			80+	47750	0.48 (42)	1.37 (121)	50542	0.49 (43)	(0.45-0.52)	1.21 (107)	(1.14-1.27)
GGT	U/L	M	19-50	31582	9	79	38692	9	(8.4-10.4)	52	(48.2-56.2)
			51-65	26758	12	95	31762	12	(10.5-13.3)	72	(66.5-77.9)
			66-80	28080	11	84	31909	12	(10.6-13.0)	53	(48.8-56.4)
			80+	13160	8	79	14901	9	(8.3-10.3)	48	(44.4-51.6)
		F	19-50	48040	7	48	51958	8	(7.3-8.7)	27	(24.9-28.3)
			51-65	35997	8	71	39690	9	(7.7-9.5)	42	(39.1-45.3)
			66-80	40412	8	65	43540	10	(8.9-10.5)	34	(32.0-36.6)
			80+	27174	7	66	29585	9	(8-9.6.0)	35	(32.6-37.4)
LDH	U/L	M + F	6-12	257	359	643	273	332	(313.2 - 350.8)	645	(613.4 - 677.4)
			13-18	340	274	531	340	288	(275.5 - 299.9)	432	(414.6 - 448.4)
			19-50	2573	256	507	2607	260	(246.1 - 273.5)	467	(445.0 - 488.6)
			51-65	1963	274	534	1983	282	(266.9 - 296.3)	503	(479.6 - 526.4)
			66-80	2039	270	551	2077	273	(257.7 - 287.7)	512	(487.1 - 536.5)

			80+	1539	266	584	1598	270	(253.6 - 286.0)	562	(532.9 - 591.3)
Magnesium	mg/dL (mmol/L)	M + F		4571	1.8 (0.72)	2.4 (1.00)	4915	1.8 (0.75)	(1.77-1.89)	2.4 (0.99)	(2.32-2.47)
Phosphate	mg/dL (mmol/L)	M	1-5	147	4.2 (1.34)	5.4 (1.74)	476	4.7 (1.49)	(4.48-4.82)	6.3 (2.00)	(6.04-6.46)
			6-12	405	4.2 (1.33)	5.3 (1.7)	763	4.2 (1.34)	(4.02-4.38)	6.3 (2.01)	(6.03-6.53)
			13-18	390	3.6 (1.16)	5.4 (1.72)	605	3.9 (1.24)	(3.70-4.06)	6.2 (2.00)	(5.98-6.50)
			19-50	3132	2.4 (0.77)	4.7 (1.51)	3172	2.4 (0.77)	(2.29-2.55)	4.6 (1.48)	(4.39-4.85)
			51-65	2917	2.2 (0.72)	4.3 (1.38)	2953	2.5 (0.79)	(2.33-2.59)	4.4 (1.40)	(4.19-4.59)
			66+	6640	2.2 (0.71)	4.2 (1.34)	6885	2.5 (0.78)	(2.33-2.57)	4.2 (1.35)	(4.04-4.42)
		F	1-5	126	4.3 (1.39)	5.4 (1.71)	430	4.5 (1.45)	(4.34-4.70)	6.5 (2.07)	(6.23-6.71)
			6-12	428	4.1 (1.32)	5.3 (1.71)	807	4.2 (1.35)	(4.05-4.39)	6.2 (1.97)	(5.92-6.38)
			13-18	729	3.5 (1.11)	5.2 (1.67)	791	3.6 (1.16)	(3.48-3.78)	5.4 (1.72)	(5.17-5.59)
			19-50	5874	2.6 (0.84)	4.7 (1.51)	5934	2.7 (0.86)	(2.55-2.83)	4.8 (1.53)	(4.55-4.99)
			51-65	7970	2.7 (0.88)	4.7 (1.5)	8041	3.0 (0.96)	(2.85-3.13)	4.8 (1.52)	(4.55-4.95)
			66+	18548	2.6 (0.84)	4.5 (1.43)	19113	2.8 (0.90)	(2.68-2.94)	4.5 (1.44)	(4.32-4.70)
Potassium	mmol/L	M + F		257189	3.6	5.1	269303	3.7	(3.54-3.82)	5.1	(4.89-5.24)
Sodium	mmol/L	M + F		256775	136	144	268873	136	(134.6 - 138.4)	144	(142.1 - 146.0)
Total protein	g/dL (g/L)	M + F		35141	6.1 (61)	8.0 (80)	36883	6.2 (62)	(6.04-6.45)	8.0 (80)	(7.75-8.24)
Urea	mg/dL (mmol/L)	M	1-5	227	15 (2.5)	45 (7.5)	618	13 (2.1)	(11.5-13.5)	41 (6.8)	(38.4-43.6)
			6-12	755	19 (3.1)	47 (7.7)	1094	18 (3.0)	(16.8-19.4)	47 (7.8)	(44.2-49.6)
			13-18	996	18 (3.0)	47 (7.8)	1001	17 (2.7)	(15.3-17.7)	45 (7.5)	(42.2-47.6)
			19-50	4709	20 (3.3)	54 (9.0)	4795	20 (3.3)	(18.3-21.1)	51 (8.6)	(48.4-54.4)
			51-65	4457	21 (3.5)	61 (10.2)	4862	20 (3.3)	(18.4-21.4)	56 (9.3)	(52.6-59.4)
			66-80	5680	23 (3.9)	75 (12.6)	7157	25 (4.1)	(22.6-26.4)	75 (12.4)	(70.0-79.4)

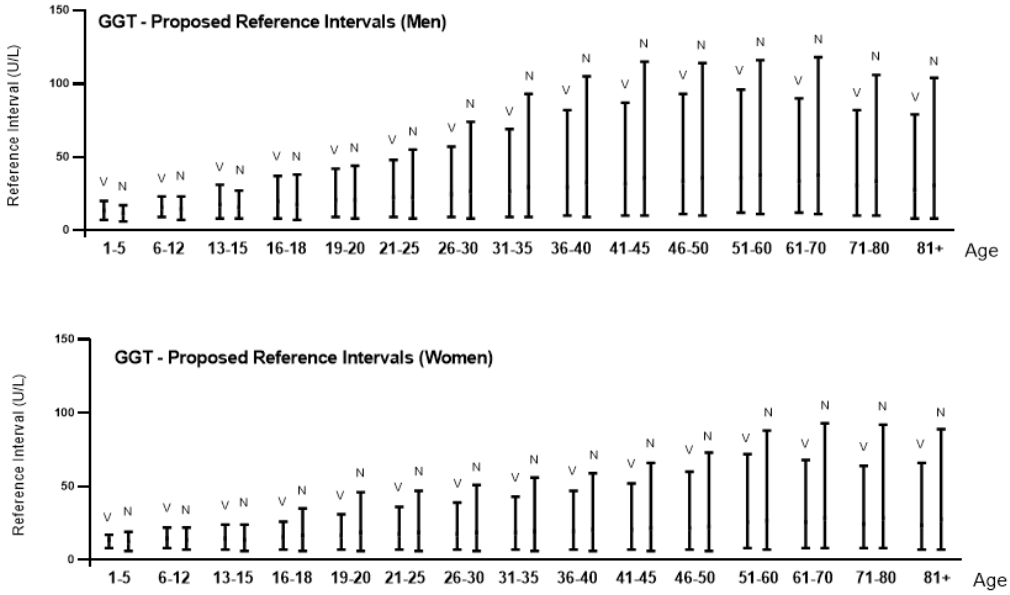
			80+	3662	27 (4.5)	93 (15.5)	6347	28 (4.6)	(25.1-30.7)	131 (21.8)	(121.4 - 140.4)
			1-5	167	16 (2.6)	43 (7.2)	499	14 (2.3)	(12.7-14.7)	37 (6.1)	(34.7-39.1)
			6-12	767	17 (2.8)	44 (7.4)	1168	16 (2.7)	(15.2-17.6)	43 (7.2)	(40.7-45.7)
			13-18	1220	16 (2.7)	42 (7.0)	1240	15 (2.5)	(13.7-15.9)	38 (6.4)	(36.2-40.6)
		F	19-50	6850	16 (2.7)	46 (7.7)	7001	16 (2.6)	(14.7-17.1)	43 (7.1)	(40.1-45.1)
			51-65	5524	20 (3.4)	58 (9.7)	5630	20 (3.4)	(18.8-21.8)	56 (9.2)	(52.2-58.8)
			66-80	6986	22 (3.7)	72 (12.0)	7447	24 (3.9)	(22.0-25.4)	65 (10.8)	(60.7-68.5)
			80+	7668	25 (4.1)	97 (16.2)	9801	24 (4.0)	(21.9-26.3)	98 (16.3)	(91.0-104.6)
CK	U/L		13-18	70	55	384	76	38	(33.5-41.9)	210	(193.9 - 226.1)
			19-50	772	50	380	1571	47	(40.6-52.8)	364	(333.4 - 394.0)
		M	51-65	1073	42	331	2057	43	(37.5-48.1)	296	(272.1 - 320.1)
			66-80	1057	33	280	1072	26	(23.4-28.2)	107	(99.1-114.1)
			80+	383	27	214	394	30	(26.6-33.0)	159	(146.6 - 170.6)
			13-18	69	37	225	67	47	(43.5-49.9)	116	(109.5 - 122.9)
			19-50	808	32	209	1579	36	(32.2-39)	155	(144.0 - 166.2)
		F	51-65	1162	34	238	2311	38	(34.0-41.0)	150	(139.9 - 160.7)
			66-80	1329	30	213	1333	28	(25.8-30.8)	102	(95.3-108.9)
			80+	726	23	199	745	17	(14.4-19.4)	172	(156.9 - 187.5)
Uric Acid	mg/dL (mmol/L)		6-12	117	2,1 (0,12)	5,6 (0,34)	115	2.2 (0.13)	(2.00-2.30)	5.3 (0.31)	(4.95-5.55)
			13-18	660	3,3 (0,20)	7,9 (0,47)	634	3.5 (0.21)	(3.28-3.72)	7.5 (0.45)	(7.11-7.91)
		M	19-50	22950	3,4 (0,20)	8,4 (0,50)	23022	3.9 (0.23)	(3.61-4.11)	8.7 (0.52)	(8.24-9.18)
			51-65	30690	3,5 (0,21)	8,7 (0,52)	30909	3.8 (0.22)	(3.52-4.02)	9.0 (0.54)	(8.50-9.52)
			66-80	38168	3,4 (0,21)	8,8 (0,53)	38597	3.8 (0.23)	(3.58-4.10)	9.5 (0.56)	(8.95-10.03)

	80+	18803	3,3 (0,20)	9,0 (0,54)	19382	3.6 (0.21)	(3.29-3.81)	9.4 (0.56)	(8.86-9.98)
	6-12	110	1,9 (0,11)	5,8 (0,35)	110	1.9 (0.11)	(1.76-2.04)	5.4 (0.32)	(5.05-5.71)
	13-18	790	2,4 (0,15)	6,1 (0,36)	771	2.5 (0.15)	(2.32-2.64)	5.7 (0.34)	(5.41-6.05)
	19-50	28283	2,2 (0,13)	6,4 (0,38)	28138	2.6 (0.15)	(2.39-2.73)	6.1 (0.37)	(5.79-6.49)
F	51-65	33556	2,4 (0,14)	7,3 (0,44)	33398	2.9 (0.17)	(2.68-3.10)	7.5 (0.45)	(7.10-7.98)
	66-80	47549	2,5 (0,15)	7,8 (0,47)	47587	3.1 (0.18	(2.86-3.32)	8.4 (0.50)	(7.92-8.92)
	80+	36558	2,5 (0,15)	8,6 (0,51)	37517	3.1 (0.18)	(2.80-3.330)	9.8 (0.58)	(9.20-10.46)

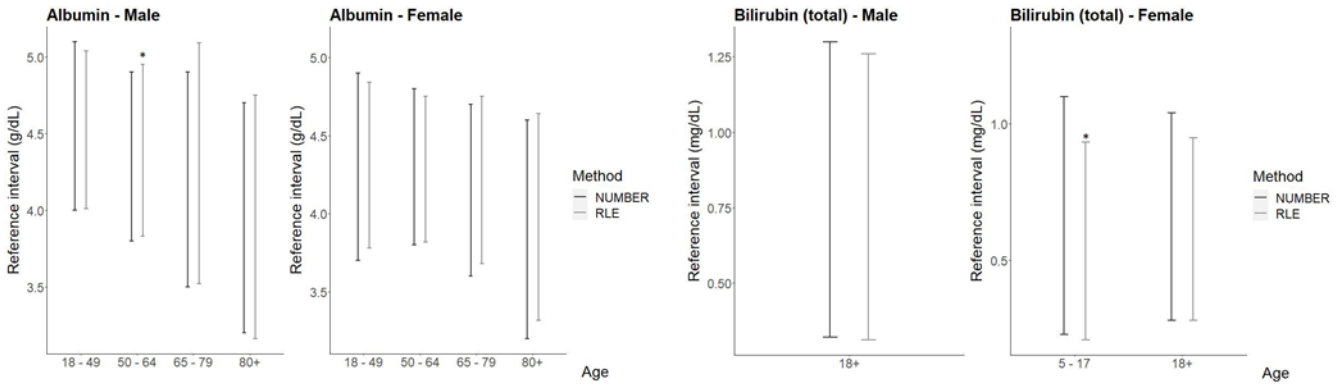
S1 Fig. Daily averages plot for calcium. Daily average is represented as points, monthly average as black lines and the average of the year as red lines. Slashed lines represent biological variation from monthly (black) or yearly (red) average and were used as an indication for person to person variation. Decisions about quality stability were made by visual inspection of the plots.

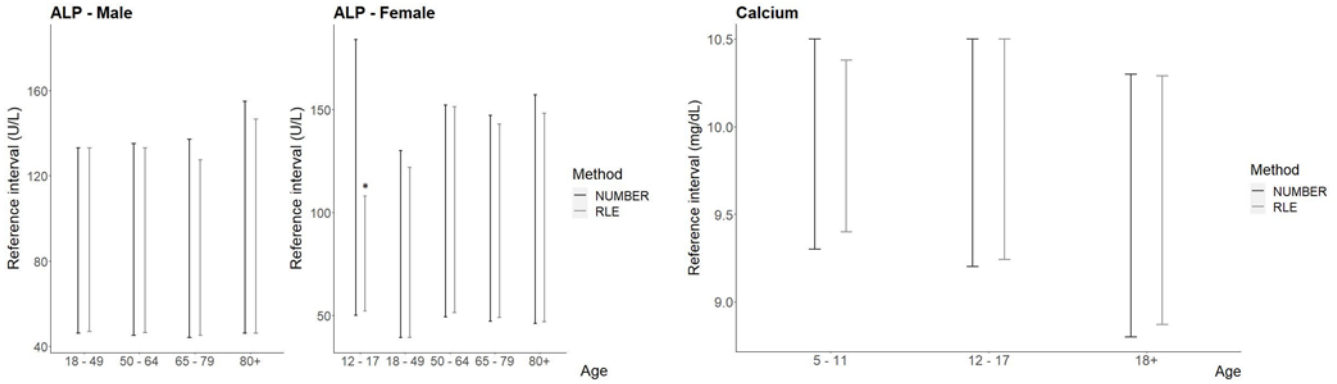


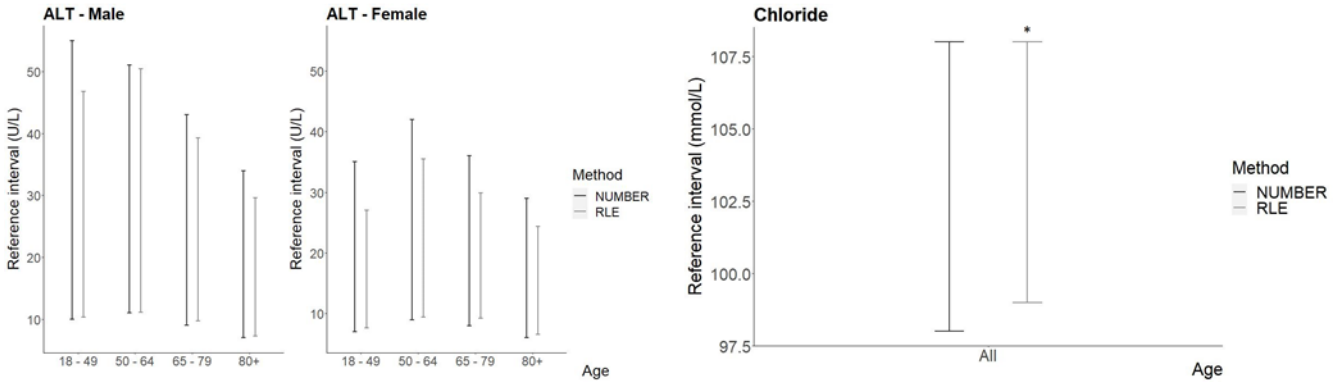
S2 Fig. GGT reference interval results by age. Different age representation for the calculated reference intervals for ALT and GGT for Vall d'Hebron (V) and NUMBER (N).

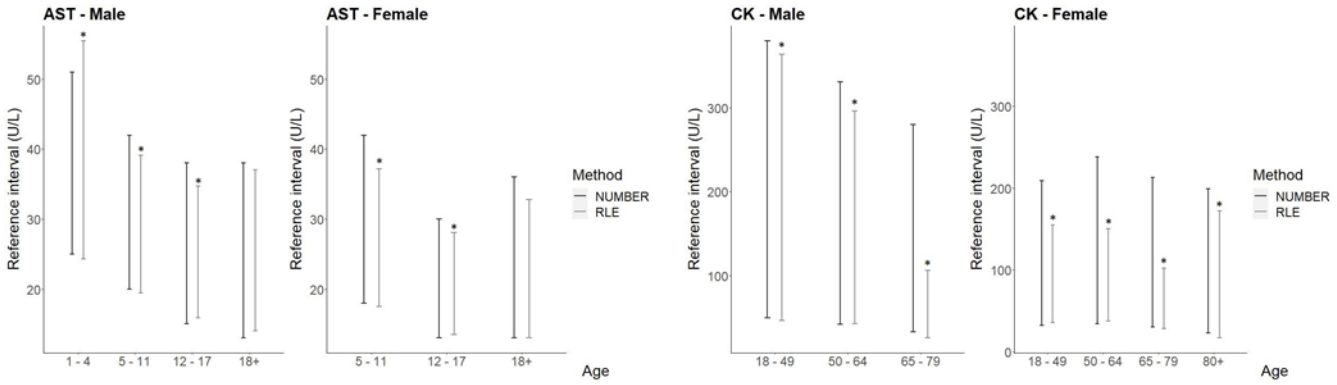


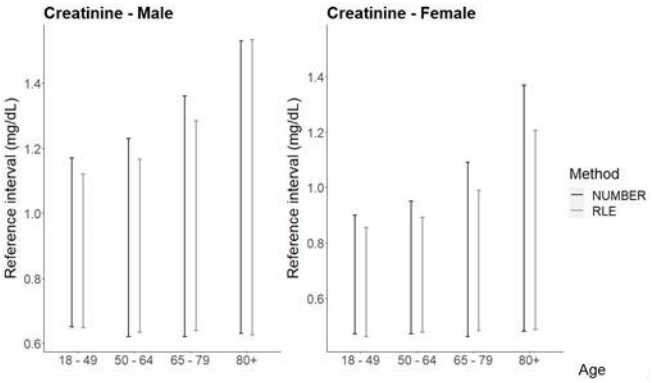
S3Fig. Comparison between indirect reference intervals using two methods. NUMBER method and reference limit estimator (RLE) method. Representation of reference intervals from S4 Table were made just when the number of data per both methods were higher than 500. *Reference interval results calculated with less data than the recommended by the RLE method (4.000).

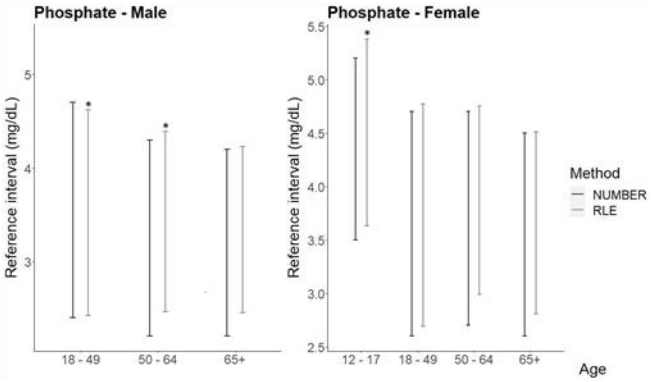

Albumin - Male
Albumin - Female
Bilirubin (total) - Male
Bilirubin (total) - Female
Reference interval (g/dL)
Reference interval (mg/dL)
18 - 49
50 - 64
65 - 79
80+
18+
5 - 17
Age
Method
NUMBER
RLE

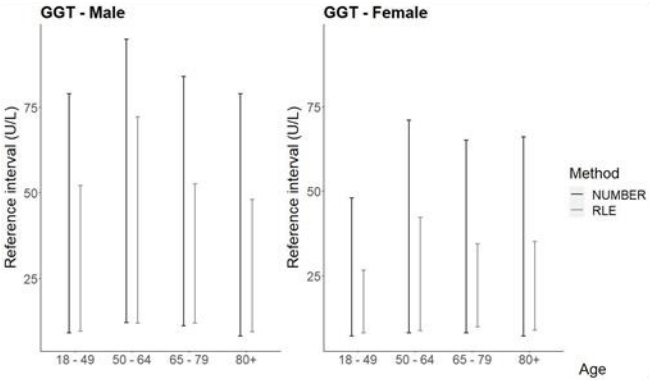

ALP - Male
ALP - Female
Calcium
Reference interval (U/L)
Reference interval (mg/dL)
12 - 17
18 - 49
50 - 64
65 - 79
80+
5 - 11
Age
Method
NUMBER
RLE

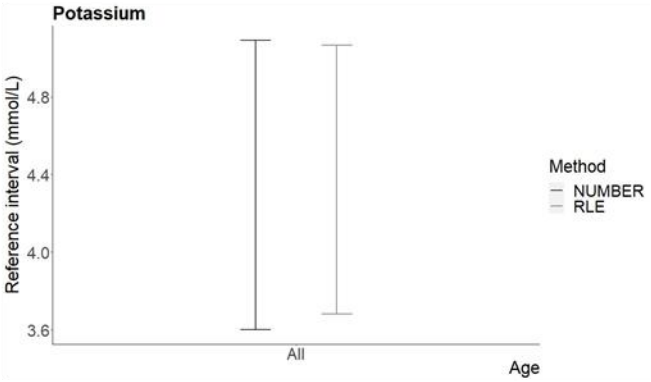

ALT - Male
ALT - Female
Chloride
Reference interval (U/L)
Reference interval (mmol/L)
All
Age
Method
NUMBER
RLE

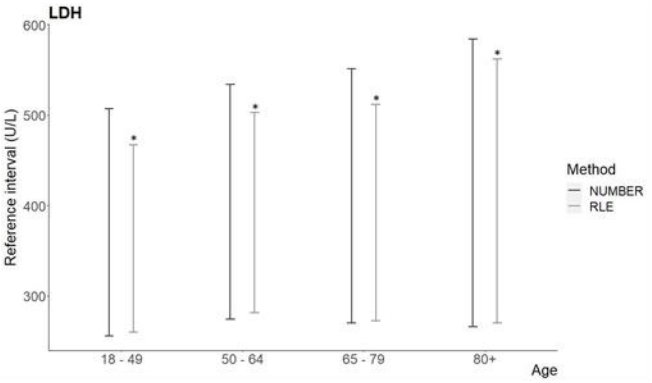

AST - Male
AST - Female
CK - Male
CK - Female
Reference interval (U/L)
1 - 4
5 - 11
12 - 17
18+
18 - 49
50 - 64
65 - 79
80+
Age
Method
NUMBER
RLE

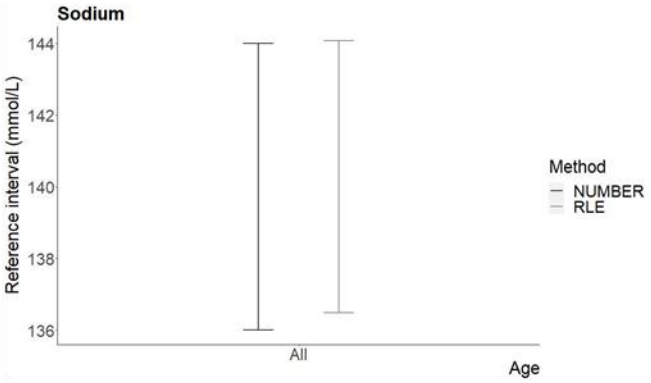

Creatinine - Male
Creatinine - Female
Reference interval (mg/dL)
18 - 49
50 - 64
65 - 79
80+
Age
Method
NUMBER
RLE

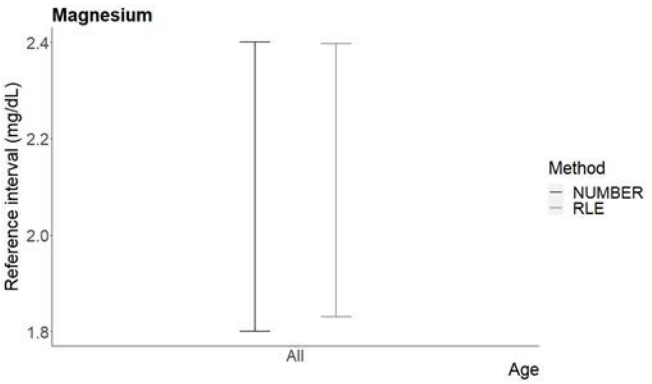

Phosphate - Male
Phosphate - Female
Reference interval (mg/dL)
18 - 49
50 - 64
65+
12 - 17
Age
Method
NUMBER
RLE

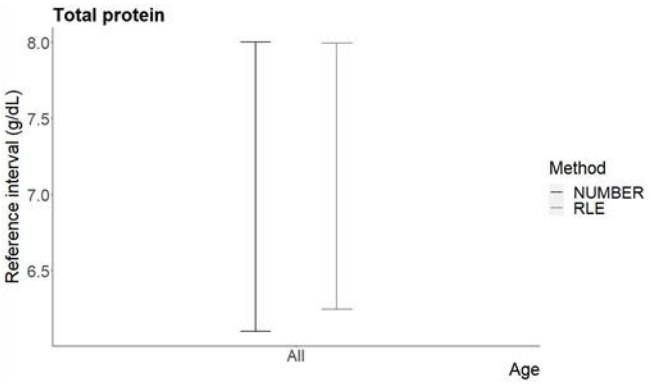

GGT - Male
GGT - Female
Reference interval (U/L)
18 - 49
50 - 64
65 - 79
80+
Age
Method
NUMBER
RLE

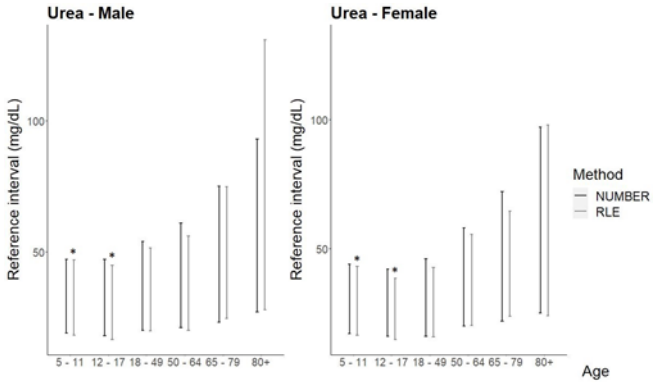

Potassium
Reference interval (mmol/L)
All
Age
Method
NUMBER
RLE

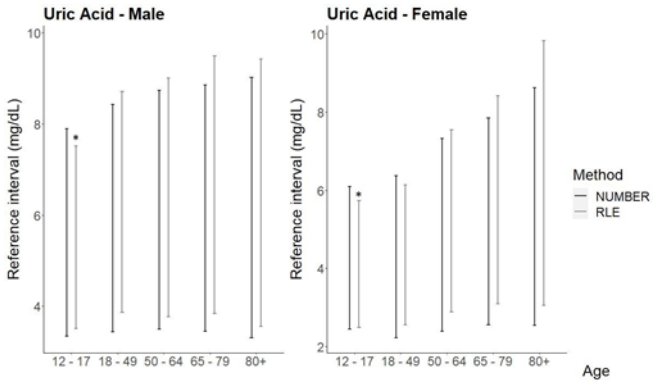

LDH
Reference interval (U/L)
18 - 49
50 - 64
65 - 79
80+
Age
Method
NUMBER
RLE


Sodium
Reference interval (mmol/L)
All
Age
Method
NUMBER
RLE


Magnesium
Reference interval (mg/dL)
All
Age
Method
NUMBER
RLE


Total protein
Reference interval (g/dL)
All
Age
Method
NUMBER
RLE


Urea - Male
Urea - Female
Reference interval (mg/dL)
100
50
5 - 11
12 - 17
18 - 49
50 - 64
65 - 79
80+
Age
Method
NUMBER
RLE


Uric Acid - Male
Uric Acid - Female
Reference interval (mg/dL)
10
8
6
4
2
12 - 17
18 - 49
50 - 64
65 - 79
80+
Age
Method
NUMBER
RLE

4. HARMONIZATION OF INDIRECT REFERENCE INTERVALS CALCULATION BY THE BHATTACHARA METHOD

4. Harmonization of indirect reference intervals calculation by the Bhattacharya method

4.1. Summary

Objectives: Bhattacharya is a graphical method for identifying a Gaussian distribution (reference population) in the midst of a complete dataset. This method is known to require the (subjective) input of the user, which results in important between user differences in calculated reference intervals. The aims of this study were to harmonize the criteria for the Bhattacharya indirect method Microsoft Excel Spreadsheet for reference interval calculation to reduce between-user variability, and to use these criteria to calculate and evaluate reference intervals for eight medical tests in two different years in the population from a single laboratory.

Material and Methods: Anonymized laboratory test results from outpatients were extracted from January 1st 2018 to December 31st 2019. To assure data quality, we examined the monthly results from an external quality control program and daily and monthly averages. Reference intervals were determined by the Bhattacharya method using the St Vincent's hospital Excel Spreadsheet firstly using original criteria provided by the Spreadsheet creator and then using additional harmonized criteria defined within this study. Variability between users was compared and evaluated using the coefficient of variation by magnitude and partition. Variability reduction was mainly focused on some user dependent variables to note bin size (size of the ranges of numerical values into which the data are sorted), total number of bins included and the number of bins inside the gaussian curve. To further reduce variability, statistical correlations between user dependent variables were analyzed. Then, consensus reference intervals using the additional harmonized criteria were calculated as the mean of four users' lower and upper reference interval results and by an independent user to test the defined criteria. Finally, flagging rates (percentages of measurements below and above the lower and upper reference limits) were calculated with an independent dataset.

Results: The extracted results for all selected laboratory tests fulfilled the quality criteria and were included in the present study. Differences between users in calculated reference intervals were frequent when using the Spreadsheet with the original criteria. Therefore, additional criteria for the Spreadsheet were proposed and applied by independent users, such as: to set central bin as the mean of all the data, bin size as small as possible, at least three consecutive bins and a high proportion of bins within the curve. Also, a recommendation was included to always calculate the reference intervals by four independent users. A reduction in within user variation were gained in 68.8% of the limits, with 40.6% of those reductions being statistically significant. The external user results slightly differed in 11.3% of the limits calculated by the initial users. A linear correlation was found between the bins or points included in the line and the number of available data which could further reduce variability. Results of the flagging rates obtained by applying the reference intervals calculated for 2018 and 2019 in a population from 2020 exceeded 5% of pathological values for all analytes except for phosphate in 2018.

Conclusions: An important reduction of between users' variability when using the tool was found for most tests after applying the criteria defined as part of the study, and therefore the proposed criteria contributed to the harmonization of reference intervals calculation between users with the Bhattacharya indirect method Spreadsheet. This system, including the additional criteria presented, could be applied in other clinical laboratories to optimize reference interval calculation by the Bhattacharya method.

4.2. Publication

Luisa Martinez-Sanchez[1,2,3,4]*, Pablo Gabriel-Medina[1,2,5]*, Yolanda Villena-Ortiz[1,2,5]*, Alba E. García-Fernández[1,5], Albert Blanco-Grau[1,2,5], Christa M Cobbaert[3], Daniel Bravo-Nieto[1,5], Sarai Garriga-Edo[1,5], Clara Sanz-Gea[1,5], Gonzalo Gonzalez-Silva[1,5], Joan López-Hellín[1,5], Roser Ferrer-Costa[1,5], Ernesto Casis[1,5], Francisco Rodríguez-Frías[1,2,5#], Wendy PJ den Elzen[4,5#].

*L. Martinez-Sanchez, P. Gabriel-Medina and Y. Villena-Ortiz contributed equally to this work.

[#] Fracisco Rodríguez–Frías and Wendy den Elzen contributed equally to this work.

[1]Biochemistry Department, Clinical Laboratories, Vall d'Hebron University Hospital, Barcelona, Spain.

[2]Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[3]Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Centre, Leiden, The Netherlands.

[4]Amsterdam UMC, University of Amsterdam, Department of Clinical Chemistry, Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[5]Clinical Biochemistry Research Team, Vall d'Hebron Institute of Research (VHIR)

Clin Chem Lab Med. 2022 Nov 17;61(2):266-274

The presented version is a pre-print from the published paper that can be found here:
10.1515/cclm-2022-0439

Harmonization of indirect reference intervals calculation by the Bhattacharya method

Short title: Indirect reference intervals using Bhattacharya method

Luisa Martinez-Sanchez[a,b,c,e*], Pablo Gabriel-Medina[a,b*e], Yolanda Villena-Ortiz[a,b*e], Alba E. García-Fernández[a,e], Albert Blanco-Grau[a,b,e], Christa M Cobbaert[c], Daniel Bravo-Nieto[a,e], Sarai Garriga-Edo[a,e], Clara Sanz-Gea[a,e], Gonzalo Gonzalez-Silva[a,e], Joan López-Hellín[a,e], Roser Ferrer-Costa[a,e], Ernesto Casis[a,e], Francisco Rodríguez-Frías[a,b,e,#], Wendy PJ den Elzen[d,e,#].

*L. Martinez-Sanchez, P. Gabriel-Medina and Y. Villena-Ortiz contributed equally to this work.

Abstract

Objectives. The aim of this study was to harmonize the criteria for the Bhattacharya indirect method Microsoft Excel Spreadsheet for reference intervals calculation to reduce between-user variability and use these criteria to calculate and evaluate reference intervals for eight analytes in two different years.

Methods. Anonymized laboratory test results from outpatients were extracted from January 1st 2018 to December 31st 2019. To assure data quality, we examined the monthly results from an external quality control program. Reference intervals were determined by the Bhattacharya method with the St Vincent's hospital Spreadsheet firstly using original criteria and then using additional harmonized criteria defined in this study. Consensus reference intervals using the additional harmonized criteria were calculated as the mean of four users' lower and upper reference interval results. To further test the operation criteria and robustness of the obtained reference intervals, an external user validated the Spreadsheet procedure.

Results. The extracted test results for all selected laboratory tests fulfilled the quality criteria and were included in the present study. Differences between users in calculated reference intervals were frequent when using the Spreadsheet. Therefore, additional criteria for the Spreadsheet were proposed and applied by independent users, such as: to set central bin as the mean of all the data, bin size as small as possible, at least three consecutive bins and a high proportion of bins within the curve.

Conclusions. The proposed criteria contributed to the harmonization of reference interval calculation between users of the Bhattacharya indirect method Spreadsheet.

Key words (3-6): reference Intervals; indirect approach; harmonization; Bhattacharya.

INTRODUCTION

Reference intervals are very important, as they support clinical decision making based on laboratory results (1,2). Laboratory test results outside the reference interval could be defined as pathological and may warrant further attention (2). In addition, the accuracy of in-range values is also important, since the unjustified absence of medical actions could also drive negative long-term consequences for patients (3). Therefore, establishing correct, updated and specific reference intervals for our population is a critical point in the clinical laboratory and assuring their quality and reliability is one of the most important tasks of specialists in clinical laboratory medicine.

Reference intervals are currently calculated using the direct approach (4). The limitations and disadvantages of this methodology have been widely discussed (5). To note: Complexity to select, contact and enrol 120 healthy random individuals, especially for those tests that require multiple partitions per sex and age or the costs of performing the study; among others. These drawbacks may become so tedious that routine laboratories frequently choose to adopt the reference intervals suggested by the manufacturer, calculated using a different population and settings.

Given these limitations, indirect methods have emerged as an alternative approach (6–10) and are increasingly used. These methods use data from thousands of individuals from already performed routine analyses, collecting the data from the laboratory information system (LIS) and subsequently analysing them statistically. Availability of a high number of test results in the LIS is an essential requirement for the calculation of reference intervals by indirect approaches. Clinical Laboratory Vall d'Hebron is one of the largest laboratories in Europe by workload and complexity as more than 60,000 tests results are produced every day and the catalogue includes more than 1,000 tests, providing "in vitro diagnostics" service to the majority of the Barcelona city public health activity. Faced with this scenario, we consider the calculation of reference intervals by indirect approaches a positive and revolutionary opportunity in our laboratory.

In 2019, the International Federation of Clinical Chemistry (IFCC) published a review encouraging clinical laboratories to participate in the development of indirect methods for reference intervals determination (11). Multiple methods have been developed using the idea of calculating reference intervals from patient populations: Hoffmann (12), Pryce (13), Bhattacharya (14), NUMBER (8), kosmic (15), truncated minimum chi-squared (TMC) (16), among others. In this study the Bhattacharya method was used due to a free access tool available online that facilitate the handling of indirect methods (often highly complex statistically) for non-statistical experts in the laboratory.

The Bhattacharya method was described in 1967 (14) and is a graphical method for identifying a Gaussian distribution (reference population) in the midst of a complete dataset with both reference individuals and non-reference individuals (non-healthy subjects). Two requirements are necessary to separate these two populations mathematically: 1) they do not highly overlap and 2) the total sample size is large enough (more than 1,500 in the original description). In the original description of the method, the Gaussian distribution of data was considered another requirement. Since most laboratory data do not show a normal distribution, Baadenhuijsen *et al.* and Oosterhuis et al. described some modifications to address some of these limitations (17,18).

The currently used spreadsheet and other online applications for the Bhattacharya method apply linear regression to shape a line of best fit for the segment that the user visually chooses as a straight-line. This line identifies the reference population. Actually, more robust and reliable reference intervals are estimated if larger numbers of individuals are included (more than 5,000) and a greater proportion of the dataset is from the reference population (11).

In the present study, Bhattacharya analyses were performed using St Vincent's hospital Spreadsheet available online (http://www.sydpath.stvincents.com.au/). This method requires the (subjective) input of the user for selecting an appropriate bin size and the points included on the graph (19). The first purpose of this study was to study between user variability when calculating reference intervals in the Excel application; then a second purpose was to standardize the criteria initially defined to reduce between-user variability in the reference interval results. Then, a third aim was to calculate and evaluate reference intervals for eight tests during two different years (2018 and 2019), based on the new criteria.

MATERIAL AND METHODS

Data selection

Anonymized laboratory test results from individuals (more than 18 years old) visiting general practitioners were extracted from January 1st 2018 to December 31st 2019 from the LIS of the Clinical Laboratory Vall d'Hebron in Barcelona.

Test results from outpatients belonging to primary care attention centres were included, since we expected a high proportion of healthy people. Haemolytic (>0.03 mmol/L haemoglobin), lipemic (>0.45 mmol/L Intralipid®) and icteric (>23.94 μmol/L bilirubin) samples were excluded. A total of 1,067,794 clinical requests were selected (509,408 from 2018 and 558,386 from 2019). A detailed description of the dataset from 2018 is provided elsewhere (20).

Analytical measurements

Samples were collected from 62 blood collection centres and were transported via 8 different routes to the laboratory (average transportation time 3 hours). Serum tubes included separating gel and coagulation activator (BD Vacutainer®). The samples were transported to the laboratory in cool boxes with a temperature monitoring system. After arriving in the laboratory, the samples were centrifuged either 12 minutes at 3,500 rpm (2,438 *g*) when handled manually outside the track or 10 minutes at 3,000 rpm (2,113 *g*) when on the track. No clinically significant differences in the test results were found when comparing the two centrifugation conditions (results not shown).

Biochemistry tests were measured on AU5800 chemistry analysers (Beckman Coulter®). The following test methods were used according to the instructions for use of the manufacturer: alanine aminotransferase (ALT), IFCC recommended method without pyridoxal phosphate traceable to Beckmann coulter master calibrator; glucose, reaction with hexokinase traceable to NIST SRM 965; calcium, reaction with arsenazo III traceable to NIST SRM 909bL1; magnesium, direct method with xylidyl blue traceable to NIST SRM 909bL2; inorganic phosphorus, reaction with ammonium molybdate traceable to Beckmann coulter master calibrator; chloride, potassium and sodium by indirect ion selective electrodes traceable to NIST SRM 919, 918 and 919 respectively.

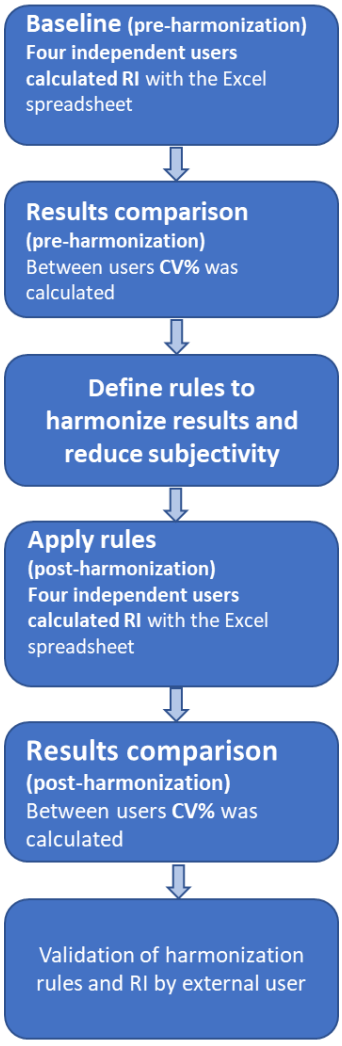


Figure 1. Workflow with the main steps followed during the project.

Quality assessment

To assure data quality, we examined the monthly results from the biochemistry specific external quality control program from the Spanish Society of Laboratory Medicine (Sociedad Española de Medicina de Laboratorio, SEQC^{ML}). In this scheme, the results from the external quality control materials obtained in our laboratory were compared with the average calculated from every laboratory participating in the program using the same analytical method and/or instrument. Alike routine laboratory practice, when our result was within two times the standard deviation from other laboratories participating in the scheme using the same method, data from this particular month and test were accepted as valid. If our result exceeded ± two standard deviations, we excluded the data from that particular test, month and instrument. In addition, uncertainty was calculated as the sum of standard uncertainty from the calibrator material, the analytical coefficient of variation and the uncertainty of the analytical system.

To assess longitudinal accuracy across lot numbers, daily averages of the extracted General Practitioner test results were investigated to check for analytical stability over time. Averages were calculated per batch of 200 results a day and were visually compared with the average per month and average for the whole year 2018 or 2019.

Reference intervals calculation and statistical analysis

The Bhattacharya method was performed to determine the reference intervals using the programmed Microsoft Excel sheet by St Vincent's hospital (available in: http://www.sydpath.stvincents.com.au/) as advised by the IFCC Committee on Reference Intervals and Decision Limits (C-RIDL) (11). A workflow with the main steps followed during the project process is presented in Figure 1.

First, identical excel templates were made, avoiding errors in data transfer. Then, four different laboratory specialists (users) worked independently to obtain the reference intervals following the initial recommendations from the original sheet. The variability between users was compared and evaluated using

the coefficient of variation (CV) of the reference interval limits calculated by magnitude and partition. To simplify results presentation, the CV is shown together for 2018 and 2019 and separating low and high interval per test. For ALT only the high interval is shown as the low intervals were not considered clinically relevant.

To reduce variability between users, we focused on the user-dependent variables based on the results obtained, i.e. in the bin size, the total number of bins included in the reference population (#points) and the number of bins graphically inside the curve (which was considered 0.5 when half bin stood outside the curve) (Figure 2). Based on our experiences with the Spreadsheet, we developed additional new consensus criteria for the use and operation of the excel sheet in order to reduce inter-user variability. Again, the same four users obtained the reference intervals independently using the new consensus criteria.

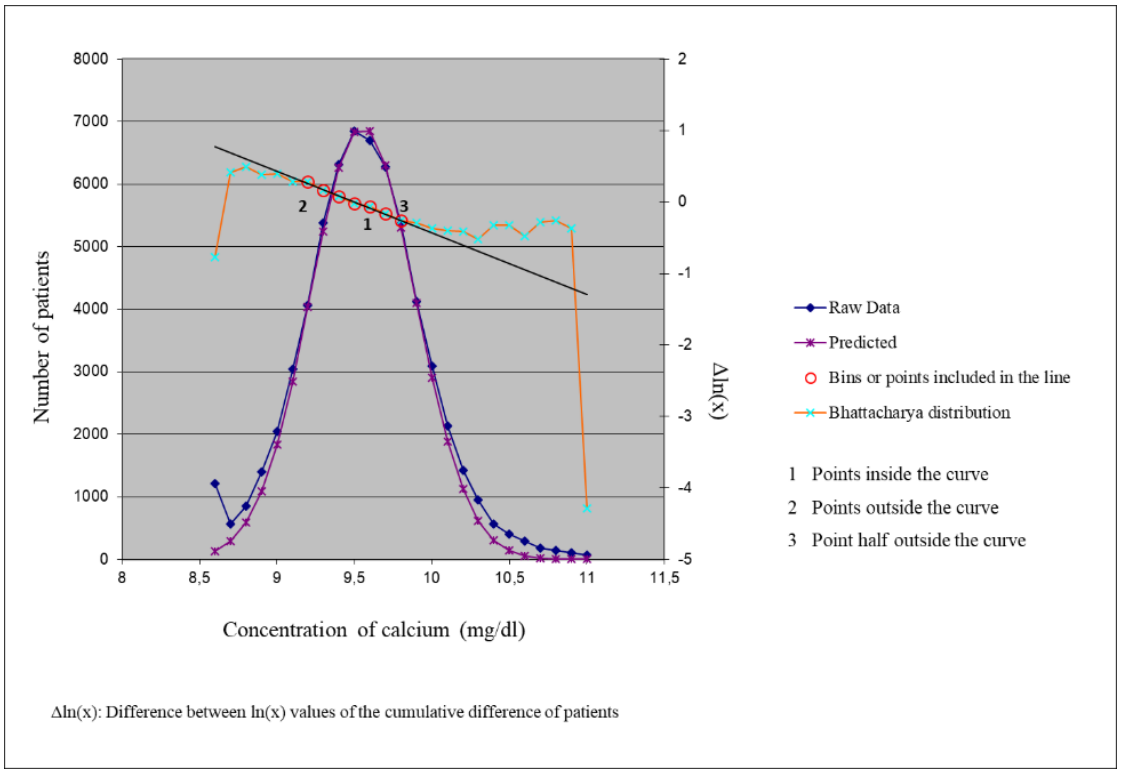


Figure 2. Representation of the graphs obtained using the St Vincent's hospital excel Spreadsheet (http://www.sydpath.stvincents.com.au/) for Bhattacharya indirect method reference interval calculation. Raw data present how the original data is distributed according to the selected bin size; predicted data present how the data would be distributed according to the bin size selected and to the number of bins included in the line and considered for the calculations.

We calculated the mean between the four users and the 95% confidence interval (CI) using the formula $\mu \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$, being μ=mean, Zα/2=1.96 and σ=standard deviation (SD) with n=4 for the four independently obtained low and high limits of the reference intervals. If, after this calculation, any of the four users, either the low or high limit of the calculated reference interval lay outside the 95% CI, then its results were considered not valid and discarded for the final calculation of the new 95% CI and reference intervals. Consensus reference intervals were calculated as the mean of the valid users' high and low results. As an example, the results obtained for the high limit for potassium in mmol/L in 2018 were: user 1 = 4.99; user

2= 5.03; user 3 = 5.11 and user 4 = 5.03. The 95% CI calculated for the 4 users (n=4) was 4.99 – 5.04. Therefore, as the result from user 3 was higher than the 95% CI, it was considered not valid and the mean reference interval was re-calculated using the remaining three users' results.

Per test and per group boxplots were visually inspected to decide whether or not subgroup differentiated reference intervals were necessary per sex and age.

Results evaluation

To test for significant differences between pre and post harmonization strategies, the F-test (21) was applied to the SD before and after harmonization. In addition, to decide upon the acceptance of the obtained CV after harmonization within-individual biological variation was used (22).

To reduce variability of the user dependent variables, statistical correlations between them were analysed. After applying the new criteria, the number of data (n), the number of decimal points and the central bin (defined by the data) were considered as independent variables. The contribution of the independent variables on the bin size and the bins included in the line (#points) were analysed in univariate and multivariate models using linear regression analysis and Pearson correlation.

To further test the operation criteria and robustness of the obtained reference intervals, an external user reproduced the Spreadsheet procedure applying the defined criteria.

Flagging rates per test and per year were calculated with an independent dataset from primary care (1st January 2020 to 31st December 2020). Results are shown as the percentage of individuals outside the reference intervals.

RESULTS

The data obtained fulfilled the quality criteria and, therefore, were included in the present study. Longitudinal accuracy was also considered fulfilled as observed in annual averages for the eight laboratory tests studied (Supplemental Table 1) and in the monthly averages plots. An example of these plots is shown in Figure 3 for potassium, where increase in potassium concentrations was observed during colder temperatures months (23).

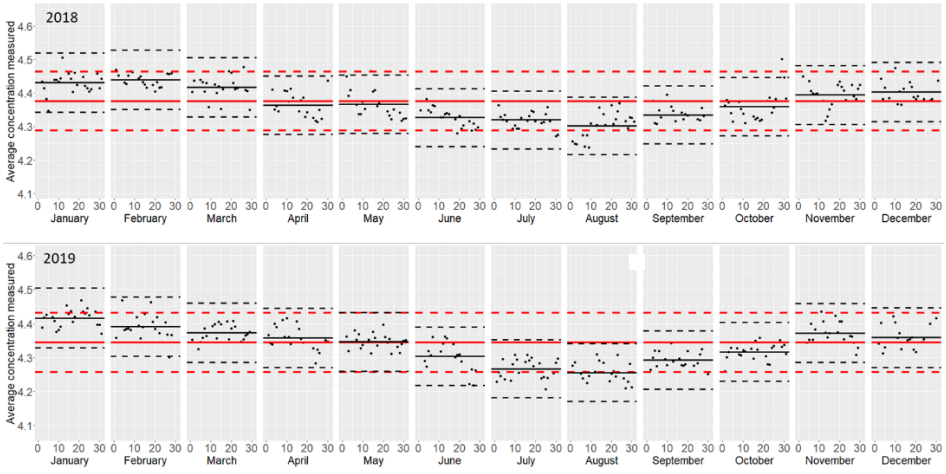


Figure 3. Daily averages of consecutive primary attention patient results in the extracted datasets (black dots) for potassium results in 2018 and 2019. Monthly average (black) and annual average (red) are represented as lines and biological variation percentage over and under the mean is represented a slashed lines.

Table 1 shows the coefficient of variation (CV) between users' reference intervals calculations when applying the original criteria (pre-harmonization), median (Q1-Q3) of 3.99% (1.49-10.95). The additional criteria defined to standardize the analysis by reducing between user variability in the reference intervals calculation and their justifications are shown in Table 2. The CVs after applying these additional criteria are also presented in Table 1 (post-harmonization). Considerably less inter-user variability was obtained in the post-harmonization results, with a median (Q1-Q3) of 2.90% (0.06-7.35). Table 1 also shows within-subject biological variation for comparison with the obtained CV and the p-value from the F-test to assess significant differences between variation of the pre and post harmonization results.

Table 1. Coefficient of variation (CV) between the reference intervals results calculated between the 4 users per each test in 2018 and 2019 with the original criteria of Microsoft Excel Bhattacharya Spreadsheet (pre-harmonization) and with the criteria proposed within this study (post-harmonization). Within-subject biological variation (CVi) and p-value for the F-test between the pre and post harmonization results are also shown.

Analytical test	RV 2018 and 2019	CV_i (%)	CV (%) pre-harmonization	CV (%) post-harmonization	p-value (F-test)
Sodium	LRL	0.5	0.19	0.21	0.809
	URL		0.35	0.34	0.931
Potassium	LRL	4.0	0.71	0.74	0.923
	URL		1.16	0.78	0.297
Chloride	LRL	1.0	0.77	0.00	**<0.001**
	URL		0.85	0.00	**<0.001**
Calcium	LRL	1.8	1.30	0.00	**<0.001**
	URL		1.11	0.00	**<0.001**
Magnesium	LRL	2.8	0.73	0.00	**<0.001**
	URL		0.94	0.00	**<0.001**
Phosphate (Males 18-50 years)	LRL	7.7	4.00	4.18	0.869
	URL		2.23	2.12	0.877
Phosphate (Females 18-50 years)	LRL		3.39	4.73	0.402
	URL		1.79	1.77	1.000
Phosphate (Males 51-65 years)	LRL		5.61	0.00	**<0.001**
	URL		2.23	0.00	**<0.001**
Phosphate (Females 51-65 years)	LRL		1.22	0.00	**<0.001**
	URL		0.93	0.00	**<0.001**
Phosphate (Males >65 years)	LRL		5.86	4.41	0.490
	URL		1.90	2.06	0.844
Phosphate (Females >65 years)	LRL		3.25	3.82	0.734
	URL		2.32	2.26	1.000
Glucose	LRL	4.9	4.29	3.36	0.679

	URL		3.42	0.75	**0.001**
ALT (Males 18-50 years)	URL	10.0	15.23	9.44	0.123
ALT (Females 18-50 years)	URL		13.64	5.71	**0.018**
ALT (Males 51-65 years)	URL		7.32	7.71	0.719
ALT (Females 51-65 years)	URL		11.04	1.04	**<0.001**
ALT (Males 66-80 years)	URL		5.99	7.65	0.711
ALT (Females 66-80 years)	URL		5.79	3.45	0.126
ALT (Males >80 years)	URL		6.35	8.38	0.603
ALT (Females >80 years)	URL		5.09	9.02	0.331
Median (Q2)			**2.28**	**1.41**	
Lower quartile (Q1)			**1.07**	**0.00**	
Upper quartile (Q3)			**5.66**	**4.24**	

Table 2. Summary of original and additional criteria defined to harmonize the analysis by reducing between user variability in the reference intervals calculated using St Vincent's hospital Spreadsheet available online (http://www.sydpath.stvincents.com.au/).

ORIGINAL CRITERIA	ADDITIONAL CRITERIA	EXPLANATION
Initial central bin (including log transformed data) should be close to the mean or median.	Set the value of the central bin as the mean of all the data.	Central bin could be fixed as the arithmetic mean if there is a little influence from pathological results in the database and this would reduce the variability.
Bin size must be equal to or larger than the reporting interval.	To adjust the bin size, use the value of the reporting interval of the data as a starting point and increase it to meet all the following criteria. Finally select the smallest possible bin size.	Higher bin sizes lead to low resolution graphs and inappropriate reference interval results.
Select data from four to six bins to include in the Bhattacharya analysis.	The line must be defined with a minimum of 4 bins, at least three of them consecutively.	The biggest possible number of bins should be selected, since it allows a larger population to be included in the calculation of reference intervals.
	If bins are not considered for the adjustment of the line, they must be placed between two included bins.	Excluding intermediate bins assumes that the subpopulation is not homogeneous with respect to the bins immediately nearby. Excluding a single bin might be permissible, if a minimum bin size is selected. Excluding more than one intermediate bin would be an error and would skew the result.
The Bhat line must be very straight. Particularly data points "steeper" than the line of best fit should be included.	R-squared value 0.99 is big enough.	A larger R-squared does not modify or ensure validity of the results obtained. Instead, looking for a larger R-squared can penalize the selection of the most important variables, the bin size and the number of bins selected.
	The maximum number of points on the line must be included within the curve.	Points included inside the curve highlight the importance of the central bunch of data for the final calculation of reference intervals, in contrast with the data found in the extremes of the distribution.
If in doubt, seek expert advice and/or another operator for validation.	The spreadsheet should be explored by independent scientists (four in our case).	Reducing the inherent subjectivity that could lead to less reliable results when obtained by a single user.

Results of the bin size, the number of total points included in the line (#points) and the number of points within the distribution curve (#points inside) obtained by each of the four users are shown in Supplemental Table 2.

Supplemental Table 3 shows the reference intervals results obtained in the years 2018 and 2019 per user, the final 95% CI between the users and the reference intervals currently used in our laboratory derived from analyser inserts (RI_{cu}) for the eight tests studied. Shaded results in Supplemental Table 3 were considered not valid (outside 95% CI, n=4) and discarded for the final calculation of the final 95% CI (n=3) and reference intervals.

A linear correlation was found by the univariate model between the bins or points included in the line (see Figure 2, y = #points) and the number of available data (x = n) ($r=0.277$; $p<0.001$). This correlation was defined by the formula: $y = 5.818 + 0.53x10^{-5}\,x$. According to that, when calculating reference intervals for a laboratory test with for example 10,000 results, the recommended points or bins over the curve based on our formula are 5.8, rounded to 6 included bins. It means that there is a proportional increase in #points with higher n. In the multiple linear regression analysis, bin size was statistically associated with the central bin ($\beta=0.071$, $p<0.001$) and decimal points (-0.0943, $p<0.001$) ($R^2_{adjusted}=0.706$). It is important to remark that the observed correlations are specific for the selected analytes, the units and the methodology.

The external user results differed in 9 out of the 80 limits calculated by the initial users (Supplemental Table 3). This was particularly the case for the lower and upper limits for chloride of 98 mmol/L and 110 mmol/L respectively, the lower limit for magnesium of 0.78 mmol/L (1.9 mg/dl), the lower limit for phosphate in males (>65 yrs) 0.68 mmol/L (2.1 mg/dl) from 2018, the upper limit of phosphate in males (>65 yrs) of 1.42 mmol/L (4.4 mg/dl) from 2019, the lower limit for ALT in males (51-65 yrs) 9 U/L for 2018, the higher limit for ALT in males (>80) 29 U/L for 2018 and the lower limits for ALT in females (18-50) 5 U/L and (51-65) 7 U/L from 2018. The remaining reference intervals calculated by the external user fell within the 95% CI calculated by the independent users.

Results of the flagging rates obtained by applying the reference intervals calculated for 2018 and 2019 in a population sample from 2020 are shown in Figure 4. The percentages of the flagging rates for 2018 and 2019 exceeded 5% of pathological values for all analytes except for phosphate in 2018.

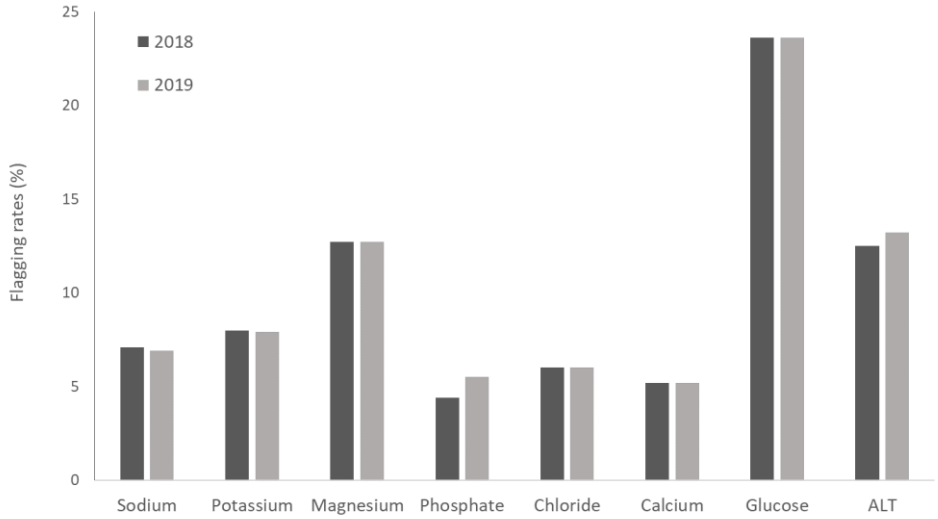


Figure 4. Validation of the calculated RIs in the 2020 laboratory dataset. Percentage of flagged patient results outside the calculated reference intervals in 2018 and 2019.

DISCUSSION

The Bhattacharya indirect method for reference intervals calculation can be performed in a simple and easy way using the Spreadsheet created by St Vincent's hospital. Initial recommendations, included in the "instruction" sheet from the excel Spreadsheet, allows the user to obtain reliable results but variability between users was found to be an issue (Table 1). Additional criteria and recommendations created within this study from the observation of these variations (Table 2), reduced subjectivity when performing the procedure for reference intervals calculation using the Spreadsheet. In addition, we observed a dependency between some (subjective) decisions the user has to face (as the number of points to be included in the line or the bin size) and other known variables as the number of results or the test units. When these relationships are taken into account, even less between user variability may be observed.

A reduction in within user variation using the Excel spreadsheet were gained in 22 out of 32 reference interval limits presented in Table 1; 13 of those reductions were statistically significant. For the 10 cases in which a reduction was not observed, the change in CV was not statistically significant. A reduction of CV appeared challenging for ALT for which extreme values are found often and a non-Gaussian distribution is present for the test results. As the Bhattacharya method does not exclude extreme values previous to reference intervals calculation, medical tests with a high proportion of extreme values will have more variability between users when using the tool. Between user variations for the calculated reference intervals were always lower than the within-subject biological variation for both pre and post harmonization, except for ALT (males and females, 18-50 and females 51-65) where CV pre harmonization was higher than 10%. To note, the pre harmonization CVs were in general very close to the within-individual biological variation threshold for those tests in which

we gained a significant reduction of CV for the post harmonization results by applying the additional harmonized criteria.

In previous studies other procedures to exclude bin selection, based on the differences in data frequency between consecutive bins, have been proposed; either by establishing a minimum data frequency regarding the mode or by graphical observation of the residues obtained from Bhattacharya graphic against test concentration (24). Since we aimed to propose a simple and objective method, this was not considered in this study.

Bin size is also an important variable for method performance. Smaller bin sizes will lead to higher random variation in the number of data per bin and therefore the complexity of the linear fit that represents the Gaussian population will be higher (25). We noticed that the reporting interval of data is also an important source of variability between users (Supplemental Table 2). A lower reporting interval of data leads to different bin sizes between users and therefore more variability in obtained reference intervals. Potassium is an example of this.

The ratio of results outside the calculated reference intervals (flagging rates) in 2018 and 2019, from a new dataset with population from 2020 has shown similar results for all analytical tests. Therefore, even with slight numeric differences in reference intervals, the method leads to coherent results attending to the ratio of pathological population detection. The expected results higher than 5% in flagging rates (based on the statistical model of reference intervals where 95% of healthy population are within the intervals (26)) are accomplished in all cases except for phosphate in 2018 (4.4%).

It is important to remark that the same dataset from 2018 was used in a previous study (20) for calculating reference intervals using two indirect methods: The Dutch NUMBER method (11) and the German reference limit estimator method (16). The calculated reference intervals were comparable with the mentioned results (20) for all the included analytical test. Comparison with other important reference interval studies such as CALIPER (direct method) (27), ARIA (indirect method) (28) and NORIP (direct method) (29) also gives comparable results for all tests, except for ALT. This is an important topic of further research.

One of the weaknesses in our study is that level one commutable external quality control was not applied yet in our laboratory in 2018 and 2019. Two important aspects from data quality should be always considered for data reuse: 1) the use of methods traceable to higher order reference materials and 2) the use of level one commutable external quality control. The fulfilment of these two requirements is a prerequisite for the application of calculated reference intervals to the clinical practice (5). If data quality is assured (30), the obtained reference intervals from different populations can be universally compared. The proposed methodology for the use of the Spreadsheet in Bhattacharya calculation is useful for data from several laboratories where these conditions are met.

To conclude, we assessed between user variability when using the Bhattacharya Excel Spreadsheet and designed additional criteria to harmonize reference intervals calculation. Considering the eight laboratory tests analysed, we conclude that the proposed additional criteria for the use of St Vincent's hospital Spreadsheet contribute to the harmonization of reference intervals calculation by the Bhattacharya method.

This system, including the additional criteria presented, could be applied in other clinical laboratories to optimize reference intervals calculation by the Bhattacharya method.

REFERENCES

1. Siest G, Henny J, Gräsbeck R, Wilding P, Petitclerc C, Queraltó JM, et al. The theory of reference values: An unfinished symphony. Clin Chem Lab Med. 2013;51:47–64.
2. Sikaris KA. Physiology and its importance for reference intervals. Clin Biochem Rev. 2014;35:3–14.
3. Placzkowska S, Terpinska M, Piwowar A. The Importance of Establishing Reference Intervals - is it still a Current Problem for Laboratory and Doctors? Clin Lab. 2020;1:1429-38.
4. Committee S, Section C, Petitclerc C, Biochimie S De, Montreal H De. International Federation of Clinical Chemistry (IFCC) 1), 2) Approved Recommendation (1987) ; on the Theory of Reference Values Part 2 . Selection of Individuals for the Production of Reference Values International Föderation of Clinical Chemistry. J Clin Chem Clin Biochem. 1987;25:2–30.
5. Martinez-Sanchez L, Marques-Garcia F, Ozarda Y, Blanco A, Brouwer N, Canalias F, et al. Big data and reference intervals: rationale, current practices, harmonization and standardization prerequisites and future perspectives of indirect determination of reference intervals using routine data. Adv Lab Med / Av en Med Lab. 2020;0:9–16.
6. Poole S, Schroeder LF, Shah N. An unsupervised learning method to identify reference intervals from a clinical database. J Biomed Inform. Elsevier Inc.; 2016;59:276–84.
7. Lykkeboe S, Nielsen CG, Christensen PA. Indirect method for validating transference of reference intervals. Clin Chem Lab Med. 2018;56:463–70.
8. Den Elzen WPJ, Brouwer N, Thelen MH, Le Cessie S, Haagen IA, Cobbaert CM. NUMBER: Standardized reference intervals in the Netherlands using a “big data” approach. Clin Chem Lab Med. 2019;57:42–56.
9. Clerico A, Trenti T, Aloe R, Dittadi R, Rizzardi S, Migliardi M, et al. A multicenter study for the evaluation of the reference interval for TSH in Italy (ELAS TSH Italian Study). Clin Chem Lab Med. 2019;57:259–67.
10. Shaw JLV, Cohen A, Konforte D, Binesh-Marvasti T, Colantonio DA, Adeli K. Validity of establishing pediatric reference intervals based on hospital patient data: A comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children. Clin Biochem. The Canadian Society of Clinical Chemists; 2014;47:166–72.
11. Jones GRD, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, et al. Indirect methods for reference interval determination - Review and recommendations. Clin Chem Lab Med. 2019;57:20–9.
12. Hoffmann RG. Statistics in the Practice of Medicine. JAMA J Am Med Assoc. 1963;185:864–73.

13. Pryce JD, Lond MD. &dag er; 1960;276:333–6.
14. Bhattacharya CG. A Simple Method of Resolution of a Distribution into Gaussian Components. Biometrics. 1967;23:115.
15. Zierk J, Arzideh F, Kapsner LA, Prokosch HU, Metzler M, Rauh M. Reference Interval Estimation from Mixed Distributions using Truncation Points and the Kolmogorov-Smirnov Distance (kosmic). Sci Rep. 2020;10:1704.
16. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. Clin Chem Lab Med. 2019;57:1933-1947.
17. Baadenhuijsen H, Smit JC. Indirect Estimation of Clinical Chemical Reference Intervals from Total Hospital Patient Data: Application of a Modified Bhattacharya Procedure. Clin Chem Lab Med. 1985;23:829–40.
18. Oosterhuis WP, Modderman Th. A, Pronk C. Reference values: Bhattacharya or the method proposed by the IFCC? Ann Clin Biochem. 1990;27:359–65.
19. Sikaris KA. Separating disease and health for indirect reference intervals. J Lab Med. 2021; 45: 55–68.
20. Martinez-Sanchez L, Cobbaert CM, Noordam R, Brouwer N, Blanco-Grau A, Villena-Ortiz Y, et al. Indirect determination of biochemistry reference intervals using outpatient data. PLoS One. 2022;17:e0268522.
21. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. Crit Rev Clin Lab Sci. 1989;27:409 – 37.
22. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, et al. The EFLM Biological Variation Database [Internet]. The European Federation of Clinical Chemistry and Laboratory Medicine (EFLM). [cited 2022 Sep 25]. Available from: https://biologicalvariation.eu/.
23. Davis KR, Crook MA. Seasonal factitious increase in serum potassium: still a problem and should be recognised. Clin Biochem. 2014;47: 283–286.
24. Hemel JB, Hindriks FR, Van Der Slik W. Critical discussion on a method for derivation of reference limits in clinical chemistry from a patient population. J Automat Chem. 1985;
25. Farrell CJL, Nguyen L. Indirect Reference Intervals: Harnessing the Power of Stored Laboratory Data. Clin Biochem Rev. 2019;
26. Wayne P. Defining , Establishing , and Verifying Reference Intervals in the Clinical Laboratory ; Approved Guideline. Third Edition. C28-A3. CLSI. 2008.
27. Adeli K, Higgins V, Nieuwesteeg M, Raizman JE, Chen Y, Wong SL, Blais D. Biochemical marker reference values across pediatric, adult, and geriatric ages: establishment of robust pediatric and adult reference intervals on the basis of the Canadian Health Measures Survey. Clin Chem. 2015;61:1049-62.
28. Tate JR, Sikaris KA, Jones GR, Yen T, Koerbin G, Ryan J, et al. Harmonising adult and paediatric reference intervals in australia and new zealand: an evidence-based approach for establishing a first

panel of chemistry analytes. Clin Biochem Rev. 2014;35:213-35.

29. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Mårtensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. Scand J Clin Lab Invest. 2004;64:271-84.

30. Jansen RTP, Cobbaert CM, Weykamp C, Thelen M. The quest for equivalence of test results: the pilgrimage of the Dutch Calibration 2.000 program for metrological traceability. Clin Chem Lab Med 2018;56:1673-84.

4.3. Supplementary material

Supplemental Table 1. Annual averages and number of results included (n) from the total extracted datasets in 2018 and 2019.

Analytical test	2018		2019	
	Average	n	Average	n
Sodium (mmol/L)	140.1	264449	139.5	280997
Potassium (mmol/L)	4.38	264877	4.34	280691
Chloride (mmol/L)	104	829	103	913
Calcium (mmol/L)	2.40 9.60[a]	60421	2.39 9.59[a]	63780
Magnesium (mmol/L)	0.82 2.0[a]	4783	0.86 2.1[a]	5062
Phosphate (mmol/L)	1.16 3.6[a]	46134	1.16 3.6[a]	48392
Glucose (mmol/L)	5.78 96[a]	441351	5.77 96[a]	483216
ALT (U/L)	20	426509	20	394085

[a]Values in mg/dL.

Supplemental Table 2. Summary of the graph variables of the Bhattacharya reference intervals method calculated by each user using the St Vincent's hospital Spreadsheet (always r2>0.99).

Analytical test	Year	General characteristics				User 1			User 2			User 3			User 4		
		n	Skew	Kurtosis	Central bin	Bin size	# points	# points inside	Bin size	# points	# points inside	Bin size	# points	# points inside	Bin size	# points	# points inside
Sodium	2018	264449	-0.353	13.2	140.2 mmol/L	0.70	10	4.5	0.50	12	2.5	0.70	8	4.5	0.80	10	5.5
	2019	280997	-0.627	4.9	139.5 mmol/L	1.00	7	4.5	0.60	9	4.5	0.40	11	4.0	0.70	9	6.0
Potassium	2018	264877	1.646	68.6	4.36 mmol/L	0.08	10	2.5	0.11	9	4.0	0.23	4	3.0	0.15	6	5.0
	2019	280691	0.334	1.3	4.34 mmol/L	0.08	12	0.0	0.14	7	3.0	0.15	6	4.5	0.15	7	4.5
Chloride	2018	829	-0.054	12.0	103 mmol/L	3.00	3	2.0	2.00	5	2.0	2.00	4	1.0	3.00	3	1.0
	2019	913	-0.884	3.1	103 mmol/L	3.00	3	2.0	3.00	4	2.0	2.00	7	2.5	2.00	4	2.5
Calcium	2018	60421	-0.159	12.6	2.4 mmol/L (9.6 mg/dL)	0.10	11	4.0	0.10	6	4.0	0.20	5	3.5	0.20	6	4.0
	2019	63780	-0.011	3.9	2.4 mmol/L (9.6 mg/dL)	0.10	10	2.5	0.20	6	3.5	0.20	8	2.5	0.20	8	3.5
Magnesium	2018	4783	-1.172	5.6	0.86 mmol/L (2.1 mg/dL)	0.10	6	4.0	0.10	6	4.0	0.10	6	4.0	0.10	6	4.0
	2019	5062	-1.317	4.8	0.86 mmol/L (2.1 mg/dL)	0.10	5	4.0	0.10	6	4.0	0.10	8	5.0	0.10	6	4.0
Phosphate Males 18-50 years	2018	3301	0.296	0.5	1.13 mmol/L (3.5 mg/dL)	0.30	7	4.0	0.40	5	2.5	0.30	7	3.0	0.40	5	2.5
	2019	3398	0.503	1.4	1.13 mmol/L (3.5 mg/dL)	0.30	8	4.0	0.40	5	3.0	0.30	10	4.0	0.30	9	4.0
Phosphate Females 18-50 years	2018	6279	0.209	0.3	1.19 mmol/L (3.7 mg/dL)	0.30	6	3.5	0.30	6	3.0	0.20	6	5.5	0.20	6	6.0
	2019	6693	0.252	0.5	3.6 mg/dL (1.16 mmol/L)	0.30	5	4.0	0.40	5	3.0	0.20	4	4.0	0.30	5	4.0
Phosphate	2018	3074	0.742	4.9	1.07 mmol/L (3.3 mg/dL)	0.30	6	3.5	0.30	6	3.5	0.30	6	3.0	0.30	6	3.5

Males 51-65 years	2019	3047	0.397	0.6	1.07 mmol/L (3.3 mg/dL)	0.30	6	3.0	0.50	4	2.5	0.20	5	3.5	0.40	5	3.0
Phosphate Females 51-65 years	2018	8406	0.17	0.6	1.19 mmol/L (3.7 mg/dL)	0.20	9	4.5	0.20	9	4.5	0.30	6	3.0	0.30	7	4.0
	2019	8680	0.156	0.6	1.19 mmol/L (3.7 mg/dL)	0.20	7	5.0	0.30	7	3.0	0.20	7	5.0	0.30	8	4.0
Phosphate Males >65 years	2018	6641	0.862	5.0	1.03 mmol/L (3.2 mg/dL)	0.20	7	4.5	0.30	7	3.0	0.20	4	4.0	0.20	6	4.0
	2019	6635	0.63	1.9	1.03 mmol/L (3.2 mg/dL)	0.30	6	3.5	0.40	5	2.5	0.20	5	4.0	0.30	7	3.5
Phosphate Females >65 years	2018	18433	0.315	1.9	1.13 mmol/L (3.5 mg/dL)	0.20	8	5.5	0.20	7	5.0	0.20	5	4.0	0.20	8	5.5
	2019	19939	0.399	2.3	1.13 mmol/L (3.5 mg/dL)	0.20	6	4.0	0.30	5	3.0	0.20	7	5.0	0.30	6	3.5
Glucose	2018	441351	1.554	3.8	0.728 [a] (1.984[b])	0.030	7	2.5	0.030	6	2.5	0.021	4	1.5	0.041	4	2.5
	2019	483216	1.590	3.9	0.728 [a] (1.983[b])	0.026	6	3.5	0.031	4	3.0	0.028	4	2.0	0.035	4	2.5
ALT Males 18-50 years	2018	56074	0.863	2.1	1.426[c]	0.086	4	2.5	0.074	4	3.5	0.072	4	4.0	0.190	5	2.5
	2019	51543	0.772	1.5	1.424[c]	0.108	4	3.0	0.074	4	3.5	0.074	4	2.5	0.190	4	2.5
ALT Females 18-50 years	2018	85341	1.329	4.0	1.217[c]	0.130	5	2.5	0.086	4	3.0	0.091	4	3.0	0.130	4	2.5
	2019	83137	1.245	3.7	1.217[c]	0.085	4	3.0	0.085	4	3.0	0.126	4	2.0	0.130	4	2.5
ALT Males 51-65 years	2018	44790	0.705	2.0	1.403[c]	0.089	4	2.0	0.089	5	3.0	0.059	4	4.0	0.091	5	3.0
	2019	40742	0.606	1.5	1.412[c]	0.070	4	3.0	0.072	4	3.0	0.064	4	4.0	0.124	4	2.5
ALT Females 51-65 years	2018	57850	1.002	2.9	1.309[c]	0.089	4	3.0	0.089	4	3.0	0.049	4	3	0.089	4	3.0
	2019	55445	0.993	2.7	1.312[c]	0.086	4	3.0	0.089	4	3.0	0.086	4	3	0.089	4	3.0

ALT Males 66-80 years	2018	46296	0.695	3.0	1.310[c]	0.088	4	2.5	0.092	5	3.0	0.079	4	4.0	0.124	5	3.0
	2019	42523	0.551	2.0	1.318[c]	0.088	4	2.0	0.095	4	3.0	0.073	4	4	0.090	4	2.5
ALT Females 66-80 years	2018	64799	1.105	4.8	1.245[c]	0.086	5	3.5	0.102	4	2.0	0.085	4	3.5	0.117	5	2.0
	2019	61040	0.969	4.0	1.255[c]	0.080	4	2.5	0.080	4	3.0	0.08	4	2.5	0.120	5	2.0
ALT Males >80 years	2018	20790	0.733	3.1	1.189[c]	0.118	4	2.0	0.081	4	3.0	0.130	4	2.0	0.130	4	2.5
	2019	19953	0.611	2.6	1.199[c]	0.091	4	3.0	0.087	5	2.5	0.078	4	3.5	0.130	4	1.5
ALT Females >80 years	2018	42569	0.930	4.3	1.138[c]	0.097	4	2.0	0.098	4	2.5	0.097	4	2.5	0.130	5	2.5
	2019	39702	0.869	3.8	1.153[c]	0.126	4	2.0	0.068	4	3.0	0.034	5	2.5	0.130	4	2.0

[a] Log(10) of average in mmol/L. [b] Log(10) of average in mg/dL. [c] Log(10) of average in U/L

Supplemental Table 3. Reference interval results obtained by four independent laboratory specialists using the Bhattacharya method and currently used reference intervals in Vall d'Hebron laboratory (RIcu). Shaded results were excluded from the final calculation of reference intervals and 95%CI.

Analytical test	RI_{cu}		Year	Bhattacharya (95%CI)	User 1	User 2	User 3	User 4
Sodium (mmol/L)	LRL	136	2018	136.4 (136.3-136.5)	136.3	136.5	136.5	136.3
			2019	135.9 (135.8-136.0)	135.8	136.0	136.0	135.8
	URL	146	2018	144.2 (144.1-144.3)	144.3	144.1	144.1	144.3
			2019	143.3 (143.2-143.4)	143.4	143.2	143.2	143.4
Potassium (mmol/L)	LRL	3.50	2018	3.63 (3.63-3.63)	3.63	3.63	3.55	3.63
			2019	3.62 (3.61-3.62)	3.62	3.61	3.62	3.62
	URL	5.10	2018	5.02 (4.99-5.04)	4.99	5.03	5.11	5.03
			2019	5.02 (5.01-5.02)	4.98	5.01	5.02	5.02
Chloride (mmol/L)	LRL	101	2018	97 (97-97)	97	97	97	97
			2019	97 (97-97)	97	97	97	97
	URL	109	2018	109 (109-109)	109	109	109	109
			2019	109 (109-109)	109	109	109	109
Calcium (mmol/L)	LRL	2.20 8.8[a]	2018	2.20 (2.20-2.20) 8.8 (8.8-8.8)[a]	2.20 8.8[a]	2.20 8.8[a]	2.20 8.8[a]	2.20 8.8[a]
			2019	2.20 (2.20-2.20) 8.8 (8.8-8.8)[a]	2.20 8.8[a]	2.20 8.8[a]	2.20 8.8[a]	2.20 8.8[a]
	URL	2.65 10.6[a]	2018	2.60 (2.60-2.60) 10.4 (10.4-10.4)[a]	2.60 10.4[a]	2.60 10.4[a]	2.60 10.4[a]	2.60 10.4[a]
			2019	2.60 (2.60-2.60) 10.4 (10.4-10.4)[a]	2.60 10.4[a]	2.60 10.4[a]	2.60 10.4[a]	2.60 10.4[a]
Magnesium (mmol/L)	LRL	0.74 M/ 1.8[a] 0.78 F 1.9[a]	2018	0.74 (0.74-0.74) 1.8 (1.8-1.8)[a]	0.74 1.8[a]	0.74 1.8[a]	0.74 1.8[a]	0.74 1.8[a]
			2019	0.74 (0.74-0.74) 1.8 (1.8-1.8)[a]	0.74 1.8[a]	0.74 1.8[a]	0.74 1.8[a]	0.74 1.8[a]
	URL	1.07 M/ 2.6[a] 1.03 F 2.5[a]	2018	0.99 (0.99-0.99) 2.4 (2.4-2.4)[a]	0.99 2.4[a]	0.99 2.4[a]	0.99 2.4[a]	0.99 2.4[a]
			2019	0.99 (0.99-0.99) 2.4 (2.4-2.4)[a]	0.99 2.4[a]	0.99 2.4[a]	0.99 2.4[a]	0.99 2.4[a]
Phosphate (mmol/L) (Males 18-50 years)	LRL	0.81 2.5[a]	2018	0.77 (0.74-0.81) 2.4 (2.3-2.5)[a]	0.74 2.3[a]	0.81 2.5[a]	0.74 2.3[a]	0.81 2.5[a]
			2019	0.74 (0.74-0.77) 2.3 (2.3-2.4)[a]	0.74 2.3[a]	0.74 2.3[a]	0.77 2.4[a]	0.74 2.3[a]
	URL	1.45 4.5[a]	2018	1.48 (1.45-1.51) 4.6 (4.5-4.7)[a]	1.51 4.7[a]	1.45 4.5[a]	1.51 4.7[a]	1.45 4.5[a]
			2019	1.51 (1.51-1.54) 4.7 (4.7-4.8)[a]	1.51 4.7[a]	1.51 4.7[a]	1.54 4.8[a]	1.51 4.7[a]
Phosphate (mmol/L) (Females 18-50 years)	LRL	0.81 2.5[a]	2018	0.87 (0.84-0.87) 2.7 (2.6-2.7)[a]	0.87 2.7[a]	0.87 2.7[a]	0.84 2.6[a]	0.84 2.6[a]
			2019	0.81 (0.77-0.84) 2.5 (2.4-2.6)[a]	0.77 2.4[a]	0.84 2.6[a]	0.84 2.6[a]	0.77 2.4[a]
	URL	1.45 4.5[a]	2018	1.51 (1.48-1.51) 4.7 (4.6-4.7)[a]	1.51 4.7[a]	1.51 4.7[a]	1.48 4.6[a]	1.48 4.6[a]
			2019	1.51 (1.48-1.54) 4.7 (4.6-4.8)[a]	1.54 4.8[a]	1.48 4.6[a]	1.48 4.6[a]	1.54 4.8[a]
Phosphate (mmol/L) (Males 51-65 years)	LRL	0.81 2.5[a]	2018	0.71 (0.71-0.71) 2.2 (2.2-2.2)[a]	0.71 2.2[a]	0.71 2.2[a]	0.71 2.2[a]	0.71 2.2[a]
			2019	0.71 (0.71-0.71) 2.2 (2.2-2.2)[a]	0.71 2.2[a]	0.71 2.2[a]	0.71 2.2[a]	0.71 2.2[a]
	URL	1.45 4.5[a]	2018	1.35 (1.35-1.35) 4.2 (4.2-4.2)[a]	1.35 4.2[a]	1.35 4.2[a]	1.35 4.2[a]	1.35 4.2[a]
			2019	1.35 (1.35-1.35) 4.2 (4.2-4.2)[a]	1.35 4.2[a]	1.35 4.2[a]	1.35 4.2[a]	1.35 4.2[a]

Phosphate (mmol/L) (Females 51-65 years)	LRL	0.81 2.5 [a]	2018	0.87 (0.87-0.87) 2.7 (2.7-2.7) [a]	0.87 2.7 [a]	0.87 2.7 [a]	0.87 2.7 [a]	0.87 2.7 [a]
			2019	0.87 (0.87-0.87) 2.7 (2.7-2.7) [a]	0.87 2.7 [a]	0.87 2.7 [a]	0.87 2.7 [a]	0.87 2.7 [a]
	URL	1.45 4.5 [a]	2018	1.51 (1.51-1.51) 4.7 (4.7-4.7) [a]	1.51 4.7 [a]	1.51 4.7 [a]	1.51 4.7 [a]	1.51 4.7 [a]
			2019	1.51 (1.51-1.51) 4.7 (4.7-4.7) [a]	1.51 4.7 [a]	1.51 4.7 [a]	1.51 4.7 [a]	1.51 4.7 [a]
Phosphate (mmol/L) (Males >65 years)	LRL	0.81 2.5 [a]	2018	0.74 (0.71-0.77) 2.3 (2.2-2.4) [a]	0.77 2.4 [a]	0.71 2.2 [a]	0.77 2.4 [a]	0.71 2.2 [a]
			2019	0.71 (0.68-0.71) 2.2 (2.1-2.2) [a]	0.71 2.2 [a]	0.71 2.2 [a]	0.68 2.1 [a]	0.71 2.2 [a]
	URL	1.45 4.5 [a]	2018	1.32 (1.29-1.35) 4.1 (4.0-4.2) [a]	1.29 4.0 [a]	1.35 4.2 [a]	1.29 4.0 [a]	1.35 4.2 [a]
			2019	1.35 (1.32-1.35) 4.2 (4.1-4.2) [a]	1.35 4.2 [a]	1.35 4.2 [a]	1.32 4.1 [a]	1.35 4.2 [a]
Phosphate (mmol/L) (Females >65 years)	LRL	0.81 2.5 [a]	2018	0.81 (0.81-0.81) 2.5 (2.5-2.5) [a]	0.81 2.5 [a]	0.81 2.5 [a]	0.87 2.7 [a]	0.81 2.5 [a]
			2019	0.87 (0.87-0.87) 2.7 (2.7-2.7) [a]	0.87 2.7 [a]	0.87 2.7 [a]	0.87 2.7 [a]	0.81 2.5 [a]
	URL	1.45 4.5 [a]	2018	1.45 (1.45-1.45) 4.5 (4.5-4.5) [a]	1.45 4.5 [a]	1.45 4.5 [a]	1.39 4.3 [a]	1.45 4.5 [a]
			2019	1.39 (1.39-1.39) 4.3 (4.3-4.3) [a]	1.39 4.3 [a]	1.39 4.3 [a]	1.39 4.3 [a]	1.45 4.5 [a]
Glucose (mmol/L)	LRL	4.1 74 [a]	2018	4.08 (3.72-4.14) 68 (67-69) [a]	4.08 68 [a]	4.08 68 [a]	3.72 67 [a]	4.14 69 [a]
			2019	4.08 (3.72-4.14) 68 (67-69) [a]	4.02 67 [a]	4.08 68 [a]	4.08 68 [a]	4.14 69 [a]
	URL	5.9 106 [a]	2018	6.66 (6.60-6.72) 111 (110-112) [a]	6.72 112 [a]	6.72 112 [a]	6.60 110 [a]	6.66 111 [a]
			2019	6.66 (6.60-6.66) 111 (110-111) [a]	6.66 111 [a]	6.60 110 [a]	6.66 111 [a]	6.60 110 [a]
ALT (IU/L) (Males 18-50 years)	LRL	10	2018	11 (11-12)	11	11	12	11
			2019	10 (10-10)	12	10	10	10
	URL	50	2018	59 (54-64)	48	61	54	62
			2019	62 (60-63)	53	61	61	63
ALT (IU/L) (Females 18-50 years)	LRL	10	2018	7 (7-8)	7	8	5	7
			2019	8 (7-8)	8	8	8	7
	URL	35	2018	31 (30-31)	31	28	31	30
			2019	29 (27-31)	27	27	30	30
ALT (IU/L) (Males 51-65 years)	LRL	10	2018	12 (11-12)	12	12	9	12
			2019	13 (13-14)	13	13	14	11
	URL	50	2018	45 (43-47)	43	46	44	47
			2019	49 (45-53)	45	46	53	52
ALT (IU/L) (Females 51-65 years)	LRL	10	2018	10 (10-10)	10	10	7	10
			2019	11 (10-11)	11	10	11	10
	URL	35	2018	34 (33-34)	34	34	33	34
			2019	34 (34-34)	34	34	34	34
ALT (IU/L) (Males 66-80 years)	LRL	10	2018	9 (8-10)	10	9	8	9
			2019	11 (10-12)	11	10	12	11
	URL	50	2018	40 (36-44)	35	42	44	40
			2019	39 (37-41)	37	42	38	38
ALT (IU/L) (Females 66-80 years)	LRL	10	2018	9 (8-10	8	10	8	9
			2019	10 (9-10)	10	10	10	9
	URL	35	2018	31 (31-31)	31	29	31	31
			2019	31 (31-31)	31	31	31	33
	LRL	10	2018	8 (7-9)	8	9	7	7

ALT (IU/L) (Males >80 years)			2019	8 (7-9)	9	8	9	7
	URL	50	2018	32 (31-34)	29	31	33	33
			2019	30 (29-32)	30	29	32	37
ALT (IU/L) (Females >80 years)	LRL	10	2018	8 (7-8)	8	7	8	7
			2019	7 (7-7)	7	7	7	7
	URL	35	2018	23 (23-24)	23	24	23	29
			2019	24 (22-25)	25	23	22	25

[a]

Values in mg/dL. M: male; F: Female; LRL: Lower Reference Limit; URL: Upper Reference Limit

5. DISCUSSION

5. Discussion

5.1. General background

Clinical laboratories and technology have rapidly evolved in the last decade to give rise to the current situation in which routine analytical tests are automatically processed and results from thousands of samples are delivered within less than 24 hours in a single site (28). In parallel to that, also data analysis strategies have evolved, which has paved the way for innovative laboratory quality assessment strategies. All in all, data science is getting closer to laboratory medicine and the strategies for data reusing will provide new opportunities for science and clinical improvements within the next years.

The definition of big data analytics is basically based on volume (64). According to MeSH, (Medical Subject Headings) big data is defined, since 2019, as "extremely large amounts of data which require rapid and often complex computational analyses to reveal patterns, trends, and associations, relating to various facets of human and non-human entities." In medical specialties, published papers about big data used to have a really large number of individuals and a really large number of variables (64). The main characteristics of big data include 3 v's: volume (size), variety (diversity) and velocity (frequency of update) (65). Some authors also add 4th and 5th v's which are veracity (66) and valorization (67). Therefore, the high-volume test result data produced by clinical laboratories could be classified as big data (68).

The application of data science, defined by MeSH as "an interdisciplinary field involving processes, theories, concepts, tools and technologies, that enable the review, analysis and extraction of valuable knowledge and information from structured and unstructured data", has also started to be fashionable in clinical medicine in the past years and it is predicted to grow fast in the coming years (69), considering the amount of data from healthy individuals available at a clinical laboratory every day. The reuse of laboratory data in particular and of health data in

general has the potential to transform healthcare research into faster and more affordable practices. The real-world data currently available, together with the new technologies for data mining and machine learning, are enabling the discovery of new data patterns that could help answer previously unaddressed questions. Nevertheless, some limitations stated below will have to be overcome before the reuse of data is properly done and applied to clinical practice.

- Harmonization and Standardization of electronic health records: Despite some international efforts (70), this is still a really important challenge not just between countries but also within countries and regions. Harmonization into a common format of health records would be an important improvement of clinical medicine, not just to everyday practice, but also to retrospective research and data quality.
- Data protection: Sensitive information could possibly be found in some collected datasets, therefore anonymization is crucial but could be an important challenge since an individual could sometimes be re-identified by date of birth, sex, postcode or other variables (71). According to the GDPR (General Data Protection Regulation, 2016/679) (72), when conducting scientific research with personal data, organizations must assess the risk of re-identification and implement appropriate measures to minimize that risk.

Reference values comprise the interval in which the results from the 95% of the reference population would lie (2). The reference population is traditionally known as healthy or "normal" population which are two concepts difficult to define (6). The direct method for reference intervals calculation is the one recommended by guidelines and include the laborious work of recruiting at least 120 individuals per partition and consider them as representative of the reference population (2,3). Because of the difficulties related to this process, most laboratories adopt reference intervals from analyser inserts or transfer them from other studies. For all that reasons, during the last years, the interest in laboratory and population specific reference intervals derived from data from the laboratory information systems has been gaining attention

(73). This is called the indirect approach of establishing reference intervals. Several methods have been proposed for this purpose. Some of these methods involve dataset cleaning to identify the reference population (group A), or more complex statistical techniques that use data analysis strategies to determine it (group B). The implementation of common and official protocols for reference interval calculation using indirect methods would allow laboratories to calculate their reference intervals adapted to their own population and methods in a simpler, faster, and cheaper manner.

Two main objectives were covered within this thesis: on the one hand, the comparison of reference intervals results obtained by three existing indirect methods using the same dataset and, on the other hand, the description and simplification of two indirect methods for reference interval calculation that could enable laboratory specialists to calculate reference intervals for their own laboratory, using their own patient data.

5.2. Review of discussion by chapter

In section 3 reference intervals were calculated for 16 biochemistry tests in the Vall d'Hebron dataset using the Dutch indirect method NUMBER (25). This method, after database cleaning using few clinical criteria, considers biochemically related tests for outlier elimination. Then, the reference intervals are calculated as the mean plus/minus two times the standard deviation from the remaining reference population. We found that, for the test results that were normally distributed (albumin, total protein, magnesium, phosphate, calcium, sodium, potassium and chloride), reference intervals were very similar compared to other previous direct and indirect studies (19,54,62,74) and also compared with the results using the TMC method through the RLE (also calculated within the study presented in section 3). The TMC method is an indirect method from group B based on the maximum likelihood estimation of the mean, the standard deviation and the variance from a power normal distribution that includes only the subjects from

the reference population ('healthy' subjects), being detected using the intersection points between a calculated density function.

We also found in section 3 important differences between the Dutch and the Vall d'Hebron reference intervals results for liver enzymes when using NUMBER method. They were found to be generally higher in the adult Dutch population, but comparable in children. Those differences where further analysed for GGT as other liver enzymes were not metrologically comparable. Several hypotheses were discussed for those differences, being the lifestyle component the most probable one. CK and Uric acid are two tests for which both in the Dutch study and in the Vall d'Hebron study the higher limit of the reference interval was much higher than previously reported and that currently being used. It is important to note that some precleaning of the dataset was differently done between both studies as in Vall d'Hebron study icteric, haemolytic and lipemic indexes were considered for all tests and AST results were also considered for CK calculations. Regarding flagging rates, too much flagging was noted when comparing analyser insert reference intervals with indirectly calculated reference intervals. Also, one of the main conclusions from section 3 is the need for: i) IFCC recommended methods traceable to higher order materials and ii) level 1 commutable quality control for trueness verification and bias assessment. When using indirect methods, this would allow the calculation of standardized reference intervals, that are easily transferable and comparable between regions.

In section 4, eight biochemistry laboratory tests were analysed for reference interval calculation by the Bhattacharya method using the Excel Spreadsheet created by St Vincent's hospital and recommended by the IFCC (29). Bhattacharya is one of the classical methods to determine reference intervals using routine laboratory data and it is based on graphical identification of the reference population from the total dataset by applying linear regression to shape a line of best fit for the segment that the user visually chooses as a straight line corresponding to a gaussian distribution of the data. This method is known to require the (subjective) input of the

user, which results in important between user differences in calculated reference intervals. An important reduction of between users' variability when using the tool was found for most tests after applying the criteria defined as part of the study. To note, Bhattacharya is a user dependent method (41) and therefore variability reduction close to zero was not expected. Regarding the presented results, it is important to highlight that despite the significant variability reduction for all tests, ALT was found to be more challenging as extreme values were often found and the distribution of test results were highly asymmetrical. Nonetheless, section 4 illustrates the need for a practical approach and additional instructions for non-statisticians, for calculating reference intervals using indirect methods.

5.3. Common discussion

In table 1 results from the common reference intervals calculated between section 3 (NUMBER and TMC, RLE method) and section 4 (Bhattacharya) are compared. In table 1 colours present, per test and limit, the higher (orange), lower (green) and middle (yellow) results. When the results were equal for all methods, all cells are shown in yellow. Also, considering the middle values as references, the bias ratio (75) is included in brackets to compare the reference intervals between users using between subject biological variation (BV) from the EFLM biological variation database (76).

According to indirect reference intervals calculation theory and based on the idea that database cleaning (based on methods such as outlier elimination) could leave more "pathological" results than statistical detection of the reference population, we would have expected lower high limit values for the methods from group B (Bhattacharya and RLE) than for the method from group A (NUMBER). These differences were expected to be higher for tests with non-gaussian distribution where reference population detection is even more challenging. Nevertheless, a clear common pattern with always higher or lower results for a certain method was not observed. Actually, phosphate results were generally the highest using the RLE method and the

lowest using the NUMBER method contrary to our initial hypothesis. For all other tests we observed a mixed pattern with not one clear method returning higher or lower limits.

Analytical test	RI_{cu} (mmol/l)		Bhattacharya	NUMBER	RLE
Sodium	LRL	136	136	136	136
	URL	146	144	144	144
Potassium	LRL	3.50	3.6	3.6	3.7 ($2.4x10^{-2}$)
	URL	5.10	5.0 ($-2.4x10^{-2}$)	5.1	5.1
Chloride	LRL	101	97 **(-0.77)**	98	99 **(+0.77)**
	URL	109	109 **(+0.77)**	108	108
Calcium	LRL	2.20	2.20	2.20	2.21 ($3.7x10^{-3}$)
	URL	2.65	2.60 ($7.4x10^{-3}$)	2.58	2.57 ($-3.7x10^{-3}$)
Magnesium	LRL	0.74 M/ 0.78 F	0.74	0.72 ($-3.4x10^{-3}$)	0.75 ($1.7x10^{-3}$)
	URL	1.07 M/ 1.03 F	0.99	1.00 ($1.7x10^{-3}$)	0.99
Phosphate (Males 18-50 years)	LRL	0.81	0.77	0.77	0.77
	URL	1.45	1.48	1.51 ($2.8x10^{-3}$)	1.48
Phosphate (Females 18-50 years)	LRL	0.81	0.87 ($9.0x10^{-4}$)	0.84 ($-1.9x10^{-3}$)	0.86
	URL	1.45	1.51	1.51	1.53 ($1.9x10^{-3}$)
Phosphate (Males 51-65 years)	LRL	0.81	0.71 ($-9.0x10^{-4}$)	0.72	0.79 ($6.5x10^{-3}$)
	URL	1.45	1.35 ($-2.8x10^{-3}$)	1.38	1.40 ($1.9x10^{-3}$)
Phosphate (Females 51-65 years)	LRL	0.81	0.87 ($-9.0x10^{-4}$)	0.88	0.96 ($7.5x10^{-3}$)
	URL	1.45	1.51	1.50 ($-9.0x10^{-4}$)	1.52 ($9.0x10^{-4}$)
Phosphate (Males >65 years)	LRL	0.81	0.74	0.71 ($-2.8x10^{-3}$)	0.78 ($3.7x10^{-3}$)
	URL	1.45	1.32 ($-1.9x10^{-3}$)	1.34	1.35 ($9.0x10^{-4}$)
Phosphate (Females >65 years)	LRL	0.81	0.81 ($-2.8x10^{-3}$)	0.84	0.90 ($5.6x10^{-3}$)
	URL	1.45	1.45 (9.0x 10^{-4})	1.43 ($-9.0x10^{-4}$)	1.44
ALT (Males 18-50 years)	LRL	5	11 ($3.4x10^{-2}$)	10	10
	URL	50	59 (0.14)	55	47 (-0.27)
ALT (Females 18-50 years)	LRL	5	7	7	8 ($3.4x10^{-2}$)
	URL	35	31	35 (0.14)	27 (-0.14)
ALT (Males 51-65 years)	LRL	5	12 ($3.4x10^{-2}$)	11	11
	URL	50	45 (-0.17)	51 ($3.4x10^{-2}$)	50
ALT (Females 51-65 years)	LRL	5	10 ($3.4x10^{-2}$)	9	9
	URL	35	34 ($-6.8x10^{-2}$)	42 (0.20)	36
ALT (Males 66 - 80 years)	LRL	5	9	9	10 ($3.4x10^{-2}$)

	URL	50	40	43 (0.10)	39 ($-3.4x10^{-2}$)
ALT (Females 66 - 80 years)	LRL	5	9	8 ($-3.4x10^{-2}$)	9
	URL	35	31	36 (0.17)	30 ($-3.4x10^{-2}$)
ALT (Males >80 years)	LRL	5	8 ($3.4x10^{-2}$)	7	7
	URL	50	32	34 ($6.8x10^{-2}$)	30 ($-6.8x10^{-2}$)
ALT (Females >80 years)	LRL	5	8 ($3.4x10^{-2}$)	6 ($-3.4x10^{-2}$)	7
	URL	35	23 ($-3.4x10^{-2}$)	29 (0.17)	24

ALT: alanine aminotransferase, LRL: Low reference limit, RIcu: Current Reference Intervals, RLE: Reference Limit Estimator, URL: Upper reference limit.
ALT is the only test with non-gaussian distribution considered.

Table 1. Reference interval results from section 3 (NUMBER and Reference Limit Estimator) and section 4 (Bhattacharya) are presented together and compared. Colours present, per test and limit, the higher (orange), lower (green) and middle (yellow) results. When two or the three results are equal for all methods, all cells are showed in yellow. Also, considering the middle values as references, bias ratio is included in brackets to compare between reference intervals using between subject biological variation from EFLM biological variation database. Results higher than the recommended cut-off for the bias ratio of 0.375 are indicated in bold.

Regarding bias ratio calculated using within-subject biological variation and considering the cut-off of 0.375 (75), only chloride would present significant differences between methods. This is due to the relatively low between-subject biological variation of 1.3%. To objectively judge differences in reference intervals obtained by different methods, we recommend assessing the clinical relevance of the differences as well. Currently, official guidelines do not offer a general method for reference interval comparison, but we consider the combination of bias ratio and clinical relevance as an appropriate standard. Nevertheless, bias ratio calculation could be modified if needed either to be more restrictive by using within-subject biological variation instead of between-subject variation or to be less restrictive by adding the contribution of the analytical variation to the equation.

As discussed in both sections, clinical tests that follow a gaussian or close to gaussian distribution have reference intervals results that are very robust and are comparable between methods and studies. Therefore, simple indirect methods would be a good approach for them. Other tests with non-gaussian distribution as liver enzymes or glucose would need more statistically

complex methods. Further studies comparing the same dataset using different indirect methods are needed to understand the clinical implications of new calculated reference intervals and the use of different methods (75,77).

Comparing the three methods regarding their possible current application in clinical laboratories, TMC is statistically more complex and difficult to be applied by non-statisticians. The available tool called RLE is based on both excel and R language and it is not simple to install or use as it works in an old Office version not currently available. We would recommend the authors to update their tool and make it accessible for users not familiar with R. The NUMBER method is described in detail in our manuscript presented in section 3 and the statistics are not complex. Even though, their application and use would depend on how skilled the user is with any statistical program that could also include Microsoft Excel or other spreadsheet software. For this thesis a code in R language for NUMBER calculation was developed that could be used in the future to produce an online free access web application. The Bhattacharya excel spreadsheet is the easiest tool for reference intervals calculation between the methods presented within this work. The tool allows the user to paste their data into a specific place within the excel sheet and then, following certain instructions, calculate the reference intervals. Despite this, choosing the variables for the correct use of the application is challenging due to the complex statistics behind the method. In addition, the already mentioned main drawback from Bhattacharya method is the high between user variation when calculating reference intervals. In this thesis those two issues were treated in section 4, allowing a better understanding and use of the excel spreadsheet by non-statisticians.

5.4. Limitations and strength

One of the limitations from the presented work is that only three of the available indirect methods are presented, tested on a dataset extracted from a single year and laboratory. It would be interesting to explore other methods with the same dataset as well as datasets from different

settings in parallel. Another limitation we found is that the data used within this study was not obtained using IFCC recommended methods nor with calibrators traceable to a higher order material for all clinical tests. In addition, level 1 EQA commutable material was not available in our laboratory during that time. Also, due to the completely anonymous dataset, clinical information was not available for the methods we used and pure statistical approaches were followed.

As our main strength, we found clear evidence that indirect reference intervals calculation overcome most of the limitations from direct methods exposed in the introduction of this thesis (Box 1). The comparison of the three methods using the same dataset highlights the great value of our study for other researchers and laboratory specialists. Also, for laboratories where level 1 EQA commutable control material is not yet available (which was one of the limitations exposed), section 3 presented an alternative method for quality assurance using data from a later time where control material is available. Using that method, prospective studies with data obtained before level 1 EQA material is available can be securely performed. In addition, the quality assurance strategy we introduce using moving averages can be applied by other researchers in the field.

5.5. Next research questions raised within the presented work

In this thesis only biochemistry tests were assessed for reference interval calculation using indirect methods. Reference intervals for other tests such as haematology, endocrinology, immunology or serology would also require to be studied using indirect methods. Several indirect methods have already been used for the calculation of haematology reference intervals (78,79) and thyroid related tests (55,80–83). The methods we propose here can be adapted also for other laboratory tests. Nevertheless, it is important to note that for some tests the calculation will be more challenging due to multiple preanalytical and analytical issues that may hamper standardization as well as comparison of reference intervals between sites. To note as

one of the most important issues, the possible equivocal definition of the measurand due to cross-reactivity, different epitope recognition, insufficient molecular characterization of the compound of interest, inadequate nomenclature, analytical selectivity issues or matrix effects. Those issues also limit the possibilities to combine test results from different sites into one dataset or into one combined analysis. For biochemistry test that is less an issue as, under the auspices of the IFCC Scientific Division, several IFCC endorsed Reference Measurement Systems (RMS) have been established. These RMS are widely used by the IVD-industry for standardization of their commercial tests and are established as the top of the traceability chain to ensure accuracy, precision, and validity of medical test results produced by field methods. Therefore, for biochemistry tests common reference intervals are preferentially deduced from accurate test results produced by standardized tests but may still differ between regions due to differences in lifestyle or other factors. For haematology, endocrinology, immunology, and serology test internationally endorsed RMS are still not in place. In addition, the distribution of some tests may differ from the gaussian/non-gaussian options we presented for the biochemistry tests, for instance if the dataset includes test results lower than the analytical quantification limit what gives a truncated distribution. For those cases, specific adaptations of indirect methods could be proposed. It is also important to note that, in certain cases for assessing particular disease states, clinical decision limits could be more appropriate than reference intervals as interpretative guide limits (4). Despite that, specifications on decision limits were not included within the scope of this thesis.

Currently, official recommendations for the development of reference intervals include mainly the direct method (84)(26). As it was previously outlined in this thesis, this has several disadvantages that precludes the calculation by individual laboratories and reduces the quality of laboratory reports if reference intervals are not appropriate. In addition, laboratories sometimes report a too limited set of reference intervals, not even stratified per age and sex in cases where it is needed. As it was already indicated, some indirect methods even allowed to

calculate continuous reference intervals per age by applying regression and cubic spline techniques (36) which would overcome that problem. In the near future, new guidelines will become available including indirect methods with data reuse. The selection of an appropriate indirect method is still challenging as the implication of selecting a suitable reference population for reference intervals calculation is an important topic for further study. As the incorporation of new indirectly calculated reference intervals will directly impact patient management in current clinical care pathways, it will demand clinical validation studies which demonstrate the benefit for patients (compared to the conventional approach) (85). Likely this can be done in a randomized control trial type of study with head-to-head comparison of the direct reference intervals and the more continuous or personalized indirect reference intervals proposed. The final goal will be evaluating the impact on patient management. In this context, current prospective research studies using information from the hospital information system (medical Electronic Health Records) may also be a possible solution. Natural language processing technologies capable of reading free text from clinical records, structuring it and combining it with laboratory results are promising for studying patient outcomes after test results are performed (86). These types of studies will result in a clearer picture of the relation between a flag per specific test and the immediate and future outcome of patients. Currently, multicentre studies where clinical records are used for scientific purposes are also performed by external companies and clinical research organizations. So another important issue to think off is, who will be responsible of producing and updating the reference intervals if the statistics behind them is complex. It will be the responsibility of the laboratory specialist to create multidisciplinary teams including data scientists and clinicians. In addition, it may be recommended to include statistics and data handling knowledge in the curriculum of laboratory specialists, for them to be able to perform their role as linking pin (data analytics translator) between the data scientists, the laboratory and the clinicians.

Another question arising from this thesis is where personalized medicine will meet laboratory reports. We are still working on generalized models (taking sex and age into account) to calculate population reference intervals based on a reference population that only exists theoretically. As mentioned before, nowadays it is feasible to work with datasets that contain laboratory and clinical information of individuals' follow-up during several years. Some machine learning classification methods such as clustering or random forest could be used to classify patients as healthy or, even in other more specific subcategories (liver disease, pancreatic disease, diabetes…), according to models that can learn from data (69,87). In order to do that, basic predictors that the model can automatically collect from the laboratory or hospital information systems would have to be available for all individuals. Then, based on those predictors, the individualized interpretation of the test result would be automatically generated into the report. The easiest predictors to start with would be age, sex, reason for consultation, previous and current diseases, drugs taken and previous test results. In later stages, other predictors such as the specific ethnic group, genetic variations or lifestyle characteristics could be included to tailor the reference intervals even more, paving the way for precision laboratory medicine. It remains to be seen how this information can become centrally available for reference interval calculation. If that information is automatically available, the model would classify the patient according to its probability of being in the healthy group or in any other group. With such an approach, more personalized theoretical values arising from machine learning or artificial intelligence models would allow the interpretation of test results per person. Conducting research on these issues is challenging because it would require a complete overhaul of the working routine of clinical laboratories and physicians.

Other important key players for adopting new indirect reference interval calculations are the clinicians. Joint validation studies must be designed to evaluate the clinical value of indirectly calculated reference intervals in the clinical care pathways. The impact of newly obtained reference intervals on patient management should be clear ahead and therefore clinicians

should not only be informed but should be part of the validation studies and the subsequent decision-making process. In order to do that, appropriate educational resources will have to be available for them as soon as possible, to make them aware of how the calculations are performed and how the new technologies are changing them.

Finally, the IVDR (In Vitro Diagnostic Medical Devices Regulation, 2017/746) (45) requires manufacturers of IVD devices to establish reference intervals scientifically sound, accurate, and representative of the population for which their devices are intended. They are required to provide clear and detailed information regarding the reference intervals, including the method used to establish the intervals and any limitations or uncertainties associated. They are also required to conduct post-market surveillance on their IVD devices, which includes monitoring the performance of the device over time and updating the reference intervals if necessary. Therefore, it is common for IVD manufacturers to include reference intervals in their product insert usually taken from direct studies or from the literature. That caused problems in the past because, even if the manufacturers are required to provide information on the population and methodology used to establish the reference intervals, this is often not detailed enough and the provided reference intervals are not specific for the population where they are going to be used. As the idea of indirect reference intervals is to be population specific and retrospectively calculated from laboratory data, IVD-manufacturers will have to: first provide, in the product insert, indirectly calculated reference intervals with relevant partitions for a population as close as possible to the population where the intervals are going to be used and, then, provide services for i) reference interval transference when installing new methods into a clinical setting and ii) reference interval calculation using indirect methods when the method is running during an appropriate time, that would depend on how often the test is requested. Currently there is already a guideline (26,88) for reference intervals transference between settings that considers two distinct issues: i) the comparability of analytical systems and ii) the comparability of the test subject population. In the case of IVD manufacturers, recommendations should be done

considering the second issue. As reference intervals calculated by indirect methods are supposed to be comparable with the directly calculated reference intervals, we believe that the current transference methodology can still be applied. Nevertheless, to improve robustness and to take advantage of the computational power, new protocols for reference intervals transference with higher amount of data could also be proposed in the near future.

All in all, specialists in laboratory medicine are responsible for ensuring that the reference intervals they use are appropriate and representative for their population. If the specialists determine that the reference intervals provided by the manufacturer are not accurate, they should update the intervals to ensure that accurate results are provided to patients. Overall, the IVDR sets out strict requirements for the establishment, maintenance and communication of reference intervals for IVD devices, but it does not prevent medical laboratories from updating them if needed.

6. CONCLUSIONS

6. Conclusions

1. Indirect determination of reference intervals is very robust for biochemistry medical tests from population that follow a gaussian or close to gaussian distribution. Therefore, very simple methods (including dataset pre-cleaning and mean/median/standard deviation calculations) would allow for appropriate reference intervals.

2. For medical tests with non-gaussian distributions and long tails, more complex statistical methods seem necessary.

3. To date, there does not seem to be one general appropriate method for indirect determination of reference intervals. According to our results and previous conclusions, test dependent methods would be an appropriate approach.

4. For the calculation of reference intervals, laboratories are recommended to use medical tests that produce test results traceable to IFCC endorsed reference methods with high order material as calibrators and level 1 EQA programs for quality control. Moving averages could be an additional useful tool as longitudinal quality control.

5. Appropriate and easy to use tools and software for indirect reference interval calculation will be necessary in the future to allow laboratory specialists to reliably calculate their own refined reference intervals.

6. The application of the presented indirect methods for reference interval calculation to hematology, endocrinology, immunology, or serology tests will be more challenging due to several issues that hamper test standardization.

7. For the selection of the most appropriate indirect methods, clinical validation studies where the impact of indirect reference intervals on patient management is evaluated, are needed. For that, the inclusion of clinical information would be crucial.

8. Clinicians will be a relevant sounding board group during the design, application and implementation of indirect reference intervals and clinical validation studies.

9. Specific research studies for the calculation of more personalized reference intervals using machine learning are needed even though their application might be highly challenging.

10. The IVD-industry should adapt their product inserts to include beyond direct reference intervals, also indirect reference interval transference and posterior reference interval calculation using indirect methods.

7. REFERENCES

7. References

1. Koerbin G, Sikaris KA, Jones GRD, Ryan J, Reed M, Tate J. Evidence-based approach to harmonised reference intervals. Clin Chim Acta 2014;432:99-107.

2. Gary LH, Sousan A, James CBM, Ceriotti F, Garg U, Horn P, et al. EP28-A3c: Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition. Clin Lab Stand Inst 2010;28:30.

3. Ozarda Y, Ichihara K, Barth JH, Klee G. Protocol and standard operating procedures for common use in a worldwide multicenter study on reference values. Clin Chem Lab Med 2013;51:1027-40.

4. Ozarda Y, Sikaris K, Streichert T, Macri J. Distinguishing reference intervals and clinical decision limits - A review by the IFCC Committee on Reference Intervals and Decision Limits. Crit Rev Clin Lab Sci 2018;55:420–31.

5. Klee GG, Ichihara K, Ozarda Y, Baumann NA, Straseski J, Bryant SC, Wood-Wentz CM. Reference Intervals: Comparison of Calculation Methods and Evaluation of Procedures for Merging Reference Measurements From Two US Medical Centers. Am J Clin Pathol.2018;150:545–54.

6. Gräsbeck, R. The evolution of the reference value concept. Clin Chem Lab Med 2004;42:692–7

7. Expert Panel in the Theory of Reference Values International Federation of Clinical Chemistry Committee on Standards, Gräsbeck, R, Siest, G, Wilding, P, et al. Provisional recommendation on the theory of reference values. Clin Chem 1979;25:1506–8.

8. Aellig, A, Albert, A, Blin, G, Buret, J, Daubrosse, E, Drosdowsky, M, et al. Société Française de Biologie Clinique. Section of physiopathology. Commission “reference values”. Utilisation of reference values. (Document J, stage 3, version 1). Ann Biol Clin 1982;40:697–708.

9. Queralto, JM, Ribo, A, Cortes, M, Domenech, MV, Ferrer, P, Fuentes, J, et al. Documento D: produccion y utilizacion de valores de referencia. Quím Clín 1987;6:49–68.

10. Alström, T, Gräsbeck, R, Lindblad, B, Solberg, HE, Winkel, P, Viinikka, L. Establishing reference values from adults: recommendation on procedures for the preparation of individuals, collection of blood, and handling and storage of specimens. Committee on Reference Values of the Scandinavian Society for Clinical Chemistry. Scand J Clin Lab Invest 1993;53:649–52.

11. Solberg HE. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. Clinica Chimica Acta. 1987;165:111–8.

12. PetitClerc C, Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. Clinica Chimica Acta. 1987;170:S1-11.

13. Solberg HE, PetitClerc C. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. Clinica Chimica Acta. 1988;177:S3-11.

14. Solberg HE, Stamm D. IFCC recommendation - theory of reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. Clinica Chimica Acta. 1991;13:231-4.

15. Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. Clinica Chimica Acta. 1987;170:S13-32.

16. Dybkœr R, Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 6. Presentation of observed values related to reference values. Clinica Chimica Acta. 1987;170:S33-41.

17. Directive 98/79/EC of European Parliament and the Council of 27 October 1998 on in vitro diagnostic medical devices. Offical J Eur Commun 7 December 1998; L331/1-L331/37.

18. International Organization for Standardization (ISO). ISO 15189:2012 Medical laboratories – requirements for quality and competence. London: ISO; 2014.

19. Tate JR, Sikaris KA, Jones GRD, Yen T, Koerbin G, Ryan J, et al. Harmonising adult and paediatric reference intervals in Australia and New Zealand: an evidence-based approach for establishing a first panel of chemistry analytes. Clin Biochem Rev 2014;354:213-35.

20. Flegar-Meštrić Z, Perkov S, Radeljak A. Standardization in laboratory medicine: Adoption of common reference intervals to the Croatian population. World J Methodol 2016;6:93-100.

21. Borai A, Ichihara K, Al Masaud A, Tamimi W, Bahijri S, Armbuster D, et al. Establishment of reference intervals of clinical chemistry analytes for the adult population in Saudi Arabia: a study conducted as a part of the IFCC global study on reference values. Clin Chem Lab Med 2016;54:843-55.

22. Rustad P, Felding P, Franzson L, Kairisto V, Lahti A, Mårtensson A, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. Scand J Clin Lab Invest 2004;64:271-84.

23. Evgina S, Ichihara K, Ruzhanskaya A, Skibo I, Vybornova N, Vasiliev A, et al. Establishing reference intervals for major biochemical analytes for the Russian population: a research conducted as a part of the IFCC global study on reference values. Clin Biochem 2020;81:41-58.

24. Parker ML, Adeli K. Pediatric and adult reference interval harmonization in Canada: an update. Clin Chem Lab Med 2018;57:57-60.

25. Den Elzen WPJ, Brouwer N, Thelen MH, Le Cessie S, Haagen IA, Cobbaert CM. NUMBER: Standardized reference intervals in the Netherlands using a “big data” approach. Clin Chem Lab Med 2018;57:42-56.

26. Henny J, Vassault A, Boursier G, Vukasovic I, Mesko Brguljan P, Lohmander M, et al. Recommendation for the review of biological reference intervals in medical laboratories. Clin Chem Lab Med. 2016;54:1893–900.

27. Horowitz, GL. Reference intervals: practical aspects. EJIFCC 2008;19:95–105.

28. Gonzalez de Buitrago JM. Automatización e informática en los laboratorios clínicos. 1995. Grafiart.

29. Jones, GRD, Haeckel, R, Loh, TP, Sikaris, K, Streichert, T, Katayev, A, et al. Indirect methods for reference interval determination – review and recommendations. Clin Chem Lab Med 2019;57:20–9.

30. Fraser, CG. Biological variation: from principles to practice. Washington, DC: AACC Press; 2001.

31. Özçürümez MK, Haeckel R. Biological variables influencing the estimation of reference limits. Scand J Clin Lab Invest. 2018;78:337–45.

32. Haeckel, R, Wosniok, W, Arzideh, F. A plea for intra-laboratory reference limits. Part 1. General considerations and concepts for determination. Clin Chem Lab Med 2007;45:1033–42.

33. Ichihara, K, Boyd, JC. IFCC Committee on Reference Intervals and Decision Limits (C-RIDL). An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48:1537–51

34. Haeckel R, Wosniok W, Torge A, Junker R. Reference limits of high-sensitive cardiac troponin T indirectly estimated by a new approach applying data mining. A special example for measurands with a relatively high percentage of values at or below the detection limit. Journal of Laboratory Medicine. 2021;45:87–94.

35. Zierk J, Arzideh F, Haeckel R, Cario H, Frühwald MC, Groß HJ, et al. Pediatric reference intervals for alkaline phosphatase. Clin Chem Lab Med. 2017;55:102–10.

36. Asgari S, Higgins V, McCudden C, Adeli K. Continuous reference intervals for 38 biochemical markers in healthy children and adolescents: Comparisons to traditionally partitioned reference intervals. Clin Biochem. 2019;73:82–9.

37. Ichihara K, Ceriotti F, Tam TH, Sueyoshi S, Poon PMK, Thong ML, et al. The Asian project for collaborative derivation of reference intervals: (1) strategy and major results of standardized analytes. Clin Chem Lab Med. 2013;51:1429–42.

38. Sennels HP, Jørgensen HL, Hansen ALS, Goetze JP, Fahrenkrug J. Diurnal variation of hematology parameters in healthy young males: the Bispebjerg study of diurnal variations. Scand J Clin Lab Invest. 2011;71:532–41.

39. Sennels HP, JØrgensen HL, Goetze JP, Fahrenkrug J. Rhythmic 24-hour variations of frequently used clinical biochemical parameters in healthy young males--the Bispebjerg study of diurnal variations. Scand J Clin Lab Invest. 2012;72:287–95.

40. Ghasemi, A, Zahedias, S. Normallity tests for statistical analysis: a guide for non-statistician. Int J Endocrinol Metabol 2012;10:486–9

41. Haeckel R, Wosniok W, Streichert T. Review of potentials and limitations of indirect approaches for estimating reference limits/intervals of quantitative procedures in laboratory medicine. Journal of Laboratory Medicine. 2021;45:35–53.

42.Zierk J, Arzideh F, Haeckel R, Rascher W, Rauh M, Metzler M. Indirect determination of pediatric blood count reference intervals. Clin Chem Lab Med. 2013;51:863–72.

43.Lippi G, Salvagno GL, Lima-Oliveira G, Brocco G, Danese E, Guidi GC. Postural change during venous blood collection is a major source of bias in clinical chemistry testing. Clin Chim Acta. 2015;440:164–8.

44.Plebani M. Harmonization in laboratory medicine: The complete picture. Clin Chem Lab Med. 2013;51:741–51.

45. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU.

46.De la Salle B. Pathology harmony moves on: Progress on implementation in haematology. Br J Haematol. 2012;158:804–5.

47.Jansen, RTP, Cobbaert, CM, Weykamp, C, Thelen, M. The quest for equivalence of test results: the pilgrimage of the Dutch Calibration 2.000 program for metrological traceability. Clin Chem Lab Med 2018;56:1673–84.

48.Cobbaert, C. Time for a holistic approach and standardization education in laboratory medicine. Clin Chem Lab Med 2017;55:311–3.

49.Cobbaert, C, Weykamp, C, Franck, P, de Jonge, R, Kuypers, A, Steigstra, H, et al. Systematic monitoring of standardization and harmonization status with commutable EQA-samples—five year experience from The Netherlands. Clin Chim Acta 2012;414:234–40.

50. Ricós, C, Fernández-Calle, P, Marqués, F, Minchinela, J, Salas, A, Martínez-Bru, C, et al. Impact of implementing a category 1 external quality assurance scheme for monitoring harmonization of clinical laboratories in Spain. Adv Lab Med 2020;1:20200008.

51.Ricós, C, Perich, C, Boned, B, González-Lao, E, Diaz-Garzón, J, Ventura, M, et al. Standardization in laboratory medicine: two years' experience from category 1 EQA programs in Spain. Biochem Med 2019;29:010701.

52. Baadenhuijsen, H, Smit, JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya Procedure. J Clin Chem Clin Biochem 1985;23:829–39.

53. Peng, X, Lv, Y, Feng, G, Peng, Y, Li, Q, Song, W, et al. Algorithm on age partitioning for estimation of reference intervals using clinical laboratory database exemplified with plasma creatinine. Clin Chem Lab Med 2018;56:1514–23.

54.Arzideh, F, Wosniok, W, Gurr, E, Hinsch, W, Schumann, G, Weinstock, N, et al. A plea for intra-laboratory reference limits. Part 2. A bimodal retrospective concept for determining reference limits from intra-laboratory databases demonstrated by catalytic activity concentrations of enzymes. Clin Chem Lab Med 2007;45:1043–57.

55.Inal TC, Serteser M, Coskun A, Özpinar A, Ünsal I. Indirect reference intervals estimated from hospitalized population for thyrotropin and free thyroxine. Croat Med J. 2010;51:124–30.

56. Lo Sasso, B, Vidali, M, Scazzone, C, Agnello, L, Ciaccio, M. Reference interval by the indirect approach of serum thyrotropin (TSH) in a Mediterranean adult population and the association with age and gender. Clin Chem Lab Med 2019;25:1587–94.

57. Iicol YO, Aslan D. Use of total patient data for indirect estimation of reference intervals for 40 clinical chemical analytes in Turkey. Clin Chem Lab Med. 2006;44:867–76.

58. Shine B. Use of routine clinical laboratory data to define reference intervals. Ann Clin Biochem. 2008;45:467–75.

59. Grossi E, Colombo R, Cavuto S, Franzini C. The REALAB project: A new method for the formulation of reference intervals based on current data. Clin Chem. 2005;51:1232–40.

60. Ma C, Cheng X, Xue F, Li X, Yin Y, Wu J, et al. Validation of an approach using only patient big data from clinical laboratories to establish reference intervals for thyroid hormones based on data mining. Clin Biochem. 2020;80:25–30.

61. Oosterhuis WP, Modderman Th. A, Pronk C. Reference values: Bhattacharya or the method proposed by the IFCC? Ann Clin Biochem. 1990;27:359–65.

62. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. Clin Chem Lab Med. 2019;57:1933-47.

63. Jones GR, Barker A, Tate J, Lim CF, Robertson K. The case for common reference intervals. Clin Biochem Rev. 2004;25:99-104.

64. Baro, E, Degoul, S, Beuscart, R, Chazard, E. Toward a literature-driven definition of big data in healthcare. BioMed Res Int 2015;2015:639021.

65. Berger, ML, Doban, V. Big data, advanced analytics and the future of comparative effectiveness research. J Comp Eff Res 2014;3:167–76.

66. Lupşe, OS, Crişan-Vida, M, Stoicu-Tivadar, L, Bernard, E. Supporting diagnosis and treatment in medical care based on big data processing. Stud Health Technol Inf 2014;197:65–9.

67. Dereli, T, Coşkun, Y, Kolker, E, Güner, Ö, Ağirbaşli, M, Özdemir, V. Big Data and ethics review for health systems research in LMICs: understanding risk, uncertainty and ignorance-and catching the black swans?. Am J Bioeth 2014;14:48–50.

68. Tolan, NV, Parnas, ML, Baudhuin, LM, Cervinski, MA, Chan, AS, Holmes, DT, et al. “Big data” in laboratory medicine. Clin Chem 2015;61:1433–40.

69. Gruson, D, Helleputte, T, Rousseau, P, Gruson, D. Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. Clin Biochem 2019;69:1–7.

70. Kim, M, Shin, SY, Kang, M, Yi, BK, Chang, DK. Developing a standardization algorithm for categorical laboratory tests for clinical big data research: retrospective study. JMIR Med Inform 2019;7:e14083.

71. Shine, B, Barth, JH. Big data in clinical biochemistry. Ann Clin Biochem 2019;56:308–9.

72. General Data Protection Regulation (2016/679). Official Journal of the European Union L 119, 04.05.2016.

73. Martinez-Sanchez L, Marques-Garcia F, Ozarda Y, Blanco A, Brouwer N, Canalias F et al. Big data and reference intervals: rationale, current practices, harmonization and standardization prerequisites and future perspectives of indirect determination of reference intervals using routine data. Adv Lab Med 2021;2:9-16.

74. Adeli K, Higgins V, Nieuwesteeg M, Raizman JE, Chen Y, Wong SL, Blais D. Biochemical marker reference values across pediatric, adult, and geriatric ages: establishment of robust pediatric and adult reference intervals on the basis of the Canadian Health Measures Survey. Clin Chem 2015;61:1049-62.

75. Ozarda Y, Ichihara K, Jones G, Streichert T, Ahmadian R. Comparison of reference intervals derived by direct and indirect methods based on compatible datasets obtained in Turkey. Clinica Chimica Acta. 2021;520:186–95.

76. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, et al. The EFLM Biological Variation Database [Internet]. The European Federation of Clinical Chemistry and Laboratory Medicine (EFLM). [cited 2023 Jan 26]. Available from: https://biologicalvariation.eu/.

77. Ichihara K, Ozarda Y, Barth JH, Klee G, Qiu L, Erasmus R, et al. A global multicenter study on reference values: 1. Assessment of methods for derivation and comparison of reference intervals. Clinica Chimica Acta. 2017;467:70–82.

78. Mrosewski I, Dähn T, Hehde J, Kalinowski E, Lindner I, Meyer TM, et al. Indirectly determined hematology reference intervals for pediatric patients in Berlin and Brandenburg. Clin Chem Lab Med. 2021;60:408–32.

79. Bracho FJ. Reference intervals of automated reticulocyte count and immature reticulocyte fraction in a pediatric population. Int J Lab Hematol. 2022;44:461–7.

80. Płaczkowska S, Terpińska M, Piwowar A. Establishing laboratory-specific reference intervals for TSH and fT4 by use of the indirect Hoffman method. PLoS One. 2022;17: e0261715.

81. Zhang G, Liang Q, Yang Y, Leng P, He E, Wang Y. Establishment of Serum TSH Reference Intervals by an Indirect Method for Healthy Population in Southwest China. Clin Lab. 2021;67:2042–7.

82. Milinković N, Ignjatović S, Žarković M, Radosavljević B, Majkić-Singh N. Indirect estimation of reference intervals for thyroid parameters. Clin Lab. 2014;60:1083–9.

83. Mokhtar KM. TSH continuous reference intervals by indirect methods: A Comparisons to Partitioned Reference Intervals. Clin Biochem. 2020;85:53–6.

84. Ozarda Y. Establishing and using reference intervals. Turkish Journal of Biochemistry. 2020;45:1–10.

85. Horvath AR, Lord SJ, StJohn A, Sandberg S, Cobbaert CM, Lorenz S, et al. From biomarkers to medical tests: the changing landscape of test evaluation. Clin Chim Acta. 2014;427:49–57.

86. Hernandez Medrano I, Tello Guijarro J, Belda C, Urena A, Salcedo I, Espinosa-Anke L, et al. Savana: Re-using Electronic Health Records with Artificial Intelligence. International Journal of Interactive Multimedia and Artificial Intelligence. 2018;4:8-12.

87. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019;20:e262–73.

88. Ozarda Y. Reference intervals: Current status, recent developments and future considerations. Biochem Med. 2016;26:5–16.

8. ANNEX

8. Annex

8.1. Abstract presented at congresses

Luisa María Martínez Sánchez, Joan López Hellin, Yolanda Villena Ortiz, Pablo Gabriel Medina, Albert Blanco Grau, Francisco Rodríguez Frías, Christa Cobbaert, Wendy Den Elzen. Copper and Zinc reference intervals calculated from laboratory data. **XIV Congreso Nacional del Laboratorio Clínico 2020.** 8 - 14 November 2020 (virtual).

Copper and Zinc reference intervals calculated from laboratory data

Luisa María Martínez Sánchez[1,2], Joan López Hellin[1], Yolanda Villena Ortiz[1], Pablo Gabriel Medina[1], Albert Blanco Grau[1], Francisco Rodríguez Frías[1], Christa Cobbaert[2], Wendy Den Elzen[2].

1. Hospital Universitario Vall d'Hebron. Laboratorios Clínicos. Servicio de Bioquímica Clínica.
2. Leiden University Medical Center. Leiden (The Netherlands)

Introduction

Indirect methods for reference intervals determination are promising tools for calculating reference intervals using patient data. Copper is a trace element usually determined for the follow up and diagnostic work-up of pathological conditions like Wilson or Menkes disease and both Copper and Zinc are useful to assess nutritional status. Current reference intervals recommended by the Spanish society of clinical chemistry (SEQC) are common for serum and plasma.

Objectives

1) To investigate potential differences between serum and plasma concentrations of copper and zinc
2) To calculate RI using two different indirect methods (Bhattacharya and NUMBER) for copper and zinc, for serum and plasma.

Material and Methods

In the first part of the study, serum (clotting activator, BD Vacutainer®) and plasma (lithium heparin, specific for trace metals S-Monovette®) samples were drawn from 18 individuals and compared for zinc and copper, measured by inductively coupled plasma mass spectrometry. Preanalytical conditions were performed in parallel. Samples were centrifuged (2438g, 10min) and serum/plasma was stored at -20 degrees until analysis. Since a Gaussian distribution was observed for both zinc and copper, paired t-test group comparison studies were performed between serum and plasma. For reference intervals calculation, data from 2018 was extracted from our laboratory information system. Bhattacharya (1) analysis was done on the raw data. For the NUMBER (2) analysis we performed outlier exclusion using the Tukey method and we calculated the central 95% of the population as mean±2SD for the total dataset and by gender. Then, boxplots for men and women separately were visually inspected to decide on final partitioning.

Results

Results from the paired t-test comparison showed significant differences between serum and plasma samples both for copper (difference -4.7 ±4.9ug/dl) and zinc (difference -10.0 ±3.5ug/dl). Observation of boxplots showed that partitioning by gender was necessary for copper and not for zinc, which is consistent with previous studies and recommendations. We did not consider age partitioning in any case. RI results are shown in tables 1 and 2.

Table 1. Calculated reference intervals for copper using indirect methods Bhattacharya (B) and NUMBER (N) (in ug/dl).

Matrix	Gender	Low limit (N)	Low limit (B)	High limit (N)	High limit (B)	n
Plasma	Women/Men	43.9	46.5	111.7	107.1	4418
Serum	Women/Men	72.4	75.5	161.0	157.0	1385

Table 2. Calculated reference intervals for zinc using indirect methods Bhattacharya (B) and NUMBER (N) (in ug/dl).

Matrix	Gender	Low limit (N)	Low limit (B)	High limit (N)	High limit (B)	n
Plasma	Men	47.9	49.5	156.6	155.9	756
Plasma	Women	58.2	66.5	166.9	168.4	841
Serum	Men	56.3	52.8	142.5	148.6	789
Serum	Women	65.8	67.5	157.4	163.0	806

References

(1) C. G. Bhattacharya. A simple method of resolution of a distribution into Gaussian components. Biometrics, Vol. 23, No. 1 (Mar., 1967), pp. 115-135

(2) Den Elzen WPJ, Brouwer N, Thelen MH, Le Cessie S, Haagen IA, Cobbaert CM. NUMBER: Standardized reference intervals in the Netherlands using a "big data" approach. Clin Chem Lab Med 2019;57:42-56.

Conclusions

In our study, different RI were found for copper and zinc for serum and plasma, indicating these reference intervals should be calculated separately for these matrices. The Bhattacharya and NUMBER methods resulted in very similar RI, which both differed from the current Spanish recommendations. Next, to establish reference intervals for copper and zinc in serum and plasma using the indirect method, further studies are needed to explore the effects of pre-analytical conditions (e.g. potential contamination) and analytical conditions (e.g. analytical performance)

Vall d'Hebron Hospital | Vall d'Hebron

Luisa M Martinez, Christa M Cobbaert, Raymond Noordam, Nannette Brouwer, Albert Blanco, Yolanda Villena, Ernesto Casis , Francisco Rodríguez-Frias, Wendy PJ den *Elzen. Indirect determination of biochemistry reference intervals using outpatient data.* **IFCC WorldLab SEOUL 2022 and Dutch NVKC congress.** 26-30 June and 24-26 October 2022.

NUMBER

Nederlandse Uniforme Bilagrenzen en Referentie intervallen

Luisa M Martinez[a,b,c], Christa M Cobbaert[b], Raymond Noordam[d], Nannette Brouwer[e], Albert Blanco[a], Yolanda Villena[a], Ernesto Casis[a], Francisco Rodríguez-Frias[a,c], Wendy PJ den Elzen[e]

a. Clinical Laboratories, Vall d'Hebron University Hospital, Barcelona, Spain.
b. Department of Clinical Chemistry and Laboratory Medicine, Leiden University Medical Center, Leiden, The Netherlands.
c. Department of Biochemistry, Molecular Biology and Biomedicine, Universitat Autònoma de Barcelona, Barcelona, Spain.
d. Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands.
e. Diagnost-IQ, Expert Centre for Clinical Chemistry, Purmerend, The Netherlands.

Indirect determination of biochemistry reference intervals using outpatient data

Introduction

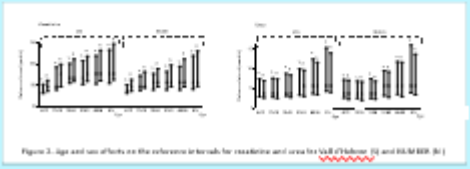
Currently used methods to establish reference intervals (direct) are laborious and expensive. To overcome these drawbacks, new (indirect) methodologies are promising tools. The aim of this study was to validate the indirect Dutch method (NUMBER) with an external database with routine biochemistry test results.

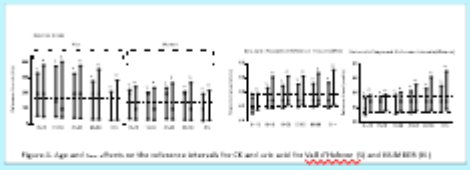
Methods

We used anonymized clinical results from individuals visiting general practitioners and analyzed in 2018 in Clinical Laboratories Vall d'Hebron, Barcelona. Analytical quality was checked by EQA performance and daily average. Per test, outliers were excluded using Tukey method, data were transformed to approximate a normal distribution (if necessary) and reference intervals were calculated for each test, stratified by age and sex, if necessary.

Results

After quality assessment and exclusion based on clinical criteria, we obtained 509,408 clinical requests. The normally distributed tests showed similar results between Barcelona and Dutch population (Table 1). Reference intervals for creatinine (Jaffe method) and urea followed the same tendency of increasing values by increasing age (Figure 2). For ALT, AST, and GGT, markedly higher results for upper limits were obtained in the Dutch population, part of which can be explained by metrological differences (e.g. no use of pyridoxyl 5 phosphate in ALT/AST assays) (Figure 3). As in the Dutch study, creatine kinase and uric acid showed higher reference intervals than traditionally used in Vall d'Hebron (Figure 4). The differences in LD could be explained by the use of the pyruvate to lactate method instead of lactate to pyruvate (IFCC recommended method) (Figure 5).





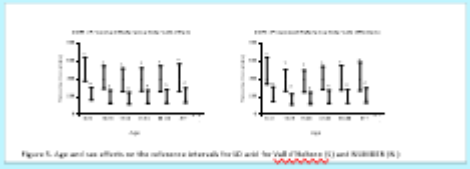


Table 1. The obtained Vall d'Hebron reference intervals for the normally distributed tests, compared with results from the Dutch NUMBER project, stratified for sex and age categories, if necessary.

Test	Unit	Gender	Age, years	Vall d'Hebron RI: Low	Vall d'Hebron RI: High	NUMBER RI: Low	NUMBER RI: High
Albumin	g/L	M	6-18	41	51	40	52
			19-50	40	51	39	51
			51-65	38	49	37	49
			66-80	35	49	36	48
			80+	32	47	36	46
		F	1-5	40	50	39	50
			6 - 18	40	51	40	51
			19-50	37	49	38	49
			51-65	38	48	38	49
			66-80	36	47	37	48
			80+	32	46	36	47
Calcium	mmol/L	M + F	6-12	2,32	2,63	2,29	2,56
			13-18	2,29	2,61	2,23	2,57
			19+	2,20	2,58	2,18	2,55
Chloride	mmol/L	M + F		98	108	97	108
Phosphate	mmol/L	M	13-18	1,16	1,72	0,88	1,53
			19-50	0,77	1,51	0,62	1,32
			51-65	0,72	1,38	0,62	1,32
			66+	0,71	1,34	0,62	1,32
		F	13-18	1,11	1,67	0,82	1,52
			19-50	0,84	1,51	0,73	1,44
			51-65	0,88	1,50	0,73	1,44
			66+	0,84	1,43	0,73	1,44
Potassium	mmol/L	M + F		3,60	5,09	3,88	5,20
Magnesium	mmol/L	M + F		0,72	1,00	0,71	0,98
Sodium	mmol/L	M + F		136	144	136	145
Total protein	g/L	M + F		61	80	61	79

Conclusions

Using an indirect approach, we determined reference intervals for 16 biochemistry tests for the Barcelona population. The reference intervals were compared with Dutch results using the same methodology (NUMBER). Although similar results were found for normally distributed tests, for kidney and liver parameters we found substantial differences which might be explained by methodological, analytical and/or population differences (e.g. lifestyle).

Vall d'Hebron Hospital · Leiden University Medical Center · Diagnost-iq · NVKC · skml · IFCC WorldLab

Contact: wpjdenelzen@lumc.nl

Luisa M Martinez, Nannette Brouwer, Christa M Cobbaert, Joost van Pelt, Jurgen Kooren, Raymond Noordam, David Beneitez, Albert Blanco, Ernesto Casis , Marc H.M. Thelen, Francisco Rodríguez-Frias, Wendy PJ den Elzen. *NUMBER-2: harmonized reference intervals for routine haematology tests in the Netherlands; an indirect data-mining approach.* **IFCC WorldLab SEOUL 2022 and Dutch NVKC congress.** 26-30 June and 24-26 October 2022.

NUMBER

Nederlandse UniforMe BEslisgrenzen en Referentie-intervallen

Luisa M Martinez[a,b], Nannette Brouwer[c], Christa M Cobbaert[b], Joost van Pelt[c], Jurgen Kooren[c], Raymond Noordam[c], David Beneitez[a], Albert Blanco[a], Ernesto Casis[a], Marc H.M. Thelen[b], Francisco Rodríguez-Frias[a], Wendy PJ den Elzen[c]

NUMBER-2: harmonized reference intervals for routine haematology tests in the Netherlands; an indirect data-mining approach

Introduction

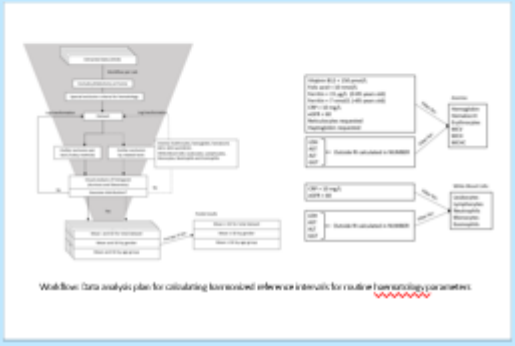
After the first NUMBER (Nederlandse UniforMe BEslisgrenzen en Referentieintervallen) initiative, focussing on chemistry tests, we initiated NUMBER-2 in 2019, to set up a sustainable system for determination and long-term monitoring of harmonized haematology reference intervals in the Netherlands.

Methods

We included medical tests from the Dutch EQA programme 'SKML Hemocytometry'. We extracted anonymous test results from laboratory databases of primary care patients from 8 laboratories using three analytical platforms (Sysmex, Beckman-Coulter, Abbott).

Results were included when EQA analytical performance specifications were met and vitamin B12, folate, ferritin, CRP, eGFR (CKD-EPI), LD, ALT, AST, GGT, and bilirubin were within reference range or not requested (Workflow). Results were excluded when phlebotomy was performed at home.

Per laboratory, per test, outliers were excluded (Tukey method), data were transformed to a normal distribution (if necessary) and means and standard deviations (SDs) were calculated. Then, average means and SDs were calculated to generate pooled (mean±2SD) reference intervals, stratified by age and sex, if necessary.



Workflow: Data analysis plan for calculating harmonized reference intervals for routine haematology parameters

Results

So far, we analysed the data from 8 participating laboratories in the Netherlands using four different platforms (Figure 1) for hemoglobin (normally distributed), hematocrit (normally distributed), erytrocites (normally distributed) and leukocytes (normally distributed after log transformation).

During the first expert meeting, results from hemoglobin, hematocrit, erythrocytes and lymphocytes were discussed. Except for lymphocites (results nor shown), agreement about results was

Figure 1. Distribution of participating laboratories in NUMBER-2

Conclusions

Nationwide harmonization of reference intervals, obtained by this indirect data-mining approach, will enable harmonized data exchange between healthcare systems and help reduce the need for repeated laboratory tests when patients are seen in different care settings.

Vall d'Hebron Hospital | LUMC Leiden University Medical Center | Diagnost-ia | NVKC | skml | IFCC TLM

Contact: wpjdenelzen@lumc.nl