

## Opinion Paper

Graham R.D. Jones\*, Rainer Haeckel, Tze Ping Loh, Ken Sikaris, Thomas Streichert, Alex Katayev, Julian H. Barth and Yesim Ozarda, on behalf of the IFCC Committee on Reference Intervals and Decision Limits

# Indirect methods for reference interval determination – review and recommendations

<https://doi.org/10.1515/cclm-2018-0073>

Received January 19, 2018; accepted March 15, 2018; previously published online April 19, 2018

**Abstract:** Reference intervals are a vital part of the information supplied by clinical laboratories to support interpretation of numerical pathology results such as are produced in clinical chemistry and hematology laboratories. The traditional method for establishing reference intervals, known as the direct approach, is based on collecting samples from members of a preselected reference population, making the measurements and then determining the intervals. An alternative approach is to perform analysis of results generated as part of routine pathology testing and using appropriate statistical techniques to determine reference intervals. This is known as the indirect approach. This paper from a working group of the International Federation of Clinical Chemistry (IFCC) Committee on Reference Intervals and Decision Limits (C-RIDL) aims to summarize current thinking on indirect approaches to reference intervals. The indirect approach has some major potential advantages compared with direct methods. The processes are faster, cheaper and do not involve patient inconvenience, discomfort or the risks

associated with generating new patient health information. Indirect methods also use the same preanalytical and analytical techniques used for patient management and can provide very large numbers for assessment. Limitations to the indirect methods include possible effects of diseased subpopulations on the derived interval. The IFCC C-RIDL aims to encourage the use of indirect methods to establish and verify reference intervals, to promote publication of such intervals with clear explanation of the process used and also to support the development of improved statistical techniques for these studies.

**Keywords:** decision limits; indirect methods; pathology testing; reference intervals; statistical approach.

## Introduction

Quantitative pathology results are generally supported by the provision of reference intervals to aid in interpretation. The concept of reference intervals is now well established and is based on including a fixed percentage of a reference population within the interval described by upper and lower reference limits (RLs). The reference population is generally made up of a statistically significant number of predefined condition-free subjects, but the concept can be applied to any defined population. Generally, it is the responsibility of laboratories to either validate a reference interval derived elsewhere or determine their own interval for use with their population and analytical methods [1]. The recommended process for defining a reference interval is the so-called “direct” approach, where subjects representing the reference population are selected and sampled and the specimen analyzed for this purpose. The reference laboratory document describing this process is the Clinical and Laboratory Standards Institute EP28-A3C [2]. An alternative approach is the “indirect” approach where results from specimens are collected for routine purposes, which have been collected for screening, diagnostic or monitoring purposes and are used to determine the reference intervals.

---

\*Corresponding author: **Graham R.D. Jones**, Department of Chemical Pathology, SydPath, St Vincent’s Hospital, Sydney, NSW, Australia; and University of NSW, Sydney, NSW, Australia, E-mail: [Graham.Jones@svha.org.au](mailto:Graham.Jones@svha.org.au)

**Rainer Haeckel:** Institute for Laboratory Medicine, Klinikum Bremen-Mitte, Bremen, Germany

**Tze Ping Loh:** Department of Laboratory Medicine, National University Hospital, Singapore, Singapore

**Ken Sikaris:** Department of Pathology, Melbourne University, Parkville, Melbourne, Australia; and Sonic Healthcare, Sydney, NSW, Australia

**Thomas Streichert:** Institut für Klinische Chemie, Uniklinik, Cologne, Germany

**Alex Katayev:** Department of Science and Technology, Laboratory Corporation of America Holdings, Elon, NC, USA

**Julian H. Barth:** Blood Sciences, Leeds General Infirmary, Leeds, UK

**Yesim Ozarda:** Department of Medical Biochemistry, Uludag University School of Medicine, Bursa, Turkey

Data mining, or “big data”, is the process of using previously generated data to identify new information. Routine pathology databases often contain many thousands or millions of results from many 100s or 1000s of patients, which can be used in this manner. Using the data for the goal of determining population reference intervals by indirect techniques is one example of data mining [3]. In addition to setting reference intervals, data in pathology databases can be used for internal quality control [4], external quality assessment [5], reference interval validation [6] and determining biological variation data [7–9]. In addition to these factors related to laboratory quality and result reporting, data mining can be used to learn about physiological changes [10], relationships between analytes, effects of interferences, laboratory utilization, epidemiological studies and many other purposes.

This review and opinion document has been prepared by the International Federation of Clinical Chemistry (IFCC) Committee on Reference Intervals and Decision Limits (C-RIDL) and describes various aspects of the indirect approach to setting reference intervals including benefits and risks, available methods and their strengths and weaknesses. It is the opinion of the authors that indirect techniques are highly valuable, either as stand-alone tools or to support other approaches, and that laboratories should be encouraged to use these techniques and further develop tools for use in this area.

## Direct vs. indirect approaches

The traditional approach to establishing reference intervals has been termed the direct approach [2]. In this process, individuals from a population (the reference population) are selected for sampling based on defined criteria. Specimens are then collected from these individuals and analyzed for the selected measurands. This approach has been subdivided into *a priori* and *a posteriori* selection processes. The *a priori* approach is to select individuals for specimen collection and analysis if they meet defined inclusion criteria. In the *a posteriori* approach, specimens collected from a population will be included in the analysis based on other factors such as clinical details or other measurement results, which were not used to define the collection (see Box 1). Thus, in the *a posteriori* approach, not all specimens that were collected would be included in the reference population for further analysis. Ideally, a direct approach would use randomly selected members of the reference population; however, this is rarely achieved with the tested population usually heavily influenced by

### Box 1: Definitions.

---

*Direct sampling.* Selection of results from a reference population by predetermined criteria, which are independent of the measurand of interest

– *A priori* selection is the application of criteria before the collection of the samples

– *A posteriori* selection is the application of criteria after the collection of the samples

*Indirect sampling.* Selection of the results from a mixed population (mixed = containing diseased and non-diseased subjects) to get the results of a predefined reference subpopulation. The selection is performed by statistical tools resolving the distribution of interest from the mixed population

---

convenience and cost factors. True randomization to seek a fully representative group requires extensive (and expensive) planning and implementation, such as was used in the Canadian Health Measures surveys [11]. Another factor with direct approaches is that every selected test result is included in the statistical analysis. This makes outlier exclusion a vital part of the process, although the exclusion process itself can significantly affect the intervals determined [12]. Known limitations for direct studies include difficulty with the definition of health and the prevalence of subclinical disease, as well as selection bias associated with relatively small cohorts.

Indirect approaches are those performed using laboratory results collected for other purposes, usually for routine clinical care, although also for screening or other purposes, where the reference intervals are usually determined by statistical methods based on identifying a distribution in the midst of the data, rather than requiring assessment of all individual results in the database as belonging to the reference population or otherwise. Although standard parametric or non-parametric processes have been used for indirect reference interval studies, these techniques suffer from influence by the more extreme results in a data set, which are also those most likely to be affected by disease.

A vital issue with any reference interval project, using direct or indirect techniques, is an understanding of the factors that influence variations in analyte concentrations. The effects of within- and between-individual variability, analytical and preanalytical variability, physiology and pathology as well as clinical decision making need to be considered when designing studies, interpreting the results and deciding on reference intervals [10]. Establishing reference intervals should not be considered just a “statistical game”, but also requires oversight from experts in laboratory medicine, physiology and disease.

Of note, reference intervals should not be confused with clinical decision limits. Reference intervals are generally considered as a distribution of test values in the pre-defined population, whereas clinical decision limits are mostly determined by assessing the patients' outcomes or response to management changes.

## Comparison of the two approaches: cost and complexity

Direct sampling requires a series of structured steps that together require significant resources [13, 14]. These steps include the following: definition of the reference population; locating/recruiting members of the reference population; obtaining informed consent; sample collection, processing and storage; sample analysis; statistical evaluation (including outlier exclusion); and development of reference intervals for routine use. The processes of identifying subjects, collecting specimens and performing analysis are, at the very least, expensive and time consuming. In some cases, it can be exceedingly difficult, e.g. extremes of age or other than blood/urine sample types. Owing to the resource requirements, it is not uncommon for laboratories to establish reference intervals using fewer than the recommended number of subjects or use other published sources of reference intervals that may not apply to the existing analytical methods and tested populations. It is also rare that studies consider all of the many factors that can influence test results [15]. As it is almost impossible to select a small number of healthy individuals that represent overall biological diversity [15], this may lead to imprecise and inaccurate definition of the reference intervals. However, this limitation may be minimized by more sophisticated statistical/computational approaches such as bootstrapping (repeated analysis of random subsamples of the data set with replacement between sampling) [16].

By contrast, the indirect approach is based on data that have already been generated as part of routine care, thus excluding the resource-intensive components, i.e. patient identification, recruiting, specimen collection and measurement, of the direct approach.

## Benefits and risks/difficulties of indirect approaches

Important benefits of the indirect approach, relative to the direct approach, include that it is faster and cheaper.

It is also based on the actual preanalytical and analytical conditions used in routine practice [17]. Additionally, the reference population is the one from which a patient is actually being distinguished from, i.e. a person presenting to a health care service who does not have the condition under consideration is compared with the person attending for medical care of that condition (See Table A). There are however risks and difficulties associated with indirect approaches. Perhaps the most important risk is the question as to whether the presence of diseased individuals influences the reference intervals. This will depend on the nature of the disease state, i.e. clearly separated or overlapping with the non-disease population, and the relative prevalence in the population.

## Selecting the population

By definition, the population will be derived from one or more routine pathology databases. Before starting any statistical analysis, some basic considerations are necessary to consider which results from the data set should be included.

### Source—inpatient vs. outpatient

If the aim is to produce “health-associated” reference intervals, then results from outpatients are clearly preferred, particularly from those in a primary care setting. The high frequency of inflammation, recumbency, intravenous fluids, medications and dietary changes, in addition to the disease(s) leading to the admission, makes inpatient samples less desirable.

### Number of subjects

There is no prescription for the number of samples required; however, “more is better” to produce robust results. If a data set is composed of nearly all unaffected results and is close to Gaussian, e.g. serum sodium or calcium in a general practice population, smaller numbers can provide reliable estimates. If the underlying distribution is skewed or heavily contaminated, larger numbers are required. If the statistical tool used produces a confidence interval around the derived RL, then it can be assessed whether the limits generated are “close enough”. In the absence of such estimates, assessing

multiple subsets of the data to demonstrate reproducibility can provide supporting evidence. The above comments are qualitative only. However, to provide a starting point for further work in this area, 1000 subjects may be considered a small number and above 10,000 as a large number, and in populations which are poorly represented in a database (e.g. extremes of age), smaller numbers may still provide useful information. It was also recommended [18] to use at minimum of 400 reference subjects for each partition for a statistically reliable reference interval calculation.

## Stability over time

Before using any data set, it is important to ensure that the analytical method and the population have been stable over the period of data collection [19]. The first assessment is an historical review, i.e. has the method been changed or the population serviced changed during the period of data collection. This can be further assessed by reviewing medians and other percentiles over the time period of data collection, and also by the assessment of QC and EQA results. Any changes in these parameters over the time of the data set generation must be investigated and the data limited to a period where the assay performance matches the stable on-going analytical performance using established quality performance goals. The review of at least 1 year of data may also reveal any circannual variation that may need to be considered.

## Consideration of partitioning

Although consideration of possible partitioning should be done based on known effects (e.g. creatinine differences with age and sex), all data sets should be assessed for the effects of readily determinable cofactors such as age and sex. This can be performed by plotting medians and selected centiles (e.g. 10th and 90th) for males and females against patient age. This may reveal a previously unconsidered need for partitioning or confirm the stability across the groups. Analysis against other co-factors such as ethnicity, body mass index (BMI) or alcohol consumption may be considered, although this information is rarely available in routine pathology databases.

The need for partitioning can be assessed by several objective criteria. Harris and Boyd [20] recommends a separate reference interval when the ratio of standard

deviation (larger over smaller) between the subgroups exceed 1.5, or when the Z-statistics between the two subgroup distribution exceeds 3. Alternately, partitioning may be justified when more than 4.1% of a subgroup falls outside of the RL [21]. More details on statistical considerations for partitioning can be found in the paper by Ichihara and Boyd [18].

In addition to considering separate partitions based on continuous variables such as age, it is also possible to generate smoothed reference intervals, which can be applied, for example, to graphical data. This approach to the data may be especially useful in the pediatric age group [22, 23].

## Exclusions

Data sets can be “biochemically filtered” to reduce the frequency of results from subjects where there is a higher likelihood of disease affecting the result. This can be based on other results (e.g. exclude thyroxine results where TSH is outside the reference interval), the location of sample collection (i.e. high altitude residents, inpatients) or supplied clinical information. Depending of the relative frequency of the samples and the nature of the statistical technique, it may not be necessary to exclude specific subgroups (e.g. lipid clinic or renal clinic). As an example, the use of statistical methods using the bulk of data near the centre of the distribution, such as the Hoffman, Bhattacharya or DGKL methods (see below), will be resistant to the inclusion of such groups, but standard parametric or non-parametric methods may be strongly influenced.

An additional recommended approach is to limit results to a single result per patient. As a diseased patient is more likely to be retested than a non-diseased patient, failure to do this is likely to lead to overrepresentation of results from unwell subjects. When selecting the single result, the last result of a patient during a “healthcare episode” (e.g. a hospital admission) is preferred as it is most likely to represent a return towards health. An extension of this approach is to only use results where a single collection has been made from a patient during the period of data collection (“solo” samples). This is based on the assumption that a result considered abnormal by the treating doctor is more likely to be repeated. Other approaches to “improve” the data are to use results from other corequested tests, such as the REALAB project [24], or by linking to clinical databases, which contain patient-specific health information [25, 26].

## Statistical techniques for the indirect approach – descriptions and assessment

### Standard parametric and non-parametric statistics

Standard parametric (mean and standard deviation) or non-parametric statistics (percentiles), such as those used in direct reference interval studies, can also be used for indirect studies. This will involve outlier removal, either before or after transformation, followed by calculation of the mean and SD or median and relevant percentiles. A parametric approach has been used in analysis of data from NHANES [27] and a non-parametric approach in a Turkish study [28].

A key difficulty of standard statistical techniques is the high likelihood (or indeed expectation) of values from diseased individuals in the data set, which has been extracted from the pathology database, to influence reference interval results. As standard statistical techniques are strongly influenced by the extremes of the data set, and these extremes are those most likely to be from affected subjects, great attention needs to be given to outlier removal. There have been a number of examples of approaches to attempt to minimize the presence of results from diseased subjects in database extracts. For example, in the study by Inal et al. [28], outliers were removed by iterative removal of results outside the interquartile range after log transformation.

Other methods have been used to reduce the contamination of the database with results from subjects with disease. As stated above, in the REALAB study, data exclusion was based on related results followed by the use of standard parametric statistics [24]. The latent abnormal values exclusion (LAVE) process, also excluding subjects based on laboratory results, has been used in reference intervals studies by the C-RIDL but has not been tested in indirect processes [29]. In a process linking pathology results with clinical diagnostic codes, Poole and colleagues [26] developed an automated system to identify clinical codes overrepresented in extreme results, and then remove related samples from the data set. After iteratively applying this process, reference intervals were established by standard non-parametric techniques. In general, standard parametric and non-parametric statistical methods are not recommended unless there are validated robust methods for outlier removal and a population with a very low probability of disease, e.g. data collected

at a community screening project or similar. The removal of probable outliers from a data set can be a useful tool, even if more robust statistical processes are used.

### Hoffmann method

The Hoffmann technique was developed in 1963 as a method to identify a homeostatically regulated population subset of test results in a data set that is assumed to follow Gaussian or near-Gaussian distribution. It was developed in the precomputer era for paper-based systems [30]. More recently, this method has been used in a computerized form [31, 32]. A limitation of the Hoffmann procedure is that it is influenced by the presence of a secondary population of significant size [33] although filtering of the data can reduce this effect. The original method used a normal probability paper. A recent revised version dispensed with this requirement [32] although this revised method has been challenged [34]. However, calculated by this method, reference intervals for tests that were expected to be heavily influenced by diseased subjects were reported to be statistically similar to those reported from published statistically robust direct studies [32].

### Bhattacharya method

The Bhattacharya method is also a graphical method for identifying a Gaussian distribution in the midst of other data [35]. Like the Hoffmann method, it was originally developed in the precomputer era using manual paper-based systems. The procedure is able to separate overlapping distributions [33, 35] giving an advantage over Hoffmann in this setting. Computer-based versions have been developed in Java and Microsoft Excel (see Appendix). The Bhattacharya method has been shown to be less influenced by data not included in the Gaussian distribution compared with the Hoffmann method [33]. This method has been subject to review [36–38] and also used in a number of published papers [39–41]. The method is user dependent, requiring selection of bin size for the data, the bin location and the number of bins included in the analysis. Typically, data from four to six bins are used to determine the line of best fit and a high degree of linearity is preferred (e.g.  $r^2 > 0.99$ ).

### Special programs (e.g. DGKL working group)

A more sophisticated procedure than those of Hoffmann and Bhattacharya was developed by Arzideh et al. [19].

In this process, a smoothed kernel density function was estimated for the distribution of the total mixed data of the sample group (combined data of non-diseased and diseased subjects). It was assumed that the “central” part of the distribution of all data represents the non-diseased (“healthy”) population.

The distribution of the non-pathological values is modeled by the power normal (PN) distribution family (Gaussian/truncated Gaussian after using a Box-Cox Transformation function). Thereby, it is assumed that the main (central) part of the data (truncated at the left and right sides), which contains almost only non-pathological values, can be modeled by a truncated PN distribution. The parameters of the PN distribution are estimated using the maximum likelihood method. A goodness-of-fit statistic (a kind of Kolmogorov-Smirnov statistic) is used to find (optimize) the main part of the data. RIs are calculated as the 2.5 and 97.5 percentiles of the estimated PN distribution for the non-pathological value. Clinical decision limits may also be separately computed as the intersection point of the non-pathological and pathological density curves (bimodal RI with the highest diagnostic efficiency).

A software program consisting of an Excel spreadsheet used as a front end and an R-script for the calculations is available from the home page of the German Society of Clinical Chemistry and Laboratory Medicine (see Appendix). This program checks a possible analytical trend during the time of data collection and considers automatic stratification according to sex and age.

## Transformations (log, Box-Cox, other)

The simplest distribution to identify within a mixed population is a Gaussian distribution as demonstrated by the Hoffmann and Bhattacharya methods. When the analyte of interest has a skewed distribution, alternative approaches are needed, generally involving transformation of the distribution to Gaussian. Log transformation is commonly used although distributions may be both more or less skewed than a log distribution, and a selected Box-Cox transformation can provide a better fit. As the derived RIs are dependent on the choice of transformation it is important to have tools to confirm that an appropriate transformation is used. One published approach was to identify best fits of transformed data for both diseased and non-diseased subgroups followed by a Bhattacharya analysis [42], a similar approach to that of Arzideh et al. [19].

A distribution may be skewed for to a number of possible reasons. It may be that in a homogenous, healthy

population, the distribution of an analyte is skewed, as shown by the varying values of the transformation parameter required for Box-Cox transformation in the IFCC multinational study [43]. Many blood hormones (e.g. TSH, LH, insulin) and blood enzymes (e.g. CK, ALP) are inherently skewed, and log transformation is often applied. It may also be that there are distributions overlapping an underlying Gaussian distribution, making it appear skewed. In the latter setting, transformation of the data set may have the effect of including a diseased subset in the population reference interval. An example of this may be the presence of fatty liver on liver enzymes. A lean population without fatty liver may seem close to Gaussian, whereas a population including the overweight and obese will be markedly skewed. This has been shown for liver enzymes by Ichihara et al. [43].

An approach to transformation can be considered as follows: prior to any transformation, seek independent information about the likely nature of the distribution. This may come from previous direct or indirect reference interval studies, especially where possible confounders have been considered. Additionally, the data set under consideration should be reviewed for subpopulations, e.g. based on sex, age, sample collection and handling or other factors, which may contribute to the shape of the distribution. Of primary importance is an understanding of the pathophysiology of the measurand to be aware of confounding factors and how the reference interval may be used in practice. Caution should be exercised when analyzing and interpreting highly skewed distributions that may not be amenable to transformation (e.g. troponin). More work is needed to examine the appropriate statistical treatment/transformation for reliable reference interval derivation. Additionally, some data sets may be based on a high percentage of pathological results, making the use of indirect methods inappropriate. An example of this may be blood gas results where the test is rarely performed in individuals without a high probability of a condition that may affect the results.

## Common reference intervals

In recent years, on behalf of IFCC, C-RIDL has performed direct reference interval studies in many countries to determine global reference intervals. This work is based on a common protocol [44] and the use of a panel of sera to harmonize measurement results [29]. Data mining approaches can support and expand this work. Data mining is also uniquely valuable for the assessment and validation of

common reference intervals. Due to the relatively low cost of data mining approaches, it is possible for multiple laboratories in a region or country to perform a study under the same conditions with the same methods, with the aim of establishing a reference interval to cover a number of laboratories. The indirect approach can be considered particularly valuable as it assesses all aspects of an interval at the same time, i.e. the population, the pre-analytical factors and the analytical factors. Thus, if a working group is considering common reference intervals, an interval can be recommended or confirmed based on data from as many participants in the group as possible, an approach used by the Australian project on common reference intervals [45]. Additionally, a proposed interval can be validated for local use, even if patient numbers are not large, using the midpoint (median) of a distribution, which is the most robust output of a data mining exercise [6].

For smaller laboratories, the use of data from multiple sites can increase the numbers of results included in a reference interval study [46]. Data from multiple sites can also be used to validate the transferability of reference intervals [6], or to investigate the possibility of significant regional differences [47]. One example of such initiative for the pediatric population is the PEDREF study ([www.pedref.org](http://www.pedref.org)).

## Verification of derived reference intervals

It is important that laboratories verify their reference intervals before applying them for routine clinical care. This requirement applies to reference intervals derived using the indirect approach. This can be achieved by the conventional approach where the laboratory analyses samples from 20 subjects without the predefined condition in the reference population. The reference intervals is considered verified if two or less results out of 20 fall outside of the reference intervals that would correspond to a 95% probability [2].

Alternately, laboratories can assess if the given reference interval is appropriate for their testing patient population and analytical method by monitoring the percentage of abnormal results (that would be typically flagged by the laboratory information system) and comparing it with the expected percentage that may be easily derived from the original indirect study calculations. When a change in the flagging rate in any direction (increased or decreased) does not exceed a predefined expected value, the reference interval under evaluation is acceptable for use. This

method does not require additional patient testing and may be programmed in the laboratory information system as a continuous quality control monitoring measure.

## Ethical considerations

It is a requirement for clinical laboratories to provide reference intervals with numerical results [1]. As with any processes involving patient data, it is important to consider any relevant ethical issues related to the process. In the case of indirect reference intervals, as there is no direct intervention with the patient, the key considerations are to ensure security and privacy of the patient results and to consider whether the use of patient data in this way may be likely to be contrary to patient wishes when the result may be linked to the original subject. As a starting point, consideration can be given to obtaining patient consent. This is essentially impossible due to the large numbers involved and the potential lack of access to patient contact details. If a consent process was required, the process would have to run throughout the period of sample collection and the issues of informing subjects and recording responses would not be inconsiderable. To avoid this situation, all data extraction and handling should be done in a way where the identity of the patient is unknown, even to the person performing the analysis. Some basic patient demographics may be required such as age, sex and BMI together with the laboratory results, but the risk of identifying an individual from this data would be very low. Just as importantly the analysts should be aware of their obligations in this area. A unique patient identifier may be required if limitations are to be made to a single result from a subject and a unique code (e.g. medical record number) is preferred to using recognizable identifiers. The nature of data mining is that the outcome is based on the distribution of results rather than individual results; however, care should be taken that no individual can be identified based on any report or publication based on the data. A laboratory should comply with local ethical and privacy requirements when performing such an analysis. Some general principles regarding ethical considerations in laboratory medicine practice and a useful guide on de-identification of patient data can be found in the paper by Burnett et al. [48] and the United States Code of Federal Regulation [49].

When compared with direct approaches to determining reference intervals, indirect approaches remove any issues of possible patient harm, e.g. bruising from sample collection, consumption of subjects' time, the use of

**Box 2:** Minimum requirements for publication of indirect reference interval studies.

- 
1. Details of study design, specifically stating the use of the indirect approach
  2. Description of population: age distribution, sex distribution, source (hospital, GP, other) and source (cities, regions)
  3. Description of available records of preanalytical processes: patient preparation, sampling time distribution, sitting or recumbent, sample type, sample processing and any storage timing
  4. Descriptions of analytical processes: method principle, precision, measurement traceability/trueness, relevant information on analytical specificity, manufacturer kit name (for kit assays); precision and accuracy information (i.e. internal QC and external QA) should cover the period of time over which data has been collected
  5. Description of any data set selection and filtering criteria: exclusions based on age, sex, pregnancy, number of samples collected from patient, presence of other tests corequested, results of other tests corequested, clinical notes or linkage to clinical databases and collection sites
  6. Description of data set: number of samples, median and other percentile values (where appropriate), kurtosis, initial analysis of partitioning, e.g. on age or sex
  7. Description of statistical process. Include outlier detection, method and transformations
  8. Results of statistical analysis: midpoint, upper and lower RIs, uncertainty of estimates (where possible) and information of “goodness of fit” for any model. This should be provided for any partitions of the data set
  9. Comparison with other statistically reliable peer-reviewed published studies (if available)
  10. Final recommendations and discussion of study
- 

resources to collect and measure samples and identification of unexpected results, which which may negatively affect a subject. With this in mind, indirect approaches can be seen to have considerably fewer ethical issues than direct methods, and suitable, fast track ethical approvals or exemptions should be sought.

## Recommendations for publications

It is important that data analysis is shared, and thus publication of reference interval studies, both direct and indirect, is recommended. It is also important that all reference interval studies are described in sufficient detail to allow a reader to fully understand the process and identify any weaknesses or strengths in a study. To this end, a checklist for publication has been developed to support authors, editors and reviewer in this process. It can be seen that many of these steps are analogous to those in the STARD guidelines for reporting diagnostic accuracy studies [50] (see Box 2).

## Conclusions

The methods and processes for determination of reference intervals using indirect methods have been in development for over 50 years. It is the belief of the IFCC C-RIDL that this approach is not only a useful adjunct to traditional direct methods but also has a number of significant benefits and advantages. These advantages include basing the outcomes on the analytical and preanalytical

procedures in use, the ability to address a wide range of populations, especially the population served by the laboratory, and importantly the relative ease and far lower costs. The processes need to be done with care and with due consideration for physiology, pathology and use of appropriate statistics. Laboratorians are encouraged to use indirect methods to evaluate their reference intervals in use, to estimate new reference intervals and to publish and share their results in appropriate detail and also to continue the search for new and improved techniques for the process.

It is also important to understand that no reference interval is absolutely accurate and is only estimation. It has inherited uncertainties and assumptions that may or may not be true. Once a second sample is collected, a comparison with the previous result may be more important than comparison with the RI. Each individual patient should be assessed using all available clinical and laboratory data. Clinicians should realize that test result is not an absolute number but rather a range that is determined by a combination of analytical and biological variations.

**Author contributions:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

**Research funding:** None declared.

**Employment or leadership:** None declared.

**Honorarium:** None declared.

**Competing interests:** The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.



## Appendix: Available computer programs for indirect reference intervals estimation

Bellview Bhattacharya analysis. Java Application. By Doug Chesher. Cross platform application designed to simplify the process of importing data and analyzing data using the Bhattacharya method. <https://sourceforge.net/projects/bellview/>.

Reference Limit Estimator. Excel with R-program. By DGKL Working Group for Guide Limits. Detailed installation and use instructions available to download (in German). [http://www.dgkl.de/PA106975\\_DE\\_VAR100?sid=n443D57v68w211](http://www.dgkl.de/PA106975_DE_VAR100?sid=n443D57v68w211).

Bhattacharya Spreadsheet. Excel application. By Graham Jones. Detailed use instructions included in spreadsheet. <http://www.sydpath.stvincents.com.au/>.

**Table A:** Comparison of direct and indirect methods for reference interval determination.

Direct	Indirect
– Difficult to get spread of subjects by age and sex	– Spread of ages and sex match the clinical use of the test
– Difficult to obtain uncommon sample types	– Data for uncommon sample types are readily available in the laboratory database
– Difficult and expensive to get statistically significant numbers	– Significant numbers readily available
– Difficult to define “healthy” status	– Defining “health” is not required
– Preanalytical conditions may not match routine conditions	– Preanalytical conditions match routine conditions
– Analytical conditions may not match routine conditions	– Analytical conditions match routine conditions
– Costs of performing the study are high	– Costs of performing the study are very low
– Hard to repeat in different locations	– Easy to repeat in different locations
– Hard to repeat with analytical changes	– Easy to repeat with analytical changes
– Ethical issues with sample collection and responses to new information identified on patient, obtaining informed consent may be difficult	– No ethical issues with sample collection and no new information identified on patient
– Statistically more easy to perform	– Requires reasonably robust statistical knowledge and expertise. Some of the statistical analysis may require subjective assessment

## References

- ISO15189:2012. Medical laboratories – requirements for quality and competence.
- CLSI EP28-A3C. Defining, establishing, and verifying reference intervals in the clinical laboratory, 3rd ed. Clinical & Laboratory Standards Institute, 2010.
- Solberg HE. Using a hospitalized population to establish reference intervals: pros and cons. *Clin Chem* 1994;40:2205–6.
- Fleming JK, Katayev A. Changing the paradigm of laboratory quality control through implementation of real-time test results monitoring: for patients by patients. *Clin Biochem* 2015;48:508–13.
- De Grande LA, Goossens K, Van Uytendange K, Stöckl D, Thienpont LM. The Empower project – a new way of assessing and monitoring test comparability and stability. *Clin Chem Lab Med* 2015;53:197–204.
- Jones GR. Validating common reference intervals in routine laboratories. *Clin Chim Acta* 2014;432:119–21.
- Cembrowski GS, Tran DV, Higgins TN. The use of serial patient blood gas, electrolyte and glucose results to derive biologic variation: a new tool to assess the acceptability of intensive care unit testing. *Clin Chem Lab Med* 2010;48:1447–54.
- Loh TP, Ranieri E, Metz MP. Derivation of pediatric within-individual biological variation by indirect sampling method: an LMS approach. *Am J Clin Pathol* 2014;142:657–63.
- Loh TP, Metz MP. Indirect estimation of pediatric between-individual biological variation data for 22 common serum biochemistries. *Am J Clin Pathol* 2015;143:683–93.
- Sikaris K. Physiology and its importance for reference intervals. *Clin Biochem Rev* 2014;35:3–14.
- Adeli K, Raizman JE, Chen Y, Higgins V, Nieuwesteeg M, Abdelhaleem M, et al. Complex biological profile of hematologic markers 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;61:1075–86.
- Hickman PE, Koerbin G, Potter JM, Abhayaratna WP. Statistical considerations for determining high-sensitivity cardiac troponin reference intervals. *Clin Biochem* 2017;50:502–5.
- PetitClerc C, Solberg HE. IFCC approved recommendation on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *Clin Chim Acta* 1987;170:S3–12.
- Solberg HE, PetitClerc C. IFCC approved recommendation on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *Clin Chim Acta* 1988;177:S1–12.
- Ritchie RF, Palomaki G. Selecting clinically relevant populations for reference intervals. *Clin Chem Lab Med* 2004;42:702–9.
- Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem* 1998;44:622–31.
- Barth JH. Reference ranges still need further clarity. *Ann Clin Biochem* 2009;46:1–2.
- 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.

19. 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.
20. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clin Chem* 1990;36:265–70.
21. Lahti A, Hyltoft Petersen P, Boyd JC, Fraser CG, Jørgensen N. Objective criteria for partitioning Gaussian-distributed reference values into subgroups. *Clin Chem* 2002;48:338–52.
22. Zierk J, Arzideh F, Haeckel R, Cario H, Frühwald MC, Groß H-J, et al. Pediatric reference intervals for alkaline phosphatase. *Clin Chem Lab Med* 2017;55:102–10.
23. Yang Q, Lew HY, Peh RH, Metz MP, Loh TP. An automated and objective method for age partitioning of reference intervals based on continuous centile curves. *Pathology* 2016;48:581–5.
24. 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.
25. Kouri T, Kairisto V, Virtanen A, Uusipaikka E, Rajamaki A, Finneman H, et al. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. *Clin Chem* 1994;40:2209–15.
26. Poole S, Schroeder LF, Shah N. An unsupervised learning method to identify reference intervals from a clinical database. *J Biomed Inform* 2016;59:276–84.
27. Lacher DA, Hughes JP, Carroll MD. Biological variation of laboratory analytes based on the 1999–2002 National Health and Nutrition Examination Survey. National health statistics reports; no 21. Hyattsville, MD: National Center for Health Statistics, 2010.
28. Inal TC, Serteser M, Coşkun A, Özpınar A, Ünsal I. Indirect reference intervals estimated from hospitalized population for thyrotropin and free thyroxine. *Clin Sci* 2010;51:124–30.
29. Ichihara K, Ozarda Y, Barth JH, Klee G, Qiu L, Erasmus R, et al. A global multicentre study on reference values: 1. Assessment of methods for derivation and comparison of reference intervals. *Clin Chim Acta* 2017;467:70–82.
30. Hoffmann RG. Statistics in the practice of medicine. *J Am Med Assoc* 1963;185:864–73.
31. Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory results. Is there a better way? *Am J Clin Pathol* 2010;133:180–6.
32. Katayev A, Fleming JK, Luo D, Fisher AH, Sharp TM. Reference intervals data mining no longer a probability paper method. *Am J Clin Pathol* 2015;143:134–42.
33. Grindler ME. Calculation of normal ranges by methods used for resolution of overlapping Gaussian distributions. *Clin Chem* 1970;16:24–8.
34. Jones GR, Horowitz G. Data mining for reference intervals – getting the right paper. *Am J Clin Pathol* 2015;144:1–3.
35. Bhattacharya CG. A simple method of resolution of a distribution into Gaussian components. *J Biometric Soc* 1967;23:115–35.
36. 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.
37. 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;7:20–30.
38. Oosterhuis WP, Modderman TA, Pronk C. Reference values: Bhattacharya or the method proposed by the IFCC. *Ann Clin Biochem* 1990;27:359–65.
39. Pottel H, Vrydags N, Mahieu B, Vandewynckele E, Croes K, Martens F. Establishing age/sex related serum creatinine reference intervals from hospital laboratory data based on different statistical methods. *Clin Chim Acta* 2008;396:49–55.
40. Hoffmann JJ, van den Broek NM, Curvers JC. Reference intervals of reticulated platelets and other platelet parameters and their associations. *Arch Pathol Lab Med* 2013;137:1635–40.
41. Farrell CJ, Nguyen L, Carter AC. Data mining for age-related TSH reference intervals in adulthood. *Clin Chem Lab Med* 2017;55:e213–5.
42. Concordet D, Geffre A, Braun J-P, Trumel C. A new approach for the determination of reference intervals from hospital-based data. *Clin Chim Acta* 2009;405:43–8.
43. Ichihara K, Ozarda Y, Barth JH, Klee G, Shimizu Y, Xia L, et al. A global multicentre study on reference values: 2. Exploration of sources of variation across the countries. *Clin Chim Acta* 2017;467:83–97.
44. Ozarda Y, Ichihara K, Barth JH, Klee G; on behalf of the CRIDL on Reference intervals and Decision Limits, IFCC. Protocol and standard operating procedures for common use in worldwide multicenter study on reference values. *Clin Chem Lab Med* 2013;51:1027–40.
45. 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.
46. Arzideh F, Brandhorst G, Gurr E, Hinsch W, Hoff T, Roggenbuck L, et al. An improved indirect approach for determining reference limits from intra-laboratory data bases exemplified by concentrations of electrolytes. *J Lab Med* 2009;33:52–66.
47. Arzideh F, Wosniok W, Haeckel R. Reference limits of plasma and serum creatinine concentrations from intra-laboratory data bases of several German and Italian medical centres: comparison between direct and indirect procedures. *Clin Chim Acta* 2010;411:215–21.
48. Burnett L, McQueen MJ, Jonsson JJ, Torricelli F, IFCC Taskforce on Ethics. IFCC position paper: report of the IFCC taskforce on ethics: introduction and framework. *Clin Chem Lab Med* 2007;45:1098–104.
49. CFR 164.514 – Other requirements relating to uses and disclosures of protected health information. (<https://www.law.cornell.edu/cfr/text/45/164.514>).
50. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.