Reference Intervals

***Graham Jones,[1,2] Antony Barker[3]**
[1]Department of Chemical Pathology, St Vincent's Hospital, Sydney, Victoria St, Darlinghurst, NSW 2010, [2]Faculty of Medicine, University of NSW, Randwick, NSW 2052, Australia, [3] LabPlus, Auckland City Hospital, Auckland, 1148, New Zealand.
*For correspondence: Dr Graham Jones e-mail: gjones@stvincents.com.au

Summary
Recommended elements of a process for establishing a reference interval:

- Define the analyte (measurand) for which the reference interval is being established, the clinical utility, biological variation and major variations in form.
- Define the method used, the accuracy base, and analytical specificity.
- Define important pre-analytical considerations together with any actions in response to the interference.
- Define the principle behind the reference interval (i.e. central 95% etc.)
- Describe the data source(s), including: number of subjects, nature of subjects, exclusions, pre-analytical factors, statistical measures, outliers excluded and analytical method.
- Define considerations of partitioning based on age, sex etc.
- Define the number of significant figures, i.e. the degree of rounding.
- Define the clinical relevance of the reference limits.
- Consider the use of common reference intervals.
- Decision and implementation.

Introduction
Reference intervals are the most common decision support tool used for interpretation of numerical pathology reports. As laboratory results may be interpreted by comparison with these intervals, the quality of the reference intervals can play as large a role in result interpretation as the quality of the result itself.

The recommended protocol for setting a reference interval is to perform a reference interval study according to standard published procedures.[1-3] In practice it is impossible for any single laboratory to perform these studies for all of the tests it performs to an appropriately high standard. We therefore need to consider alternative processes of obtaining data for setting reference intervals. This data may come from the literature, manufacturers, data mining, or other laboratories. Analytical factors and pre-analytical factors need to be considered, along with partitioning on the basis of sex or age, either as part of a reference interval study or when interpreting other studies. The number of significant figures for the reference limits, the possible adoption of common reference intervals and clinical consultation also need consideration after the data has been obtained.

It is essential that the process used for setting the reference interval is clearly documented, including summaries of the data on which the interval is based and the people involved in setting the interval. This is now clearly stated in the National Association of Testing Authorities, Australia (NATA) Field Application Document for ISO 15189 section 5.5.5[4] and reads as follows:

The sources of biological reference intervals and/or medical decision points must be documented and should include references to the information used in deciding the intervals, any statistical processes used, literature studies considered and the personnel involved in deciding the intervals. Where possible and relevant, customers of the laboratory with appropriate expertise should also be involved in the determination of reference intervals. Consideration should be given to adopting intervals/decision points consistent with those in other laboratories, where possible and appropriate.

It is also a requirement of ISO 15189[5] under the same section heading that (biological) reference intervals shall be reviewed periodically, and whenever a particular interval is thought to be no longer appropriate, or where a pre-examination or

examination (analytical) procedure is changed. The aim of this paper is to expand on the key factors which should be considered when setting a reference interval and then documented as part of the quality system.

Analyte Description

It is important to define the analyte (measurand) for which the reference interval is being established, and the common reasons for measurement of the analyte, in order to ensure that any intervals are appropriate for those purposes. Data on within- and between-person biological variability can also be useful background information. Important biological forms should be briefly described and their relevance considered. For example when considering setting a reference interval for serum prolactin, an awareness of macro-prolactin is required and consideration as to whether it is desirable or possible to exclude samples containing this form from the interval setting process. The analyte description should also include the units to be used and it may be useful to provide conversion factors from other units which may be encountered in the literature.

Documentation Requirements

There should be a brief description of the analyte, clinical utility, biological variation, and major variations in forms. This information may be already held in other documents such as method manuals and may only require cross referencing rather than repetition.

Methods

The method used in the laboratory to produce patient results is an important factor in setting reference intervals. If data are obtained from the literature or other laboratories the relationship between the methods must be understood. The key factors are the accuracy base and analytical specificity. The accuracy base may be described as traceability to a method or reference material and ideally this would be expressed in a quantitative manner. The analytical specificity is important to ensure that data from different methods are appropriately compared. When a local reference interval study is performed attention should be given to ensuring the assay is performing as specified and the accuracy can be maintained over time. In general, assay precision has a smaller effect than bias on setting reference intervals, especially when the precision is small compared to the combined within and between-person biological variation. It is however appropriate to use an assay in setting reference intervals with the same performance characteristics used for routine practice.

Documentation Requirements

There should be a description of the method used in the laboratory, the accuracy base used for that assay, and relevant issues regarding analytical specificity and, if possible, evidence that the assay is working as specified by the manufacturer and remains stable over time.

Pre-Analytical Factors

Some pre-analytical factors will affect results and so should be considered when performing reference interval studies, reviewing the literature, or when applying the intervals to patient results. Examples include serum versus heparin plasma when measuring potassium or total protein; time of day for collection for serum cortisol or testosterone; sample handling, such as time until centrifugation for potassium measurement; and common interferences, such as haemolysis for potassium, CK, AST and LD. If relevant, these factors should be considered when performing reference interval studies or assessing data from other sources and ideally matched with the routine practice for the laboratory.

Documentation Requirements

Any important pre-analytical considerations should be documented together with any actions in response to the interference.

Reference Interval Principle

The reference interval principle is usually to take the central 95% of a reference population.[1,2] There are accepted variations from this principle such as the 99th centile of a healthy population for cardiac specific troponins,[6] the glucose concentration associated with risk for the development of diabetes and of macrovascular disease,[7] or therapeutic intervals for therapeutic drug monitoring. It is necessary to review the literature for any current and relevant guidelines that may recommend the adoption of an interval or decision point other than a statistically derived population interval.

The principles used in defining a reference interval should be available to the user of the test, particularly when factors other than a healthy population distribution are used, or where there are important limitations with regard to age or other factors such as pregnancy or menopause.

Documentation Requirements

The principle behind the reference interval should be documented and available to clinicians if required.

Data Sources

The key data source for setting a reference interval is the reference interval study performed according to the Clinical and Laboratory Standards Institute (CLSI) and IFCC recommendations.[1,2] Ideally this should be performed by the laboratory establishing the reference interval. However data from one study is rarely sufficient and in any case should be compared with data from other sources. Other such data sources include peer-reviewed literature, posters and meeting

abstracts, manufacturer's information, and unpublished data from other laboratories. Data mining studies, for example using the Bhattacharya method to extract information from large patient result databases,[8] may also provide valuable information.

When performing a reference interval study there are several key factors to be addressed. Firstly, the subjects being tested (the reference population) should be as similar as possible to that for which the test will be applied, with the exception of the presence of disease. While some tests may be significantly different due to racial[9] or environmental factors, the main effects to consider are age, sex and common factors such as obesity or diabetes. It may be appropriate to partition for these factors (see below) but consideration should be given to the population where the test is likely to be used. For example the use of university students to set a reference interval for BNP, a test most commonly used in the geriatric population, may produce an inappropriate interval. The pre-analytical factors (e.g. patient preparation and sample handling) should reflect the usual practice in the laboratory. Special handling for the study, e.g. very rapid separation of the plasma, may bias a reference interval compared to normal practice.

It is commonly stated that 120 is the number of data points needed for a reference interval study. This is the number needed to calculate the 90% confidence limit of the upper and lower reference limits when using non-parametric statistics. It is important to calculate this range for any reference interval study as this is a measure of the "experimental error" of the study due to the statistics. This can highlight the need for more data, the significance of a difference between similar studies, and provide guidance on the number of significant figures to which a reference interval limit should be reported. More data will give better certainty about the upper and lower reference limits but there is no simple way to know how many subjects are enough. Be aware also that this consideration also applies to subgroups being considered for separate intervals, e.g. 120 each for males and females or different age groups.

It should be recognised that even well-designed, large reference interval studies do not provide all the required information with regard to setting reference intervals. There are subjective factors in the design and interpretation, such as the choice of population, the numbers included, the statistical techniques used, and the method of outlier exclusion which require professional judgement and may be done differently in different locations, even when the same data set is being considered.

Other sources of reference interval studies should be assessed even if a local study has been performed. Where possible, studies using the same analytical method should be identified although results from studies using other methods can be useful if the relationship between the methods is understood. Note that the use of Google Scholar[10] can be a useful addition to other search engines as the full text can be searched for proprietary words such as Centaur or ElecSys which are not search terms on systems such as Medline. Reviewing other studies allows confirmation of the local results or may identify a problem with the local study. If a difference is seen between studies, as commonly happens, careful consideration should be given to the factors mentioned above to try and identify the cause. If no cause can be identified it is useful to assess the results from your assay in routine use to examine the effect of a proposed reference interval on the abnormal flagging rate.

If it is not possible to perform a local study then it is necessary to rely on published data and data mining techniques. If a literature or manufacturer's range is to be used in your laboratory a local validation (for example, using 20 "normal" samples) is required.[1] An assessment using a patient database, especially when there is a significant proportion of results in the database unaffected by disease, will give further reassurance about an external source reference interval.

Documentation Requirements

The data source(s) should be described, including where possible the following factors: number of subjects; nature of subjects (how identified); important exclusions; relevant pre-analytical factors; statistical measures (mean, median, 2.5th and 97.5th centiles and 90% confidence limits); number of outliers excluded; analytical method and traceability of method.

Consideration of Partitioning

Partitioning of a reference interval is the use of separate intervals for different sub-populations, and may be necessary to take account of sex, age (in both the paediatric and geriatric populations), reproductive status (puberty, menstrual cycle, stage of pregnancy, menopause) and race (e.g. PSA and eGFR in African Americans). The type of sample (e.g. plasma vs. whole blood, or random vs. first morning spot urine sample) may also require separate reference intervals.

While strict statistical criteria have been published on the need for partitioning,[11] a pragmatic approach is also required to minimise confusion due to an unnecessary plethora of intervals. A key consideration is the likely effect of partitioning on clinical decision making.

Documentation Requirements

The consideration of partition for sex and age in the paediatric and geriatric age groups should be documented, including when partitioning is not performed. Other partitioning relevant to the analyte should be documented.

Significant Figures Used for Reference Intervals

There is now a range of literature describing the process for determining the number of significant figures used to report patient results,[12,13] and this is dealt with elsewhere in this volume of the Journal.[14] There are however no guidelines on the apparent accuracy with which we should report reference interval limits. The estimation of the uncertainty of a reference interval study can guide us in the number of significant figures, which should never be more than is used to report results and the use of an appropriately rounded limit may facilitate recall by practitioners. A slightly wider interval due to "rounding out" of the interval may also allow for slight changes in analytical performance over time. The number of significant figures of the results can markedly influence the impact of a reference interval or decision point. For example reporting HDL cholesterol to the nearest 0.1 mmol/L with a decision point of <0.9 mmol/L will flag data at 0.85 mmol/L or below. If two decimal places are used data will be flagged at 0.895 mmol/L and below.

Documentation Requirements

The degree of rounding of the reference interval limits should be stated together with the reasons for the conclusion.

Clinical Significance

The use to which intervals may be put is an important consideration, and requires discussion with clinicians experienced in the relevant area. The need for sensitivity or specificity may indicate that a reference interval limit should be rounded in or out. Specifically in cases where there is an overlap between health and disease, or where very common conditions such as obesity or diabetes can affect the interval, discussion with clinicians is important. An example of the latter is the influence of these two factors on liver enzymes. Some important confounding factors may be raised, for example, subclinical hypothyroidism in setting TSH intervals.

Whatever process is used to define a reference interval, it will be necessary at some stage to confirm the validity of the proposed reference interval with clinical colleagues, based on their experience using that test to manage their patients. Without this consultative step, and their acceptance, it is possible that they will simply ignore the quoted reference interval and continue to use published guidelines for the interpretation of results, even though this practice may not be absolutely valid for the analytical method being used.

Documentation Requirements

It is valuable to record the people involved in the decision-making process. This ensures that they may be involved again in any re-consideration of the interval at a later time.

Common Reference Intervals

If there are no significant differences in populations or accuracy between two laboratories, it is preferable that the same reference intervals be used. We have noted that even when the same data are assessed, different people will come to different values for the reference interval limits. Thus, in order to achieve common reference intervals, a process of discussion and co-operation is required. This may be on local, regional, manufacturer or organisational lines but may best be done at the national level. The use of common reference intervals where possible is now recommended in the NATA ISO 15189 field application document.[4] In order to share intervals there needs to be criteria for sharing, agreement on the issues described above, and a method for ensuring sufficient analytical agreement between laboratories.

For analytes with a Gaussian distribution, Fraser and others have determined optimal, desirable and minimum bias limits which can be expressed as 0.03, 0.06 and 0.09 × the width of the reference interval respectively.[15] The principle is that the minimal standard will allow a flagging rate (asterisks) of about 5.7% at one end compared to the expected 2.5%. An alternative approach may be to set intervals to a number greater than the central 95% in some cases.

Documentation Requirements

If a common interval is used, it may be possible for a single document to be produced for the interval, with the only requirement for each laboratory to show that their analytical performance allows them to use the interval. This approach has the possibility to markedly reduce the documentation workload on participating laboratories.

Decision and Implementation

The final decision on the upper and lower reference limits must take all of the factors listed above into account. This implies the application of professional judgement and weighing the importance of different factors against each other.

A brief written description of the factors considered in setting the interval is vital to ensure that future reviews will also take them into account.

Any change in reference interval must be communicated with the clinicians in an appropriate manner. This may be with a temporary footnote on results, a flyer to clients or other methods.

Documentation Requirements

The factors used to determine the reference interval and the manner in which they were considered should be briefly described. This should also include the people involved in the

final decision. The process of communicating the change to clinicians should also be described.

Competing Interests: None declared.

References

1. Clinical and Laboratory Standards Institute. How to define and determine reference intervals in the clinical laboratory: approved guideline – second edition. CLSI document C28-A2. Wayne, PA, USA: CLSI; 2000.
2. Solberg HE. A guide to IFCC recommendations on reference values. J Int Fed Clin Chem 1993;5:162-5.
3. Solberg HE. Establishment and use of reference values. In: Burtis CA, Ashwood ER, eds. Tietz Textbook of Clinical Chemistry. 3rd ed. Philadelphia, PA, USA: WB Saunders Co; 1999.
4. National Association of Testing Authorities. Application document: supplementary requirements for accreditation in the field of medical testing (AS 4633, ISO 15189). Rhodes, Australia: NATA; 2007.
5. Australian Standard. Medical laboratories - particular requirements for quality and competencies (AS 4633, ISO 15189). Homebush, Australia: Standards Australia; 2004.
6. Thygesen K, Alpert JS, White HD: Joint ESC/ACCF/AHA/WHF Task Force for the Redefinition of Myocardial Infarction. Universal definition of myocardial infarction. Circulation. 2007;116:2634-53.
7. Twigg SM, Kamp MC, Davis TM, Neylon EK, Flack JR. Prediabetes: a position statement from the Australian Diabetes Society and Australian Diabetes Educators Association. Med J Aust 2007;186:461-5.
8. Bhattacharya CG. A simple method of resolution of a distribution into gaussian components. Biometrics 1967;23:115-35.
9. Ichihara K, Itoh Y, Lam CW, Poon PM, Kim J-H, Kyono H, et al. Sources of variation of commonly measured serum analytes in 6 Asian cities and consideration of common reference intervals. Clin Chem 2008;54:356-65.
10. Google Scholar: http://scholar.google.com (Accessed 26 May 2008).
11. Lahti A, Hyltoft Petersen P, Boyd JC, Fraser CG, Jorgensen N. Objective criteria for partitioning Gaussian-distributed reference values into subgroups. Clin Chem 2002;48:338-52.
12. Hawkins RC, Johnson RN. The significance of significant figures. Clin Chem 1990;36:824.
13. Badrick T, Wilson SR, Dimeski G, Hickman PE. Objective determination of appropriate reporting intervals. Ann Clin Biochem 2004;41:385-90.
14. Badrick T, Hickman PE. Significant Figures. Clin Biochem Rev Suppl (i) August 2008;29:S91-S93.
15. Fraser CG. Biological variation: from principles to practice. Washington, DC, USA: AACC Press; 2001.