Commentary

The Case for Common Reference Intervals

*Graham RD Jones,[1] Antony Barker,[5] Jill Tate,[2] Chen-Fee Lim,[3] Ken Robertson[4]

[1]Department of Chemical Pathology, St Vincent's Hospital, Sydney, NSW, and Faculty of Medicine, University of NSW, [2]Department of Chemical Pathology, Princess Alexandra Hospital, Queensland Health Pathology Service, Brisbane, QLD, [3]Department of Chemical Pathology, The Alfred Hospital, Melbourne, VIC, [4]Core Clinical Pathology and Biochemistry, Royal Perth Hospital, Perth, WA, Australia, [5]LabPlus, Auckland, New Zealand. On behalf of the AACB Working Party on Common Reference Intervals. *For correspondence: Dr Graham Jones e-mail: gjones@stvincents.com.au

Abstract

The current paradigm for pathology reference intervals is for each laboratory to determine its own interval for use with each test offered by the laboratory. It is our contention that this approach does not best serve the medical community, especially at a time when electronic databases of health information are being expanded and integrated. We also believe that this approach is not performed well in many laboratories and is excessively expensive in practice. In contrast, we believe that the preferable option is to develop and apply common reference intervals throughout Australia and New Zealand, together with common reporting formats and assay standardisation wherever this is possible.

We are aware that these are neither trivial nor simple issues, however we believe that failure to achieve this goal where technically possible will be a failure of the pathology profession to meet the challenges of the modern health community. (Clin Biochem Rev 2004; 99-104)

The Current Situation

The current approach to establishing reference intervals follows recommendations from the NCCLS[1] and the IFCC.[2] In Australia this process is encoded in the NATA summary of ISO/IEC guide 17025,[3] which specifies that laboratories may perform their own detailed reference interval studies or may validate reference intervals published elsewhere for their own methods and populations. This recommendation for local validation of reference intervals is repeated in Product Information from most suppliers of in-vitro diagnostic equipment and reagents. The benefit of this approach is that method differences between laboratories should be included in the locally determined reference interval, as would any differences in the population served by the laboratory compared to other locations. The current paradigm requires that results should be reported together with accompanying information, including units and reference intervals, from which it follows that result interpretation should only be made when this supporting information is available. This approach is easily implemented with paper reports where these elements are all part of the report printed by the laboratory. This can also be supported by electronic data transfer in fixed format (e.g. PIT format) where paper versions are reproduced electronically, and is suitable for atomised data transfer when results are obtained from only one laboratory.

Assigning the task of setting or validating reference intervals to each laboratory reduces the requirement for strict control of bias of results compared to external standards as differences in standardisation are allowed for in the local reference intervals. Long-term control of precision, i.e. a constancy of bias, is important to ensure no change in results from the time of reference interval determination. The current approach also does not prescribe precisely such matters as the reporting units, where more than one format of SI units may be acceptable, or the number of decimal places to report. While there is literature related to these issues,[4,5] there is no body currently empowered to provide ongoing review of these matters.

Difficulties with the Current Approach

There are many difficulties with the current approach to reference intervals and these are summarised in Table 1. Some of these issues are long-standing however some are recent developments which require a response from the profession. The most important issue is that relating to optimal patient care. Patients do move between hospitals and geographic locations and are seen by multiple doctors and have tests performed at more than one laboratory. In these circumstances the differences in laboratory results and the associated reference intervals make the assessment and monitoring of the patient more difficult. The worst case may be the incorrect interpretation of a result, either due to unanticipated differences in standardisation between laboratories or because the wrong reference intervals are applied. This is more likely to occur when results are separated from the laboratory's reference intervals as may happen when data is transferred by telephone, written into patient notes, or compared to reference intervals remembered from other locations.

Electronic databases which may receive results from more than one laboratory are becoming more common. These include doctors' desktop or practice-based systems but also under development are regional, state-wide or national electronic health records (EHR) which may receive pathology results. In any of these systems the most benefit which can be gained from the pathology data relies on appropriate assessment of all the data, e.g. in reviewing changes in results over time with graphical or other techniques. This benefit can only be realised if results are directly comparable through appropriate standardisation and use of common reference intervals. Indeed the capacity of such databases to handle results from different sources with different reference intervals is not a current design feature. Accumulation of standardised data from many sources, for example in a centralised EHR, will enable more valuable research to be performed on the database than if the data is not directly comparable.

Problems with the current approach also include the cost and difficulty of performing appropriate reference interval studies for all assays performed in a laboratory and maintaining this with changes in methodology over time. In this regard the increased number of tests available in even medium-sized laboratories has increased the size of this task. With immunoassays the reagent costs can become significant and the time required to collect and analyse samples and then perform the statistical analysis makes the process expensive irrespective of the analytical costs. It is the experience of the authors that the majority of laboratories with which they have had professional contact do not comply with the requirement to have recent, documented local reference interval data for the assays provided.

It may not always be appreciated that the limits derived from a reference intervals study themselves have errors in their estimation. For example a normal distribution with a sample size of 40 will have a 90% confidence interval for both the upper and lower reference limits (+ and − 2SD) which are each 22% of the overall interval.[6] This reduces to about 13% for 120 samples. With non-parametric statistics it is not possible to estimate the confidence limits for sample sizes below 120 and with this number of samples the 2.5th centile is taken from the 4th lowest sample and the 90% confidence limit is defined by the lowest and the 7th lowest samples, with the same consideration defining the accuracy of the upper reference limit.[7] Given this data it can be seen that considerable numbers are required for accurate estimation of reference intervals and even larger numbers are required when partitioning on the basis of age, sex or other factors is considered. Data-mining techniques such as the Bhattacharya method have recently been used to determine reference intervals in a large private practice.[8,9] While no additional testing or sample collection is required, several conditions need to be present to use this approach. These include a large database of results, stable methods over the time of data collection, a large proportion of the results to be unaffected by the health status of the patient and a normal or log-normal distribution of the results. This may provide a valuable

**Table 1.** Difficulties with laboratory-based reference intervals.

1. Patient management may be compromised due to lack of direct comparability of results from different laboratories.
2. Combining results from different laboratories in electronic databases is difficult to support.
3. The expense of performing effective reference interval studies in all laboratories for all analytes is prohibitive.
4. Evidence from IMEP-17 suggests that differences in reference intervals between laboratories is not related to assay standardisation.
5. Development of some reference intervals, e.g. sex hormones, is beyond all but the most well-resourced laboratories.

resource but is unlikely to be a suitable method for all laboratories.

Additional difficulty is associated with generation of reference intervals where special patient selection is required. For example the recent study supported by Andrology Australia has shown that current reference intervals for luteinising hormone and testosterone in men vary widely between laboratories.[10] This group have performed a reference collection where each member of the reference population has had a normal physical examination and a normal sperm count. Obviously the performance of this type of reference interval determination is impractical to perform at every laboratory measuring these analytes. Although standardisation of the various available immunoassay systems and common reference intervals was considered in this study, the main outcome is reference intervals for each analyser system based on the same well-defined reference population.

The International Measurement Evaluation Program (IMEP) is a program for interlaboratory comparisons. The program is founded, owned and co-ordinated by the European Commission's Institute for Reference Materials and Measurements (EC, IRMM).[11] IMEP-17, subtitled Trace and Minor Constituents in Human Serum, involved sending a serum sample containing 19 analytes with concentrations determined by reference methods to 1037 laboratories world-wide, including 56 volunteer laboratories in Australia and New Zealand. In addition to analysing the sample, participating laboratories supplied reference interval data for the measured analytes. The measurements were performed in duplicate on five consecutive days and the reference interval data requested was for a 40 year old male patient. The raw data was obtained from the RCPA-AACB Chemical Pathology QAP which administered the program in Australia and New Zealand. An example of the data is shown in Figure 1 for magnesium. It can be seen that there is no apparent relationship between the measured analyte concentration and either of the reference limits. Inspection of the data revealed no obvious correlation between the reference limits and the measured concentration for any of the analytes. These observations are supported by statistical analysis, with the range of r values for correlation between reference limits and analyte concentration ranging between -0.24 and +0.25 with no p values less than 0.05. Additionally both the upper and lower reference limits showed much higher between-laboratory variation than was seen for the measurement results (Figure 2). These data clearly demonstrate that amongst this self-selected group of Australian and New Zealand laboratories the reference intervals do not compensate for method differences and have a variability unrelated to the measurement.
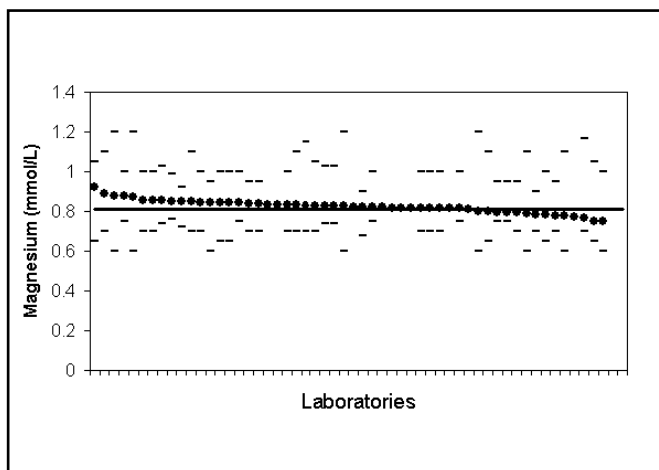
Figure 1. IMEP-17 data for Australian and New Zealand participants for magnesium. Magnesium concentration of supplied sample, average of 10 measurements (filled circles); upper and lower reference limits (dashes), and the IMEP target value (unbroken line). The data is sorted in order of decreasing values for measured magnesium. Note there is no correlation between measured magnesium concentration and the stated reference limits.
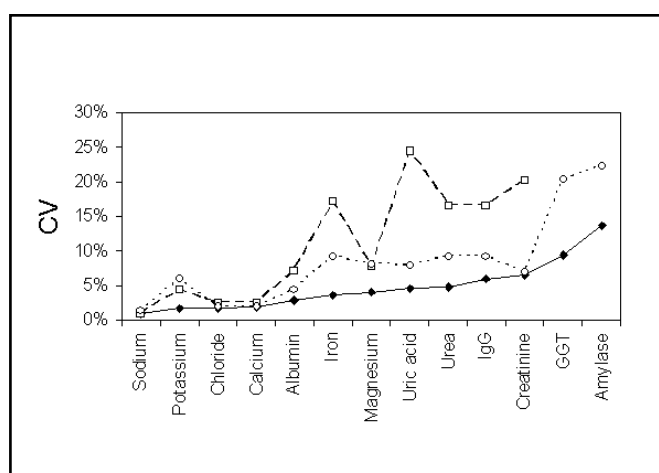
Figure 2. IMEP-17 data for Australian and New Zealand participants. Plot of between-laboratory variation for analyte concentration (closed diamonds) and upper (open circles) and lower reference interval values (open squares). Variation expressed as the CV. Note no data is supplied for lower reference intervals for GGT and amylase due to the use of "less than" formats in some laboratories.

Considerations Supporting the Generation and Adoption of Common Reference Intervals

A number of factors are developing which support attempts to adopt common reference intervals and these are summarised in Table 2. The concept is not new in Australian and New Zealand laboratories in that common reference

Table 2. Factors supporting common reference intervals.

1. Several assays already have widely accepted common decision points which are not determined or validated in individual laboratories, e.g. glucose and lipid measurements.
2. Common reference intervals are being developed by another international clinical chemistry organisation (Nordic Reference Interval Project[15]) and have been in practice in Auckland for many years.
3. This approach has been discussed favourably in an international meeting.[17]
4. Advances in assay standardisation (e.g. European Union).[18]

intervals, or at least common decision points, are in place for a number of analytes. These include decision points for diabetes diagnosis using serum glucose,[12] for cardiac risk assessment and treatment goals with serum lipids[13] and all therapeutic ranges for common therapeutic drugs.[14] Other areas where reference intervals are taken from the literature, often without local validation are for paediatric patients. Implicit in the acceptance of these data is the responsibility of the laboratory involved to minimise bias for the measurement of these analytes compared to the method used to determine the interval or decision point.

These issues have been considered by other Clinical Chemistry Societies and common reference interval programs have been established in some locations. These include the Nordic Reference Interval Project[15] covering 25 common analytes and a Spanish initiative.[16] A recent international meeting also addressed this issue and concluded that common reference intervals have much to commend them.[17] An important local example is the Auckland Regional Quality Assurance Group, which has been meeting on a monthly basis for the past 27 years and has standardisation of reference intervals as one of its main goals. Analysis of split patient samples is used to establish the degree of bias between dissimilar methods, and to assist in resolving analytical issues for standardised methods. Agreement has been reached on the reference intervals for most routine tests, which in turn has facilitated the establishment of a regional electronic database of patient results.

There are also a number of international developments aimed at addressing assay standardisation. Within the European Union all assays must have standardisation back to a definitive method and statements of the uncertainty of the final result produced by the laboratory method.[18] This is supported by programs such as IMEP-17 and analyte-specific programs such as the NGSP.[19] Additionally reference methods can be used to validate material for use in proficiency testing programs allowing assessment of bias with these programs.

An example of this is the measurement of many of the analytes in the RCPA-AACB QAP General Chemistry program with reference methods, or comparison with reference intervals.

Difficulties with Common Reference Intervals

The development and application of common reference intervals is not without considerable difficulties and some of these are listed in Table 3. Firstly, common reference intervals can only be considered if there is assay standardisation. While laboratories will never produce exactly the same results, laboratories wishing to share a common reference interval must determine limits to variation in bias and precision that allow interval sharing. The statistics will not be further discussed here. There are however results which are not currently standardised due to chemical differences in the analyte detected and the calibrators used (e.g. troponins), or to value assignment or choice of standard by the manufacturing company. We need to recognise these differences and ensure that a common reference interval is not inappropriately applied. It may be that several different reference intervals are needed for a single analyte, depending on the method in use. In such cases the laboratory would need to indicate on the report the method used to generate the results.

Table 3. Difficulties with common reference intervals.

1. True standardisation differences.
2. True local population differences.
3. Agreement on format of results and reference intervals.

Another assumption with the proposal of common reference intervals is that the populations served by the different laboratories are similar in respect to the concentrations of the analytes concerned. We are unaware that this assumption has been formally tested for any analytes across Australia and

New Zealand however we are also unaware of any data refuting this assumption. As stated above it may take studies of considerable size to detect any such differences and application of data from a well-performed local study may be preferable to comparison with overseas derived reference intervals determined by manufacturers and other groups. One important population difference for some analytes is the difference between ambulant and hospital patients. It can be argued that a hospital laboratory should have lower reference intervals compared to those for healthy ambulant individuals for such analytes as albumin and sodium which tend to be lower in recumbent and unwell persons. We do not support this concept for several reasons. Firstly reference intervals are commonly described as "health-associated" and this is not generally the case with inpatients and secondly the concept of an average hospital patient seems flawed as there is a wide variation in the illness of the group of admitted patients. Another possible cause of differences in local reference intervals may be the racial make-up of the local population. While it is known that there are some racial differences in concentrations of some common analytes, the inclusion of these differences in local reference intervals requires very careful attention. Firstly the racial makeup of the reference population must match that of the population served by the laboratory and even then an individual may not benefit from the application of such intervals. If a separate group is thought to exist, it should be studied separately and either partitioning performed,[20] or knowledge of the difference used in interpretation of results.

Acceptance of common reference intervals by laboratories also requires adoption of common reporting formats and acceptance of assumptions concerning the intervals. Thus use of the same units, reporting to the same number of decimal places, and rounding of the reference interval limits (e.g. are creatinine reference intervals reported to the nearest µmol/L or 10 µmol/L) needs to be agreed before common intervals can be adopted. While there is no current clear guidance in this area it would seem inappropriate to imply greater precision in reference interval limits than can be supported by the data on which it is based.

Practical Issues of Developing and Applying Common Reference Intervals.

Some of the issues related to practical development and application of common reference intervals are listed in Table 4. Firstly a body needs to be convened and supported to develop methodologies and criteria for common reference intervals. The body needs to develop the theoretical background for adoption of common reference intervals and consider the best reporting formats. This body then also

Table 4. Practical issues with common reference intervals.

1.	Organise and support a body to oversee the project.
2.	Develop agreed statistical approaches to development and application of common reference intervals.
3.	Obtain quality local data for reference intervals.
4.	Publish common reference intervals and criteria for use by laboratories.
5.	Overcome inertia in laboratories and encourage wide-spread adoption.

needs to gather currently available data for reference intervals and, where necessary, commission further studies. Australia and New Zealand are very well supported by the RCPA-AACB QAP and data from this body may provide the confirmation that a laboratory is performing within-specifications for adoption of a common reference interval. Where matrix effects limit the value of the QAP data it may be necessary to investigate other ways of assessing laboratory performance. Once a common reference interval has been developed and recommended it must be published and actively promoted for adoption by laboratories. It would be hoped that laboratories which have invested in high quality local reference intervals would have relatively minor changes to make to adopt the common intervals. Laboratories with older, less well-validated intervals may have more marked changes to make, however the benefits for their clients may be greater. Obviously the project only achieves maximum benefit if such intervals are adopted by the majority of Australasian pathology laboratories.

Conclusions

It is our belief that the current paradigm for generation and maintenance of laboratory reference intervals is difficult and expensive to implement correctly, in general is poorly performed, and does not meet the needs of patients and doctors. The adoption of common reference intervals where appropriate, and addressing the issues of between laboratory bias that underlies this issue, may provide improvement in all of these areas. The first steps in this process are recognition of the problems, a willingness amongst laboratories to work together towards these goals and the development of an organisational structure to provide leadership.

Abbreviations:

NCCLS:The National Committee for Clinical Laboratory Standards.
NATA:National Association of testing Authorities, Australia
IMEP:International Measurement Evaluation Program
NGSP:National Glycohaemoglobin Standardization Program
ISO/IEC: The International Organisation for Standardisation /the International Electrotechnical Commission.
PIT: Pathology Information Transfer
RCPA: The Royal College of Pathologists of Australasia
AACB: Australasian Association of Clinical Biochemists
IFCC: International Federation of Clinical Chemistry and Laboratory Medicine.

References

1. National Committee for Clinical Laboratory Standards. How to define and determine reference intervals in the clinical laboratory: Approved guideline - second edition. NCCLS document C28-A2. Wayne USA, 2000, NCCLS.
2. Solberg HE. A Guide to IFCC recommendations on reference values. JIFCC 1993;5:160-4.
3. ISO/IEC 17025 Application Document. Supplementary requirements for accreditation in the field of medical testing. National Association of Testing Authorities. 2000.
4. Barker AN, Garcia-Webb P, Hughes WG, McPherson VJ, Rothfield D, Whitfield JB. SI Units revisited. Royal College of Pathologists. Broadsheet No. 29. August 1989.
5. Hawkins RCW, Johnson RN. The significance of significant figures. Clin Chem 1990;36:824.
6. Solberg HE. Approved recommendation on the theory of reference values (part 5). Statistical treatment of collected reference values. Determination of Reference Limits. J Clin Chem Clin Biochem 1987;25:645-56.
7. Reed AH, Henry RJ, Mason WB. Influence of statistical method used on the resulting estimate of normal range. Clin Chem 1971;17:275-84.
8. Taylor N, Meerkin M, Sikaris K, McNeil A, Garcia-Webb P, Guerin M. Patient Data Defined LFT Reference Intervals. Clin Biochem Rev 2001;23:89 (abstract)
9. Bhattacharya CG. A simple method of resolution of a distribution into Gaussian components. J Biometric Soc 1967;23:115-35.
10. de Kretser DM, McLachlan RI, Handelsman DJ et al. Andrology Australia's study of testosterone reference intervals Clin Biochem Rev 2003;24(4, suppl)
11. IMEP-17. Trace and Minor Constituents in Human Serum, report to participants, European Commission, Institute for Reference Materials and Measurements, June 2003.
12. New Classification and Criteria for diagnosis of Diabetes Mellitus. Position statement from the Australian Diabetes Society, the New Zealand Society for the Study of Diabetes, the Royal College of Pathologists of Australasia and the Australasian Association of Clinical Biochemists. Med J Aust 1999;170:375-8.
13. Lipid Management Guidelines - 2001. Med J Aust 2001;175: supplement S1-S85.
14. Australian Medicines Handbook. http://www.amh.hcn.net.au/ (accessed 1st April 2004).
15. Nordic Reference Interval Project. http://www.furst.no/norip/ (accessed 1st April 2004).
16. Ferre-Masferrer M, Fuentes-Arderiu X, Alvarez-Funes V, Guell-Miro R, Castineiras-Lacambra MJ. Multicentric reference values: shared reference limits. Eur J Clin Chem Clin Biochem 1997;35:715-8.
17. Cruz-Carlos LM, Monge-Azemar N, Fuentes-Arderiu X. Report of the II European symposium on clinical laboratory and in vitro diagnostics industry: "physiological reference values: a shared business?" Clin Chim Acta 2003;338:165-9.
18. ISO/FDIS. In-vitro diagnostic systems - Measurement of quantities in samples of biological origin - Requirements and layout of reference measurement procedures ISO/FDIS 15193, 2002.
19. NGSP, Harmonising Glycated Haemoglobin Testing.http://www.missouri.edu/~diabetes/ngsp.html. (accessed 1st April 2004).
20. Lahti A, Petersen PH, Boyd JC, Fraser CG, Jorgensen N. Objective criteria for partitioning Gaussian-distributed reference values into subgroups. Clin Chem 2002;48:338-52.

Disclaimer

The opinions expressed in this paper are those of the authors, and do not necessarily represent the position of the AACB.