Western Pathologist Quality Assurance Association

M. Omar Shokeir, MD

Reference Intervals

What is a reference interval?

A reference interval is a range of values for a given analyte, usually bounded by 2 reference limits, used for comparison with an observed value from a patient under investigation. Reference intervals often represent the central 95% of values obtained from a reference sample group, subjects who are similar to the patient, except that they lack the disease state for which the patient is being investigated.

Why have reference intervals?

Despite the World Health Organization's absolute definition of health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity," health is a relative concept operationally, requiring normative data for assessment. For example, a physician can only diagnose a patient with hyponatremia knowing that the patient's serum sodium level is lower than that found in healthy individuals. In short, a patient's observed value must be compared with reference intervals for a meaningful ascertainment of health or disease.

Why is the term "reference interval" preferred over "normal range"?

The term "normal" is ambiguous because it has several meanings. First "normal," can mean common or usual. Second, "normal," when used in astatistical sense, refers to a Gaussian probability distribution rather than a state of being. In fact, many clinically relevant analytes do not exhibit a symmetrical, bell shaped, "normal" distribution. Third, "normal" can denote absence of disease or risk factors for disease. In contrast to "normal range," the term "reference interval" carries no vague denotations or connotations, but describes only a set of values intended for comparison with the observed result.

What is the difference between subject- and population- based reference values?

Subject based reference values refer to levels of analytes obtained from the patient when he or she is healthy and used for comparison when the patient is being investigated for a disease. Since most of us are unfortunate (or fortunate) enough not to have every lab test performed on us when we are healthy just in case those values would be useful if we ever became sick, population-based reference values are generally used. A sample of results from well defined reference individuals are used to generate population-based reference intervals, Reference intervals are population-based, unless otherwise qualified.

What is the difference between uni- and multi-variate reference intervals?

When each analytic result from a patient is compared independently to a healthy population, univariate reference intervals are used. When a panel or group of test results is simultaneously and collectively assessed, a multivariate reference interval is required. Employing complex mathematical models to develop, multi-variate reference intervals utilize a separate spatial dimension for each analyte, defining a "result space" (or hyperspace if more than 3 analytes are concurrently examined) that healthy individuals

inhabit. Such an approach, while not commonly invoked, is useful to determine how a given pattern of patient test results compares to healthy individuals.

How is the reference population selected?
The reference population, a usually hypothetical group, consists of all people (reference individuals) meeting well-defined inclusion and exclusion criteria. Typically, only a subset (the reference sample group) of the reference population is tested for determination of reference intervals.

What is meant by exclusion?
Certain characteristics, termed exclusion criteria (e.g., medication use, recent surgery), may be used to eliminate individuals from the reference sample group, especially if the targeted reference population is composed of healthy people. Typically, interviews and questionnaires are the tools used for exclusion.

What does partitioning mean?
Partitioning is the division of reference individuals into subclasses to obtain more homogenous groups. The most common partitioning criteria are age and gender. The number of partitioning criteria should be kept small enough to create subclasses large enough for statistical validity. Partitioning criteria should be demographic features or physiologic states that materially affect the level of an analyte. The means and/or standard deviations of the subclasses should be substantially different to justify partitioning.

How is the reference group selected from the reference population?
Two temporally different approaches may be used to select the reference group from the reference population. In the a priori method, exclusion and partitioning are performed before testing. This approach is best suited to small studies on analytes for which the sources of biologic variation are well characterized. Conversely, in the a posteriori method, exclusion and partitioning are performed after testing. Essentially, individuals used for the reference sample group are selected from a database, the existence of which is a necessary pre-requisite.

How many reference values are required?
Using non-parametric statistics (which are based solely on the rank order of observations), n reference values are required to distinguish between two percentiles P% apart, where n(1 00/P) - 1. A minimum of 39 values are therefore required to establish a 95% reference interval, in which the 2.5 and 97.5 percentiles are distinct from the 5 and 95 percentiles. However, to establish the 2.5 and 97.5 percentiles with 90% confidence requires 120 samples, the minimum suggested by most authors and NCCLS to determine a reference interval. To increase the confidence in those limits to 95% and 99%, 153 and 198 reference values are required, respectively. Practical considerations maytemper the desire for statistical rigor for difficult to obtain or uncommon specimens (e.g., CSF or neonatal specimens).

What preanalytic factors affect reference values?

Both biological (e.g., medication, fasting, physical activity) and methodological (e.g., tourniquets, centrifugation, storage, delays in testing) factors may affect reference values. Standardized patient preparation and specimen handling should be employed to minimize variation in these preanalytic factors. Perhaps more importantly, actual patient specimens should be procured in exactly the same fashion as samples for a reference interval study, otherwise the validity of comparing results to reference intervals is degraded.

Why are confidence intervals used for reference limits?
Since usually only a very small proportion of the reference population is selected as the reference sample group, the resulting reference interval may differ, by chance alone, from the reference interval that would be obtained if the entire reference population were tested. The confidence interval refers to the probability that a reference limit derived from a sample group is close to the true (but obviously unknown) reference limit for the entire population. For example, suppose that the lower reference limit for serum sodium obtained by analysis of a sample group is 135 mmol/L, a 90% chance might exist that the true reference limit (if the entire population were analyzed) would lie between 134 and 136 mmol/L. 134-136 mmol/L would be the 90% confidence interval. The calculation of confidence intervals assumes that the selection of the reference sample group from the reference population is completely random. Thus, confidence intervals can only account for random error in obtaining the sample group. Confidence intervals do not measure systematic error (bias) (e.g., using only volunteers) in choosing the sample group from the reference population. Confidence intervals are useful reminders that the reference limits derived from the sample group are only estimates of the true population reference limits. Enlarging the sample size produces a commensurate tightening of the confidence intervals.

Is transferring reference intervals acceptable? If so, when and how?
It is impractical, and often impossible, for every laboratory to develop its own reference intervals for every analyte, especially when many tests require age and gender specific reference intervals. Transference refers to the use of reference values from a manufacturer or another laboratory. Although convenient, common, and often necessary, reference intervals must not be transferred mindlessly; several conditions must be fulfilled. First, the laboratorian must be satisfied that the original reference interval study was performed properly. Second, the original reference value study must have been performed on an identical or comparable analytic system. To be comparable, the analytic systems should report in the same units, produce similar absolute values, exhibit similar imprecision, be vulnerable to the same interferences, and utilize like calibrators. Third, the reference population should be comparable. Fourth, preanalytic factors such as the subject preparation, specimen procurement, handling, and storage should be similar. If no reference interval study is available that fulfils these 4 criteria, then the laboratory is obliged to perform its own.

How is the transference of reference values validated?
Even if the 4 above guidelines are applied, desire or compulsion may oblige the laboratorian to validate the donor institution’s reference intervals before adopting them. Two approaches have merit, both involving the recipient institution testing a smaller

number of reference individuals than the minimum of 120 which the donor facility should have assayed. First, the recipient laboratory could test 60 reference individuals and determine the reference interval. If those reference limits do not differ significantly from the values obtained from the donor institutions full scale study, then the reference interval can be transferred. A second approach that does not defeat the economy of transference as much requires the testing of only 20 individuals at the recipient institution. Any subjects with outlying values should be excluded and additional reference individuals should be recruited to yield a full complement of 20. If more than 2 of the recipient laboratory's subjects' values fall outside the donor facility's reference interval, then the study should be repeated. If 3 or more values again fall outside the donor laboratory's limits, then the reference interval is not transferable and the recipient facility should conduct its own full reference interval study. If 2 or less of the 20 subjects fall outside the reference limits, then the reference interval may be transferred. Using this 20-person transference validation test, the chance of falsely rejecting the donor institution's reference interval is only 5-7%.

How should laboratory reports present reference intervals?

All laboratory reports of quantitative results should display reference intervals. Ideally, the reference intervals exhibited should be appropriate for the age and sex of the patient. If so, the patient's pertinent demographic information (e.g., male, age 25) must also be present on the report. Preprinted reports listing all of the reference intervals for every age and gender subclass are discouraged. Results outside the reference interval should be flagged. Displaying HIGH and LOW next to such results may be helpful. Laboratories should make information on the development of each of its reference ranges available to its clients. Laboratory users should also be advised of any material change in reference ranges.

Acknowledgement:

The author would like to thank Ms. Alison Grady for her assistance with research for this paper.

Selected Bibliography:

- Borenstein, M. (1994). The case for confidence intervals in controlled clinical trials. Controlled Clinical Trials, 15,411-428.
- Kouri, T, Kairisto, V, et al. (1994). Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. Clinical Chemistry, 40, 2209-2215.
- Peterson, P, Fraser, C, et al. (1999). The index of individuality is often a misinterpreted quantity characteristic. Clinical Chemistry in Laboratory Medicine, 37, 655-661.
- Sasse, E, Doumas, B. How to define and determine reference intervals in the clinical laboratory; approved guideline—second edition. NCCLS Document C28-A2, 20, 1-36.

- Solberg, H. Establishment and use of reference values. In: Burtis, C, Ashwood, E. Tietz, eds. Textbook of Clinical Chemistry. Philadelphia, Penn: W.B. Saunders Company; 1999: 336-356.
- Swift, M. (1995). Simple confidence intervals for standardized rates based on the approximate bootstrap method. Statistics in Medicine, 14, 1875-1888.
- Wright, E, Royston, P. (1999). Calculating reference intervals for laboratory measurements. Statistical Methods in Medical Research, 8, 93-112.
- Young, D. (1992). Determination and validation for reference intervals. Archives of Pathology and Laboratory Medicine, 116, 704-709.