Reference intervals are the most widely used decision-making tool in laboratory medicine and serve as the basis for many of the interpretations of pathology results. Although laboratories are well versed in method verification and validation to assess if methods are fit-for-purpose, less importance is commonly put on selecting the most appropriate reference interval to determine whether an individual is healthy.

**What Are the Sources of Reference Intervals Used by Routine Clinical Laboratories?**

Sources of reference intervals can vary from the lowest level of current publications on methodology, e.g., manufacturers’ package inserts, to published professional recommendations by national or international expert bodies, expert local groups, or individuals, to the more robust reference interval derived from apparently healthy populations. Health-associated reference intervals including shared (i.e., common or harmonized) reference intervals can be determined by formal (direct) studies using a local population and specified preanalytical conditions or by (indirect) data mining drawn from a local population in which specified preanalytical conditions are used.

The advantage of large, well-conducted direct studies is the well-defined reference population with exclusion of diseased subjects, optimum control of preanalytical variables, robust statistical analysis to remove outlier values, and narrow confidence limits around the obtained reference limits. Such is the case with the joint collaboration of the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER)

and the Canadian Health Measures Survey (CHMS), which collected 11,999 biospecimens representative of 96% of Canada’s household population covering the ages 3–79 years and supported by other physical measures and general health information. Before the calculation of intervals, subjects were first excluded according to various parameters including increased body mass index, history of chronic or metabolic disease, acute illness, or use of prescribed medication a month before sample collection. Next there was statistical removal of outlier values followed by non-parametric ranking analysis to calculate pediatric (3–18 years), adult (19–59 years), and geriatric (60–79 years) reference intervals. Generally data from >500 females and >500 males were collected for most of the partitioned age groups, resulting in very robust data with narrow 90% CIs.

Adel and colleagues present data from the CHMS-CALIPER formal reference interval study in 3 separate articles in this issue of Clinical Chemistry (2–4). Partitioned reference intervals are described for 24 routine biochemistry analytes measured by Vitros 5600 FS (Ortho Clinical Diagnostics) (2); 13 immunoassay-based biomarkers measured by Immulite 2000 (Siemens Healthcare), Advia Centaur XP (Siemens Healthcare), and Vitros 5600 FS or Liaison (Diasorin) (3); and 16 hematological parameters measured by Coulter Hm-X (Beckman Coulter) and Sysmex CA-500 Series (Siemens Healthcare (4). The authors note that further transference and validation studies for other assays/platforms are required before these reference intervals can be used to support clinical interpretation of patient test results generated on other platforms.

It is the responsibility of individual laboratories or laboratory networks to use reference intervals that are appropriate for their methodologies and the population they serve, as specified by International Organization for Standardization (ISO) standard 15189 (5). Advice on how to do this is found in guidelines from the CLSI. The key question is, “Is this reference interval suitable for my collection processes, my method, and my population?” Thus there is a need to demonstrate similar preanalytical processes (e.g., specimen collection, transportation, and handling); method comparability including precision, bias, and analytical specificity according to predefined acceptance criteria; and no relevant population differences. When transferring reference intervals from formal studies, the results from the assays/platforms used in the study need to be shown to be equivalent to those produced in the local laboratory, or the interval needs to be transformed as though it had been performed by the local method. These comparisons must be performed using patient samples to ensure a valid comparison. Validation
of the transferred reference interval is then required using individual samples from healthy individuals from the local population served by the laboratory or other techniques (Table 1).

An initial assessment of whether methodology and calibration traceability of the laboratory’s assays are similar to those used to establish the reference interval will prevent wasted efforts. Where a laboratory uses a method known to be biased compared with the method used to set the reference interval, this interval cannot be used. Rather, for assays with established traceability, traceable assays should be used to both set and use the interval (6). Therefore, preliminary information gathered from the manufacturer, external quality assurance (EQA) programs, and other published data is a first step to determining an assay’s traceability claims and suitability for transfer of the reference interval.

The next step is to perform a method comparison study with another laboratory that has the setup assays, according to CLSI EP09-A3 (7), to determine the statistical validity of a reference interval transfer. Samples from healthy individuals with concentrations spanning the reference interval should be tested. In the transference of CALIPER pediatric reference intervals from the Abbott Architect assays/platforms to 4 other commonly used chemistry platforms, Estey et al. (8) performed method comparisons using 200 pediatric pooled serum specimens obtained from leftover pediatric sera that covered most age and sex partitions. Lower and upper reference limits were determined for the 4 platforms from slope and y-intercept values determined by either least squares or Deming regression analysis. However, if the correlation coefficient was <0.70, the data were considered inadequately correlated and could not reliably transfer the Abbott Architect reference intervals. It should of course be remembered that correlation coefficients are highly dependent on the concentration range of the samples. EQA samples were measured at the same time and mimicked the similarities and differences in reference intervals among the 5 assays (8).

Koerbin et al. (9) did a similar method comparison in the Australasian harmonized reference interval study, in which aliquots of biobanked sera from healthy individuals were used to determine the bias between 8 major chemistry platforms in use throughout Australia and New Zealand. Frozen aliquots from 33 healthy individuals who participated in the “Aussie Normals” direct reference interval study (10) were distributed to 24 public and private laboratories throughout Australasia. By use of this approach, the effect was determined of methodological differences on bias and the sharing of a common reference interval. When the data were analyzed, the average result for each analyte was compared with the mean of all results. Specified performance limits based on biological variation were then applied to determine whether bias would prevent the use of a common reference interval by assessing whether all results fell within the allowable limits of agreement and whether regression lines were all within allowable limits for the 8 measurement procedures. Of 27 tested analytes, 19 gave acceptable bias for a common reference interval (11).

The next step in the transference process and potential adoption of an interval is validation of the proposed reference interval, which takes into account preanalytical, analytical, and population differences. Many biomarkers are under tight biological regulation and have little population variation, whereas other analytes may be affected by diet, infection, and local factors. The CLSI guideline C28-A3 (12) allows for subjective validation of a reference interval by laboratory assessment of population demographics and preanalytical and analytical parameters. A more formal validation recommends using 20 (60 for a more robust analysis) samples from local healthy individuals. If ≤2 of 20 values (≤10%) fall outside the reference interval then the interval can be accepted. If >2 of 20 samples (>10%) fall outside the interval, then the comparison is repeated with an additional 20 samples. If ≤2 values then fall outside the interval, accept the reference interval. If this fails, an investigation is required to assess the cause. When an analyte requires age or sex partitioning, CLSI suggests that validation of 1 partition may lead to accepting the validity of other partitions. However, this may depend on various factors such as analyte concentrations, different forms, or matrix issues, in particular for pediatric partitions (13). Such statistical tests are not a replacement for considered judgment of the available data.

An alternate approach for validation of a proposed reference interval against the local method and population is to mine your laboratory’s existing data. This method has advantages over the direct reference interval validation process by providing large amounts of data on the local population being tested and reflects the actual analytical and preanalytical conditions for the tested population. This approach is valid only if there is a majority of unaffected results from the general practice population such that the healthy distribution of values can be clearly identified in the midst of a smaller number of the non-healthy. The most useful parameter is the midpoint of the extracted data, which can be used to assess analytical or population bias by comparison with the corresponding midpoint of the data used to set the reference interval. At a more complex level, Bhattacharya analysis can be used to assess the proposed intervals (14).

Any significant bias will result in misclassification of too many patients. The expected information derived from the combination of assay and reference interval must meet the appropriate clinical sensitivity and specificity for each test. The use of data mining also allows for an assessment of the expected number of results outside
<table>
<thead>
<tr>
<th><strong>Table 1.</strong> Procedures for transfer and validation of reference intervals.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Procedure</strong></td>
</tr>
<tr>
<td>1. Do initial investigation of local laboratory assay to assess its suitability to transfer proposed reference interval obtained in formal study</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2. Use method comparison study to assess whether: Methods are similar enough to use proposed reference interval as is; Methods are similar enough to develop a transfer equation; or Bias will prevent reference interval transfer</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3. Perform a validation of proposed reference interval to demonstrate: Similar preanalytical processes and method comparability including precision, bias, and analytical specificity according to predefined acceptance criteria; and No relevant population differences; Also consider use of a common reference interval across multiple assay/platforms for analyses with acceptable bias</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
resulting in a wider reference interval (14). The laboratory can then compare the expected flagging rates with their current rates (15). Where a reference interval is shared, the analytical variation for more analyzers in more laboratories using more methods will be larger than a singly derived interval, resulting in a wider reference interval (14).

The CHMS-CALIPER collaboration for adult, pediatric, and geriatric reference intervals provides statistically the most robust data to date for a total of 53 different chemistry, immunoassay-based, and hematology analytes and parameters in a Canadian population. The strength of the data lies in the large numbers per age and sex partition and the stringent exclusion of diseased individuals. The comprehensiveness of the population analyzed (ages 3–79 years) means the authors provide a considerably more meaningful account of biological changes to every analyte by age and sex, a feat rarely achieved in reference interval examination.

The advantage of the collaboration is the potential to harmonize reference intervals across Canada and, no doubt, further afield globally. As noted by the authors (4), the chemistry analytes are largely standardized and have calibration traceability to reference materials and measurement procedures. By contrast, the hematology parameters lack certified reference materials, although several are traceable to reference measurement procedures, which has resulted in acceptable equivalence of measurement results between different platforms. To maximize use of this database of reference intervals, laboratories are encouraged to do transference and local validations to demonstrate the validity of intervals.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

References