3. PEDIATRIC REFERENCE INTERVALS: Critical Gap Analysis and Establishment of a National Initiative

Kareena Schnabl, Man Khun Chan, and Khosrow Adeli*
Clinical Biochemistry, Department of Laboratory Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, M5G 1X8, CANADA

*To whom correspondence should be addressed:
Email: khosrow.adeli@sickkids.ca;
Phone: 416-813-8682

The practice of pediatric laboratory medicine involves unique challenges related to development, nutrition, growth and diseases during different periods of infancy, childhood and adolescence. Pediatric patients require a unique medical approach, as significant differences exist in disease frequencies, specimen collection, test performance and test interpretation. In the last two decades, clinical diagnostic laboratories have witnessed a tremendous increase in the variety of biomarkers and major technology changes. Screening, diagnosis and monitoring of almost all pediatric diseases requires the measurement of a wide range of disease biomarkers with varying degree of clinical specificity and sensitivity. These biomarkers are commonly measured in clinical laboratories and guide important clinical decisions. Physicians rely on the availability of appropriate and reliable reference (or normal) intervals to accurately interpret laboratory test results with data collected during medical interview and clinical examination. According to the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) guidelines, reference intervals are defined as the central 95% of values from a reference population bound by lower and upper reference limits (2.5% above and below the interval) at designated percentiles. In order to validate a marker for use in the diagnosis of clinical disorders, the hormones or chemical substances must be measured in large healthy populations of various ages, ethnicities and both genders to establish appropriate reference intervals. Although health professionals understand the importance of reference intervals, many laboratories still do not have comprehensive reference interval data, especially for the pediatric population. Pediatric specimen collection provides challenges and is a major concern for health care providers. It is frequently difficult to obtain sufficient volumes of blood or urine from pediatric patients. Common reference intervals have not been implemented in Canada due to lack of harmonization of methods and differences in patient populations. Consequently, clinical laboratory accreditation organizations and licensing agencies require that each laboratory verify or establish reference intervals for each method. To provide such reference intervals requires selection criteria for suitable reference individuals, defined conditions for specimen collection and analysis, method selection to determine reference limits and validation of the reference interval.

Establishing Population-Based Reference Intervals

i. Selection of Reference Individuals

Ideally, reference intervals should be determined by sampling a healthy population that span the whole age range (1-3). Potential sources of healthy volunteers include the blood bank, hospital employees and students. Clinical Laboratory Standards Institute (CLSI) recommends (4) a minimum of 120 individuals, the minimum sample size required to determine 90% confidence intervals for the 95th percentile reference limits (2.5th and 97.5th percentiles). There are 2 sampling methods; direct and indirect. Direct techniques involve selection of individuals from a population using defined criteria. A priori sampling uses well-defined inclusion/exclusion and partition criteria for reference

individuals before specimens are collected whereas with a posteriori sampling the process of exclusion and partition of reference individuals occurs after specimen collection and analysis. The indirect technique involves application of statistical methods to analytical values in a laboratory database without selection of reference individuals. The direct technique agrees with IFCC recommendations and is preferred; however, the challenge and cost of obtaining a representative group of reference individuals may be overcome with the indirect technique. The indirect technique has clinical utility in select situations including pediatrics and the elderly population. Participation requires informed consent and completion of a questionnaire that ensures confidentiality. Individuals are selected based on analyte-dependent criteria and health assessment. Reference individuals are most commonly partitioned by age and gender. Many reference intervals are based on individuals between 21 and 60-65 years of age. Thus, a need for establishing pediatric-specific reference intervals still exists.

ii. Defining Pre-Analytical and Analytical Conditions

The quality of laboratory testing and valid reference intervals require careful consideration of pre-analytical and analytical variables (1-3). Inclusion/exclusion criteria, partitioning factors and the questionnaire must also consider biological variation and analytical interferences for each analyte. Depending on the analyte, subject preparation may include a specified diet (fasting or non-fasting), abstinence from pharmacologic agents, a prescribed drug regimen, sampling time in relation to biological rhythms, limited physical activity (rest period) or stress. Specimen collection considerations include environmental conditions during collection, specimen type, time of day, posture, sample volume, anti-coagulants, additives, collection site preparation, blood flow, equipment and phlebotomy technique. In the pediatric population, patient age is a significant pre-analytical factor and specimen collection and handling is unique. Phlebotomy for infants and children is technically challenging and requires special training and skill, particularly for sites such as scalp, jugular veins and umbilical artery catheters. Pediatric patients have frequent blood draws and are at risk for anemia necessitating small specimen volumes. Liquid anticoagulant in collection tubes may potentially dilute the specimen when sample volumes are small. Skin puncture specimens (capillary blood) are more commonly obtained from children and consist of a mixture of blood from arterioles, venules, and capillaries with interstitial and intracellular fluids. The results from various analytes deviate from those obtained with arterial or venous samples. Specimen handling and processing variables include time and temperature for transport, clotting, plasma/serum separation, storage and sample preparation for analysis. Since automated laboratory equipment frequently cannot handle small volume specimens or samples directly from pediatric-sized tubes, manual processing of pediatric samples is common in larger laboratories making it difficult to standardize specimen processing across laboratories and to maintain positive sample identification for pediatric specimens throughout all testing phases. Small pediatric samples may also preclude repeat testing to confirm abnormal results. Analytical variables required when determining reference intervals are quality control procedures and analytical performance. The analytical method should be controlled in the same way as routine patient samples including equipment and instrument preventative maintenance and function checks, reagents, calibrators, controls and calculation methods. The analytical validity of the method is critical and the method chosen for analysis must clearly state accuracy, precision, minimum detection limit, reportable range, recovery, and interference characteristics. Analytical interference is encountered more commonly in neonatal patients than in adults and represents a significant technical challenge. Specifically, high concentrations of

bilirubin, lipids, and fetal hemoglobin are present in many neonatal specimens. Bilirubin absorbs light at wavelengths at which many spectrophotometric methods measure a variety of analytes and can cause spurious test results in both chemistry and hematology laboratories. Intravenous nutrition supplemented with fat emulsions can result in specimen turbidity, which interferes with both spectrophotometric and nephelometric methods based on changes in light scatter. Fetal hemoglobin, normally present in significant quantities only in newborns and infants, interferes with accurate measurement of several hemoglobin derivatives. Carboxyhemoglobin and methemoglobin are critical to oxygen transport in the newborn and are important parameters for the management of neonates in the intensive care unit, but large amounts of fetal hemoglobin invalidate their conventional spectrophotometric measurement. Instrumentation appropriate to the neonatal setting can eliminate these technical issues.

ii. Statistical Evaluation of Reference Values

Visual inspection of a frequency distribution histogram of values helps determine whether the distribution is Gaussian (normal, symmetric) or non-Gaussian (skewed, kurtosis, bimodal, polymodal) and identifies outliers. Non-Gaussian distributions suggest that unhealthy individuals are included or the values require partitioning. The standard normal deviate test partitions by age and gender, one of which is the Harris and Boyd method for partitioning (5). A simple test for outlier exclusion is the Dixon/Reed rule. The suspected outlier is rejected if the distance between the outliers is > 1/3 the range of all values. Another method of outlier detection, proposed by Tukey (6) involves the labeling of extreme values by using only the middle 50% of the sample, thus reducing, or even eliminating, the possible masking effect of multiple outliers on one side of the distribution. Since the majority of analytes do not have normal Gaussian distributions, the non-parametric method is recommended by IFCC and CLSI with a minimum of 120 individuals required per partition. Reference limits can be easily calculated using non-parametric ascending rank order statistics. The conventional 95th percentile reference limits are determined by calculating the rank numbers for the 2.5th and 97.5th percentiles. If laboratories cannot establish reference intervals based on the requirement of 120 reference individuals, the bootstrap or robust methods may be applied to get a good estimate of reference intervals. The bootstrap method (7) uses a computer to provide robust percentile estimates by re-sampling from a single subset of reference values. Horne and Pesce (8) describe the robust method where more weight is given to central values of a distribution than to distant values for the calculation of reference intervals. The advantage of this approach is that it is more tolerant of outliers in the reference population data, it does not require as large a sample size as the non-parametric calculation method and it does not require reference data transformation to a Gaussian distribution. Normal values can be presented as reference intervals, decision limits or statistical cut-off values. For example, troponin uses a cut-off at the 99th percentile of the healthy population while lipid levels, hemoglobin A1c, and therapeutic/toxic drug levels employ medical decision limits based on outcomes or diagnostic performance.

iv. Validation of Reference Intervals

Transfer of reference intervals is not ideal given the lack of test method harmonization and differences in patient populations. Because there are few analytes where test methods have been harmonized, it is often necessary to establish reference intervals for each test system. Laboratories establishing a reference interval must ensure comparability of the reference population and pre-analytical/analytical variables. The CLSI guidelines require sample collection from at least 20 reference individuals and performance of a formal

outlier test. If no more than 2 (of 20) are outside the proposed reference interval, the proposed reference interval is acceptable for use. If greater than 2 are outside the proposed reference intervals, then samples for 20 additional individuals are collected and measured. However, if three or more again fall outside the limits (or if five or more in the original set fall outside the limits), the user should reexamine the analytical procedures used, consider possible differences in the biological characteristics of the two populations sampled, and consider whether the receiving laboratory should develop its own reference interval according to the full-scale study guidelines.

Challenges and Gaps in Pediatric Reference Intervals

It is well recognized that pediatricians and pediatric specialists have special needs and requirements for the laboratory testing of their patients. Age-specific reference intervals and specimen collection issues are particular concerns in pediatric patient care and evaluation. Frequently, adult reference intervals are not appropriate for pediatric patients. To assist physicians in treating their pediatric patients, attempts have been made by many laboratories to establish age-specific reference intervals for many traditional disease biomarkers with limited success. Unfortunately, critical gaps currently exist in available pediatric reference intervals (normal values) for accurate interpretation of laboratory tests. These critical gaps may subject infants and children unnecessarily to extra blood collection, infectious risk, stressful diagnostic procedures or inappropriate treatment. Moreover, inadequate pediatric reference intervals may be costly and devastating potentially contributing to misclassification or erroneous/delayed diagnosis of many diseases of childhood and adolescence. Children are not small adults. Therefore, establishing pediatric reference intervals requires a unique approach, considering not only age, gender and ethnicity but reflecting the significant differences that exist in nutrition, metabolism, growth, development, disease frequencies, specimen collection, test performance and test interpretation.

Age-specific reference intervals are important for all types of pediatric laboratory testing. Many endocrinology, chemistry, serology, coagulation and hematology markers are subject to age-specific variability. Sex hormones, growth hormones and bone alkaline phosphatase levels vary with age and gender during childhood growth/development and puberty. Serology test results are influenced by the transplacental passage of maternal IgG and by immunization responses in infants and children. Coagulation tests are optimized for anti-coagulation monitoring and not for childhood genetic diseases such as hemophilia. In hematology, the automated differential uses algorithms in children that differ from those used in adults. Currently available laboratory information systems do not permit the automatic calculation of a patient's age related to prematurity and cannot separate reference intervals by gestational age. Unfortunately, most of the available reference intervals are incomplete covering a limited pediatric age interval that does not always cover both genders. Clearly, determining age- and gender-specific reference intervals is crucial for screening, diagnosis and monitoring of many pediatric disorders.

Children often acquire diseases that differ from adults and are lower in frequency. Newborns are "immunologically naïve", have little if any infectious disease history and are exposed to relatively unique infections. They respond to infections in a different way and often require special testing. Premature birth can result in a set of diseases due to incomplete organ system development, especially those affecting the lungs and the central nervous system. Cancer in childhood comprises a set of neoplasms for which

genetic alterations are much more important than environmental factors. Metabolic disorders and genetic diseases such as cystic fibrosis and sickle cell disease are examples of diseases diagnosed during childhood. Neonatal screening programs established in some countries are becoming more widespread with test menu expansion. Reference intervals are not available for new and emerging pediatric disease biomarkers. Cutoff ranges for health and disease are challenging and data are limited or difficult to find. Consensus guidelines established for adults may not apply to children, as is the case for cholesterol and hyperlipidemias. Drugs-of-abuse testing in children is another area in which data are very limited but could have important implications for growth, development and overall patient care. Finally, metabolism differs in children and will affect therapeutic drug monitoring and therapy. For instance, different antibiotics are indicated for children and not adults.

Many of the problems associated with establishing good reference intervals for adults also exist for the pediatric population. Current pediatric reference intervals have been derived predominantly from samples collected on hospitalized infants and children of the Caucasian population and may not reflect levels in healthy multicultural populations. Collecting samples from healthy children is difficult, particularly from premature infants, term infants and toddlers. The influences of prematurity on laboratory test results are poorly understood and have not been assessed systematically. Different rates of organ maturation occur after premature birth; however, due to concerns about blood volume depletion in premature infants, it has been difficult to study reference intervals. Recruiting children as research subjects poses ethical dilemmas and protocols with institutional review board approval and informed consent are required for reference interval studies on normal, healthy children. Alternative ways of developing reference intervals for established tests require further study and could include mathematical approaches, extrapolation and establishment of decision limits. There is also a need to include a diversity of ethnic groups into pediatric reference interval databases. Application of reference ranges specific to other ethnic groups is clinically inappropriate for many biomarkers.

Another problem with most available pediatric reference intervals, is that they were determined over two decades ago using older/less accurate instrumentation and methodologies that are no longer relevant to testing technology used by clinical laboratories today. Changes in instrumentation and methodology are often not accompanied by reference interval modification and literature data are often adopted without critical appraisal. Manufacturer supplied reference intervals are often not accompanied with a detailed description of how reference intervals were derived including partitioning factors, sub-class differences, sample size, age, gender, ethnicity, race, percentiles used or traceability. Moreover, different reference intervals for the same analyte exist between different laboratories. For instance, ALT has 90 different reference intervals for the male population and some laboratories use the same reference interval for males and females. Establishment of up-to-date reference intervals for current and emerging instrumentation and methodology using standardized pre-analytical and analytical phases is essential. Both the IFCC and CLSI are working together towards a more standardized method for establishing reference intervals. At the 2007 AACC meeting in San Diego a session was dedicated to the issue of reference intervals. The new CLSI guidelines (version C28-A3) were unveiled to the laboratory community. The main differences between the new C28-A3 versus the C28-A2 are as follows:

- Decision limits as well as reference intervals should be used for certain analytes eg. cholesterol, glycated haemoglobin and neonatal bilirubin.
- Validation vs. establishment of reference intervals by a laboratory is acceptable.
- Establishment of reference intervals through multi-center trials. Using the "robust" statistical method as well as the non-parametric method to calculate reference intervals is valid.
- Laboratories should make reports easier to understand with abnormal values highlighted.
- Laboratories should quote either the decision limit or reference intervals, not both.

Establishment of a National Collaborative Initiative on Pediatric Reference Intervals

The concept of common reference intervals was presented by the IFCC as a possible solution to the establishment of up-to-date and comprehensive reference intervals. Since individual laboratories do not have the resources to adequately reassess all pediatric reference intervals, multi-center trials are required. The components of a successful multi-center trial include: adequate numbers of reference individuals, goals for total allowable error, use of methods that are traceable to primary reference methods or commutable reference materials, comparable pre-analytical/analytical phases and an ad hoc quality control program. Improvement in method standardization will potentially eliminate the need for method-specific reference intervals. With a good level of standardization, only population differences justify different reference intervals and the introduction of common reference intervals might be feasible. Currently, there are few laboratory tests that are standardized against a reference method or material. Much work needs to be done before implementing common reference intervals. Do population differences exist? Does ethnicity or different lifestyles affect reference intervals? True population differences are demonstrated for only a few analytes; creatine kinase, creatinine and some proteins (orosomucoid, IgG, C3, C4 and CRP). For most common analytes the differences may be small or insignificant. All these questions can only be answered by multi-center trials with large numbers of healthy reference individuals and the cooperation of international laboratory communities facilitated through an organization such as the IFCC.

Despite the challenges that exist in establishing up-to-date, comprehensive pediatric reference intervals, an ambitious Canadian team of investigators from the pediatric focus group of the Canadian Society of Clinical Biochemists has recognized the urgent need and have assembled to lead the establishment of a reliable database across the country. This national research initiative, working under the name of Canadian Laboratory Initiative on Pediatric Reference interval database (CALIPER database) (9), is critically needed to fill important gaps in the clinical laboratory reference values database and will benefit every Canadian and world-wide clinical laboratory serving a pediatric population. The objective of the proposed laboratory project is to develop a database of patient demographics and laboratory results for the pediatric biochemistry tests that are in urgent need of having accurate reference intervals established for the entire pediatric age range from birth to 18 years old. The database includes traditional as well as emerging biomarkers of pediatric disease and will be utilized to derive comprehensive ageand gender-specific reference intervals that encompass the major ethnic groups of Canada's

diverse population including indigenous Canadians. The reference intervals will be shared among pediatric hospital laboratories across Canada. The major benefit of the study is an accurate and reliable determination of what is "normal" considering a child's age and gender in clinical diagnosis.

Healthy children are currently being recruited in several sites across Canada from hospital nurseries, day cares, schools and families of hospital volunteers to ensure a representative sampling distribution. Samples are sent to a central location to monitor specimen collection and to assess the gap in reference ranges. Developed guidelines are applied for all tests where on-site analysis is required to preserve specimen stability, to avoid delays in analyte measurement and to achieve accurate results. Following specimen collection, analyte levels are measured to establish reference ranges. Analysis is carried out at specified centres using a number of different instruments commonly used in clinical laboratories across Canada. The goal is to generate instrument-specific reference intervals and determine variations in methods across laboratories.

Following data collection, reference intervals are established using parametric and nonparametric methods in collaboration with a clinical biostatistician. Based on a detailed gap analysis (published in the June 2006 issue of the 'Clinical Biochemistry' journal) (10-13), a series of cardiac/metabolic markers, bone markers, genetic metabolic markers, thyroid hormones, the growth hormone-insulin axis, creatinine and C reactive protein have been selected for the initial phase of reference interval determination. CALIPER has recently finished several pilot studies in collaboration with IVD industries. Data generated from these pilot projects has been found to be beneficial for the laboratory community. The last stage of the project is to implement standardized age-specific reference ranges in various pediatric healthcare centres across Canada. CALIPER strives to extend this valuable pediatric reference interval database worldwide in 2 to 3 years as significant gaps exist in pediatric reference intervals in both developed and developing countries. The project seeks to improve the health care of children and to promote academic, clinical and industrial collaborations among pediatric laboratory medicine specialists, pediatricians and others caring for children.

References

1. Editors William Clarke and D Robert Dufour, Contemporary Practice in Clinical Chemistry, Chapter 2 Reference Intervals AACC Press, Washington, DC 2006 13-19
2. Carl Burtis, Edward Ashwood and David Bruns Tietz Textbook of Clinical Chemistry and Molecular Diagnostics Elsevier Saunders 2006 St. Louis Missouri; Chapter 16 Establishment and Use of Reference Values 425-446
3. Coffin C, Pediatric Laboratory Medicine Current Challenges and Future Opportunities American Society for Clinical Pathology 2002 117: 683-690.
4. Clinical Laboratory Standards Institute - Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Proposed Guideline-Third Edition; Proposed Guideline. C28-P3. http://www.clsi.org/source/orders/free/c28-p3.pdf
5. Horn PS, Pesce AJ. Reference intervals: an update. Clin Chim Acta. 2003 Aug;334(1-2):5-23.
6. Solberg HE, Lahti A. Detection of outliers in reference distributions: performance of Horn's algorithm. Clin Chem. 2005 Dec;51(12):2326-32.
7. Henderson AR. The bootstrap: a technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. Clin Chim Acta. 2005 Sep;359(1-2):1-

26.
8. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. Clin Chem 1998;44:622-31
9. Editorial, Pediatric reference intervals: Critical gap analysis and establishment of a national initiative. Clin Biochem. 2006 Jun;39(6):-559-560.
10. Mansoub S, Chan MK, Adeli K. Gap analysis of pediatric reference intervals for risk biomarkers of cardiovascular disease and the metabolic syndrome. Clin Biochem. 2006 Jun;39(6):569-87.
11. Liju Yang, Vijaylaxmi Grey. Pediatric reference intervals for bone markers. Clin Biochem. 2006 Jun;39(6):561-568.
12. Edgard E Delvin, Vijay Laxmi Grey, Zulfikarali Verjee and for the CALIPER Working Group. Gap analysis of pediatric reference intervals related to thyroid hormones and the growth hormone-insulin growth factor axis. Clin Biochem. 2006 Jun;39(6):588-594.
13. Nathalie Lepage, Dailin Li, Peter A. Kavsak, Fiona Bamforth, John Callahan, Kent Dooley, Murray Potter. Incomplete pediatric reference intervals for the management of patients with inborn errors of metabolism. Clin Biochem. 2006 Jun;39(6):595-599.