

# Reference intervals for common biochemistry laboratory tests in the Saudi population by a direct a priori method

Xuejiao L. Hu, Huda Hassan, Fouad Hassan Al-Dayel

From the Department of Pathology & Laboratory Medicine, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia

Correspondence: Dr. Fouad Hassan Al-Dayel · King Faisal Specialist Hospital and Research Center, Pathology & Laboratory Medicine, Riyadh, 11211, Saudi Arabia · T: 966-11-442-7224 F: 966-11-442-4280 · dayelf@kfshrc.edu.sa

Ann Saudi Med 2017; 37(1): 16-20

DOI: 10.5144/0256-4947.2017.16

**BACKGROUND:** Reference intervals (RI) for biochemistry laboratory tests are now based on Caucasian rather than Saudi populations. Test parameters may vary because of race, lifestyle, population structure and geographic location.

**OBJECTIVES:** To establish reference intervals for common clinical chemistry laboratory tests for the Saudi population.

**DESIGN:** Direct a priori method.

**SETTING:** Tertiary care hospital.

**MATERIALS AND METHODS:** Blood samples were taken from 625 individuals aged from 2 to 87 years from different geographic areas for 93 biochemistry tests. RIs were established following the International Federation of Clinical Chemistry guideline.

**MAIN OUTCOME MEASURE(S):** Reference values for common biochemistry lab tests.

**RESULTS:** Ninety-three age- or gender-stratified reference intervals (RIs) based on the Saudi population were established. There were 72 non-partitioned tests. Most of the tests were similar to RIs from manufacturer's inserts. For some sex hormones (estrogen, luteinizing hormone, follicle-stimulating hormone, progesterone and 17 $\alpha$ -Hydroxyprogesterone) only male RIs were established as there were not enough samples to stratify for females based on physiologic status.

**CONCLUSION:** The RIs are reliable and applicable to a general Saudi population.

**LIMITATIONS:** Due to the sample size, RIs were not generated for some sex hormones for females.

Reference intervals are critical for clinical decision making. Due to differences in age, sex, ethnic group, life styles, geographic location and population structures in the region served by the laboratory, we hoped to establish complete reference intervals based on representative population tested by the medical laboratory. However, because of the high cost, stringent requirements, the time required and lack of manpower, establishing reference intervals (RIs) is a daunting task. Most Saudi clinical labs, including our laboratory at KFSHRC rely solely on manufacturer inserts or published peer review data, which are mainly based on Caucasian populations. The drawback is that these test results may not be representative of the Saudi population or may be based on old data that do not reflect current popu-

lation groups. Some well-designed reference interval studies have been completed in Western societies,<sup>1-3</sup> but in regions such as the Middle East, few laboratories have established their own reference values, especially comprehensive ones.<sup>4</sup> To fill this gap, we decided to establish reference values for some commonly used biochemistry laboratory tests following the proposed 2008 Clinical and Laboratory Standards Institute/ International Federation of Clinical Chemistry (CLSI/IFCC) document EP28-A3 guideline.<sup>5</sup> There are direct (a priori and a posteriori) and indirect methods to establish RIs. The direct method is favored due to controversy over the indirect method.<sup>6</sup> Arguably, it is not necessary to establish RIs for all tests, especially for the tests in which RIs have been replaced by decision limits by international con-

sensus, such as for cholesterol, troponin T or glycated hemoglobin (HbA<sub>1c</sub>), and others. However, the data obtained could be used to compare the accuracy of the RIs; therefore, reference ranges for these parameters were still calculated. The data collected for other tests were used for establishing our own RIs. The tests were grouped as electrolytes, liver function, renal profile, lipid profile, thyroid function, cardiac panel, anemia profile, immune function, cancer markers, metabolic panel, and sex hormones for the convenience of application. There were 72 non-partitioning RIs, 21 for either age or gender partitioning RIs, including 6 male-only RIs.

## MATERIALS AND METHODS

### Reference population

The reference population were 625 healthy individuals from the north, west, and east of Saudi Arabia and the Riyadh area. Adults were recruited as well as children under 17 after receiving permission from their parents. Some children were middle school students and others were recruited from daycare centers. In this way, the geographic limitations of the reference population were minimized and the age scope was broadened. This reference population consisted of 316 (50.6%) females and 309 (49.4%) males, ranging from 2 to 87 years of age. There were 201 children younger than 17 years of age (32.2%) including 105 males from 3 to 17 years of age and 96 females from 2 to 17 years of age. Participation was voluntary.

A questionnaire similar to the document C28-A3 (Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition)<sup>5</sup> was used in selecting individuals. The individuals should have no of signs of disease specifically related to the biochemical laboratory measure. A history of diabetes, hypertension, chronic kidney disease, cancer, taking prescription drugs and other chronic illness was grounds for exclusion. After selection, test samples were drawn after 10 hours of fasting. Samples were taken using standardized test tubes by routine phlebotomy. Blood samples were allowed to clot at room temperature, and serum was obtained after brief centrifugation. Blood in anticoagulated tubes was also centrifuged to collect the plasma except for a few that needed whole blood for tests. All specimens were placed into a -70°C freezer until analysis. Due to variations in the number of available tests, not all the same number of individuals tested varied from test to test. The study was funded by King Faisal Specialist Hospital and Research Centre (KFSHRC) and approved by the Ethics Committee of KFSHRC.

### Testing instruments

Roche MODULAR ANALYTICS, E170 module, an electrochemiluminescence based assay was used to perform most tests. The Dade Behring BNII nephelometer was used for apolipoprotein B, apolipoprotein A1, transferrin, haptoglobin, homocysteine, IgG, IgM, IgA, C3, C4, beta 2 microglobulin, a1 antitrypsin, ceruloplasmin and prealbumin; DiaSorin Liaison chemiluminescent immunoassay for calcitonin and somatomedin C (IGF-1); DPC immulite for insulin-like growth factor binding protein 3; Gamma counter radioimmunoassay for aldosterone and Vitamin 1,25 D; Radial immunodiffusion test for IgD; Cobas Mira for angiotensin converting enzyme, Hb plasma and glucose-6-phosphate dehydrogenase. Routine calibration and controls on all testing instruments were performed following the internal procedure protocol.

### Statistics

Reference ranges were determined by the test values that fell within 2 standard deviations or 95% of a normal distribution of the sampled reference population. The lower 0.025 and upper 0.975 percentiles of the test results were excluded. This distribution is expressed by equation:  $RI\ min = \alpha * 2.5 + \beta$ ;  $RI\ max = \alpha * 97.5 + \beta$ . Final sample size of each test for the RI was determined by elimination of outliers using the modified Thompson Tau method<sup>7</sup> or due to variation in the number of tests performed on each sample. Each test parameter was plotted to a histogram to view its distribution and outliers. RIs were determined either by nonparametric parameters or using the log transformed data if the distribution was skewed. Occasionally parametric values were adopted if a test values were close to a perfect normal distribution with a better confidence ratio (CR). To stratify tests for gender, a Harris-Boyd partitioning model built in EP Evaluator (Data Innovations LLC, South Burlington, Vermont) was used to calculate a critical z value and SD ratio. If the SD ratio >1.5 or z max >critical z, partitioning would be done or the manufacturer's instructions would be followed for the purpose of comparison. Age partitioning was based on a t test correlation by dividing the test results into  $\geq 18$  adult group and  $\leq 17$  of children group using IBM SPSS Statistics 20 software or it was based on published test data. When there was a statistically significant difference between RIs in two different age groups ( $P < .05$ ), RIs would be established separately. The data were analyzed using the EP Evaluator software. Based on the IFCC requirement, 90% confidence intervals (CIs) of the mean for each test parameter were calculated and listed along with reference values. The 90% CIs were

calculated using the equation:  $\bar{x} \pm z\alpha/2 [\sigma/\sqrt{n}]$ , where  $\bar{x}$  represents sample mean, plus and minus margin error, where  $z$  stands for z-scores,  $\alpha$  is the significance level of the test, 0.10 for 90% CI and  $\sigma$  equals the standard deviation. The confidence intervals mean there is a 90% chance that the true population mean would fall within the confidence interval, which is a reliable indicator of the uncertainty of RIs. The confidence ratio, expressed as  $0.5*(URLU-URLL+LRLU-LRLL)/(URL-LRL)$ , is the ratio of the average confidence interval width to the reference interval width, which is related to the sample size (0.1 or less is desirable and less than 0.3 is acceptable).

## RESULTS

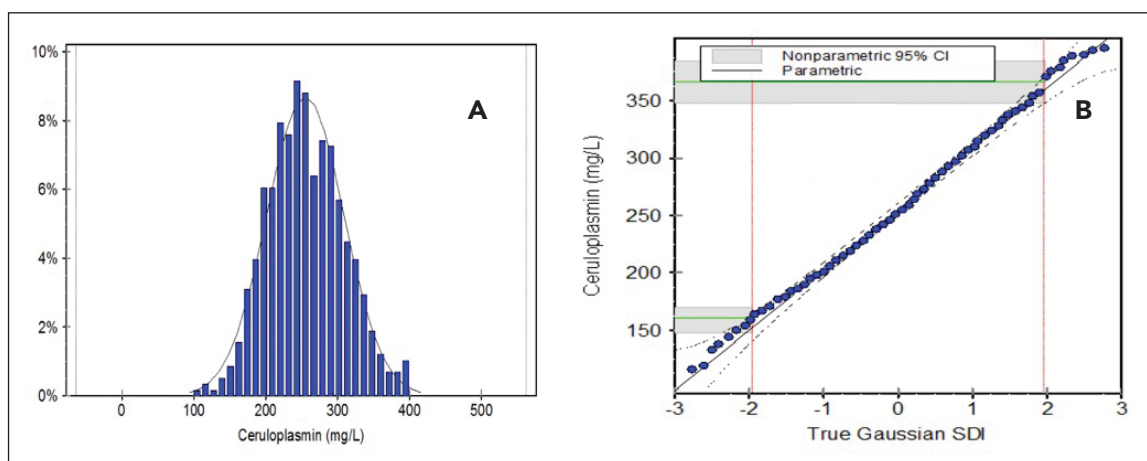
The total of 93 biochemistry tests were divided into 11 groups for the convenience of indexing. There were 72 non-partitioned tests with an age span of 2 to 87 years so that these reference values would be applicable for both males and females for this age group. Test samples for aldosterone were obtained in a sitting position, not in a supine or standing position so partitioning for position was unnecessary. Cortisol was not partitioned as the majority of samples were taken in the morning, so the test results were taken as random and no gender stratification was indicated by the partitioning calculation. Among 21 partitioned tests, 5 were sex hormones (estrogen, luteinizing hormone, follicle-stimulating hormone, progesterone and  $17\alpha$ -Hydroxyprogesterone) and only for males as there were not enough samples to stratify for females based on physiologic status. Our samples were mixed with narrower reference ranges than the insert; therefore, the test would be applicable for both population groups. Detailed results are listed in **Appendix 1. Figures 1 and 2** are examples

of parametric and nonparametric distributions on a (A) histogram and (B) probability plot for original and transformed data, respectively.

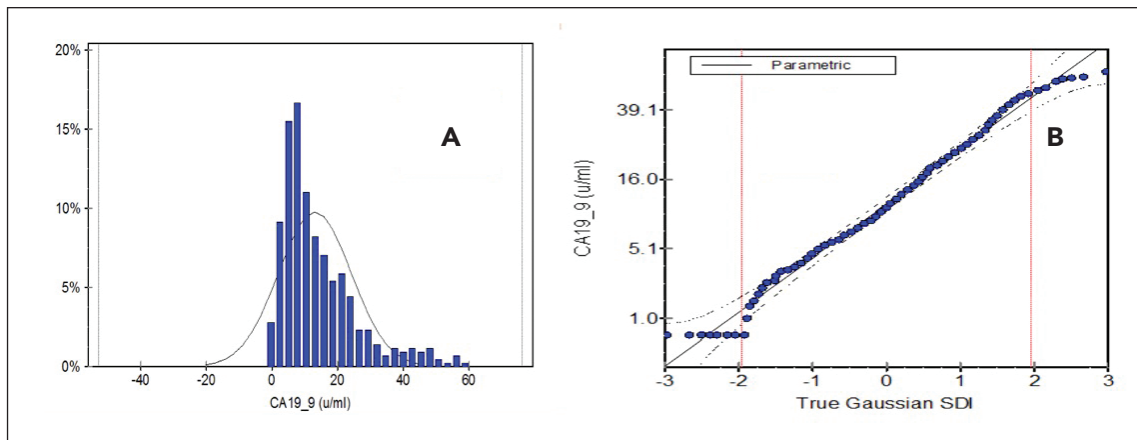
For 108 confidence ratios calculated for 93 tests, the average was 0.097 (desirable  $<0.1$ ). To further verify the sensitivity and usefulness of the RIs we established, the index of individuality was calculated (data not shown). The index of individuality was below 0.6 for only 4 tests, above 1.4 for 43, and 27 were in between. According to Harris<sup>15</sup> if the ratio of variation within (CVi) or between subjects (CVi/CVg ratio) is  $>1.4$  the RIs will be sensitive and useful, whereas if the ratio is  $<0.6$  the utility of the RI is low. We calculated the CVg (between subject variation) for 74 tests for which data was available.<sup>16</sup>

## DISCUSSION

Because many factors impact reference ranges, such as selection of the reference population, test methods, sample size, statistical methods, partitioning, and others, we paid close attention to pre-analytical, analytical and the post-analytical phases to ensure the reliability and accuracy of the reference values. We adopted a standardized questionnaire for the reference population, choosing adequate samples sizes, using traceable test methods, partitioning as necessary, and applying suitable statistical methods, strictly following the internal procedure protocols for handling of all specimens. As there are no universally accepted criteria to define outliers, we based our calculations on different equations, including the Dixon Q test,  $X(n)-X(n-1) > X(n)-X(1)/3$  to reject the largest value and  $X(2)-X(1) < X(n)-X(1)/3$  to reject the smallest value. Since the calculations were not all in accordance, we took comprehensive measures to review the results, including histograms, examining



**Figure 1.** Example of normally distributed histogram of ceruloplasmin values and probability plot after removing outliers. The central 95% of the data is the reference interval. SDI: standard deviation index.



**Figure 2.** Example of histogram of skewed distribution of CA19.9 values that was logarithmically transformed to a linear probability plot and then 95% of reference values were selected. SDI: standard deviation index.

parameters before and after eliminating outliers, and confidence ratios (CR) to extrapolate the best fit RIs. We undertook these measures because skewed distributions were present. Even after elimination of outliers, the CR was unacceptable, even with nonparametric calculations. Conversely, log-transformed parametric values provided much better CRs and reference ranges (e.g., anti-TPO, glucose). For tests that had Gaussian distributions, parametric methods were adopted (e.g., C3, IgG and TT3). If CRs were all the same, we choose nonparametric values (e.g., IGFBP3 and TSAT). All the data used to determine RIs were then verified using a function for verification of RIs by the EP Evaluator; all passed. Although the CLSI C28-A3 recommends the nonparametric method, the RIs calculated by the parametric and nonparametric methods were compared in a recent IFCC, C-RIDL study, 10 which concluded that the results of the two methods were very close, concluding that parametric methods can also be used as a first choice. In theory, parametric methods will produce more accurate and precise estimates than non-parametric methods if assumptions are met. Parametric methods may also have an advantage over non-parametric methods in allowing identification and exclusion of extreme values when computing RIs.<sup>11</sup>

In comparison with package inserts, many RIs we established were very close to the manufacturer's RIs. For instance, electrolytes, which are known for having low biological variation, and for liver enzymes and tests in immune function panels, some RIs and inserts are identical (e.g., homocysteine, troponin T, direct bilirubin, female gamma-glutamyl transferase and male human chorionic gonadotropin). However, several RIs were much higher than values in inserts, especially where the manufacturer used decision limits, such as

the lipid profile. This may reflect the prevalence of hyperlipidemia in the Saudi population. The same was true of glucose and HbA<sub>1c</sub>, which apparently overlap with the decision limits for the diagnosis of diabetes. This may reflect that in our "healthy" reference population, there is a less than "healthy" group with impaired glucose tolerance. Similar observations have been reported in other countries and ethnic groups,<sup>12,13</sup> which is compatible with the growing global diabetes epidemic.<sup>14</sup> It is important that RIs are not confused with clinical decision limits (CDLs). The RIs are calculated specifically for health whereas CDLs indicate sensitivity of disease. In general, CDLs are determined by consensus. They are the thresholds above or below which a specific medical decision is recommended and are derived from receiver operating characteristic curves and predictive values.<sup>10</sup> Specifically, CDLs are based on the diagnostic question and are obtained from clinical trials designed to define the probability of the presence of a certain disease. These limits lead to decisions about how individuals with values above or below the CDLs should be treated. Therefore, RIs of those few tests we calculated should not be used for clinical decision making in place of following CDLs. Some of our RIs had much higher or lower values than that of the package inserts; LDH, for example, is known for high biological variation. Serum folate RIs may reflect the diet of Saudi population, which involves less consumption of green leafy vegetables in addition to there being biological variation. Also, some tests may need further stratification or a different stratification, because results of partitioning varied, depending on method. We adopted the partitioning specified in manufacturer's inserts in some cases even though no or a different partitioning was indicated.

Limitations of the study were that all biochemistry laboratory tests were not included. To reduce cost and produce more homogeneous RIs, multicenter collaboration may be preferable for establishing RIs in the future. For females, we were unable to generate progesterone and 17 $\alpha$ -Hydroxyprogesterone. Human chorionic gonadotropin was only generated for non-pregnant women; some tests lacking partitioning by time (morning or afternoon), position (standing or supine) or lifestyle (smoker or non-smoker), deserve fur-

ther investigation.

Though the index of individuality was below 0.6 for only 4 tests, some studies<sup>17,18</sup> have shown that if a single sample is taken, the index of individuality has no influence on the usefulness of the RIs. Therefore, these results are reliable and applicable to a general Saudi population.

#### Conflict of interest

The authors declared no conflict of interest.

## REFERENCES

1. The REALAB Project: A New Method for the Formulation of Reference Intervals Based on Current Data E. Grossi, R. Colombo, S. Cavuto, C. Franzini Clinical Chemistry. 2005;51(7):1232-1240.
2. NORDIC Reference Interval Project. P, Ruscad. 2002. <http://www.furst.no/norip/>, 2003.
3. Suggested Reference Intervals For 8 Serum Enzymes Based On Data From NORIP Database Stromme J. et.al.: [http://www.ifcc.org/ifcc-communications-publications-division-\(cpd\)/ifcc-publications/ejifcc-\(journal\)/e-journal-volumes/ejifcc-2008-vol-19/vol-19-n-2/accurate-reference-intervals-are-required-for-accurate-diagnosis-and-monitoring-of-patients/](http://www.ifcc.org/ifcc-communications-publications-division-(cpd)/ifcc-publications/ejifcc-(journal)/e-journal-volumes/ejifcc-2008-vol-19/vol-19-n-2/accurate-reference-intervals-are-required-for-accurate-diagnosis-and-monitoring-of-patients/)
4. Borai A, Ichihara K, Al Masaud A, et al. Establishment of reference intervals of clinical chemistry analytes for the adult population in Saudi Arabia: a study conducted as a part of the IFCC global study on reference values. Clin Chem Lab Med. 2016 May;54(5):843-55.
5. CLSI and IFCC. C28-A3 document; Defining, establishing and verifying reference intervals in the clinical laboratory: approved guideline-third edition, 2008; 28:1-76.
6. Ferruccio C, Hinzmann R and Panteghini M Reference intervals: the way forward. Ann Clin Biochem 2009; 46: 8-17.
7. Home Page for Professor John M. Cimbala of the Department of Mechanical and Nuclear Engineering of The Pennsylvania State University. ME 345: Instrumentation, Measurements, and Statistics; Outliers (<http://www.mne.psu.edu/cimbala/me345/Lectures/Outliers.pdf>).
8. Thomas L. Blut glucose. In: Thomas L, ed. Labor und Diagnose, 6th ed. Frankfurt/Main:TH-Books,2005;193-199.
9. Klein G, Berger A, Bertholf R, et al. Abstract: multicentre evaluation of liquid reagents for CK, CK-MB and LDH with determination of reference intervals on Hitachi systems. Clin Chem 2001; 47: suppl. A30.
10. Ozarda Y. Reference intervals: current status, recent developments and future considerations Biochem Med (Zagreb). 2016 15; 26(1): 5-1.
11. Ichihara K, Boyd JC IFCC committee An appraisal of statistical procedures used in derivation of reference intervals Clin Chem Lab Med. 2010 Nov;48(11):1537-51.
12. Bimenya G S, Byarugaba W, Kalungi S, et al. Initial attempt to establish population reference values for blood glucose and lipids in Makerere university undergraduate students Afr Health Sci. 2006; 6(4): 247-253.
13. Canadian Diabetes Association Clinical Practice Guidelines Expert Committee. Targets for glycemic control clinical practice guidelines for WHO, Global report on diabetes [www.who.int/diabetes/global-report](http://www.who.int/diabetes/global-report) April 7, 2016; 1-88.
14. Harris EK. Effects of intra- and inter-individual variation on the appropriate use of normal ranges. Clin Chem 1974; 20:1535 - 42.
15. C. Ricós, V. Alvarez, F. Cava, J. V. Garcia-Lario, A. Hernández, C. V. Jiménez, J. Minchinela, C. Perich, M. Simón. Current databases on biological variation: pros, cons and progress. Scandinavian Journal of Clinical and Laboratory Investigation 2009;59:491-500 (<http://www.tandfonline.com/doi/abs/10.1080/00365519950185229>)
16. Westgard J. 2009. <https://www.westgard.com/biodatabase1.htm>. Accessed on Feb 2 2017/
17. Petersen PH, Fraser CG, Sandberg S, et al. The index of individuality is often a misinterpreted quantity characteristic. Clin Chem Lab Med 1999; 37:655-61.
18. Petersen PH, Sandberg S, Fraser CG et al. Influence of index of individuality on false positives in repeated sampling from healthy individuals. Clin Chem Lab Med. 2001; 39(2):160-5.