Generation of Pathology Reference Intervals for Indian Population

Abhishek Shah, Rushabha Maru, Kinjal Shah, Khushali Deulkar

ABSTRACT- Almost all reference intervals currently used in India are developed by Western, European and other Asia Pacific countries. The use of these reference intervals can be misleading as India is a huge nation with enormous racial and ethnic diversity. The international guidelines on reference intervals suggest the generation of new reference intervals for local homogeneous population. This paper illustrates the literature review done on various papers having similar subject and also enlightens a solution for generation of new reference interval.

General Terms- Big data processing, Data mining, Hadoop application for clinical laboratory

Keywords- CLSI, clinical laboratory, Reference Interval Generation, Reference Population

I. INTRODUCTION

Reference intervals are the most common decision support tool used by clinical pathology laboratories for interpretation of numerical pathology reports. Health professionals use these reports to decide normality of a person. A person is said to be normal if the test results lie within the range specified by the reference interval. Thus, reference ranges play vital role in disease diagnosis. A reference interval is impacted by various factors such as age, sex, food habits, religion, ethnicity, environmental conditions, etc. It is not recommended to use reference interval developed by other countries in a country like India where there exists enormous diversity. International guidelines have recommended every laboratory that they should establish or verify the reference intervals. The laboratories can establish reference ranges on their own by testing the large number of healthy population and figuring out what appears to be "Normal" for them. However, it is critical to define the reference population. Demographically, it should match the population whose laboratory results will be compared to this reference range. Also this establishment should be done every time after a specific time period because reference ranges may change with time and also the methods. If this seems to be a difficult option for a laboratory, they can use alternative by verifying the reference ranges developed by other laboratory in the same population.

Revised Version Manuscript Received on January 04, 2016.

Abhishek Shah, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai - 400056, (Maharashtra). India.

Rushabha Maru, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai - 400056, (Maharashtra). India.

Kinjal Shah, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai - 400056, (Maharashtra). India.

Khushali Deulkar, Department of Computer Engineering, Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W), Mumbai - 400056, (Maharashtra). India.

II. LITERATURE SURVEY

Almost all laboratories in India use the reference values published either from available textbooks or from diagnostic kit inserts without giving details of the original source of the data.

It is therefore necessary to define specific reference intervals for Indian population in view of country's enormous ethnic and racial diversity.

It has been noted that geographical, ecological and personal factors affect what is 'normal' for a homogeneous group of people [2] [3]. T Malati [2], in his research, has divided Indian population in various groups on the basis of ethnicity, language, religion, gender, age and economical conditions. He has stated the importance of generating the regional reference range using varied facts and examples. Despite the progress that has been made in the conceptual aspects of reference intervals, in practice their use is still not completely satisfactory in India. Whereas, various countries like China, Saudi Arabia, Japan, America have started establishing their domestic reference ranges [4] [5].

Various experiments have been conducted for the same in the past. The techniques used in these experiments to calculate RI have been included in this paper. A brief introduction on these researches is given as follows. A brief comparison of the same is given in Table1. *Establishing Reference Intervals for Clinical Laboratory Test Results (Alex Katayev et. al.)*[6]

A computerized indirect Hoffmann method was studied for accuracy and reproducibility. The study used data collected retrospectively for 5 analytes without exclusions and filtering from a nationwide chain of clinical reference laboratories in the United States. The accuracy was assessed by the comparability of reference intervals as calculated by the new method with published peer-reviewed studies, and reproducibility was assessed by the comparability of 2 sets of reference intervals derived from 2 different data sets. There was no statistically significant difference between the calculated and published reference intervals or between the 2 sets of intervals that were derived from different data sets. A computerized Hoffmann method for indirect estimation of reference intervals using stored test results is proved to be accurate and reproducible

Reference Values of Lipid Profile for Population of Haryana Region (Yuthika Agrawal et. al.)[5]

A total of 120 apparently healthy individuals coming to a tertiary govt. hospital in Haryana for regular health check up were included in this study.

They have excluded individuals having diabetes mellitus, excessive body weight, dyslipidemias, smoking, hypertension, alcohol abuse, cardiovascular diseases, coronary bypass graft, any other chronic disease, recent surgery, diseases causing alterations in lipids, hypothyroid, hyperthyroid, drugs affecting lipid concentrations, strenuous exercise, renal diseases, hormone therapy, women on oral contraceptive and medication .They analyzed lipid profile of these apparently healthy individuals. Out of total 120 individuals, were 60 females and 60 were males. The age of reference individuals ranged from 30-85years. Average age was 55.46±1.30 yrs. They have calculated mean and Standard deviation . A reference interval for each parameter was calculated from the 95% reference intervals ranging from 2.5% and 97.5% percentiles and, arithmetic mean + 2 SD were also calculated. *CLSI-Derived Hematology Reference Intervals for Healthy Males in Eastern India (Abhijit Banerjee et. al.)*[3]

A prospective cross-sectional study was carried out in Kolkata on 528 male individual to establish reference haematological ranges for male population of ages between 20 to 59. Blood samples were collected of blood donors representing a healthy population. A reference range was obtained following standard guidelines, showing the 2.5-97.5 percentile intervals and median values for each of the haematological parameters determined for 528 male subjects. The above data were classified into three age groups of 20-29 (n=193), 30-39 (n=201), and 40- 59 years (n=134) as per NCCLS specification which requires at least 120 subjects to construct a reference range. Age group specific variation of haematological parameters was evaluated by one-way analysis of variance (ANOVA) for independent samples. Statistical difference between the obtained mean and international data for each parameter was compared by Chi-square test. Clinical acceptability or bias percentage was checked between the obtained mean of haematological parameters with international data for male subjects.

An Improved Auto-Generation System to Obtain Reference Intervals for Laboratory Medicine (Hyung Hoi Kim et. al.)[8]

In order to obtain standardized reference intervals, they developed an integrated program that can calculate, by a nonparametric method, reference intervals with using the Clinical and Laboratory Standards Institute (CLSI) processes as its guideline. they also developed a grouping interface that enables users to customize classification of each group (age, gender, blood group, race, etc) when calculating reference intervals. To verify their developed program, they compared the reference intervals of the data on 281 persons for 8 total areas, and the reference intervals was already calculated beforehand using this new program.

The Origin of Reference Intervals (Richard C. Friedberg et. al.)[7]

The study was conducted according to the Q-Probes study format , which relies on a convenience sample of clinical laboratories that subscribe to the CAP Q-Probes benchmarking program. After refinement of a standardized data collection instrument, CAP Q-Probes subscribers were mailed data collection instructions in late 2005.Participants were asked to provide their laboratories' low and high values for reference intervals for 7 analytes .Adult values and pediatric values were collected. In addition, hemoglobin reference intervals were collected for female adult and pediatric inpatients .For each analyte, the following additional information was collected: unit of measure, primary specimen type, analytic instrument manufacturer, year the reference interval was originally established, year of the most recent revalidation of the reference interval, and year the primary instrument was placed into service. The methods the laboratory used to determine reference intervals for each analyte were also ascertained .One hundred sixty-three clinical laboratories provided information about their reference intervals to them .Approximately half the laboratories reported conducting an internal study of healthy individuals to validate reference intervals for adults. Most laboratories relied on external sources to establish reference intervals for pediatric patients. There was slight variation in intervals used by the laboratories, In some cases the intervals used by 2 laboratories had no overlap. For example, one laboratory considered a hemoglobin of 13.8 g/dL in a woman to be “low” while another considered the same value to be “high.” Three percent of reference intervals contained a limit that qualified as an “outlier” using standard statistical tests; they could not identify any practice associated with adoption of outlier intervals.

III. COMPARATIVE STUDY

The study shows that there is room for doubt that the reference intervals currently used in many hospitals or health institutions are appropriate. It can be seen that the selections of reference individuals are not valid universally. Almost all the experiments show significant change in new calculated reference interval and existing international reference interval as shown in Table 1. If the inappropriate reference intervals are adopted in the decision making of examinees’ health status, it can lead to false-negatives or false-positives. Despite progress in the conceptual aspects of reference values, in practice their use is still not entirely satisfactory. The main reason behind this seems the following disadvantages:

Expensive – The procedure of reference range generation involves complex statistical analysis which must be done by a professional statistics tool which follows CLSI guidelines.

Time consuming – The process also involves selection of reference individuals from the population for different criteria (e.g. age groups, gender, various test parameters, etc.) This process is too long and needs to be repeated for new tests and testing instruments.

Unfeasible – This reality is compounded by the fact that for certain tests establishment of a reference range for different age groups and for different samples becomes impractical for small scale laboratories.

Table 1. Comparative Study Of Various Experiments Conducted To Produce Reference Range

Authors / Parameters	Alex katayev et al.[6]	Yuthika agarwal et al.[5]	Abhijit banerjee et al.[3]	Hyung hoi kim et al.[8]	Richard friedberg et al.[7]
Reference value analysis method	Hoffmann	Parametric	Anova	Non parametric	Non-parametric kruskal-wallis
No. of samples	>50,000	120	528	281random samples	163 institutions
Outlier exclusion Method	Chauvent criteria	Manual screening	Manual screening	D/R ratio	Tukey, sd procedure
Experiment location	U.s	Haryana, India	Kolkata, India	Busan, korea	97%from u.s
Selection method for reference individual	Indirect	Direct	Direct	Indirect	Indirect
Standard followed	-	IFCC	CLSI	CLSI	CLSI
Difference from existing RI	Not significant	Significant	Very significant	Significant	Significant
Reference population selection	Laboratory's Database	Manual selection	Manual selection	Hospital's Database	Laboratory's Database
Age group	>18	30-85	20-59	No restriction	No restriction

IV. PROPOSED SOLUTION

The project will help pathology laboratories to improve their quality of service given to their customers. The proposed software will help the pathologists to determine Reference Intervals (RI) for various tests conducted in their laboratories. This will be calculated on the visiting population of the respective laboratories. The software will follow international standards (CLSI) specified for the laboratory. The proposed software solution can be used by laboratories in conjunction with the existing software used by them. The software will perform statistical analysis on the data set and then provide reference interval for each laboratory parameter. To make the system reliable, all the statistical techniques described in International guidelines will be implemented and the results of all the techniques would be compared. The new reference intervals will help to improve the accuracy of test results, reducing false positives and false negatives. This would be of worth importance for the future diagnosis and treatment of linked diseases. The developed system will work with 100 GB of data approximately and the data is supposed to grow exponentially year by tear. Hence the solution is proposed to be implemented using Hadoop cluster. In Fig. 1, the client provides details asked in the GUI form and submits the job to Hadoop core node. The node divides the task into sub tasks (map function) and sends its to various nodes. These nodes compute the result and return those results to Hadoop core node. The core node combines these results into 1 resultset (reduce function). Later this resultset is returned to the user.

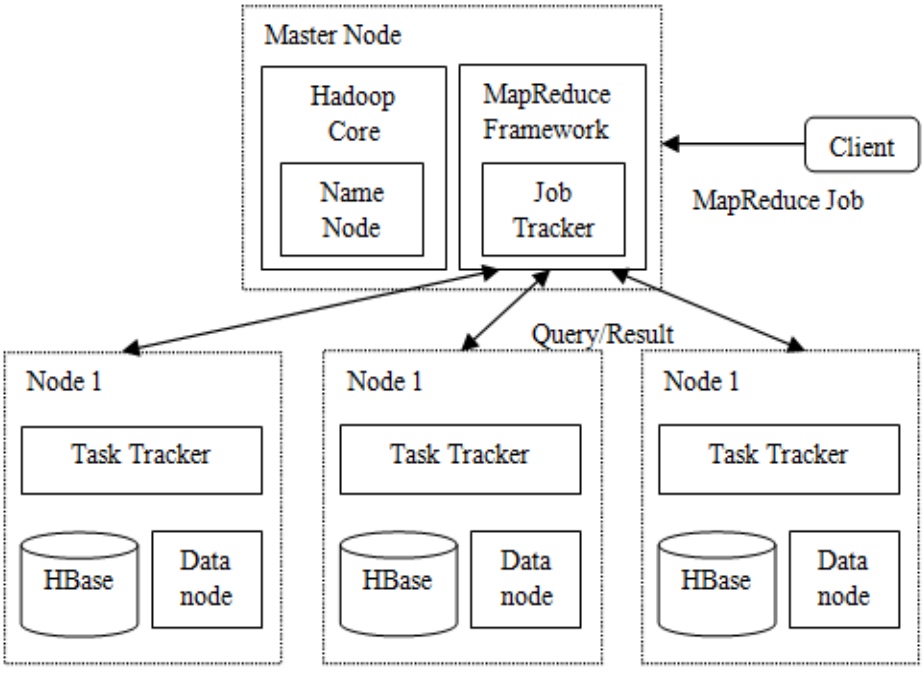

Figure 1: Architecture diagram for R.I. generation software using hadoop and hbase

The steps of R.I. calculation are described in Fig. 2. User is

loading new database in order to generate R.I. He will also select the outlier exclusion method and R.I generation method. The system will calculate R.I. based on that.

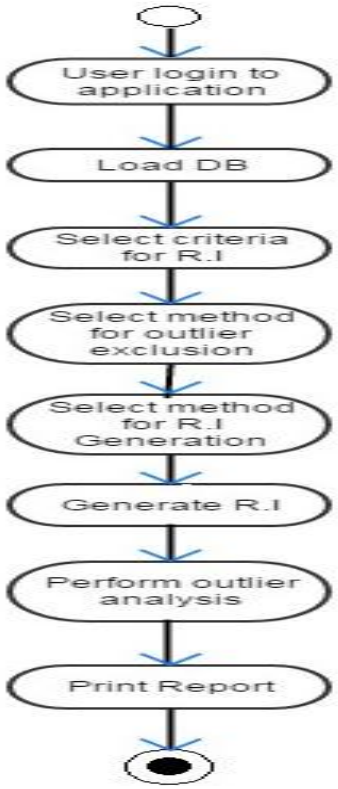

Figure 2: General steps of R.I. calculation

V. CONCLUSION

There is a room for doubt that the reference intervals currently used in many hospitals or health institutions are appropriate, some scientists do not agree that the selections of reference individuals were valid universally. This paper shows comparative study of some attempts that have been made to calculate reference intervals in respective regional areas. Only the experiments which follow CLSI guidelines are considered here. Almost all the experiments show significant change in new calculated reference interval and existing international reference interval. This proves the necessity of a software tool which can support laboratories to calculate reference intervals on their own from their existing database.

ACKNOWLEDGEMENT

Our thanks to the pathology laboratories who have shared their patients' test data, which served as the basis for this paper.

REFERENCES

1. "Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory", Third Edition, C28 – A3c, Vol. 28 No. 30.
2. T Malati, "Whether Western Normative Laboratory Values Used For Clinical Diagnosis Are Applicable To Indian Population? An Overview On Reference Interval", Indian Journal of Clinical Biochemistry, 2009.
3. Abhijit Banerjee, Diganta Dey, Parbati Banerjee, Sudarshan Ray, Ratnamala Ray, Banasri Hazra, "CLSI-Derived Hematology Reference Intervals for Healthy Males in Eastern India", Global Journal Of Medicine And Public Health, 2013.
4. Tanzeel Huma, Usman Waheed , "The Need To Establish Reference Ranges", Journal of Public Health and Biological Sciences, Vol. 2, No. 2, ISSN 2305-8668 (Print) 2307-0625 (Online), 2013
5. Yuthika Agrawal, Vipin Goyal, Kiran Chugh, Vijay Shanker , "Reference Values of Lipid Profile for Population of Haryana Region", Scholars Journal of Applied Medical Sciences, 2014.
6. Alex Katayev, MD, Claudiu Balciza,and David W. Seccombe, MD, PhD , "Establishing Reference Intervals for Clinical Laboratory Test Results - Is There a Better Way?", American Journal for Clinical Pathology, 2010.
7. Richard C. Friedberg, MD, PhD; Rhona Souers, MS; Elizabeth A. Wagar, MD; Ana K. Stankovic, MD, PhD, MPH; Paul N. Valenstein, MD, "The Origin of Reference Intervals A College of American Pathologists Q-Probes Study of ''Normal Ranges'' Used in 163 Clinical Laboratories", Archives of Pathology & Laboratory Medicine —Vol 131, March 2007.
8. Hyung Hoi Kim, MD, PhD , Hae Sook Hong, RN, PhD , Shine Young Kim, MD, MS, Tung Tran, PhD, Ji Min Lee, RN, MS, Hwa Sun Kim, RN, PhD, Hune Cho, PhD, "An Improved Auto-Generation System to Obtain Reference Intervals for Laboratory Medicine", Healthcare Informatics Research, 2010.
9. Yuthika Agrawal, Vipin Goyal, Kiran Chugh, Vijay Shanker , "Reference Values of Lipid Profile for Population of Haryana Region", Scholars Journal of Applied Medical Sciences,2014.
10.