"Data mining" for Common Reference Intervals – Serum Sodium as an Example

Graham RD Jones
Department of Chemical Pathology, St Vincent's Hospital, Sydney, Australia.
gjones@stvincents.com.au

Introduction

- Reference interval studies are time-consuming, expensive and may be difficult to perform.
- "Data Mining" techniques such as Bhattacharya analysis may allow rapid determination of population reference intervals based on existing laboratory data.
- Common reference intervals may improve laboratory performance by providing high quality intervals with reduced work for each laboratory.
- Reference intervals for serum sodium were derived using Bhattacharya analysis of outpatient data for a range of laboratories. The derived intervals were compared for consideration of a common reference interval.

Methods

- Participating laboratories submitted serum sodium results from outpatient samples for periods up to 4 months from January 2005.
- Results were not partitioned for age, sex or sample type (serum / heparin) as literature references do not support such partitioning.
- Bhattacharya analysis (see below) was used to determine the midpoint and upper and lower reference limits.
- Bhattacharya analysis was performed using an in-house spreadsheet application in Microsoft Excel. See figure 1 for example analysis.
- Laboratories were also asked to supply results for serum sodium from the RCPA-AACB General Chemistry QAP for the period of data collection (Cycle 68) and their reference intervals in routine use for adult patients. QAP data comparisons were made using the average of 4 results within standard reference intervals.

Bhattacharya Analysis

- Bhattacharya analysis is a method for identifying a Gaussian distribution in the midst of other data (1).
- This method has been used previously for setting reference intervals (2).
- Unlike standard reference interval studies using parametric or non-parametric statistics there is no requirements to exclude results from members of an "unhealthy" population prior to analysis.
- The assumptions that must be satisfied for Bhattacharya analysis to be valid are as follows:
  - A Gaussian (or Log Gaussian) distribution for results from the reference population.
  - The majority of results in the data set from patients where the analyte under analysis are unaffected by the condition for which the patient is being investigated.
  - A sufficiently large number of results, although the number is not established.

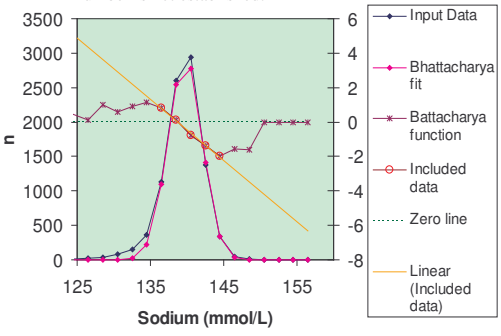

Figure 1. Example Bhattacharya analysis using the Excel application.

Data Sets

- Data sets were received from 20 laboratories, ranging in size from 541 results to 78,311 results giving a total of over 280,000 individual results.
- Results for serum sodium from the RCPA-AACB General Chemistry QAP were received from 14 laboratories covering the period of patient data collection.
- Reference interval data was obtained from 15 laboratories.
- Data was obtained from laboratories using instruments from the following manufacturers:
  - Roche (n=13)
  - Ortho-Clinical Diagnostics (n=2)
  - Beckman-Coulter (n=2)
  - Dade (n=2)
  - Abbott (n=1)

Results (1)

- Bhattacharya analysis identified the following estimates from all laboratories:
  - Average LRL: 136.1 mmol/L (SD 1.2, CV 0.9%)
  - Average midpoint: 140.1 mmol/L (SD CV 0.83%)
  - Average URL: 145.0 mmol/L (SD 1.0, CV 0.7%).
  - Average Range: 8.8 mmol/L (SD 0.6, CV 6.5%)
- There was a strong correlation between the LRL and URL suggesting bias was the main difference between the derived reference intervals (figure 2). This relationship was true for all size of data sets.
- The bias of Bhattacharya results was compared with analytical bias determined by QAP results (figure 3). The between-method analytical bias accounted for a significant component of the difference in Bhattacharya results for all data (figure 2, blue data, r^2=0.42), and an even higher component when results were limited to one manufacturer (figure 2, red data, r^2=0.79).

COMMENT: The interval size of approximately 8.8 mmol/L indicates a total SD of 2.2 mmol/L for the combined variation of analytical and within- and between-individual biological variation. This is the same as the calculated variation taking figures from Westgard's website (3) and an analytical CV of 1%.
2.2 mmol/L = 140 mmol/L x $\sqrt{(1\%^2+0.7\%^2+1\%^2)}$

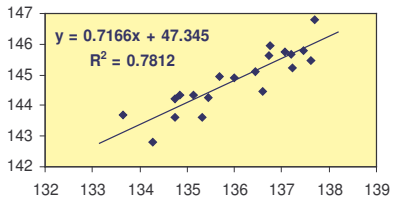

Figure 2. Upper reference limit plotted against lower reference limit for all 20 laboratories. Limits determined by Bhattacharya analysis

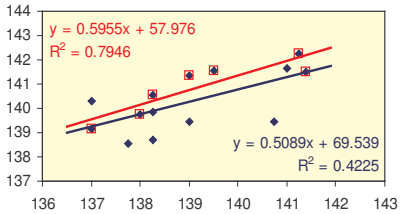

Figure 3. The center of the Bhattacharya intervals plotted against the QAP average. Blue data – all laboratories with available data. Red data – results from laboratories with Roche Hitachi analysers. Solid lines are lines of best fit.

Results (2)

- The Bhattacharya derived intervals were compared with the supplied reference intervals (figure 4a and 4b). A poor correlation was seen.

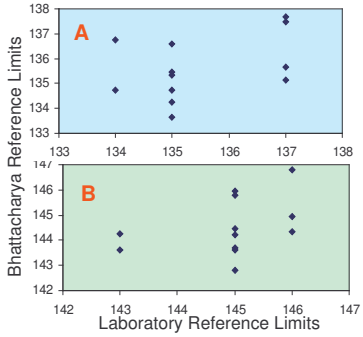

Figure 3. Comparison of Bhattacharya reference intervals with supplied laboratory reference intervals. A - Lower Reference Limits. B - Upper Reference Limits

Interpretation

If is assumed that the Bhattacharya method is a valid method for determination of sodium reference intervals from outpatient data, a number of observations can be made:

- The approximate SD of 1 mmol/L for upper and lower reference limits gives a 95% coverage of approximately 4 mmol/L (+/- 2 mmol/L) fpr variation between laboratories.
- Using the partitioning criteria of Fraser (4) to decide on the requirement for separate intervals for different labs, only those with a bias of +/- 0.8 mmol/L would be acceptable (1.0 to 5.7% of healthy population excluded). Thus the average intervals are not generally applicable unless between laboratory bias can be improved.
- An alternative view of using a common reference interval of the central 99% (+/- 2.5 SD, 135 to 146 mmol/L) would allow bias of up to 1.8 mmol/L without increasing the flagging rate (percent outside interval) but with a reduction in sensitivity for abnormalities is sodium concentration.

Conclusions

- This study suggests that data mining is a robust, simple technique for determining reference interval information, generating data without additional laboratory testing.
- The variability of estimated reference limits for serum sodium from 20 analysers from 5 manufacturers is too wide to allow adoption of common reference intervals with standard central 95% limits.
- Alternate interval definitions may need to be considered.
- Control of between-laboratory bias will be important for adoption of common reference intervals.

Acknowledgements / references

I thank the scientists at many laboratories who provided data for this analysis.

This poster produced as part of the workings of the AACB-RCPA Working Party on Common Reference Intervals.

(1) Bhattacharya, LG. Journal of the Biometric Society. 1967;23:115-135.

(2) Taylor N et al. Clin Biochem Rev 2001;23:89 (abstract)

(3) www.westgard.com.au

(4) Fraser GC. Biological variation: from principles to practice. AACC press. 2001.