Bhattacharya analysis is a useful method for isolating Gaussian distributions in the presence of other data (1) and has been used for reference interval determination using pathology databases (2).

Biological analytes are often non-Gaussian in their distributions. Examples include serum ALT, CK, proINP and amylase. These skewed distributions are sometimes commonly log-transformed to allow the use of parametric statistics or Bhattacharya analysis.

The distributions of some analytes have been shown to be skewed to a greater or lesser extent than can accurately be described by log transformation (3).

Assumptions about the underlying analyte distribution in healthy populations may influence the process of estimating reference intervals using the Bhattacharya was evaluated.

The aim of the study is the examine the effect of assumptions concerning the underlying distribution of serum ALT on reference intervals derived using Bhattacharya analysis using transformed data.

Data sets for serum ALT (males only) were obtained for healthy individuals from the NORIP and NHANES III studies as well as from outpatients from our pathology service.

Varying assumptions on the underlying distributions were evaluated using the Box-Cox transformation methods.

Bhattacharya analysis was performed with an Excel spreadsheet application. The best fit was determined by optimisation of the linearity of the Bhattacharya curve (determined by the correlation coefficient) and the percent of the data included in the analysis.

ALT Data sets:

- NHANES III
  - N=2593, highest health category only
  - ALT: IFCC method, Roche Hitachi Analyser
- NORIP (Nordic Reference Interval Project)
  - Data Source: www.furst.no/norip
  - ALT: DGRKC 1997 standardisation
  - n = 1080, healthy by pre-test questionnaire (2002)
  - NORIP conclusion: 10 - 70 U/L
- SydPath data
  - Results from patients attending General Practitioners.
  - ALT: IFCC method, Olympus analyser and reagent, Roche calibrator (2006 data)
  - Current SydPath Interval: 0 - 30 U/L

Results

Inspection of all three ALT data sets revealed heavily right-skewed distributions (figure 2).

For all three data sets, the value of $\lambda$ giving the best fit using Bhattacharya analysis, indicated by the percent of data included in the transformed Gaussian distribution and the correlation coefficient, was -0.5. (figure 3, table 1). This indicates a native distribution which is more right skewed than a logarithmic distribution.

The same transformation was optimal for the SydPath GP samples, which comprises patients presenting with unknown health issues, for the two “healthy” data sets.

The reference intervals determined were highly dependent on the assumed distribution, with the assumption of a more skewed distribution giving markedly higher upper reference limit (figure 3, table 1).

Discussion

The study has demonstrated that assumptions made about the distribution of ALT has a profound effect on the reference intervals produced using Bhattacharya analysis.

ALT can be considered an example of other skewed distributions where data transformation may be used prior to reference interval determination.

The healthy populations show no difference to the outpatient population in the nature of the distribution however the different results may indicate analytical or population differences.

Factors other than pure statistical analysis are likely to be required for reference interval determination of analytes with skewed distributions.

References

3. Home and Pesecke A. Reference Intervals, a user guide. ACC Press 2005