

Table 1: Classification of two simulated s-TSH values from 100,000 individuals according to the bivariate distribution and the combination of RCVs and univariate reference limits. Classified as normal according to the bivariate distribution were the pairs of measurement within the central 95 % of that distribution. Classified as normal according to the combination of RCVs and univariate reference limits were those pairs of measurements where the difference between them was within RCVs and both measurements were within the univariate reference range.

Combination of non-parametric RCVs and reference limits	Bivariate distribution		Total
	Normal	Abnormal	
Normal	87,608	992	88,600
Abnormal	7,392	4,008	11,400
Total	95,000	5,000	100,000

and 0.960 (95 % CI 0.957 to 0.962) for those with x_1 above the median.

The graphical study of the s-sodium data is presented in a supplementary figure. The sensitivity of the combination of univariate reference limits and RCVs for identifying pairs of x_1 and x_2 lying outside the central 95 % of the bivariate distribution of s-sodium was 100 %. The specificity for identifying pairs of x_1 and x_2 lying inside the central 95 % of the bivariate distribution was 92.5 %.

Discussion

If the physician does not know the patient's setpoint value of s-TSH and wants to judge the clinical condition from s-TSH in two samples taken a time apart, we believe that the physician basically wants to assess whether the patient is healthy and stable. Then the two s-TSH values could be compared against the bivariate distribution in Figure 1, which represents a stable, euthyreot population. As clearly shown in Figure 1, the lines of the 2.5 and 97.5 percentile univariate reference limits in combination with the 2.5 and 97.5 percentile RCVs do not accurately delineate the central 95 % of the points of the bivariate distribution of x_1 and x_2 . The space between the RCV lines contains 95 % of the points, as do the space between each set of reference limits, but the space between the RCVs and the reference limits is not congruent with the ellipse marking the central 95 % of the bivariate distribution. The RCV lines are approximately tangent to the ellipse marking the central 50 % of the distribution and cut through the other ellipses. Compared to the central 95 % of the bivariate distribution, the combination of univariate reference limits and RCVs had a fair specificity of 92.2 % but a lower sensitivity of 80.2 %. Without the assistance of univariate reference limits, the RCVs showed a particularly low sensitivity. Obviously, RCVs are not designed to detect healthiness, as the

area between the limits of RCVs includes analyte concentrations from zero to infinity (Figure 1). These considerations are not limited to s-TSH; probability density contour plots for bivariate distributions are not straight lines for any analyte, as indicated in the Supplementary Figure. In the example of s-sodium, the diagnostic accuracy of the univariate reference limits and RCVs was considerably better than for s-TSH. Obviously, how well the combination of univariate reference limits and RCVs delineate the corresponding bivariate distribution must be studied for each analyte.

Looking at the ellipses marking the various central proportions of the bivariate distribution and the line of equality (Figure 1), it is obvious that regression towards the mean does occur in this scenario. If the measured value in the first specimen (x_1) is relatively low, the measured value in the second specimen (x_2) is most likely to be higher, and vice versa. The median values of the difference $x_2 - x_1$ and the ratio x_2/x_1 for pairs with x_1 below and above the median value of x_1 showed the same phenomenon, as expected, because a difference in percent is equivalent to a ratio. We prefer ratios in this setting.

Thus, the idea of RCVs as a constant fraction of the first measurement is flawed, a finding in accordance with a previous study [7]. We estimated the RCVs both parametrically and non-parametrically, to see whether the two methods gave different results. They did not; the two methods of estimation gave almost identical RCV lines. They were symmetrical about the equality line, and asymmetrical in the y direction. We simulated the s-TSH-values assuming a Gaussian distribution around the setpoints; still, the non-parametrically derived RCVs based on the simulated values were asymmetrical in the y direction.

We believe our data set was theoretically sound. Data were derived from a Gaussian distribution truncated at ± 3 standard deviation, and transformed to lognormally distributed data as coming from an euthyreot (healthy) population. Each data pair was generated from the same, individual setpoint value, so the data represented a stable population. All data pairs were generated with the same CV_I of 17.9 % [4], thus the variance was homogeneous. The value of 3.4 % for CV_A is the total CV_A in our laboratory, estimated from quality control values over several months. Truncation at ± 3 standard deviation when generating population setpoint and individual values were done because those distributions are not really Gaussian (they do not include values from minus to plus infinity) and often values outside ± 3 standard deviation are regarded as outliers.

Note that this work deals with how the physician might interpret two s-TSH values from the same patient when the patient's setpoint value is unknown. It has no relevance if the physician needs to judge a time series of three or more measured values of the same analyte. Neither does it go