University of York Department of Health Sciences

Measurement in Health and Disease

Suggested answers: Reference ranges

Question 1.

a) *What do they mean by 'reference range'?* This is the range of values which should contain the central 95% of observations from apparently normal healthy subjects.

b) *What method do you think they have used to calculate the reference range?* There are two standard methods. One is to use the mean and standard deviation of the observations and calculate mean ± 2SD, the other is to rank the observations and use the centiles of the cumulative frequency distribution of the data directly. Here they could not use the mean ± 2SD method because the lower limit would be negative, 7.9% – 2×4.0% = –0.1%. They could have used a log transformation of the data, found the reference range on the log scale, and transformed back, which would work if the logs of the observations followed a Normal distribution. Alternatively, they could have used the direct method. The lower limit would then be at the observation ranked 0.025×(120+1) = 3.025, the 3rd observation.

Question 2.

a) *The authors found a regression equation with the LC-MS/MS method of measuring UFC as the y, dependent, or outcome variable and the ACS:180 method as the x, independent, or predictor variable to be*

$$LC\text{-}MS/MS = 0.408\ (ACS:\ 180) + 2.65$$

What is the slope and intercept for this line? What do they tell us about the agreement between these two methods of measurement? The slope is 0.408, the intercept is 2.65. If there were no bias in the methods and the points were scattered evenly about the line of equality on a scatter diagram, we expect the slope to be less than 1.0 and the intercept greater than 0.0, just as they are here, so these coefficients do not tell us much about the agreement.

b) *The correlation coefficient between the LC-MS/MS method and the ACS:180 method is given as the square, $r^2 = 0.6664$. This gives us $r = 0.82$. What does the correlation coefficient tell us about the agreement between these two methods of measurement?* Very little. Correlation depends on the variability of the sample and completely ignores bias.

c) *Why were the data log transformed?* They tell us that the distribution was skew, presumably to the right. They transformed the data to reduce this skewness.

d) *What method do you think they used to estimate the centiles?* They must have used mean ± 2SD, otherwise there is no need to transform. The direct method gives the same estimates whatever transformation is used. Also, they calculate a 95^{th} centile for only 25 men, which would be difficult to do using the cumulative frequency distribution, as it would be the highest value.

e) *What possible reason could they have for estimating the 97.5th centile using the log transformed data and the 95th centile using the untransformed data?* If they want a range which includes 95% of observations, they can get this from mean ± 2SD. On the log scale, the observations outside the range will be equally distributed above and below the limits. Hence the top of the 95% range will be the 97.5th centile. On the natural scale, it is quite likely that all the observations outside the range will above the upper limit, with none at all below the lower limit. The lower limit may even be negative. Hence the top of the 95% range will be the 95th centile.