Measurement in Health and Disease

Solution to exercise: Measurement error and observer variation

a) ICC tells us about the information content of the measurement. The closer to 1.00 the ICC is, the more information it contains. The SEM tells us about the interpretation of an individual measurement on a subject. It tells us 95% of measurements will be within 1.96×SEM of the subject's average value, and that two measurements on the same subject will differ by at most 1.96√2×SEM.

b) For the elbow measurement, we can conclude from the ICC that the elbow measurement has quite a high degree of repeatability and that it doesn't matter which observer makes the measurement. From the SEM, the observed value is likely to be within 3 (=1.96×1.57) kgf of the subject's average value. For the knee measurement, the repeatability is quite high provided the same observer is used, but it is not so good if different observers make the measurements. For a given observer, the measurement is likely to be within 3.6 (=1.96×1.84) kgf of the average measurement on that subject by the same observer, but only within 6.2 (1.96×3.16) kgf if either observer might be used.

c) If the population ICC=1, then each observation on a subject will be the same. We know that this is not true in the sample, so it cannot be true in the population from which the sample comes. Hence the upper 95% confidence limit must be less than 1.

d) For the elbow, there is virtually no difference in the ICC or the SEM for the intra-rater and inter-rater error. Using different raters does not add any variability. For the knee, there is a big difference. Using a different rater produces a substantial extra error. It is reasonable to conclude that they are not measuring the same thing, and since the patient has not changed to conclude that the observers are not measuring in the same way.

e) They are right. The ICC is calculated from the variance between subjects divided by the sum of the variances between and within subjects. The variance within subjects is fixed and the variance between subjects depends on the sample. Thus a very variable sample will produce a big ICC. The ICC is only meaningful for the population for which our subjects form a representative sample.

f) Both must be positively skew (skew to the right). There are two clues to this. First, the mean is closer to the lower end of the range than to the upper end. Second, the standard deviation is too large. For the elbow it is greater than half the mean, implying negative values, and for the knee the mean minus twice the standard deviation is considerably less than the lower limit of the range. Correlation coefficients are difficult to interpret when the variables do not follow a normal distribution, because we cannot easily calculate a valid confidence interval. There is therefore some doubt over the ICCs.

g) We are drawing conclusions about whether observers can agree from a sample of only two observers. The study would be improved by increasing the number of observers. We could also add more subjects, but it is the observers who are critical.