STATISTICS IN MEDICINE, VOL. 7, 1203 (1988)

LETTER TO THE EDITOR

HOW MUCH OF THE PLACEBO 'EFFECT' IS REALLY STATISTICAL REGRESSION?

by C. J. McDonald, S. A. Mazzuca and G. P. McCabe, *Statistics in Medicine*, **2**, 417–427 (1983)

From: *S. J. Senn*
Department of Mathematics and Computer Studies
Dundee College of Technology
Bell Steet
Dundee, Scotland

There are three points in this interesting article upon which I should like to comment.

The first concerns the formula (2) which the authors quote on p. 418 for the expected change between baseline and repeat observations. This formula is incorrect. They have

$$\Delta = E(X_2 - X_1 | X_1 \geqslant x_1) = (\mu - x_1)(1-\rho)$$

whereas the formula should read[1]

$$\Delta = E(X_2 - X_1 | X_1 \geqslant x_1) = c\sigma(1-\rho)$$

where $c = \phi(a)/[1-\Phi(a)]$, $a = (x_1 - \mu)/\sigma$, $\phi(a)$ is the ordinate of the standard Normal probability density function and $\Phi(a)$ is its distribution function. Alternatively, one could redefine Δ so that

$$\Delta = E(X_2 - X_1 = x_1) = (\mu - x_1)(1-\rho)$$

but in this case this formula cannot be used to calculate the values in Table II.

The second point concerns the authors' observation that 'by using the average of enough pretreatment measures one can reduce the size of regression to any predetermined level. 'This is based upon an optimistic view of the correlation structure for repeated observations. It assumes that a sample of repeated measurements taken in a short period before treatment can be regarded as a simple random sample from a period covering the initial assessment and the post treatment assessment.

A similar criticism is the substance of my third point. According to McDonald, Mazucca and McCabe, 'the regression effect can be eliminated entirely by two pretreatment observations of the outcome variable. In this case, the first observation is used to select for treatment and the second to measure change for treatment.' The model which justifies this observation by the authors is as follows. We have X_1 as 'selection' measurement, X_2 as 'baseline' measurement and X_3 as 'evaluation' measurement. Then,

$$E(X_2 | X_1 \geqslant x_1) = \mu + c\sigma\rho_{12}$$

and

$$E(X_3 | X_1 \geqslant x_1) = \mu + c\sigma\rho_{13},$$

therefore,

$$E[X_3 - X_2 | X_1 \geqslant x_1] = c\sigma(\rho_{13} - \rho_{12}),$$

where ρ_{12} is the correlation between X_1 and X_2 and ρ_{13} is the correlation between X_1 and X_3.

Hence, if $\rho_{12} = \rho_{13}$ there is no regression effect. McDonald, Mazzuca and McCabe themselves implicitly recognize the difficulty in this assumption saying that it is only valid if 'the baseline variation in the observation is due only to random noise, and not to drifts in the baseline, circadian rhythms or other cycles'. Since, however, we are dealing with a time series, can we ever reasonably expect that $\rho_{12} = \rho_{13}$?

REFERENCES

1. Davies, C. E. 'The effect of regression to the mean in epidemiologic and clinical studies', *American Journal of Epidemiology*, **104**, 493–498 (1976).