

Some examples of regression towards the mean

J Martin Bland, Douglas G Altman

This is the seventh in a series of occasional notes on medical statistics.

We have previously shown that regression towards the mean occurs whenever we select an extreme group based on one variable and then measure another variable for that group (4 June, p 1499).¹ The second group mean will be closer to the mean for all subjects than is the first, and the weaker the correlation between the two variables the bigger the effect will be. Regression towards the mean happens in many types of study. The study of heredity¹ is just one. Once one becomes aware of the regression effect it seems to be everywhere. The following are just a few examples.

Treatment to reduce high levels of a measurement—In clinical practice there are many measurements, such as weight, serum cholesterol concentration, or blood pressure, for which particularly high or low values are signs of underlying disease or risk factors for disease. People with extreme values of the measurement, such as high blood pressure, may be treated to bring their values closer to the mean. If they are measured again we will observe that the mean of the extreme group is now closer to the mean of the whole population—that is, it is reduced. This should not be interpreted as showing the effect of the treatment. Even if subjects are not treated the mean blood pressure will go down, owing to regression towards the mean. The first and second measurement will have correlation $r < 1$ because of the inevitable measurement error and biological variation. The difference between the second mean for the subgroup and the population mean will be approximately r times the difference between the first mean and the population mean. We need to separate any genuine reductions due to treatment from the effect of regression towards the mean. This is best done by using a randomised control group, but it can be estimated directly.²

Relating change to initial value—We may be interested in the relation between the initial value of a measurement and the change in that quantity over time. In anti-hypertensive drug trials, for example, it may be postulated that the drug's effectiveness would be different (usually greater) for patients with more severe hypertension. This is a reasonable question, but, unfortunately, the regression towards the mean will be greater for the patients with the highest initial blood pressures, so that we would expect to observe the postulated effect even in untreated patients.³

Assessing the appropriateness of clinical decisions—Clinical decisions are sometimes assessed by asking a review panel to read case notes and decide whether they agree with the decision made. Because agreement between observers is seldom perfect the panel is sure to conclude that some decisions are "wrong." For example, Barrett *et al* reviewed cases of women who had had a caesarean section because of fetal distress.⁴ The percentage agreement between pairs of observers in the panel varied from 60% to 82.5%. They judged a caesarean section to be "appropriate" if at least four of the five observers thought a caesarean should have been done. Because there was poor agreement among the panel, judgments by panel members and the actual obstetricians doing the sections must also be poorly related and not all caesareans will be deemed appropriate by the panel. The authors concluded that 30% of all caesarean sections for fetal distress were un-

necessary, but what the study actually showed was that decisions about whether women should have emergency surgery for fetal distress are difficult and that obstetricians do not always agree.⁵

Comparison of two methods of measurement—When comparing two methods of measuring the same quantity researchers are sometimes tempted to regress one method on the other. The fallacious argument is that if the methods agree the slope should be 1. Because of the effect of regression towards the mean we expect the slope to be less than 1 even if the two methods agree closely. For example, in two similar studies self reported weight was obtained from a group of subjects, and the subjects were then weighed.^{6,7} Regression analysis was done, with reported weight as the outcome variable and measured weight as the predictor variable. The regression slope was less than 1 in each study. According to the regression equation, the mean reported weight of heavy subjects was less than their mean measured weight, and the mean reported weight of light subjects was greater than their mean measured weight. We have a finding which allows a simple and attractive, but misleading, interpretation: those who are overweight tend to underestimate their weights and those who are excessively thin tend to overestimate their weights. In fact we would expect to find a slope less than 1, as a result of regression towards the mean. If self reported and measured weight were equally good measures of the subject's true weight then the slope of the regression of reported weight on measured weight will be less than 1. But the slope of the regression of measured weight on reported weight will also be less than 1. Now we have the opposite conclusion: people who are heavy have overestimated their weights and people who are light have underestimated theirs. Elsewhere we describe a better approach to such data.⁸

Publication bias—Rousseeuw notes that referees for papers submitted for publication do not always agree which papers should be accepted.⁹ Because referees' judgments of the quality of papers are therefore made with error, they cannot be perfectly correlated with any measure of the true quality of the paper. Thus when an editor accepts the "best" papers for publication the average quality of these will be less than the editor thinks, and the average quality of those rejected will be higher than the editor thinks. Next time you are turned down by the *BMJ* do not be too despondent. It could be just another example of regression towards the mean.

- 1 Bland JM, Altman DG. Regression towards the mean. *BMJ* 1994;308:1499.
- 2 Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *Am J Epidemiol* 1976;104:493-8.
- 3 Hayes RJ. Methods for assessing whether change depends on initial value. *Statistics in Medicine* 1988;7:915-27.
- 4 Barrett JFR, Jarvis GJ, Macdonald, HN, Buchan PC, Tyrrell SN, Lilford RJ. Inconsistencies in clinical decision making in obstetrics. *Lancet* 1990;336:549-51.
- 5 Esmail A, Bland M. Caesarian section for fetal distress. *Lancet* 1990;336:819.
- 6 Kuskowska-Wolk A, Karlsson P, Stolt M, Rössner S. The predictive value of body mass index based on reported weight and height. *Int J Obesity* 1989;13:441-3.
- 7 Schlichting P, Høiland-Carlson, Quaade F. Comparison of self reported height and weight with controlled height and weight in women and men. *Int J Obesity* 1981;5:67-76.
- 8 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;ii:307-10.
- 9 Rousseeuw PJ. Why the wrong papers get published. *Chance* 1991;4:41-3.

Department of Public Health Sciences, St George's Hospital Medical School, London SW17 0RE
J Martin Bland, reader in medical statistics

Medical Statistics Laboratory, Imperial Cancer Research Fund, London WC2A 3PX
Douglas G Altman, head

BMJ 1994;309:780