Regression to the Mean

Stephen Senn, Competence Centre for Methodology and Statistics

Introduction

Regression to the mean is the tendency for individuals to have values on re-measurement that are closer to the average than they were when originally measured. It is a powerful potential source of bias in uncontrolled studies.

It arises as a consequence of the way that values are selected and followed up. As such it is a purely statistical phenomenon. It is extremely common and is responsible for many errors in interpreting study outcomes.

Example

Figure 1 shows simulated data for a particular population for which diastolic blood pressure (DBP) has been measured on two occasions, *baseline* (on the X axis) and *outcome* (on the Y axis). Any given subject's two readings can be represented as a point in the XY plane. A common but arbitrary standard of 95mm Hg has been chosen as the boundary between normotensive and hypertensive. Subjects who are normotensive on both occasions are plotted using a blue circle. Those who are hypertensive on both occasions are plotted using a red diamond. Those who are normotensive on one occasion and hypertensive on another are represented by orange stars. The figure represents readings for 1000 subjects in total.

The figure is roughly symmetrical and, indeed, since it is simulated from a distribution that *is* symmetrical, any departure from symmetry is purely due to chance. The consequence of the symmetry is that on average the blood pressure reading at outcome is the same as that at baseline, which is to say about 90mmHg. In fact, it is a pretty pointless thing to do, but just for the record, if one does a t-test on the differences, outcome minus baseline, the result is not significantly different from zero (P=0.61).

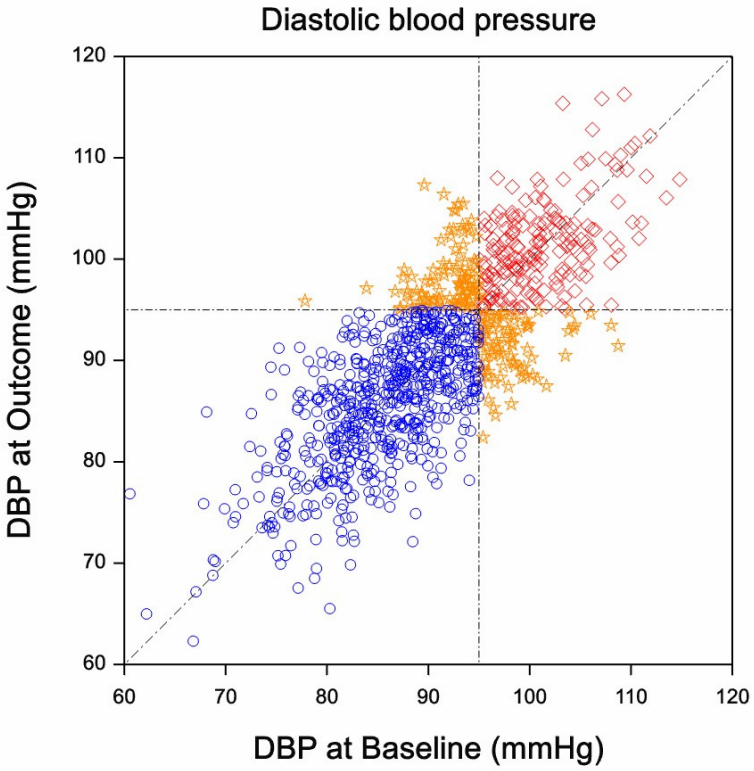

Figure 1 DBP readings for 1000 subjects on two occasions

Now suppose that we decide to remove all subjects who are normotensive on both occasions. If we do that, we obtain the rather curious plot in Figure 2. Although it is very strange, at least one can say it is still symmetric and in fact, the mean at baseline for the subjects so selected is 98.0 mmHg and at outcome is 98.2 mmHg, so very similar.

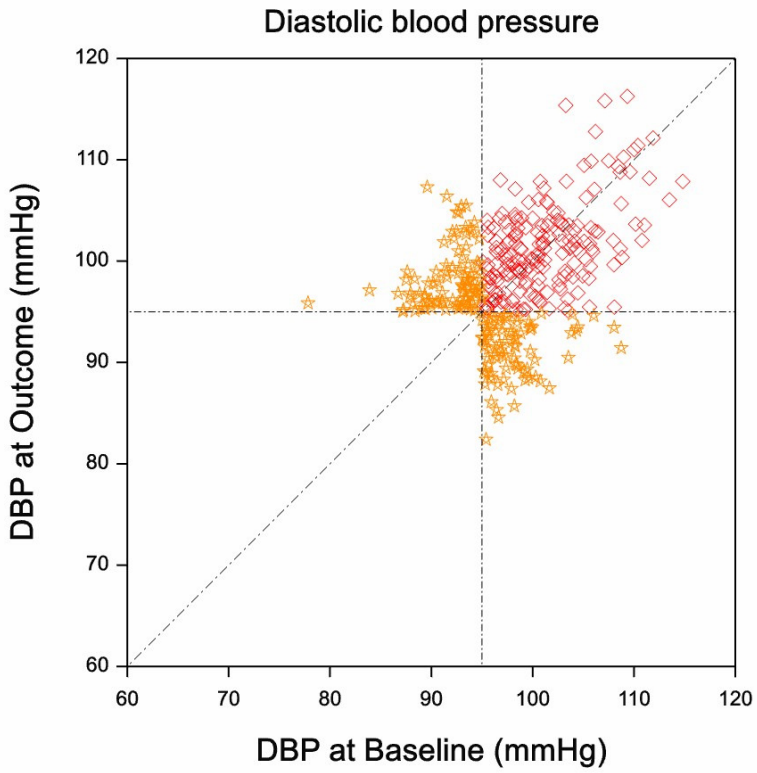

Figure 2. DBP readings for subjects who were hypertensive on at least one occasion.

In fact, however, Figure 2 is a figure we would hardly ever see in practice. Instead we would be far more likely to see something like Figure 3. This figure represents what we would see if , having measured subjects at baseline, we only chose to follow up those that had a baseline measured value in excess of 95 mmHg. This is the sort of thing we might do if we were only interested in hypertensive subjects.

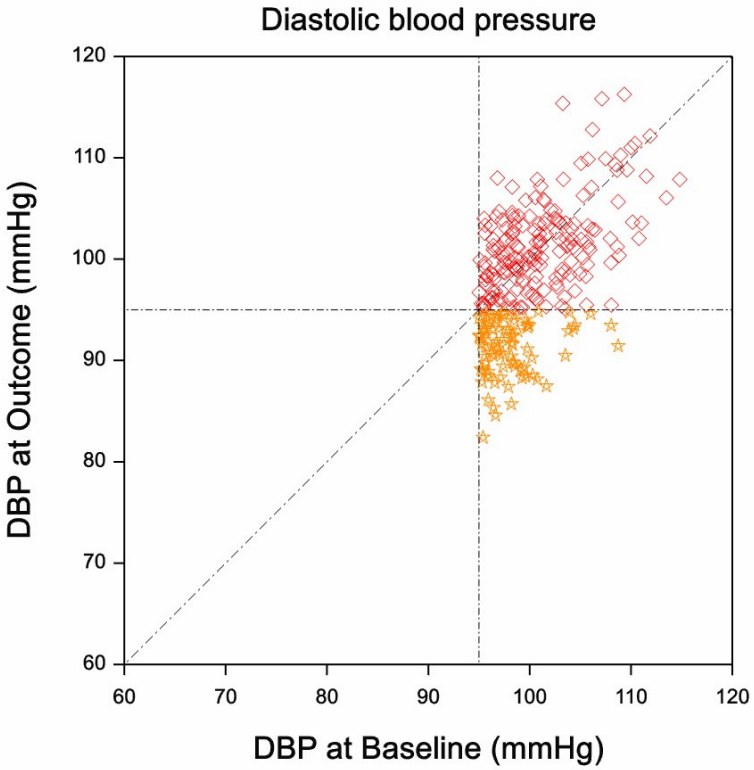

Figure 3. DBP readings at baseline and outcome for individuals who have been chosen because apparently hypertensive at baseline.

Now the picture is no longer symmetric. The mean value of the readings at outcome is no longer the same as the mean at baseline. Indeed, if a t-test of the differences is performed then the result is highly significant ($P < 0.001$) and the 95% confidence interval is well away from zero being -2.63 to -1.36 mmHg.

The consequence is that data like those shown in Figure 3 can be very misleading unless great care is taken to interpret them. However, unfortunately, data like these are extremely common. They are likely to occur in any study with inclusion criteria based on baseline values.

Explanation

There are two factors contributing to the errors of interpretation that are commonly made. The first is that the inevitable random variation in measurements tends to be forgotten. An individual's blood pressure will vary from occasion to occasion. This means that some individuals will appear to have an increase over time and some a decrease. The second factor is that the method of selection of data makes it more likely that we will see 'improvers' rather than 'worseners'. Consider Figure 4. These show individuals who were 'normotensive' at baseline and 'hypertensive' at outcome. The typical method of choosing individuals for study represented by Figure 3 excludes such individuals. However the method of selection does not exclude individuals who were hypertensive at baseline but normotensive at outcome. Hence there is an inherent bias *if the way that effects are judged is by looking over time*.

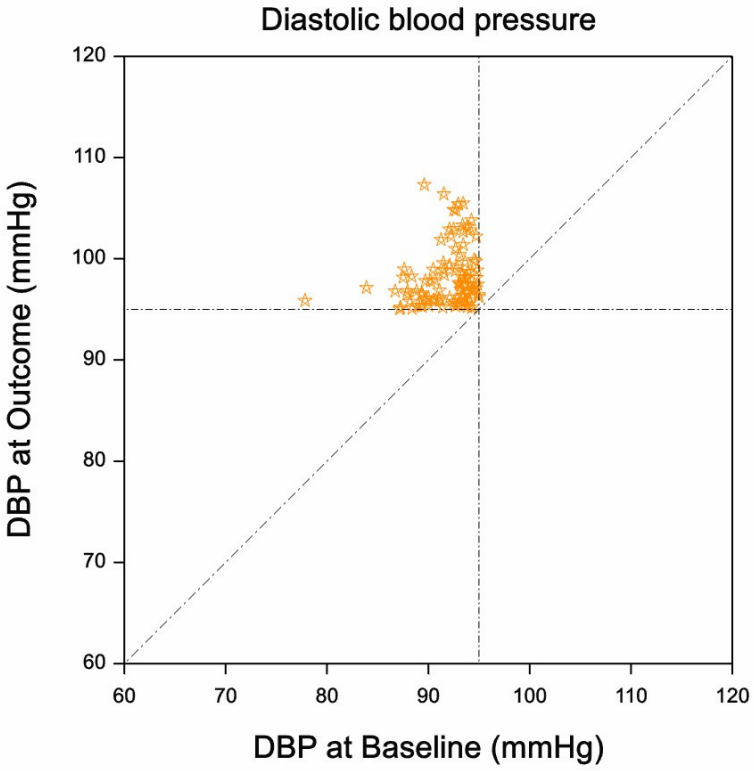

Figure 4 DBP readings for those subjects who were 'worseners'.

If we were able to add the individuals represented by Figure 4 to Figure 3 we would obtain Figure 2 and hence have again symmetry between baseline and outcome. However the nature of study designs means we will not be able to do so. Hence the difference to baseline is misleading.

Discussion

Regression to the mean is an extremely common phenomenon. Luckily it is not a problem in controlled experiments *provided that* interpretation is restricted to comparing outcomes between treated and control groups. Unfortunately trialists often perform within-group comparisons back to baseline. Regression to the mean is one reason, amongst many, as to why this is a very foolish thing to do. If the result is significant it may mean no more than '*you chose the values in a biased way that makes this sort of comparison significant even if nothing interesting is happening*'.

If individuals are chosen for study because they have extreme measured values then when measured again they make be expected to have values that are closer to the mean. If the investigator then falls into the regression trap of regarding the baseline values as being an adequate way to define response, such natural statistical improvement is likely to be extravagantly attributed to the effect of treatment.

Further reading

Senn, S. (2011). "Francis Galton and regression to the mean." Significance **8**(3): 124-126.

Senn, S. J. (2009). "Three things every medical writer should know about statistics." The Write Stuff **18**(3): 159-162.

Senn, S. J. (2007). Statistical Issues in Drug Development. Hoboken, Wiley.

Statistical Appendix

Suppose that we have a stable population of values. This will not always be exactly true but it is a reasonable enough approximation in many cases and it suffices to show the problem. In that case the mean and variances at outcome are the same as at baseline so that we may write $E[Y]=E[X]=\mu$ and $Var[X]=Var[Y]=\sigma^2$. Now suppose that *Y* is linearly related to *X* but imperfectly so, so that we have $Y=a+bX+\varepsilon$, where ε is a disturbance term independently distributed of *X* with $E[\varepsilon]=0, Var[\varepsilon]=\sigma_\varepsilon^2$. We then have that $E[Y]=a+bE[X]+E[\varepsilon]=a+b\mu$. However, we also know that this is equal to μ so that we have $\mu=a+b\mu, a=(1-b)\mu$.

However, we also have that $Var[Y]=b^2Var[X]+Var[\varepsilon]$ so that $\sigma^2=b^2\sigma^2+\sigma_\varepsilon^2$. Dividing by σ^2 and rearranging terms we have $b^2=1-\sigma_\varepsilon^2/\sigma^2$. Since b^2 cannot be less than 0 we must have that $\sigma_\varepsilon^2<\sigma^2$ and (except when *Y* is perfectly related to *X* so that $\sigma_\varepsilon^2=0$) we must also have $b^2<0$ from which it follows that $-1<b<1$ with, in practice, $0<b<1$.

This is not really surprising since we know that in general the regression coefficient may be written $b=\rho\sigma_y/\sigma_X$ where $-1\le\rho\le1$ is the correlation coefficient and since we have here that $\sigma_Y=\sigma_X=\sigma$ it follows that $b=\rho$.

Finally, we can re-arrange our original equation to read

$$
\begin{aligned}
Y &= a+bX+\varepsilon \\
Y &= (1-b)\mu+bX+\varepsilon \\
Y &= \mu+b(X-\mu)+\varepsilon \\
Y-\mu &= b(X-\mu)+\varepsilon.
\end{aligned}
$$

Since we have just shown that $b<1$ we have that we may *expect* that any randomly chosen value will be closer to the mean at outcome than it was at baseline. Only if we choose the X values so that on average they are equal to μ will the *Y* values on average be equal to the average of the X. If we choose the X values so that on average they are not equal to the population mean, the *Y* values will be closer to this mean than the X values were.

This is the regression to the mean phenomenon.