Technical Note: Survivor treatment selection bias and all that

The article about Oscar winners living longer mentioned *lead time bias* and *survivor treatment selection bias*. Both of these are usually defined with reference to studies of new screening tests for diseases that take a long time to surface, such as cancer and heart disease. I looked in the index of my medical statistics text and found a long list of possible biases, and on a handy web page`http://www.gpnotebook.co.uk/simplepage.cfm?ID=1624244292` found some more. Here is a list, with abbreviated definitions.

lead time bias (also called length/time bias): Patients screened for a disease appear to have better survival times than unscreened patients, but only because their survival is measured from the time of screening, whereas for the unscreened patients survival is measured from the time that the disease appeared. A fair comparison of the two groups subtracts this 'lead time' from the survival times of the screened patients. Redelmeier and Singh describe this as an 'unstable definition of time zero'.

survivor treatment selection bias (selection bias for short): This is very closely related to lead time bias, but in my opinion a bit more subtle. Patients destined to live longer have more opportunity for treatment, so the treated patients appear to have better survival than the untreated patients. The heart transplant example is the clearest: patients who got heart transplants had to live long enough to get a donor heart, so if they are compared to the non-transplant group they will appear to have better survival. The solution is to include eventually transplanted patients in the 'no-transplant' group until the day of their transplant. The technical term for this is 'time dependent covariates'. The criticism of Redelmeier and Singh's work hinged on the fact that they did not control very well for this bias. They did do an analysis using time-dependent covariates (and their "Oscar advantage" nearly vanished when they did) but their main analysis was indeed subject to this bias. On the Chance News website[1]this is explained as follows: "We note that 100 percent of the Oscar winners live to be at least 30 years old. Of course this is not surprising because they are known to be Oscar winners. Thus we know ahead of time that the Oscar winners will live longer than a traditional life table would predict. This gives them an advantage in their life expectancy. This is called a selection bias or Immortal bias. Of course the controls also have an advantage because we know that were in a movie at about the same age as a nominee. But there is no reason to believe that these advantages are the same."

In lead time bias the correction needed is to subtract the lead time, so that the two sets of times are fairly compared. With selection bias the correction

[1] `http://chance.dartmouth.edu/chancewiki/index.php/Oscar_winners_do_not_live_longer`

needed is to define the 'groups' differently. The group "Oscar winners" has a time advantage over the non-winners, because membership in the group is defined retrospectively.

recall bias: In a case-control study, cases (patients) are asked about their exposure to a potential risk factor in their past. Patients tend to recall their exposure differently than healthy controls, possibly because they are looking for a possible cause.

publication bias: Studies that show a beneficial effect of treatment are more likely to be published than studies showing no difference between treatment and control.

surveillance bias: High risk groups may be studied more intensively than low risk groups, leading to a higher probability of detection of some disease or condition simply by chance, not due to the risk factor.

ascertainment bias: Screening appears to increase the rate of illness in the population, but only because it is detected more often. Some people think the rise in the incidence of autism is at least partially due to ascertainment bias. Ascertainment bias and surveillance bias are very similar to lead time bias, but they refer to studies that try to assess the incidence of a condition in the population, whereas lead time bias and selection bias refer usually to comparison of two groups.

No wonder people find statistics confusing!! The bottom line is, if you are comparing two (or more) groups in an observational study, you need to think of all the ways the two groups could be different, besides the fact that they are in two different groups. Those differences can affect the outcome you want to measure in subtle ways. This is why the 'randomized, controlled, double-blind clinical trial' remains the gold standard.