6.2: Truth Inflation

Medical trials also tend to have inadequate statistical power to detect moderate differences between medications. So they want to stop as soon as they detect an effect, but they don't have the power to detect effects.

Suppose a medication reduces symptoms by 20% over a placebo, but the trial you're using to test it does not have adequate statistical power to detect this difference. We know that small trials tend to have varying results: it's easy to get ten lucky patients who have shorter colds than usual, but much harder to get ten thousand who all do.

Now imagine running many copies of this trial. Sometimes you get unlucky patients, and so you don't notice any statistically significant improvement from your drug. Sometimes your patients are exactly average, and the treatment group has their symptoms reduced by 20% – but you don't have enough data to call this a statistically significant increase, so you ignore it. Sometimes the patients are lucky and have their symptoms reduced by much more than 20%, and so you stop the trial and say "Look! It works!"

You've correctly concluded that your medication is effective, but you've inflated the size of its effect. You falsely believe it is much more effective than it really is.

This effect occurs in pharmacological trials, epidemiological studies, gene association studies ("gene A causes condition B"), psychological studies, and in some of the most-cited papers in the medical literature.[30, 32] In fields where trials can be conducted quickly by many independent researchers (such as gene association studies), the earliest published results are often wildly contradictory, because small trials and a demand for statistical significance cause only the most extreme results to be published.[33]

As a bonus, truth inflation can combine forces with early stopping rules. If most drugs in clinical trials are not quite so effective to warrant stopping the trial early, then many trials stopped early will be the result of lucky patients, not brilliant drugs – and by stopping the trial we have deprived ourselves of the extra data needed to tell the difference. Reviews have compared trials stopped early with other studies addressing the same question which did not stop early; in most cases, the trials stopped early exaggerated the effects of their tested treatments by an average of 29%.[3]

Of course, we do not know The Truth about any drug being studied, so we cannot tell if a particular study stopped early due to luck or a particularly good drug. Many studies do not even publish the original intended sample size or the stopping rule which was used to justify terminating the study.[43] A trial's early stoppage is not automatic evidence that its results are biased, but it *is* a suggestive detail.