

DATA DREDGING



Slice your data in enough different ways and you'll observe some correlations purely as a result of chance. Data dredging is the failure to acknowledge that the correlation was in fact the result of chance.



Tests for statistical significance only work if you've defined your hypothesis upfront. Historically, this has been a problem with clinical trials where researchers have 'data-dredged' their results and switched what they were testing for. It explains why so many results published in scientific journals have subsequently been proven to be wrong. To avoid this, it's now becoming standard practice to register clinical trials, stating in advance what your primary endpoint measure is.

