

Lecture notes:

The decision is (not) obvious: [https://docs.google.com/presentation/d/1o-WUGyT74xgY4iRP\\_G1Z89VB7znMh15fUsfLrR\\_vrf4/edit?usp=sharing](https://docs.google.com/presentation/d/1o-WUGyT74xgY4iRP_G1Z89VB7znMh15fUsfLrR_vrf4/edit?usp=sharing)

Estimation target is

relevant: [https://docs.google.com/presentation/d/1tzGIR42QSvxw5OqLOdreBBoqGW1pguUsnhFb\\_SjH8FE/edit?usp=sharing](https://docs.google.com/presentation/d/1tzGIR42QSvxw5OqLOdreBBoqGW1pguUsnhFb_SjH8FE/edit?usp=sharing)

Ideally, when one is done with an analysis, the decision to be made from the data is obvious. Here we discuss two common instances where the decision is far from obvious.

First, we consider the instance where the results are equivocal. For example, p-values are around 0.05 where a 5% error rate is the standard. Secondly, we consider the instance where even if the decision is clear, the outcome can't be measured so that the analysis was performed on a surrogate variable.

These concerns over the reliability of data for making decisions are on top of all of the other concerns already addressed, such as unmeasured confounders, sampling bias and other forms bias that may have caused cloud a significant or non-significant finding. However, in previous lectures we gave some remedies for these problems so let's confine ourselves to the first two concerns from above.

It is interesting to note that there is no universal consensus on the role of sample size in the significance of a statistical finding. Does a marginally significant finding gain credibility from a large sample size or does it suffer because it should have been more clear with this much data to interrogate hypotheses? Richard Royall wrote a fascinating paper on this idea titled [The effect of sample size on the meaning of significance tests](#). It's a somewhat challenging paper, but it covers this odd quirk of hypothesis testing well if you'd like to read it.

One statement about sample size is clear, if your results are null and your sample size was very low, the null results are not surprising. This is because the study wasn't set up for success. This is the idea of **power**. Power is the probability of rejecting the null hypothesis when it's false. You want more power. Low powered studies are likely not going to reject regardless of whether the null is true, just because of variability. Power is under control at the time of design of an experiment via the sample size. However, after the experiment is performed, there isn't much to be done about power. Often people get the idea of calculating power after the experiment has been done to try to differentiate between non-significant results due to lack of power or true non-significance. This is a bad idea (see this manuscript <http://amstat.tandfonline.com/doi/abs/10.1198/000313001300339897>). Potentially at your disposal, however, is conducting new studies, or getting more data for the existing study.

The second problem that we consider is one where the desired outcome is not measurable, but some surrogate is. This is extremely common. For example, BMI for body fat percentage, GDP for economic health, food frequency questionnaires instead of actual calorie consumption and so on. The use of a surrogate can occur for the outcome or the predictors or both. If the surrogate variable is used as a predictor, the field of **measurement error** has several tools available (see this great book <http://www.amazon.com/Measurement-Error-Nonlinear-Models-Perspective/dp/1584886331>). A particularly useful one is called **SimEx**. As an outcome, the problem is called **surrogate outcomes** (see these nice notes <http://depts.washington.edu/ssbiost/PRESENTATIONS/DeMets.pdf>). Ideally, you'll know that the surrogate is unbiased around the true outcome and know the variance. This could occur if you could do a gold standard

study. If you don't know it, the second best case would be to have the data to estimate it. For example, in our BMI example, you could measure body fat percentage and BMI on a subset of your sample.

In the absence of any calibration data to evaluate your surrogate, you are left with either modeling via assumptions or **sensitivity analyses**.

Finally, if your surrogate variable is such an unreliable of an estimate of your actual outcome, one must come to the conclusion that it's better to not conduct the study at all.