Data Support for Swedish Pharmacy Residents

Jack B. Huber, Ph.D.

May 2022

Purpose of this Document

The purpose of this document is to assemble some advice and resources to support Swedish Pharmacy residents in their journey from project proposal to a completed research project.

Planning your Project

The causal logic of your project

Fundamentally, the resident research project is an effort to gather causal evidence for the claim that a particular treatment or way of utilizing medicine "worked" by being more effective than a rival or pre-existing treatment. To plan data collection for the strongest causal evidence is the purpose of research design [Campbell and Stanley, 1963].

The ideal research design would be a pure experiment with random assignment. The resident would randomly assign patients to treatment and control conditions, then compare outcomes following treatment. In such a scenario the resident could attribute any significant difference in outcomes to the treatment because the groups would differ only by chance in all other ways. Any pre-existing differences would be statistically insignificant.

Because the resident cannot randomly assign patients to conditions going forward, the project falls under the category of ***quasi-experiment***. This means it is more vulnerable to confounding explanations of any difference in outcomes.

The resident research project will culminate in a comparison between two groups: a pre-implementation group and a post-implementation group. One difference between the groups is time, and recent history has made a big difference. Compare a pre-implementation group from the onset of the pandemic with a high incidence of COVID-19 to a more recent post-implementation group with much lower incidence of the virus. This difference between the two groups in COVID-19 infection may combine with the difference in drug dosing to influence treatment outcomes. This difference in COVID-19 infection could therefore be a ***confound***.

The point is to clarify:

- What is the outcome to improve? This is the dependent variable.
- What is the difference in treatment intended to cause the improvement in the post-implementation group? This is the independent variable.
- All other variables are control variables. They should differ only by chance. If there is a noticeable pre-existing difference, and the level of that factor

in one group affects the outcome, it is a confound.

How many patients?

Possibly the most pressing question for resident research projects is: How many patients do I need? The resident research project will culiminate in a series of comparisons between the pre-implementation and post-implementation groups. Outcomes of the two samples will differ by at least some quantity. The resident researcher expresses this difference as an effect, like this:

Assume that this effect suggests more favorable outcomes for the post-implementation group. This effect raises several questions:

- How do we evaluate this effect?
- Could we attribute it to chance? (because it would be very unlikely for both groups to have exactly the same outcomes)
- Or is it larger than that?

In statistical terms, this is a question of **statistical power**. Power is the ability to isolate a treatment effect when it really does exist [Cohen, 1988]. Power is a function of effect size, sample size, and statistical significance. In order to decide on a number of patients we need to have a sense of what size of effect we want to reliably detect.

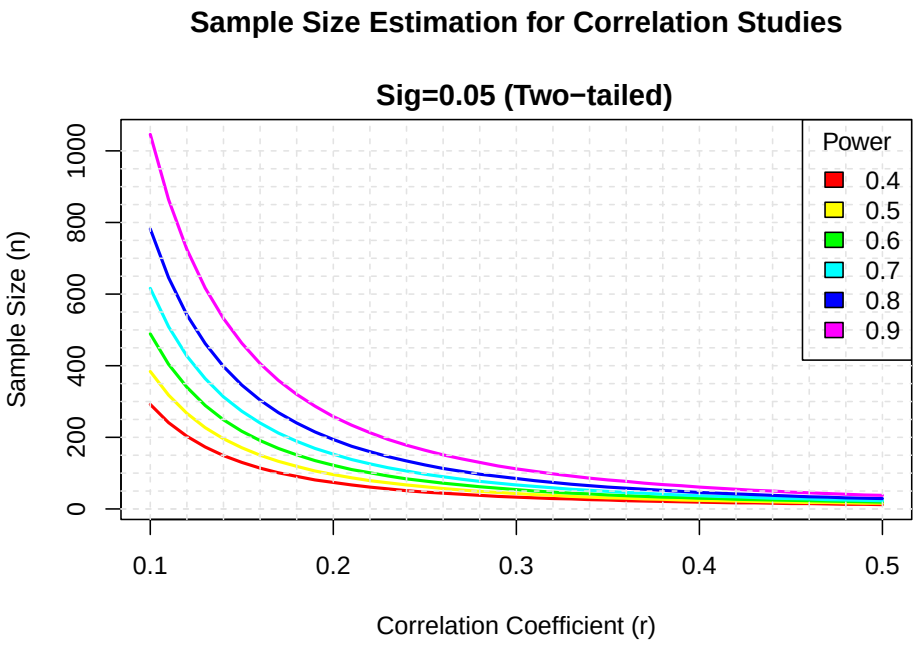


Figure 1: Power Analysis

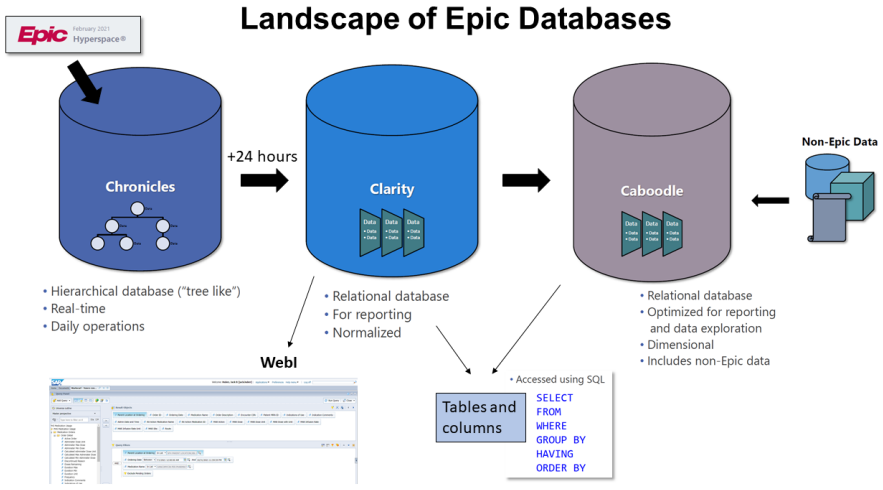
Power is a function of sample size, effect size, and statistical significance. Figure (#fig:power) is a plot of power rates against these other variables. A large effect ($r \sim 0.5$) is detectable with a sample of any size. But only large samples have the power to detect a statistically significant ($p < .05$) small correlation ($r = 0.1$).

Collecting your Data

Data Sources

For collecting data from Epic, it might be helpful to have a sense of the landscape of its different databases. There are three primary databases:

1. **Chronicles**. This is the database that is collecting data from Hyperspace in real time. For reporting, Reporting Workbench pulls data directly from Chronicles, but it is otherwise not designed very well for historical reporting.
2. **Clarity**. This is the relational database for reporting. Through a nightly process known as ETL ("extract-transform-load"), it extracts data from Chronicles and stores it in a thousand bazillion tables, and these tables are linked together by keys. To pull data from Clarity requires the use SQL language and the identification of the correct tables and fields. The SQL query selects the fields to report, from the appropriate tables (joined by common fields) with the appropriate selection criteria.



3. **Caboodle**.

Granularity

- Patient-level data
- Encounter level data
- Medication administration level data
- Lab results level data

Advice for Data Collection

1. Define the most critical variables.
2. Study an exemplar (provided).
3. We need to figure out version control. Data proliferation and overload.

Analyzing your Data

You have data. Congratulations! Let's analyze!

Preparing your data for analysis

The first step is to complete a number of tasks to prepare your data for analysis. One data matrix. See the exemplar.

Choosing appropriate statistics

It is important to know the measurement level of your variables. How do you express the outcome by which to compare your pre- and post- samples? Is it…

- Mortality rate (% surviving)? In such a case you would be comparing two proportions.
- "Time to…" a therapeutic level? In such a case you would be comparing two different quantities of time.

Odds ratios

Chi-square test of independence

Two-sample t-test

Z test of the difference of proportions

Cohen's *d* effect size

Reporting your Results

Advice on graphs

Advice on tables

Bibliography

Donald T. Campbell and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin Company, 1963.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2nd edition, 1988.