ACCP Critical Care PRN: Guide to Sample Size Calculation

PURPOSE
To provide a brief introduction to pharmacy residents completing sample size calculations.

DISCLAIMER
This document is to be used as an introductory guide and is not intended to replace a trained professional, statistician, or experienced researcher and is not intended to be the sole resource for individualized resident research projects. More complex sample size calculations (e.g., non-inferiority, >2 samples) are not included in this guide.

ACKNOWLEDGEMENTS
This guide was created by members of the Critical Care PRN Research Committee with advanced education in statistical analysis in 2018. We would like to acknowledge the following members for their assistance in creating this resource: Melissa Thompson Bastin, PharmD, BCPS; Brittany Bissell, PharmD, BCCCP; Benjamin Hohlfelder, PharmD, BCPS; Joshua DeMott, PharmD, MSc, BCPS, BCCCP; Kendall Gross, PharmD, BCPS, BCCCP; Zachary Smith, PharmD, BCPS, BCCCP; Adrian Wong, PharmD, MPH, BCPS, BCCCP (chair).

INTRODUCTION
When completing research using null hypothesis significance testing (NHST), sample size calculations are necessary to demonstrate that the observed differences in outcomes are not a result of chance (or error).

CORE CONCEPTS
Sample size calculations are a function of power, significance level, and effect size:

Power ($1-\beta$):
- Definition: The probability that the study will detect a true difference in the outcome between treatments
- Typically set at 80%
- Resident pro-tips:
    - Can set higher than 80% (e.g., 90%) if expected drop out rates are high or estimated effect size from previous literature differs from expected clinical rates.
    - Type II error = β = "humble error"
        - Accepting the null hypothesis when it is false (e.g., being modest by saying that the effect isn't significant when in reality it is)

Significance level (α):
- Definition: The probability of observed difference being from random chance alone
- In medical literature, almost always set at $\alpha < 0.05$ for statistical significance (unless evaluating composite endpoints, considering other multiplicity issues, etc.)
- Resident pro-tips:
    - Has other names that mean about the same thing: *p*-value, type I error, or "statistical significance"
    - Type I error = "cocky error"
        - Rejecting the null hypothesis ($p<0.05$) when it is true (e.g., boasting to your friends that your study is the greatest of all time and there is a statistically significant difference when in reality there is not)

Effect size (δ):

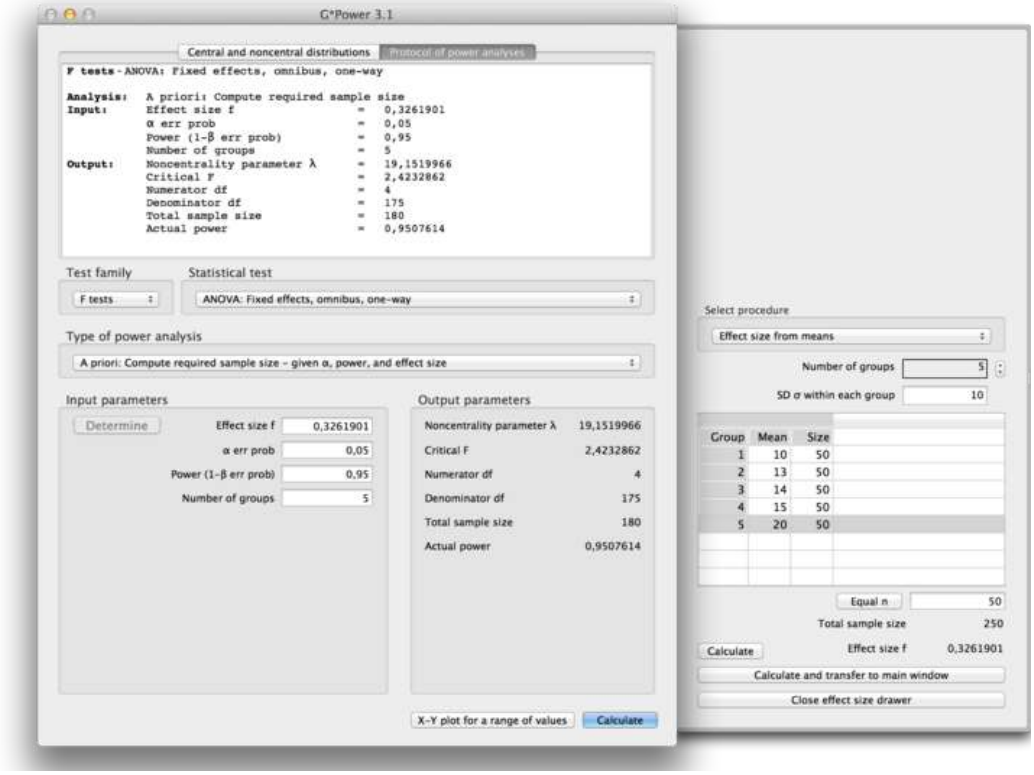
- Definition: Effect size quantifies the estimated difference between groups.
  - Tries to put numbers on *clinical significance* (e.g., estimated difference in means and standard deviations)
- Based "loosely" on prior literature and researchers prior experience
- The smaller the effect, the bigger the sample size needed to detect a difference
- Can consider standardized differences (e.g., Cohen's d) if little previous information is known
  - Small effect = 0.2; Medium effect = 0.5; Large effect = 0.8
- Resident pro-tip:
  - This tends to be the hardest part for resident pharmacists (and really all researchers) because there is no 100% right answer. You have to take a look at similar types of studies, other literature, and the experience of clinicians to make the best "educated guess" at what sort of difference you *expect* to find.

SAMPLE SIZE STEPS

Step 1: Set accepted level of type II error (β) or power (1- β)
Step 2: Set accepted level of type I error (α)
Step 3: Determine treatment groups, data type, and statistical test
Step 4: Estimate an effect size (δ)

G*POWER

G*Power is a free-to use software used to calculate statistical power. The program offers the ability to calculate power for a wide variety of statistical tests including t-tests, F-tests, and chi-square-tests, among others. Available for download with help manual: http://www.gpower.hhu.de/



EXAMPLES
The examples below are fictious and are meant to provide a framework for how to do a basic sample size calculation in two different trial designs. G*Power is used below, but may other websites/programs can be used as well (listed in "Other Free Power Calculation Resources" section at end of this guide).

Example 1
A PGY-1 resident rotating in the emergency department receives a patient with end stage renal disease (ESRD) on hemodialysis (HD) and mild hyperkalemia and wonders if giving a single dose of sodium polystyrene sulfonate (SPS, Kayexalate®) as monotherapy will lower the serum potassium sufficiently.

After doing a thorough literature review, the resident discovers that there is no published literature that specifically looks at single dose SPS in ESRD patients and they decide to do a research study. After careful review and discussions with research mentors they determine their aim and primary outcome.

- Study Aim: To evaluate the effects of single-dose SPS monotherapy for the treatment of acute hyperkalemia in patients with ESRD.
- Primary outcome: The primary end point of the study was decrease in serum potassium after a 30g oral SPS dose, measured as the difference between potassium before SPS administration and the first re-check following SPS administration.

They then set out to calculate a sample size.

Step 1: Set accepted level of type II error (β) or power ($1-\beta$)

- Could set power at 80%, but because of the minimal data in ESRD, the effect size will be estimated and therefore, resident decides to set at 90%

Step 2: Set accepted level of type I error (α)

- Typically set at $\alpha < 0.05$ and we have no known reasons to change it for this study

Step 3: Determine treatment groups, data type, and statistical test

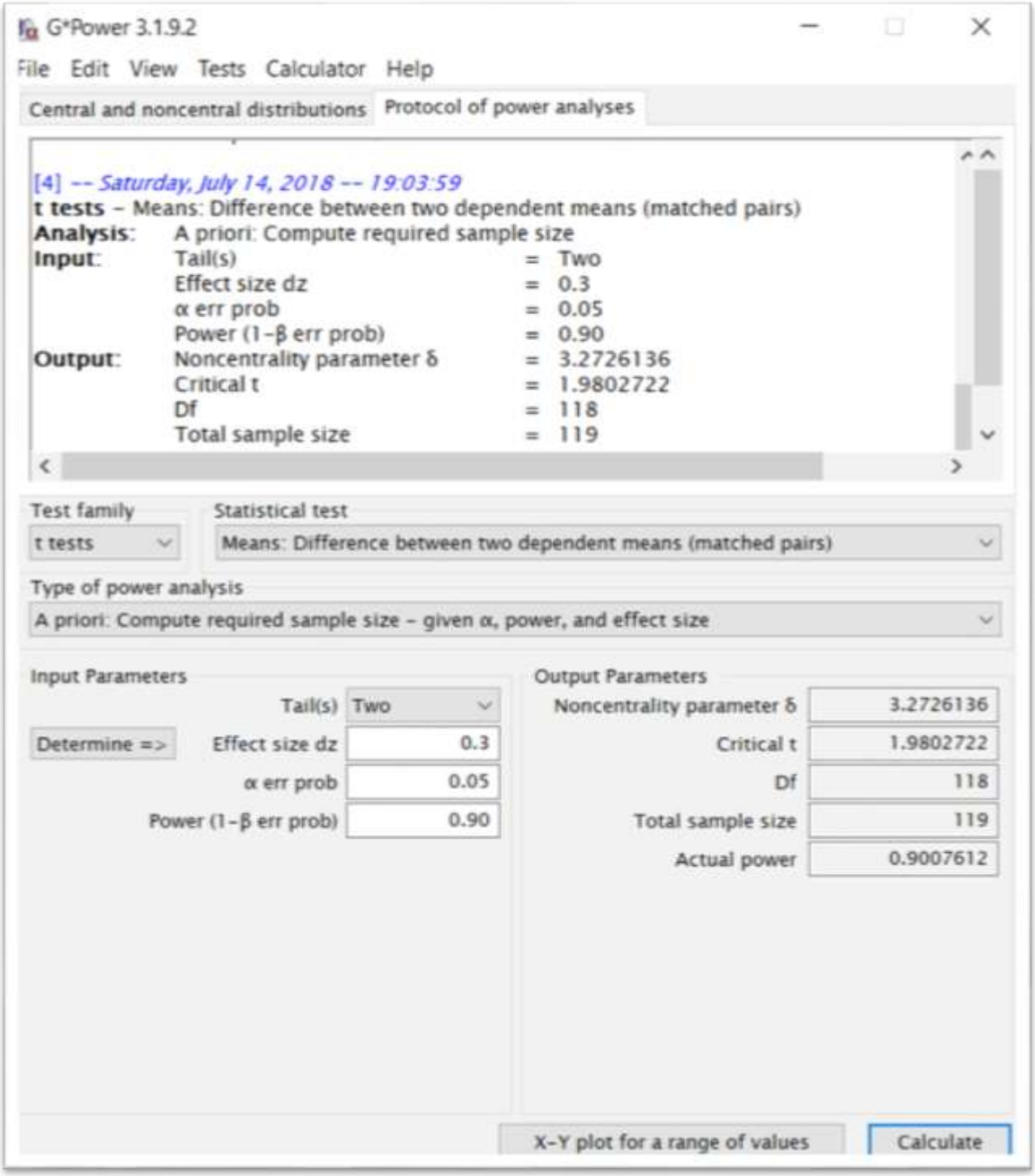
- Treatment groups: will be one group with two measures (pre- and post- potassium) and measures will be from the same patient making them *dependent* groups
- Data type: pre- and post- serum potassiums are continuous data (mean and standard deviation)
- Statistical test: As we are interested in doing one group, two dependent measures on continuous data, the resident decides to do a dependent samples t-test.

Step 4: Estimate an effect size (δ)

- The resident now looks at prior literature to determine an effect size. They find a couple retrospective studies on SPS, but they excluded patients with CKD IV, V, and ESRD and/or gave concomitant medications such as insulin + dextrose, albuterol, and sodium bicarbonate (which may confound prior results). In these studies the average decrease in potassium was 0.69 and 0.93 mEq/L.
- Because the patients in these studies had good renal function and natural potassium elimination plus other medications that help decrease potassium, the resident and research mentor decide that the expected potassium decrease in ESRD patients after only single dose-SPS will be much less. They estimate (or guess) that the observed potassium decrease will be on average only 0.15 and that they might see a small to medium effect size (Cohen's $d = 0.3$).

The resident now puts it all together, runs the numbers, and reports their calculation: "Using a t-test to detect a difference between two dependent means for the primary endpoint of potassium levels at

different time points with an effect size of 0.3, a 90% power with an alpha level of 0.05 would require a total sample size of 119 patients."



Example 2
After reviewing the pharmacokinetic changes in the critically ill while on a MICU rotation, a PGY-2 critical care resident notices that ceftriaxone is typically dosed once-daily, however, this dosing regimen may not be ideal in critically ill patients with altered pharmacokinetic properties such as hypoalbuminemia.

They notice when reading the literature that several studies have measured ceftriaxone levels in critically ill patients with hypoalbuminemia, but have not evaluated the clinical impact of the corresponding drug levels. The resident then decides to study split-dose ceftriaxone and clinical resolution (i.e., "clinical cure").

- Study Aim: The aim of this study was to determine if split-dosing of ceftriaxone leads to better clinical outcomes in critically ill, bacteremic patients with hypoalbuminemia.
- Primary Outcome: The primary outcome of this study was to assess clinical cure (resolution of signs of infection – normalization of temperature, WBC, serial negative blood cultures) rates at the end of therapy.

They then set out to calculate a sample size.

Step 1: Set accepted level of type II error (β) or power ($1- \beta$)

- Decide to set at the usual 80%

Step 2: Set accepted level of type I error (α)

- Typically set at $\alpha < 0.05$ and we have no known reasons to change it for this study

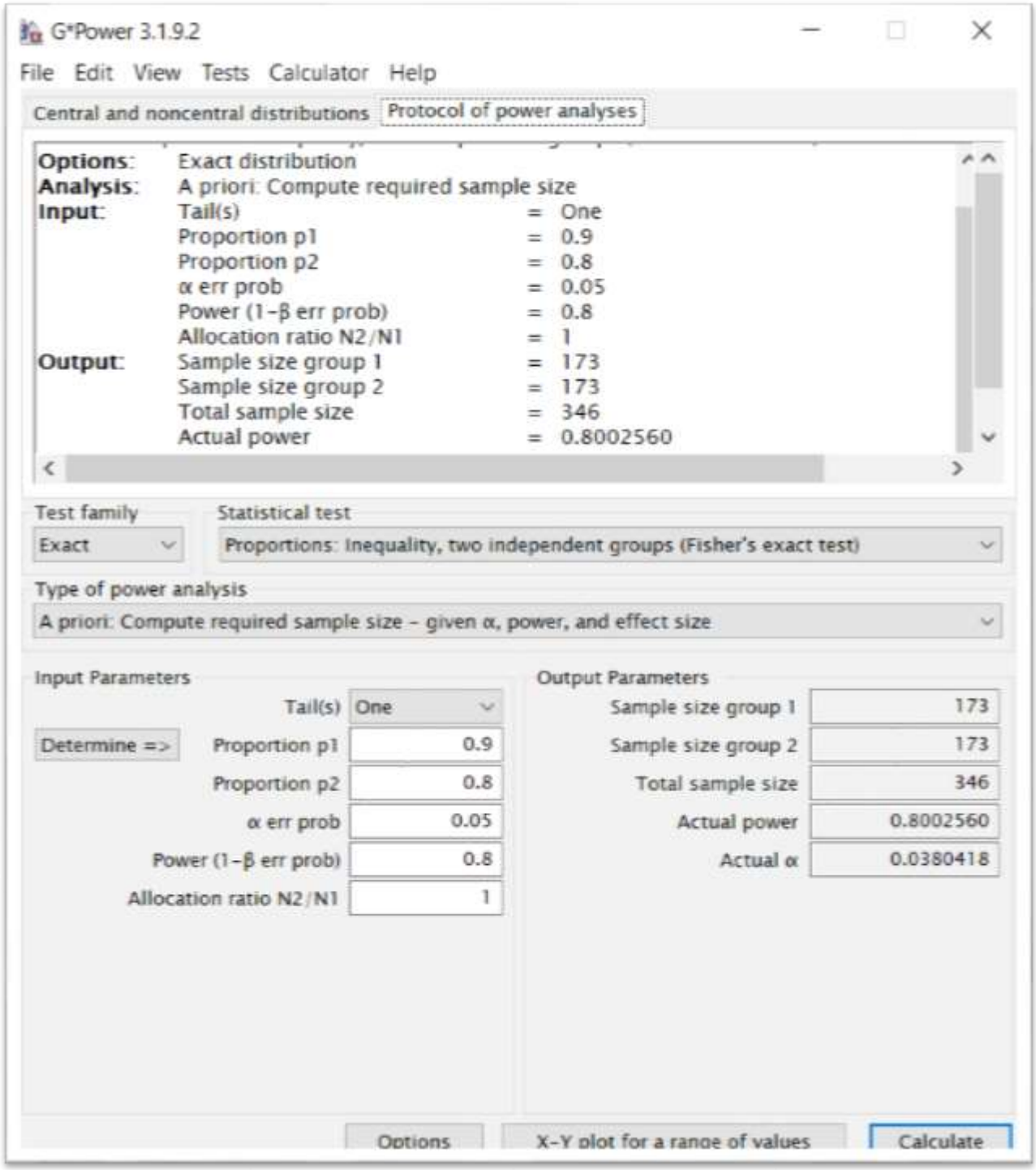
Step 3: Determine treatment groups, data type, and statistical test

- Treatment groups: Two study groups, with one group receiving treatment with the ceftriaxone dose split q12h and the other group receiving treatment with standard q24h dosing
- Data type: Clinical cure will be assessed as either 'yes' or 'no' = data is categorical (n, %)
- Statistical test: Comparing two proportions (n, % clinical cure yes/no) between 2 groups = Pearson chi-square test (or Fisher's exact test)

Step 4: Estimate an effect size (δ)

- The resident now looks at prior literature to determine an effect size. Upon review they don't find any data that specifically have addressed different ceftriaxone dosing and clinical outcomes. However, after looking at infectious disease research, they discover that most studies consider a therapy to be different if the cure rate is greater than 5-10%. The resident therefore estimates that patients will achieve a clinical cure of 90% in the q12h group and 80% in the q24h group.

The resident now puts it all together, runs the numbers, and reports their calculation: "A total of 346 patients (173 in each group) are needed to detect a 10% difference in the proportion of clinical cure between the q12h and q24h ceftriaxone groups using a Fisher's Exact test, α of 0.05, and power of 80%."



OTHER FREE POWER CALCULATION RESOURCES

- Web-based R application: https://designingexperiments.com/shiny-r-web-apps/
- Power and Sample Size Website: http://powerandsamplesize.com/Calculators/
- UCSF Clinical & Translational Science Institute: http://www.sample-size.net/sample-size-means/
- Vanderbilt PS program: http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize