Web-Based Sample Size Calculation

By: Sergey Tarima, PhD; Kwang Woo Ahn, PhD; Alexis Visotcky, MS

The Medical College of Wisconsin is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

The Medical College of Wisconsin designates this Live activity for a maximum of 1.0 *AMA PRA Category 1 Credit(s)* ™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Hours of Participation for Allied Health Professionals

The Medical College of Wisconsin designates this activity for up to 1.0 hours of participation for continuing education for allied health professionals.

Financial Disclosure

- In accordance with the ACCME® standard for Commercial Support Number 6, all in control of content disclosed any relevant financial relationships. The following in control of content had no relevant financial relationships to disclose.

Name:	Role in Meeting:
Ruta Brazauskas, PhD	Planning Committee
Haley Montsma, BBA	Planning Committee
Alexis Visotcky, MS	Speaker
Kwang Woo Ahn, PhD	Speaker
Sergey Tarima, PhD	Editor

Learning Objectives

- Familiarize the audience with basic power calculations, sample size concepts and terminology
- Introduce commonly used power calculation methodology
- Describe and illustrate the benefits of careful power and sample size analysis

Evaluation Forms

Your opinion matters!

Help us plan future meetings, by completing and submitting your evaluation forms.

Thank you.

The need for sample size

- Why do we need sample size calculation?
  - too small sample
  - too large sample
  - we need an optimal sample size
- Optimal sample size
  - there are many different optimality criteria for sample size calculations (the desired width of a confidence interval, or the desired power against some treatment effect)
- The optimality criteria for sample size calculation comes from the research question. In this presentation we consider two possibilities:
  - Continuous outcome; comparing means between 2 groups
  - Binary outcome; comparing proportions of between 2 groups

Ingredients of sample size

- The null hypothesis
- Alternative hypothesis
- Statistical method for testing the null
- Significance level
- Statistical power
- Sample size formula

The null hypothesis

- The statistical null hypothesis is formulated on the basis on the research question.

- Consider the research question "Is the weight management therapy A more effective than the well established weight management therapy B"?

- This research question leads to the following null hypothesis "H_0: $W_A = W_B$", where W_A is the group mean weight change after 12 months of therapy A and W_B is the group mean weight change after 12 months of therapy B. Here we have a clear outcome W (the weight change in a 12 month period).

The alternative hypothesis

- Obviously, we expect that the new therapy A will be more effective as the previously available therapy B.
- The question remains: What does it mean “More effective”? We needed to define a minimal clinically relevant improvement in weight change, suppose d.
- Then, we formulate the alternative hypothesis H_A: W_A-W_B=d.
- If for example, we want to see an improvement of 5 lbs, then H_A: W_A-W_B=5.

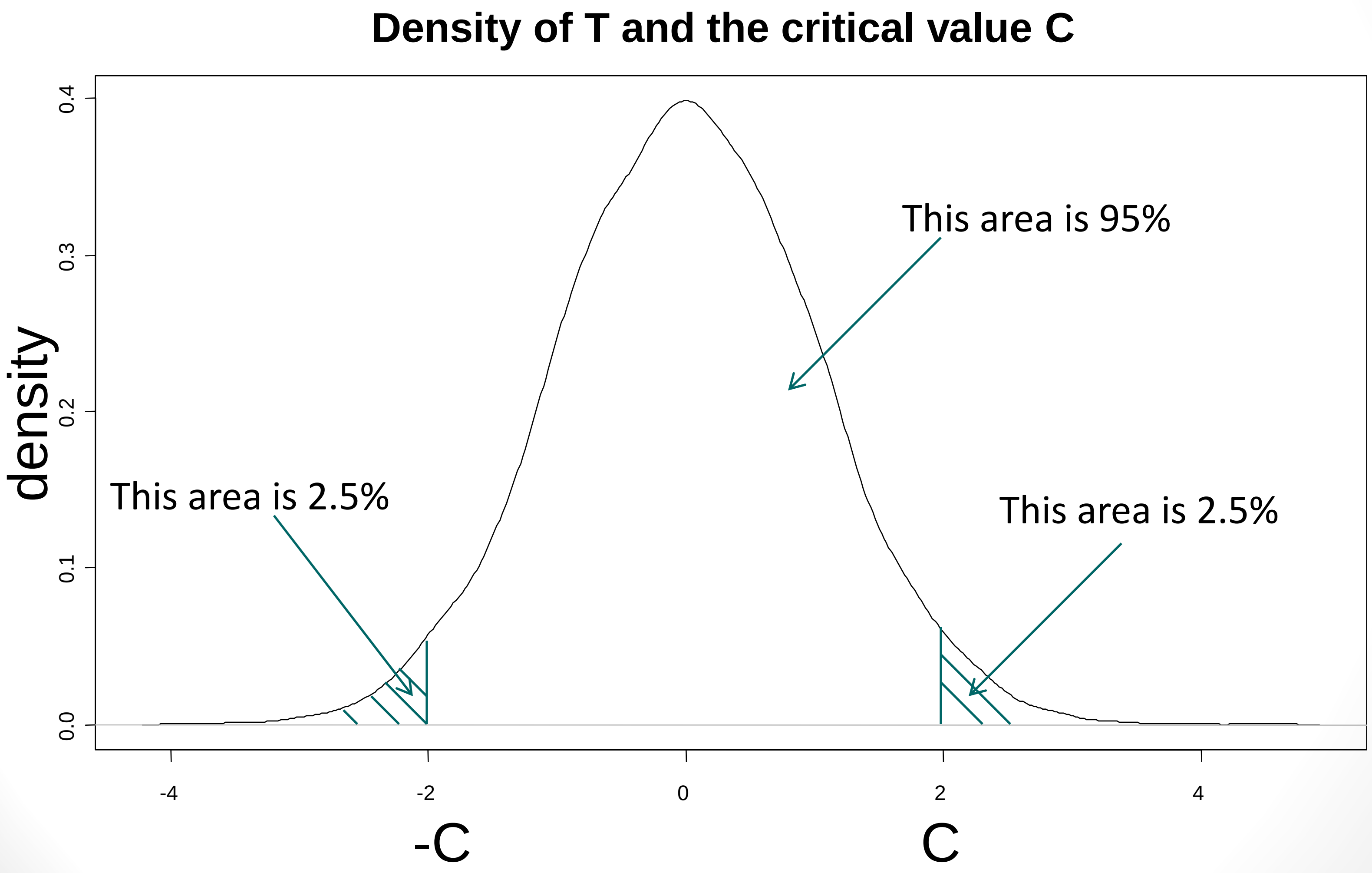
Statistical method for testing the null

- What method is to be used for comparing mean group changes? The two sample t-test is a popular solution.
- The two sample t-test tests the *simple* null hypothesis H_0: $W_A = W_B$ (no difference between therapies)
- The two sample t-test calculates the test statistic T and the test is performed with comparing this T against a critical value C.
- If $T > C$ or $T < -C$ the null hypothesis H_0 is rejected in favor of the alternative H_A , otherwise we “fail to reject the null”
- Where does this C comes from? Significance level!

Significance level

- Significance level defines the probability of type I error. The error that “the null hypothesis is rejected when it actually is true”
- Significance level is usually = 5%, which for the two sample t-test defines a critical value of approximately 2. More accurate values are tabulated in books or can be calculated using a statistical software.
- Significance level defines a *critical region* (the values of T when H_0 is rejected). In our case T< -C and T> C is the critical region.

Significance level



Statistical Power

- Statistical power is the probability to reject H_0 if the alternative H_A is true. Power is usually defined against a simple alternative hypothesis.
- For example, if H_A: W_A-W_B = d then the probability that we will be able to claim that the groups are different is the POWER
- This d is called *the effect size.*
- In sample size calculations POWER is traditionally = 80%

Sample size formula

- For the two-sample t-test the formula for sample size is

$$N = 4\sigma^2 \frac{7.85}{d^2}$$

- In this formula we assumed that the significance level = 5%, POWER=80%, d is selected a minimally clinically relevant effect size… the problem is in σ.
- The σ is usually unknown and is estimated from a pilot data, or extracted from literature review, or estimate from a previously completed study with similar outcomes.

Example: Weight management (1)

- Suppose we are designing a study to compare a new weight management therapy A with a previously established therapy B.
- Previous therapy is known to reduce weight (on average) by 5 lbs after 3 months on this therapy. From a previous study we know that the standard deviation of this reduction is 5 lbs ($\sigma=5$) .
- We expect that our new therapy will decrease the weight by 10 lbs in 3 month. Then, d=10 – 5 = 5 lbs.
- Then the null hypothesis is H_0: $W_A = W_B$
- Then the alternative hypothesis is H_A: $W_A - W_B = 5$

Example: Weight management (2)

- The use of our sample size formula leads to

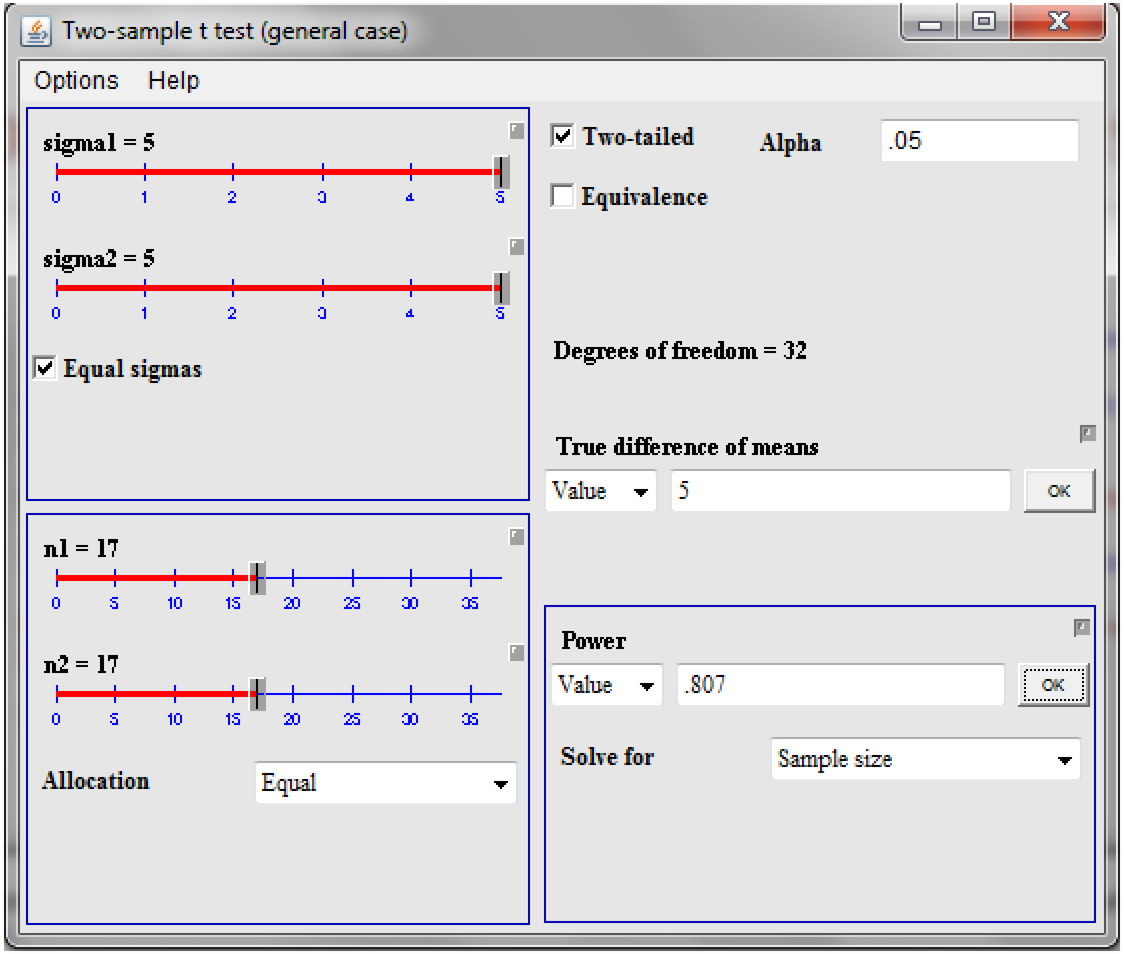
$$N = 4\sigma^2 \frac{7.85}{d^2} = 4 \cdot 5^2 \frac{7.85}{5^2} = 31.4$$

- This means that we will need to enroll 32 subjects into the study and randomize them: 16 to one group and 16 to another group. This sample size secures approximately 80% power at 5% significance.
- The above formula is only a rough approximation of how many subjects we will need. Statistical software: standalone or WEB-based would give a more accurate sample size estimate.

Example – Two sample mean

- Consider http://www.stat.uiowa.edu/~rlenth/Power/
- Assume that the significance level is 0.05.
- Assume that we want to compare two independent group means.
  - The null hypothesis: $\mu 1 = \mu 2$
  - The alternative hypothesis: $\mu 1 \neq \mu 2$
- We want to detect at least 5 mean difference with SD 5 with 80% of statistical power in the weight management example.

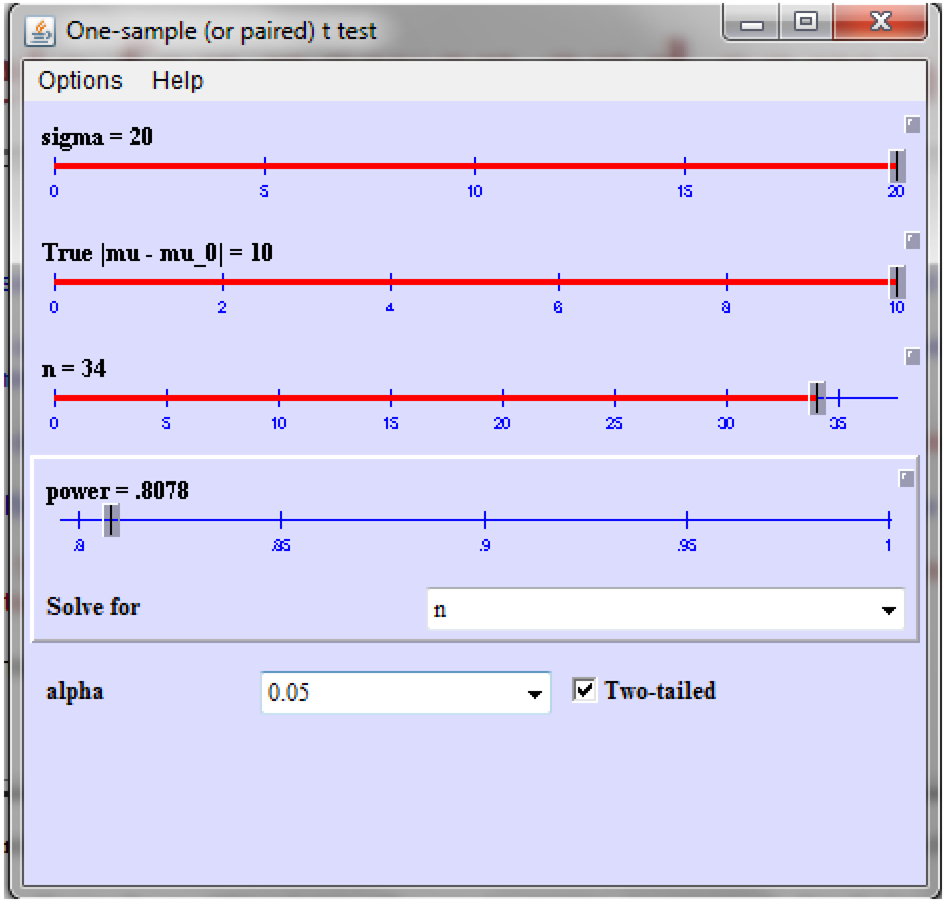
Example - continued



Example – One sample mean

- Assume that we want to test whether the group mean μ is equal to μ_0.
  - The null hypothesis: $\mu = \mu_0$
  - The alternative hypothesis: $\mu \neq \mu_0$
- Suppose that we want to test to see whether a new weight management therapy is effective in reducing weight.
- Suppose that the null hypothesis is that $\mu=130$.
- We want to detect at least 10 difference between μ and 130 with SD 20 with 80% of statistical power.

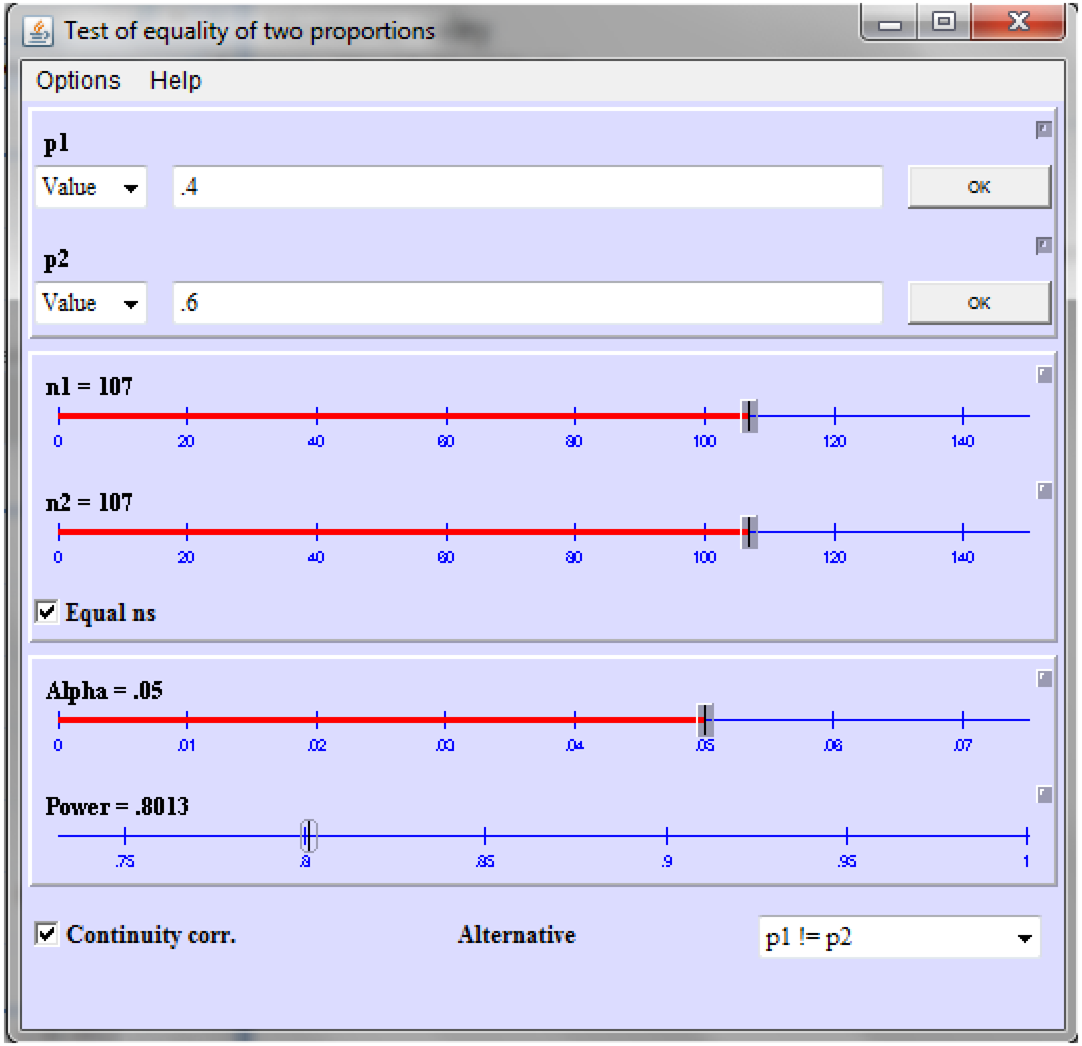
Example - continued



Example – Two proportions

- Assume that we want to compare two independent group proportions.
  - The null hypothesis: $p1 = p2$
  - The alternative hypothesis: $p1 \neq p2$
- We want to detect difference when $p1=0.4$ and $p2=0.6$ with 80% of statistical power.

Example - continued



Recommendation

- The Division of Biostatistics provides free consulting service including study design, sample size and power calculation, analysis, etc for MCW, Marquette, and UWM faculty and staff.
- Consulting service applications can be submitted here:
  - http://www.mcw.edu/biostatsconsult/ConsultingApplication/Application.htm

Thank You!!!

- Questions???