How Many Subjects Do I Need? A Crash Course in Sample Size Calculations
November 19, 2024 – FINAL version

Scenario 1: Single Proportion

A primary care provider working in long-term care wants to estimate the prevalence of a recent (in the past year) major depressive episode in adults aged 65+ years. She will take a random sample and, after obtaining informed consent (assuming cognitive capacity), the subjects will be evaluated.

This investigator will report the prevalence along with a 95% confidence interval (CI). The prevalence is a proportion. She does not want to compare groups, for example, the prevalence in men compared to the prevalence in women, so this scenario involves a single proportion. Based on her review of the literature, the primary care provider believes the prevalence of a recent major depressive episode is 10% in her long-term care home. She would like a margin of error of 3 percentage points. The margin of error is half the width of the CI. The margin of error is also known as the absolute precision.

Using OpenEpi: Point your browser to www.OpenEpi.com and please follow these steps:

1. On the left-side of the screen click on **Sample Size**.
2. Under **Sample Size**, click on **Proportion**.
3. Click on the **Enter New Data** button.
4. Enter the value of 10 in the box labeled **Anticipated % frequency(p)**.
5. Enter 3 in the box that is to the left of the label **Absolute precision %**.
6. Then click on **Calculate** at the top of the screen.

What is the required sample size given these scenario elements for a 95% confidence level?

What is the required sample size given these scenario elements for a 99% confidence level?

In step 5 above we specified a half width of 3 percentage points. Now click on the **Enter** tab. Change the absolute precision to 5 percentage points. Click on Calculate. What is the required sample size for a 95% confidence level given this new half width and assuming none of the other scenario elements have changed?

Scenario 2: Comparison of Two Means

An interdisciplinary team consisting of family physicians and dieticians will randomize a group of breast cancer patients to one of two groups using a 1:1 design: The intervention group or the control group. The intervention group will receive a whole-food, plant-based diet while the control group will continue with their usual diet.

The outcome is dietary cholesterol intake (a continuous variable) 8 weeks after randomization. A two-tailed Student's two-sample *t*-test will be performed to test the null hypothesis that the population means are equal. For the sample size calculation assume the following: Alpha = 0.05, Power = 80%, and an equal number of study subjects in both arms of the study.

This team of investigators believes that the mean dietary cholesterol intake will be 180 mg/day with a standard deviation (SD) of 40 mg/day at the end of the observation period in the control group. The anticipated mean dietary cholesterol intake and SD in the intervention group are 10 mg/day and 4 mg/day, respectively, at the end of the observation period.

Using OpenEpi: Point your browser to www.OpenEpi.com and please follow these steps:

1. On the left-side of the screen click on **Sample Size**.
2. Under **Sample Size**, click on **Mean Difference**.
3. Click on the **Enter New Data** button.
4. Enter the value of 180 for the **Mean** of **Group 1**.
5. Enter 40 for the **Std. Dev.** for **Group 1**.
6. Enter the value of 10 for the **Mean** of **Group 2**.
7. Enter 4 for the **Std. Dev.** for **Group 2**.
8. Then click on **Calculate** at the top of the screen.

What is the total required sample size given these scenario elements?

Now click on the **Enter** tab. Change the **Mean** of **Group 2** to 100 mg/day. Then click Calculate. What is the new total required sample size?

Scenario 3: Comparison of Two Proportions

A team of emergency medicine physicians and surgeons is interested in the risk of 30-day mortality in trauma patients who are at risk of hemorrhage. The study team is planning to conduct a retrospective cohort study using existing medical records. The exposed group will be a sample of injured patients who received tranexamic acid (TXA) in the prehospital setting within one hour of the time of the injury (the "early" group). The unexposed group (the comparison group) will be a sample of injured patients who received TXA in the prehospital setting more than one hour after the time of the injury (the "delayed" group).

The outcome is death within 30 days of the injury. This is a binary (also known as dichotomous) variable: Either the patient died or did not die within 30 days of the injury. The statistical analysis will involve the comparison of two proportions. A chi-square test will be performed.

The anticipated incidence proportion (the risk) of 30-day mortality is 4% in the early group while the anticipated incidence proportion of 30-day mortality is 9% in the delayed group. The research team believes that there will be twice as many patients in the delayed group than in the early group. Therefore, the expected ratio of the number of unexposed to exposed patients is 2. Here are the final additional assumptions: Two-sided testing, an alpha of 0.05, and a power of 80% (so beta is 0.20).

Using OpenEpi: Point your browser to www.OpenEpi.com and please follow these steps:

1. On the left-side of the screen click on **Sample Size**.
2. Under **Sample Size**, click on **Cohort/RCT**.
3. Click on the **Enter New Data** button.
4. Enter the value of 2 for the **Ratio of Unexposed to Exposed in sample**.
5. Enter 9 for the **Percent of Unexposed with Outcome**.
6. Enter 4 for the **Percent of Exposed with Outcome**.
7. Then click on **Calculate** at the top of the screen.

What is the required sample size given these scenario elements using the method of Fleiss?

Number of patients in the early group (the exposed group) ☐

Number of patients in the delayed group (the unexposed group) ☐

Scenario 4: Least Detectable Odds Ratio from an Unmatched Case-Control Study

An obstetrician and a maternal health epidemiologist want to study preeclampsia, a serious hypertensive disorder of pregnancy. They will conduct a case-control study of the association between periodontal disease and the outcome of preeclampsia. An odds ratio (OR) for preeclampsia comparing women with periodontal disease with women who do not have periodontal disease will be calculated. To clarify, the exposure variable is periodontal disease (present vs. absent). Cases will be women with preeclampsia while controls will be women free of preeclampsia.

Here are the elements for this sample size scenario:

- Alpha = 0.05.
- 80% power.
- Two-tailed testing.
- Prevalence of periodontal disease in controls will be 6%.
- Equal number of cases and controls (so the ratio of controls to cases is 1).
- No matching.
- Least extreme OR to be detected: 1.80.

Using OpenEpi: Point your browser to www.OpenEpi.com and please follow these steps:

1. On the left-side of the screen click on **Sample Size**.
2. Under **Sample Size**, click on **Unmatched CC**.
3. Click on the **Enter New Data** button.
4. Enter the value of 6 for the **Percent of controls exposed**.
5. Enter 1.8 for the **Odds ratio.**
6. Then click on **Calculate** at the top of the screen.

What is the required total sample size given these scenario elements using the method of Fleiss?

Now click on the **Enter** tab and change the **Odds ratio to 1.20**. Now click on Calculate. What is the required sample size given these scenario elements using the method of Fleiss?

Please contact Zuber D. Mulla, Ph.D., at zuber.mulla@ttuhsc.edu with any comments or questions.