

Sample Size Determination

The Fundamentals of International
Clinical Research Workshop

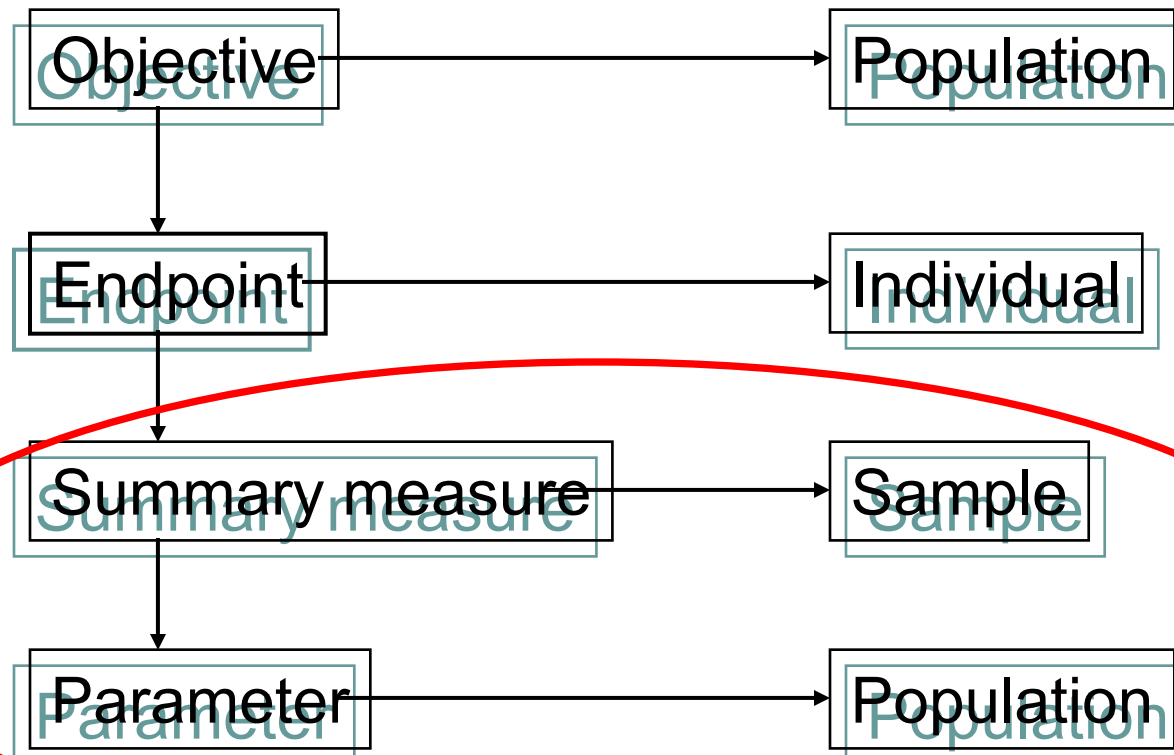
Mario Chen

DMID/ICSSC

Framework

- Objectives & Endpoints
 - Type of Inference:
 - **Estimation**
 - **Hypothesis Testing**
- Study Design:
 - Type of Study
 - **Sample Size**
- Data Collection & Data Management
- **Data Analysis**

Selection of summary measure



Sample Size for Estimation

Necessary Components:

1. Summary measure of interest
 - Proportions or means
2. Desired Confidence Level ($1-\alpha$)
3. Desired Precision Level (d)
4. Expected Variability in the study population:
 - For means (σ)
 - For proportions (P)

Sample Size for Estimation (Proportions)

Example:

1. Summary measure: Prevalence of Vibrio Cholerae
2. Confidence Level: 95% n = 523
3. Precision Level:
d = 3 percent points, i.e. 0.03
4. Variability: Estimate of P = 15% (from previous studies)

Recommendation: If there is not a good estimate for P, use P closest to 50% to be conservative

Sample Size for Estimation (Means)

Example:

1. Summary measure: Mean hydrocele size among filariasis infected patients
2. Confidence Level: 95%
3. Precision Level: $d = 1.5$ mm.
4. Variability: Estimate for $\sigma = 5.7$ mm.

$$n = 58$$

To estimate σ we can use previous studies, expert advice, or a pilot study

Sample Size for Hypothesis Testing

Necessary Components:

1. Summary measure of interest

- proportions or means

2. Statistical Hypotheses

- Null hypothesis (H_0) – the hypothesis to be tested, often includes hypothesis of no difference
 - H_0 : Avg. BP in group A \geq Avg. BP in group B
- Alternative hypothesis (H_A) – corresponds to the research hypothesis
 - H_A : Avg. BP in group A $<$ Avg. BP in group B
- H_0 and H_A - mutually exclusive and exhaustive

Sample Size for Hypothesis Testing

Necessary Components (cont'd):

3. Significance Level (α)

4. Desired Power ($1-\beta$)

Sample Size for Hypothesis Testing

- Consequences of Decision Making

Study Decision	Null Hypothesis	
	True	False
Do not Reject Null Hypothesis	Correct Decision (1- α)	Type II error (β)
Reject Null Hypothesis	Type I error (α)	Correct Decision (1- β)

α = Significance Level

(1- α) = Confidence Level

(1- β) = Power

Sample Size for Hypothesis Testing

Necessary Components (cont'd):

5. Effect Size: Smallest difference worth detecting (clinically)

- If the difference is greater then changes will occur
- If the difference is smaller, groups are considered “equivalent” (status quo preserved)

6. Variability expected in the population

- For means (σ_1 , σ_2)
- For proportions (P1, P2)

Sample Size for Hypothesis Testing

Example for proportions:

1. Summary measure: Proportion Cured

2. Statistical Hypotheses:

H_0 : Proportion cured is the same with both the new treatment and the standard treatment

H_1 : Proportion cured is higher with the new treatment than with the standard treatment

One sided vs. two sided Alternative Hypothesis.

Sample Size for Hypothesis Testing

(cont-d)

3. Significance Level: $\alpha = 5\%$

4. Power: $1 - \beta = 90\%$
(never lower than 80%)

5. Effect size: $P_2 - P_1 = 15\%$

$n = 109$

6. Variability:

Estimate for $P_1 = 75\%$

Estimate for $P_2 = 90\%$

Recommendation:

a) Define level for P for the “baseline” group

b) Use effect size to obtain level of P for the “study” group

Main Determinant of Study Size



Recommendations when budget is not enough:

1. (Estimation) Lower desired precision.
2. (Hypothesis Testing) Lower desired power or increase minimum detectable effect size.
3. It is not recommended to change confidence levels, significance levels, or variance estimates.
4. If after all these changes, budget is still insufficient, one has to decide between:
 - Not conducting the study until enough budget has been obtained, or
 - go ahead with the study knowing that the results are likely to be inconclusive (pilot study or exploratory).



Adjustments to Sample Size

- Non-Response (and attrition):

$$n_2 = n_1 / (1 - NR)$$

n_2 = final size, n_1 = effective size

NR = Non-response (and attrition) rate

Other Considerations

- **Data dependencies** (e.g., Matching, Repeated Measures, Clustering)
- **Multivariate methods** (e.g., control for confounding)
- **Multiplicity issues** (e.g., Multiple testing, endpoints, treatments, interim analyses)
- **Other endpoints** (e.g., time to event)
- **Other hypotheses** (e.g., equivalence)

Software

- **PASS**
(www.ncss.com/pass.html)
- **nQUERY**
(www.statsolusa.com/nquery/nquery.htm)
- **EPI-INFO**
(www.cdc.gov/epiinfo)
- **EPIDAT**
(www.paho.org/English/SHA/epidat.htm)

A Final note on Interpretation

- If your final analysis comes out not significant (i.e., $p > 0.05$)
 - Correct decision. H_0 is true!
 - Bad luck. Type II error
 - Lack of power:
 - Effect is smaller than effect size used in SS
 - Incorrect assumptions (e.g., σ)
 - Don't do post-hoc power calculations
- Correct interpretation:
 - Do not conclude Null Hypothesis is true
 - Insufficient evidence to conclude the Null Hypothesis is not true

**Sample Size Determination:
Foibles and Nuances of Estimating
Sample Sizes:**

Falgone, Case Scenario 5

DMID/ICSSC

Dar es Salaam, Tanzania

January 17-22, 2010

Kenneth F. Schulz, PhD

α -error (Type I error, p-value)?

0.05

β -error (Type II-error)?

0.10

Power = 1 - [β -error] = 1 - 0.10 = 0.90

What is a clinically important difference to detect?

- Standard drug has 30% treatment failure
- And falgone . . .

□ 29%? . . . 28%? . . . 27%? . . . 26%?

23%? . . . 22%? . . . 21%? . . . 20%?

Falgone incidence?

- “Ridiculous notion” . . . “Ludicrous concept” . . .
- “How would I know the incidence – I’m doing the study to estimate the incidence.”
- “I have no data on which to base an estimate.”

-----True, but-----

- You do not estimate the incidence!
- You determine the incidence reflecting the clinically important difference to detect, which may have no relation to the actual incidence.

Falgone vs. standard for treatment of moderately severe malaria

- **Standard drug**
 - ❑ **Cost of \$30.00 for 28 days**
 - ❑ **2 oral doses per day**
 - ❑ **Minimal side effects**
- **Falgone**
 - ❑ **Cost of \$500.00 for 28 days**
 - ❑ **A single oral dose per day**
 - ❑ **Neutropenia, blurred vision, and proteinuria**

What is a clinically important difference to detect?

- Standard drug has 30% treatment failure
- And falgone . . .

□ 29%? . . . 28%? . . . 27%? . . . 26%?

23%? . . . 22%? . . . 21%? . . . 20%?

17%? . . . 16%? . . . 15% ?

We believe **15%** represents a clinically important difference to detect.

Remember

- **Basically a clinical/content decision**
 - **Not a statistician's decision**
- **Set Type I error and Type II error (power)**
- **Feed into sample size software**

Sample Size Assumptions

α -error = 0.05

β -error = 0.10 (Power = 0.90)

TF incidence with the standard drug = 30%

TF incidence with falgone = 15 %