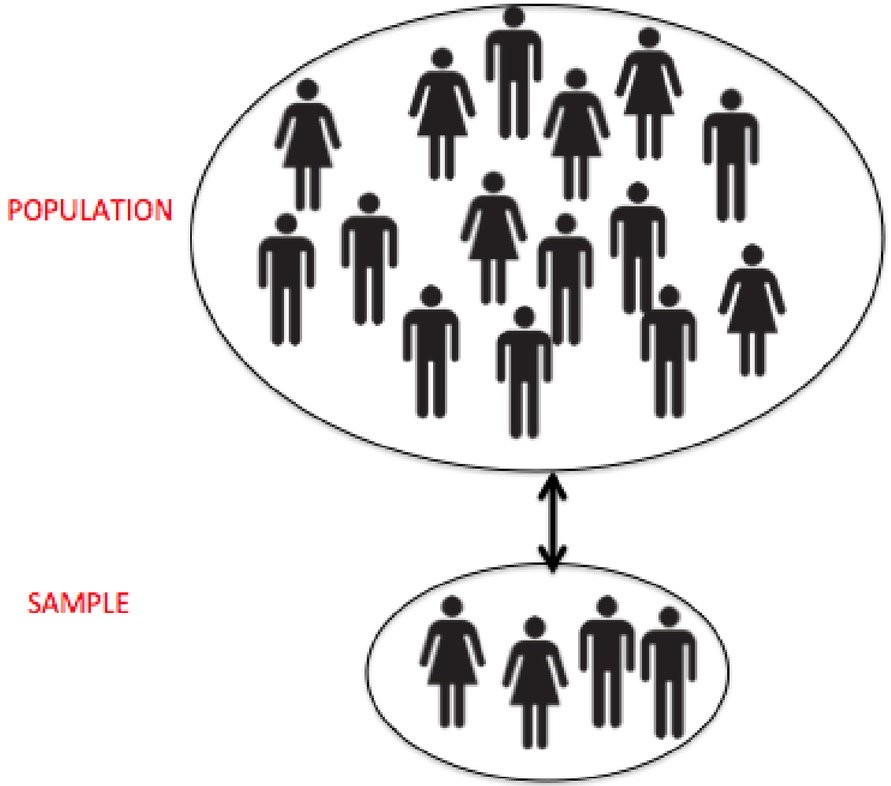
DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Introduction to Sample Size and Power

Tamuno Alfred, PhD
Biostatistician

Statistical inference



Importance of correct sample size

- Costs
- Study completion time
- Exposure to experimental drug
- Patients receiving no treatment
- Ability to reject null hypothesis

Requirements

Requirements

- Trial purpose (compare weight loss between drug and placebo)

Requirements

- Trial purpose (compare weight loss between drug and placebo)
- Primary endpoint (difference in mean % weight loss)

Requirements

- Trial purpose (compare weight loss between drug and placebo)
- Primary endpoint (difference in mean % weight loss)
- Statistical analysis of primary endpoint (two-sample t-test)

Requirements

- Trial purpose (compare weight loss between drug and placebo)
- Primary endpoint (difference in mean % weight loss)
- Statistical analysis of primary endpoint (two-sample t-test)
- Smallest meaningful difference, δ (3)

Requirements

- Trial purpose (compare weight loss between drug and placebo)
- Primary endpoint (difference in mean % weight loss)
- Statistical analysis of primary endpoint (two-sample t-test)
- Smallest meaningful difference, δ (3)
- Variability (standard deviation=10)

Requirements

- Trial purpose (compare weight loss between drug and placebo)
- Primary endpoint (difference in mean % weight loss)
- Statistical analysis of primary endpoint (two-sample t-test)
- Smallest meaningful difference, δ (3)
- Variability (standard deviation=10)
- Significance level, α (0.05)

Requirements

- Trial purpose (compare weight loss between drug and placebo)
- Primary endpoint (difference in mean % weight loss)
- Statistical analysis of primary endpoint (two-sample t-test)
- Smallest meaningful difference, δ (3)
- Variability (standard deviation=10)
- Significance level, α (0.05)
- Power to detect treatment effect (80%)

Hypothesis testing

H0: μ1 = μ2, i.e. no treatment difference

Type I error, α (False positive): Falsely reject H_0

Type II error, β (False negative): Do not reject H_0 when H_0 is false

Power: Probability of correctly rejecting H_0

Power=1-Prob(Type II error)

	H_0 True	H_0 False
Reject H_0	Type I error, α	True positive
Do not reject H_0	True negative	Type II error, β

Two-sample t-test

```
power.t.test(delta=3, sd=10, power=0.8,
        type = "two.sample",  alternative = "two.sided")
```

```
     n = 175.3851
 delta = 3
    sd = 10
sig.level = 0.05
 power = 0.8
alternative = two.sided

NOTE: n is number in *each* group
```

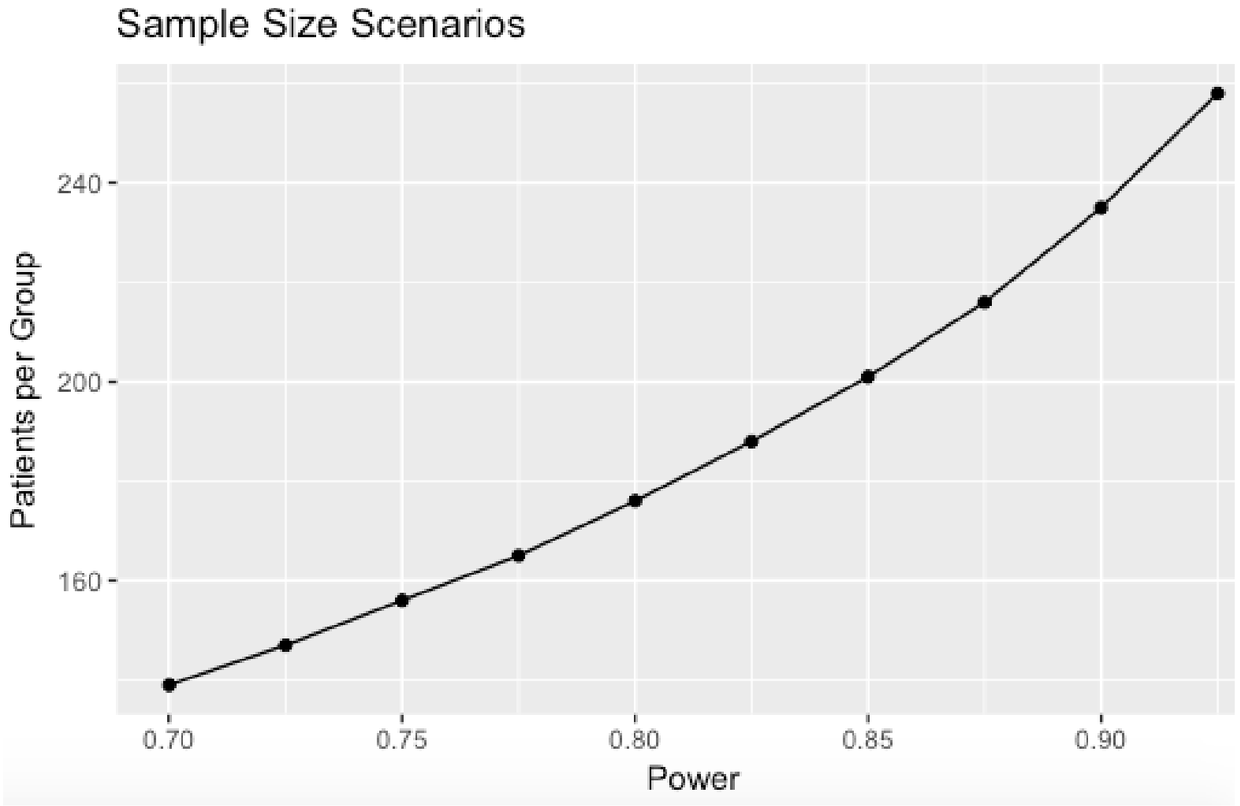
Relationship between power and sample size

```
power.t.test(delta=3, sd=10, power=0.9,
        type = "two.sample",  alternative = "two.sided")
```

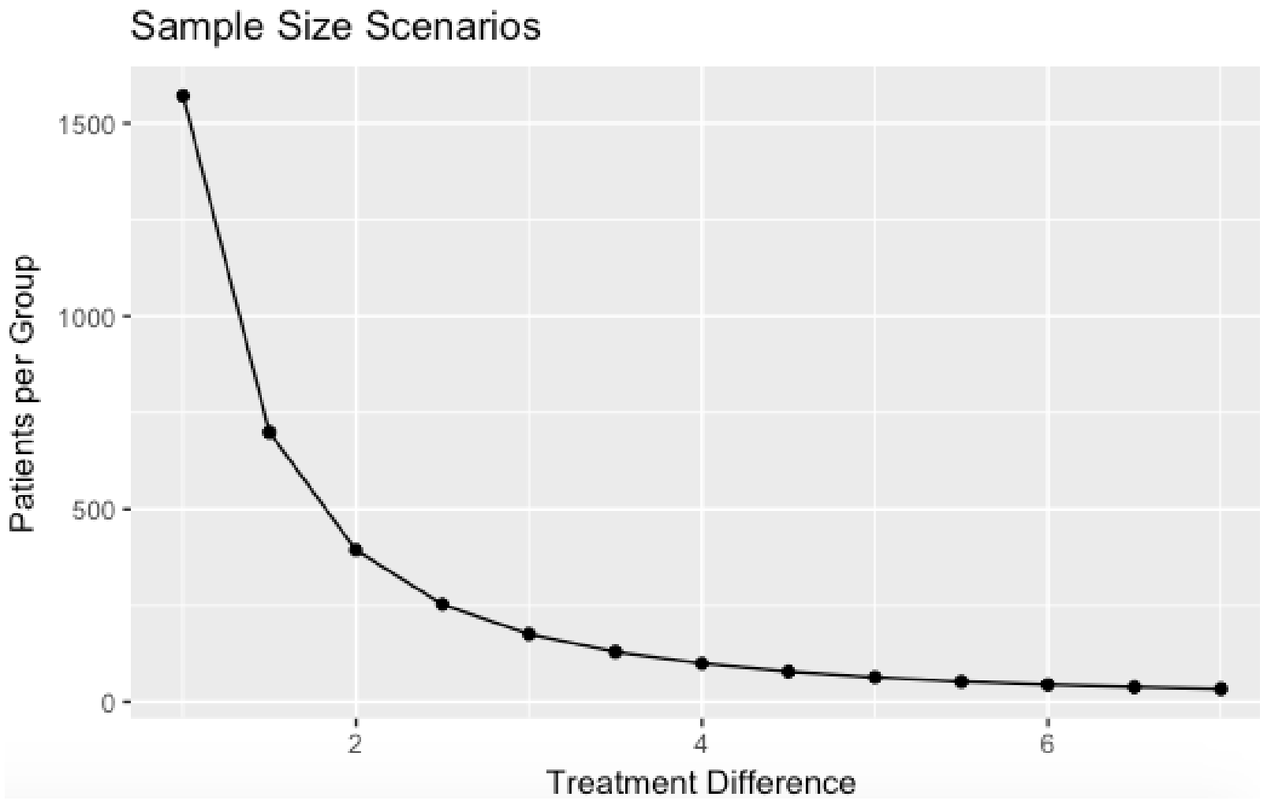
```
     n = 234.4628
 delta = 3
    sd = 10
 sig.level = 0.05
 power = 0.9
 alternative = two.sided

NOTE: n is number in *each* group
```

Relationship between power and sample size



Relationship between treatment difference and sample size



Test of proportions

```
power.prop.test(p1=0.3, p2=0.15, power=0.8)
```

```
     Two-sample comparison of proportions power calculation

              n = 120.4719
             p1 = 0.3
             p2 = 0.15
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Let's practice!

DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Sample Size Adjustments

Tamuno Alfred, PhD
Biostatistician

One-sided tests

Null Hypothesis:

$H_0: \mu_1 = \mu_2$

Alternative Hypothesis (two-sided):

$H_A: \mu_1 \neq \mu_2$

Alternative Hypothesis (one-sided):

$H_A: \mu_1 > \mu_2$

One-sided tests

```
power.t.test(delta=3, sd=10, power=0.8,
             type = "two.sample",  alternative = "two.sided")
```

```
     n = 175.3851
 delta = 3
    sd = 10
 sig.level = 0.05
 power = 0.8
 alternative = two.sided

NOTE: n is number in *each* group
```

One-sided tests

```
power.t.test(delta=3, sd=10, power=0.8,
             type = "two.sample",  alternative = "one.sided")
```

```
     n = 138.0715
 delta = 3
    sd = 10
 sig.level = 0.05
 power = 0.8
 alternative = one.sided

NOTE: n is number in *each* group
```

Unequal group sizes

Unequal group sizes

```
library(samplesize)
n.ttest(power = 0.8, alpha = 0.05, mean.diff = 3, sd1 = 10, sd2 = 10,
            k = 0.5, design = "unpaired", fraction = "unbalanced")
```

Unequal group sizes

```
library(samplesize)
n.ttest(power = 0.8, alpha = 0.05, mean.diff = 3, sd1 = 10, sd2 = 10,
            k = 0.5, design = "unpaired", fraction = "unbalanced")
```

```
$`Total sample size`
[1] 396

$`Sample size group 1`
[1] 264

$`Sample size group 2`
[1] 132

$Fraction
[1] 0.5
```

Unequal variances

$$SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

Unequal variances

$$SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

```
n.ttest(power = 0.8, alpha = 0.05, mean.diff = 3, sd1 = 9.06, sd2 = 9.06,
        k = 1, design = "unpaired", fraction = "balanced")
```

Unequal variances

$$SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

```
n.ttest(power = 0.8, alpha = 0.05, mean.diff = 3, sd1 = 9.06, sd2 = 9.06,
        k = 1, design = "unpaired", fraction = "balanced")
```

```
$`Total sample size`
[1] 290

$`Sample size group 1`
[1] 145

$`Sample size group 2`
[1] 145
```

Loss to follow-up

Q: anticipated dropout rate

Multiply original sample size by

$$\frac{1}{1-Q}$$

Loss to follow-up

Q: anticipated dropout rate

Multiply original sample size by

$$\frac{1}{1-Q}$$

```
orig.n <- power.t.test(delta=3, sd=10, power=0.8,
             type = "two.sample",  alternative = "one.sided")$n
orig.n
ceiling(orig.n/(1-0.1))
```

DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Let's practice!

DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Interim Analyses and Stopping Rules

Tamuno Alfred, PhD
Biostatistician

Motivation

- Patients recruited over time

Motivation

- Patients recruited over time
- Data accumulated gradually

Motivation

- Patients recruited over time
- Data accumulated gradually
- Safety and efficacy can be monitored regularly

Motivation

- Patients recruited over time
- Data accumulated gradually
- Safety and efficacy can be monitored regularly
- Investigators must safeguard patients' interests

When to stop a trial early

- Efficacy

When to stop a trial early

- Efficacy
- Safety

When to stop a trial early

- Efficacy
- Safety
- Futility

When to stop a trial early

- Efficacy
- Safety
- Futility
- Other
  - Cost
  - Inability to recruit enough patients
  - Poor trial design

Stopping rules

- Interim analyses often require

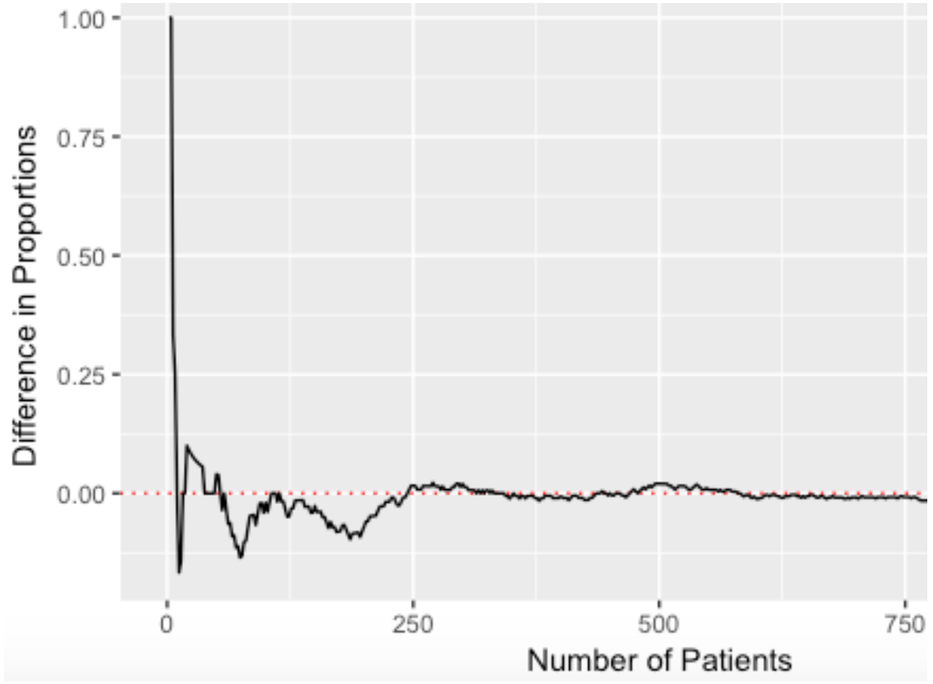
  increased sample size

Stopping rules

- Interim analyses often require increased sample size
- Multiple testing increases chance of Type I error

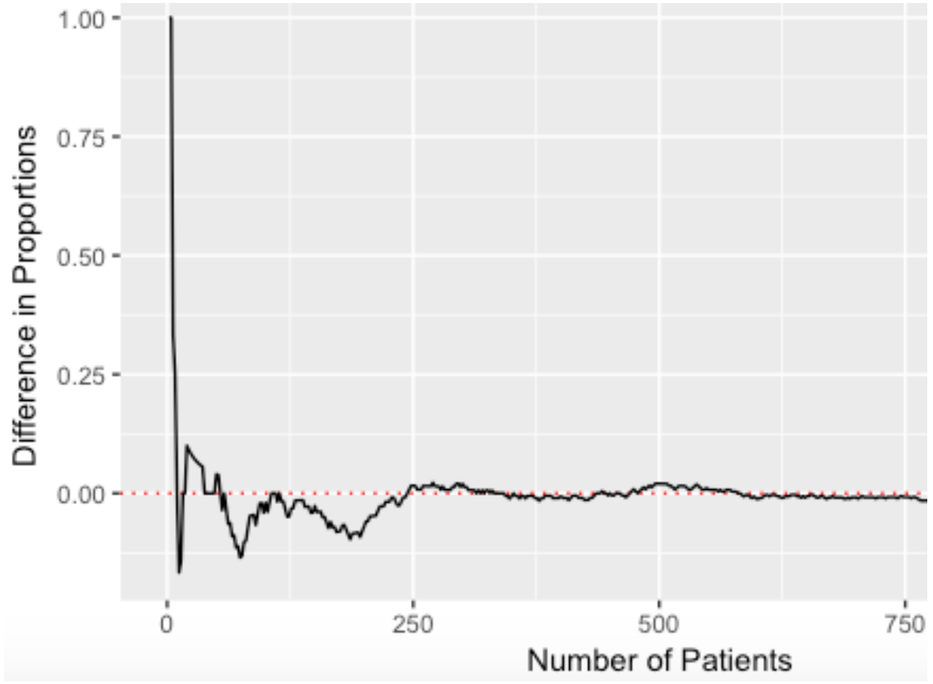
Stopping rules

- Interim analyses often require increased sample size
- Multiple testing increases chance of Type I error



Stopping rules

- Interim analyses often require increased sample size
- Multiple testing increases chance of a Type I error
- Stopping rules use p-values or test-statistics



Pocock (Fixed Nominal) Rule

K, number of planned analyses	Nominal significance level, α
1 (No interim analysis)	0.05
2 (1 interim + 1 final analysis)	0.029
3	0.022
4	0.018
5	0.016

Pocock (Fixed Nominal) Rule

K, number of planned analyses	Nominal significance level, α
1 (No interim analysis)	0.05
2 (1 interim + 1 final analysis)	0.029
3	0.022
4	0.018
5	0.016

```
library(gsDesign)
Pocock <- gsDesign(k=3, test.type=2, sfu="Pocock")
2*(1-pnorm(Pocock$upper$bound))
```

Pocock (Fixed Nominal) Rule

K, number of planned analyses	Nominal significance level, α
1 (No interim analysis)	0.05
2 (1 interim + 1 final analysis)	0.029
3	0.022
4	0.018
5	0.016

```
library(gsDesign)
Pocock<- gsDesign(k=3, test.type=2, sfu="Pocock")
2*(1-pnorm(Pocock$upper$bound))
```

```
[1] 0.02205159 0.02205159 0.02205159
```

Pocock (Fixed Nominal) Rule

K, number of planned analyses	Nominal significance level, α
1 (No interim analysis)	0.05
2 (1 interim + 1 final analysis)	0.029
3	0.022
4	0.018
5	0.016

```
library(gsDesign)
Pocock<- gsDesign(k=3, test.type=2, sfu="Pocock")
2*(1-pnorm(Pocock$upper$bound))
```

```
[1] 0.02205159 0.02205159 0.02205159
```

```
Pocock.ss <- gsDesign(k=3, test.type=2, sfu="Pocock", n.fix=200, beta=0.1)
ceiling(Pocock.ss$n.I)
```

Pocock (Fixed Nominal) Rule

K, number of planned analyses	Nominal significance level, α
1 (No interim analysis)	0.05
2 (1 interim + 1 final analysis)	0.029
3	0.022
4	0.018
5	0.016

```
library(gsDesign)
Pocock<- gsDesign(k=3, test.type=2, sfu="Pocock")
    2*(1-pnorm(Pocock$upper$bound))
```

```
[1] 0.02205159 0.02205159 0.02205159
```

```
Pocock.ss<- gsDesign(k=3, test.type=2, sfu="Pocock", n.fix=200, beta=0.1)
    ceiling(Pocock.ss$n.I)
```

```
[1]  77 154 231
```

O’Brien-Fleming Rule

O’Brien-Fleming Rule

```
OF <- gsDesign(k=3, test.type=2, sfu="OF")
2*(1-pnorm(OF$upper$bound))
```

```
[1] 0.0005183476 0.0141107255 0.0450662542
```

O'Brien-Fleming Rule

```
OF <- gsDesign(k=3, test.type=2, sfu="OF")
2*(1-pnorm(OF$upper$bound))
```

```
[1] 0.0005183476 0.0141107255 0.0450662542
```

```
OF.ss <- gsDesign(k=3, test.type=2, sfu="OF", n.fix=200, beta=0.1)
ceiling(OF.ss$n.I)
```

```
[1]  68 136 204
```

DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Let's practice!

DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Sample Size for Alternative Trial Designs

Tamuno Alfred, PhD
Biostatistician

Equivalence trials

N- New treatment
S- Standard treatment

$$H_0: |\pi_N - \pi_S| \geq \delta$$
$$H_A: |\pi_N - \pi_S| < \delta$$

Equivalence trials: binary outcomes

- Response rate in existing drug: 70%
- Delta: 5%
- Power: 90%
- Two one-sided tests (TOST)
- 90% sure that two-sided 90% CI excludes delta +/-5%

Equivalence trials: binary outcomes

- Response rate in existing drug: 70%
- Delta: 5%
- Power: 90%
- Two one-sided tests (TOST)
- 90% sure that two-sided 90% CI excludes delta +/-5%

```
library(TOSTER)
powerTOSTtwo.prop(alpha = 0.05, statistical_power = 0.9, prop1 = 0.7,
        prop2 = 0.7, low_eqbound_prop = -0.05, high_eqbound_prop = 0.05)
```

Equivalence trials: binary outcomes

- Response rate in existing drug: 70%
- Delta: 5%
- Power: 90%
- Two one-sided tests (TOST)
- 90% sure that two-sided 90% CI excludes delta +/-5%

```
library(TOSTER)
powerTOSTtwo.prop(alpha = 0.05, statistical_power = 0.9, prop1 = 0.7,
        prop2 = 0.7, low_eqbound_prop = -0.05, high_eqbound_prop = 0.05)
```

```
The required sample size to achieve 90 % power with equivalence bounds of -0.05 and 0.05 is 1819
[1] 1818.125
```

Equivalence trials: continuous outcomes

- Delta: 3 units
- Pooled standard deviation: 15
- Power: 80%
- Two one-sided tests (TOST)
- 80% sure that two-sided 90% CI excludes delta +/-3

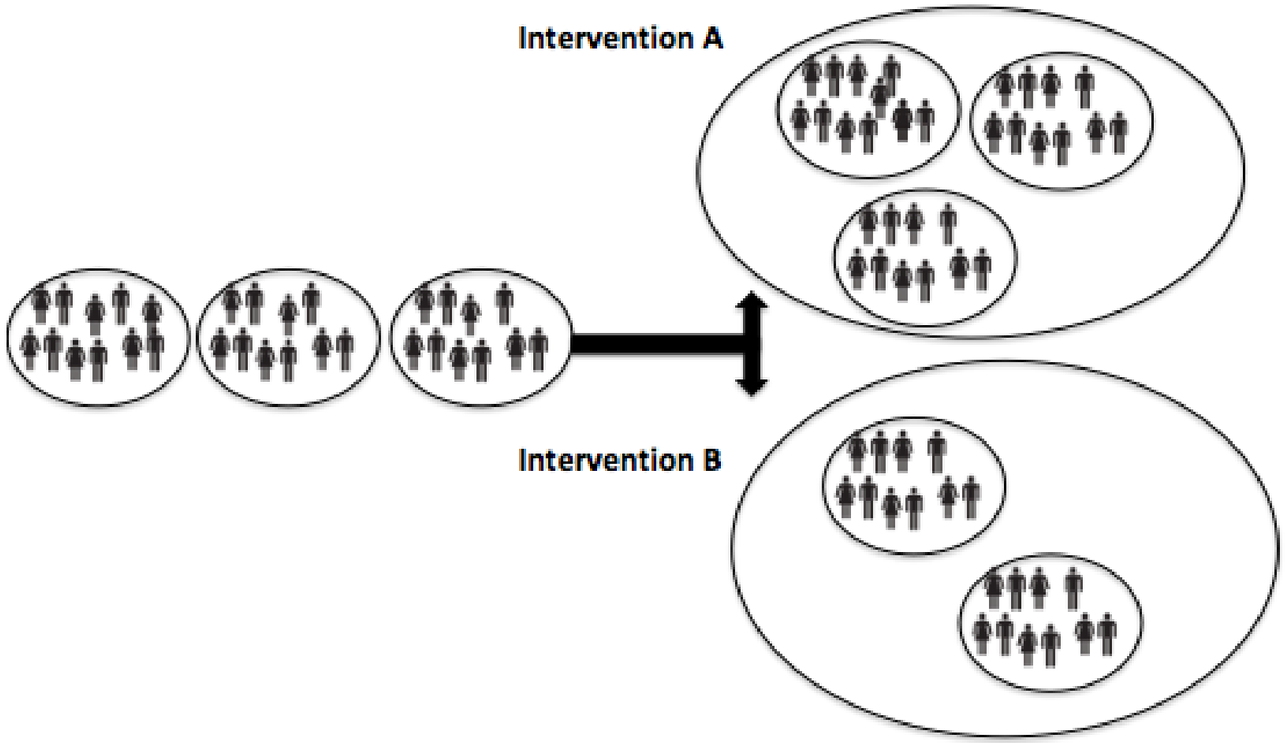
Equivalence trials: continuous outcomes

- Delta: 3 units
- Pooled standard deviation: 15
- Power: 80%
- Two one-sided tests (TOST)
- 80% sure that two-sided 90% CI excludes delta +/-3

```
library(TOSTER)
powerTOSTtwo.raw(alpha=0.05, statistical_power=0.8,
         sdpooled=15, low_eqbound=-3,high_eqbound=3)
```

```
The required sample size to achieve 80 % power with equivalence bounds of -3 and 3 is 428.1924 per group,
or 858 in total.
[1] 428.1924
```

Cluster randomized trials



Cluster randomized trials

Cluster randomized trials

- Delta: 1 unit
- Pooled standard deviation: 2.5
- Average cluster size: 25
- Intraclass correlation coefficient (ICC): 0.1
- Power: 90%

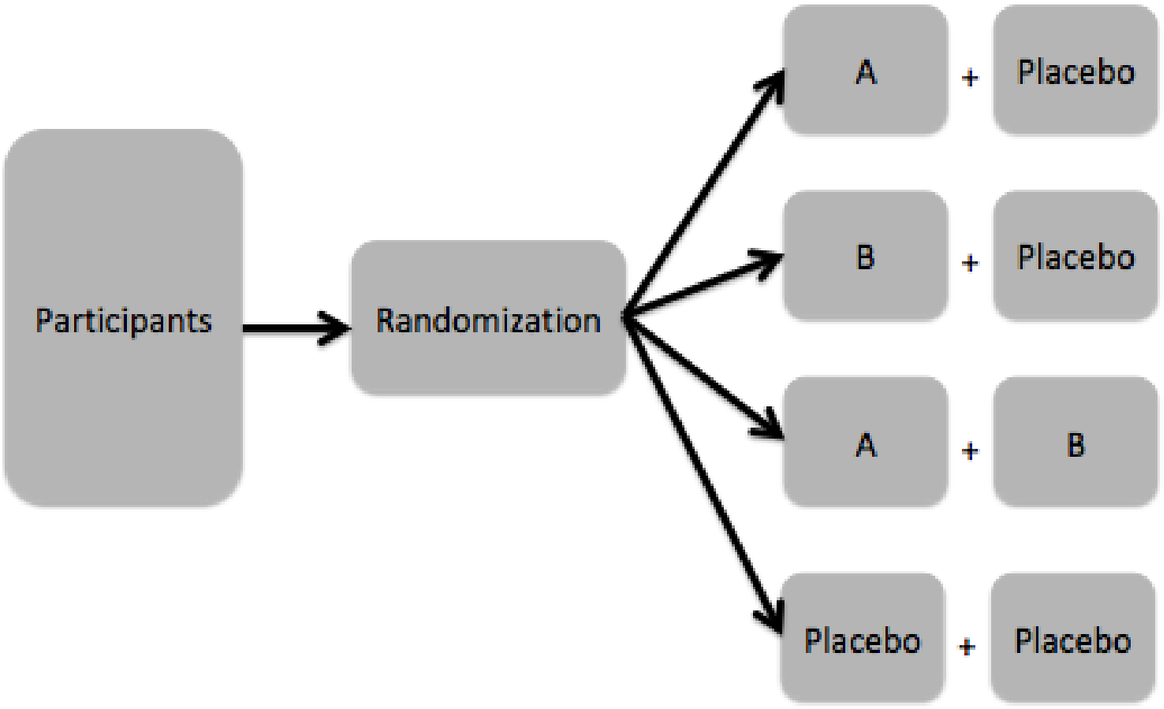
Cluster randomized trials

- Delta: 1 unit
- Pooled standard deviation: 2.5
- Average cluster size: 25
- Intraclass correlation coefficient (ICC): 0.1
- Power: 90%

```
libarary(CRTSize)
n4means(delta=1, sigma=2.5, m=25, ICC=0.1, alpha=0.05, power=0.90)
```

```
The required sample size is a minimum of 19 clusters of size 25 in the Experimental Group
 and a minimum of 19 clusters (size 25) in the Control Group.
```

Factorial designs



Factorial designs

- Assume independence
- Expected rate in placebo: 40%
- Expected rate in intervention A: 25%
- Expected rate in intervention B: 23%
- Power: 90%

Factorial designs

```
power.prop.test(p1=0.40, p2=0.25,
       power=0.9)
```

```
     Two-sample comparison of proportions power calculation

              n = 202.8095
             p1 = 0.4
             p2 = 0.25
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```

```
power.prop.test(p1=0.40, p2=0.23,
   power=0.9)
```

```
     Two-sample comparison of proportions power calculation

              n = 154.8146
             p1 = 0.4
             p2 = 0.23
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```

DESIGNING AND ANALYZING CLINICAL TRIALS IN R

Let's practice!