MAHIDOL UNIVERSITY
Wisdom of the Land

Sample Size Estimation & Randomization

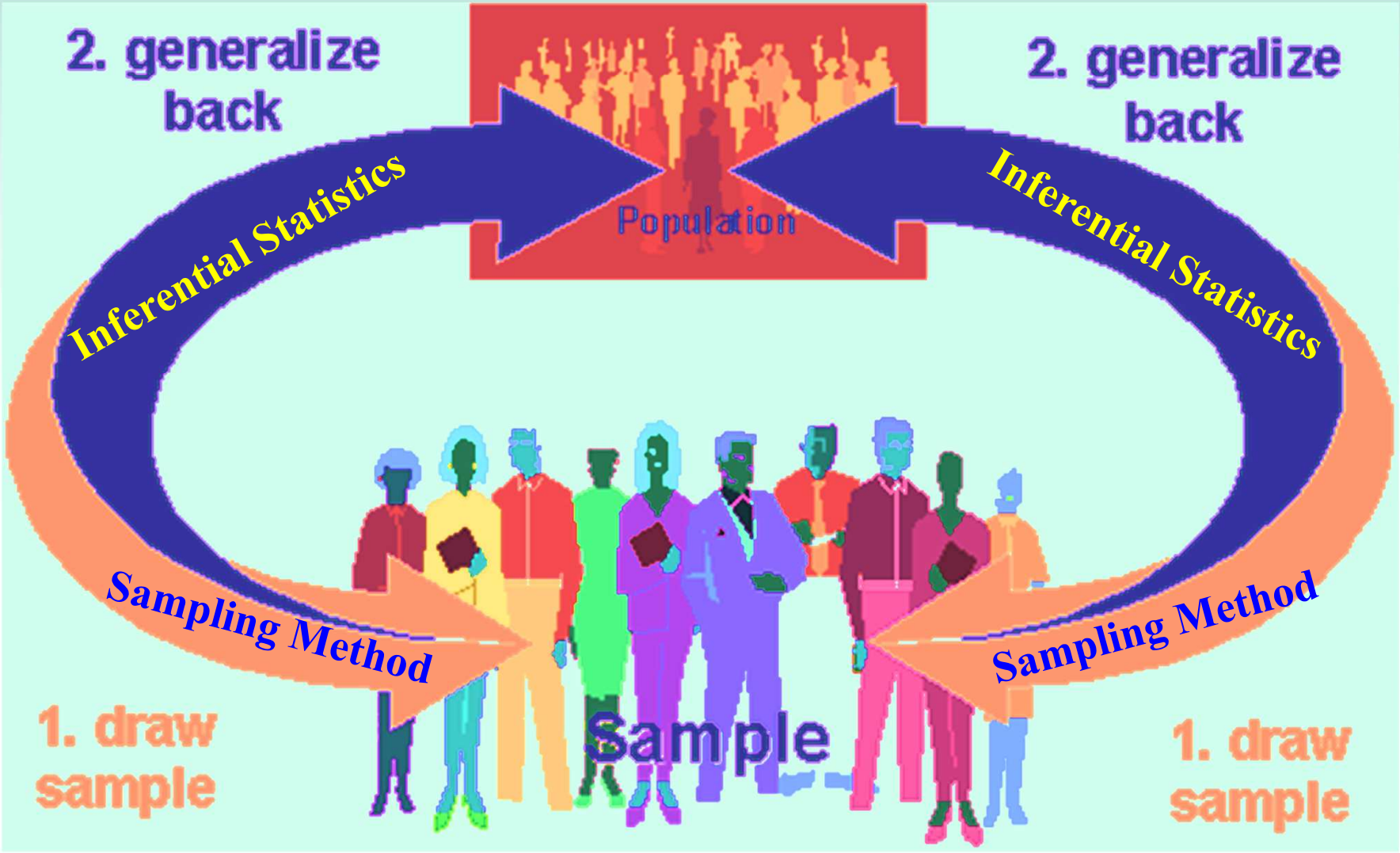
Jaranit Kaewkungwal

Faculty of Tropical Medicine

BIOPHICS
Center of Excellence for Biomedical and Public Health Informatics

Population & Sample

Sample vs. Population



Sample vs. Population

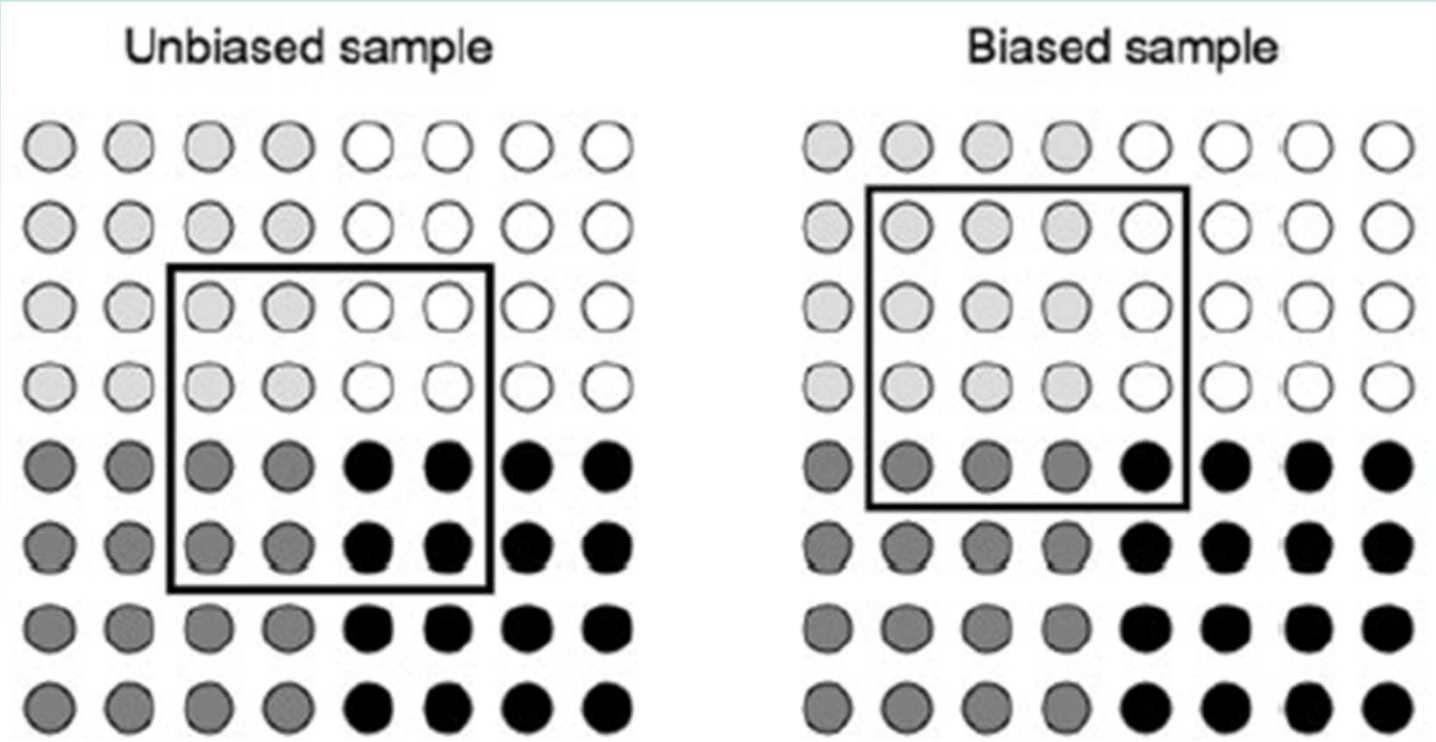


Fig. 3. Sampling bias. An unbiased sample is representative of and has the same characteristics as the population from which it has been drawn. A biased sample is not representative of the target population because its characteristics have different distribution as compared with the original population.

Sample Specification

- **Inclusion Criteria**

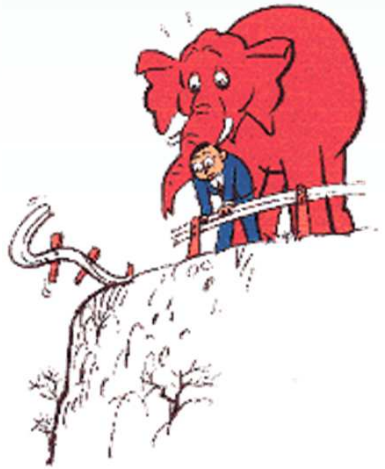
Specifying the characteristics that define populations that are relevant to the research question and efficient for the study:

- *Demographic characteristics*
- *Clinical characteristics*
- *Geographic (administrative) characteristics*
- *Temporal characteristics*

- **Exclusion Criteria**

Specifying subset of the population that will not be studied because of:

- *high likelihood of being lost to follow-up*
- *unable to provide good/complete data*
- *ethical barriers*
- *subject's refusal to participate*



Methods of Sampling

- **Probability Sampling** -- methods that utilizes some form of *random selection*

  1. **Simple Random Sampling**
  2. **Stratified Random Sampling**
  3. **Systematic Random Sampling**
  4. **Cluster (Area) Sampling**
  5. **Multi-stage Sampling**

Chance

EPS: Equal Probability of Selection

PPS: Proportionate to Size

- **Non-probability Sampling** - methods that based on either accidental or purposive; usually approach the sampling problem with a specific plan in mind.

  1. **Accidental Sampling**
  2. **Purposive Sampling**
     - 2.1 **Expert Sampling**
     - 2.2 **Quota Sampling**
     - 2.3 **Heterogeneity Sampling**
     - 2.4 **Snowball Sampling**

Relevancy

Representativeness

Specific Characteristics

Sample Size Estimation

Important questions in sample size estimation

What is the *key outcome of interest* which is to be evaluated statistically?

Cured/Not Cured, BP, Glucose conc., Survival time, No. of E. coli, …

How will the key outcome be measured?

Rate, Percent, Prevalence, Incidence, Mean, Median, etc.

What kind of study does one have?

Descriptive (Parameter estimation), Analytic (Hypothesis testing)

Are there explicit or implicit dependencies in the data which need to be accounted for?

Completeness, Non-responses, Follow-up rate,
Fixed /Limited sample size, Screening etc.

Factors in sample size estimation

- ***A priori*** **information about parameters of interest**
- **Effect size**
- **Confidence level** ***(in parameter estimation)*** **/ Tail of the test** ***(in hypothesis testing)***
- **Type I error (**α***, in parameter estimation)*** **/ Type I (**α**) & Type II (**β**) errors** ***(in hypothesis testing)***

Factors in sample size estimation

- ***A priori* information about parameters of interest**
  - ***Literature Review***

    From previous report, it was shown that cure rate of Drug A = 70%

  - ***Pilot Study***

    A pilot survey from 30 bottles of drinking water in the market shows that there are E. coli in 5 bottles.

  - ***Expert Opinion***

    3 out of 5 experts say that about 10% of workers in the XXX factory have health problem related to toxic chemicals.

Factors in sample size estimation

- **Example of *a priori* information about parameters of interest**

Definition of Pimary Outcome: PID

- **Tenderness: abdominal direct, motion of cervix and uterus, and adnexal**
- **GC+ or fever > 38◦C or leucocytosis >10,000 WBC/μl or purulent material from peritoneal cavity on culdocentesis or pelvix abscess or inflammatory complex on bimanual exam**

Estimating the Incidence of PID for Sample Size Calculations

- **Government officials estimated 40%**
- **Ob/GYN from Med School estimated 12%**
- **Pilot study found 4%**
- **We conservatively set initially at 6%**

Factors in sample size estimation

- **Effect size**
  - *Clinical/ Public Health Importance*
  - *Not Statistical Significance*
- **Examples:**

Current cure rate = 70%

New drug should be 10% better => 80%

Previous survey found infected rate = 15%

New survey expected to find infected rate not different from previous survey at ± 3% => 12-18%

Relationship Between Priori Info and Effect Size

- Sample size is function of the
  - α type I error allowed
  - β type II error allowed
  - actual predicted risk
  - expected reduction of risk
- The estimated sample size of each arm of a clinical trial, if the tolerated α type I error is 0.05 and β type II error is 0.1?

	Predicted Risk					
	1%	2%	3%	4%		10%
10% risk reduction	197,750	97,924	64,649	48,011		18,064
50% risk reduction	6,253	3,100	2,049	1,524		578

10% - 1% 9%

10% - 5% 5%

Factors in sample size estimation

- **Type I & Type II errors**

Ho: G1 = G2

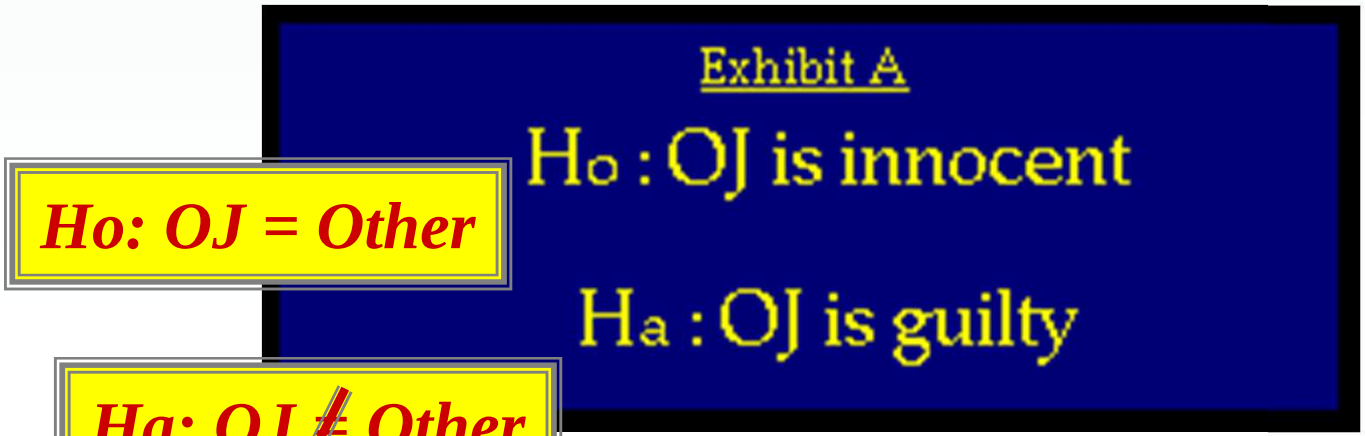
Decision	Reality/Truth: Ho True *(G1=G2)*	Reality/Truth: Ho False *(G1<>G2)*
Accept Ho	**Correct** Confidence : 1 - α .99, .95	**Type II Error** β .10, .20
Reject Ho	**Type I Error** α .01, .05	**Correct** Power : 1 - β .90, .80

Factors in sample size estimation

- **Type I & Type II errors**

The Decision Matrix on Trial

The OJ Simpson Trial Analogy



Ho: OJ = Other

Ha: OJ ≠ Other

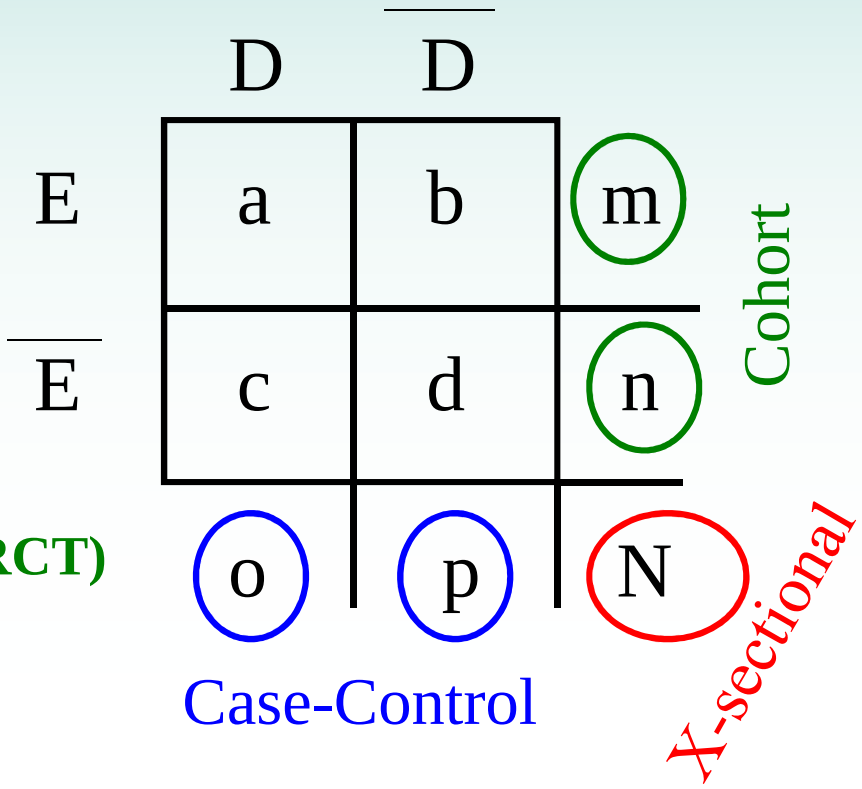
Factors in sample size estimation

- **Type I & Type II errors**

	Reality Truth: Ho True (Ho: OJ = Other)	Reality Truth: Ho False (Ha: OJ ≠ Other)
Accept Ho (Ho: OJ = Other) — What jury con-cludes	CONFIDENCE LEVEL OJ is Innocent Not Guilty Verdict	TYPE II Error OJ is Gulity Not Guilty Verdict
Reject Ho (Ha: OJ ≠ Other)	TYPE I Error OJ is Innocent Guilty Verdict	POWER OJ is Gulity Guilty Verdict

Factors in sample size estimation

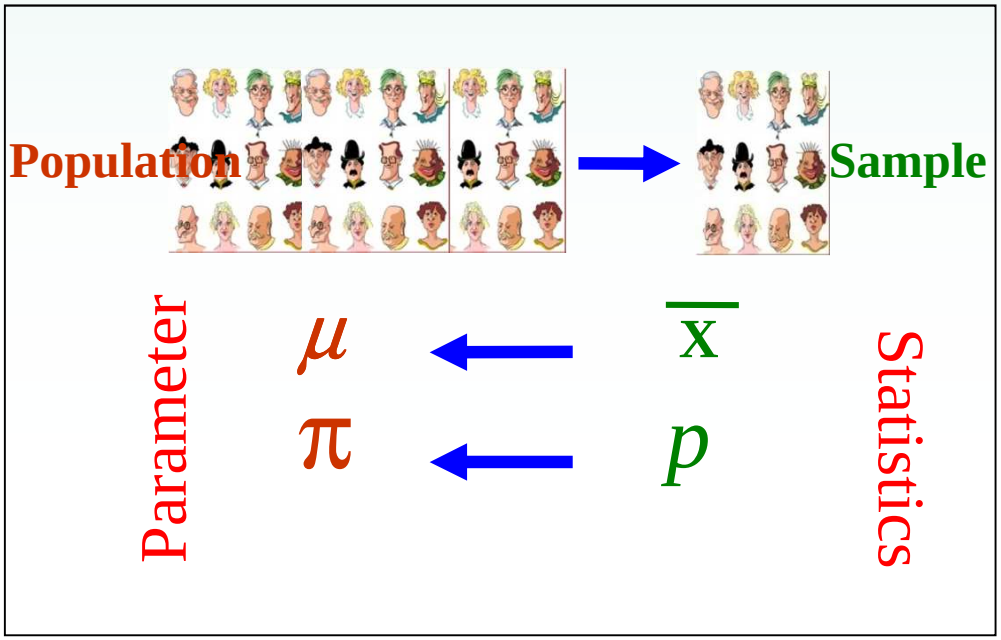
- **Type of Research Study**
  - **Descriptive**
    (Parameter Estimation)
  - **Analytic**
    (Hypothesis Testing)
- **Research Design**
  - **Experimental**
    - **True Experimental (e.g. RCT)**
    - **Quasi Experimental**
  - **Observational**
    - **Cross-sectional**
    - **Case-Control**
    - **Cohort**



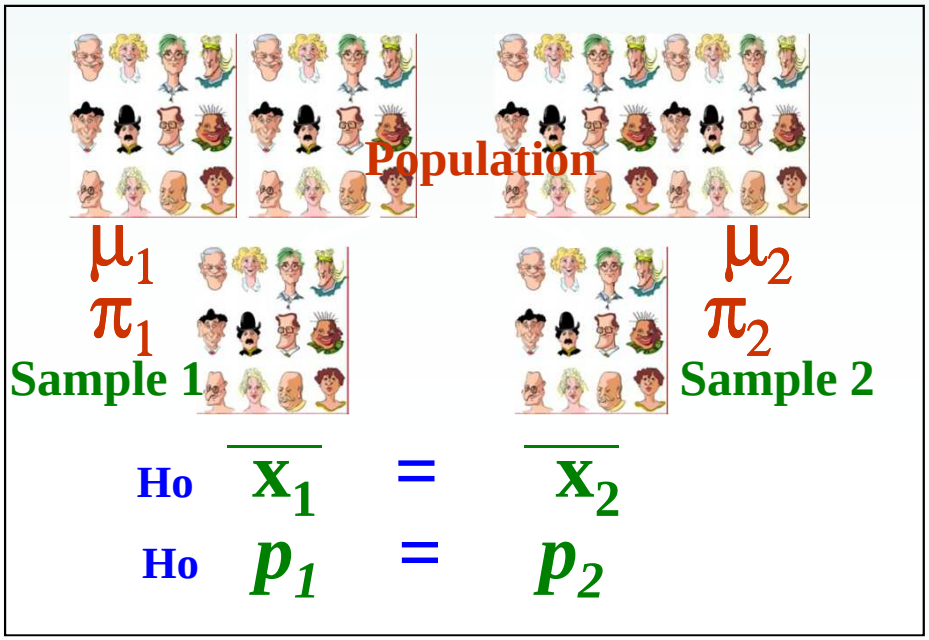
Examples:

Sample Size Estimation

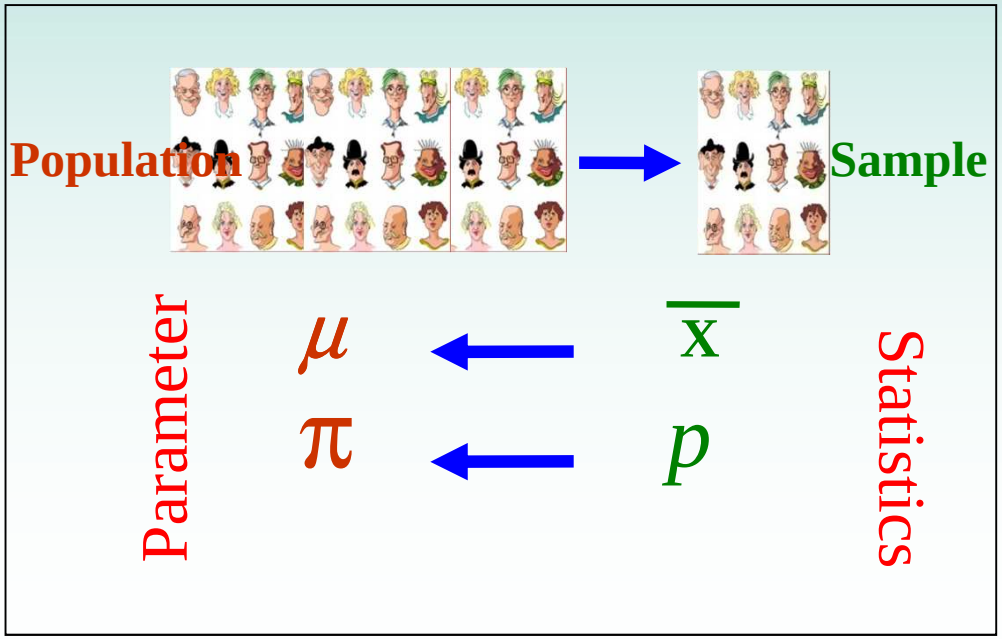
• Parameter Estimation



• Hypothesis Testing



Sample size for Parameter Estimation



Categorical outcome: $\pi = p \pm Z_{\alpha/2}\sqrt{p(1-p)/n}$

Continuous outcome: $\mu = \overline{x} \pm Z_{\alpha/2}\sigma/\sqrt{n}$

Determinants of sample size (Parameter Estimation)

Three factors determine the required sample size:

1. Standard deviation, σ (continuous); or, the proportion or prevalence rate of the outcome of interest, p (categorical).
2. The difference of the estimate that we wish to detect, δ.
3. The confidence interval level (usually set at 95% CI).

Formula:

Priori Info.

$Z_{\alpha/2} = Z_{0.05/2}$

Effect Size

Categorical outcome: $n = 1.96^2 \pi(1-\pi) / \delta^2$

Continuous outcome: $n = (1.96\ \sigma / \delta)^2$

Example: Descriptive study

Target organ damage and cardiovascular complications in patients with hypertension and type 2 diabetes in Spain: a cross-sectional study

Luis Cea-Calvo*[1], Pedro Conthe[2], Pablo Gómez-Fernández[3], Fernando de Alvaro[4], Cristina Fernández-Pérez[5] and RICARHD investigators[6]

Categorical outcome
(prevalence - p)

Background: Target organ damage (mainly cardiac and renal damage) is easy to evaluate in outpatient clinics and offers valuable information about patient's cardiovascular risk. The purpose of this study was to evaluate, using simple methods, the prevalence of cardiac and renal damage and its relationship to the presence of established cardiovascular disease (CVD), in patients with hypertension (HT) and type 2 diabetes mellitus (DM).

Statistical analysis $n = 1.96^2 \pi (1-\pi) / \delta^2$

The sample size was calculated according to the main objective of the study and based on the expected prevalence of heart and kidney damage. For an expected prevalence of <10% (π), a sample size of 2401 hypertensive diabetic patients was estimated for a 95% confidence interval (CI) (1.96) and an error of 1.2% (δ). The sample was increased 4% to cover data losses, yielding a definitive size of 2500 patients.

Example: Descriptive study

Cross-sectional survey on hantavirus seroprevalence in Canton St. Gallen, Switzerland

Detlev Schultze[a], Walter Fierz[b], Hans C. Matter[c], Sergej Bankoul[d], Matthias Niedrig[e], Andreas Schmiedl[f]

Categorical outcome
(prevalence - p)

Background and objectives: In 2002 the first endemic hantavirus infection in Switzerland was detected only by chance following a broad spectrum of diagnostics. This raised the question, whether Hantavirus infection should be included in the differential diagnosis of febrile illness of patients in Switzerland. In order to estimate the frequency of hantavirus infections in Switzerland, this survey on hantaviral seroprevalence was conducted in the Canton St. Gallen.

Example: Descriptive study

Cross-sectional survey on hantavirus seroprevalence in Canton St. Gallen, Switzerland

Detlev Schultze[a], Walter Fierz[b], Hans C. Matter[c], Sergej Bankoul[d], Matthias Niedrig[e], Andreas Schmiedl[f]

$$n = 1.96^2 \pi (1-\pi) / \delta^2$$

The sample size calculation was based on a comparison of exposure to hantaviruses measured by seroprevalence in different populations. Based on known seroprevalences in other Central European countries [22–27] an exposure of 5% in the populations with higher risk of exposure to hantaviruses and of 0.5% in blood donors was assumed. With a power of 0.8, an alpha value of 0.05, and a ratio of seropositive samples for the higher risk group to the blood donors of 10:1, 94 volunteers from each population and 940 blood donors would have been needed [28].

Finally, a total of 1710 sera from 1029 blood donors, 382 farmers, 104 forestry workers, 104 soldiers, and 91 hunters were collected.

Example: Descriptive study

Effect size = Relative precision

Suppose we are trying to estimate the prevalence of a certain disease, which we suspect to be about 3%, and

We want the 95% confidence interval of the estimate to be 0.3% (10% of 3%) on either side

$Z_{\alpha/2} = Z_{0.05/2}$ (Priori Info.) (Effect Size)

$$n = 1.96^2 \pi (1-\pi) / \delta^2$$

$$n = 1.96^2 \times 0.03(1-0.03) / 0.003^2 = 12421$$

i.e. 12421 subjects required!

Example: Descriptive study
Effect size = Absolute precision

A survey is being planned to estimate the prevalence of secondary infertility amongst couples aged 20-45. The investigators expect the prevalence to be 10%, and

They would like to estimate it to within 5% of the true value (with 95% confidence).

How many couples are required?

$Z_{\alpha/2} = Z_{0.05/2}$ Priori Info. Effect Size

$$n = 1.96^2 \pi(1-\pi) / \delta^2$$

$$n = 1.96^2 \times 0.1 \times 0.9 / 0.05^2 = 139$$

Example: Descriptive study
Multiple outcomes of interest

Title: Prevalence of gonoccocal, chlamydial, hepatitis B virus and syphillis infections among pregnant women attending Hung Vuong Hospital, Ho Chi Minh City, Vietnam

Sample size:

Using EPI-Info6 (epitable) calculation for sample size & power of single proportion

Given:

- Size of the population (pregnant women attending Hung Vuong Hospital) = 60 /day, or approximately = 22000 /yr
- Confidence level = 90%

Thus,

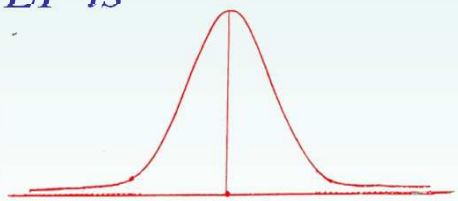
- Prevalence of the Chlamydia trachomatis carriers among pregnant women expected to be = 2.5% (Citation ?) with desired precision % = 1 %; sample size = 641
- Prevalence of Treponema pallidum carriers among pregnant women expected to be = 0.5% (Citation ?) with desired precision % = 0.3 %; sample size = 1401
- Prevalence of HBsAg carriers among pregnant women expected to be = 10.3% (Citation ?) with desired precision % = 1.5 %; sample size = 1058
- Prevalence of Gonorrhoea carriers among pregnant women expected to be = 0.7% (Citation ?) with desired precision % = 0.4 %; sample size = 1117

If the desired precision is acceptable, the minimum sample size for this study should be 1400

Example: Descriptive study
Estimate continuous outcome

Peak Expiratory Flow Rate (PEFR)

This is a simple method of measuring airway obstruction and it will detect moderate or severe disease. The simplicity of the method is its main advantage. It is measured using a standard Wright Peak Flow Meter or mini Wright Meter. The needle must always be reset to zero before PEF is measured.



(Data from Gregg et al, BMJ, 1973)

Standard deviation of PEFR: 48 litres/min

$\overline{X}$ = ? ± 20 litres/min
SD = 48 litres/min

Desired 95% confidence interval width: ± 20 litres/min

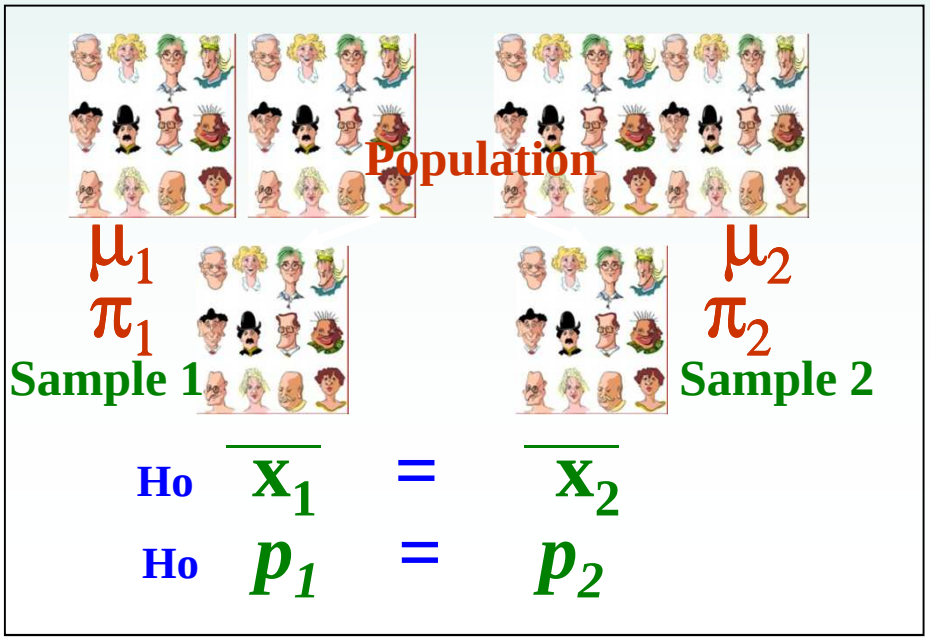
$$n = (1.96\ \sigma / \delta)^2$$

$$n = (1.96 \times 48 / 20)^2 \approx 22$$

i.e. a sample of 22 would enable us to estimate the population PEFR mean to within 20 litres/min (with 95% probability).

Sample size for Hypothesis Testing

(Differences between independent groups)



Categorical outcome:

Ho: $\pi 1 = \pi 2$

Continuous outcome:

Ho: $\mu 1 = \mu 2$

Determinants of sample size (Hypothesis Testing)

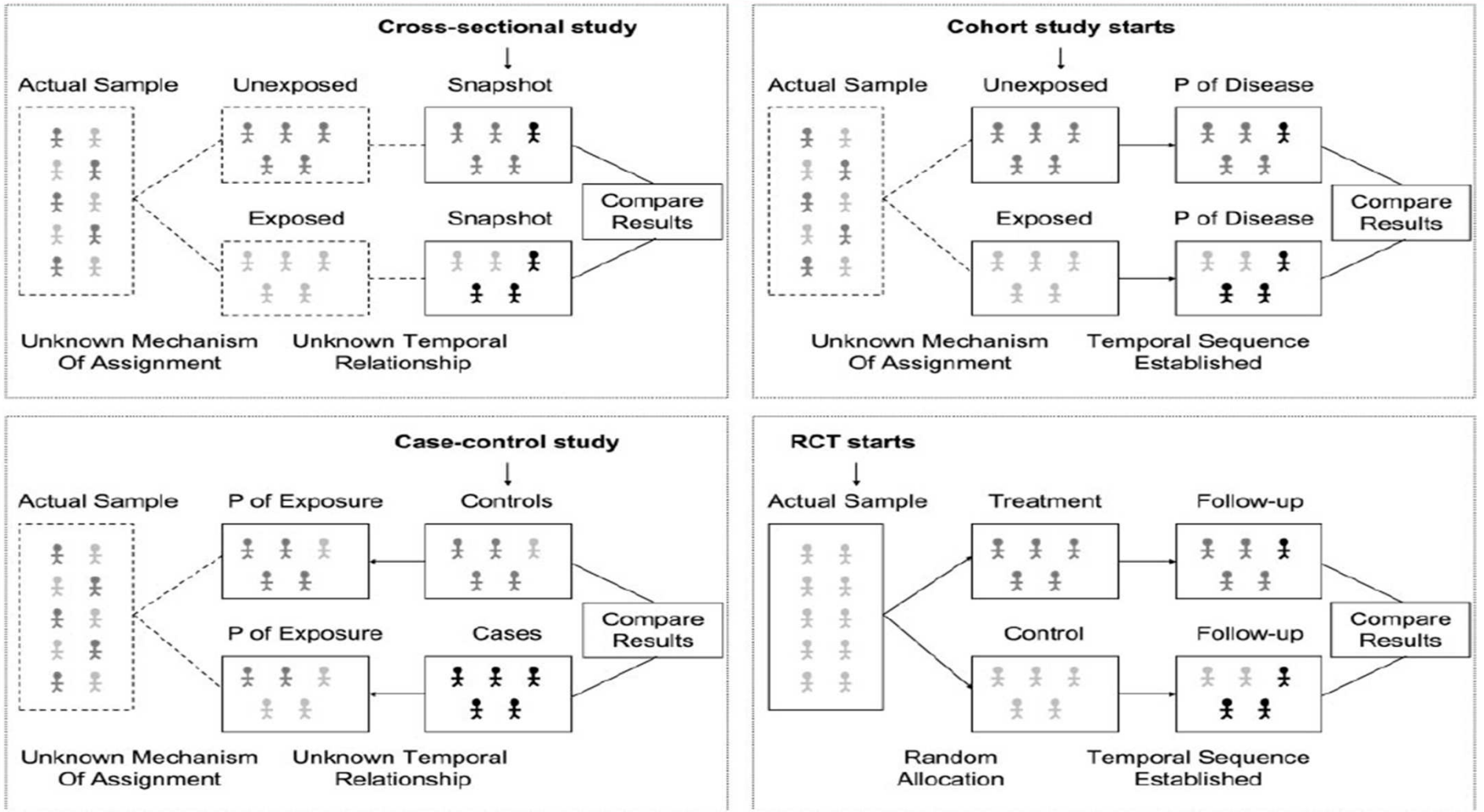


Fig. 1. Examples of study designs. In cross-sectional studies inputs and output are measured simultaneously and their relationship is assessed at a particular point in time. In case-control studies participants are identified based on presence or absence of the disease and the temporal direction of the inquiry is reversed (retrospective). Temporal sequences are better assessed in longitudinal cohort studies where exposure levels are measured first and participants are followed forward in time. The same occurs in randomized controlled trials (RCTs) where the assignment of the exposure is under the control of the researcher. P: Probability (or risk).

Determinants of sample size (Hypothesis Testing)

MAHIDOL UNIVERSITY
Wisdom of the Land

BIOPHICS
Center of Excellence for Biomedical and Public Health Informatics

Four factors determine the required sample size:

1. Standard deviation, σ (continuous), or the expected success rate for the control group, p_1 (categorical).

2. The difference between groups that we wish to detect, δ.

Note: Effect size: the size of the smallest effect that is clinically important.

- *Difference between two groups = (p1-p2) or ($\mu 1 = \mu 2$)*
- *OR/RR of two groups*

  e.g., 1.5 for risk of CHD in patients with hypertension (50% increased risk)

  H_o: RR=1.0; H_1: RR>=1.5

3. The false positive error rate, or significance level of the test, α (usually 0.05).

4. The false negative error rate, β (usually 0.1 or 0.2), more commonly expressed as (1-β), the power (0.8 or 0.9). This is for a specified δ (alternative hypothesis).

Difference between two groups

Formula: Categorical outcome

Ho: π1 = π2

Type I err

Power (1-Type II err)

Priori Info.

$$n = \frac{\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2 [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

Effect Size

Where

p_1 = expected/known proportion in the control group

p_2 = expected proportion in the intervention group (= $p_1 + \delta$)

Formula: Continuous outcome

Ho: μ1 = μ2

Type I err

Power (1-Type II err)

Priori Info.

$$n = \frac{2\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2 \sigma^2}{\delta^2}$$

Effect Size

Where

σ = expected/known standard deviation in the control group

™= expected different standard deviation in the intervention group

Difference between two proportions
Ho: $\pi 1 = \pi 2$

Formula: Categorical outcome

Type I err

Power (1-Type II err)

Priori Info.

$$n = \frac{\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2 [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

Effect Size

Where

p_1 = *expected proportion in the control group*

p_2 = *expected proportion in the intervention group* ($= p_1 + \delta$)

Thus

For 5% significance and 80% power:

$$n = \frac{7.849 \times [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

Note: This is a slight underestimate to the n.

More accurate formula:

$$n = \frac{\{z_{1-\frac{\alpha}{2}} \sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)}\}^2}{(p_1 - p_2)^2}$$

Where

p bar is the mean of p_1 and p_2.

Example: Analytic study

Intermittent versus continuous renal replacement therapy for acute kidney injury patients admitted to the intensive care unit: results of a randomized clinical trial

Robert L. Lins[1], Monique M. Elseviers[2], Patricia Van der Niepen[3], Eric Hoste[4], Manu L. Malbrain[5], Pierre Damas[6] and Jacques Devriendt[7] for the SHARF investigators

Categorical outcome

In the SHARF 4 study, we compared prospectively the outcome of different modes of therapy [daily intermittent renal replacement therapy (IRRT) versus continuous renal replacement therapy (CRRT)]. The SHARF score was used to control for disease severity. This article will focus on the comparison between both treatment options in a randomized clinical trial with the short-term outcome on hospital mortality and renal recovery at hospital discharge used as end-points.

Example: Analytic study

Intermittent versus continuous renal replacement therapy for acute kidney injury patients admitted to the intensive care unit: results of a randomized clinical trial

Robert L. Lins[1], Monique M. Elseviers[2], Patricia Van der Niepen[3], Eric Hoste[4], Manu L. Malbrain[5], Pierre Damas[6] and Jacques Devriendt[7] for the SHARF investigators

Categorical outcome

$$n = \frac{\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2 [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

Sample size calculation

The sample size calculation was based on the assumption that the overall mortality would be 50% as in the former SHARF studies [19,20] and that a difference of 10% in mortality between IRRT and CRRT had to be detected to be clinically relevant. With a first-order error of 5% and a power of 80% a sample size of 407 patients was needed in each treatment group.

P1

P1 – P2

Example: Analytic study

Single-Dose Rufloxacin versus 3-Day Norfloxacin Treatment of Uncomplicated Cystitis: Clinical Evaluation and Pharmacodynamic Considerations

G. DEL RÍO,[1]* F. DALET,[1] L. AGUILAR,[2] J. CAFFARATTI,[1] AND R. DAL-RÉ[2]

Nephrology Department, Microbiology Department, and Urology Department, Fundación Puigvert, Barcelona,[1] and Medical Department, SmithKline Beecham Pharmaceuticals, Madrid,[2] Spain

Categorical outcome

The purpose of this study was to evaluate the efficacy and safety of a single dose of rufloxacin (400 mg), compared with those of a norfloxacin standard treatment (400 mg twice a day [BID] for 3 days).

Sample size and statistical analysis. The sample size was calculated on the basis of the expected norfloxacin cure rate of 96% (28) and a delta value of 10% to show no differences between treatments (22). With type I and II errors of 0.05 and 0.2, respectively, 61 evaluable patients per treatment should be recruited (22). Sample size was, however, extended to 150 evaluable patients, following the IDSA/FDA guidelines for the clinical evaluation of anti-infective drug products (29).

Treatment group comparisons were done with chi-square tests and Fisher's exact test for qualitative parameters and Mann-Whitney and Spearman correlation for quantitative parameters. Two-tailed P values of <0.05 were considered statistically significant.

Example: Analytic study

Eur J Dermatol 2009; 19 (5): 461-8

Clinical efficacy of basic fibroblast growth factor (bFGF) for diabetic ulcer

Categorical outcome

In the present trial, we compared the efficacy of two doses of bFGF (0.001% and 0.01%) with a placebo in patients with diabetic ulcers.

The sample size was determined based on the Grade 1 achievement rate. In a 12-week preliminary trial conducted using 0.01% bFGF in patients with diabetic ulcers (data unpublished), it was observed that 5 of 8 patients achieved Grade 1 after 8 weeks' administration. Assuming that the Grade 1 achievement rate in the placebo group and 0.01% bFGF group were 30% and 60%, respectively, the planned sample size of 50 patients on each group would provide a power of 80% to detect a percent difference of 30% at a significant level of 0.05. If, however, the percent difference in the Grade 1 achievement rate was 25%, the power would decrease to 64%. We set the minimum number of patients which could have been shown the statistical significant difference because this study was positioned as an exploratory study.

Example: Analytic study
RCT- Placebo vs. Treatment

In a randomised clinical trial, the placebo response is anticipated to be 25%, and the active treatment response 65%.

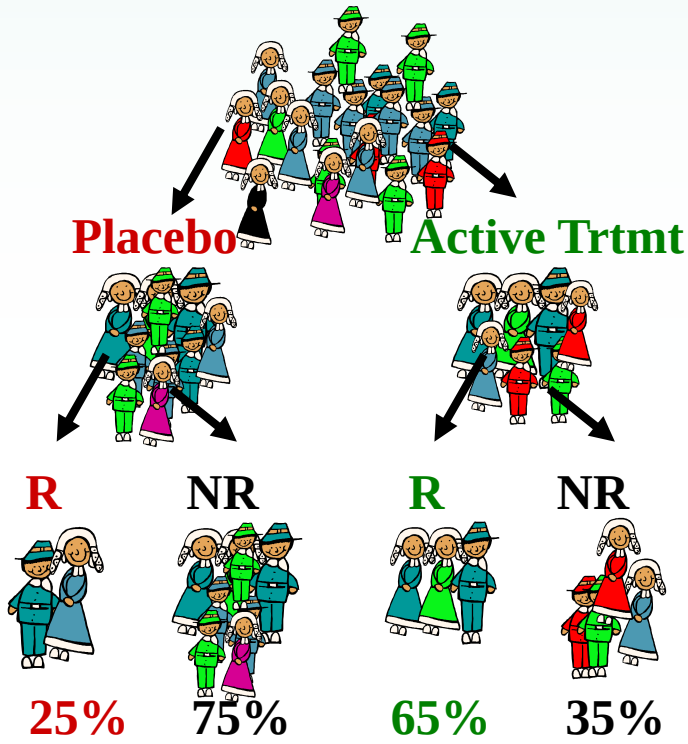
How many patients are needed if a two-sided test at the 5% level is planned, and a power of 80% is required?

$$n = \frac{7.849 \times [p_1(1-p_1) + p_2(1-p_2)]}{(p_1 - p_2)^2}$$

$$n = \frac{7.849 \times (0.25(1-0.25) + 0.65(1-0.65))}{(0.25-0.65)^2}$$

$$= \frac{7.849 \times 0.415}{0.16} = 20.4$$

so n=21 per group.



Example: Analytic study
Case-Control Design

Does oral contraceptive use cause myocardial infarction?

p_1 = proportion of exposure among controls = 10% = 0.10

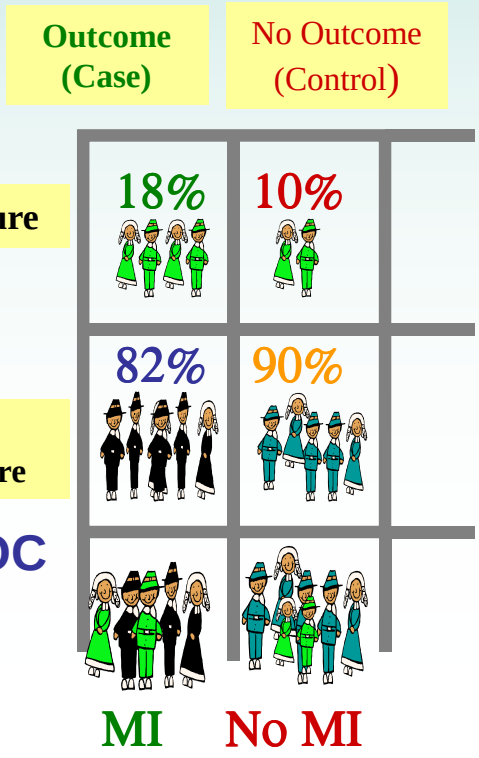
p_2 = proportion of exposure among cases = $(p_1)(PR)$
= (0.10)(1.8)=0.18 if assume Prev Ratio = 1.8

$q_2 = 1 - p_2 = 1 - 0.18 = 0.82$

$q_1 = 1 - p_1 = 1 - 0.10 = 0.90$

$z_{1-\alpha/2}$ = value of the standard normal distribution corresponding to alpha: e.g., 1.96 for 2-sided test at 0.05

$z_{1-\beta}$ = value of the standard normal distribution corresponding to desired power level: e.g., 0.84 for a power of 80%



$$n \text{ (each group)} = \frac{[(0.1)(0.9) + (0.18)(0.82)]\,[1.96 + 0.84]^2}{(0.18 - 0.10)^2} = \frac{(0.2376)\,(7.84)}{0.0064} = 291.06$$

Example: Analytic study
Case-Control Design

Table for N for OC and MI Study

Postulated Prevalence Ratio	Required n/group
1.2	3834
1.3	1769
1.5	682
1.8	291
2.0	196
2.5	97
3.0	59

Assumes 10% use of OC in population, power=80%, alpha=0.05

Difference between means

Ho: μ1 = μ2

MAHIDOL UNIVERSITY
Wisdom of the Land

BIOPHICS
Center of Excellence for Biomedical and Public Health Informatics

Formula: Continuous outcome

Type I err

Power (1-Type II err)

Priori Info.

$$n = \frac{2\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2 \sigma^2}{\delta^2}$$

Effect Size

The most commonly used value for significance (α) is 0.05, giving $z_{1-\alpha/2} = 1.96$

The most commonly used value for power (1-β) is 80%, giving $z_{1-\beta} = 0.84$

$$n = \frac{2\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2 \sigma^2}{\delta^2} = \frac{2\{1.96 + 0.84\}^2 \sigma^2}{\delta^2} = \frac{2 \times 7.849 \sigma^2}{\delta^2} \approx \frac{15.7 \sigma^2}{\delta^2}$$

Example: Analytic study

Continuous outcome

Proof-Of-Concept Trial In Patients With Dominant Polycystic Kidney Disease:

Study Aim

The primary objective of the SUISSE ADPKD study is to assess the effectiveness of sirolimus to retard kidney volume growth and to prevent the loss of renal function in young patients with ADPKD and preserved renal function. Patients with ADPKD and kidney volume growth that can be documented within 6 months are randomized to treatment with sirolimus 2 mg/day for 18 months (Figure 1) or standard treatment. The secondary objectives are to follow renal function and blood pressure and to monitor for the occurrence of proteinuria. Safety and tolerability of sirolimus treatment in ADPKD patients will also be assessed.

Example: Analytic study

Continuous outcome

Proof-Of-Concept Trial In Patients With Dominant Polycystic Kidney Disease:

$$n = \frac{2\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2 \sigma^2}{\delta^2}$$

Sample Size Considerations

In a large cohort of ADPKD patients, the mean annual kidney volume growth rate was 5.27% ± 3.92% (SD).[7] Because patients with lack of progression during the pre-randomisation period will be excluded from our study, we expect to select for a higher progression rate in our study population. Due to a shorter observation interval compared to the mentioned observational study, the standard deviation might be higher. Presuming an annual kidney growth rate of 6% ± 4.75% (SD) in the control group, a sample size of 40 patients per group will have 80% statistical power to detect a 50% relative reduction of kidney volume growth using a two-sided α-level of 0.05. To account for a drop out rate of up to 20%, we plan to randomise a total of 100 patients.

Example: Analytic study
RCT - Continuous outcomes

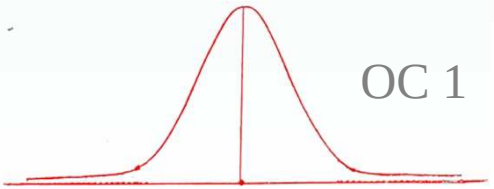
In a trial to compare the effects of two oral contraceptives on blood pressure (over one year), it is anticipated that one drug will increase diastolic blood pressure by 3 mmHg, and the other will not change it. The SD (of the changes in blood pressure) in both groups is expected to be 10 mmHg.

How many patients are required for this difference to be significant at the 5% level (with 80% power)?

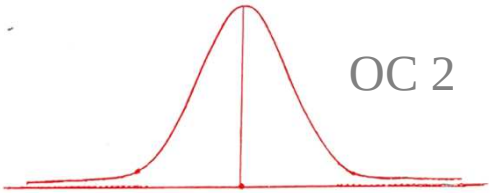
$$n = \frac{2\{z_{1-\frac{\alpha}{2}} + z_{1-\beta}\}^2\sigma^2}{\delta^2} = \frac{2\{1.96+0.84\}^2\sigma^2}{\delta^2} = \frac{2\times 7.849\sigma^2}{\delta^2}$$

$$n = \frac{2\times 7.849\times 100}{9} = 175$$

So n = 175 women per group



$\overline{X}$post-pre = 0 mmHg
SD = 10 mmHg

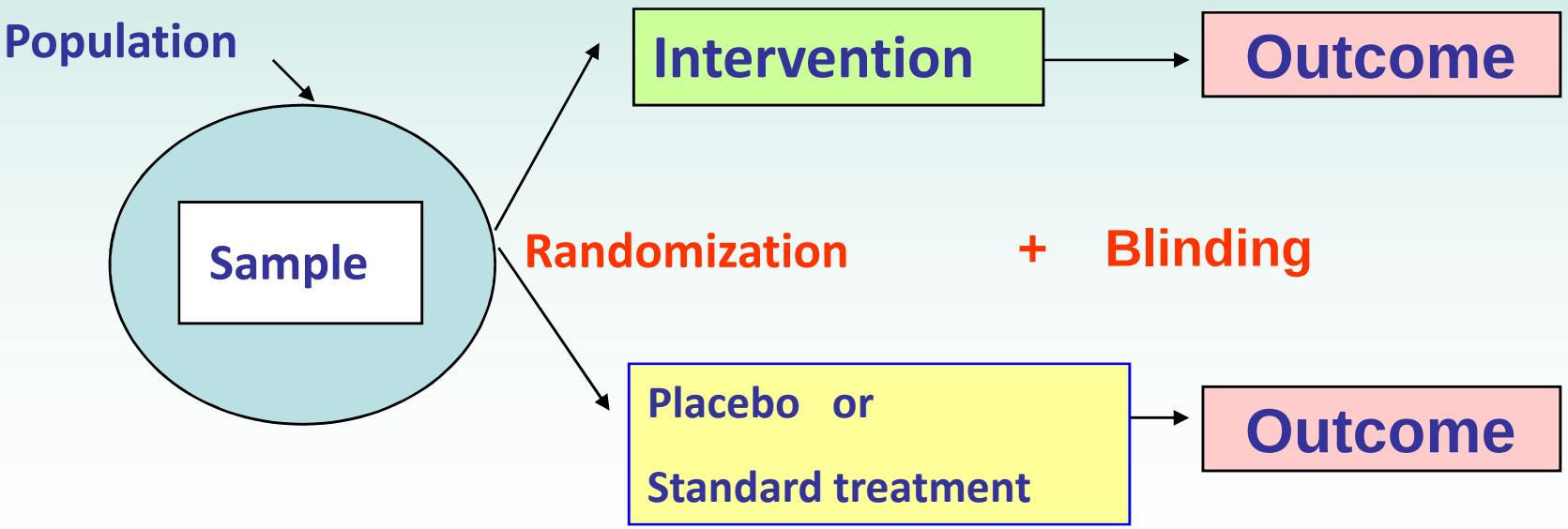


$\overline{X}$post-pre = 3 mmHg
SD = 10 mmHg

Variations of CT

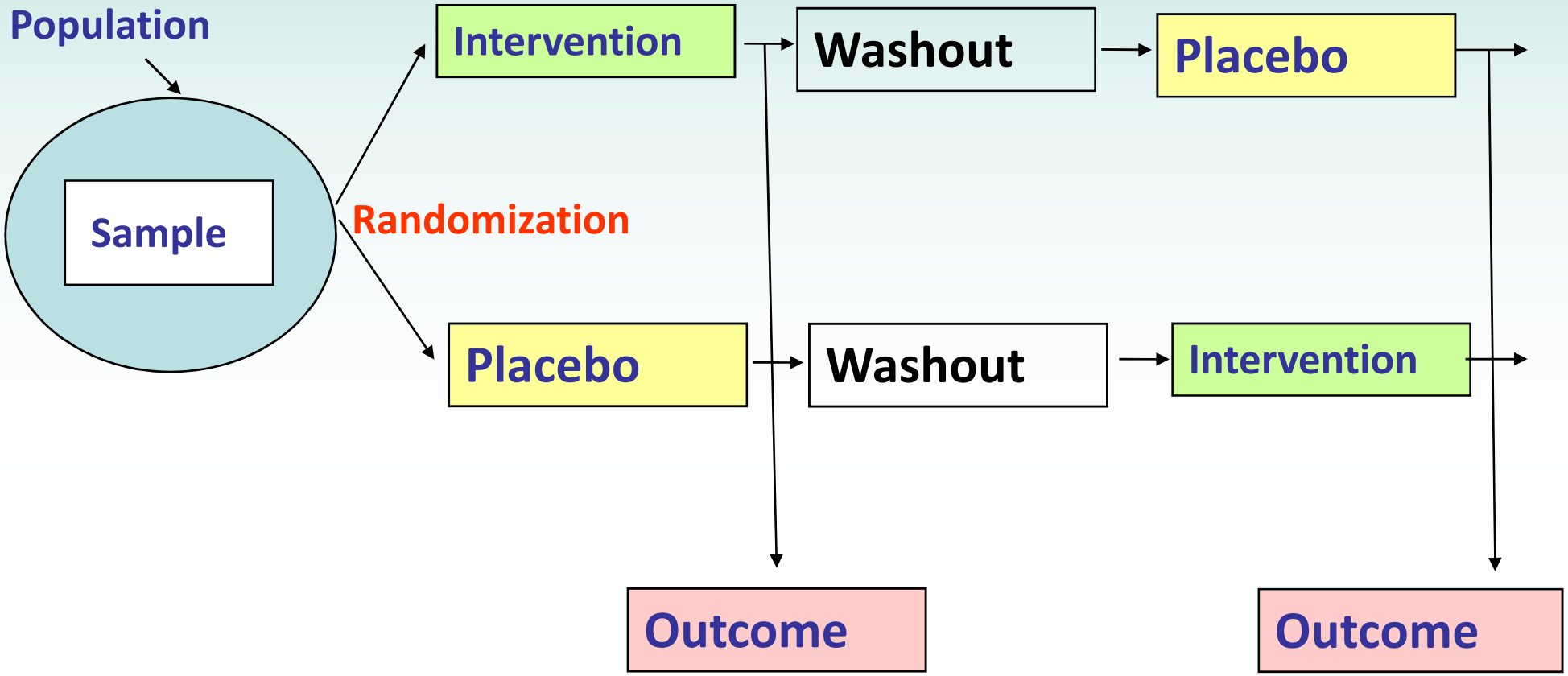
การทดสอบเพื่อแสดง	Null Hypothesis	Alternative Hypothesis
ความเท่ากัน (Equality)	$H_0: \mu_T - \mu_c = 0$	$H_a: \mu_T - \mu_c \neq 0$
ความเหนือกว่า (Superiority)	$H_0: \mu_T - \mu_c \leq \delta$	$H_0: \mu_T - \mu_c > \delta$
ความไม่ด้อยกว่า (Non – inferiority)	$H_0: \mu_T - \mu_c \geq \delta$	$H_0: \mu_T - \mu_c < \delta$
ความเสมอภาคกัน (Equivalence)	$H_0: \lvert \mu_T - \mu_c \rvert \geq \delta$	$H_0: \lvert \mu_T - \mu_c \rvert < \delta$

Basic Trial Design



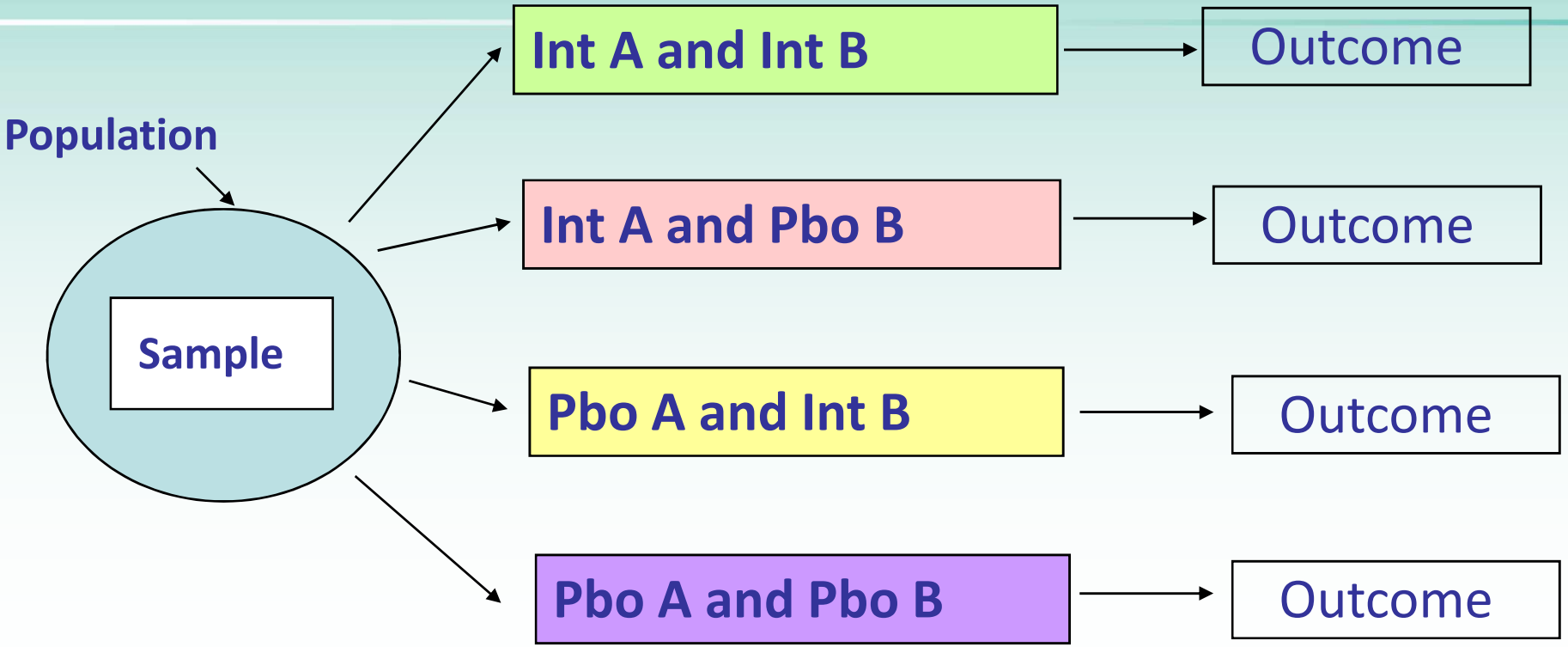
Source: Deborah Grady, Introduction to Randomized Clinical Trials

Cross-over Design



Source: Deborah Grady, Introduction to Randomized Clinical Trials

Factorial Design



	Int A	Plb A
Int B	AB	B
Plb B	A	Pbo

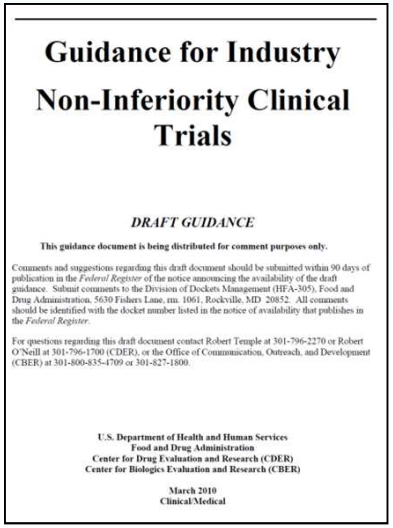
	Int A	Int B
Dose 1		
Dose 2		

Source: Deborah Grady, Introduction to Randomized Clinical Trials

Equivalence / Non-inferior Trials

Table 1. Hypotheses Associated with the Different Types of Studies when Comparing a New Therapy Against a Current Therapy with Respect to Efficacy

Type of study	Null hypotheses	Research hypothesis
Traditional comparative	There is no difference between the therapies	There is a difference between the therapies
Equivalence	The therapies are not equivalent	The new therapy is equivalent to current therapy
Noninferiority	The new therapy is inferior to the current therapy	The new therapy is not inferior to the current therapy



Guidance for Industry

Non-Inferiority Clinical Trials

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact Robert Temple at 301-796-2270 or Robert O'Neill at 301-796-1700 (CDER), or the Office of Communication, Outreach, and Development (CBER) at 301-800-835-4709 or 301-827-1800.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

March 2010
Clinical/Medical

Source: E Walker & A S. Nowacki, Understanding Equivalence and Noninferiority Testing

Equivalence / Non-inferior Trials

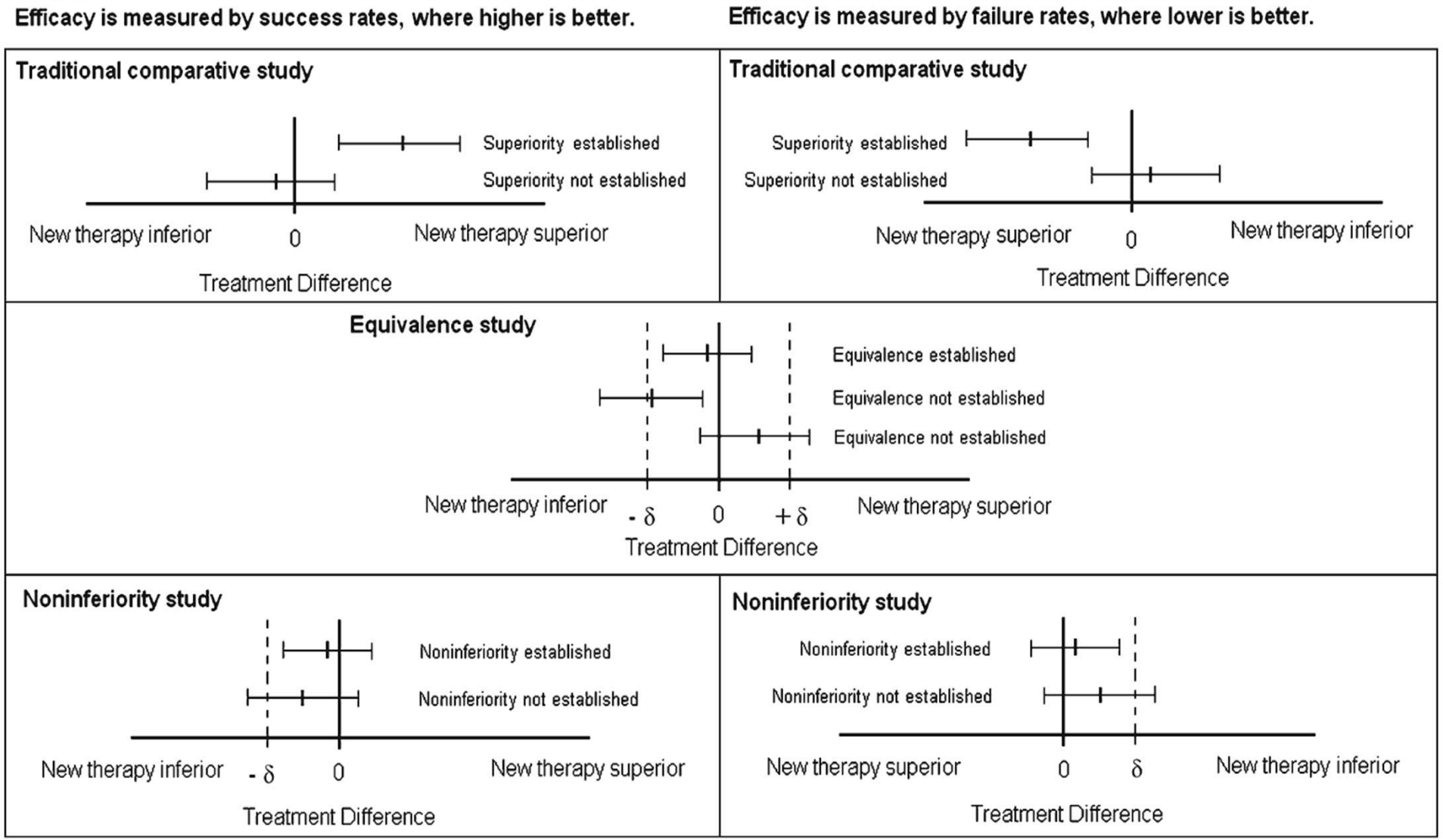
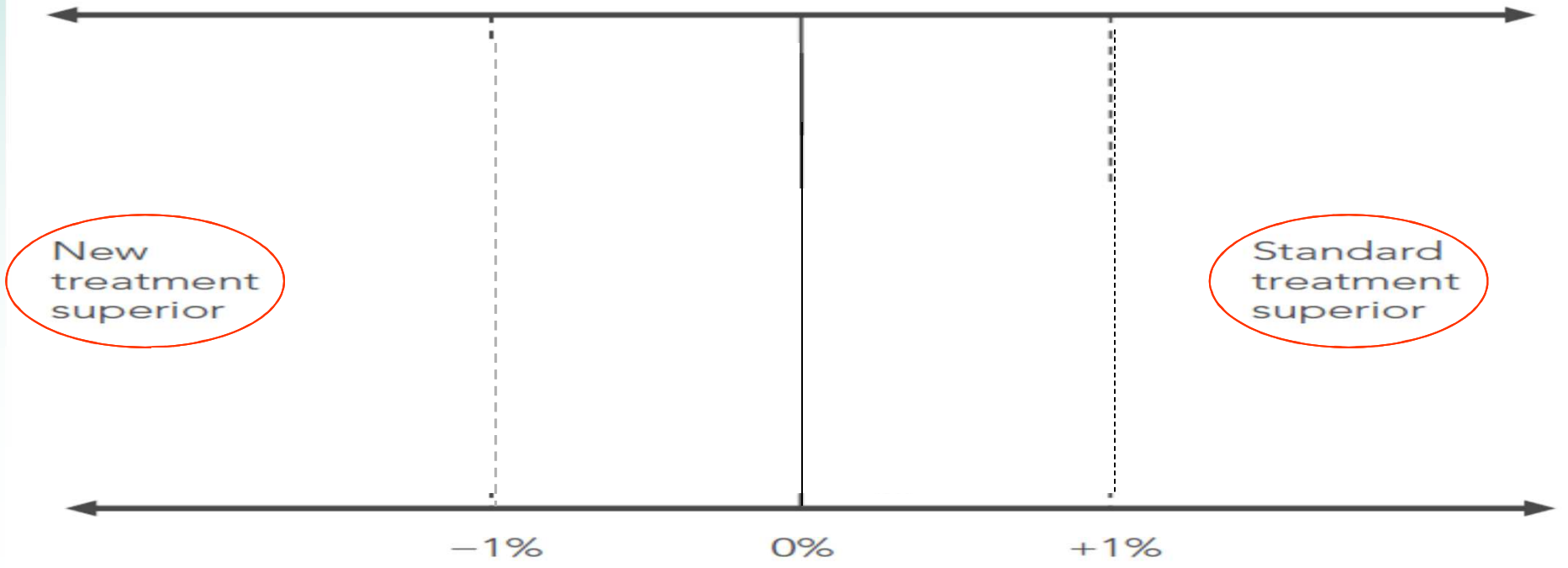


Figure 1. Two one-sided test procedure (TOST) and the equivalence margin in equivalence/noninferiority testing.

Source: E Walker & A S. Nowacki, Understanding Equivalence and Noninferiority Testing

Equivalence / Non-inferior Trials

2 Comparison of superiority, equivalence and non-inferiority* hypotheses based on a 2% margin of difference in event rates

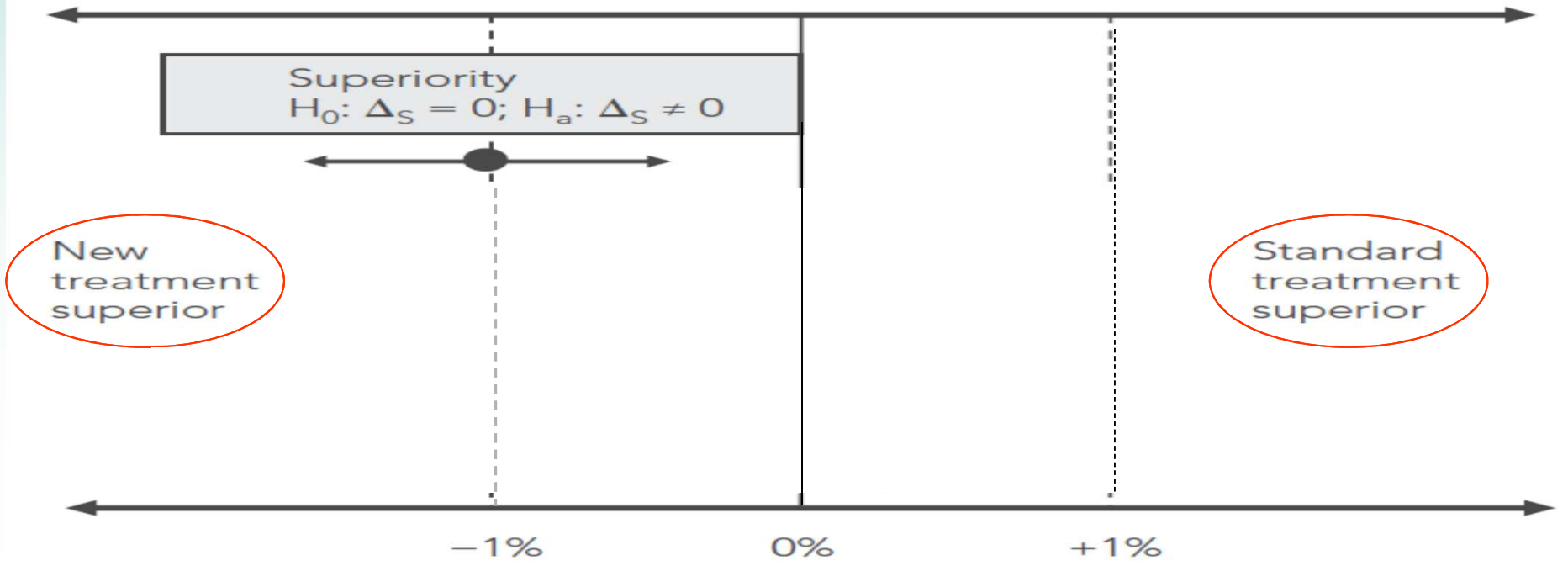


H_0 = null hypothesis. H_a = alternative hypothesis.
Δ = difference in event rates between new and standard treatments.
S = Δ in superiority trials. E = Δ in equivalence trials.
NI = Δ in non-inferiority trials.
* Testing for non-inferiority is in one direction only — even if superiority exists (dashed arrow), it is not the hypothesis being tested. ◆

Source: Ian A Scott, MJA • Volume 190 Number 6 • 16 March 2009

Equivalence / Non-inferior Trials

2 Comparison of superiority, equivalence and non-inferiority* hypotheses based on a 2% margin of difference in event rates

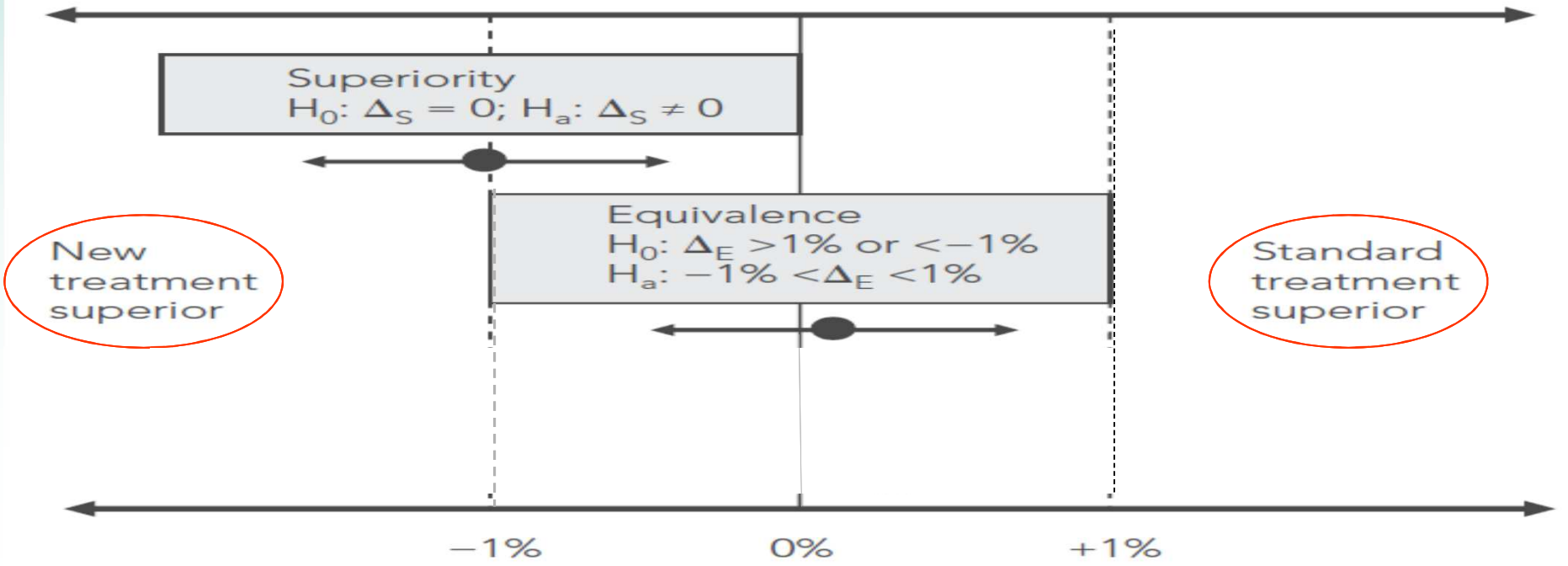


H_0 = null hypothesis. H_a = alternative hypothesis.
Δ = difference in event rates between new and standard treatments.
S = Δ in superiority trials. E = Δ in equivalence trials.
NI = Δ in non-inferiority trials.
* Testing for non-inferiority is in one direction only — even if superiority exists (dashed arrow), it is not the hypothesis being tested. ◆

Source: Ian A Scott, MJA • Volume 190 Number 6 • 16 March 2009

Equivalence / Non-inferior Trials

2 Comparison of superiority, equivalence and non-inferiority* hypotheses based on a 2% margin of difference in event rates

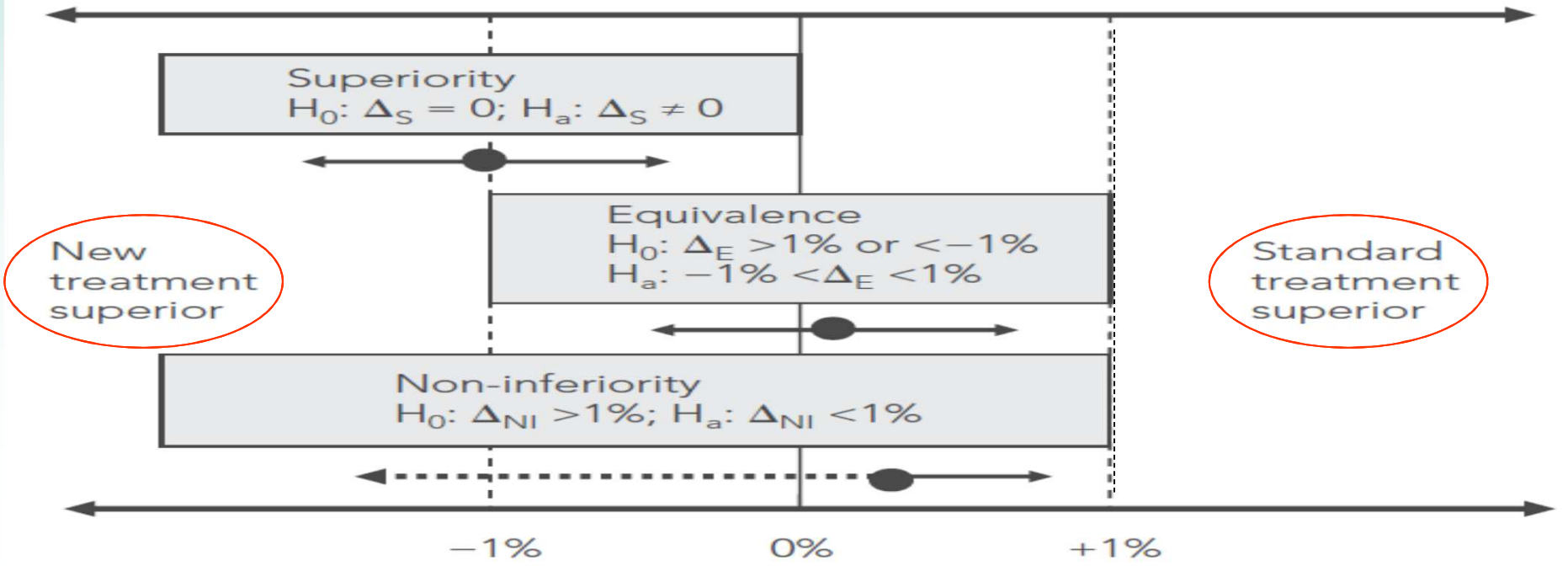


H_0 = null hypothesis. H_a = alternative hypothesis.
Δ = difference in event rates between new and standard treatments.
S = Δ in superiority trials. E = Δ in equivalence trials.
NI = Δ in non-inferiority trials.
* Testing for non-inferiority is in one direction only — even if superiority exists (dashed arrow), it is not the hypothesis being tested. ◆

Source: Ian A Scott, MJA • Volume 190 Number 6 • 16 March 2009

Equivalence / Non-inferior Trials

2 Comparison of superiority, equivalence and non-inferiority* hypotheses based on a 2% margin of difference in event rates



H_0 = null hypothesis. H_a = alternative hypothesis.
Δ = difference in event rates between new and standard treatments.
S = Δ in superiority trials. E = Δ in equivalence trials.
NI = Δ in non-inferiority trials.
* Testing for non-inferiority is in one direction only — even if superiority exists (dashed arrow), it is not the hypothesis being tested. ◆

Source: Ian A Scott, MJA • Volume 190 Number 6 • 16 March 2009

New Approaches: Bayesian / Adaptive Designs

Figure 11. Paradigm Shift

	Old	New
Inferential Process	Hypothesis Testing (Attempt to reject *null hypothesis*)	Continuous Learning (Update probabilities of alternative *hypotheses*)
Question Being Addressed	How likely are the trial results, given there really is no difference among treatments?	How likely is there a true difference among treatments, given the trial data?
Drug Approval	Pivotal Trial Distinct Phase 0-IV Trials	Weight of Evidence Continuous Trials
Trial Designs	Single Stage	Adaptive
Statistics	Traditional	Bayesian

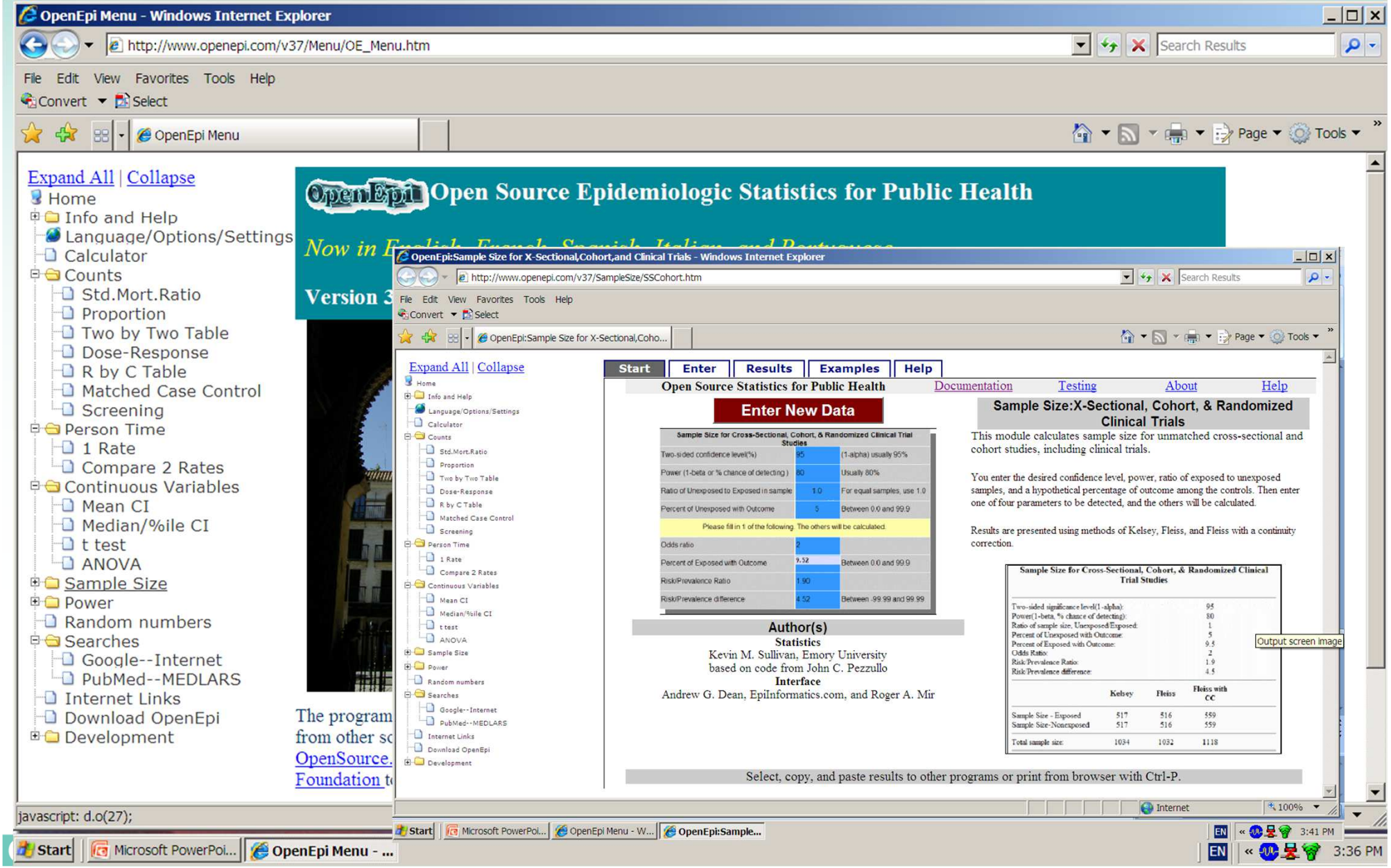
Sample Size Formula for Variations of CT – Continuous Outcomes

รูปแบบการวิจัย	สมมติฐาน	สมมติฐานและสูตรคำนวณขนาดกลุ่มตัวอย่าง		
		H_0	H_a	สูตรพื้นฐาน
กลุ่มเดี่ยว (One – Sample)	Equality	$\mu - \mu_0 = 0$	$\mu - \mu_0 \neq 0$	$n = \frac{\left(z_{\frac{\alpha}{2}} + z_\beta\right)^2 \sigma^2}{(\mu - \mu_0)^2}$
	Superiority	$\mu - \mu_0 \leq \delta$	$\mu - \mu_0 > \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu - \mu_0 - \delta)^2}$
	Equivalence	$\lvert \mu - \mu_0 \rvert \geq \delta$	$\lvert \mu - \mu_0 \rvert < \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\lvert \mu - \mu_0 \rvert - \delta)^2}$
สองกลุ่มคู่ขนาน (Two – sample Parallel)	Equality	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$n_i = \frac{2\left(z_{\frac{\alpha}{2}} + z_\beta\right)^2 \sigma^2}{(\mu_1 - \mu_2)^2}$
	Non-inferiority	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$
	Superiority	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$
	Equivalence	$\lvert \mu_1 - \mu_2 \rvert \geq \delta$	$\lvert \mu_1 - \mu_2 \rvert < \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(\lvert \mu_1 - \mu_2 \rvert - \delta)^2}$
สองกลุ่มไขว้ (Two sample Crossover)	Equality	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$n_i = \frac{\left(z_{\frac{\alpha}{2}} + z_\beta\right)^2 \sigma^2}{2(\mu_1 - \mu_2)^2}$
	Non-inferiority	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{2(\mu_1 - \mu_2 - \delta)^2}$
	Superiority	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{2(\mu_1 - \mu_2 - \delta)^2}$
	Equivalence	$\lvert \mu_1 - \mu_2 \rvert \geq \delta$	$\lvert \mu_1 - \mu_2 \rvert < \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{2(\lvert \mu_1 - \mu_2 \rvert - \delta)^2}$

Sample Size Formula for Variations of CT – Categorical Outcomes

		สมมติฐานและสูตรคำนวณขนาดกลุ่มตัวอย่าง	
รูปแบบการวิจัย	สมมติฐาน	H_0	สูตรพื้นฐาน
กลุ่มเดี่ยว (One – Sample)	Equality	$\pi - \pi_0 = 0$	$n = \frac{\left(z_{\frac{\alpha}{2}} + z_\beta\right)^2 \pi(1-\pi)}{(\pi - \pi_0)^2}$
	Superiority	$\pi - \pi_0 \le \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \pi(1-\pi)}{(\pi - \pi_0 - \delta)^2}$
	Equivalence	$\lvert\pi - \pi_0\rvert \ge \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \pi(1-\pi)}{(\lvert\pi - \pi_0\rvert - \delta)^2}$
สองกลุ่มคู่ขนาน (Two – sample Parallel)	Equality	$\pi_1 - \pi_2 = 0$	$n_i = \frac{2\left(z_{\frac{\alpha}{2}} + z_\beta\right)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\pi_1 - \pi_2)^2}$
	Non-inferiority	$\pi_1 - \pi_2 \ge \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\pi_1 - \pi_2 - \delta)^2}$
	Superiority	$\pi_1 - \pi_2 \le \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\pi_1 - \pi_2 - \delta)^2}$
	Equivalence	$\lvert\pi_1 - \pi_2\rvert \ge \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\lvert\pi_1 - \pi_2\rvert - \delta)^2}$
สองกลุ่มไขว้ (Two sample Crossover)	Equality	$\pi_1 - \pi_2 = 0$	$n_i = \frac{\left(z_{\frac{\alpha}{2}} + z_\beta\right)^2 \sigma_d^2}{2(\pi_1 - \pi_2)^2}$
	Non-inferiority	$\pi_1 - \pi_2 \ge \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma_d^2}{2(\pi_1 - \pi_2 - \delta)^2}$
	Superiority	$\pi_1 - \pi_2 \le \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma_d^2}{2(\pi_1 - \pi_2 - \delta)^2}$
	Equivalence	$\lvert\pi_1 - \pi_2\rvert \ge \delta$	$n_i = \frac{(z_\alpha + z_{\beta/2})^2 \sigma_d^2}{2(\lvert\pi_1 - \pi_2\rvert - \delta)^2}$

Example:
Websites for Sample Size Calculation



Sample Size Adjustment

1. Adjusting for loss of follow up
   - If p is the proportion of subjects lost to follow-up, the number of subjects must be increased by a factor of 1/(1-p).
     - N_{adj} = N x 1/(1-p)
2. Adjusting for Non-adherence
   - Ro =drop out rate
   - Ri=drop in rate

   $Nadj = N / (1 - R_O - R_I)^2$
   - If Ro=0.20, Ri=0.05

   Nadj =1.78N

Example: Sample Size Adjustment

Statistics Guide for Research Grant Applicants: D. Sample size calculations - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media Links

Scenario: The prevalence of dysfunctional breathing amongst asthma patients being treated in general practice is to be assessed using a postal questionnaire survey (Thomas *et al.* 2001).

Required information: -

- Primary outcome variable = presence/absence of dysfunctional breathing
- 'Best guess' of expected percentage (proportion) = 30% (0.30)
- Desired width of 95% confidence interval = 10% (i.e. +/- 5%, or 25% to 35%)

The formula for the sample size for estimation of a single proportion is as follows: -

$n = 15.4 * p * (1-p)/W^2$

where n = the required sample size p = the expected proportion - here 0.30 W = width of confidence interval - here 0.10

Inserting the required information into the formula gives: -

$n = 15.4 * 0.30 * (0.70)/ 0.10^2 = 324$

324 x 1/(1-.3) = 463

Suggested description of this sample size calculation: -

"A sample of 324 patients with asthma will be required to obtain a 95% confidence interval of +/- 5% around a prevalence estimate of 30%. To allow for an expected 70% response rate to the questionnaire, a total of 480 questionnaires will be delivered."

Back to top

Done Internet

Example: Sample Size Adjustment

Statistics Guide for Research Grant Applicants: D. Sample size calculations - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media Links

Scenario: A randomised controlled trial has been planned to evaluate a brief psychological intervention in comparison to usual treatment in the reduction of suicidal ideation amongst patients presenting at hospital with deliberate self-poisoning. Suicidal ideation will be measured on the Beck scale; the standard deviation of this scale in a previous study was 7.7, and a difference of 5 points is considered to be of clinical importance. It is anticipated that around one third of patients may drop out of treatment (Guthrie *et al.* 2001)

Required information: -

- Primary outcome variable = The Beck scale for suicidal ideation. A continuous variable summarised by means.
- Standard deviation = 7.7 points
- Size of difference of clinical importance = 5 points
- Significance level = 5%
- Power = 80%
- Type of test = two-sided

The formula for the sample size for comparison of 2 means (2-sided) is as follows: -

$n = [A + B]^2 * 2 * SD^2 / DIFF^2$

where n = the sample size required in each group (double this for total sample).

SD = standard deviation, of the primary outcome variable - here 7.7.

$DIFF$ = size of difference of clinical importance - here 5.0.

A depends on desired significance level (see table) - here 1.96.

Significance level	A
5%	1.96
1%	2.58

Power	B
80%	0.84
90%	1.28
95%	1.64

Inserting the required information into the formula gives: -

$n = [1.96 + 0.84]^2 * 2 * 7.7^2 / 5.0^2 = 38$

This gives the number required in each of the trial's two groups. Therefore the total sample size is double this, i.e. 76.

To allow for the predicted dropout rate of around one third, the sample size was increased to 60 in each group, a total sample of 120.

Suggested description of this sample size calculation: -

"A sample size of 38 in each group will be sufficient to detect a difference of 5 points on the Beck scale of suicidal ideation, assuming a standard deviation of 7.7 points, a power of 80%, and a significance level of 5%. This number has been increased to 60 per group (total of 120), to allow for a predicted drop-out from treatment of around one third"

Back to top

Internet

$38 * (1 / (1-.33) = 57$

Example: Sample Size Adjustment

Randomised, double-blind, placebo-controlled study to determine whether steroids reduce the incidence and severity of nephropathy in Henoch Schonlein Purpura (HSP)

Main research questions

Do steroids reduce the incidence and severity of nephropathy in childhood HSP?

Are ACE genotype polymorphisms predictive of progressive nephropathy in HSP?

Henoch-Schonlein Purpura (HSP) is the commonest small vessel vasculitis of childhood. Long term prognosis is related to progressive renal insufficiency. There is no conclusive evidence that steroids will alter the course of the disease. We will address this. In conjunction, we will evaluate the association between insertion and deletion polymorphisms of the ACE gene and progressive nephropathy in treated and untreated groups.

Data analysis/Sample size

Formal statistical input into the study has been provided by the Research and Development Support Unit at Southmead hospital, Bristol. To test the hypothesis that treatment with prednisolone 2mg/Kg for a period of 14 days reduces the incidence of proteinuria at a set point (12 months) after initial presentation. We will require a study of 320 patients (160 in each group). This calculation is based on the premise that 15% of children in the untreated group are likely to develop proteinuria during the 12 month period, compared with 5% in the treated group. This sample size will provide 80% power for testing the hypothesis at the 5% level of statistical significance, and assumes the difference will be analysed using a continuity corrected chi-squared test. Allowing a dropout rate of 15%, 184 patients will need to be randomised to each treatment arm (prednisolone or placebo).

Randomization
Random Allocation
Random Assignment

BACK

Random allocation

- **Assures subjects have same probability of being assigned to either experimental or control groups**

- Has effect of increasing comparability of groups - Groups similar with regard to distribution of anticipated, and unanticipated, confounders in terms of -
  - all factors other than the intervention being applied
  - essentially eliminates selection bias

Randomization Process

Randomization Rules:

- **Use a procedure that really allocates randomly**.
  - 30–50% of 287 RCTs did *not* describe an appropriate randomization procedure.
- **Use a procedure that is tamperproof**.
  - 25% of 287 RCTs did *not* provide adequate concealment.

Block Randomization

- **Assures equal distribution**
- Blocks of 4: randomly arrange the order of these six possible groupings.

  TTCC TCTC TCCT CCTT CTCT CTTC
- Problem: easier to guess next assignment

Example: Randomization Process/Coding

Example: 3 Blocks for Part A1 (12 volunteers)

Random Number Sequence	Permuted Block
1	P L L L
2	L P L L
3	L L P L
4	L L L P

Randomization List		Study ID
3	L	9 11001
	L	9 11002
	P	9 11103
	L	9 11004
2	L	9 11005
	P	9 11006
	L	9 11107
	L	9 11008
4	L	9 11009
	L	9 11010
	L	9 11011
	P	9 11012

Statistician & GPO Pharmacist know the coding

Example: Randomization Process/Coding

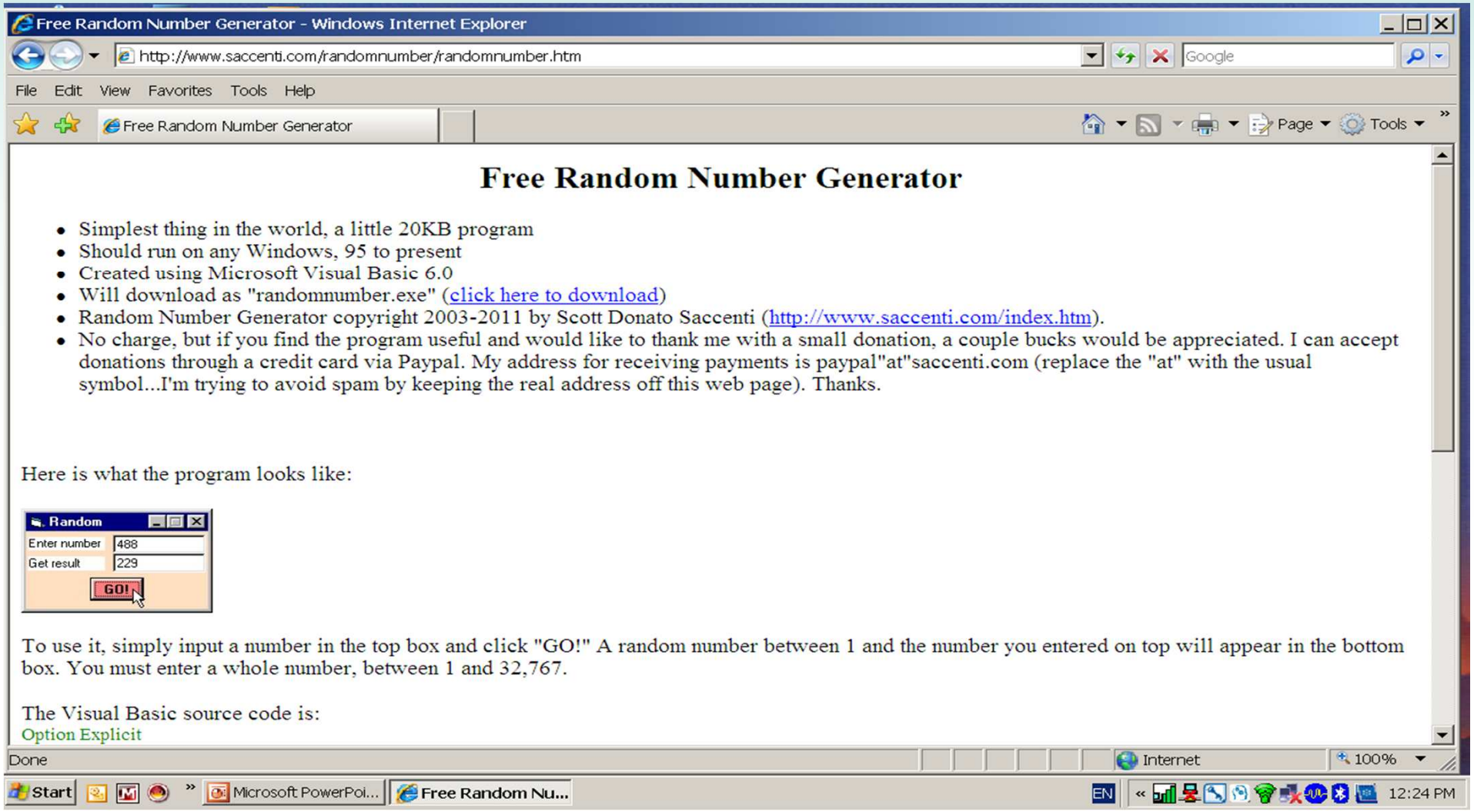
Example: 3 Blocks for Part A1 (12 volunteers)

Study ID
9 11001 9 11002 9 11103 9 11004
9 11005 9 11006 9 11107 9 11008
9 11009 9 11010 9 11011 9 11012

Coding is blinded at study site

Example:
Random number generator program

http://www.saccenti.com/randomnumber/randomnumber.htm



Example:
Random number generator program

http://www.randomizer.org/

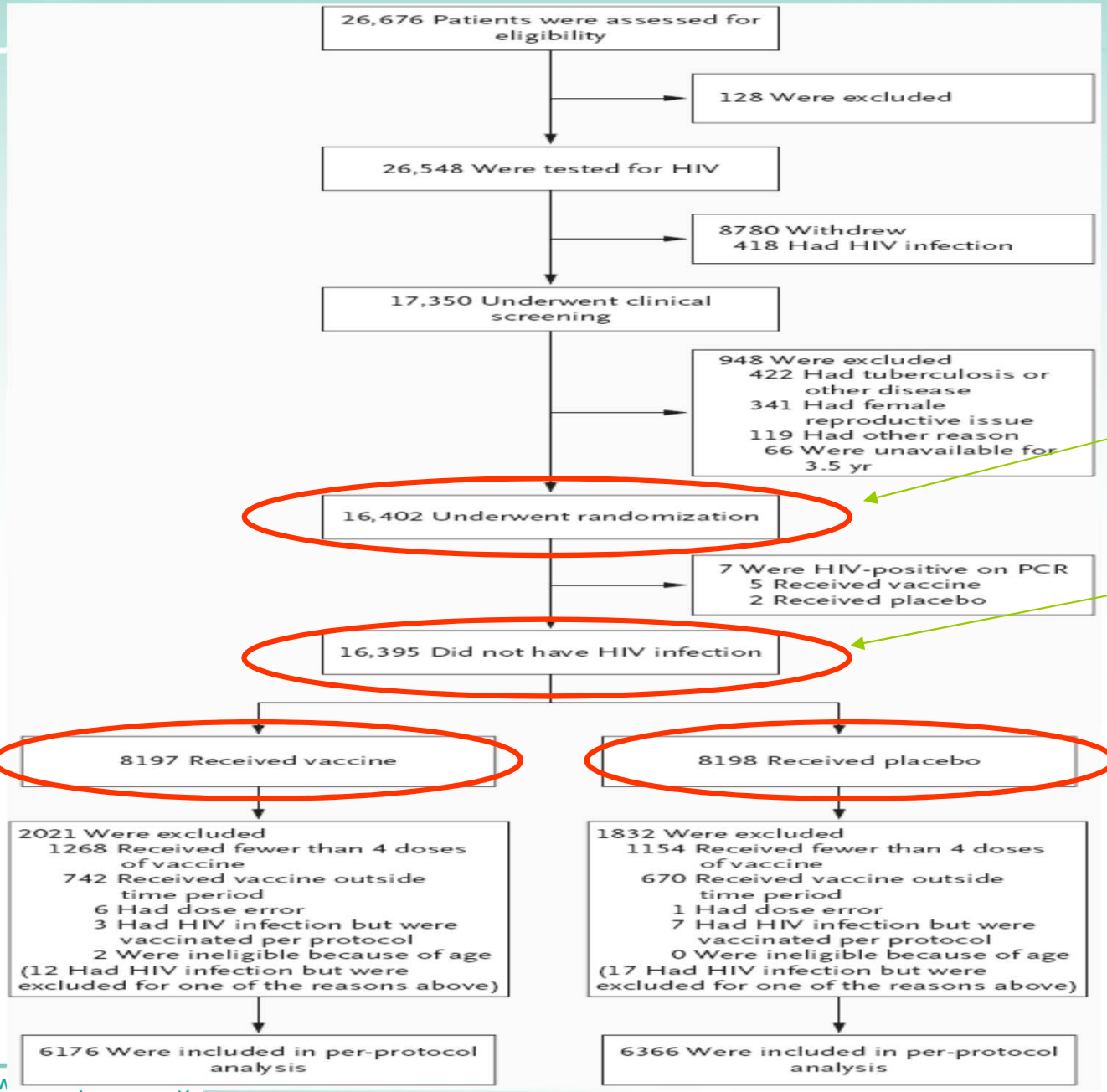
The End

BIOPHICS
Center of Excellence for Biomedical and Public Health Informatics

Sample Size Estimation & Randomization Process

Artwork © 2003 by Don Mayne. All Rights Reserved. Unauthorized Duplication Prohibited. Contact: dontoon@aol.com

Example: RV144 (2009)



The NEW ENGLAND JOURNAL of MEDICINE

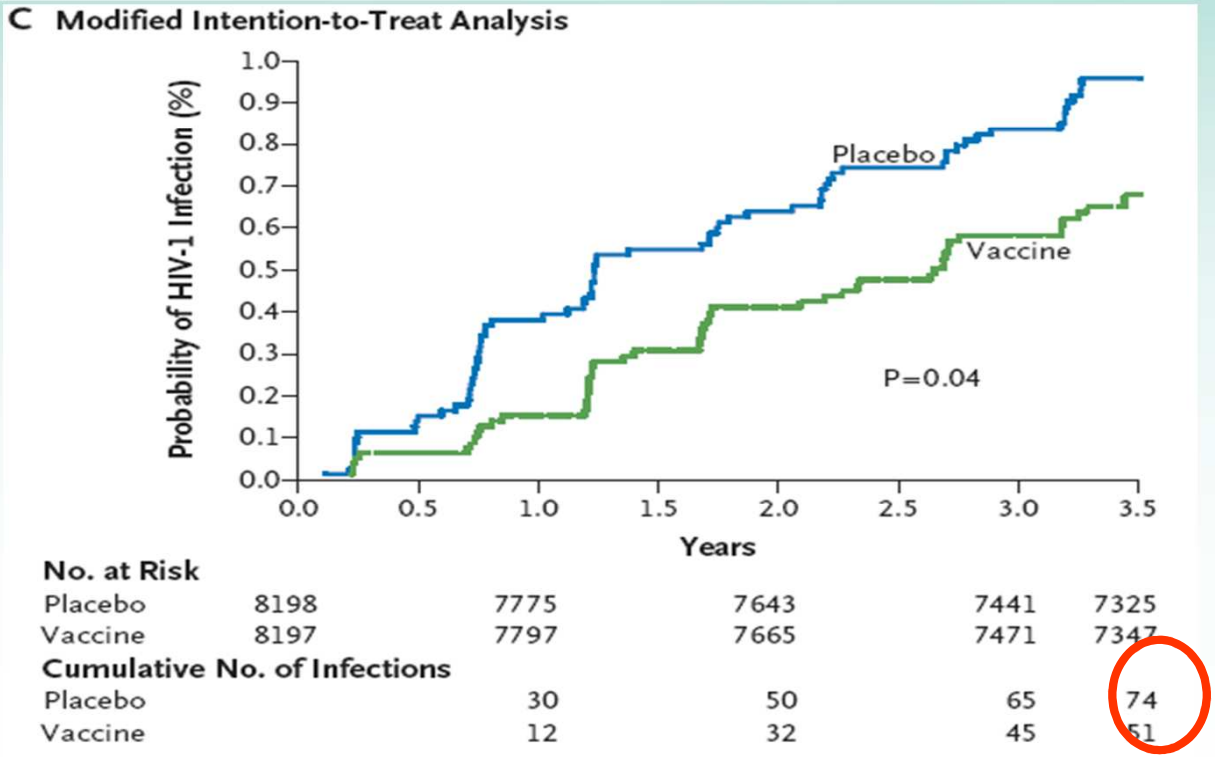
ESTABLISHED IN 1812 DECEMBER 3, 2009 VOL. 361 NO. 23

Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand

ITT Population

Modified ITT Population

Example: RV144 (2009)



The NEW ENGLAND JOURNAL of MEDICINE
ESTABLISHED IN 1812 DECEMBER 3, 2009 VOL. 361 NO. 23

Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand

ประสิทธิผล

$$\frac{0.279 - 0.192}{0.279}$$

Table 2. Rate of HIV Infection and Vaccine Efficacy, According to Selected Baseline Variables (Modified Intention-to-Treat Population).

Variable	Vaccine (N=8197)				Placebo (N=8198)				Vaccine Efficacy
	No. Evaluated	No. with Infection	No. of Person-Years	Rate	No. Evaluated	No. with Infection	No. of Person-Years	Rate	
				no./person-yr				*no./person-yr*	*% (95% CI)*
All subjects	7960	51	26,507	0.192	7988	74	26,478	0.279	31.2 (1.7 to 51.8)