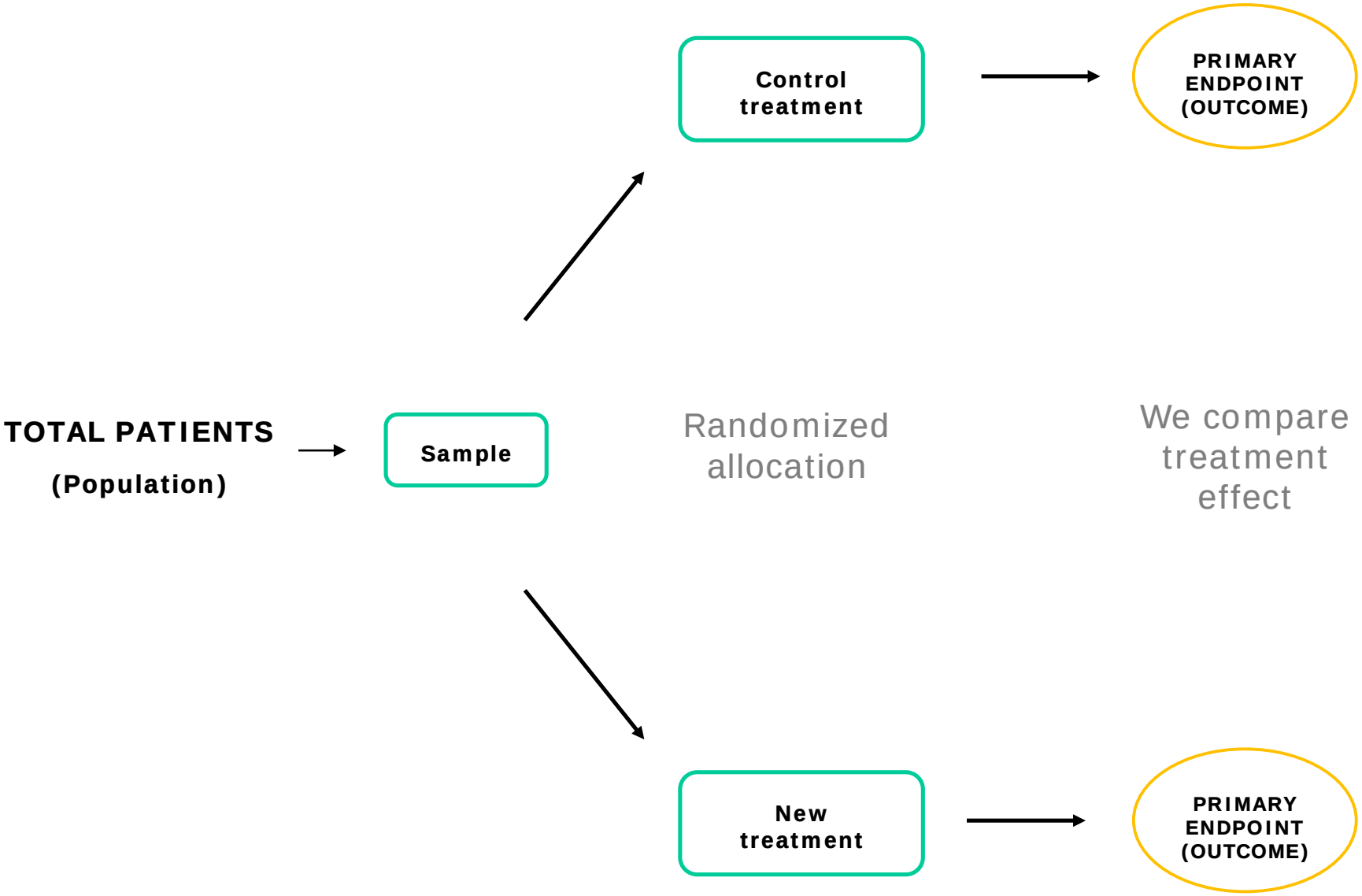
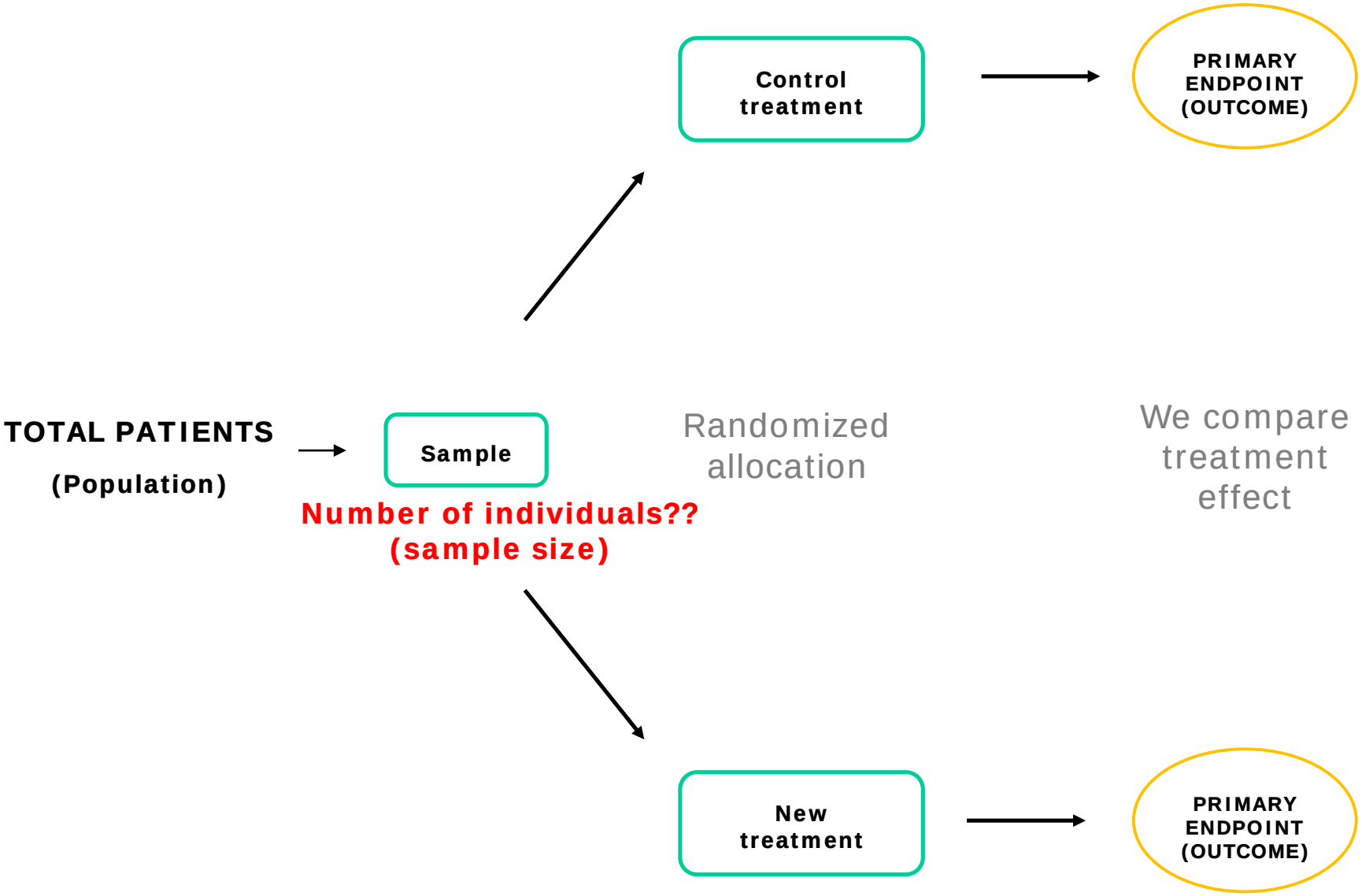
Biosciences: 3.- Health

3.2 Clinical Trials:

Sample size


Control treatment
PRIMARY ENDPOINT (OUTCOME)
TOTAL PATIENTS
(Population)
Sample
Randomized allocation
We compare treatment effect
New treatment
PRIMARY ENDPOINT (OUTCOME)


Control treatment
PRIMARY ENDPOINT (OUTCOME)
TOTAL PATIENTS
(Population)
Sample
Randomized allocation
We compare treatment effect
Number of individuals??
(sample size)
New treatment
PRIMARY ENDPOINT (OUTCOME)

Sample size to compare 2 means (independent samples)

$$\begin{cases} H_0: \ \mu_A - \mu_B = 0 \\ H_1: \ \mu_A - \mu_B = \Delta \end{cases}$$

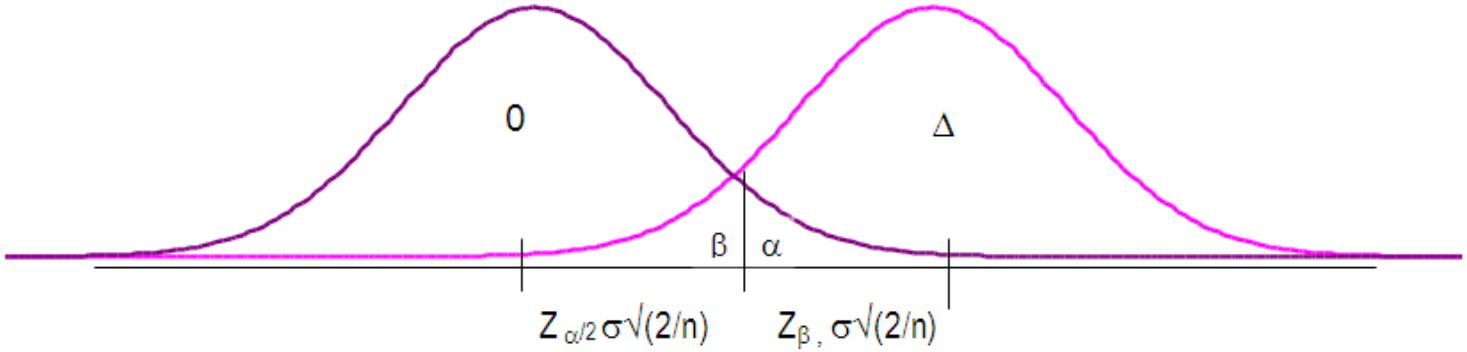
Let be: alpha risk, one or two-sided (usually $\alpha = 0'05$, two-sided),

some desired power (usually $1-\beta = 0'8$),

some given standard deviation (σ), and

some expected effect size Δ to be tested

$V(\bar{Y}_A - \bar{Y}_B) = V(\bar{Y}_A) + V(\bar{Y}_B) = \sigma^2/n_1 + \sigma^2/n_2 = 2\sigma^2/n$ (given $n_1 = n_2$)



$\Delta = Z_{\alpha/2}\ \sigma\sqrt{(2/n)} + Z_{1-\beta}\sigma\sqrt{(2/n)}$

And so, <u>each sample</u> size should be: $\mathbf{n = [\ 2\ \sigma^2\ (z_{1-\alpha/2} + z_{1-\beta})^2\ ]\ /\ \Delta^2}$ **(Bilateral)**

(Unilateral (one-side test): use $z_{1-\alpha}$)

Sample size to compare 2 means (independent samples)

Example:

Sample size to proof a gender height difference of $\Delta = 10$ cm?

Be $\sigma = 8$ cm and usual risk values ($\alpha = 0'05$; $\beta = 0'2$), two-side test.

$n = [2 \cdot 8^2 (1'96 + 0'84)^2] / 10^2 = 10'05 \approx 11$ units per sample (round **up**)

$\rightarrow N = 11x2 = 22$.

*Note: round up **n** (sample size per patient), not **N** (N must always be pair).*

Exercises:

1) What is needed to assess sample size? (What values do we need?)
2) What does Δ mean? **Desired** or expected effect size?
3) Can an observed $(\bar{Y}_A - \bar{Y}_B)$ effect smaller to Δ still being statistically significant?

Sample size to compare 2 means (independent samples)

Altman Nomogram

Overall size N
Group size n
N = 2·n

In the gender example, standardized difference 10/8=1.25

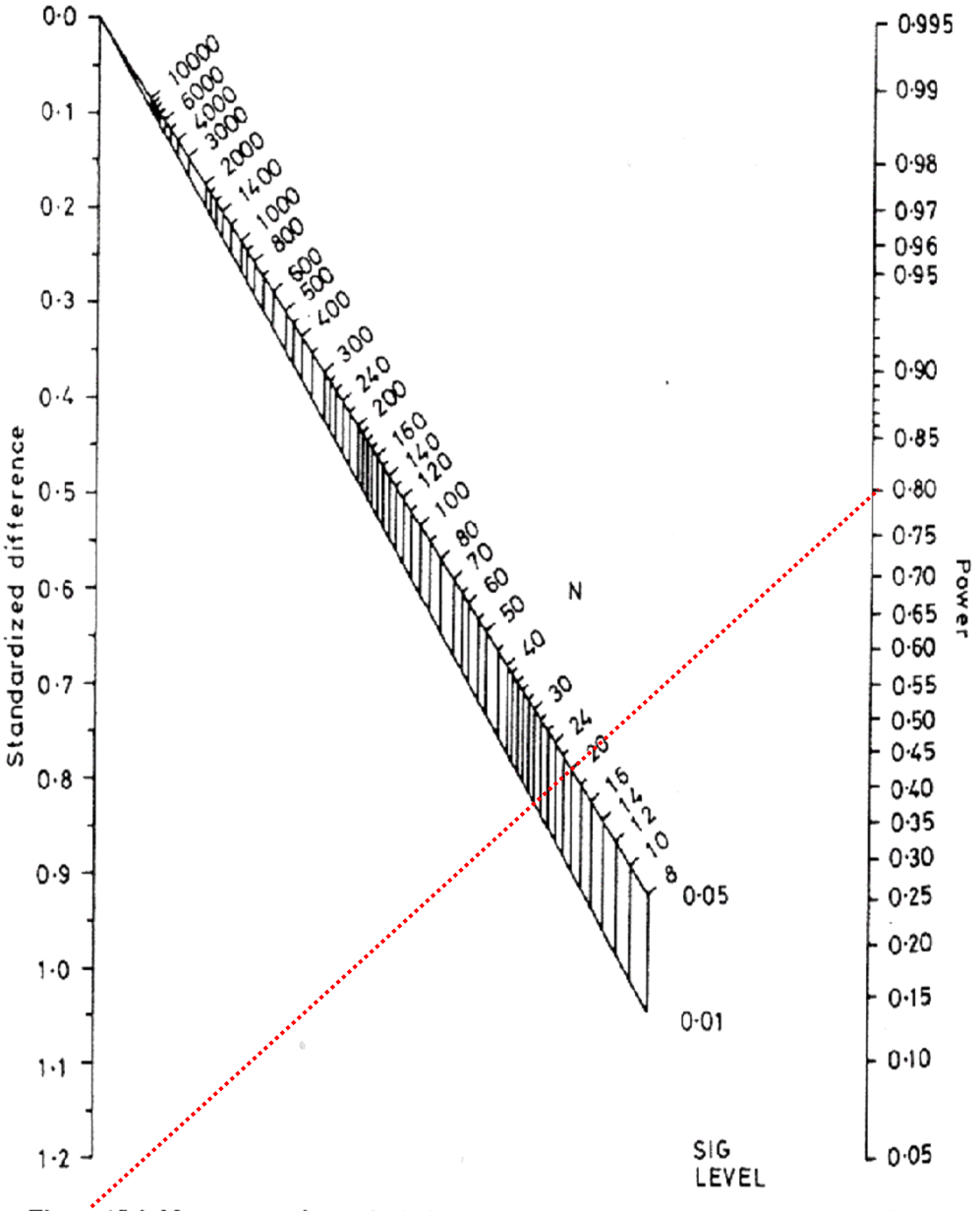


Figure 15.2 Nomogram for calculating sample size or power (reproduced from Altman, 1982b, with permission).

Exercise:

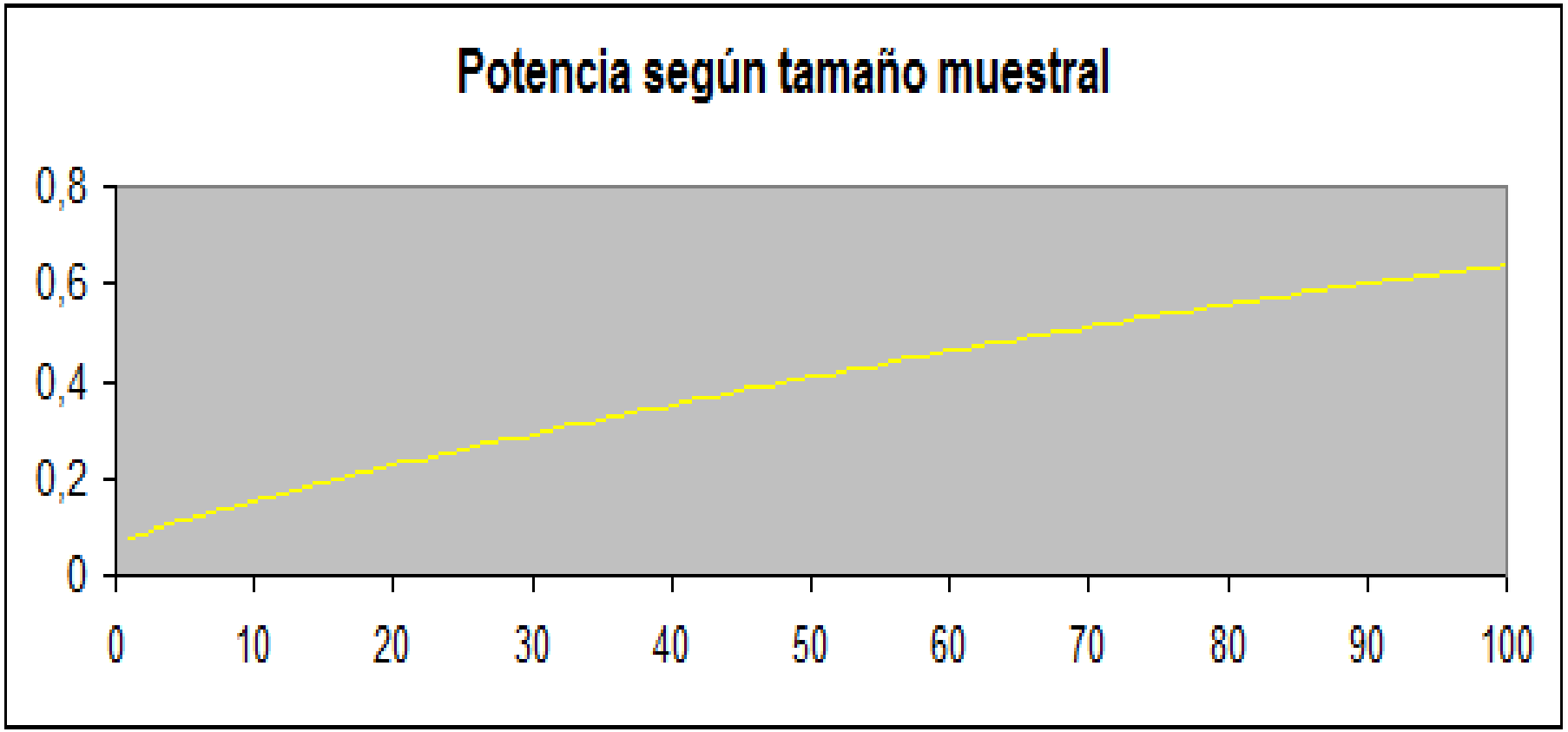
Assume that $\sigma=50$ and you need to contrast H_0: $\mu=0$ versus H_1: $\mu=10$

1. What is your power with a sample size of n=10?

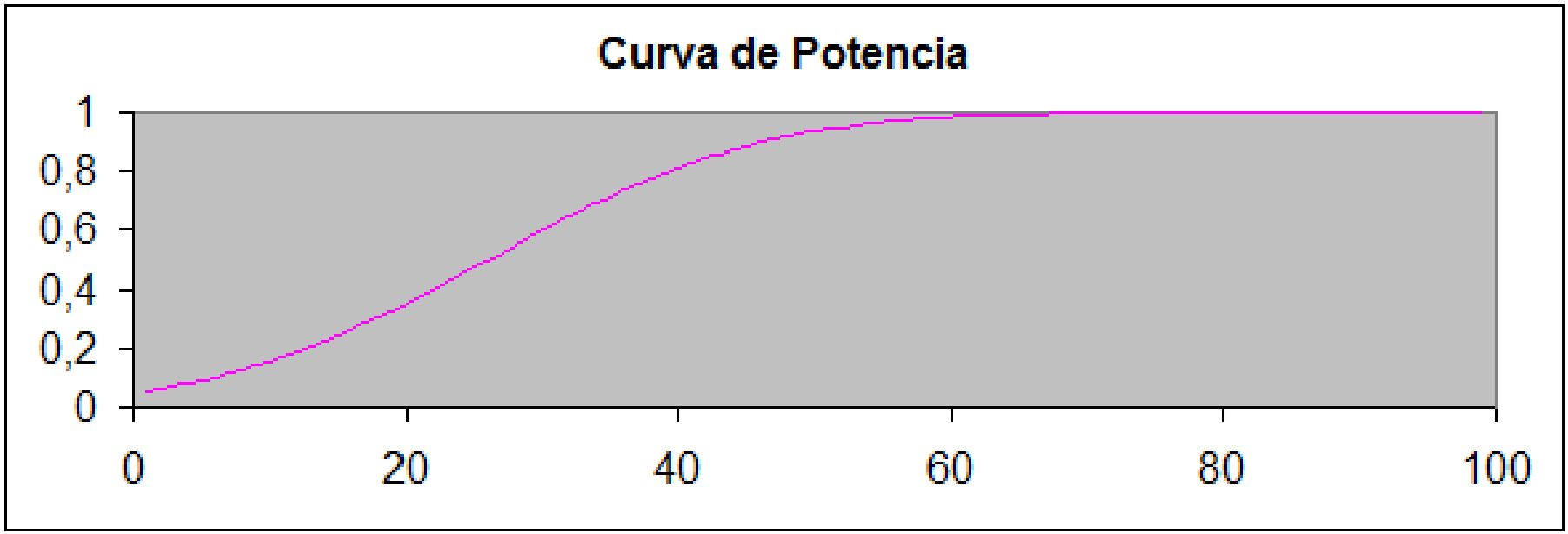
2. Draw a graph for the power (Y-axis) for
   samples from n=1 to n=100 (X-axis).
   Interpret.

3. Draw a graph for the power (Y-axis) of a sample of n=10
   for H_1: $\mu=1$ to 80 (X-axis).
   Interpret.

2.- La potencia según el tamaño ilustra cómo crece al aumentar el tamaño (n), ya que disminuye la oscilación aleatoria ($\sigma/\sqrt{n}$)



3.- Al mirar la potencia en función de Δ (magnitud de H_1) se ve el incremento de potencia que implica representa querer probar valores de Δ más alejados de H_0



OPTIMAL DESIGN

Let: overall available cases N, K treatments and 1 Placebo, number of cases in treatment k=n and in placebo=n_P, so $N = n_P + K \cdot n$

We want to optimize the K comparisons against P

Find the number of cases n to be allocated to each study arm

$V(\bar{Y}_A - \bar{Y}_B)$ is $\sigma^2 \left(\frac{1}{n} + \frac{1}{n_P}\right) \Rightarrow minimize\ f(n) = \left(\frac{1}{n} + \frac{1}{N-Kn}\right)$

$$f(n) = n^{-1} + (N - Kn)^{-1}$$

$$f'(n) = -n^{-2} + K(N - Kn)^{-2}$$

$$f'(n) = 0 \rightarrow n = \frac{(N - Kn)}{\sqrt{K}}$$

For $K=1 \rightarrow n = n_P = N/2$

Power and Sample size is related to σ

Consider a baseline, before treatment, determination Z of the outcome variable Y:

$$Z_{ij} = \mu_z + \alpha_i + \varepsilon'_{ij}$$

$$Y_{ij} = \mu_y + \tau_j + \alpha_i + \varepsilon_{ij}$$

Being:

$\alpha_i \sim N(0, \sigma_\alpha^2)$ or patient "idiosyncrasy" (among patients differences that remain constant on the two measured periods) with variance σ_α^2

$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ or any measurement error and any patient non-stability either on the baseline (ε'_{ij}) or on the outcome (ε_{ij}) measurement period

So the outcome standard deviation (σ) can be decomposed into:

$$V(Y_{ij}) = \mathrm{V}(\alpha_i) + V(\varepsilon_{ij}) = \sigma_\alpha^2 + \sigma_\varepsilon^2$$

[Sometimes called '*between*' and '*within*'.]

Power: correlation between Z and Y

If we assume that ε'_{ij}, ε_{ij} and α_i are mutually independent:

$$\rho_{Y_{ij},Z_{ij}} = \frac{Cov(Y_{ij},Z_{ij})}{\sqrt{Var(Y_{ij})}\sqrt{Var(Z_{ij})}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\varepsilon^2}$$

Exercise: proof it

And so, the correlation coefficient represents the proportion of variance among individuals to the total variance.

The correlation coefficient (amount of total variance shared by both determinations) can be interpreted as the amount of patient 'idiosyncrasy' α_i

If we further assume that the patient status doesn't change, that is, that εij is only the result of the measurement error, the correlation coefficient ρ corresponds to the reliability coefficient.

Please, note that this modelling implies that ρ should be positive.

Power: strategies (1) to diminish σ by controlling σ_α

The change or outcome-baseline difference is defined as: $D_{ij} = Y_{ij} - Z_{ij}$

and its variance is: $V(D_{ij}) = V(Y_{ij} - Z_{ij}) = \sigma^2_{\varepsilon'} + \sigma^2_{\varepsilon}$

Usually, we assume $\sigma^2_{\varepsilon'} = \sigma^2_{\varepsilon}$.

In a two-arms (X=1, treated; and X=0, control) randomized clinical trial, the treatment effect is usually estimated by:

$$\tau(b) = (\bar{Y}_1 - \bar{Y}_0) - b \cdot (\bar{Z}_1 - \bar{Z}_0)$$

Depending upon the b value, we obtain:

the "final" estimator $\hat{\tau}(0) = (\bar{Y}_1 - \bar{Y}_0)$

the "change score" $\hat{\tau}(1) = (\bar{Y}_1 - \bar{Y}_0) - (\bar{Z}_1 - \bar{Z}_0)$

or the "ANCOVA" $\hat{\tau}(\beta) = (\bar{Y}_1 - \bar{Y}_0) - \beta \cdot (\bar{Z}_1 - \bar{Z}_0)$

being β the least mean squares estimator with residual variance:

$V(e_{ij}) = (1-\rho)V(Y_i) = (1-\rho)(\sigma^2_\alpha + \sigma^2_\varepsilon)$

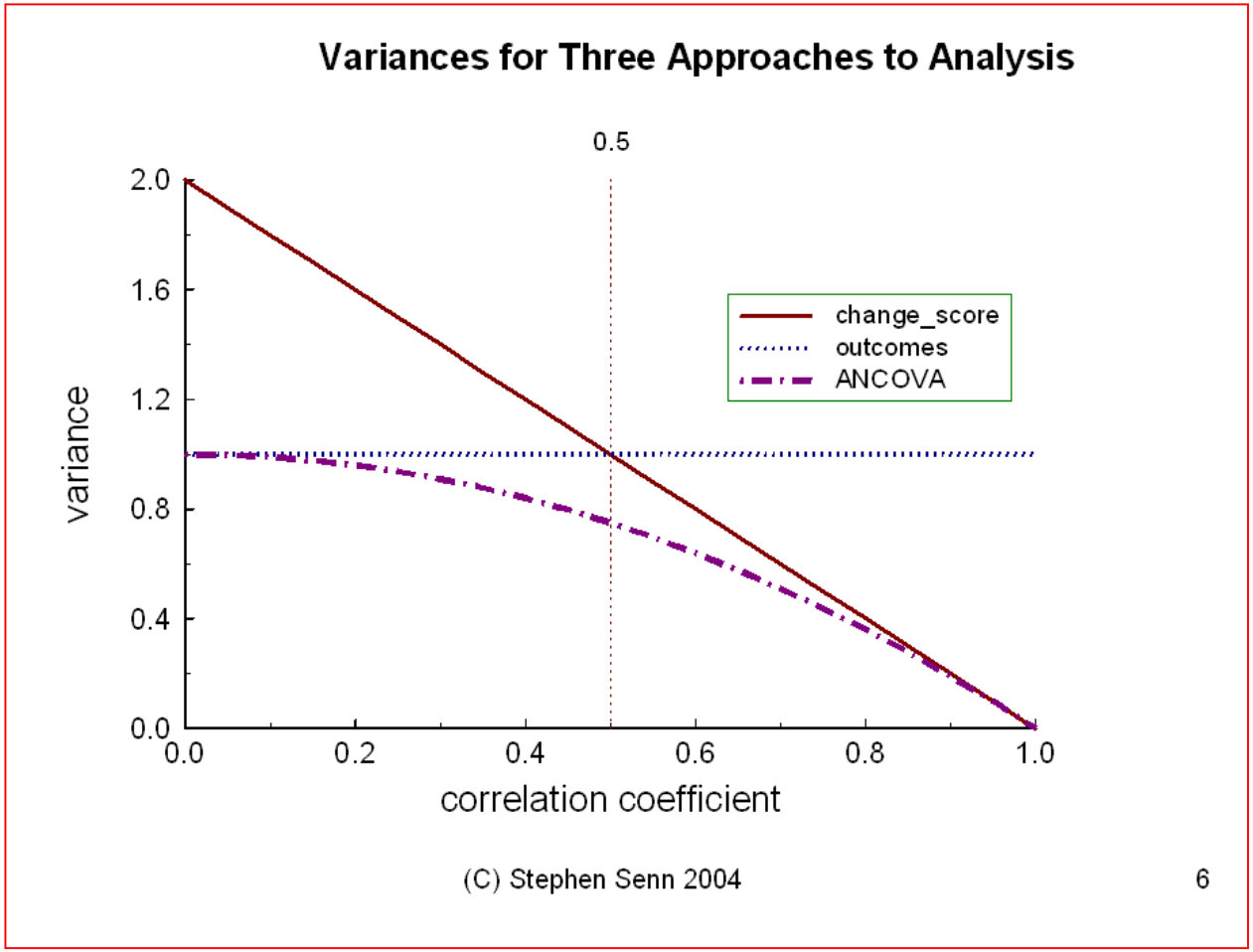
Variance of each effect estimator

$\hat{\tau}(0)$: Final outcome	$\hat{\tau}(1)$: Change from baseline	$\hat{\tau}(\beta)$: ANCOVA
$q\sigma_y^2$	$q(\sigma_y^2 + \sigma_z^2 - 2\rho\sigma_y\sigma_z)$	$q(1-\rho^2)\sigma_y^2$
	$2q(1-\rho)\sigma_y^2$ *	
$q = 1/n_0 + 1/n_1$	*Provided $\sigma_Y = \sigma_Z$	

It implies that $\hat{\tau}(\beta)$ is the most efficient estimator for any value of ρ and that $\hat{\tau}(1)$ is more efficient than $\hat{\tau}(0)$ for ρ>0.5, but less efficient in the opposite situation.

Variance of each outcome

$V(Y_{ij}) = V(\alpha_i) + V(\varepsilon_{ij}) = \sigma_\alpha^2 + \sigma_\varepsilon^2$	$V(D_{ij}) = V(Y_{ij} - Z_{ij}) = \sigma_{\varepsilon'}^2 + \sigma_\varepsilon^2$	$V(D_{ij}) = V(Y_{ij} - Z_{ij}\beta) = (1-\rho^2)\sigma_Y^2$
	(Usually, we assume $\sigma_{\varepsilon'}^2 = \sigma_\varepsilon^2$.)	, where $\sigma_Y^2 = V(Y_{ij})$


Variances for Three Approaches to Analysis
0.5
2.0
1.6
1.2
0.8
0.4
0.0
variance
change_score
outcomes
ANCOVA
0.0
0.2
0.4
0.6
0.8
1.0
correlation coefficient
(C) Stephen Senn 2004
6

Power: strategies (2) to diminish σ by controlling σ_ε

If we obtain K independent repeated measures of the outcome Y,

and we average them:

$$m(Y) = \sum_{k=1}^{k=K} Y_k / K$$

Then, the variance of m(Y) is:

$$V[m(Y)] = \frac{1}{K^2} V\left[\sum_{k=1}^{k=K} Y_k\right] = \frac{1}{K^2}\left\{V\left[\sum_{k=1}^{k=K} \alpha_{ik}\right] + V\left[\sum_{k=1}^{k=K} \varepsilon_{ik}\right]\right\}$$

$$= \frac{1}{K^2}\left\{V[K \cdot \alpha_{ik}] + V\left[\sum_{k=1}^{k=K} \varepsilon_{ik}\right]\right\} = \frac{1}{K^2}\left\{K^2\sigma_\alpha^2 + \sum_{k=1}^{k=K} \sigma_{\varepsilon k}^2\right\} = \frac{1}{K^2}\{K^2\sigma_\alpha^2 + K\sigma_\varepsilon^2\}$$

$$= \sigma_\alpha^2 + \sigma_\varepsilon^2/K$$

Furthermore: $V[m(Y) - m(Z)] = \frac{2\sigma_\varepsilon^2}{K}$ (We assume $\boldsymbol{\sigma_{\varepsilon'}^2 = \sigma_\varepsilon^2}$)

Exercise:

Systolic blood pressure (SBP) is an usual outcome variable in cardiovascular CTs.

In one particular population, with some specific measurement procedure, SDP has σ^2_α = (15mm/Hg)² and σ^2_ε = (5mm/Hg)².

Assume trial objective Δ =5mm/Hg and usual risk values (α=0'05 two-sided y β=0'2 one-sided).

Calculate outcome variance and sample size in a 2 arms CT with:

1) A single SBP determination day 8 after treatment.

2) The average of 7 repeated SBP determination from days (D) 8 to 14.

3) The change score from day 0 to day 8.

4) The ANCOVA from day 0 to day 8.

5) The change score for two averages of 7 days.

6) Another alternative could be to study both treatments in the same patients, so that each patient provide two outcomes, one for each treatment: Y_T and Y_P both averaged over 7 days.

7) Does it make sense to study the difference from baseline Z in the last design?

1) $V(Y) = VE(Y)+VI(Y) = \sigma^2_\alpha + \sigma^2_\varepsilon = 225+25 = 250$

$n = [2\sigma^2 (Z_{\alpha/2} + Z_\beta)^2]/\Delta^2 = [2 \cdot 250 (1.96 + 0.84)^2]/5^2 \approx 156.8 \rightarrow 157$ each arm

2) $m(Y) = \sum_{k=1}^{k=7} Y_k/7$; $V[m(Y)] = \sigma^2_\alpha + \frac{\sigma^2_\varepsilon}{7} = 225 + 25/7 \approx 228'57$

$n = [2 \cdot 228.57 (1.96 + 0.84)^2]/5^2 \approx 143.4 \rightarrow 144$ each arm

3) $V(Y_8 - Z_0) = 2\,\sigma^2_\varepsilon = 2 \cdot 25 = 50$

$n = [2 \cdot 50 (1.96 + 0.84)^2] / 5^2 \approx 31.4 \rightarrow 32$ each arm

4) $\rho = 225/250 = 0.9 \rightarrow V = (1 - \rho^2)\,\sigma^2_Y = (1-0.9^2) \cdot 250 \approx 47.5$

$n = [2 \cdot 47.5 (1.96 + 0.84)^2] / 5^2 \approx 29.8 \rightarrow 30$ each arm

5) $V[m_7(Y) - m_7(Z)] = 2\,\sigma^2_\varepsilon /7 = 2 \cdot 25/7 \approx 7.14$

$n = [2 \cdot 7.14 (1.96 + 0.84)^2] / 5^2 \approx 4.48 \rightarrow 5$ each arm

6) $V[m_7(Y_T) - m_7(Y_P)] = 2\,\sigma^2_\varepsilon /7 = 2 \cdot 25/7 \approx 7.14$. But now, just 1 (paired) sample:

$n = [2 \cdot 7.14 (1.96 + 0.84)^2] / 5^2 \approx 4.48 \rightarrow 5$ patients in the only arm

7) No benefit: σ^2_α is already cleared, but 1 additional baseline $\sigma^2\varepsilon$ is added.

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al*. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**(27).

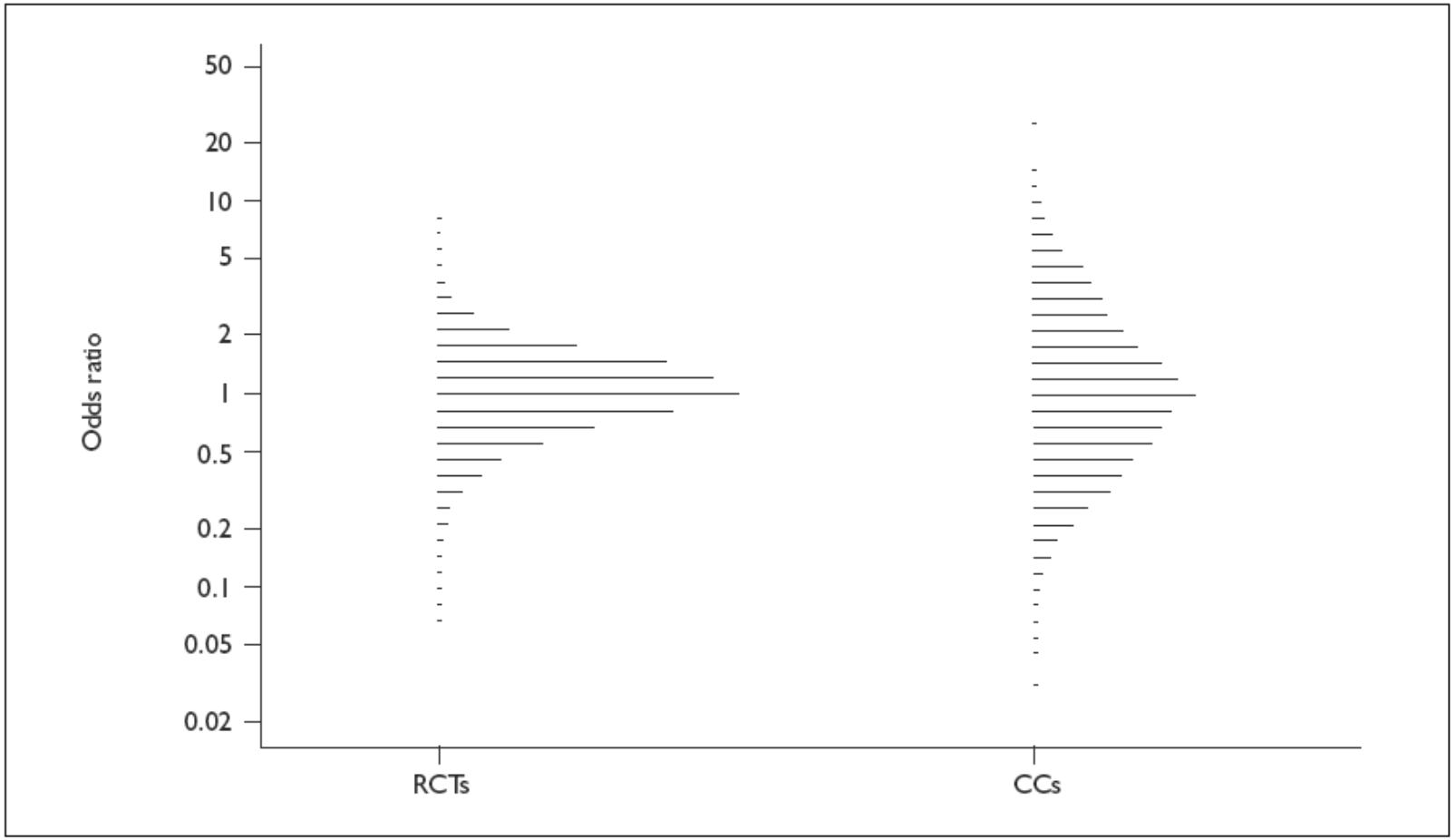


FIGURE 9 *Comparison of distributions of results of 14,000 concurrently controlled studies (CCs) and 14,000 RCTs resampled from 14 regions within the IST*

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al*. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**(27).

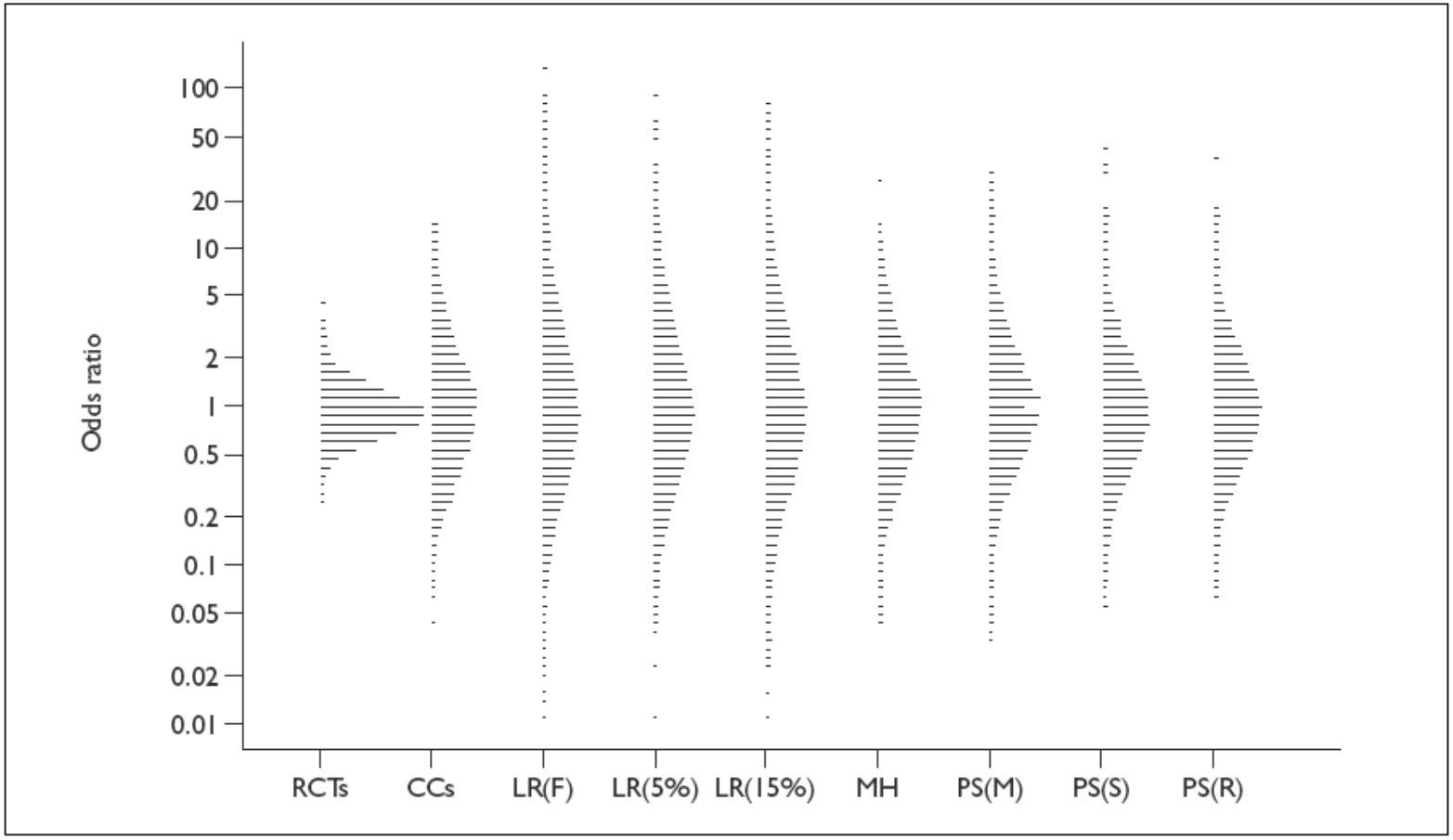


FIGURE 16 *Comparison of methods of case-mix adjustment applied to results of concurrently controlled studies resampled from 14 regions within the IST. RCT: results from corresponding randomised controlled trials; CCs: unadjusted concurrent controls without adjustment, LR(F): adjustment using full logistic regression analysis; LR(5%): adjustment with stepwise logistic regression with* $p_r = 0.05$*; LR(15%): adjustment with stepwise logistic regression with* $p_r = 0.15$*; MH: adjustment by Mantel–Haenszel stratification; PS(M): adjustment by matching on propensity score; PS(S): adjustment by stratification on propensity score; PS(R): regression adjustment based on propensity score.*


PARA SABER CÓMO ESTÁ DE SAL, ¿COJO UNA CUCHARA MUY GRANDE O REMUEVO LA SOPA ANTES?