Power and sample size

Objectives

- Explain why sample size is important
- Explain what makes up a sample size calculation
- Demonstrate sample size calculations
- Explain the level of understanding that you need for the FCEM
- Hope that you retain some/any of this information once you leave the room

FCEM – what do you need?

- Any comparative or diagnostic study needs a power calculation
- The number of study participants must reach this number
- If it doesn't ask yourself why not
  - incorrect assumptions?
  - untoward events?

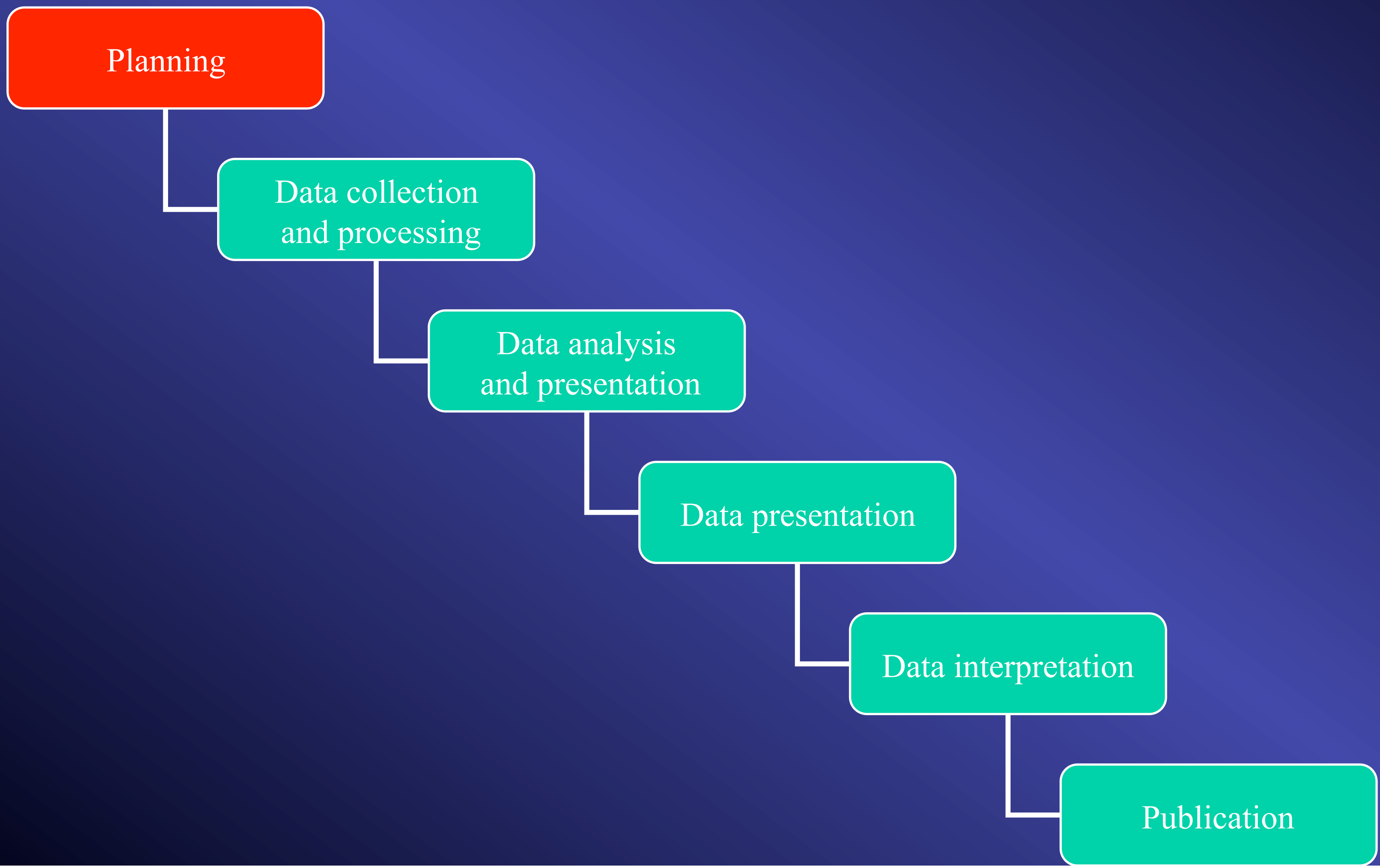
Altman

"...a trial should be big enough to have a high chance of detecting, as statistically significant, a worthwhile effect if it exists and thus be reasonably sure that no benefit exists if it is not found.."

Why is sample size important?

1. Need to get *an* answer
2. Need to get the *right* answer
3. Need to be *sure* we get the right answer
4. Avoid exposing *too many* participants than is necessary to get the answer

Study process



Hypotheses

- Research looks to answer a hypothesis
- Hypotheses are (statistically) easier to "prove" if they start from a null hypothesis:
  - "there is no difference between treatment A and treatment B in treating X"
  - "there is no difference between test C and test D in diagnosing Y"

Answers

- Four ways to get an answer:
  - Correct answer for the correct reason
  - Correct answer for the wrong reason
  - Incorrect answer for the correct reason
  - Incorrect answer for the wrong reason

Truth table

		Truth	
		There is a difference	There is no difference
Null Hypothesis	Rejected i.e. a difference was found	☺	☹ **Type I error**
	Accepted i.e. no difference was found	☹ **Type II error**	☺

Truth table

		Truth	
		There is a difference	There is no difference
Null Hypothesis	Rejected i.e. a difference was found	☺	☹ **Type I error**
	Accepted i.e. no difference was found	☹ **Type II error**	☺

Generic sample size equation

$$n \sim \frac{2(z1-\alpha/2 + z1-\beta)2\ \sigma 2}{\Delta 2}$$

Sample size

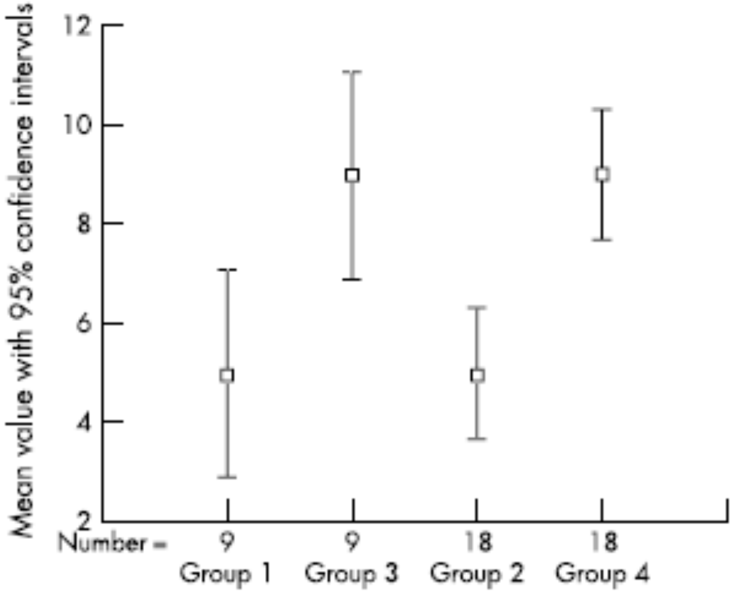
$$n \sim \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{\Delta^2}$$

- n is the sample size
- α is the significance level - often set at 0.05 so we accept a 5% chance of making a type I error
- 1-β is the power – often set at 0.8 so we accept an 80% chance of avoiding a type II error
- Δ is the effect size
- σ is the variance within the population

Factors affecting sample size

- The precision and variance of measures within any sample
- Magnitude of a clinically significant difference
- How certain we want to be of avoiding a type I error
- The type of statistical test we are performing

Precision and variance



Frequency table

Group 1	Group 2	Group 3	Group 4
1	1	5	5
2	2	6	6
3	3	7	7
4	4	8	8
5	5	9	9
6	6	10	10
7	7	11	11
8	8	12	12
9	9	13	13
	1		5
	2		6
	3		7
	4		8
	5		9
	6		10
	7		11
	8		12
	9		13

Clinically significant difference

- Very small differences require very precise estimates of the true population values
- But is it *clinically* important?
- At great effort we could demonstrate a 2mmHg difference in blood pressure between two drugs
- But is it *clinically* important?

Standardised difference

- Based upon the ratio of the difference of interest to the standard deviation of those observations
- Calculated in a different way depending on whether the data is continuous or categorical

Continuous data

$$\text{Standardised difference} = \frac{\text{difference between the means}}{\text{population standard deviation}}$$

- So if we were assessing an antihypertensive and wanted a 10mm difference between the drugs and the population standard deviation was 20mm then the standardised difference would be 0.5

Categorical data

$$\text{Standardised difference} = \frac{p_1 - p_2}{\sqrt{[P(1 - P)]}}$$

- P_1 is the baseline mortality
- P_2 is the “new” mortality we expect
- P is $0.5(P_1 + P_2)$

Standardised difference and power

Table 3 How power changes with standardised difference

Sdiff	Power level (pβ) 0.99	0.95	0.90	0.80
0.10	3676	2600	2103	1571
0.20	920	651	527	394
0.30	410	290	235	176
0.40	231	164	133	100
0.50	148	105	86	64
0.60	104	74	60	45
0.70	76	54	44	33
0.80	59	42	34	26
0.90	47	34	27	21
1.00	38	27	22	17
1.10	32	23	19	14
1.20	27	20	16	12
1.30	23	17	14	11
1.40	20	15	12	9
1.50	18	13	11	8

Sdiff, standardised difference.

Gore and Altman nomogram

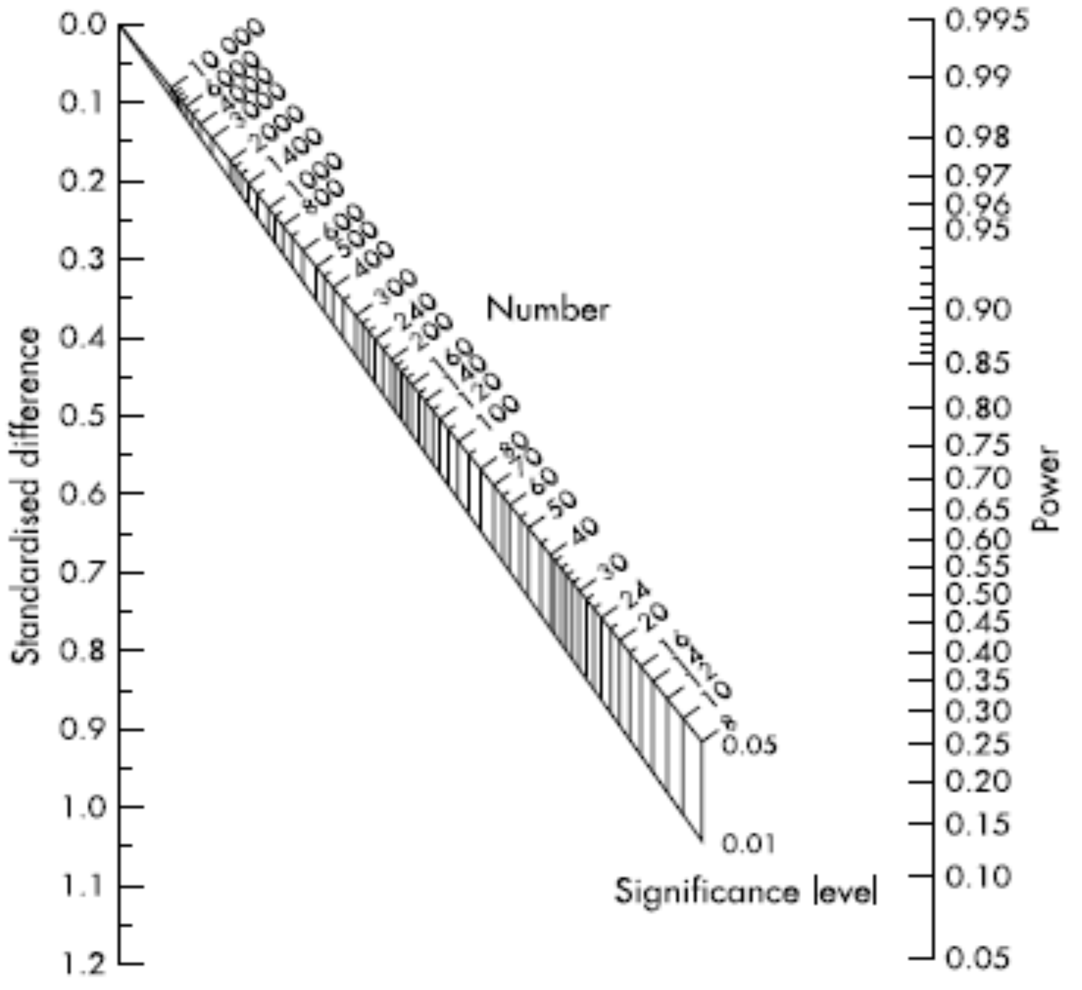


Figure 3 Nomogram for the calculation of sample size.

Gore and Altman nomogram

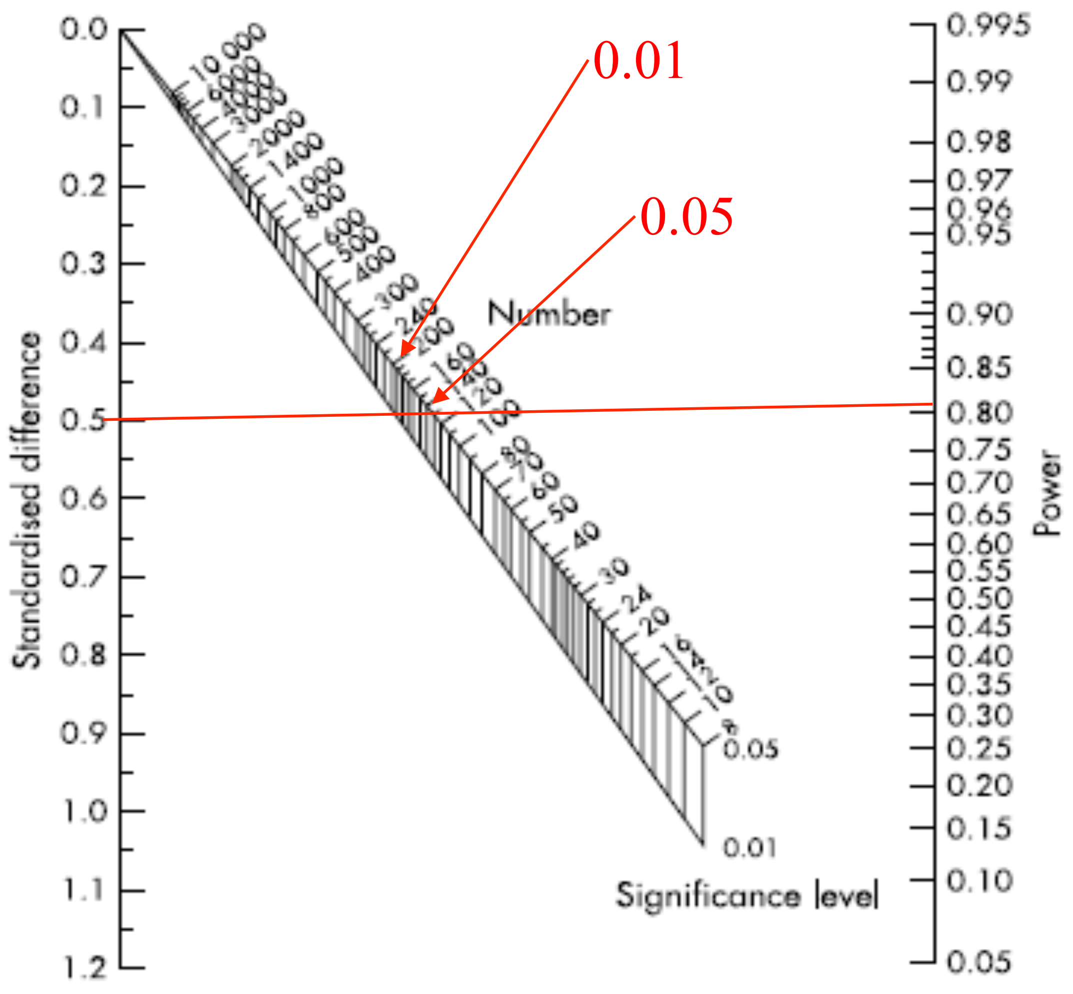


Figure 3 Nomogram for the calculation of sample size.

Diagnostic studies - sensitivity

$$\mathrm{TP} + \mathrm{FN} = z^2 \times \frac{(\mathrm{SN}\,(1 - \mathrm{SN}))}{W^2}$$

$$= 1.96^2 \times \frac{(0.95(1 - 0.95))}{0.05^2}$$

$$= 3.842 \times \frac{0.0475}{0.0025}$$

$$N(sN) = \frac{\mathrm{TP} + \mathrm{FN}}{P} \qquad = \frac{72.998}{0.3}$$

TP = true positive rate, FN = false negative rate,

SN = sensitivity, P = prevalence

Diagnostic studies - specificity

$$FP + TN = z^2 \times \frac{(SP\,(1 - SP))}{W^2}$$

$$= 1.96^2 \times \frac{(0.80\,(1 - 0.80))}{0.05^2}$$

$$= 3.842 \times \frac{0.16}{0.0025}$$

$$N\,(sp) = \frac{FP + TN}{(1 - P)} \quad = \frac{245.888}{(1 - 0.3)}$$

FP = false positive rate, TN = true negative rate

SP = specificity, P = prevalence

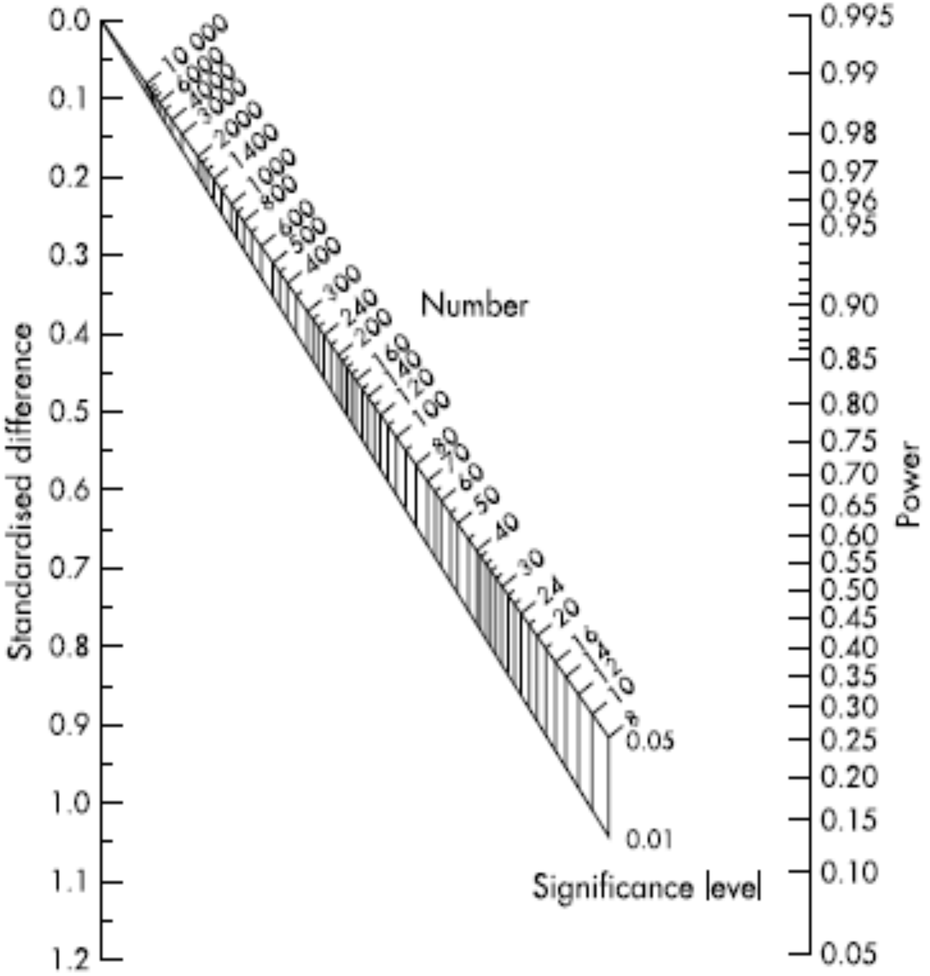


Figure 3 Nomogram for the calculation of sample size.

Table 3 How power changes with standardised difference

	Power level (pβ)			
Sdiff	0.99	0.95	0.90	0.80
0.10	3676	2600	2103	1571
0.20	920	651	527	394
0.30	410	290	235	176
0.40	231	164	133	100
0.50	148	105	86	64
0.60	104	74	60	45
0.70	76	54	44	33
0.80	59	42	34	26
0.90	47	34	27	21
1.00	38	27	22	17
1.10	32	23	19	14
1.20	27	20	16	12
1.30	23	17	14	11
1.40	20	15	12	9
1.50	18	13	11	8

Sdiff, standardised difference.

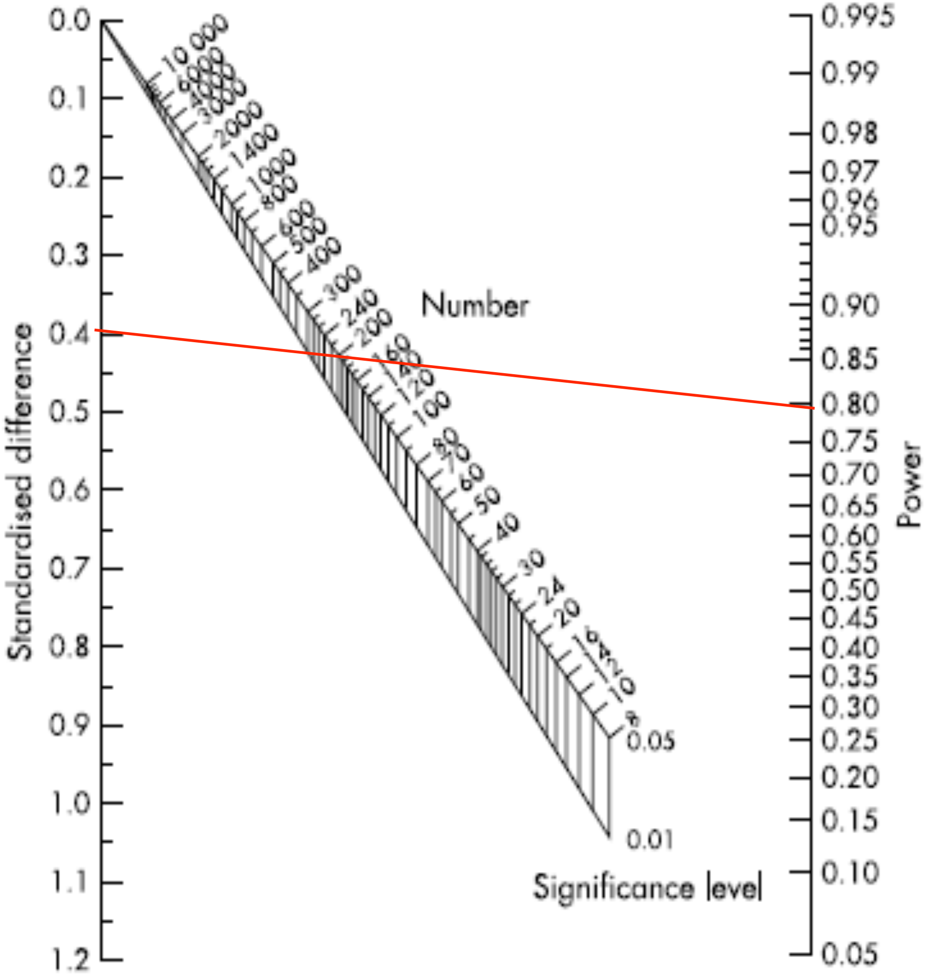


Figure 3 Nomogram for the calculation of sample size.

Table 3 How power changes with standardised difference

Sdiff	Power level (pβ) 0.99	0.95	0.90	0.80
0.10	3676	2600	2103	1571
0.20	920	651	527	394
0.30	410	290	235	176
0.40	231	164	133	100
0.50	148	105	86	64
0.60	104	74	60	45
0.70	76	54	44	33
0.80	59	42	34	26
0.90	47	34	27	21
1.00	38	27	22	17
1.10	32	23	19	14
1.20	27	20	16	12
1.30	23	17	14	11
1.40	20	15	12	9
1.50	18	13	11	8

Sdiff, standardised difference.

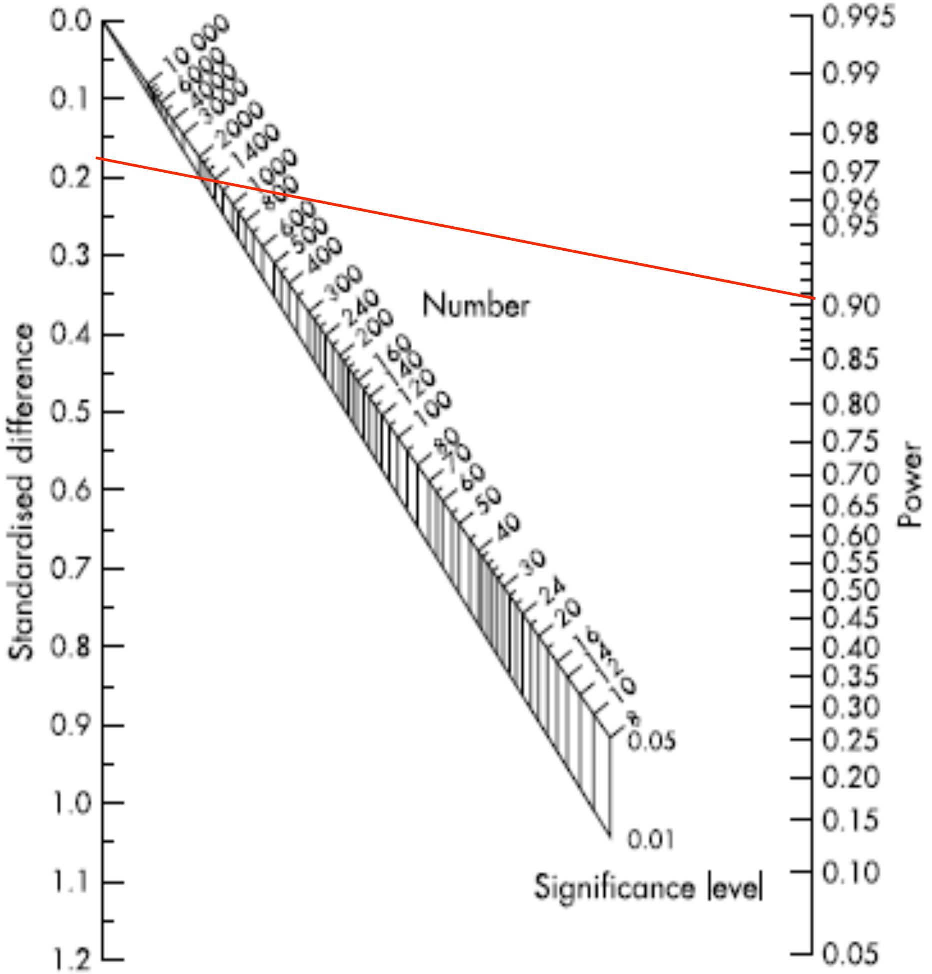
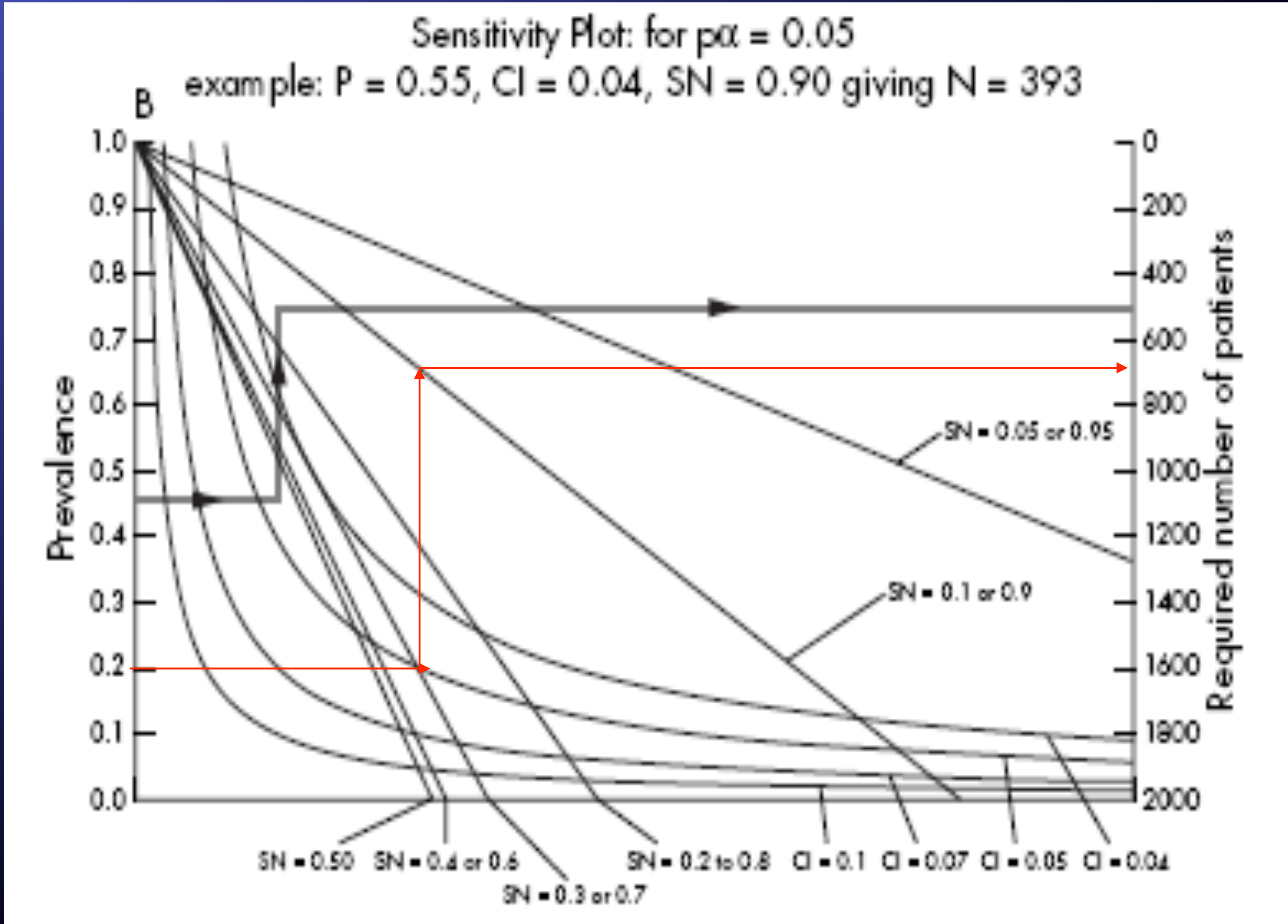


Figure 3 Nomogram for the calculation of sample size.

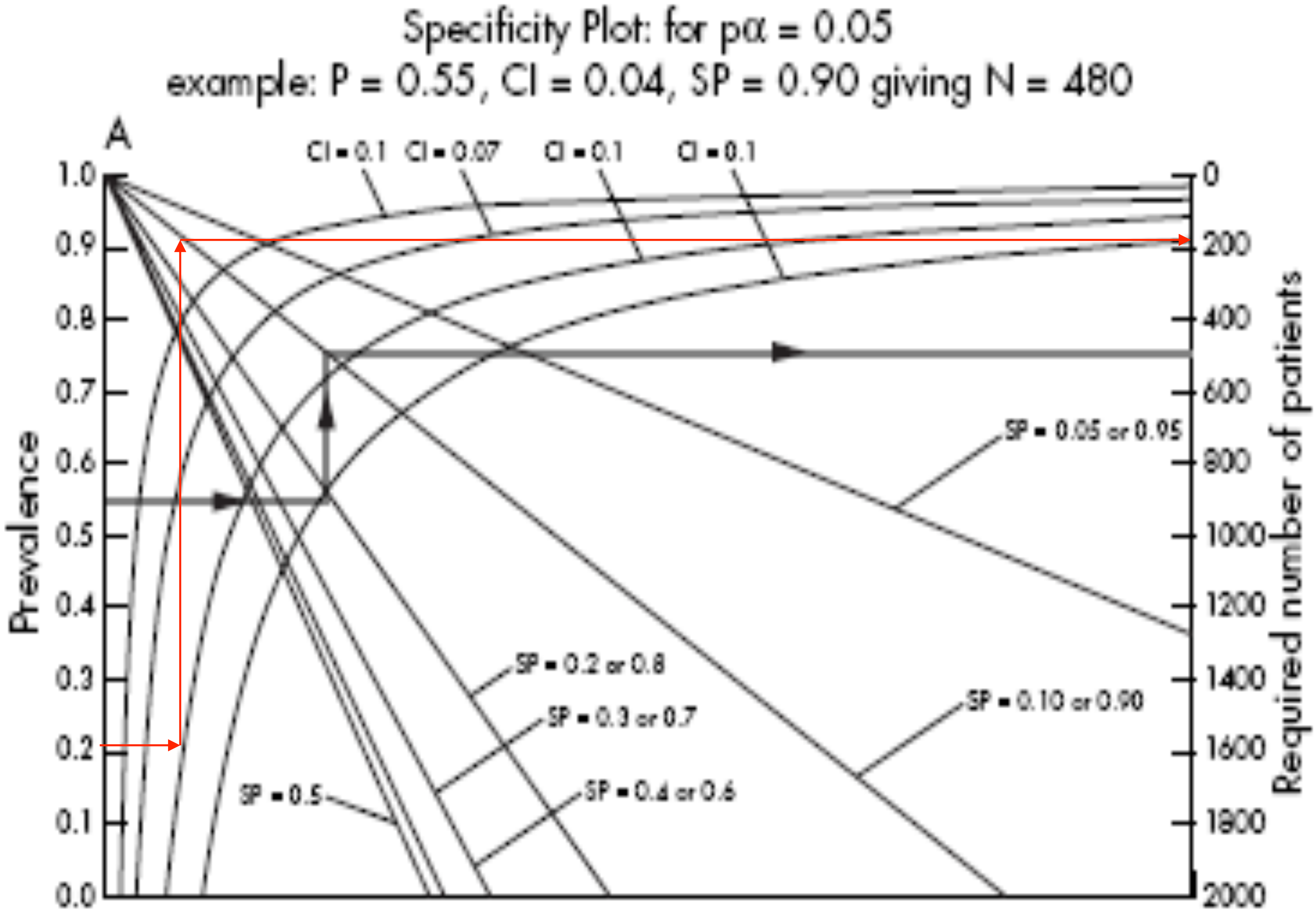
Table 3 How power changes with standardised difference

Sdiff	Power level (pβ) 0.99	0.95	0.90	0.80
0.10	3676	2600	2103	1571
0.20	920	651	527	394
0.30	410	290	235	176
0.40	231	164	133	100
0.50	148	105	86	64
0.60	104	74	60	45
0.70	76	54	44	33
0.80	59	42	34	26
0.90	47	34	27	21
1.00	38	27	22	17
1.10	32	23	19	14
1.20	27	20	16	12
1.30	23	17	14	11
1.40	20	15	12	9
1.50	18	13	11	8

Sdiff, standardised difference.



Sensitivity Plot: for pα = 0.05
example: P = 0.55, CI = 0.04, SN = 0.90 giving N = 393
B
Prevalence
Required number of patients
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
SN = 0.05 or 0.95
SN = 0.1 or 0.9
SN = 0.50
SN = 0.4 or 0.6
SN = 0.3 or 0.7
SN = 0.2 to 0.8
CI = 0.1
CI = 0.07
CI = 0.05
CI = 0.04



Specificity Plot: for pα = 0.05
example: P = 0.55, CI = 0.04, SP = 0.90 giving N = 480
A
CI = 0.1
CI = 0.07
CI = 0.1
CI = 0.1
SP = 0.05 or 0.95
SP = 0.2 or 0.8
SP = 0.3 or 0.7
SP = 0.4 or 0.6
SP = 0.5
SP = 0.10 or 0.90
Prevalence
Required number of patients
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

Questions?

FCEM – what do you need?

- Any comparative or diagnostic study needs a power calculation
- The number of study participants must reach this number
- If it doesn't ask yourself why not
  - incorrect assumptions?
  - untoward events?