# Statistics in Medicine

**Statistical advice supplied by**

**Annie Herbert**

**Formerly, Medical Statistician**

**Pennine Acute Hospitals NHS Trust**

# Outline

- Population and sample
- Randomisation
- Study example
- Average values
- Distributions
- Variability
- Sample size
- P-values
- Confidence intervals
- Types of data
- Statistical tests
- Reporting results
- Methods of analysis
- Multiplicity

# 1601 James Lancaster

- 4 ships sailing from India to England
- 1 ship issues 3 teaspoonfuls of lemon juice each day to its crew
- On three ships 40% mortality rate due to scurvy
- On Lancaster's ship 0% mortality.
- The result?

# 1747 James Lind

- HMS Salisbury
- Six arm RCT to test efficacy of cider, elixir of vitriol, vinegar, sea water, oranges and lemons, and a purgative mixture on scorbutic sailors
- Only the patients in the citrus fruit arm recovered
- The result?

# Widespread dissemination and take up…

- 1795 (48 years post Lind) Royal Navy orders citrus fruit to become part of sailors' diet. Scurvy disappears overnight
- 1865 (118 years post Lind) Board of Trade makes the same decision
- Elapsed time from Lancaster's original discovery to universal application: 264 years

# ‘Population’ & ‘Sample’

- Studying population of *interest.* Usually would like to know typical value and spread of outcome measure in population.
- Data from entire population usually impossible or inefficient/expensive so take a sample
  (even census data can have missing values).
- Sample must be *representative* of population.

# Sampling methods

- Opportunity/convenience samples: bales of jute; Hill and Doll cohort study on 50,000 medical doctors linking smoking with cancer
- Judgement samples
- Random sampling

# What happens when you don’t randomise properly?

# Types of randomisation

Simple random allocation

Block random allocation

Systematic allocation

Cluster random allocation

# Random allocation I

A randomised controlled trial investigated whether differences exist between alternating pressure overlays and alternating pressure mattresses in the development of new pressure ulcers, healing of existing pressure ulcers, and patient acceptability. The trial recruited 1972 patients admitted to hospital for elective surgery. After giving informed consent patients were randomly assigned to the alternating pressure overlays or the alternating pressure mattresses group.

## Which one of the following statements defines the process of random allocation used in the study?

(a) for every patient allocated to the alternating pressure overlays group a patient admitted for the same procedure was allocated to the alternating pressure mattresses group

(b) patients were allocated alternately to each group

(c) patients were allocated to the clinician's group of choice

(d) patients had an equal chance of being allocated to either group

Answer d is correct

# Random allocation II

A randomised single blind, controlled trial investigated the effectiveness of a home based early intervention on children's body mass index (BMI) at age two. Participating mothers were recruited in batches of fifty and allocated to one of two groups in a 1:1 allocation ratio by a computer generated random number program.

**Which one of the following statements best describes the method by which patients were allocated to an intervention?**

a) Simple random allocation

b) Block random allocation

c) Systematic allocation

d) Cluster random allocation

Answer b is correct

# Allocation concealment

- Designed to ensure that the randomisation process is not subverted by study staff
- Frequently attempted using sealed envelopes, but method is not infallible
- Failure to conceal properly typically results in a 41% overestimate of benefit[1]

[1] Schulz et al Empirical evidence of bias: dimensions of methodological quality associated with estimates of effects in controlled trials. *JAMA* 1995; 273: 408-12

# Blinding

- Designed to prevent researchers (single blinded) or researchers and participants (double blinded) knowing which treatment a patient is receiving
- Prevents assessment bias by either group, (and encourages patients not to drop out!)
- In trials with assessors failure to blind typically exaggerates treatment effect by 36%[1]

[1] Hrobjartsson et al. Observer bias in randomised controlled trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ. 2012 344: e1119

# Blinding and placebo

- Placebo treatments often fail to blind
- Difficult to obatin dummy pills that look the same as genuine ones (Pharmaceutical companies often charge exorbitantly for these)
- Genuine pills are often accompanied with unmistakable side effects (e.g. dry mouth in antidepressants)

# Allocation concealment versus blinding

A randomised controlled trial was designed to evaluate if a specific exercise strategy, targeting the rotator cuff and scapula stabilisers, improves shoulder function and pain more than unspecific exercises in patients with subacromial impingement syndrome, thereby decreasing the need for arthroscopic subacromial decompression.

An independent physiotherapist prepared the random allocation sequence beforehand. Equal numbers of the two treatment alternatives, 55 of each, were prepared and concealed in opaque envelopes. These were then mixed by hand and numbered. Treatment allocation was performed at the first visit to the physiotherapist. The research physiotherapist received the envelope with the corresponding code revealing the assigned treatment alternative out of a central locked location just before the participants presented for one of the two treatments: specific exercises (specific exercise group) or unspecific exercises (control exercise group).

# Which of the following statements, if any, are true?

(a) Allocation concealment ensured that the sequence in which patients would be allocated to treatment was not disclosed before random allocation

(b) Allocation concealment minimised allocation bias

(c) Allocation concealment meant that patients were blind to which treatment they had been allocated after random allocation

(d) Allocation concealment ensured that the trial was double blind

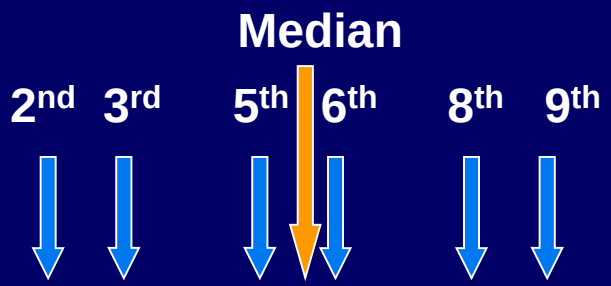
Statements *a* and *b* are true, whereas *c* and *d* are false.

# Study example – Does dolphin therapy help reduce depression more than standard care?

- BMJ – Antonioli & Reveley, 2005;331:1231 (26 Nov)
- A selection of patients with mild to moderate depression recruited via newspaper and radio were randomly allocated to either sessions of dolphin therapy or an outdoor nature snorkelling programme.
- Effectiveness of the dolphin therapy/nature programme were assessed using the change in Hamilton depression score.

# ‘Average’ values

H.D. scores in Dolphin group: 10, 8, 7, 7, 1, 7, 6, 5, 3, 4




Ordered: 1, 3, 4, 5, 6, 7, 7, 7, 8, 10

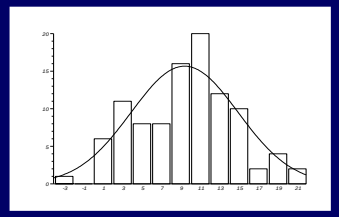
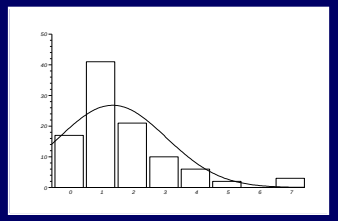
Mean = (1 + 3 + 4 + … + 10) ÷ 10 = 5.8

Median = (6+7) ÷ 2 = 6.5

Mode = 7

Range = 10-1

# Symmetric data

- Also known as ‘Normally distributed’.
- When put in a histogram :  not: 
- Useful for deciding between a mean or median and which statistical tests to use (‘parametric’/ ‘non-parametric’).
- Can be achieved using transformations.

# Abdelbaset al-Megrahi (convicted for the Lockerbie bomb). Barbara Moss (Avastin miracle survivor of bowel cancer). How do we get it so wrong?

# Variability

Standard deviation (SD):

- Indicates the spread of the data around the mean. Should only be quoted for normally distributed data
- E.g., Mean and SD of Hamilton depression score in dolphin group = 15.8 (3) implies that a typical score for a person in that group would be 15.8, but about two-thirds of the values fall between 12.8 and 18.8
- More precisely, whilst about 68% of values fall within one SD of the mean, about 95% of values fall within two SDs of the mean. In this example that means that 95% of the values fall between 9.8 and 21.8

# Sample Size

- ‘Power’ of study: Probability of finding a difference (of a given size) if it exists. Also written as ‘(1-ß)’.
- Paper should include how this sample size/power calculation performed

  (e.g. what variation was assumed)
- Is the given size of difference realistic?
- Has drop-out been accounted for?

# Sample size calculations

$n = 2 [(z_1+z_2) \times (\sigma/\delta)]^2$ where

n = sample size needed (for each group)

$z_1$ is a constant dependent on the desired P value. To reach a P value of 5% $z_1=1.96$

$z_2$ is a constant dependent on the desired power. To reach a power of 90% $z_2=1.28$

σ (sigma)=standard deviation of the measured variable

δ (delta)= clinically relevant effect size (this is a clinical decision, not a statistical calculation)

# Example – comparative RCT of two treatments for hypertension

Minimum clinical significant difference between the two would be 5mm of mercury between the improvements they each brought about. Variability of improvements between individuals (Standard Deviation) is 12mm. You want to use a P value of 5% and a power of 90%.

# Worked calculation

**$z_1 = 1.96$**

**$z_2 = 1.28$**

**σ (sigma) = 12**

**δ (delta) = 5**

**Thus**

$2 \times [(1.96+1.28) \times 12/5]^2 = 120.9$

i.e. 121 per group

# Sample size calculations

An externally randomised, pragmatic, parallel, superiority trial investigated whether communication therapy in the first four months after stroke was more effective than social contact alone for future functional communicative ability of people with aphasia or dysarthria.

The original protocol proposed a total sample size of 300 participants for 90% power to detect a difference of 0.5 points on the primary outcome of the Therapy Outcome Measure. Later, the sample size was recalculated to give 80% power at the 5% significance level to detect a difference of 0.5 points.

# Which of the following, if any, are true?

(a) The specified difference of 0.5 points on the primary outcome of the Therapy Outcome Measure is called the smallest effect of clinical interest.

(b) Power is the probability of detecting the specified difference in the Therapy Outcome Measure , if it exists in the population

(c) Reducing the power from 90% to 80% means that the sample size would increase.

# Answers

Answers *a* and *b* are true; *c* is false.

# Dolphin Study – Results (1)

| Hamilton Depression Score | Treatment Group N=15 | Control Group N=15 |
|---|---|---|
| Baseline Mean (SD) | 14.5 (2.6) | 14.5 (2.2) |
| 2 Weeks Mean (SD) | 7.3 (2.5) | 10.9 (3.4) |
| Reduction Mean (SD) | 7.3 (3.5) | 3.6 (3.4) |

# Is this enough evidence?

Difference in reduction is 7.3 – 3.6 = 3.7 points between groups. Could this just be by chance?

# P-values

- Definition: probability of obtaining these results if there really was no difference in the population. *The lady drinking tea*
- This idea of ‘no difference’ is also referred to as the ‘null hypothesis’.
- P-value in dolphin study = 0.007
- By convention, less than 0.05 said to be sufficiently small (or a ‘statistically significant’ result). In this case 0.05 is the ‘significance level’ or ‘α’)

# P-values

A randomised parallel trial investigated whether a lifestyle integrated approach to balance and strength training was effective in reducing the rate of falls in older people. Over the twelve months of the study there was a greater improvement in right ankle strength in the LiFE program than in the control program (P=0.035). However, there was no significant difference in improvements reported in right knee strength (P=0.45). All statistical tests were one sided, and the critical level of significance was set at 0.05 (5%).

# Which of the following statements, if any, can be concluded?

(a) The P value represents the strength of the evidence in support of the null hypothesis

(b) There was a statistically significant difference between treatments in ankle strength at the 0.05 level of significance

(c) There is no difference in mean improvements in right knee strength between treatments in the total population

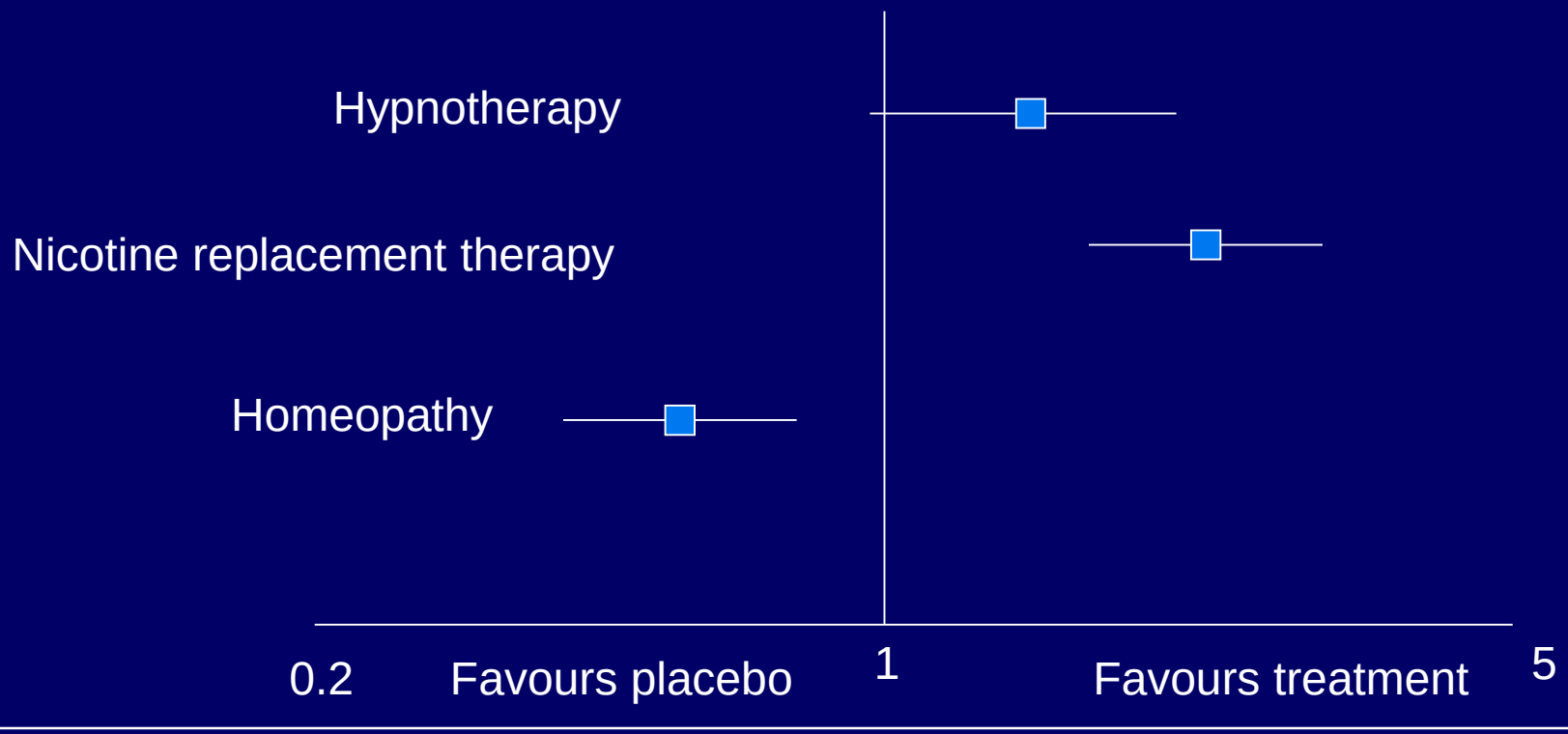
(d) The null hypothesis for the statistical test of mean right knee strength between treatments is true.

Answers *a* and *b* can be concluded; answers *c* and *d* cannot.

# Confidence Intervals

- 95% Confidence Interval (CI): range of values we can be 95% confident contains the true population statistic.
- 95% CI for difference in mean reduction = (1.1, 6.2)
- Gives more information than a p-value.

# What does this diagram tell us about the probability of quitting after five years in these three smoking cessation studies?



# Which, if any, of these odds ratios confidence intervals are significant?

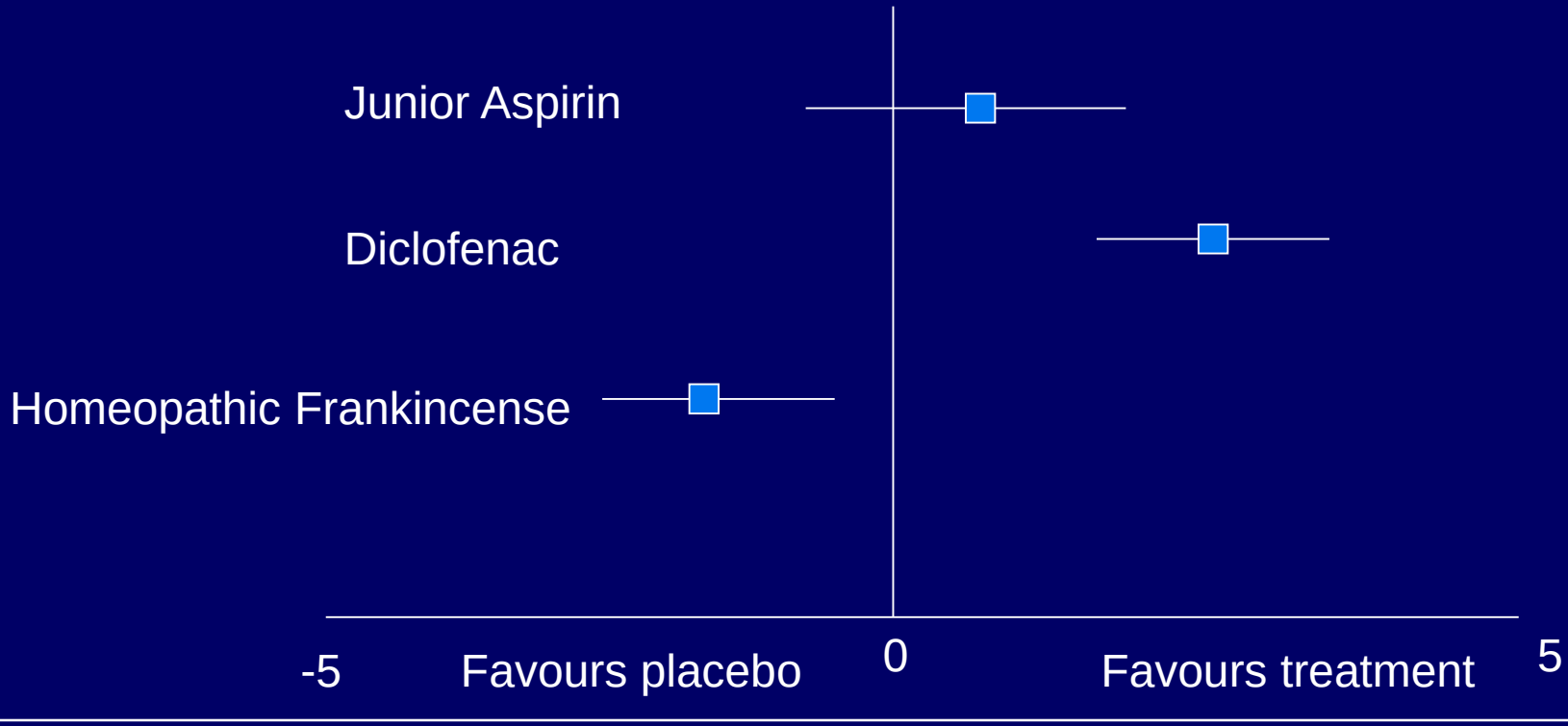
(a) (2.04, 2.97)

(b) (0.55, 0.85)

(c) (0.98, 2.45)

# What can we say about the comparative changes in reported postoperative pain scores following use of these three drugs?

## Mean differences



# Which, if any, of these confidence intervals for the pain reduction studies are significant?

(a) (-2.55, -0.05)

(b) (1.55, 3.15)

(c) (-0.95, 1.85)

Can you match the confidence intervals to the studies in the previous slide?

# Confidence intervals: predicting uncertainty

An eight arm randomised controlled trial assessed the effectiveness of a range of weight management programmes in terms of weight loss. 740 obese participants were recruited and randomly allocated to one of six twelve week weight loss programmes, a choice of programme, or a control option of twelve vouchers enabling free access to a leisure centre. At the end of the twelve weeks the participants who completed the Weight Watchers programme showed a mean weight loss of 5.15 kg. (CI 4.2 kg to 6.1 kg)

## Which one of the following statements best describes the information provided by the 95% confidence interval for mean weight loss at twelve weeks for the Weight Watchers group?

(a) 95% of sample participants in the Weight Watchers programme achieved a weight loss between 4.2 kg and 6.1 kg.

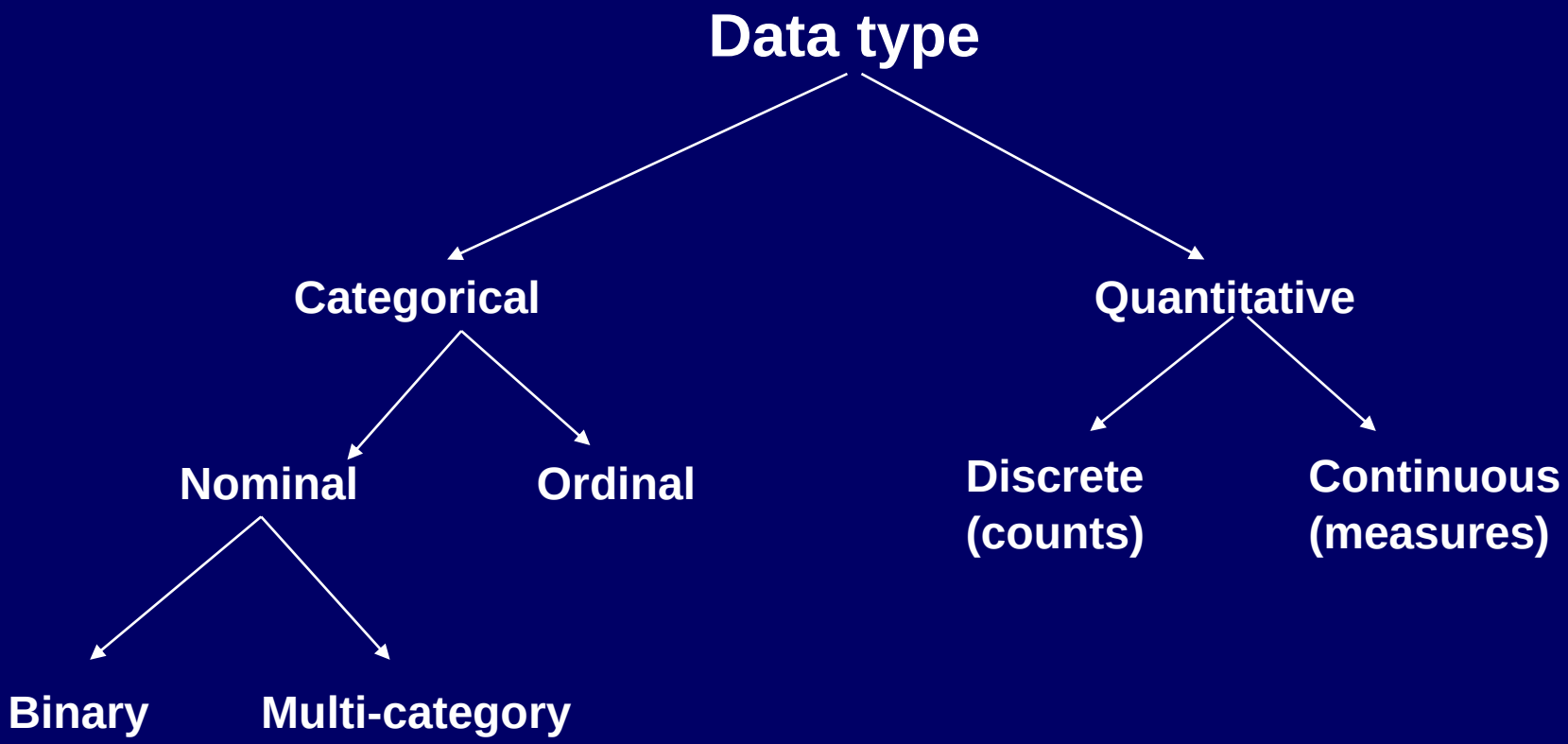
(b) 95% of the obese population would achieve a weight loss between 4.2 kg and 6.1 kg if they completed a twelve week Weight Watchers programme

(c) There is a probability of 0.95 that the obese population mean weight loss at twelve weeks on the Weight Watchers programme would be between 4.2 kg and 6.1 kg

(d) There is a probability of 0.95 that the sample mean weight loss for the Weight Watchers programme group was between 4.2 kg and 6.1 kg

Statement c is correct.

# Types of data



# Statistical Tests

- Test is enough evidence for e.g., a difference in mean between groups, that a statistic is sufficiently different from 0, etc.
- Each test has a null hypothesis, with an associated p-value, but most outputs should also provide an appropriate confidence interval.
- Most common statistical tests used are t-test (null hypothesis that no difference in means) and Chi-squared test (null hypothesis that no difference in proportions).
- T-tests assume data is continuously variable, e.g. blood pressure. Chi-squared tests assume data is discrete, e.g. number of siblings

# ‘Independent’ & ‘Paired’ data

- Independent (Unpaired) Samples:
  - The data in each sample come from a separate set of patients.
  - The outcomes in one sample do not affect the outcomes of another sample.
  - E.g. Recording the length of stay of patients in different hospitals.

- Paired Samples:
  - The data in each sample come from the same patients.
  - There is a relationship between the outcomes of one sample and the outcomes in another sample.
  - E.g. Recording blood glucose levels in each person once after they have received treatment A and once after they have received treatment B.

| | Categorical data | Non parametric data | Parametric data |
|---|---|---|---|
| One sample | χ2 test, Fisher's exact test | Wilcoxon signed rank test | One sample t-test |
| Comparing two groups | χ2 test[U], McNemar's test [P] | Mann-Whitney U test[U], Wilcoxon rank sum test [P] | t-test [U] or [P] |
| Comparing more than two groups | χ2 test[U], McNemar's test [P] | ANOVA [U], Friedman test[P] | ANOVA [U] or [P] |

**Key** [U] unpaired [P] paired

# Statistical tests 1

A randomised controlled trial compared the effects of Hydrofiber and alginate dressings on surgical wounds following hip replacement. Presence of blister was a key primary outcome.

The percentage of participants with blister in the Hydrofiber group was higher than that in the alginate group (18/100 (18%) v 7/100 (7%)). This difference was statistically significant (P=0.03).

**Which one of the following statistical tests would most likely have been used to compare the treatment groups with regard to the percentage of participants with blistering?**

a) The $\chi 2$ test
b) Independent samples t test
c) Paired samples t test
D) Mann-Witney U text

# The answer is a

# Statistical tests 2

The same study also reported the amount of pain from the dressing reported by patients during mobilisation. Pain was measured on a VAS score where 0 = no pain, and 10 = unbearable pain. The mean pain score reported by the Hydrofiber group was 0.34 compared with 0.42 in the Alginate group, although the difference was not statistically significant (P=0.6)

**Which one of the following statistical tests would most likely have been used to compare the treatment groups with regard to the mean pain scores during mobilisation reported by the two groups?**

a) The $\chi 2$ test

b) Unpaired samples t test

c) Paired samples t test

d) Mann-Whitney U test

The answer is b

# Non-normally distributed data (1)

| Statistical test | Data | Explanation |
|---|---|---|
| Sign test<br>Wilcoxon's signed rank test | One sample data | The median of the sample is compared with a hypothetical mean |
| Mann-Whitney U test | Two unpaired samples | The median of one sample is compared with the median of another sample |
| Wilcoxon's matched pairs test | Two paired samples | The median of one sample is compared with the median of another sample |

# Non-normally distributed data (2)

| Statistical test | Data | Explanation |
|---|---|---|
| Kruskall-Wallis ANOVA test | Three or more samples of unpaired data | The medians of samples from three or more groups are compared |
| Friedman's test | Three or more samples of paired data | The medians of samples from three or more groups are compared |

# Normally distributed data

| Statistical test | Data | Explanation |
|---|---|---|
| One sample t test | One sample data | The mean of the sample is compared with a hypothetical mean |
| t test | Two unpaired samples | The mean of one sample is compared with the mean of another sample |
| Paired t test | Two paired samples | The mean of one sample is compared with the mean of another sample |
| Analysis of variance: ANOVA | Three or more samples of unpaired or paired data | The means of samples from three or more groups are compared |

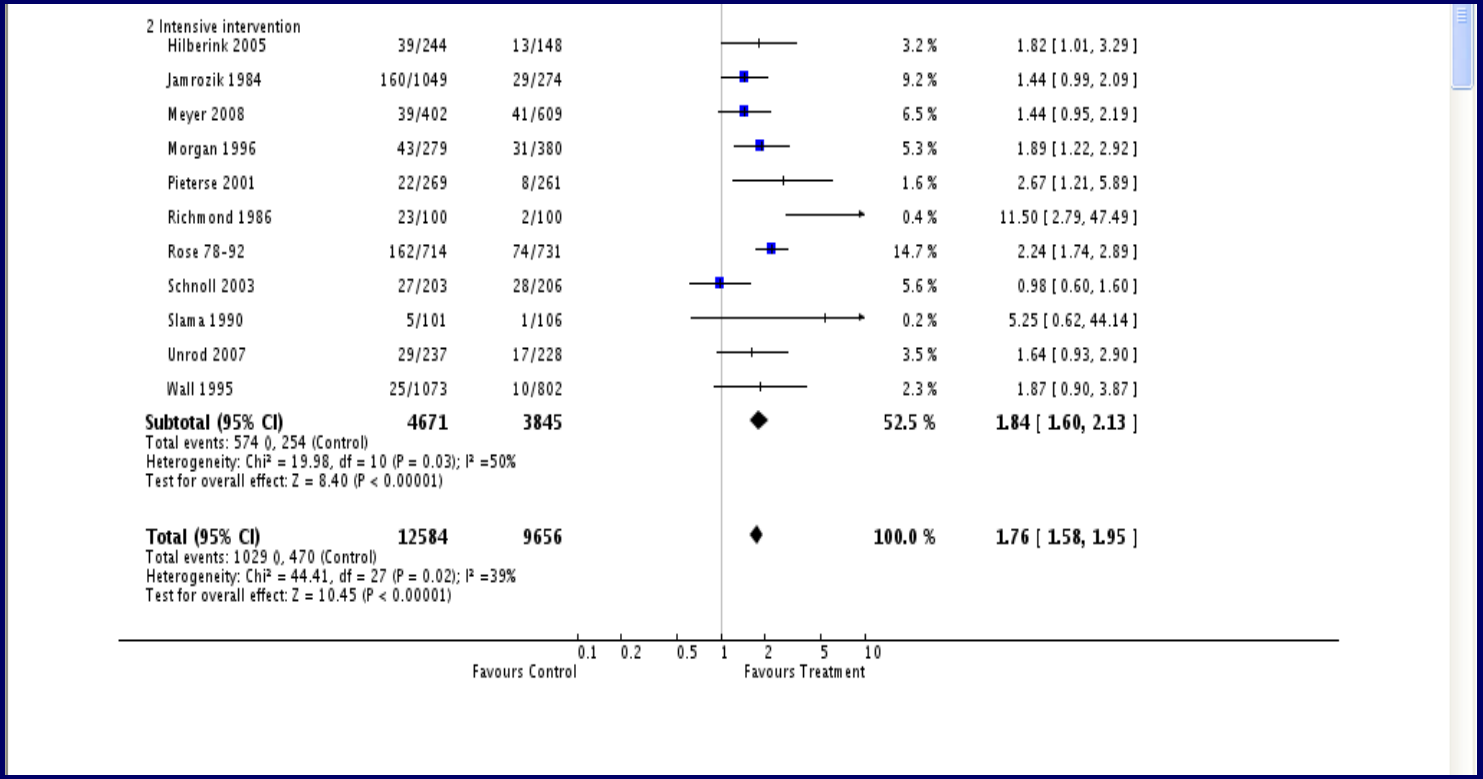
# Alternative methods of reporting results

- Relative risk reduction
- Absolute risk reduction
- Number needed to treat

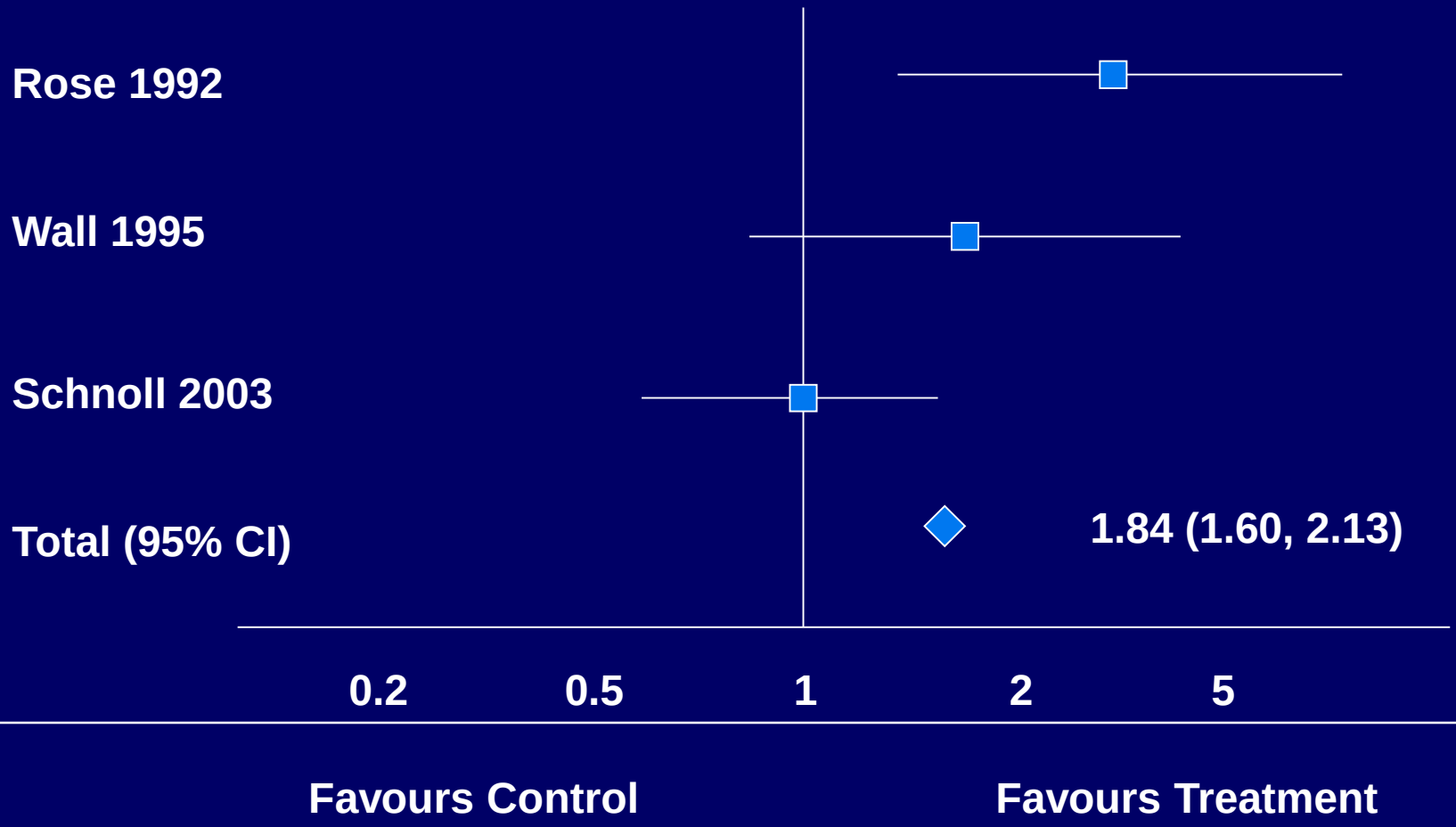
# How effective is it for GPs to spend time encouraging their patients to give up smoking?

Stead LF, Bergson G, Lancaster T. Physician advice for smoking cessation. *Cochrane Database of Systematic Reviews* 2008, Issue 2. Art. No.: CD000165. DOI: 10.1002/14651858.CD000165.pub3.

# Effectiveness of intensive advice on numbers of quitters



# Effectiveness of intensive advice on numbers of quitters: close up view



Risk ratio i.e. how much more effective than control (in this case placebo) is this intervention?

1.84 (C.I. 1.60, 2.13)

# Relative risk: how *many times more* likely* are you to quit following intensive advice than with placebo?

3.7% (quit rate with intensive advice)

2% (quit rate without intensive advice)

Relative Risk = 3.7/2 = 1.85

i.e. Smokers receiving intensive advice are 1.85 times more likely to quit

* Or less!

Absolute risk (effectiveness) i.e. *how much higher** is the quit rate in the intervention group than in the control group?

Placebo group: 2%

Intensive intervention group: 3.7%

Added value of intervention = (3.7-2)% =1.7%

* Or lower!

Number needed to treat: How many smokers would need to receive intensive advice in order to produce one extra quitter?

Formula: NNT =100/Absolute Risk

NNT = 100/1.7 = 59 (C.I. 62, 48)

# The Scandinavian Simvastatin Survival Study (4S)

| (4444 Patients) | Simvastatin 2221 | Placebo 2223 |
|---|---|---|
| Alive after 5 years | 2039 | 1967 |
| Dead after 5 years | 182 | 256 |

# Proportion dead after five years

- Placebo: 256/2223
- = 0.115
- = 11.5%

- Simvastatin: 182/2221
- = 0.082
- = 8.2%

# Relative risk of death

- Death rate in simvastatin group as proportion of death rate in placebo group
- = 8.2/11.5
- = 0.71
- =71%

# Relative risk reduction

- 1-0.71
- = 0.29
- = 29%
- Simvastatin patients had a reduction in their relative risk of death within 5 years of 29%

# Absolute risk reduction

- 0.115-0.082 = 0.033
- Or 11.5%-8.2% = 3.3%
- i.e 3.3% fewer of the patients in the simvastatin group died than in the placebo group

# NNT: number needed to treat

- = 3.3% of 100
- = 100/3.3
- = 30
- i.e for every 30 patients treated with simvastatin rather than placebo, one extra patient will be alive after five years

# Cost to save one extra life after five years

- Simvastatin 10 mg, net price 28-tab pack = 95p
- Cost per patient per year: 365/28 x £.95 = £12.38
- Cost per 30 patients for five years: £12.38 x 5 x 30 = £1857

# Does this improvement come at a cost?

Statins can cause various muscular side-effects, including myositis, which can lead to rhabdomyolysis. Muscular effects are rare but often significant. Statins can cause gastro-intestinal disturbances, and very rarely pancreatitis. They can also cause altered liver function tests, and rarely hepatitis and jaundice; hepatic failure has been reported very rarely. Other side-effects include sleep disturbance, headache, dizziness, depression, paraesthesia, asthenia, peripheral neuropathy, amnesia, fatigue, sexual dysfunction, thrombocytopenia, arthralgia, visual disturbance, alopecia, and hypersensitivity reactions (including rash, pruritus, urticaria, and very rarely lupus erythematosus-like reactions). In very rare cases, statins can cause interstitial lung disease; if patients develop symptoms such as dyspnoea, cough, and weight loss, they should seek medical attention.

# Number needed to treat I

In another trial comparing simvastatin with placebo for the prevention of repeat myocardial infarction in post MI patients, the five year incidence of repeat myocardial infarction was 8% in the simvastatin group compared with 12% in the group receiving placebo. Calculate the number needed to treat to prevent one extra case of MI within this group of patients.

The answer is 25

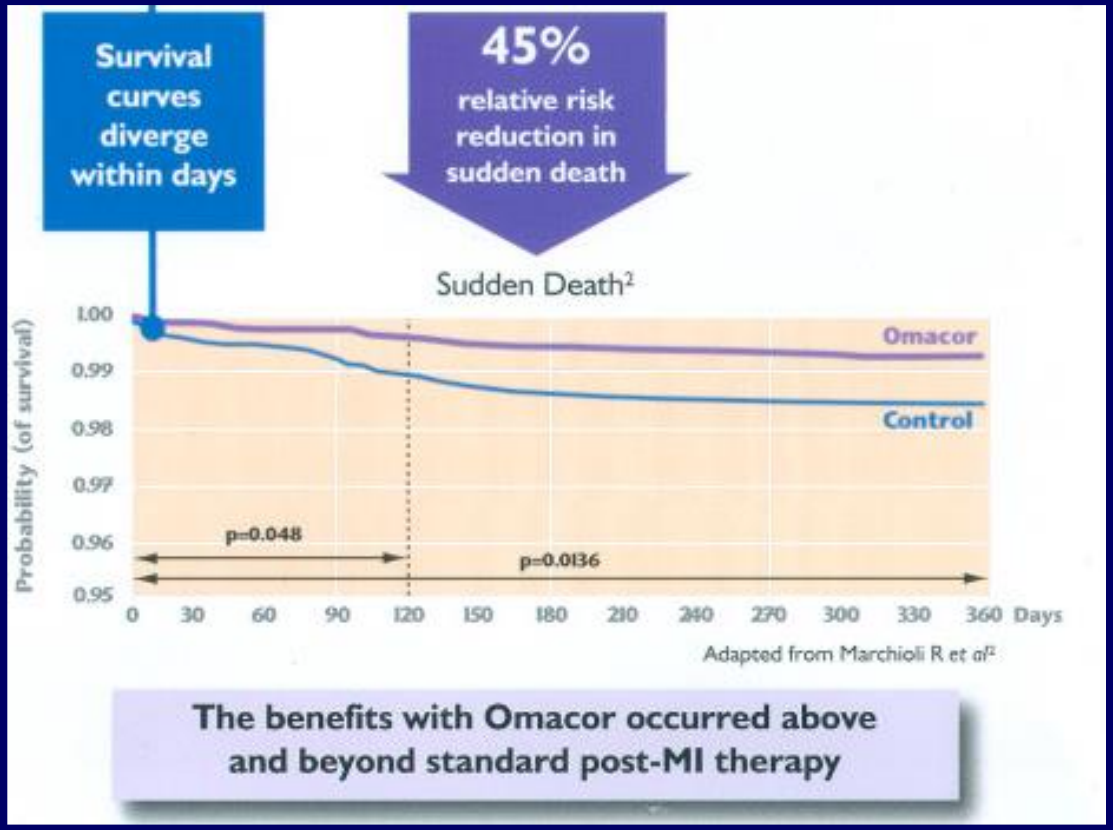
# Number needed to treat II

A trial compared glucosamine with placebo for the treatment of osteoarthritis. The main outcome was a change in the Lequesne index - which combines pain severity, maximum distance walked, and activities for daily living. After the four week trial 35% of participants who received placebo showed an improvement on the Lequesne index, compared with 55% of patients who received glucosamine. Calculate the number needed to treat with glucosamine to bring about one extra improved patient.

The answer is 5.

# Omacor saves more lives when added to standard post-MI therapy

and a Mediterranean diet



# Proportion dead after one year

- Placebo ~1.4%
- Omacor ~0.7%

# Relative risk of death

- Death rate in Omacor group as proportion of death rate in placebo group
- = ~ 0.7/1.4
- = ~ 0.5
- = ~ 50%

# Absolute risk reduction

- = ~1.4% - ~0.7%
- = ~0.7%
- Let’s be generous and round it up to 1%

# Number needed to treat

- If we use “accurate” reading = 100/0.7% = 142
- Using generous reading = 100/1% = 100
- Cost to save one additional life?

# Cost to save one extra life:

- 100-cap pack = £50.84
- 3.65 packs per patient per year = £185.56
- Multiplied by the 100 patients taking Omacor per life saved = £18,556.00

# Methods of Analysis

## Per-protocol (scientific)

- Analyse patients according to the treatment they received.
- "What is the effect of using the treatment?"

## Intention-to-treat (clinical)

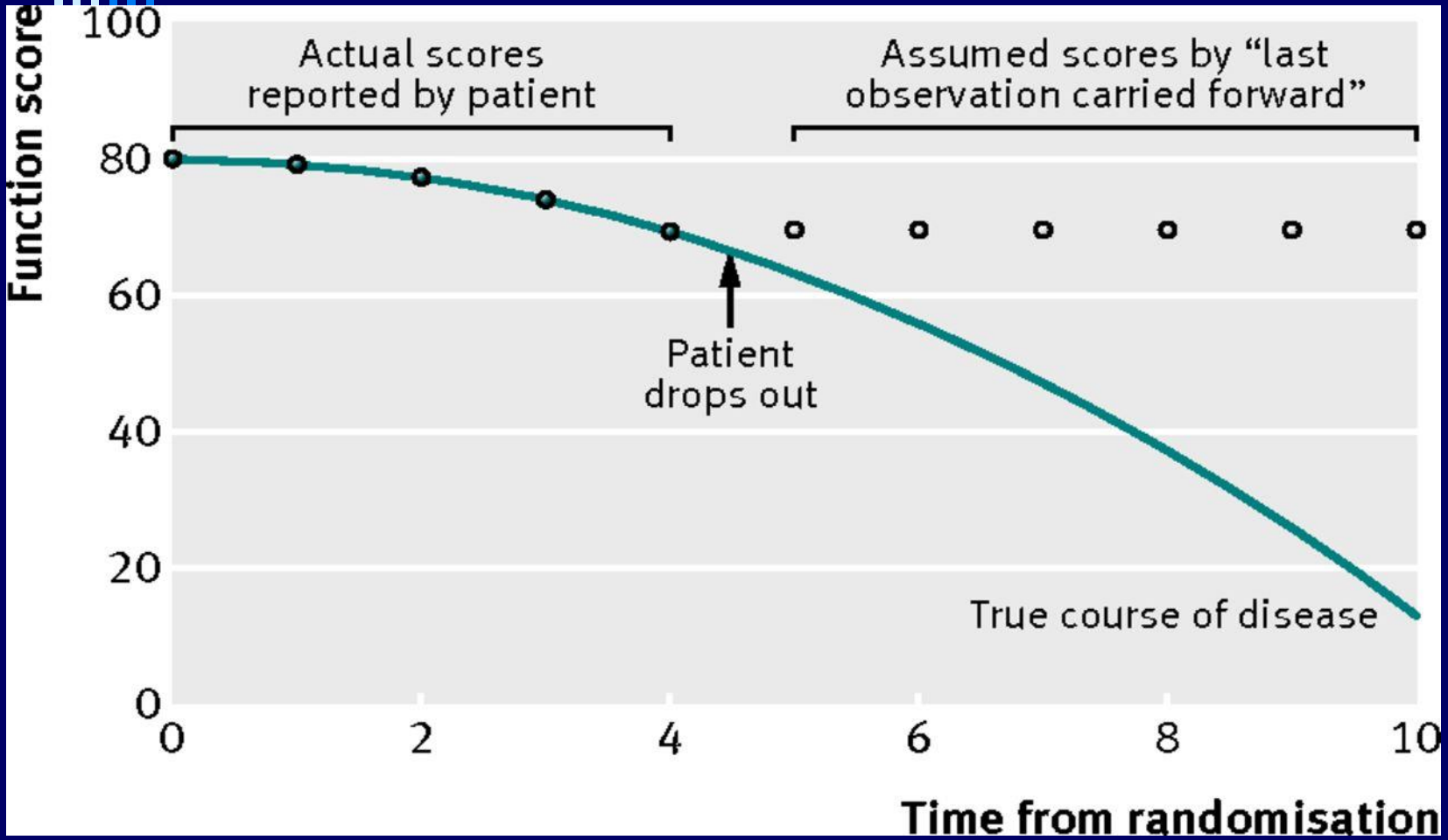
- Analyse patients according to the group they were allocated to.
- "What is the effect of prescribing the treatment?"

# How to cater for missing values

- First observation carried forward
- Last observation carried forward
- Multiple imputation: “Rubin’s rules”

But remember: a guess is still a guess.

# Function scores over time for patient with chronic degenerative disease.




Vickers A J , and Altman D G BMJ 2013;346:bmj.f3438

# Intention to treat analysis

A randomised double blind controlled trial investigated whether postextubation respiratory support via heated, humidified, high-flow nasal cannulae (HHHFNC) results in a greater proportion of infants younger than 32 weeks' gestation being successfully extubated after a period of endotracheal positive pressure ventilation compared with conventional nasal continuous positive airway pressure (NCPAP).

Owing to transfers out of the ward three babies in the Vapotherm group and four babies in the NCPAP group were lost to follow up. Statistical analyses were carried out using intention to treat and showed that there were fewer cases of nasal trauma in the HHFNC than the NCPAP group ($p<0.01$).

# Which one of the following statements best describes the principle of intention to treat?

a) All infants compared received the treatment the recruiting clinician had originally intended prior to trial recruitment.

b) The two groups of infants were compared on the basis of the treatment they eventually received.

c) The two groups of infants were compared on the basis of the treatment regimen assigned at start of the trial.

d) Infants were only included in the analyses if they completed the treatment originally allocated.

# Dolphin Study – Results (2)

"We conducted preliminary *t* tests for equality of means for two independent groups of observations for the Hamilton, Beck, and Zung's scores, to evaluate the significance of the changes in the scores from baseline to the end of treatment. The primary analysis was a modified analysis by intention to treat and last observation carried forward."

# Dolphin Study – Results (3)

"[The t-test] was highly significant in the patients who completed treatment (animal care programme, n = 13; outdoor nature programme, n = 12). For the Hamilton scale (95% confidence interval 1.66 to 6.11, P = 0.002; equal variances not assumed), the mean differences in change scores for the animal care programme and the outdoor nature programme were 8.38 (SD 1.98) and 4.50 (SD 3.15), respectively."

# Multiplicity

| Number of tests | Chance of at least one false significant value |
|---|---|
| 1 | 0.05 |
| 2 | 0.10 |
| 3 | 0.14 |
| 5 | 0.23 |
| 10 | 0.40 |
| 20 | 0.64 |

- E.g. Significance level = 0.05

  1/20 tests will be ‘significant’, even when no difference in target population.

- Can correct for this issue by lowering the significance level.

- Needs to be accounted for in a sample size calculation.

# Lancet 1988: ISIS-2 trial (RCT of oral streptokinase and aspirin on patients with suspected myocardial infarction)

expected. For example, subdivision of the patients in ISIS-2 with respect to their astrological birth signs appears to indicate that for patients born under Gemini or Libra there was a slightly adverse effect of aspirin on mortality (9% SD 13 increase; NS), while for patients born under all other astrological signs there was a strikingly beneficial effect (28% SD 5 reduction; $2p < 0{\cdot}00001$). It is, of course, clear

# Useful References

An introduction to medical statistics. Martin Bland.

Practical statistics for medical research. Douglas G. Altman.

Medical statistics made easy. Michael Harris and G Taylor

How to read a paper. Trisha Greenhalgh.

‘Endgames’ or ‘Statistics Notes’ on BMJ Online.