Clinical trials: Reading between the lines

Dr Andrew Hitchings

Senior lecturer in clinical pharmacology / Consultant in neuro-intensive care

A seminal study (1980)

■ 1073 patients with angiographically-proven CAD

■ Randomised to two management strategies

■ Planned analyses:

– Primary endpoint: all-cause mortality

– Multivariable survival analysis over 5 years of follow-up

– Biologically-relevant subgroup analyses

Baseline characteristics

- Generally well-balanced
- Slightly higher prevalence of LV impairment in Group 2

TABLE 1. *Distribution of Selected Baseline Characteristics*

	Prevalence (%)		
	Total (n = 1073)	Group 1 (n = 539)	Group 2 (n = 534)
Males	85	84	85
Age > 50 years	53	53	53
History of previous MI	51	49	53
History of CHF	14	14	14
Cardiomegaly on chest x-ray	20	18	22
Diagnostic Q waves on ECG	43	41	46
Resting ST-T-wave abnormalities	47	47	47
LVEDP > 18 mm Hg	15	14	15
AVO$_2$D > 5.5 vol %	19	19	18
Single-vessel disease	24	24	24
Three-vessel disease	51	51	51
Abnormal LV contraction	60	57	63
Significant mitral insufficiency	8	6	10
Left main stenosis ≥ 50%	16	17	15

Abbreviations: MI = myocardial infarction; CHF = congestive heart failure; LVEDP = left ventricular end-diastolic pressure; AVO$_2$D = arteriovenous oxygen difference; LV = left ventricular.
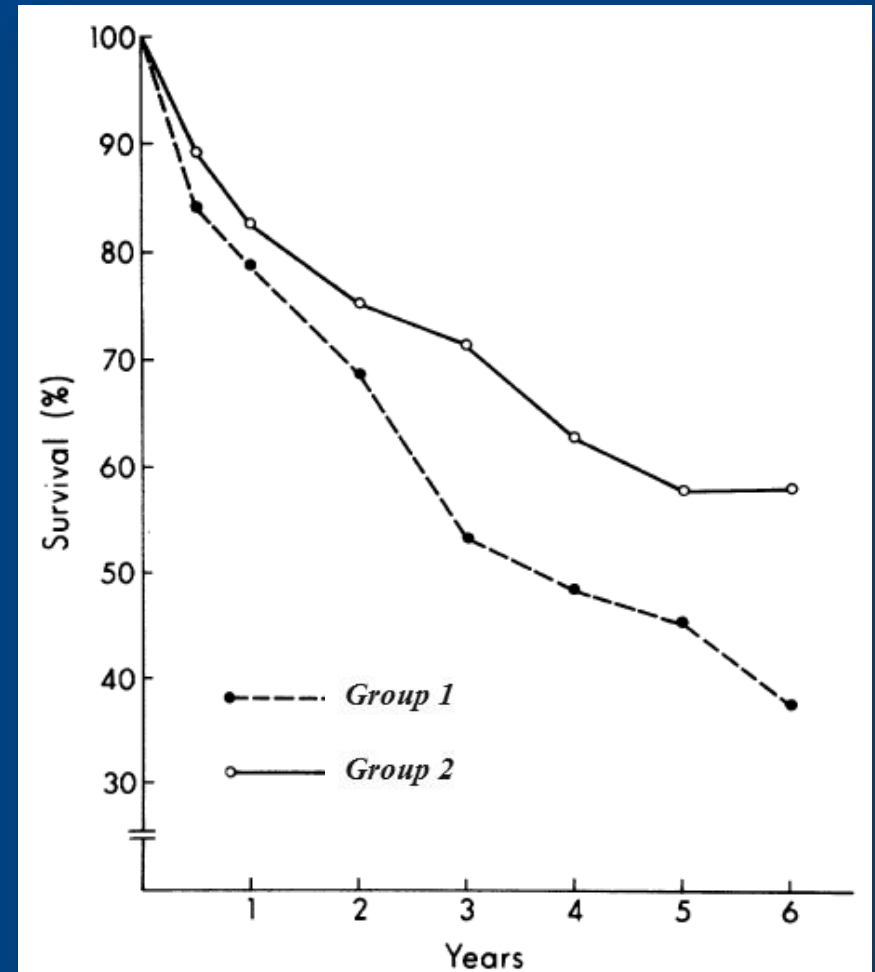
Primary analysis

- Overall survival similar

Subgroup analyses

- Number of significantly diseased vessels
- Presence or absence of LV impairment
- Symptoms of congestive cardiac failure

Subgroup analyses

- Triple-vessel disease and LV impairment (n=397)

Subgroup analyses

- …And no established symptoms of CCF (n=298)
  - 3-year survival: 60% *vs.* 80% (P<0.01)
  - Independent of other variables (P<0.01)
  - Still significant after correction for multiple comparisons

Study conclusion

- Treatment approach made no difference to survival in the population as a whole
- But there was a clinically and statistically significant difference in a sizable minority:
  - 20% absolute difference at 3 years (NNT 5)

Interpretation

- What treatment was studied in this trial?
- What do you think about the analysis and findings?

There was no treatment…

508   CIRCULATION   VOL 61, No 3, MARCH 1980

Lessons from a Simulated Randomized Trial in Coronary Artery Disease

KERRY L. LEE, PH.D., J. FREDERICK MCNEER, M.D., C. FRANK STARMER, PH.D.,

PHILIP J. HARRIS, M.B., D.PHIL., AND ROBERT A. ROSATI, M.D.

SUMMARY   A simulated randomized clinical trial in coronary artery disease was conducted to illustrate the need for clinical judgment and modern statistical methods in assessing therapeutic claims in studies of complex diseases. Clinicians should be aware of problems that occur when a patient sample is subdivided and treatment effects are assessed within multiple prognostic categories. In this example, 1073 consecutive, medically treated coronary artery disease patients from the Duke University data bank were randomized into two groups. The groups were reasonably comparable and, as expected, there was no overall difference in survival. In a subgroup of 397 patients characterized by three-vessel disease and an abnormal left ventricular contraction, however, survival of group 1 patients was significantly different from that of group 2 patients. Multivariable adjustment procedures revealed that the difference resulted from the combined effect of small imbalances in the distribution of several prognostic factors. Another subgroup was identified in which a significant survival difference was not explained by multivariable methods.

    These are not unlikely examples in trials of a complex disease. Clinicians must exercise careful judgment in attributing such results to an efficacious therapy, as they may be due to chance or to inadequate baseline comparability of the groups.
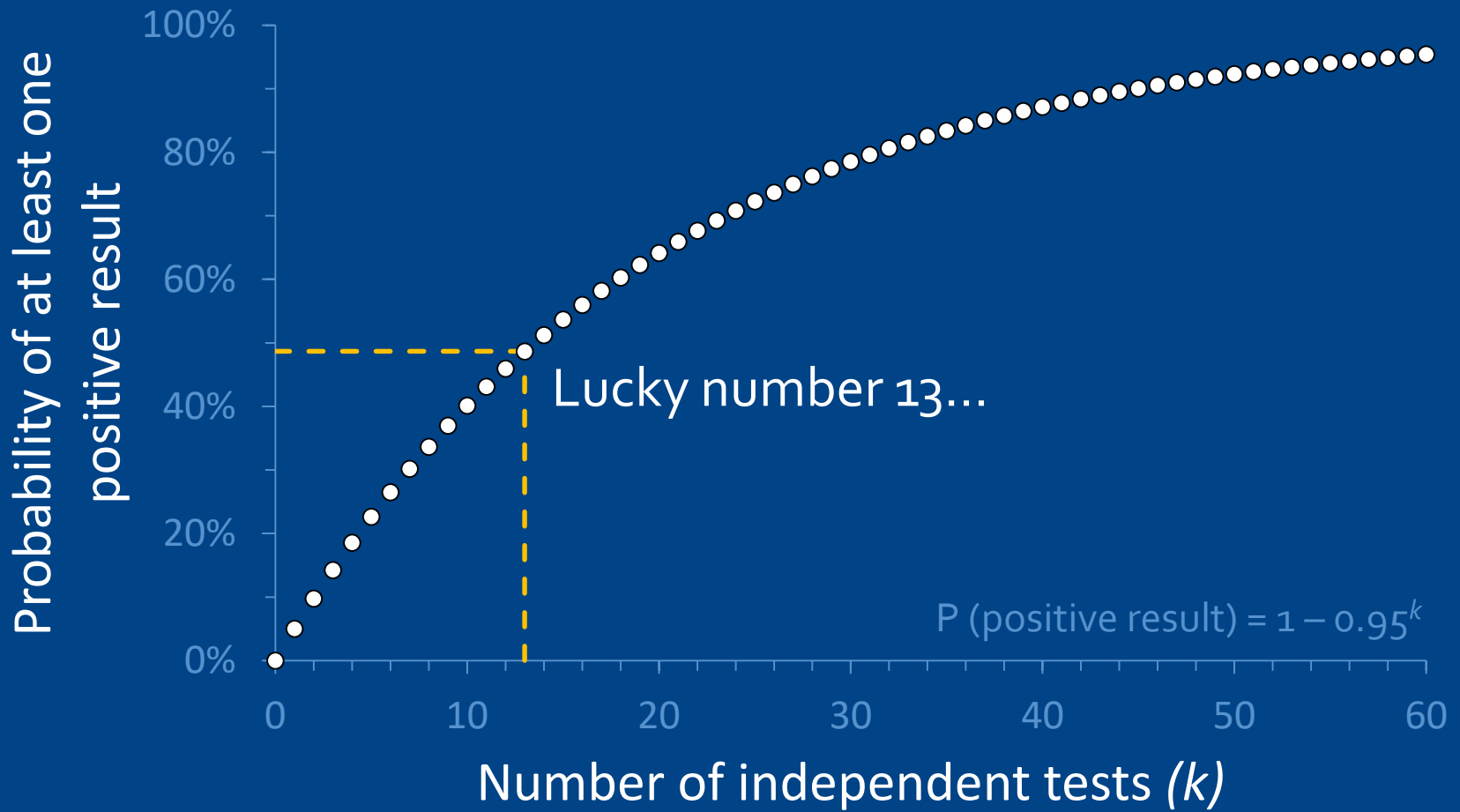
Clinical trials: Reading between the lines

- **Multiplicity**

The problem of multiplicity

Probability of at least one positive result (y-axis): 0% to 100%

Number of independent tests *(k)* (x-axis): 0 to 60

$$P\text{ (positive result)} = 1 - 0.95^{k}$$

The problem of multiplicity

Probability of at least one positive result (y-axis): 0%, 20%, 40%, 60%, 80%, 100%

Number of independent tests *(k)* (x-axis): 0, 10, 20, 30, 40, 50, 60

Lucky number 13…

$$P\,(\text{positive result}) = 1 - 0.95^{k}$$

The problem of multiplicity

- Multiplicity is everywhere, both open and hidden
  - Multiple questions, subgroups and endpoints
  - Multiple methods of analysis
  - Multiple trials, published and unpublished
- Multiplicity 'threatens the validity of every statistical conclusion'[1]

1. *Pharmaceutical Statistics* 2007;6:155-60

The problem of multiplicity
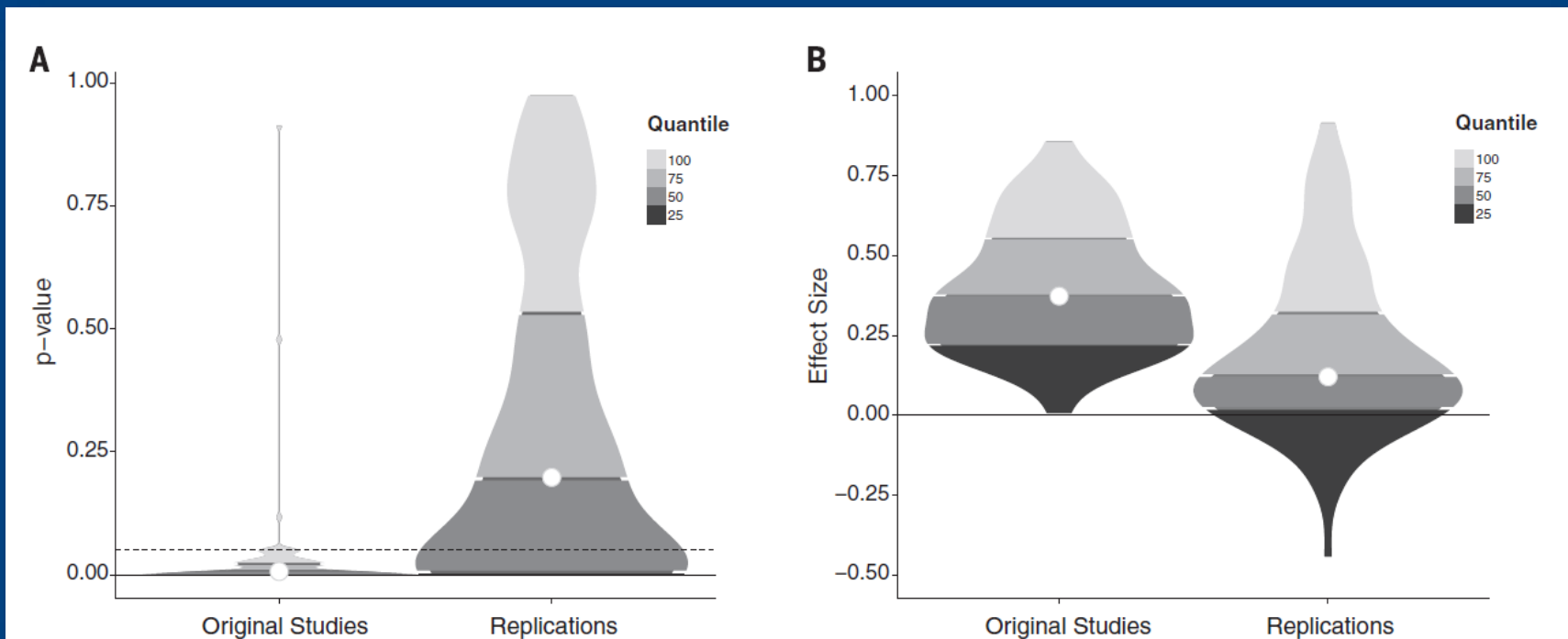
Clinical trials: Reading between the lines

1. **Multiplicity**

Clinical trials: Reading between the lines

1. Multiplicity
2. **Heterogeneity**

The tacit homogeneity assumption

- Therapeutic effects are evenly distributed among trial participants
- Spread of treatment effects in the trial reflects the spread in the population from which it was drawn

Selection biases

Region of indifference

Population density

Frequency

0

Treatment effect
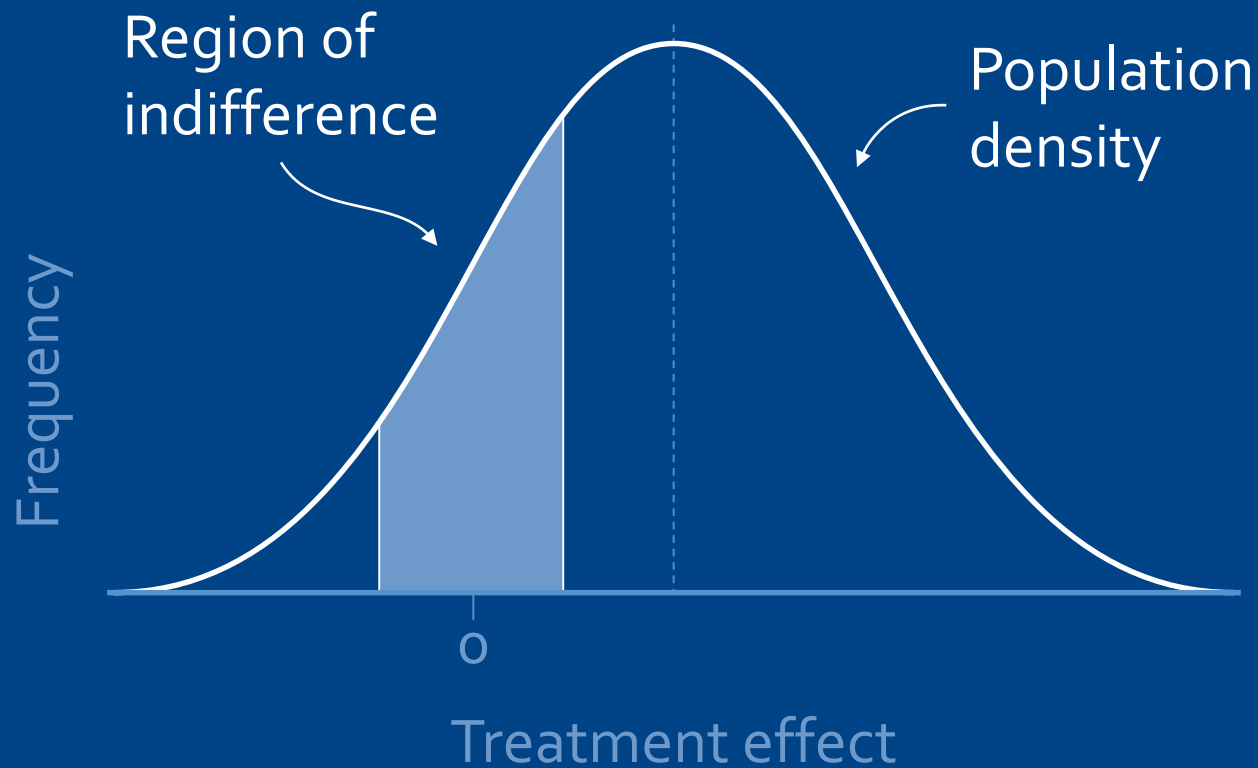
Simple random sampling
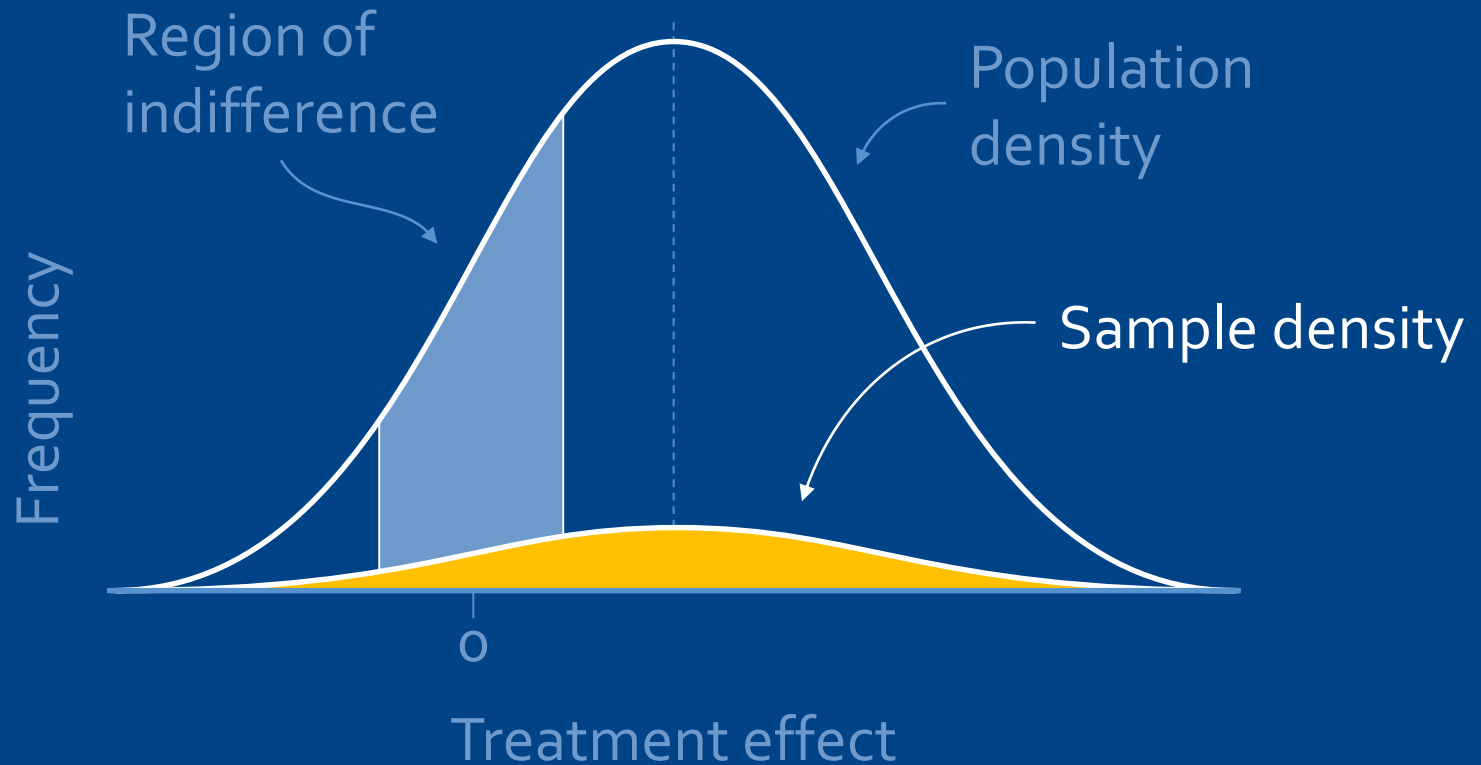
Centre-biased sample

Frequency

0

Treatment effect
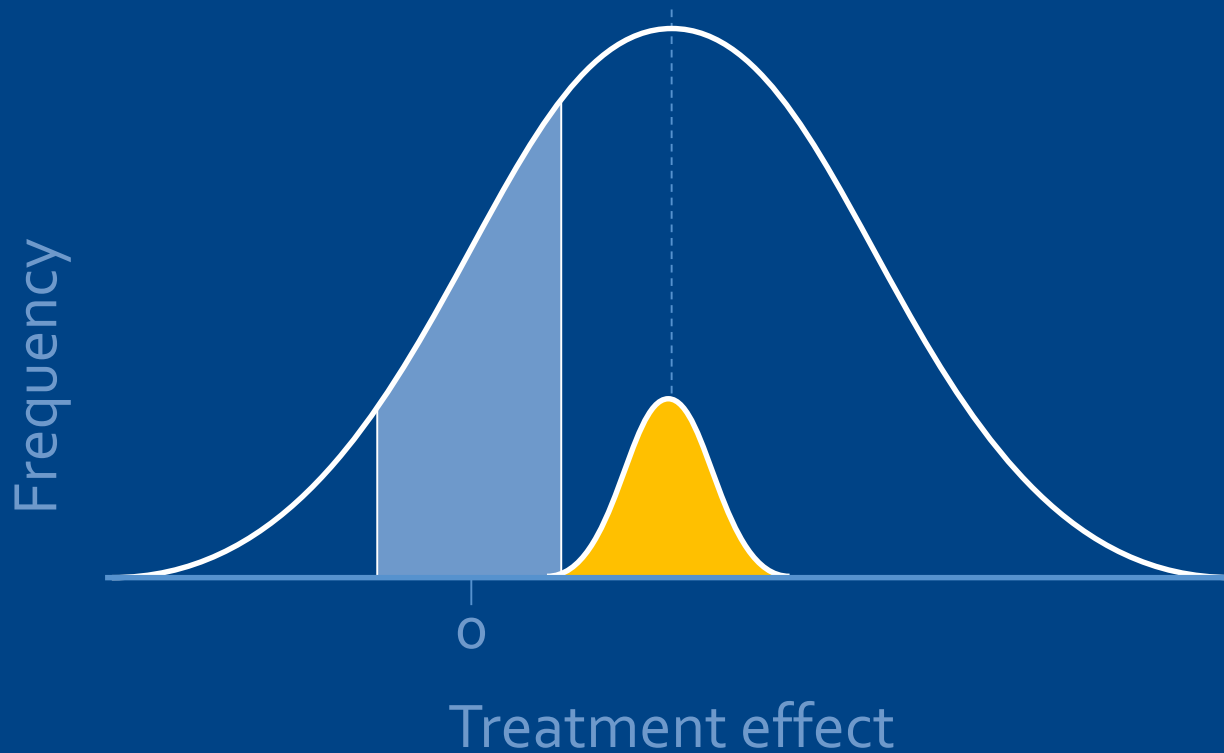
Tail-biased sample
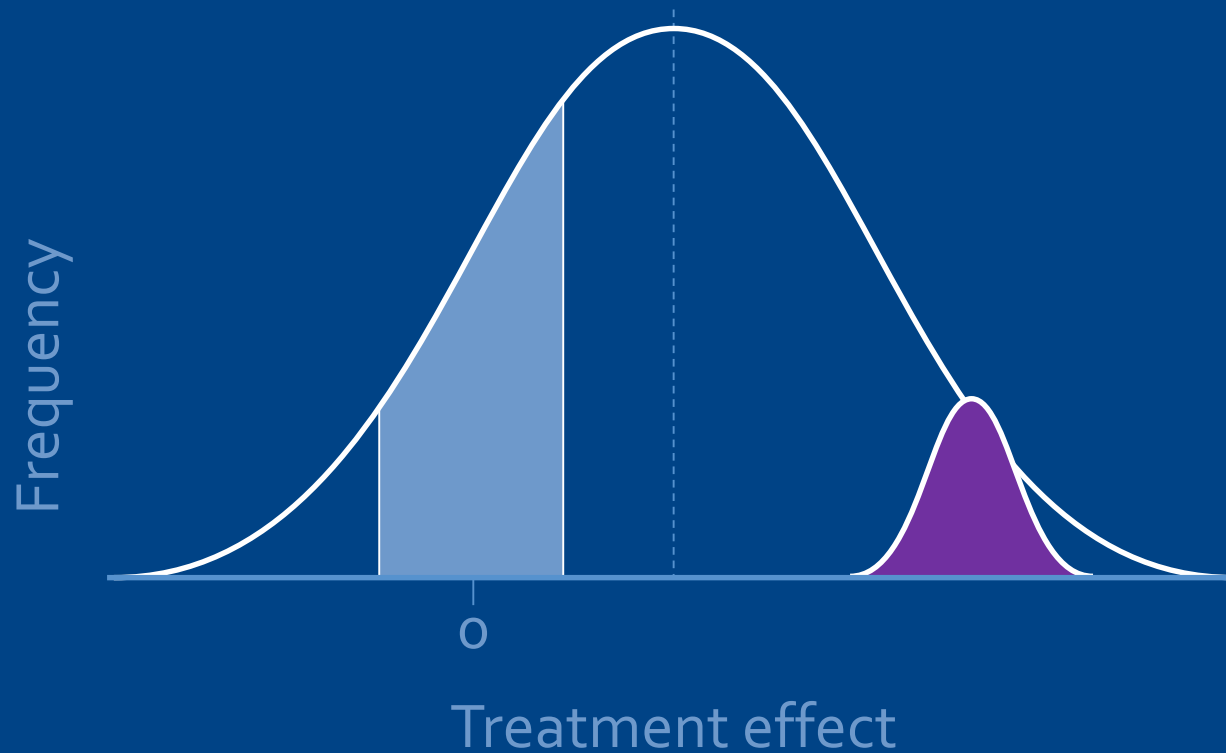
Heterogeneity in clinical trials

- **Benefit** from treatment depends on baseline risk
- **Harm** from treatment is distributed fairly randomly

Heterogeneity in clinical trials
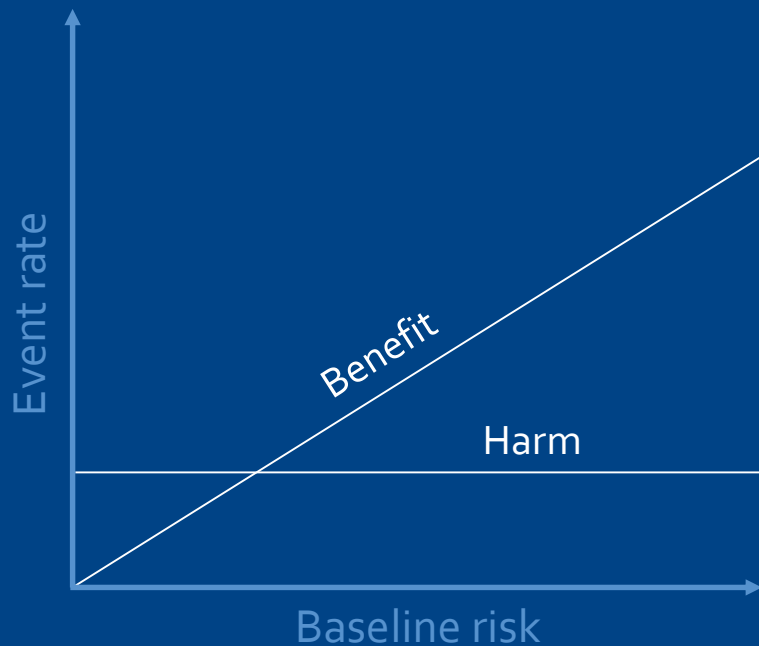
Heterogeneity in clinical trials

Net harm
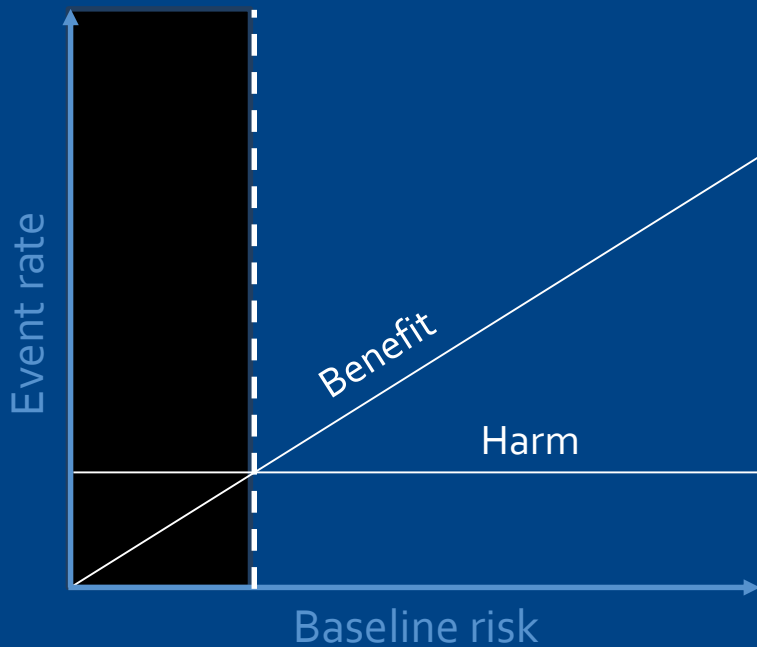
Net benefit

Heterogeneity in clinical trials
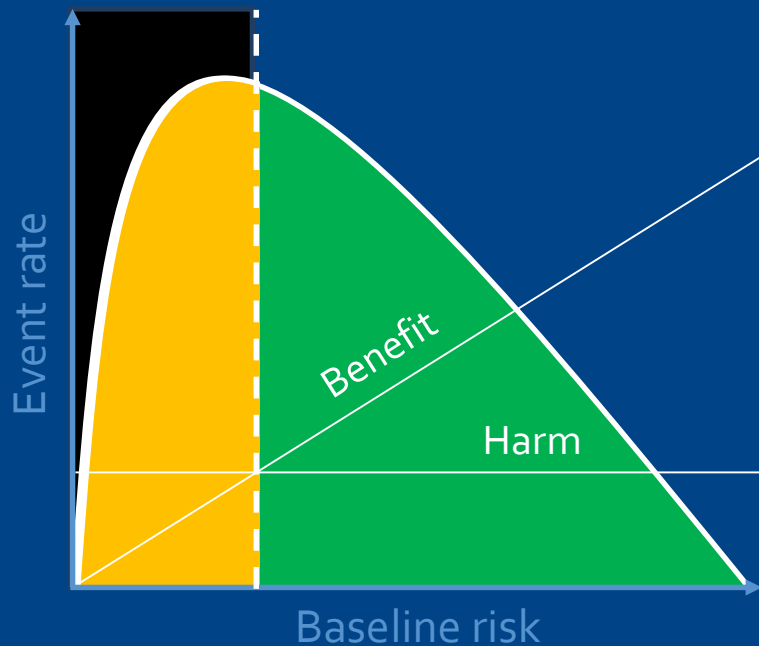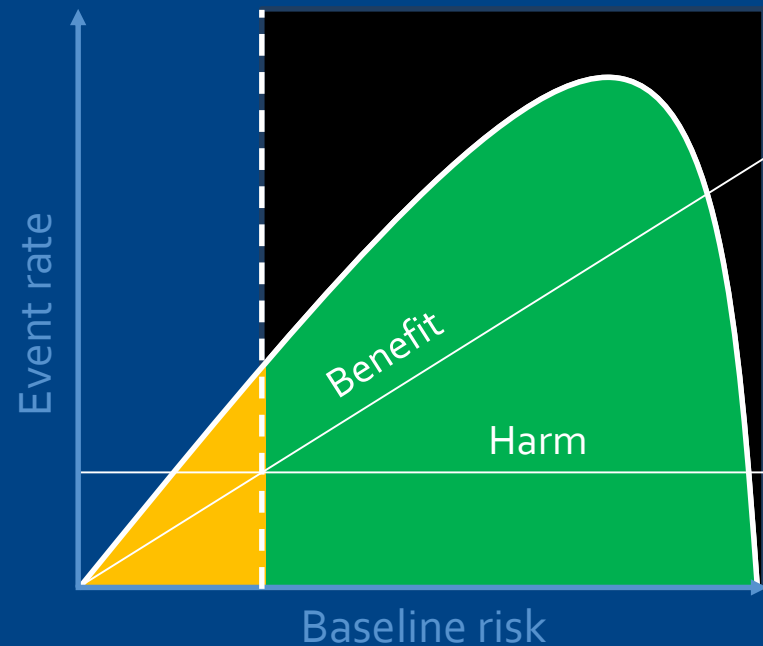
Treatment of carotid stenosis

Any stroke or operative death

(Left panel: <30% stenosis)

Logrank=4·5; p=0·03
5-year ARR=−2·6% (−6·2 to 0·9)

Proportion with event — Years after randomisation

Surgical
Medical

<30% stenosis

(Right panel: 70–99% stenosis)

Logrank=22·2; p<0·00001
5-year ARR=15·3% (9·8 to 20·7)

Years after randomisation

Surgical
Medical
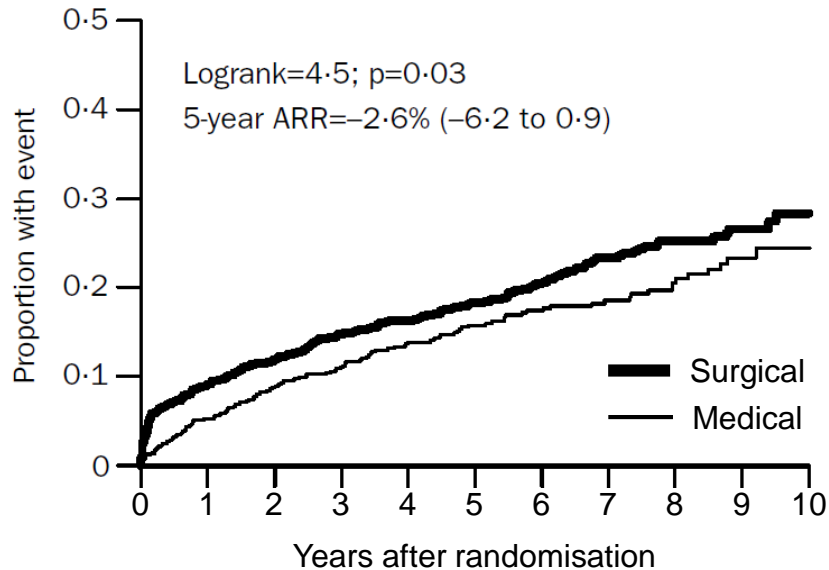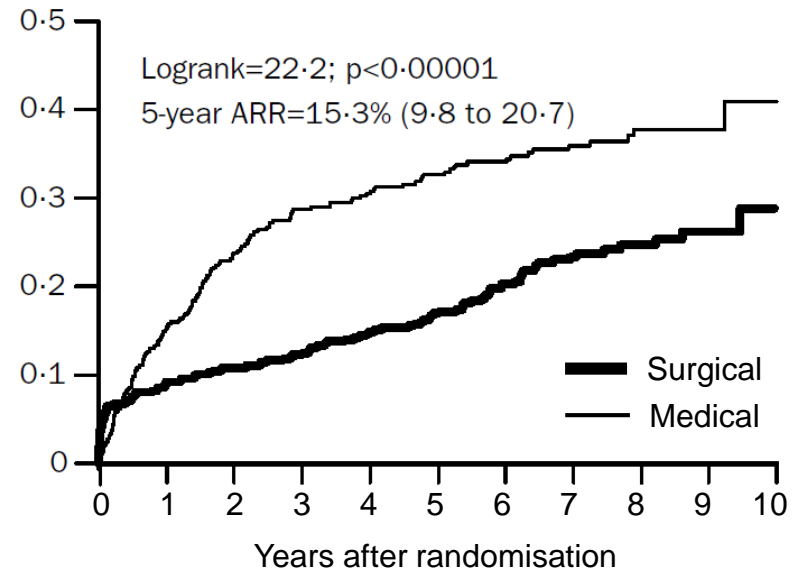
70–99% stenosis

Implications of heterogeneity

- Heterogeneity of treatment effect within the trial sample

Look for the *test for interaction*
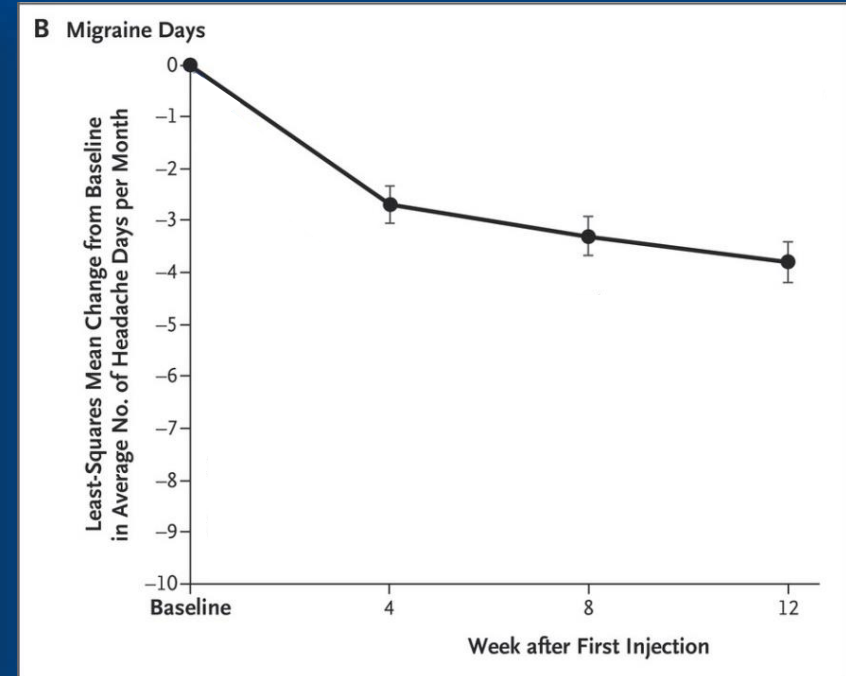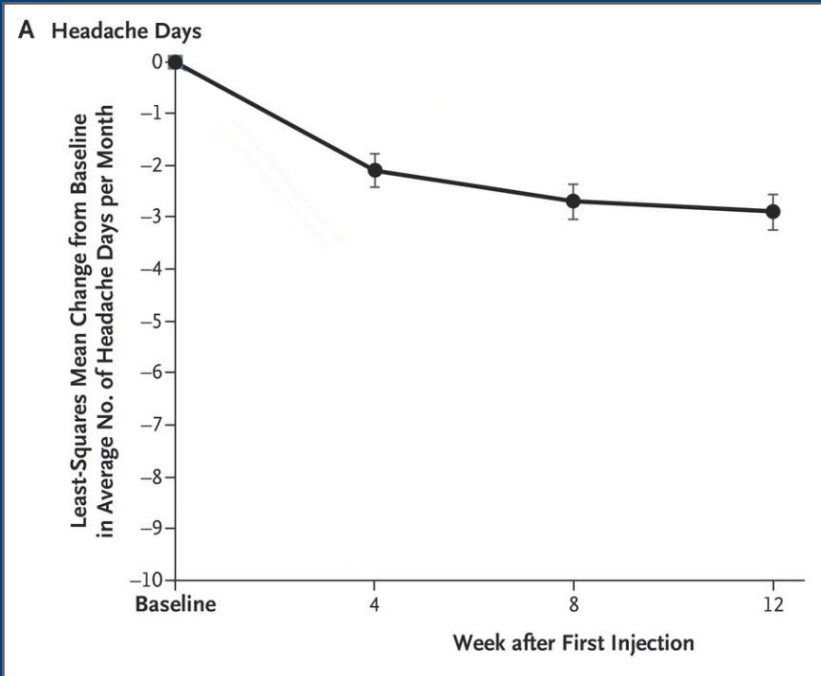
Implications of heterogeneity

- Heterogeneity of treatment effect within the trial sample
- Estimates of population parameters
    - Biased estimate of mean treatment effect
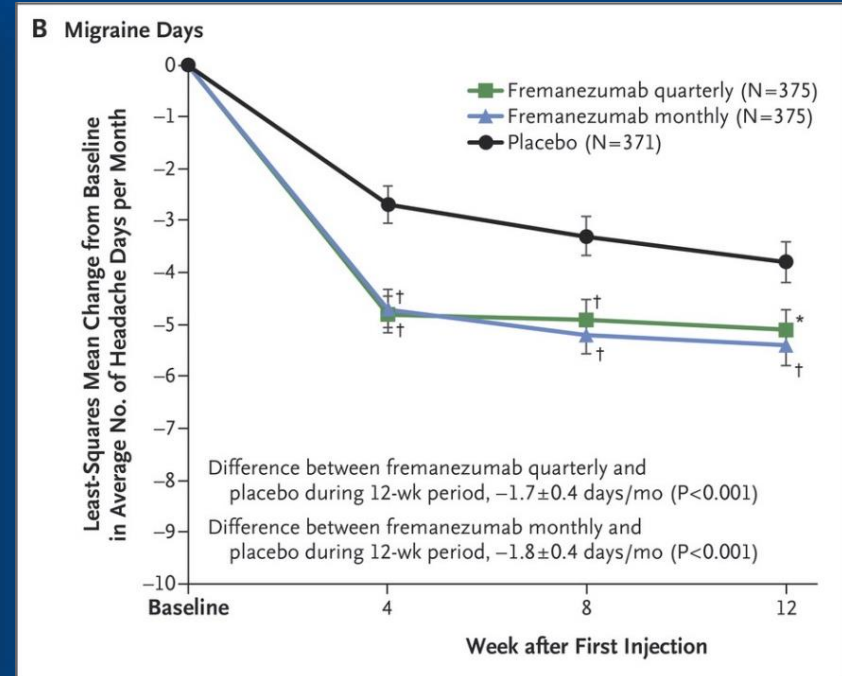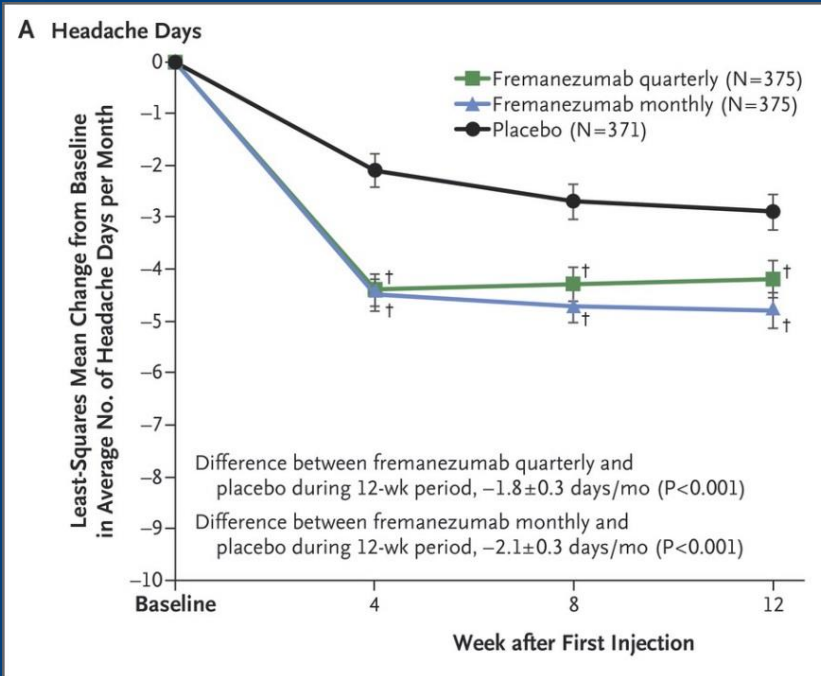   - Underrepresentation of population hetereogeneity

Clinical trials: Reading between the lines

1. Multiplicity
2. **Heterogeneity**

A new agent to treat migraine

A new agent to treat migraine

Clinical trials: Reading between the lines

1. Multiplicity
2. Heterogeneity
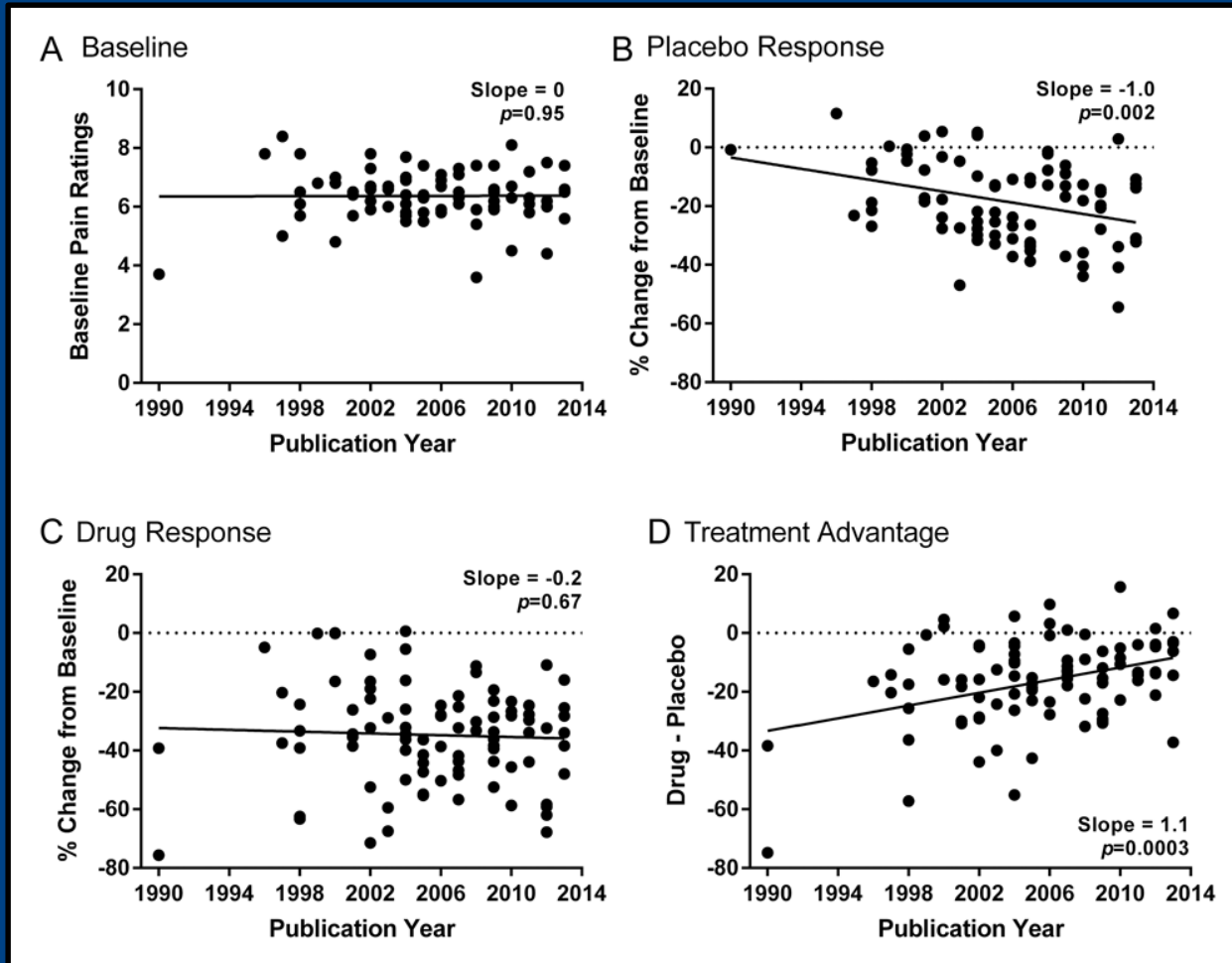3. **'Placebo' effects**

Phenomena contributing to 'placebo' effects

- **Hawthorne effects**
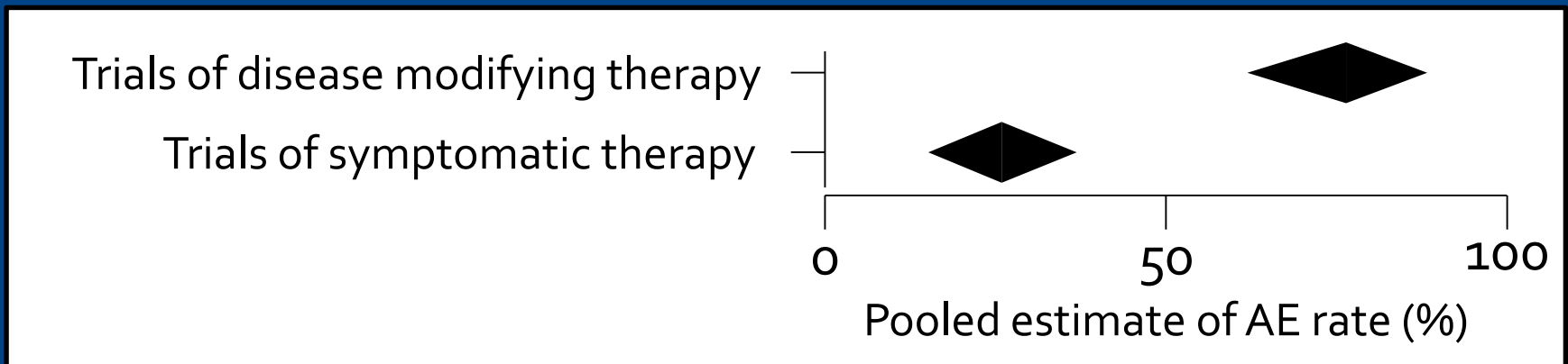- Expectation effects
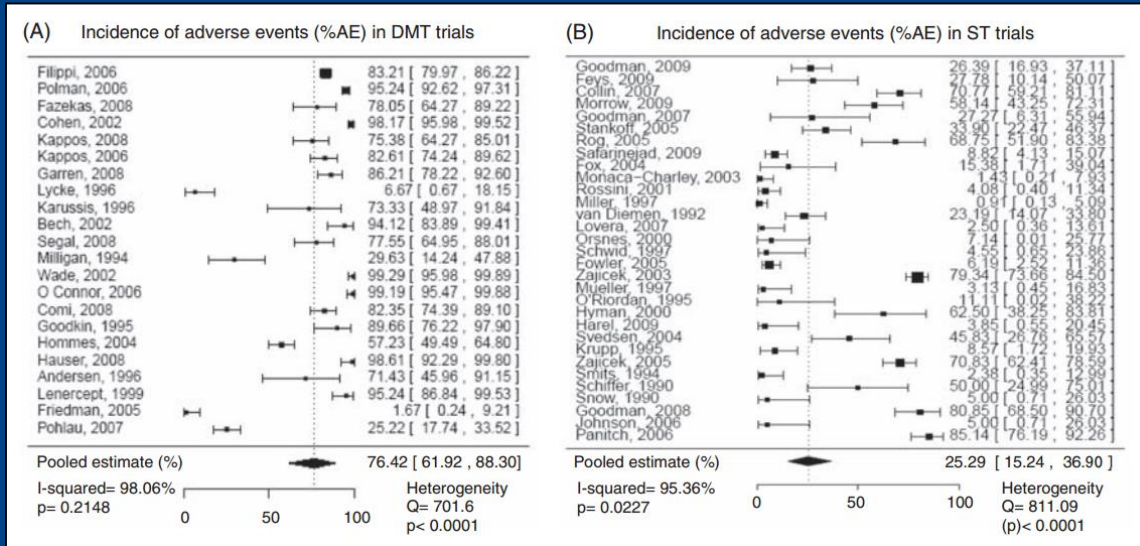 - Placebo effects
  - Nocebo effects

Phenomena contributing to 'placebo' effects

■ Hawthorne effects

■ **Expectation effects**
  - **Placebo effects**
  - **Nocebo effects**

Changing behaviour of the placebo group

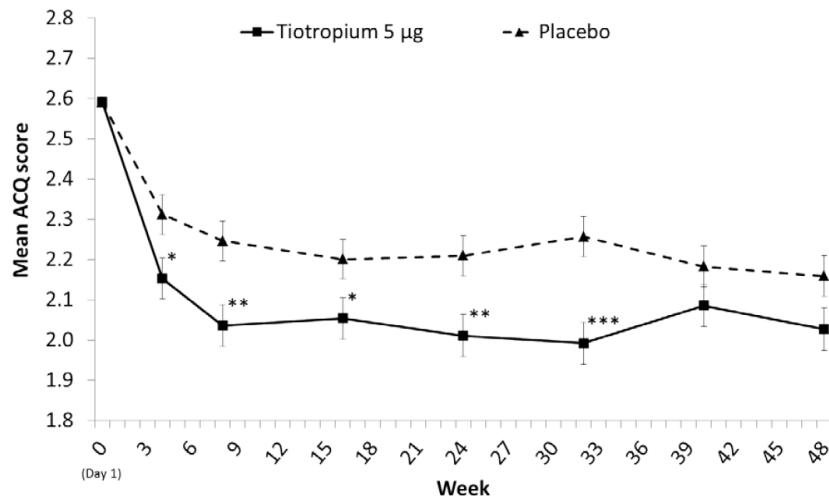# Nocebo effects in multiple sclerosis
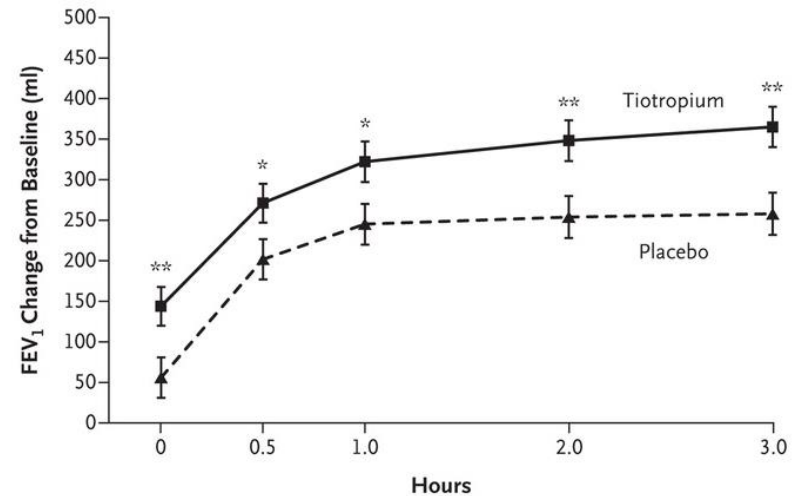
Why even bother with drugs…?

ORIGINAL ARTICLE

Tiotropium in Asthma Poorly Controlled with Standard Combination Therapy

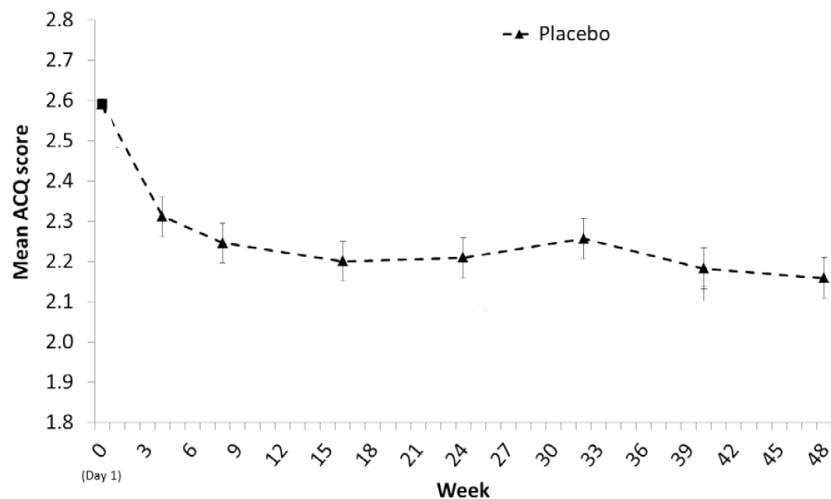Asthma control questionnaire
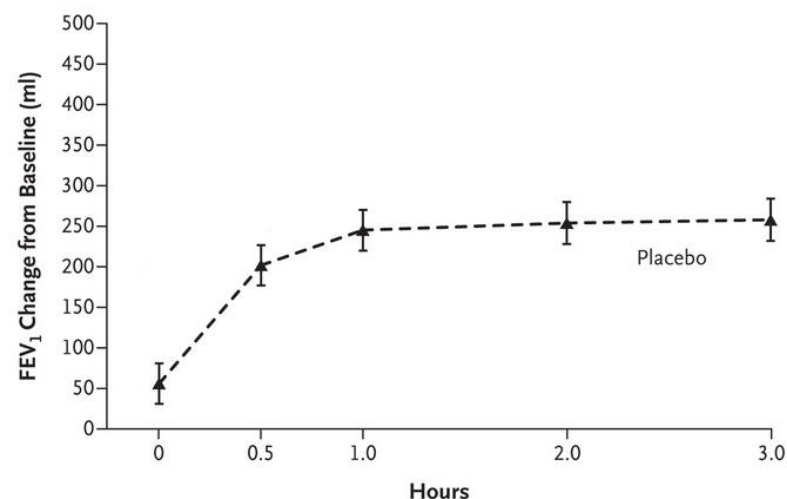
Change in FEV1

Why even bother with drugs…?

ORIGINAL ARTICLE

Placebo in Asthma Poorly Controlled with Standard Combination Therapy

Asthma control questionnaire

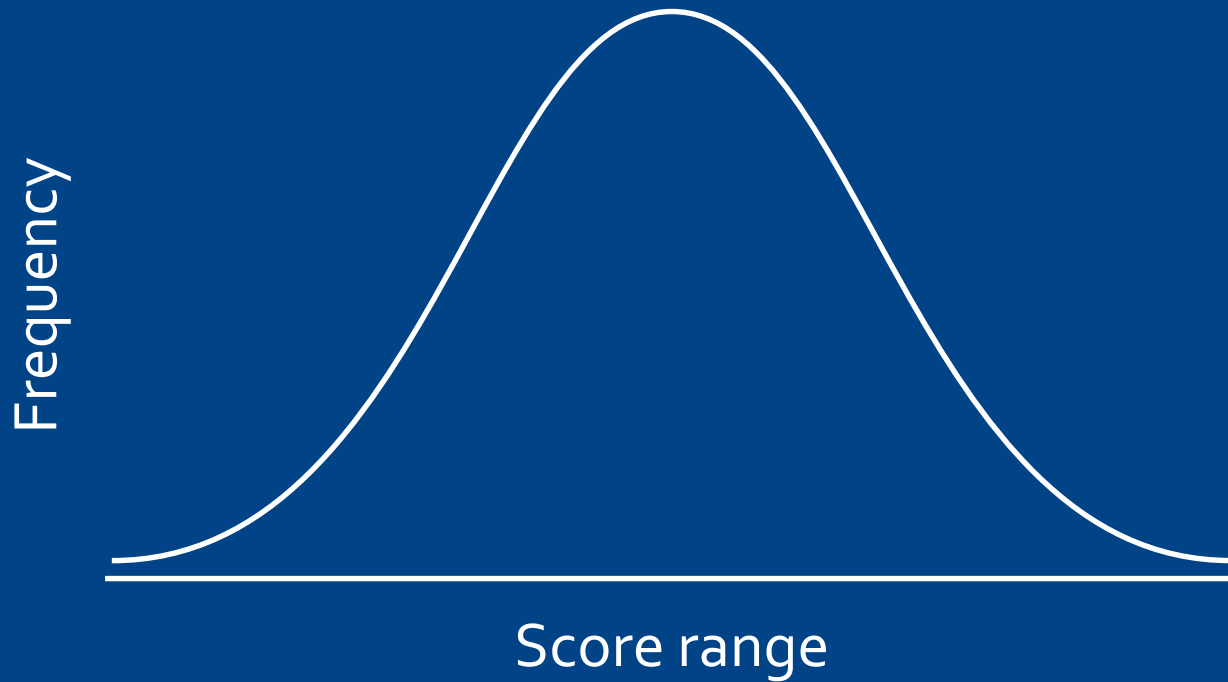Change in FEV1

Regression to the mean

PATIENT CHARACTERISTICS

Eligible patients were between the ages of 18 and 75 years and had a 5-year or longer history of asthma that was diagnosed before the age of 40 years. Patients were required to have a score of 1.5 or higher on the Asthma Control Questionnaire 7 (ACQ-7), which consists of seven questions, each scored on a range from 0 (no impairment) to 6 (maximum impairment), with a minimal clinically important difference of 0.5 units[8]; and to have persistent airflow limitation, which was defined as a post-bronchodilator forced expiratory volume in 1 second (FEV_1) of 80% or less of the predicted value[9] and 70% or less of forced vital capacity (FVC) 30 minutes after the inhalation of four puffs of 100 μg of salbutamol or of albuterol at the screening visit, despite therapy with inhaled glucocorticoids (≥800 budesonide or the equivalent) and LABAs. Patients were required to have had at least one exacerbation that was treated with systemic glucocorticoids in the previous year and to be either lifelong nonsmokers or to have a smoking history of fewer than 10 pack-years, with no smoking in the year before enrollment.
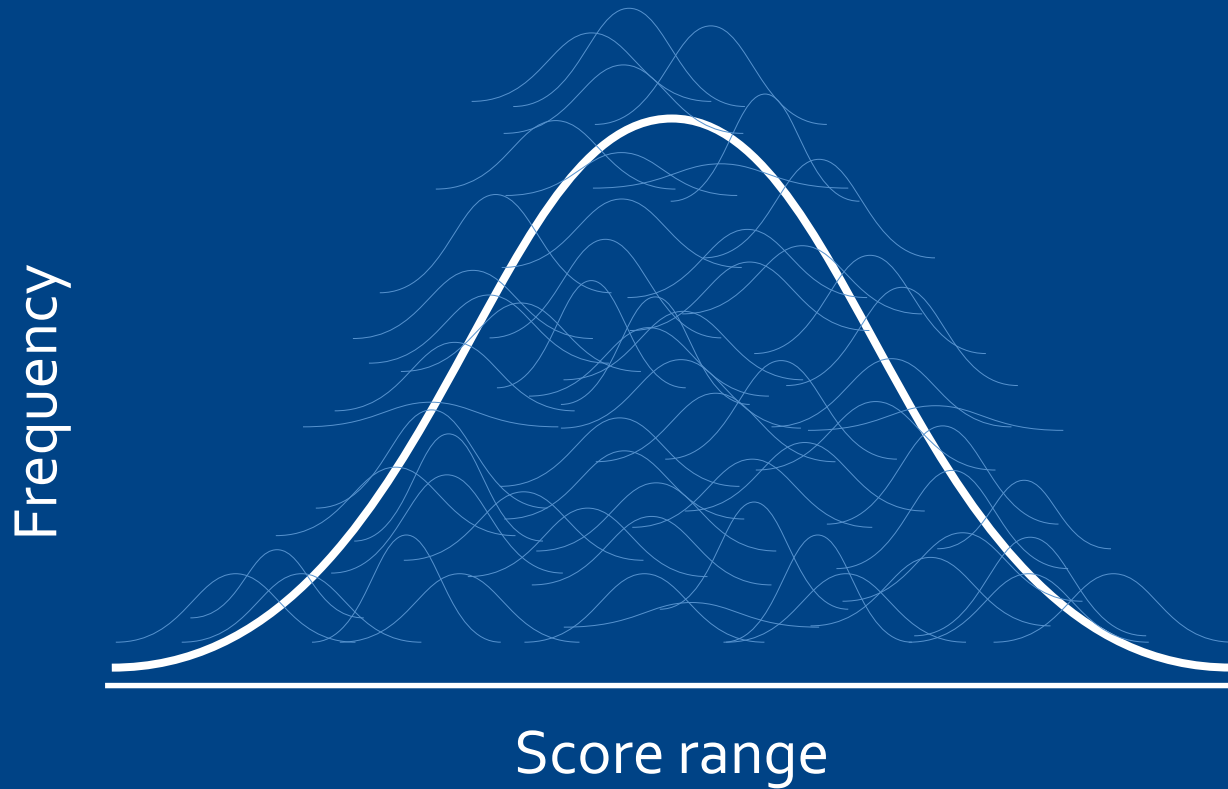
40 years. Patients were required to have a score of 1.5 or higher on the Asthma Control Questionnaire 7 (ACQ-7), which consists of seven questions,

defined as a post-bronchodilator forced expiratory volume in 1 second (FEV_1) of 80% or less of the predicted value[9] and 70% or less of forced vital capacity (FVC) 30 minutes after the inhalation
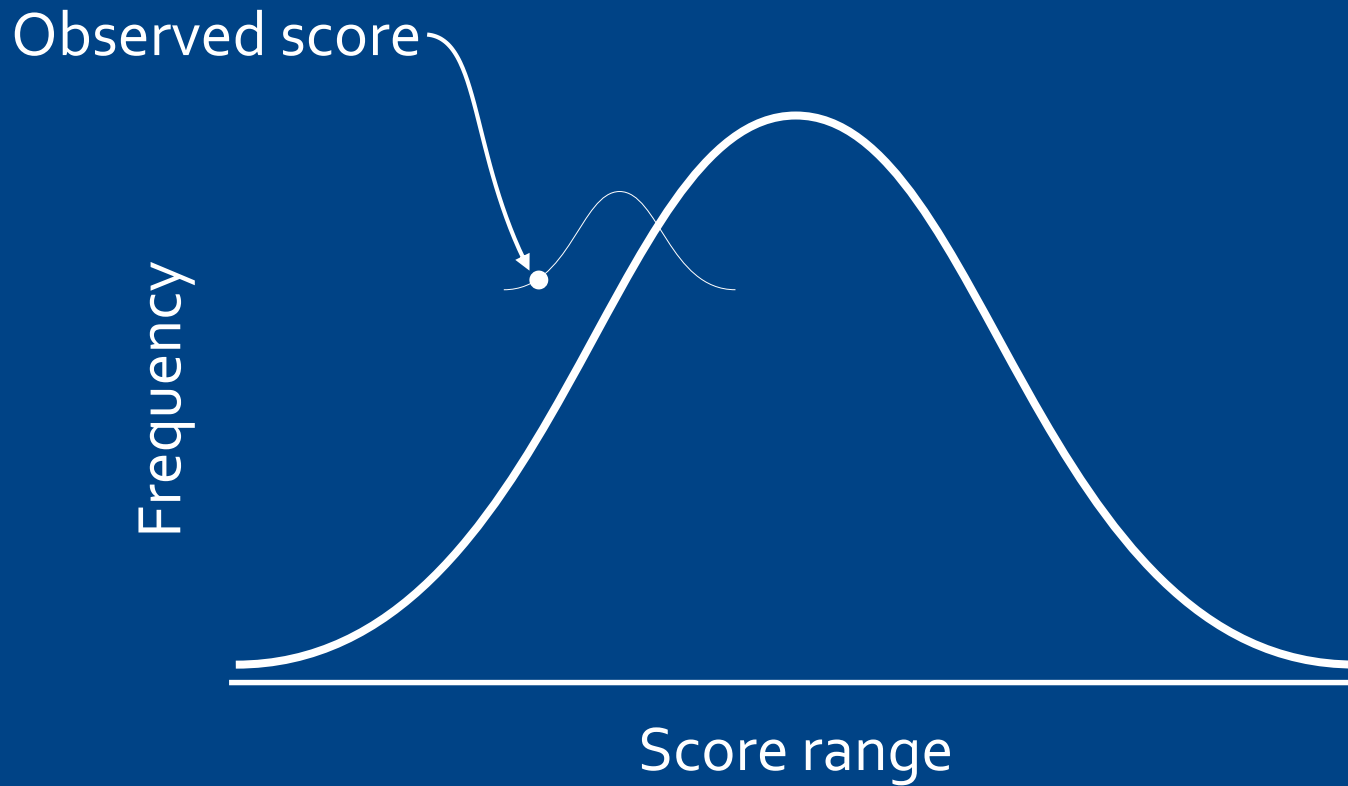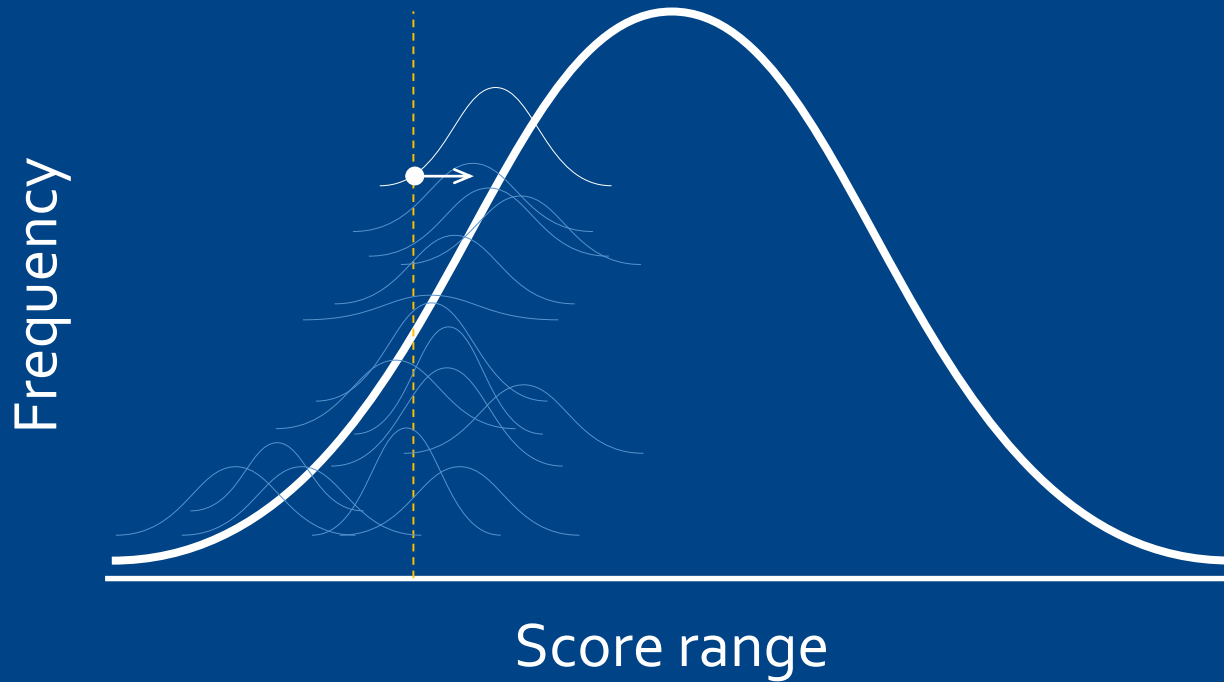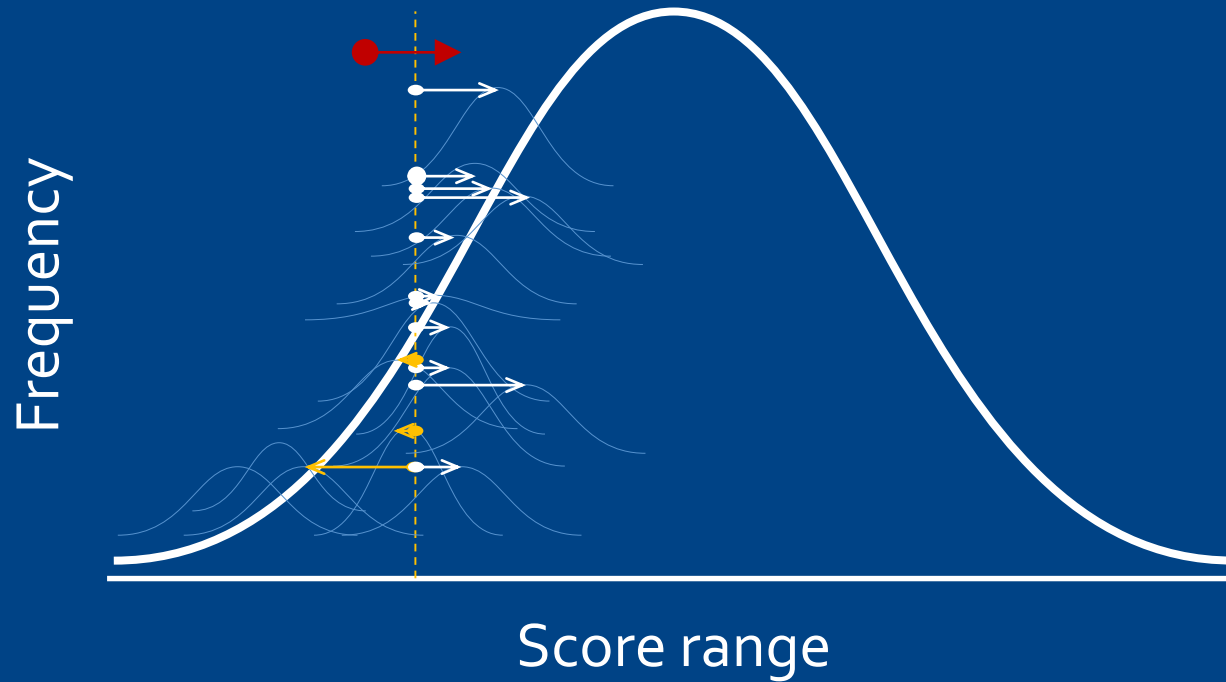
Regression to the mean

Frequency

Score range

Regression to the mean

Regression to the mean

Observed score

Frequency

Score range

Regression to the mean

Frequency

Score range

Regression to the mean

Frequency

Score range

Phenomena contributing to 'placebo' effects

■ Hawthorne effects

■ Expectation effects

   – Placebo effects

   – Nocebo effects

■ **Regression to the mean**

Regression to the mean

- A purely statistical phenomenon
- Occurs whenever a population is:
  - Asymmetrically sampled
  - Measured more than once
  - Correlation between the measurements is imperfect
- Best handled by comparing to a placebo group

Summary

- The problems of **multiplicity** are serious and all-pervasive
- Understand the implications of **heterogeneity** of treatment effect
- Understand the factors that contribute to **'placebo' effects**