Study design and sample size

Peter J Diggle

School of Health and Medicine,
Lancaster University
and
Department of Biostatistics,
Johns Hopkins University School of Public Health

Find me in room B7a, Faraday, or call 594622

A Simple Comparative Trial

- **Formoterol (F) and Salbutamol (S) are two drugs used to treat chronic asthma.**
- **trial was carried out to compare efficacy of F and S:**
  - **asthmatic children recruited;**
  - **Peak Expiratory Flow (PEF) measured after administration of F;**
  - **PEF measured after administration of S;**
  - **one week between two measurements (ordering chosen at random)**

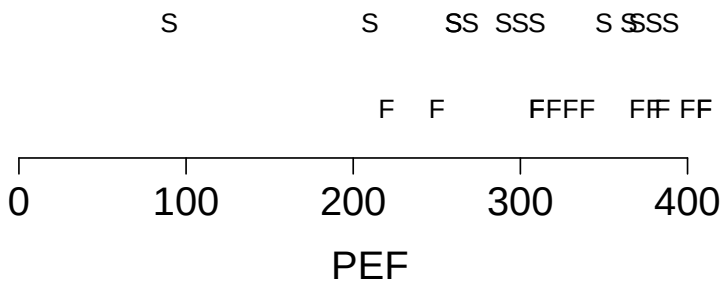
Data from the Asthma Trial

Formoterol (sorted)

220 250 310 310 320 330 340 370 380 385 400 410 410

Salbutamol (sorted)

90 210 260 260 270 290 300 310 350 365 370 380 390

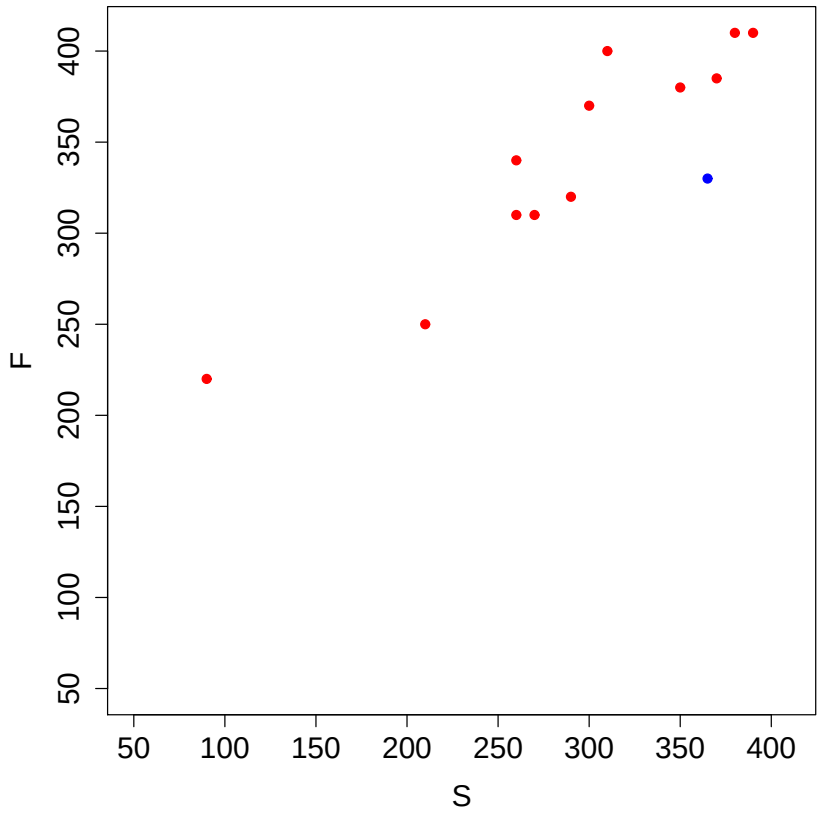


- **one very small S result**
- **considerable overlap between F and S results**
- **comparison between F and S inconclusive?**

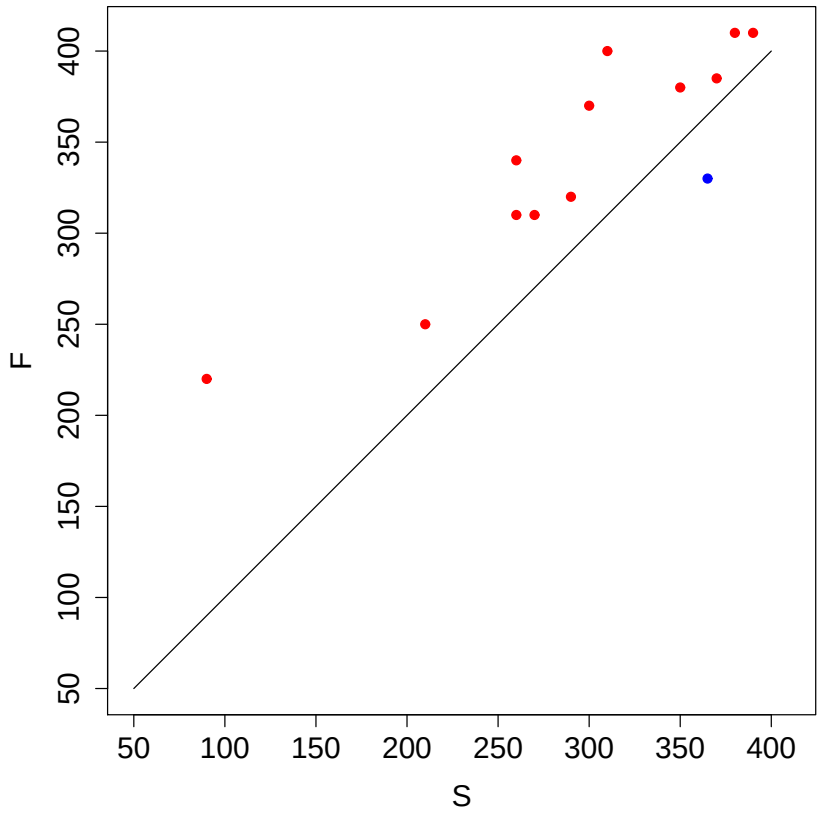
Data from the Asthma Trial: continued

Child	Drug F	Drug S	Child	Drug F	Drug S
1	310	270	8	385	370
2	310	260	9	400	310
3	370	300	10	410	380
4	410	390	11	320	290
5	250	210	12	340	260
6	380	350	13	220	90
7	330	365			

If a Picture Paints a Thousand Words ...



If a Picture Paints a Thousand Words ...



Comparison between F and S now very clear

What does this example tell us?

Study design is important.

- Decide what comparisons are of interest
  - F vs S
- Design the experiment to eliminate extraneous sources of variation
  - pairing eliminates variation between children
  - randomisation eliminates time-order effect

Testing and estimation

Q1. **Testing:** **do the data provide evidence that there is a difference (however small) between the average response to F and to S?**

Q2. **Estimation:** **what reasonable range of values do the data indicate that we should assign to the difference between the average response to F and to S?**

A1. **Reject the null hypothesis of no difference between the true average responses to F and to S if the probability that a difference as large as the observed difference could arise by chance is less than 0.05 (5%)**

A2. **Estimate the true difference as a range of values chosen in such a way that the range will include the true value with probability 0.95 (95%)**

Testing and estimation: the asthma data

- Calculate $D = F - S$ for each child

  $40 \quad 50 \quad 70 \quad 20 \quad 40 \quad 30 \quad -35 \quad 15 \quad 90 \quad 30 \quad 30 \quad 80 \quad 130$

- Sample size, mean and standard deviation

  $n = 13 \qquad \bar{D} = 45.4 \qquad SD = 40.59$

- Standard error, $SE = SD/\sqrt{n} = 40.59/\sqrt{13} = 11.26$

  - testing: $Z = \bar{D}/SE$, reject null hypothesis if $|Z| > 2$

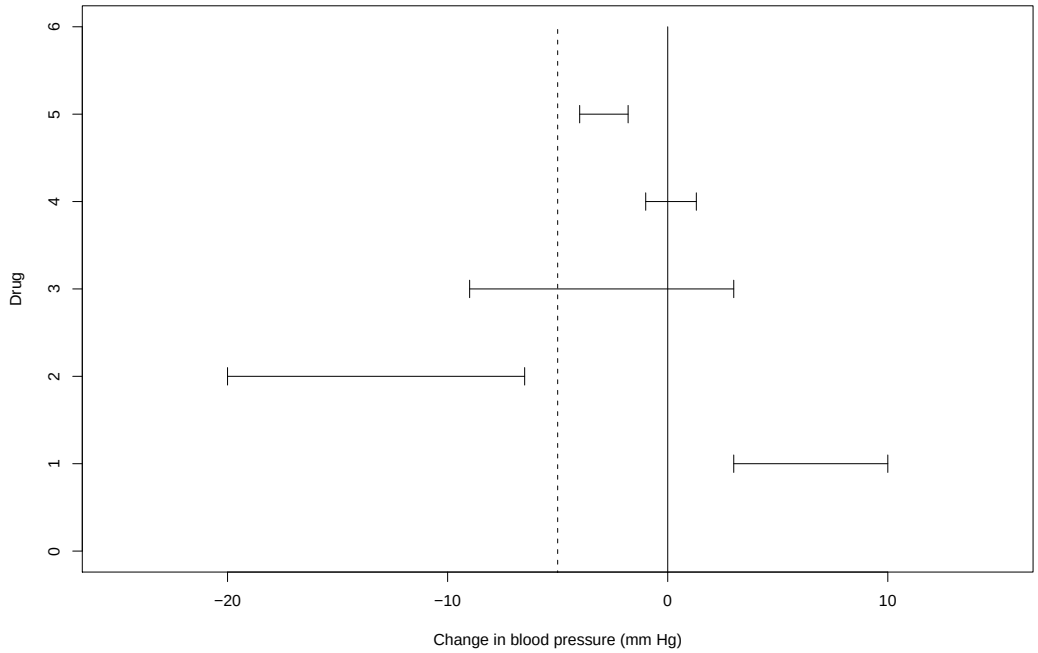
    $$Z = 45.4/11.26 = 4.03$$

  - estimation: $\bar{D} \pm 2 \times SE$ (95% confidence interval)

    $$45.4 \pm 2 \times 11.26 = (22.9, 67.9)$$

Statistically or clinically significant?

Summarising the results of many experiments:

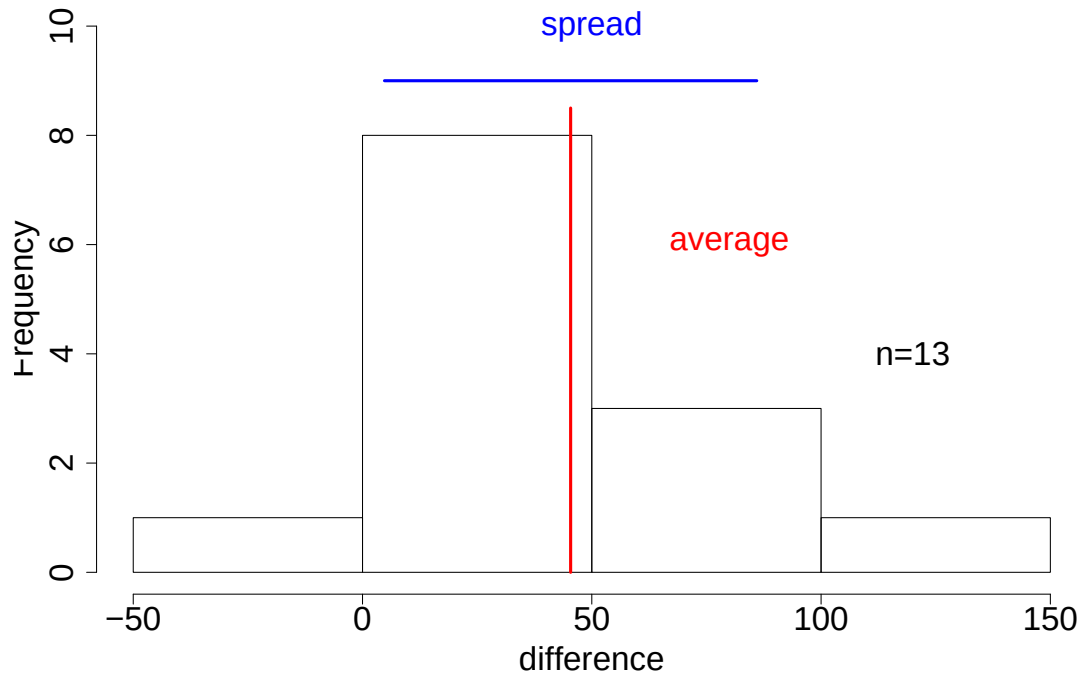


- which experiments gave a statistically significant reduction in blood pressure?
- which experiments gave a clinically significant reduction?

What affects precision?

Another look at the asthma data, differences between results on F and on S, for each of 13 children:

40 50 70 20 40 30 −35 15 90 30 30 80 130



Testing and estimation

(continued)

The following statements are equivalent:

- **the $(100 - p)\%$ confidence interval for the true mean excludes zero**
- **the sample mean is significantly different from zero at the $p\%$ significance level**

A typical statement about a clinical trial

The trial was designed to achieve a power of 80% to detect a difference of 10 mmHg between mean systolic blood pressure in the two treatment arms, using a 5% significance test

- **80% is the chance of declaring a significant difference when the true difference is 10 units**
- **10 units is the clinically significant difference**
- **5% is the chance of declaring a significant difference when there is in fact NO difference between the two treatments**

Significance, power and sample size

1. the significance level of a statistical test is the probability that you will (incorrectly) reject a true hypothesis

2. the power of a statistical test is the probability that you will (correctly) reject a false hypothesis

No free lunch

The only way to increase power when using a small significance level is to reduce the standard error, $SE = SD/\sqrt{n}$, which you can do in two ways:

- come up with a better design for the study (reduce SD)
- use a larger sample size (increase n)

How do you calculate power or sample size?

1. choose a statistical significance level (5% is conventional)

2. specify the clinically significant difference between your two treatment arms

3. specify the standard deviation, (SD) of your outcome variable (this is usually the tricky part!)

- **ideally, use data from a pilot study**
- **failing that, roughly 95% of the data should lie within plus or minus two SD's of the mean, so try to imagine what range you might expect to see for your primary outcome variable**

Example: asthma data

40 50 70 20 40 30 −35 15 90 30 30 80 130

- The values lie between −35 and 130, a range of 165, so a crude guess would be $SD = 165/4 = 41.25$
- In fact, $SD = 40.59$

  (it doesn't usually work as well as this!)

4. Go to `http://statpages.org/`

- click on Power, sample size and experimental design
- click on power/sample size calculator
- select the analysis you want (ask for advice first if need be)

You can also download the software to your own computer.

In summary

1. Be very clear about:
   (a) what question you are trying to answer
   (b) what primary outcome variable is best equipped to address this question
   (c) what difference between a new and existing treatment would make the improvment in the primary outcome clinically worthwhile
2. good design of clinical studies is at least as important as good data analysis
3. precision of estimation is at least important as power
4. 5% significance, 95% confidence and 80% power are widely accepted conventions, but no more than that

Slides available at:

`www.lancs.ac.uk/staff/diggle/CTM_lecture.pdf`