The Challenge of "Small Data"
Rare Diseases and Ways to Study Them

Stephen Senn

LUXEMBOURG INSTITUTE OF **HEALTH**

RESEARCH DEDICATED TO LIFE

Acknowledgements

Many thanks for the invitation

This work is partly supported by the European Union's 7th Framework Programme for research, technological development and demonstration under grant agreement no. 602552. "IDEAL"

Some of this work is joint with Artur Araujo and Sonia Leite in my group and Steven Julious at Sheffield University

Outline

- The roots of modern statistics
  - Small data
  - Careful design of experiments
- Some examples of problems with judging causality from associations in the health care field
- Rare diseases
- N-of-1 trials as a possible solution (in some cases)
- Some statistical issues

Warning

I am a statistician

This means that whatever you believe in, I don't

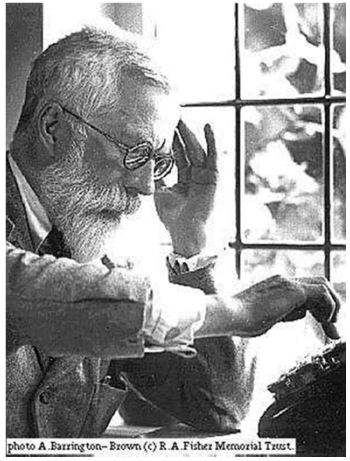
William Sealy Gosset 1876-1937

- Born Canterbury 1876
- Educated Winchester and Oxford
- First in mathematical moderations 1897 and first in degree in Chemistry 1899
- Starts with Guinness in 1899 in Dublin
- Autumn 1906-spring 1907 with Karl Pearson at UCL
- 1908 publishes ‘The probable error of a mean’
- First method available to judge ‘significance’ in small samples

Ronald Aylmer Fisher 1890-1962

- Most influential statistician ever
- Also major figure in evolutionary biology
- Educated Harrow and Cambridge
- Statistician at Rothamsted agricultural station 1919-1933
- Developed theory of small sample inference and many modern concepts
  - Likelihood, variance, sufficiency, ANOVA
- Developed theory of experimental design
  - Blocking, Randomisation, Replication,

Characteristics of development of statistics in the first half of the 20th century



- Numerical work was arduous and long
  - Human computers
  - Desk calculators
  - Careful thought as to how to perform a calculation paid dividends
- Much development of inferential theory for small samples
- Design of experiments became a new subject in its own right developed by statisticians
  - Orthogonality
    - Made calculation easier (eg decomposition of variance terms in ANOVA)
    - Increased efficiency
  - Randomisation
    - "Guaranteed" properties of statistical analysis
    - Dealt with hidden confounders
  - Factorial experimentation
    - Efficient way to study multiple influences

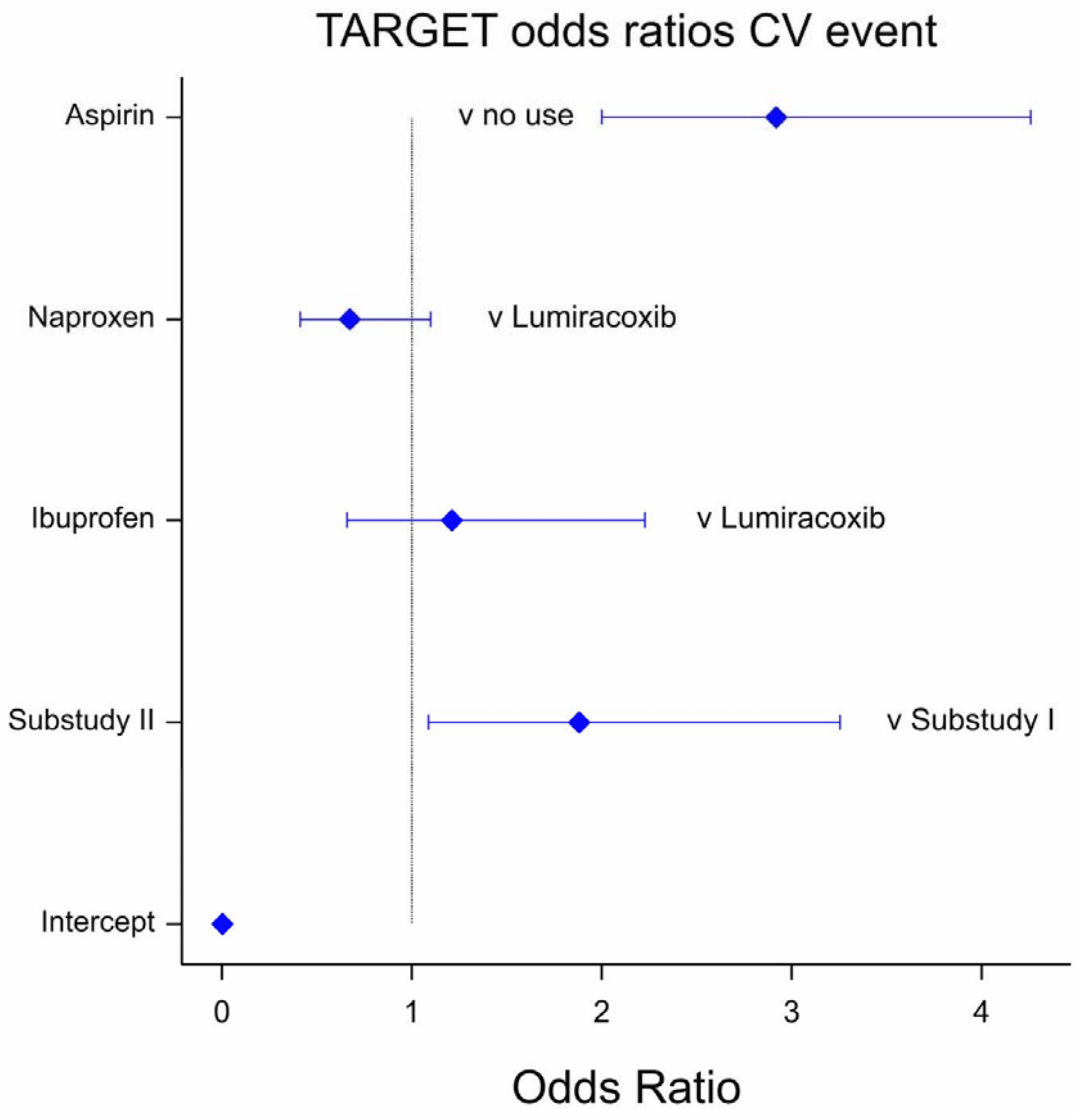
A big data analyst is an expert at reaching misleading conclusions with huge data sets, whereas a statistician can do the same with small ones

TARGET study

- Trial of more than 18,000 patients in osteoarthritis over one year or more
- Two sub-studies
  - Lumiracoxib v ibuprofen
  - Lumiracoxib v naproxen
- Stratified by aspirin use or not
- Has some features of a randomised trial but also some of a non-randomised study

	Sub-Study 1		Sub Study 2	
Demographic Characteristic	**Lumiracoxib n = 4376**	**Ibuprofen n = 4397**	**Lumiracoxib n = 4741**	**Naproxen n = 4730**
Use of low-dose aspirin	975 (22.3)	966 (22.0)	1195 (25.1)	1193 (25.2)
History of vascular disease	393 (9.0)	340 (7.7)	588 (12.4)	559 (11.8)
Cerebro-vascular disease	69 (1.6)	65 (1.5)	108 (2.3)	107 (2.3)
Dyslipidaemias	1030 (23.5)	1025 (23.3)	799 (16.9)	809 (17.1)
Nitrate use	105 (2.4)	79 (1.8)	181 (3.8)	165 (3.5)

	Model Term		
Demographic Characteristic	**Sub-study (DF=1)**	**Treatment given Sub-study (DF=2)**	**Treatment (DF=2)**
Use of low-dose aspirin	< 0.0001	0.94	0.0012
History of vascular disease	< 0.0001	0.07	<0.0001
Cerebro-vascular disease	0.0002	0.93	0.0208
Dyslipidaemias	<0.0001	0.92	<0.0001
Nitrate use	< 0.0001	0.10	<0.0001



	Statistic	
Outcome Variables	**Deviance**	**P-Value**
Total of discontinuations	13.61	0.0002
CV events	2.92	0.09
At least one AE	1.73	0.19
Any GI	21.31	<0.0001
Dyspepsia	47.34	< 0.0001

Data Filtering Some Examples

- Oscar winners lived longer than actors who didn't win an Oscar
- A 20 year follow-up study of women in an English village found higher survival amongst smokers than non-smokers
- Transplant receivers on highest doses of cyclosporine had higher probability of graft rejection than on lower doses
- Left-handers observed to die younger on average than right-handers
- Obese infarct survivors have better prognosis than non-obese

Moral

- What you don’t see can be important
- For some purposes just piling on data does not really help
- What helps are
  - Careful design
  - Thinking!

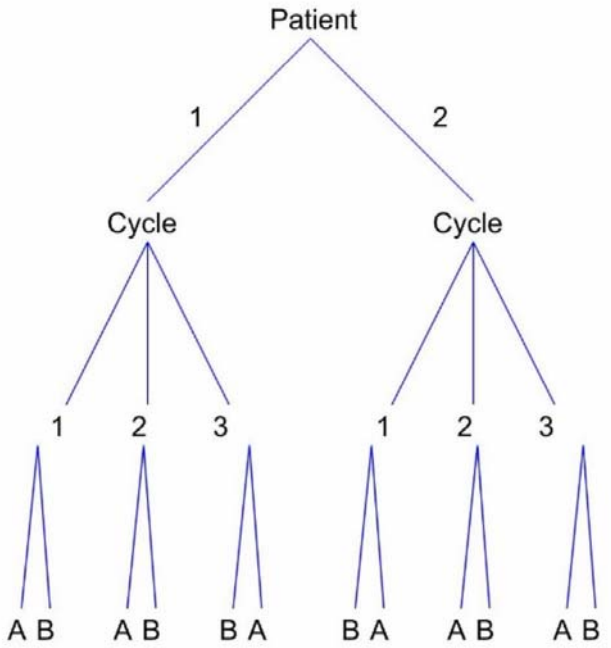
We tend to believe “the truth is in there”, but sometimes it isn’t and the danger is we will find it anyway

Rare Diseases

- As far as the Food and Drug Administration is concerned anything that affects fewer than 300,000 people in the US
- However many diseases are much rarer than this
- But there are at least 7,000 rare diseases
- Thus the total number of persons effected is considerable

N-of-1 studies

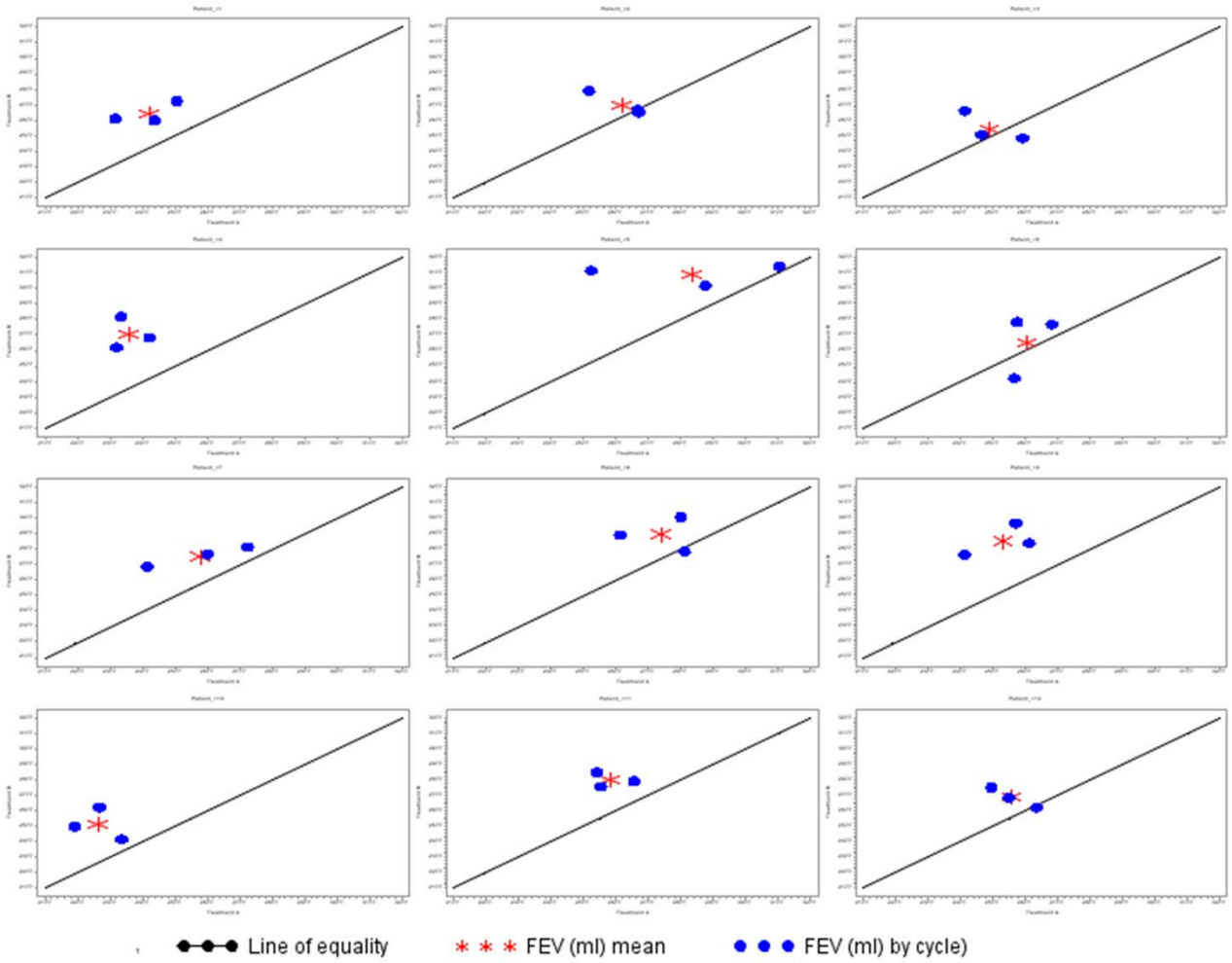
- Studies in which patients are repeatedly randomised to treatment and control
- Increased efficiency because
  - Each patient acts as own control
  - More than one judgement of effect per patient
- However, only possible for chronic diseases
- Possible randomisation in *k* cycles of treatment
- Implies 2^k possible sequences



A simulated example

- Twelve patients suffering from a chronic rare respiratory complaint
  - For example cystic fibrosis
- Each patient is randomised in three pairs of periods, comparing two treatments A and B
- Adequate washout is built in to the design
- Thus we have 12 x 3 x 2 = 72 observations altogether
- Efficacy is measured using forced expiratory volume in one second (FEV_1) in ml
- How should we analyse such an experiment?

Trellis Plot of FEV per cycle grouped in patients



Possible objectives of an analysis

- **Is one of the treatments better?**
  - **Significance tests**
- What can be said about the average effect in the patients that were studied?
  - Estimates, confidence intervals
- **What can be said about the average effects in future patients?**
- **What can be said about the effect of a given patient in the trial?**
- What can be said about a future patient not in the trial?

Two different philosophies

Randomisation philosophy

- The patients in a clinical trial are taken as fixed
- The population about which inference is made is all possible randomisations
- The patients don't change, only the pattern of assignments of treatments change

Sampling philosophy

- The patients are regarded as a sample from some possible population of patients
- This is usually handled by adding error terms corresponding to various components of variance
- This approach is much more common

Is one of the treatments better? Significance tests

Rothamsted School

- Leading statisticians such as Fisher, Yates, Nelder, Bailey
- Developed analysis of variance not in terms of linear models but in terms of symmetry
- High point was John Nelder's theory of general balance (1965)

General Balance

1) Establish and define block structure
2) Establish and define treatment structure
3) Given randomisation the analysis then follows automatically

Here the block structure is

Patient/Cycle GenStat®

`Patient(Cycle)` SAS®

The treatment structure is

`Treatment`

The general balance approach

```
BLOCKSTRUCTURE Patient/Cycle
TREATMENTSTRUCTURE Treatment
ANOVA[FPROBABILITY=YES;NOMESSAGE=residual] Y
```

.

Analysis of variance

Variate: FEV_1 (mL)

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Patient stratum	11	1458791.	132617.	10.04	
Patient.Cycle stratum	24	316885.	13204.	1.04	
Patient.Cycle.*Units* stratum					
Treatment	1	641089.	641089.	50.57	<.001
Residual	35	443736.	12678.		
Total	71	2860501.			

Comparing two models

The first is without a patient by treatment interaction

Source	DF	Type II SS	Mean Square	F Value	Pr > F
patient	11	1458791.444	132617.404	10.46	<.0001
patient*cycle	24	316884.667	13203.528	1.04	0.4479
Treatment	1	641089.389	641089.389	50.57	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
mean effect	188.722222	26.5394469	7.11	<.0001

NB Analysis with proc glm of SAS®

The second is with a patient by treatment interaction

Source	DF	Type II SS	Mean Square	F Value	Pr > F
patient	11	1458791.444	132617.404	11.20	<.0001
patient*cycle	24	316884.667	13203.528	1.11	0.3960
Treatment	1	641089.389	641089.389	54.13	<.0001
patient*Treatment	11	159516.278	14501.480	1.22	0.3241

Parameter	Estimate	Standard Error	t Value	Pr > |t|
mean effect	188.722222	25.6498562	7.36	<.0001

Any damn fool can analyse a clinical trial and frequently does

Two more difficult questions

The average effects in future patients?

- This may require a mixed effects model
- Allow for a random treatment-by-patient interaction
  - The possibility that there may be variation in the effect from patient to patient
- Strong assumptions my be involved

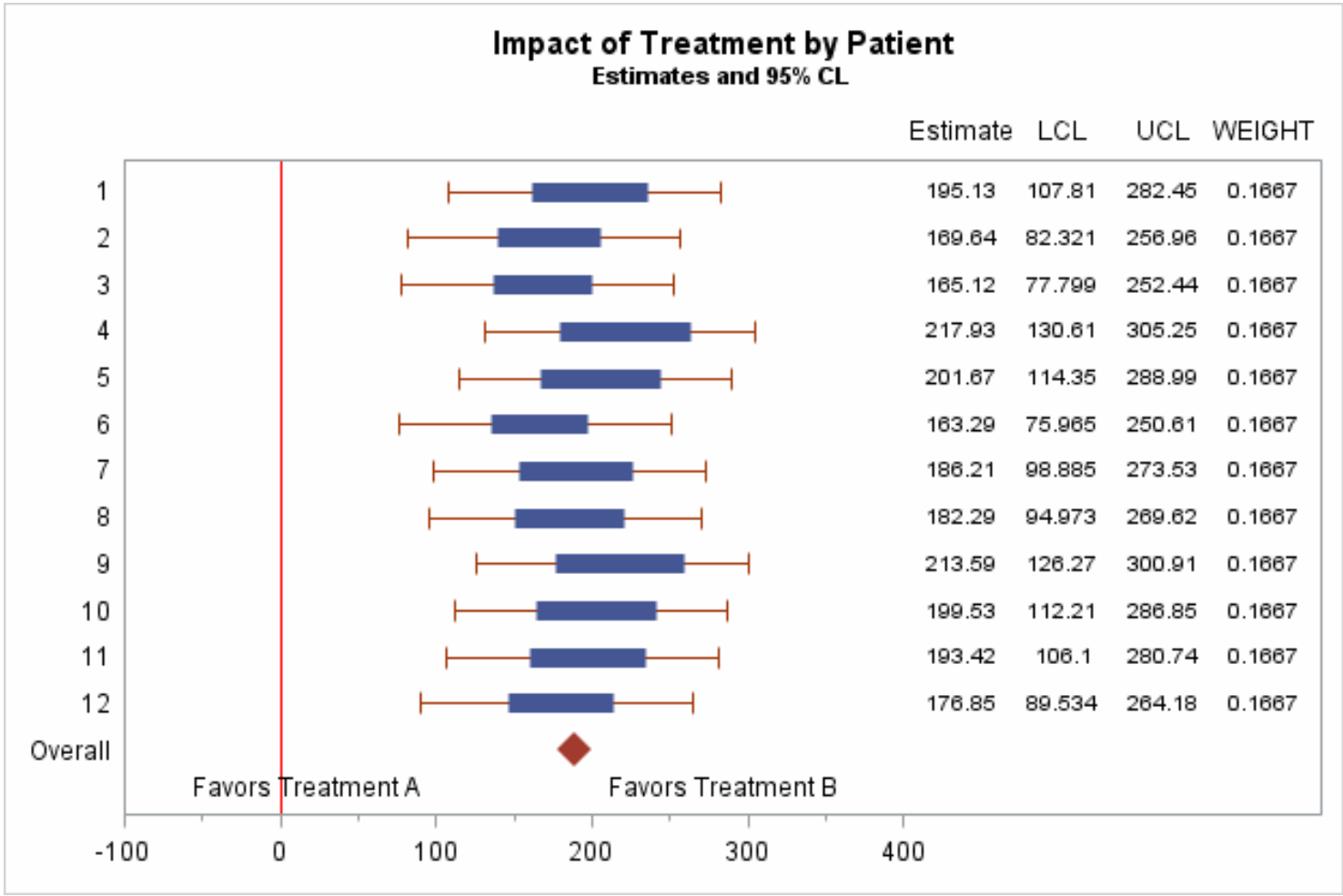
The average effect for a given patient?

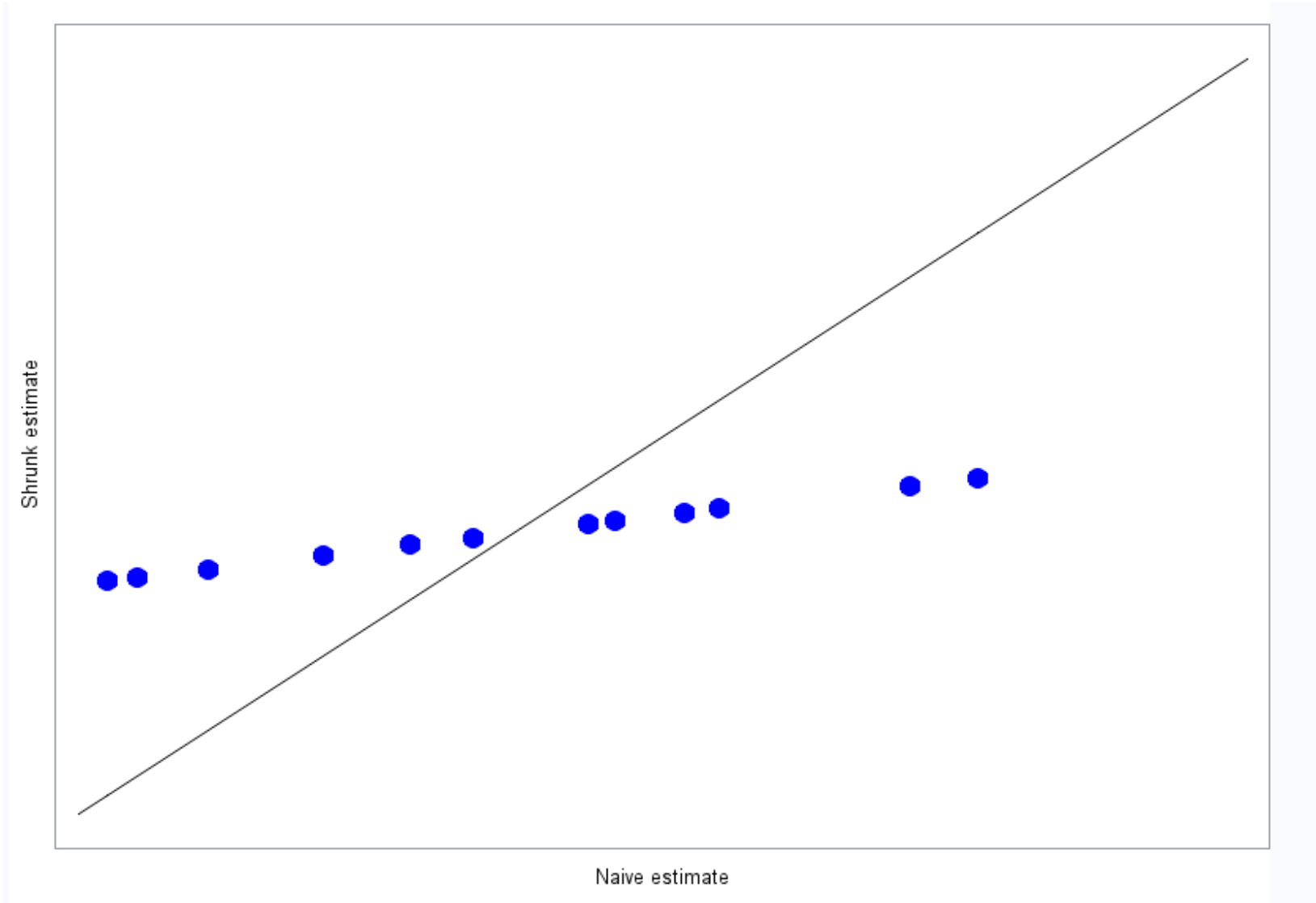
- The same random effect model can be used to predict long-term average effects for patients in the trial
- A weighted estimate is used whereby the patient's only results are averaged with the general result

Analysis using `proc mixed` of SAS®

Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
mean effect	188.72	28.3838	11	6.65	<.0001	0.05	126.25	251.19
treatment effect 1	195.13	44.5523	11	4.38	0.0011	0.05	97.0706	293.19
treatment effect 2	199.53	44.5523	11	4.48	0.0009	0.05	101.47	297.59
treatment effect 3	193.42	44.5523	11	4.34	0.0012	0.05	95.3592	291.48
treatment effect 4	176.85	44.5523	11	3.97	0.0022	0.05	78.7956	274.91
treatment effect 5	169.64	44.5523	11	3.81	0.0029	0.05	71.5834	267.70
treatment effect 6	165.12	44.5523	11	3.71	0.0035	0.05	67.0605	263.18
treatment effect 7	217.93	44.5523	11	4.89	0.0005	0.05	119.87	315.99
treatment effect 8	201.67	44.5523	11	4.53	0.0009	0.05	103.61	299.73
treatment effect 9	163.29	44.5523	11	3.67	0.0037	0.05	65.2269	261.35
treatment effect 10	186.21	44.5523	11	4.18	0.0015	0.05	88.1470	284.27
treatment effect 11	182.29	44.5523	11	4.09	0.0018	0.05	84.2353	280.35
treatment effect 12	213.59	44.5523	11	4.79	0.0006	0.05	115.53	311.65

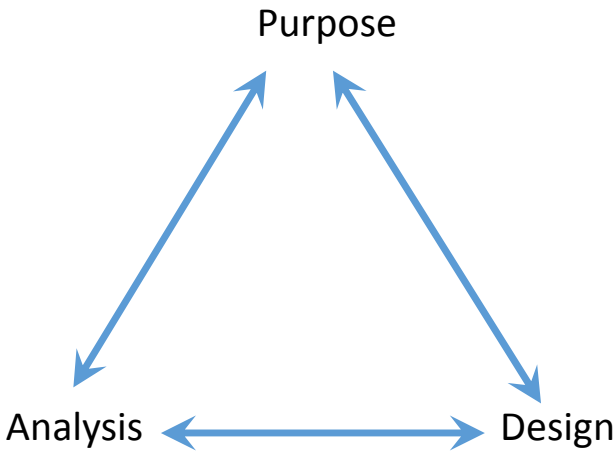
The difference between mathematical and applied statistics is that the former is full of lemmas whereas the latter is full of dilemmas


Impact of Treatment by Patient
Estimates and 95% CL
Estimate LCL UCL WEIGHT
1 195.13 107.81 282.45 0.1667
2 169.64 82.321 256.96 0.1667
3 165.12 77.799 252.44 0.1667
4 217.93 130.61 305.25 0.1667
5 201.67 114.35 288.99 0.1667
6 163.29 75.965 250.61 0.1667
7 186.21 98.885 273.53 0.1667
8 182.29 94.973 269.62 0.1667
9 213.59 126.27 300.91 0.1667
10 199.53 112.21 286.85 0.1667
11 193.42 106.1 280.74 0.1667
12 176.85 89.534 264.18 0.1667
Overall
Favors Treatment A
Favors Treatment B
-100 0 100 200 300 400


Shrunk estimate
Naive estimate

Morals

- There is still a role for small data analysis
- Design is crucial
- Analysis depends on purpose
- And also on design and vice versa
- Results depend on philosophical framework
- Calculation is difficult, yes, but so is thinking



To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of
RA Fisher