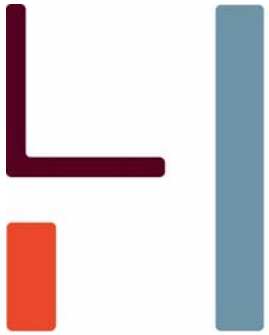
Combining evidence : Purpose is everything

Stephen Senn



LUXEMBOURG INSTITUTE OF **HEALTH**

RESEARCH DEDICATED TO LIFE

Acknowledgements

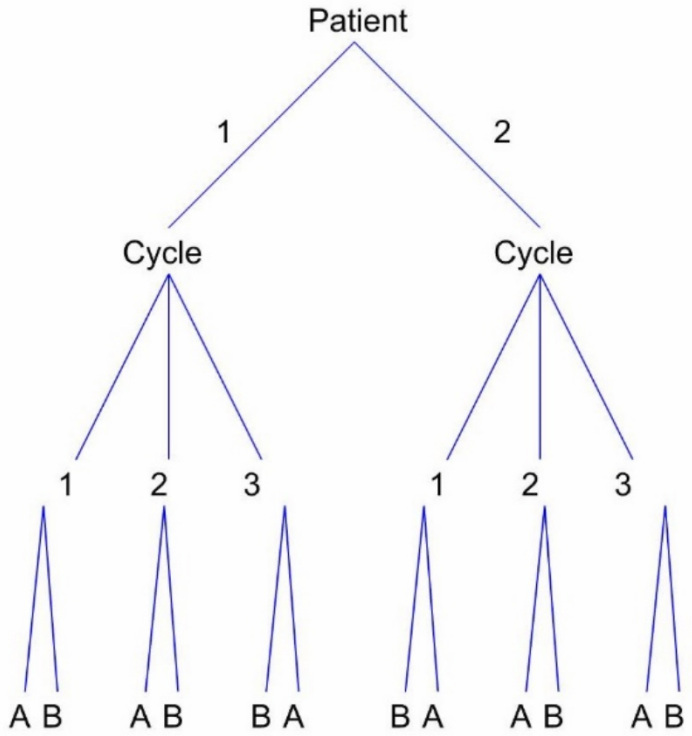
Many thanks for the invitation

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 & Steven Julious at Sheffield University

N-of-1 studies

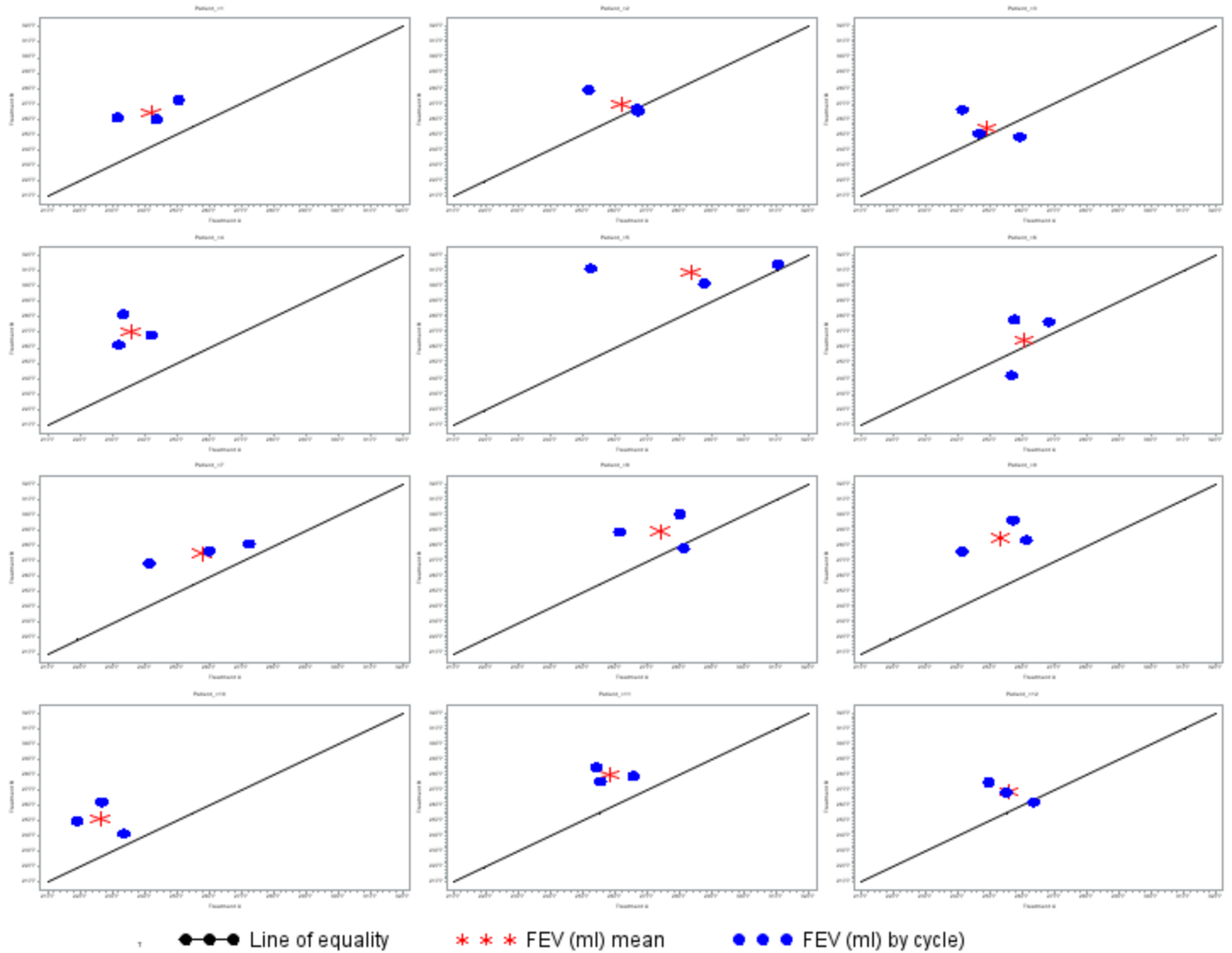
- Studies in which patients are repeatedly randomised to treatment and control
- Increased efficiency because
  - Each patient acts as own control
  - More than one judgement of effect per patient
- However, only possible for chronic diseases
- Possible randomisation in *k* cycles of treatment
- Implies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

* This is analogous to a fixed effects meta-analysis

Two simple analyses

Based on 36 pairs (3 per patient)

The TTEST Procedure

Variable: DIFFERENCE

N	Mean	Std Dev	Std Err	Minimum	Maximum
36	188.7	159.2	26.5394	-147.0	592.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
188.7	134.8	242.6	159.2	129.2	207.7

DF	t Value	Pr > |t|
35	7.11	<.0001

$7.11^2 = 50.57$

Based on 12 averages (one per patient)

The TTEST Procedure

Variable: DIFFERENCE_Mean (Mean)

N	Mean	Std Dev	Std Err	Minimum	Maximum
12	188.7	98.3242	28.3838	50.0000	348.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
188.7	126.2	251.2	98.3242	69.6524	166.9

DF	t Value	Pr > |t|
11	6.65	<.0001

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 may 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

Any damn fool can analyse a clinical trial and frequently does

Covariance Parameter Estimates

Cov Parm	Estimate
Patient_	1772.67
Residual	23685

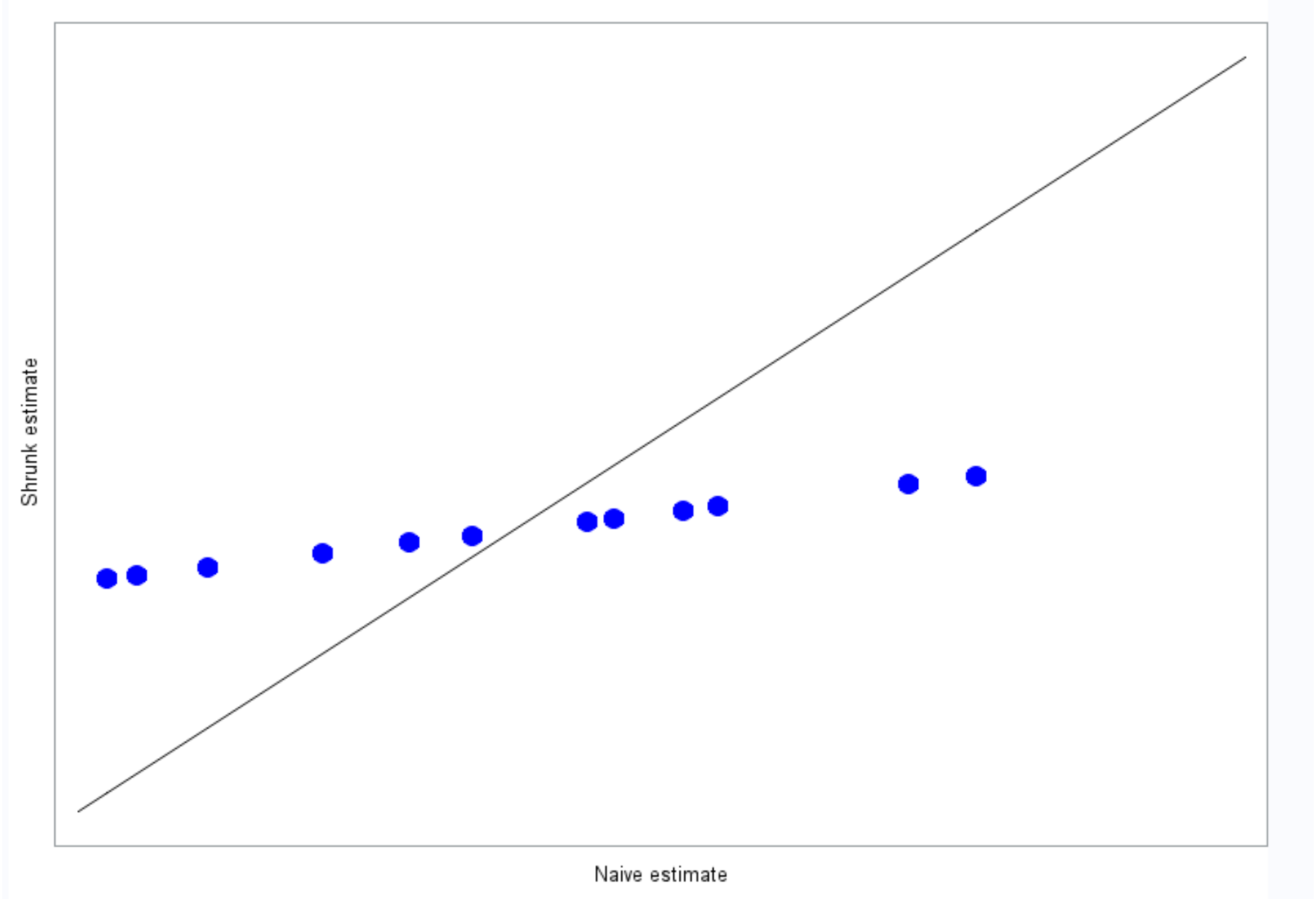
Not how the analysis using proc mixed is identical as regards inference about the mean effect to the summary measures approach using one contrast per patient

Fit Statistics

-2 Res Log Likelihood	457.7
AIC (Smaller is Better)	461.7
AICC (Smaller is Better)	462.1
BIC (Smaller is Better)	462.6

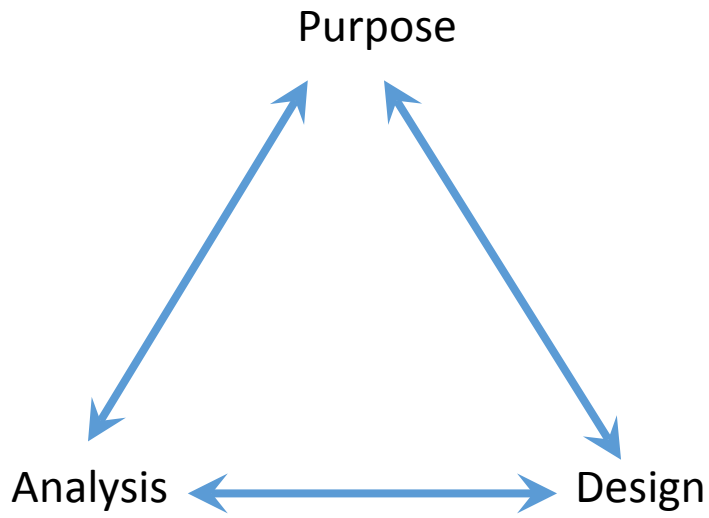
Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	188.72	28.3838	11	6.65	<.0001


Shrunk estimate
Naive estimate

Morals

- 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