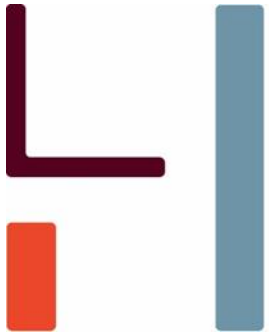
A little bit me, a little bit you

N of 1 trials, random effects and shrinkage estimators

Stephen Senn



LUXEMBOURG INSTITUTE OF **HEALTH**

RESEARCH DEDICATED TO LIFE

Acknowledgements

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 in my group and Sonia Leite formerly of my group and Steven Julious at Sheffield University

Rare Diseases

- As far as the Food and Drug Administration is concerned anything that affects fewer than 200,000 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

Rare Diseases Program

Mission Statement:

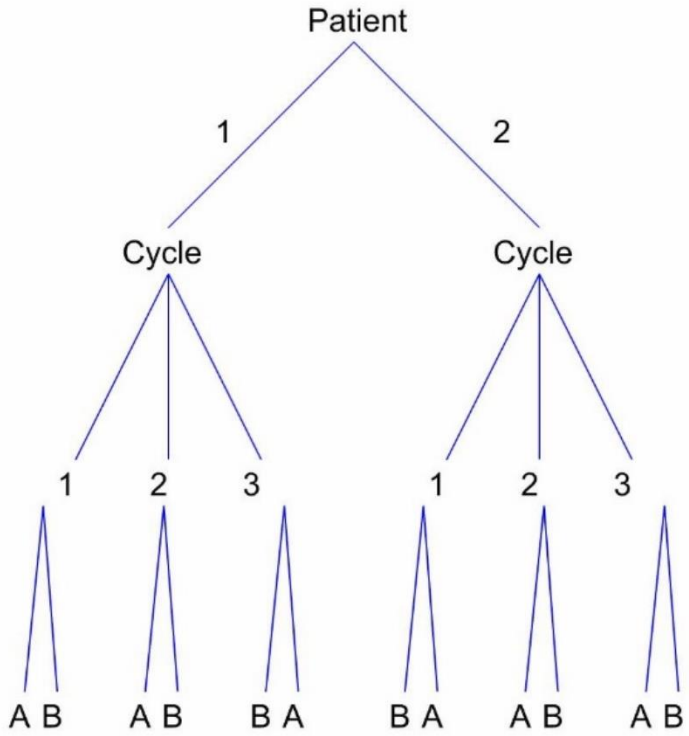
To facilitate, support and accelerate the development of drug and biologic products for the benefit of patients with rare disorders.

Overview:

- Coordinate the development of CDER policy, procedures and training for the review of treatments for rare diseases.
- Assist in outside development and maintenance of good science as the basis for the development of treatments for rare diseases.
- Work collaboratively with external and internal rare disease stakeholders to promote the development of treatments for rare disorders.
- Maintain collaborative relationships with CDER's review divisions to promote consistency and innovation in the review of treatments for rare disorders.

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



Reasons for conducting n-of-1 trials
(It is assumed that the disease is stable)

Rare disease

- Patients are few or otherwise difficult to recruit
- Within-patient studies are more efficient
- Increasing the number of periods is a way to increase the number of measurements and reduce the variance

Personalised response

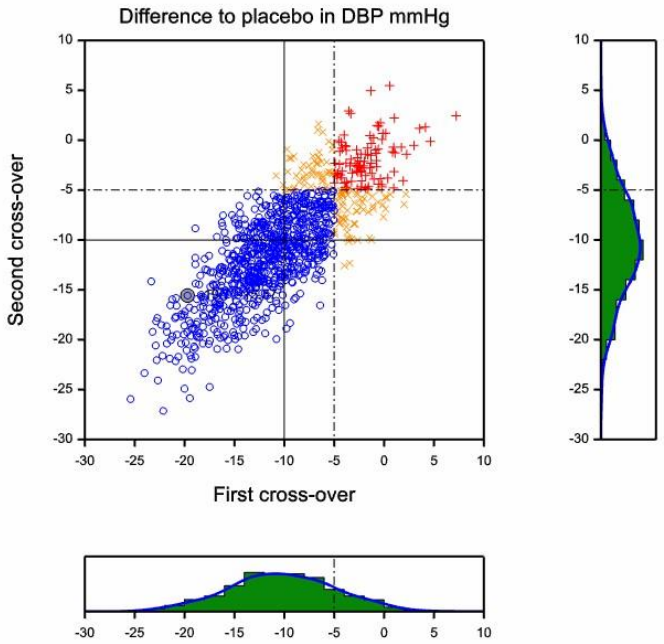
- It is desired to study personalised response to treatment
- It is necessary to separate out the components of variation
  - Within-patient
  - Treatment by patient interaction
- Designs when each patient is treated at least twice are particularly good at this

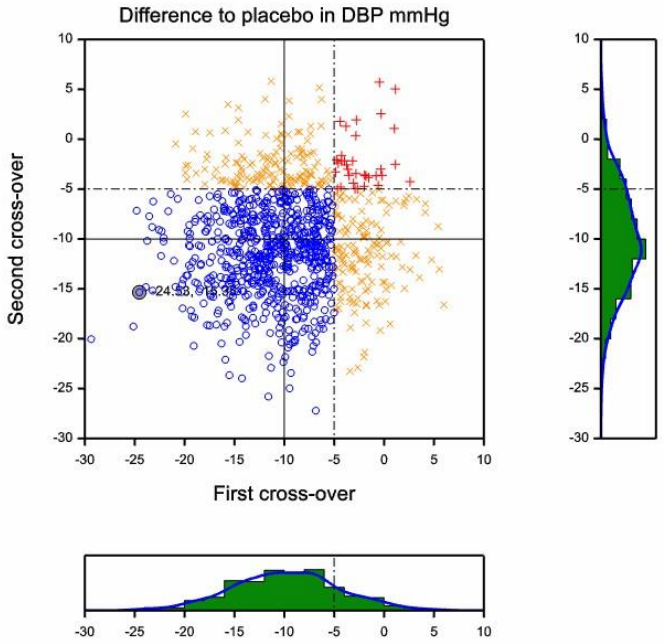
A Thought Experiment

- Imagine a cross-over trial in hypertension
- Patients randomised to receive ACE II inhibitor or placebo in random order
- Then we do it again
- Each patient does the cross-over twice
- We can compare each patient's response under ACE II to placebo twice

Design

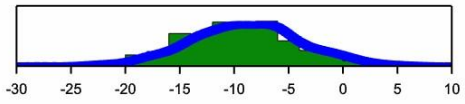
	First Cross-over		Second Cross-over	
	Period			
Sequence	1	2	3	4
I	A	B	A	B
II	B	A	B	A
III	A	B	B	A
IV	B	A	A	B





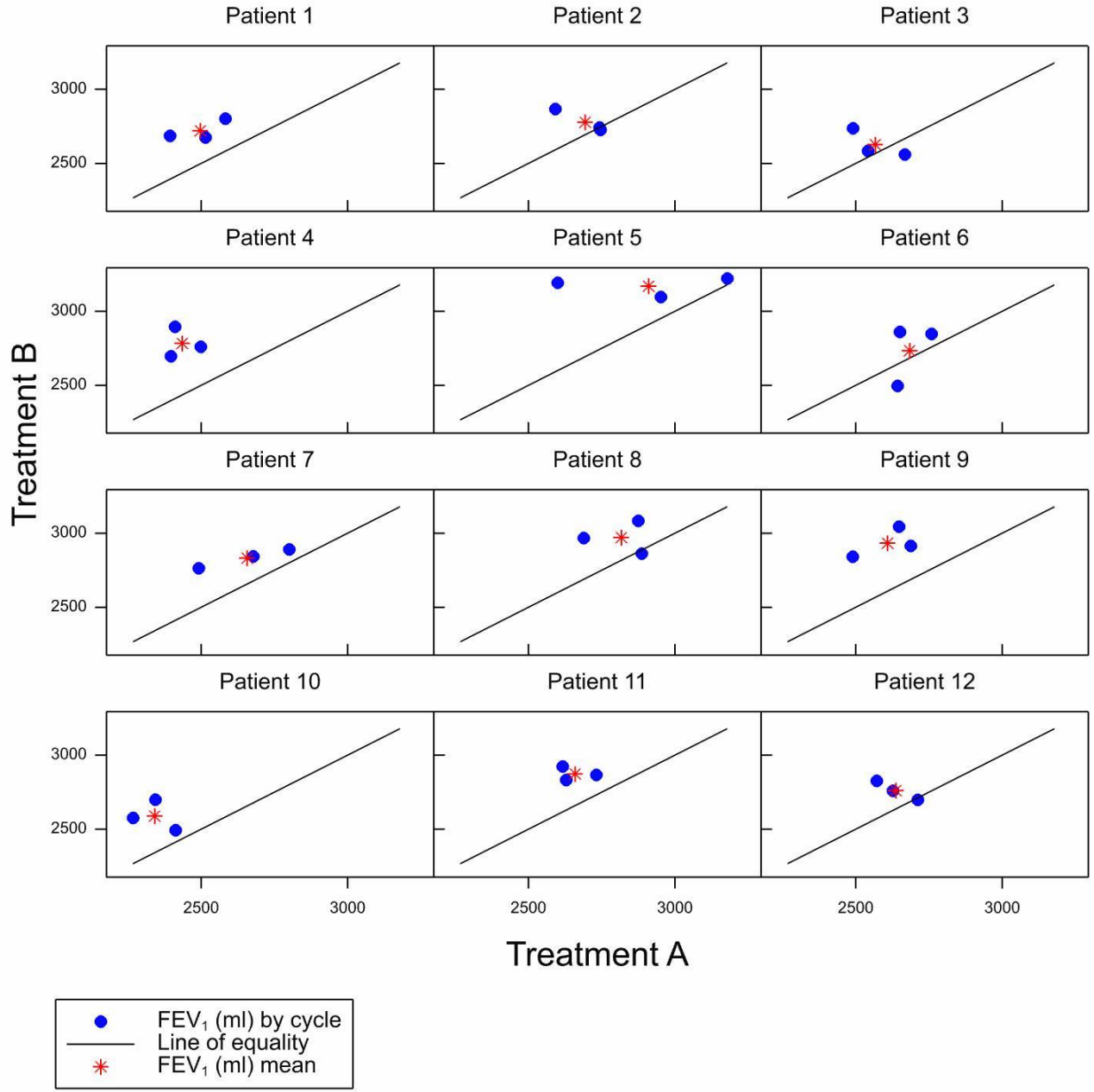
Patients are treated in two cross-over trials , thus permitting two estimates of the difference between active treatment and placebo. The difference on the second occasion is plotted against the first. Blue = response on both occasions, red = non-response on both occasions, orange = response on one occasion but not the other.

The marginal distributions are given as green histograms. LHS response on first occasion predicts response on second. RHS response on first occasion does **not** predict response on second. If you had only carried out one cross-over you would have the picture below. Which case does it apply to?



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?


Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
Patient 9
Patient 10
Patient 11
Patient 12
3000
2500
Treatment B
Treatment A
FEV1 (ml) by cycle
Line of equality
FEV1 (ml) mean

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
- Randomisation is what injects an element of stochasticity
- This requires fewer assumptions
- This approach is rare

Sampling philosophy

- The patients are regarded as a sample
- The population of relevance is some possible population of patients
- Error terms in a model capture stochasticity
- This is more ambitious
- This approach is common

Chen and Chen (2014)

- Paper on n of 1 trials in *PLOS One*
- Considered appropriate analyses of n of 1 trials randomised in cycles
- Compared performance of various analyses and compared them via simulation
- Amongst the various approaches they investigated was the matched pairs design where ‘pair’ was defined by a cycle
  - Found this performed well
- My initial reaction on seeing this is that this is wrong but on closer examination there is a sense in which it can be justified
- We shall now look at this

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
.
```

NB This is equivalent to the matched pairs approach using the 36 cycles to provide the pairs

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.			

Matched pairs t using cycles to define pair

One-sample t-test

Variate: Diff.

Summary

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
Paired difference	36	188.7	25356	159.2	26.54

95% confidence interval for mean: (134.8, 242.6)

Test of null hypothesis that mean of Paired difference is equal to 0

Test statistic t = 7.11 on 35 d.f.

Probability < 0.001

$7.11^2 = 50.57$

Consequences

- The matched pairs t-test examined by Chen and Chen (2014) *is* a valid analysis
- It is justified by the randomisation theory of the Rothamsted School and by John Nelder's approach
- However, one must be careful
  - It is a valid analysis for testing a specific null
    - That the two treatments are identical
  - In this case under the null hypothesis, the interaction is zero
- This raises the question can we do better?

The answer is 'Yes'

- We can go one step further and remove the treatment by patient interaction from the residual sum of squares
- Under the null hypothesis the expected value of the interaction is no different from the residual
- However if the alternative is true we can use a smaller residual sum of squares
- This is analogous to the following idea when carrying out a two-sample t-test
  - Despite the fact that under the *null hypothesis* a variance estimate using $n_1 + n_2$-1 is more accurate than one using $(n_1 - 1) + (n_2 - 1)$, we use the latter because it is better under the *alternative hypothesis*

```
BLOCKSTRUCTURE Patient/Cycle
TREATMENTSTRUCTURE Treatment+Treatment.Patient
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.	1.11	
Patient.Cycle.*Units* stratum					
Treatment	1	641089.	641089.	54.13	<.001
Patient.Treatment	11	159516.	14501.	1.22	0.324
Residual	24	284219.	11842.		
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 second approach is identical to fixed effect meta-analysis

The analogy to fixed effects meta-analysis

- The total degrees of freedom that we have for error are as given in the table on the right
- However if we treat each patient as a trial with *k* pairs we have *k*-1 DF for each patient
- This gives us exactly the same total
- Hence the analysis is equivalent to a fixed effects meta-analysis *provided we pool the variance estimate*

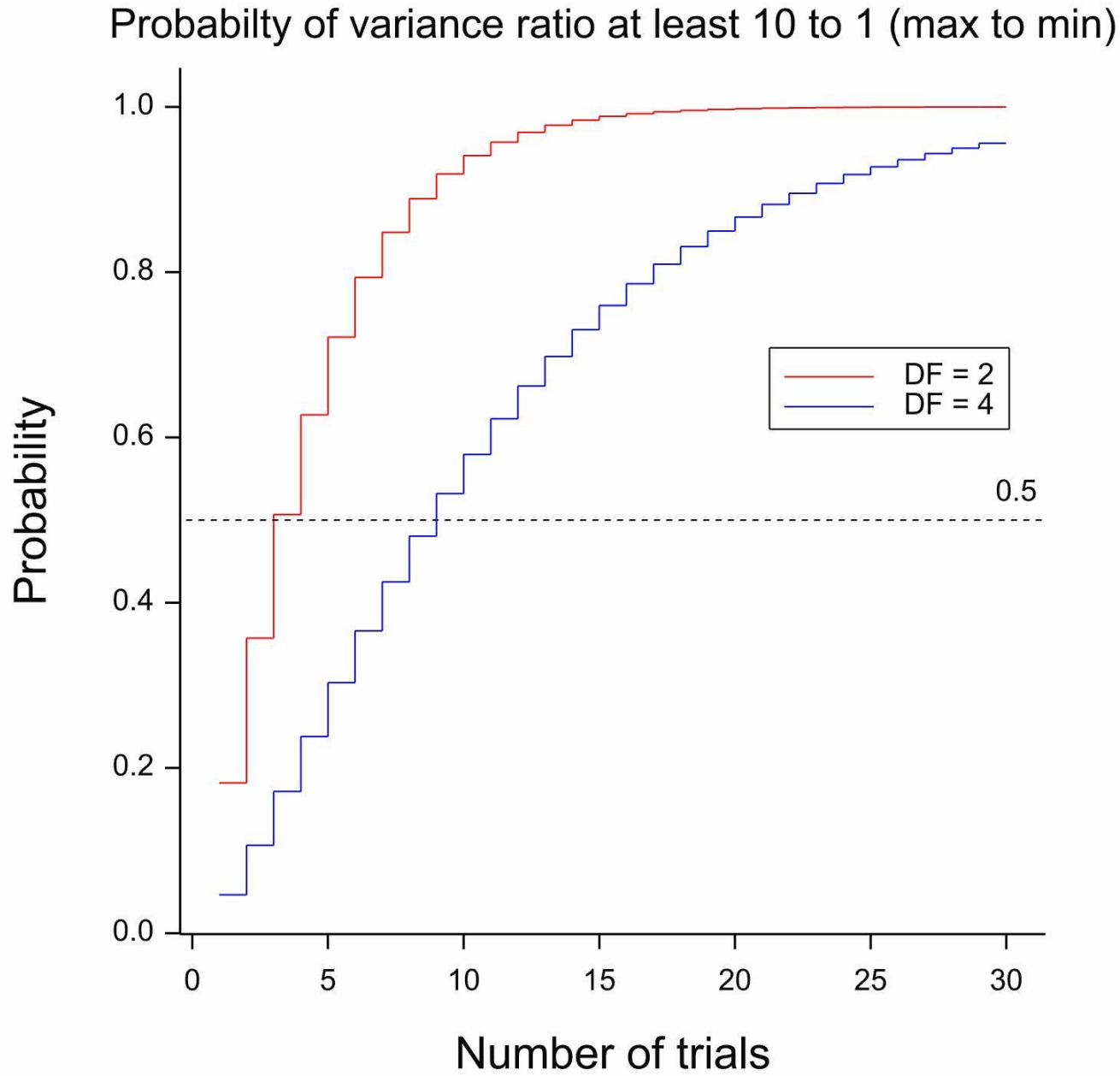
Source		
Patient	n-1	11
Cycle by Patient	n(k-1)	24
Treatment	1	1
Treatment by Patient	n-1	11
Residual	n(k-1)	24
Total	2nk-1	71

Fixed effects meta-analysis recipe

1. Calculate the differences B-A for each patient
2. Calculate the mean difference for each patient, *i*, as an estimate of the treatment effect $\hat{\delta}_i$
3. Calculate the DF for each patient. (In the balanced case these equal $k-1$. More generally we have $k_i - 1$.)
4. Calculate the corrected sum of squares for the differences for each patient
5. Sum the corrected sum of squares over all patients
6. Divide this sum by the total degrees of freedom to obtain an estimate of the variance, $\hat{\sigma}^2$
7. For each patient produce an estimate of the variance of the treatment effect as $V(\hat{d}_i) = \hat{\sigma}^2/k_i$
8. Use the estimates of the treatment effects and their variances as input to a fixed meta-analysis routine

One important difference to conventional meta-analysis

- In a conventional meta-analysis the variance would be estimated independently within each trial
- Here a pooled variance has been used
- Because the degrees of freedom are so few, independent variance estimation would be a bad idea
- Even when true variances are identical they can easily vary randomly very greatly as the next slide shows
- This shows the probability that the highest to lowest will vary by a ratio of at least 10 to 1 as a function of the number of patients and for two cases
  - Degrees of freedom = 2 and Degrees of freedom =4


Probabilty of variance ratio at least 10 to 1 (max to min)
Probability
1.0
0.8
0.6
0.4
0.2
0.0
0.5
DF = 2
DF = 4
0
5
10
15
20
25
30
Number of trials

Two more difficult questions

The average effects in future patients?

- This may require a mixed effects model
- Allows 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 effects model can be used to predict long-term average effects for patients in the trial
- A weighted estimate is used whereby the patient's own results are averaged with the general result

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

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

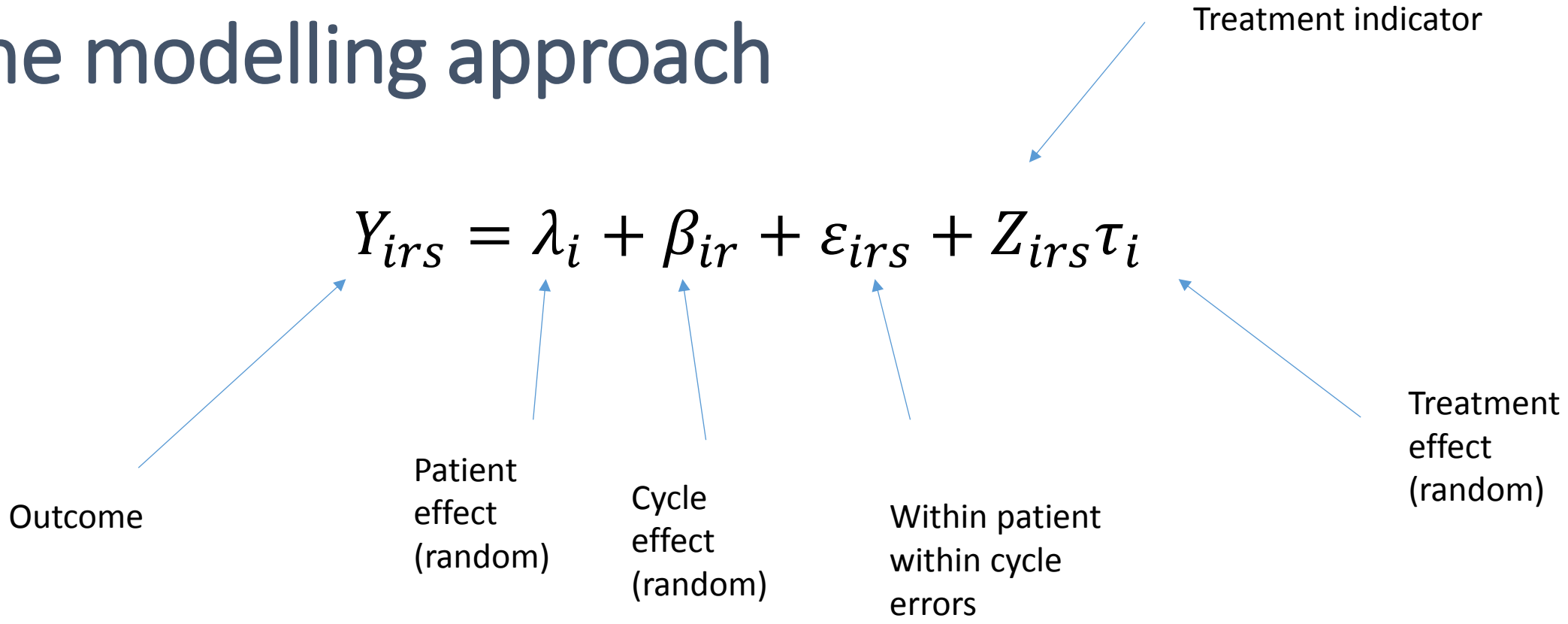
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
188.7	126.2	251.2	98.3242	69.6524	166.9

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

Moving to mixed models

- The previous approaches were two-stage approaches
- First stage reduce the data to meaningful contrasts
  - Mean difference per cycle *or*
  - Mean difference per patient
- Second stage is to analyse the contrasts using a matched-pairs t
  - Based on 36 within cycle differences or
  - Based on 12 within-patient mean differences
- An alternative is to use a mixed model
- Provided there are no incomplete cycles the following two approaches are equivalent
  - A model for the whole data or
  - A model for the differences

The modelling approach



$$\varepsilon_{irs} \sim N(0, \sigma^2), \beta_{ir} \sim N(0, \gamma^2), \lambda_i \sim N(\Lambda, \phi^2) and\ \tau_i \sim N(T, \psi^2)$$

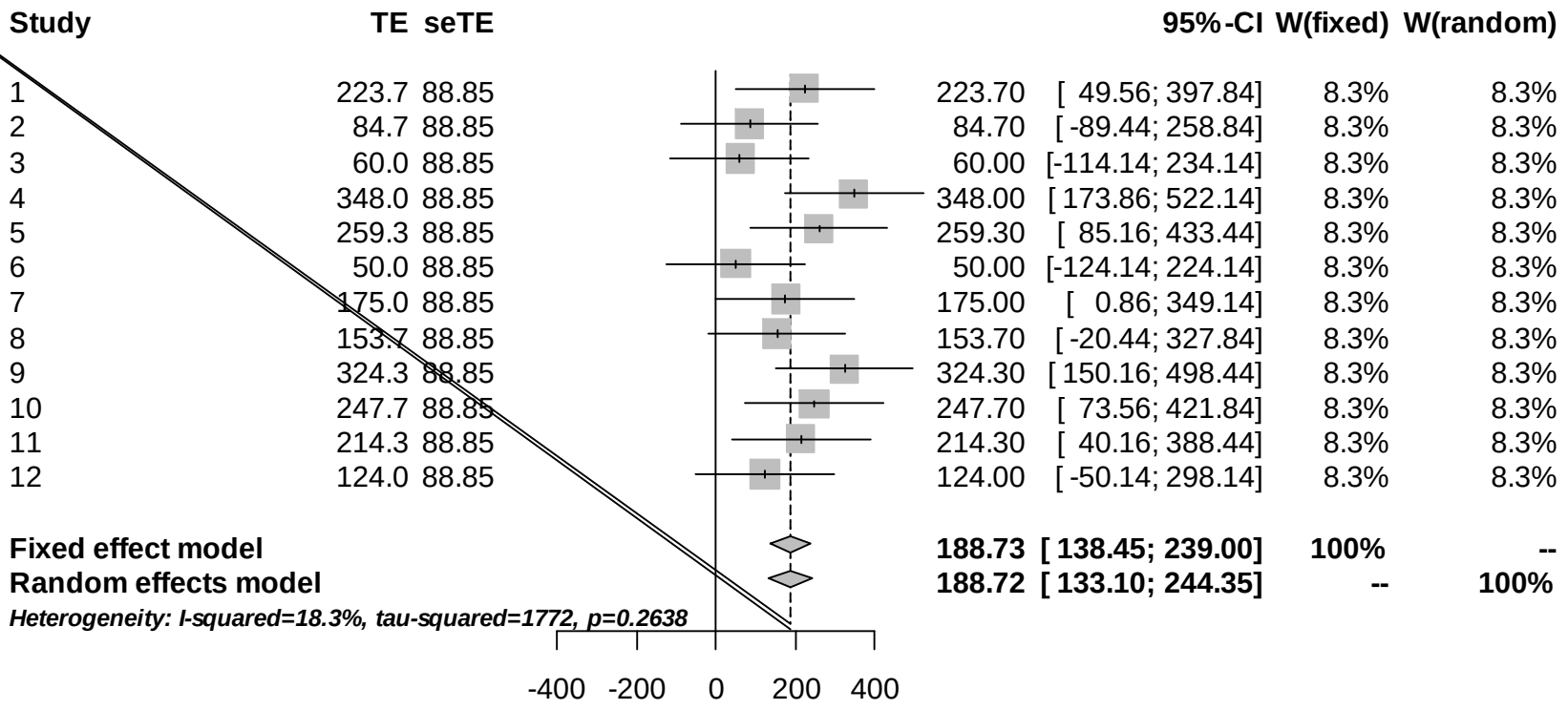
Solution for Fixed Effects

Effect	Treatment_	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		2625.75	45.2031	11	58.09	<.0001
Treatment_	B	188.72	28.3838	11	6.65	<.0001
Treatment_	A	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Treatment_	1	11	44.21	<.0001

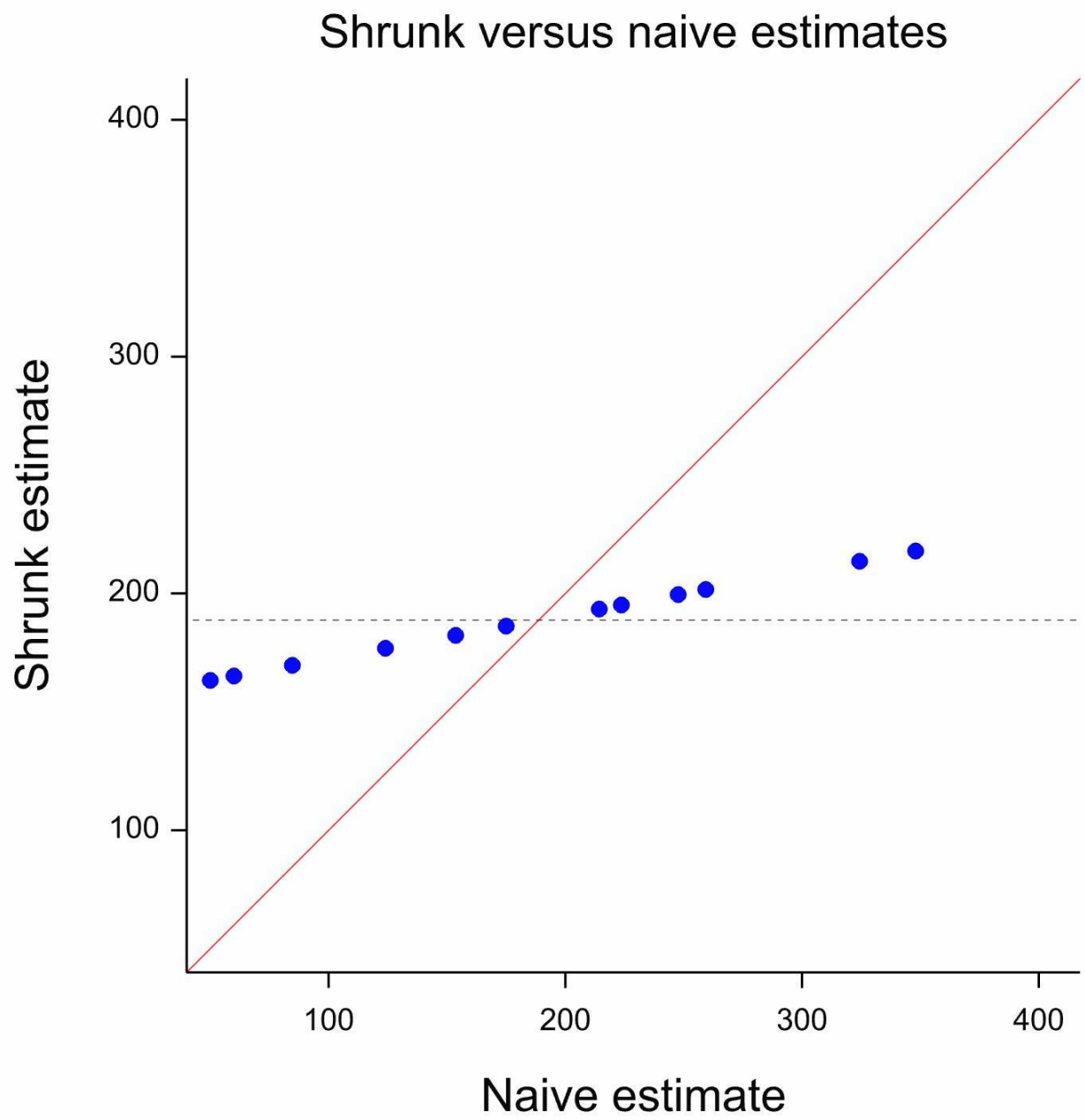
Analysis using meta package in R

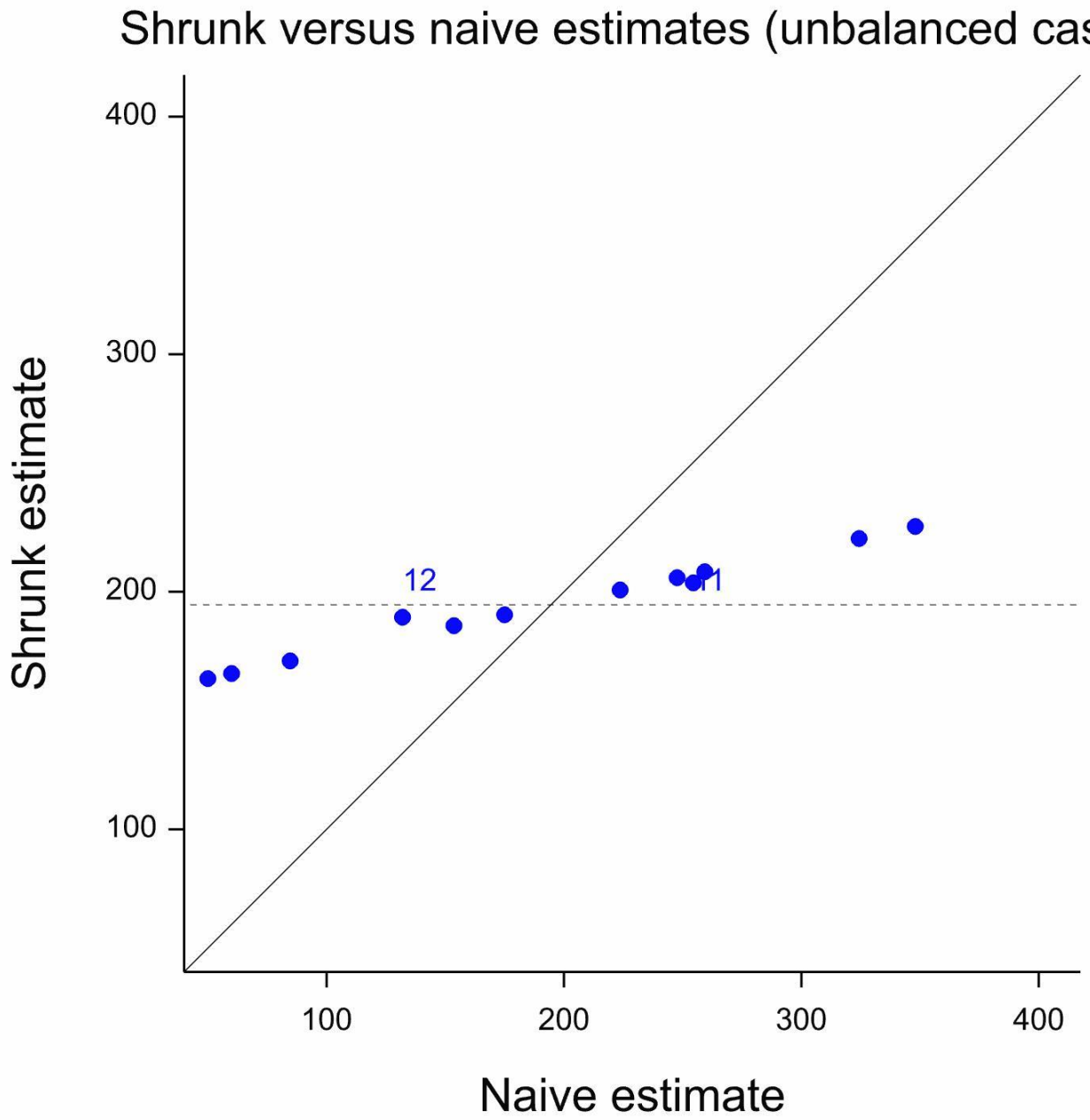


Conclusion

- Several different approaches to analysis give the exactly the same estimate and standard error of the estimate
- The question is why?

Method	Explanation
Mixed model approach	Estimates the two components of variation and then adds them together to calculate variance of the treatment estimate
Summary measures approach	Uses the combined effect at the patient level of the two variances to estimate their total contribution correctly without having to partition it
Random effects meta-analysis	Starts by using the wrong variance but then inflates it empirically so that the total matches what the other two methods would show


Shrunk versus naive estimates
Shrunk estimate
400
300
200
100
100
200
300
400
Naive estimate



Patients 1-10 with 3 cycles
Patient 11 with 2 cycles
Patient 12 with 1 cycle

Conclusions

- Very different purposes justify very different analyses
- Proving that there is a difference between treatments (causal)
  - Randomisation based
  - Fixed effects meta-analysis
- Attempting (with difficulty) to estimate effects in patients and predict them for future patients
  - Mixed models
  - Shrinkage estimators
  - Random effects meta-analysis