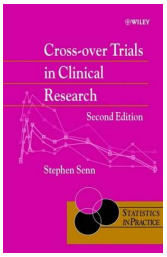
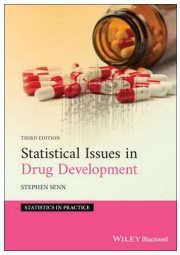
Design and Analysis of N of 1 Trials

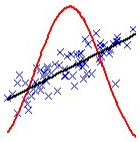
Lecture 1





Stephen Senn

Consultant Statistician, Edinburgh



DIAMOND
Development of generalisable methodology for n-of-1 trials delivery for very low volume treatments

1

Acknowledgements

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

SEVENTH FRAMEWORK PROGRAMME

DIAMOND
Development of generalisable methodology for n-of-1 trials delivery for very low volume treatments

No funders are responsible for any of the content.

Some of this work was joint with Artur Araujo and Sonia Leite, formerly of my group in Luxembourg and Steven Julious at Sheffield University

2

What I hope to do in this *course*

- Lecture 1 Introduce the topic and convince you it is interesting and perhaps even useful
- Lecture 2 Develop randomisation based approaches
  - Main purpose to establish treatment under investigation *can* work
- Lecture 3 Develop mixed-model based approaches
  - *Main purpose investigating components of variation to aid prediction and for personalising treatment*
- Lecture 4 Cover analysis of binary data and also some planning issues
- Overall: convince you that carefully thinking about purpose is important when trying to understand n-of-1 trials

(C)Stephen Senn 2023

3

What I am not going to cover

- I am not going to cover practical aspects of running and reporting n-of-1 trials
- This does not, imply, however, that these are not important
- They are covered in great detail in the Diamond report
- See https://www.sheffield.ac.uk/scharr/research/centres/ctru/diamond

Section 1: When is it appropriate to undertake n-of-1 trials:

- Scope
- Prevalence of the health condition
- Type and attributes of the health technology
- Questions that can be addressed

Section 2: Design and analysis conditions:

- Choice of outcome
- Choice of comparator
- Target of treatment
- Number of health technologies and periods
- Blinding
- Randomisation
- Analysis

(C)Stephen Senn 2023

4

What I hope to do in this *lecture*

- Define n of 1 trials
- Give some history of their use
- Explain what they can be used for
  - Increasing efficiency
  - Studying individual response
- Give some background regarding difference purposes of clinical trial generally to motivate further development of statistical issues in the course

5

Three quotes to start with

No man is an island entire in itself
John Donne, Devotions Upon Emergent Occasions, and severall steps in my Sicknes

There are probably no two men in existence on whom the drug acts in exactly the same manner.
Wilkie Collins, *The Moonstone*

If we could, this year, exactly reproduce, in your case, the conditions as they existed last year, it is physiologically certain that we should arrive at exactly the same result. But this – there is no denying it – is simply impossible.
Wilkie Collins, *The Moonstone*

6

What are n of 1 trials?

An 'n of 1' trial is a trial in which a number of episodes of treatment are studied in a single patient, usually with a view to making inferences about the effect of treatment in that patient.

Where a number of patients are studied, then, n of 1 trials taken as a set have the structure of 'extra period' cross-over trials.

However, historically, the emphasis in n of 1 trials has been upon independent study of effects, patient by patient.

As will be shown in the course, this is not a good idea

7

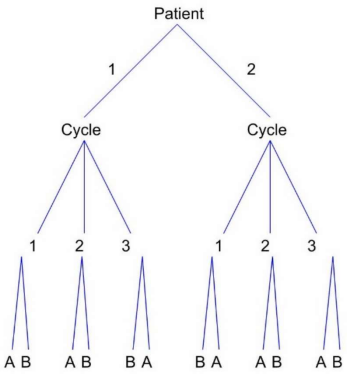
Warning

- Although I shall frequently illustrate the use of SAS® and R I am not just going to use SAS® and R
- I shall also use Genstat®
  - It is good not to be reliant on one package
  - Different packages have different strengths and Genstat® is particularly good on designed experiments
  - Looking at different ways of doing things pays dividends in terms of understanding and robustness
- I am not an expert on SAS®
- I am not an expert on R
- I am not even an expert on Genstat®!

8

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



9

Warning

There are two very different traditions

Medical

- Has tended to emphasise independent analysis of single patients
  - There are some early exceptions but this has been the modern trend
- Each time a new patient is studied one starts from scratch
- This is not the subject of this course

Statistical

- Has tended to emphasise simultaneous analysis of results from all patients treated under a common n-of-1 protocol
- This is the approach that will be taken in this course

10

Student in 1908

Illustration I. As an instance of the kind of use which may be made of the tables, I take the following figures from a table by A. R. Cushny and A. R. Peebles in the *Journal of Physiology* for 1904, showing the different effects of the optical isomers of hyoscyamine hydrobromide in producing sleep. The sleep of ten patients was measured without hypnotic and after treatment (1) with D. hyoscyamine hydrobromide, (2) with L. hyoscyamine hydrobromide. The average number of hours' sleep gained by the use of the drug is tabulated below

The conclusion arrived at was that in the usual dose 2 was, but 1 was not, of value as a soporific.

11

Additional hours' sleep gained by the use of hyoscyamine hydrobromide.

Patient	1 (Dextro-)	2 (Laevo-)	Difference (2–1)
1.	+ ·7	+ 1·9	+ 1·2
2.	− 1·6	+ ·8	+ 2·4
3.	− ·2	+ 1·1	+ 1·3
4.	− 1·2	+ ·1	+ 1·3
5.	− 1	− ·1	0
6.	+ 3·4	+ 4·4	+ 1·0
7.	+ 3·7	+ 5·5	+ 1·8
8.	+ ·8	+ 1·6	+ ·8
9.	0	+ 4·6	+ 4·6
10.	+ 2·0	+ 3·4	+ 1·4
	Mean + ·75	Mean + 2·33	Mean + 1·58
	S. D. 1·70	S. D. 1·90	S. D. 1·17

1) There was no Dextro form but a racemate (A Dextro-Laevo mixture)
2) The column headed Dextro is in any case the other molecule L-Hyosciamine

12

As a general rule a tablet was given on each alternate evening, and the duration of sleep and other features noted and compared with those of the intervening control night on which no hypnotic was given. Hyoscyamine was thus used on three occasions, and then racemic hyoscine, and then lævo-hyoscine. Then a tablet was given each evening for a week or more, the different alkaloids following each other in succession. The results may be given shortly in tabular form, details being reserved for publication elsewhere.

TABLE I.

	Controls (no hypnotic)		0·6 mg. L-Hyoscyamine HBr.			0·6 mg. L-Hyoscine HBr.			0·6 mg. R-Hyoscine HBr.		
Patient	No. of observations	Average hours of sleep	No. of observations	Average hours of sleep	Increase over controls	No. of observations	Average hours of sleep	Increase over controls	No. of observations	Average hours of sleep	Increase over controls
1	9	0·6	6	1·3	0·7	6	2·5	1·9	6	2·1	1·5
2	9	3·0	6	1·4	−1·6	6	3·8	0·8	6	4·4	1·4
3	8	4·7	6	4·5	−0·2	6	5·8	1·1	6	4·7	0·0
4	9	5·5	3	4·3	−1·2	3	5·6	0·1	3	4·8	−0·7
5	9	6·2	3	6·1	−0·1	3	6·1	−0·1	3	6·7	0·5
6	8	3·2	4	6·6	3·4	3	7·6	4·4	3	8·3	5·1
7	8	2·5	3	6·2	3·7	3	8·0	5·5	3	8·2	5·7
8	7	2·8	6	3·6	0·8	6	4·4	1·6	5	4·3	1·5
9	8	1·1	5	1·1	0·0	6	5·7	4·6	5	5·8	4·7
10	9	2·9	5	4·9	2·0	5	6·3	3·4	6	6·4	3·5
11	—	—	2	6·3	—	2	6·8	—	2	7·3	—

From these results it is evident that hyoscyamine is of no value in the dose given as a hypnotic, while the lævorotary and racemic forms of hyoscine have about the same influence in inducing sleep. In one or two cases the patient complained of thirst or dryness of the mouth. The pulse generally became slower, and slight dilation of the pupil appeared in some instances from each of the drugs. In other cases acceleration of the pulse was noted. It may be questioned whether these changes in the pulse and pupil were direct effects of the drug or merely accompanied drowsiness.

See The James Lind Library

and

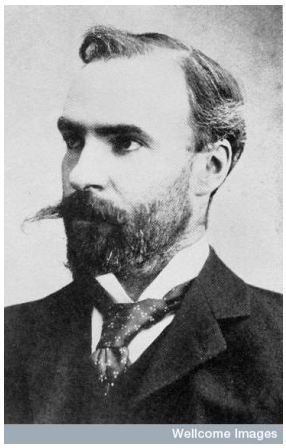
Cushny AR, Peebles AR (1905). The action of optical isomers. II. Hyoscines. J Physiology 32:501-510.

13

13

Arthur Cushny (1866-1926)

- Born Fochabers 1866
- Studied medicine at Aberdeen
- Worked in Bern (Kronecker) & Strasburg (Schmiedeberg)
- Chairs
  - Michigan 1893-1905
  - UCL 1905-1918
  - Edinburgh 1918-1926
- Died Peffermill House 1926



Wellcome library London

L0012929

(C)Stephen Senn 2023

14

14

Cushny and Peebles (1905)

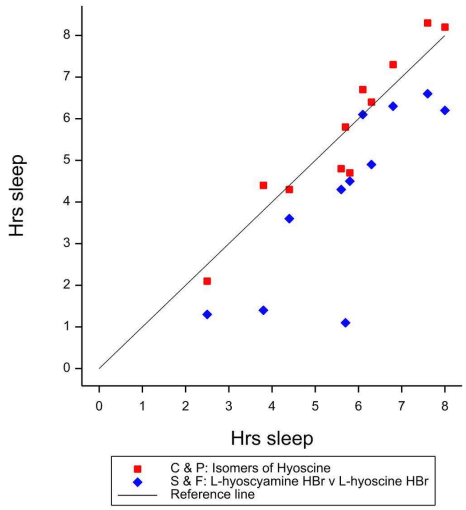
"…a number of trials of their usefulness for this purpose were made in the Michigan Asylum for Insane at Kalamazoo. The harmlessness of small doses of both alkaloids were first ascertained on ourselves, and then a number of tablets each containing 0.6mg of L-hyoscine or
R-hyoscine hydrobromate were used as hypnotics in the wards of Drs Richards and Light under the general supervision of Dr. W. M. Edwards. We are much indebted to these physicians for the results recorded by them. Instead of hyoscine, a certain number of tablets contained 0.6 mg of hyoscyamine hydrobromate, as its usefulness as a hypnotic has not yet been determined. In all, ten patients were treated with the tablets."
(pp. 508-509).

15

"As a general rule a tablet was given on each alternate evening and the duration of sleep and other features were noted and compared with those of the intervening control night on which no hypnotic was given. Hyocyamine was thus used on three occasions, and then racemic hyoscine, and then laevohyoscine. Then a tablet was given each evening for a week or more, the different alkaloids following each other in succession. The results may be given shortly in tabular form, details being reserved for publication
elsewhere. (Reference 3, p. 509)"

If we use the symbol - to stand for 'control night', H for hyoscyamine, R for R-hyoscine and L for L-hyoscine, a typical sequence seems to have been of the form
H – H – H – R – R – R – L – L – L - H R L H R L H R L

16



C & P = Cushny & Peebles (1905)
S&F = Student (1908) & Fisher (1925)

The red squares compared isomers
The blue diamonds compare the L isomer and the control

Student & Fisher mislabelled the latter as two isomers

17

Is it really Student's test?

No

- The modern form we use with the ratio of the mean to the standard error is due to Fisher
- Student was anticipated by Jakob Luroth (1844-1910) who derived (essentially) the same result in 1876

Yes!

- Student derived it himself
- Because he needed it
- And he used it
- Although Luroth anticipated him and Fisher improved (and proved) his test (in 1925) Student was the true original implementer of the t-test

18

The Trial at Kalamazoo

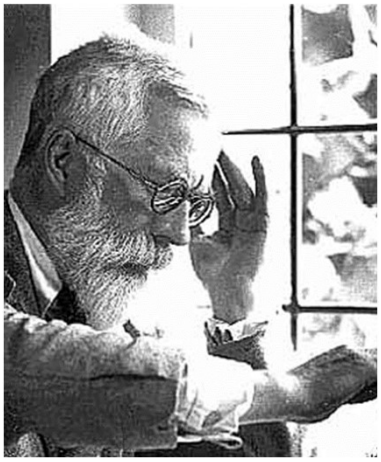
- An innovative investigation
  - Interesting scientific hypothesis
  - Clever use of control
  - Intelligent analysis
    - Without benefit of statistics!
  - Sensible conclusion
  - Ethical ground prepared
- Only randomisation is missing
- The design is of a series of n-of-1 trials
- Unfortunately we don't have the original data

Paul Martini (1889-1964)

- Born in Frankenthal
- Studied medicine in Munich and Kiel
- 1927 became medical director of St Hedwig clinic in Berlin
- Bemoaned the fact that there was far more research in diagnosis than treatment
- 1932 monograph 'Methods of Therapeutic Examination' advocated careful assessment of treatment including n-of-1 trials

RA Fisher (1890-1962)

- In 1935 published the first book devoted to experimental design *per se*
- The second chapter of *The Design of Experiments* introduces the famous tea-tasting example
- Fisher uses a randomisation argument to analyse this *as designed*



The Tea-Tasting Experiment 1

- The subject was Muriel Bristol (1888-1950) an algologist at Rothamsted
- She had objected to Fisher having poured milk in last at tea-time
- He claimed it made no difference and set her to the test

The tea-tasting experiment 2

- There are 8 cups in total, 4 with milk in first and 4 with tea in first
- The order is randomised and she is told it is randomised
- Her task is to guess which cups are which

Fisher calculates as follows

The total number of sequences is $\frac{8!}{4!4!} = 70$ so her probability of guessing all cups correctly if she cannot tell the difference is 1/70

Modern Development

- Early use in psychology, for example by Skinner
- Covered in Sidman's book of 1960
- 1980s McMaster group starts using them
- Guyatt et al report their use of them in a 1986 article in The *New England Journal of Medicine*
- By the end of the 1980s McMaster group had run 57 such trials

"The department was multidisciplinary and very tightly integrated. So there were … statisticians and psychologists and people with behavioral backgrounds, physicians and epidemiologists getting together on a regular basis. And for a while, one of the psychologists would say, "Oh, that would be very interesting for an n-of-1 trial." And we said, "Thank you very much" and would go on. Then at one point it clicked, and we started to get out the psychology literature and found three textbooks full of n-of-1 designs from a psychology perspective. … It was totally old news."

See What Ever Happened to N-of-1 Trials? Insiders' Perspectives and a Look to the Future, Kravitz et al, *The Milbank Quarterly*, Vol. **86**, No. 4, 2008 (pp. 533–555)

Since 1990

- Papers on n of 1 trials have begun to trickle into the literature
- However, generally, there is little serious work in these papers in terms of methodology
- Exceptions
  - Single trial analyses
    - Rochon (1990) using time-series measures for individual patients
  - Sets of n-of.1 trials
    - Zucker at al (1997, 2010) using "Bayesian" methods for analysing sets of n-of-1 trials
      - Effectively this is equivalent to using mixed models
    - Chen & Chen (2014) review paper in *PLOS One*

25

Irwig et al, *BMJ*, 1994

- 25 patients with osteoarthritis
- Each randomised in three four-week cycles
- In each four-week cycle, paracetamol given for two weeks, diclofenac for two-weeks, with order being random
- Most severely affected joint used for assessment
- First week of every two-week treatment discarded (to avoid carry-over)
- Results compared using a t-test with 2 df
  - 3 pairs per patient yield 3 differences & hence 2 df

26

Summary of results for 15 patients who completed study

Case No	Treatment before study	Pain (mm)* NSAID	Paracetamol	P (one tailed)	N
2	NSAID	7·5	9·3	0·23	
3	NSAID	50·1	54·2	0·13	
5	NSAID	0·4	14·1	0·19	
6	NSAID	62·5	62·1	0·63	
7	NSAID	38·4	34·3	0·63	
8	NSAID	4·8	7·2	0·14	
9	NSAID	5·8	23·8	0·001	
11	Paracetamol	4·3	5·8	0·36	
12	None	35·0	40·5	0·33	
13**	NSAID	12·5	9·7	0·63	
15	NSAID	36·4	65·9	0·01	
19	NSAID	0·8	2·1	0·11	
20	Paracetamol	7·3	7·2	0·55	
21	NSAID	9·5	43·3	0·003	
23	NSAID	3·1	2·7	0·67	

***Conclusions*—n Of 1 studies—that is, randomised trials in individual patients—are clinically useful in deciding treatment in heterogeneous conditions which require long term symptomatic relief. In osteoarthritis many patients currently receiving or being considered for non-steroidal anti-inflammatory drugs may achieve adequate control with paracetamol.**

27

Criticisms

- Each t-test only has 2 degrees of freedom
  - Variance estimates are extremely unstable
  - Would a pooled variance be better?
- The default is taken to be paracetamol
  - Diclofenac has to prove itself to be better
  - Why not have the reverse test also?
- No attempt is made to analyse the results as a whole
  - Matched pairs t using summaries?
  - Meta-analysis

28

One-sample t-test

Variate: Difference.

Summary

	Sample Size	Mean	Variance	Standard deviation	Standard error of mean
Difference	15	-6.920	135.1	11.62	3.001

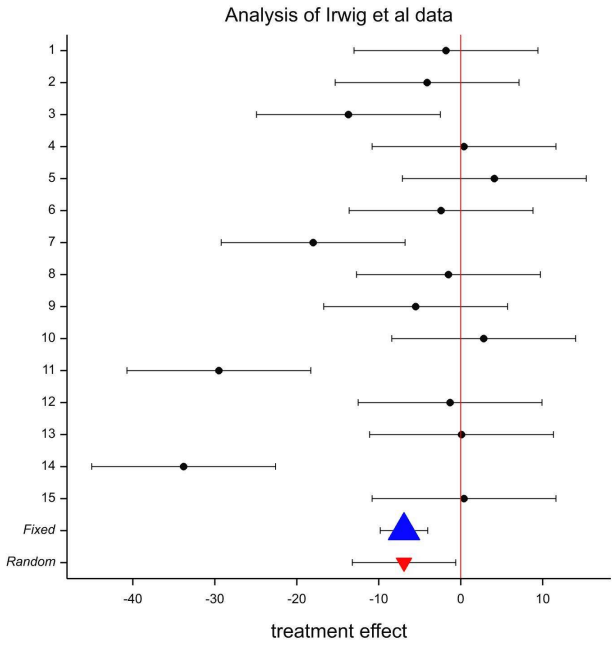
95% confidence interval for mean: (-13.36, -0.4838)

Test of null hypothesis that mean of Difference is equal to 0

Test statistic t = -2.31 on 14 d.f.

Probability = 0.037

29



Meta-analysis using pooled variance

30

Why do we do n-of-1 trials?

To increase efficiency

- Between patient (main effect) variation eliminated
- Increasing the number of times we study individual patients reduces the within-patient variance
- Studying more episodes within patients is an alternative to studying more patients

To identify components of variation

- By repeatedly studying the same patient whilst varying the treatment we can estimate various sources of variation
  - Between patient
  - Within patient
  - Treatment by patient interaction
- The idea behind this is explained in the next few slides

31

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

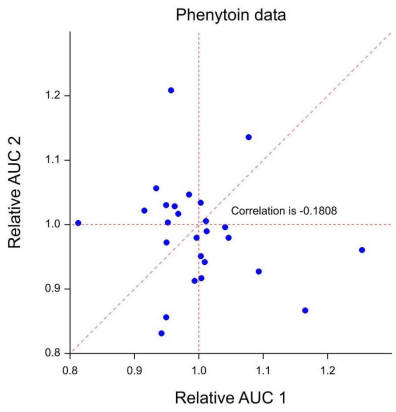
A rare disease is a life-threatening or chronically debilitating disease that affects 5 people or fewer in 10,000 and requires special, combined efforts to enable patients to be treated effectively.

UK Strategy for Rare Diseases 2013

32

Identifying components of variation

- This has been little explored in drug development
- There is one notable exception and that is a case where it is unlikely to be important – individual bioequivalence
- This example is from Shumaker & Metzler 1998



33

Sources of Variation in Clinical Trials

Label	Source	Description
A	Between treatments	The difference between treatments averaged over all patients
B	Between patients	The difference between patients given the same treatment
C	Treatment-by-Patient Interaction	The extent to which the effect of treatment varies from patient to patient
D	Within patients	The extent to which the results vary from occasion to occasion for patients given the same treatment

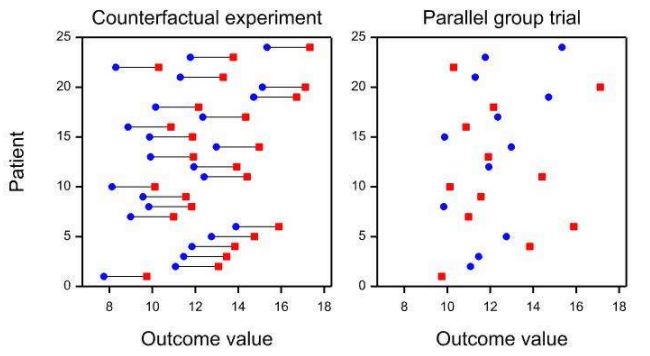
Senn SJ. Individual Therapy: New Dawn or False Dawn. *Drug Information Journal* 2001;35(4):1479-1494.

34

Identifiability and Clinical Trials

Type of Trial	Description	Identifiable Effects	Error Term
Parallel	Each patient is randomised to receive one treatment	A	B+C+D
Cross-over	Each patient receives each treatment in one period only	A and B	C+D
Series of n of 1 Trials (Repeated cross-overs)	Each patient receives each treatment in at least two periods	A and B and C	D

35



Left-hand panel: what you would see if patients could be treated both ways
Note how difference active-placebo is constant

Right-hand panel: what you will see in a parallel group trial (one outcome unobservable)
You cannot judge the effect patient by patient
A common error is to look at RH panel and judge that not everybody benefits

36

On the Normal Distribution

Tout le monde y croit cependant, me disait un jour M. Lippmann, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental. Henri Poincaré (p171)

On individual response

The trialists think genetics shows it to be inevitable and the geneticists think the trialists have demonstrated it is a fact

Stephen Senn

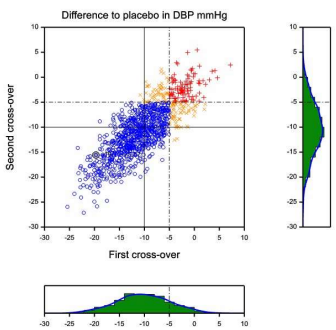
A Thought Experiment

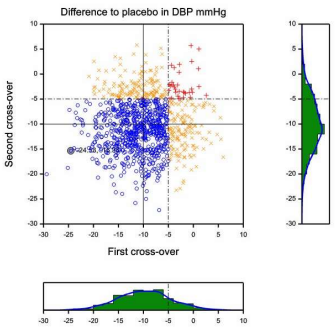
- Imagine a cross-over trial in hypertension
- Patients randomised to receive ACE II inhibitor or placebo in random order
- Then we do it again
- Each patient does the cross-over twice
- We can compare each patient's response under ACE II to placebo twice

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

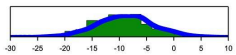
39





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?



40

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?

(C)Stephen Senn 2023

41

To sum up

- N of 1 trials have a rather chequered history
- Much of the development has been driven by medics and the statistical techniques used have often been naïve
- There are two important uses to which n-of-1 trials can be put
  - Increasing efficiency
  - Personalising treatment (or at least studying variation in personal response)
- These purposes frequently require different analyses
- Explaining *what* it appropriate *when* is one of the major objectives of this course

(C)Stephen Senn 2023

42