Why random is good:

The statistics of clinical trials

Christopher Jennison

Dept of Mathematical Sciences,

University of Bath, UK

http://people.bath.ac.uk/mascj

Mathscon

University of Bath

November 2018

Outline of talk

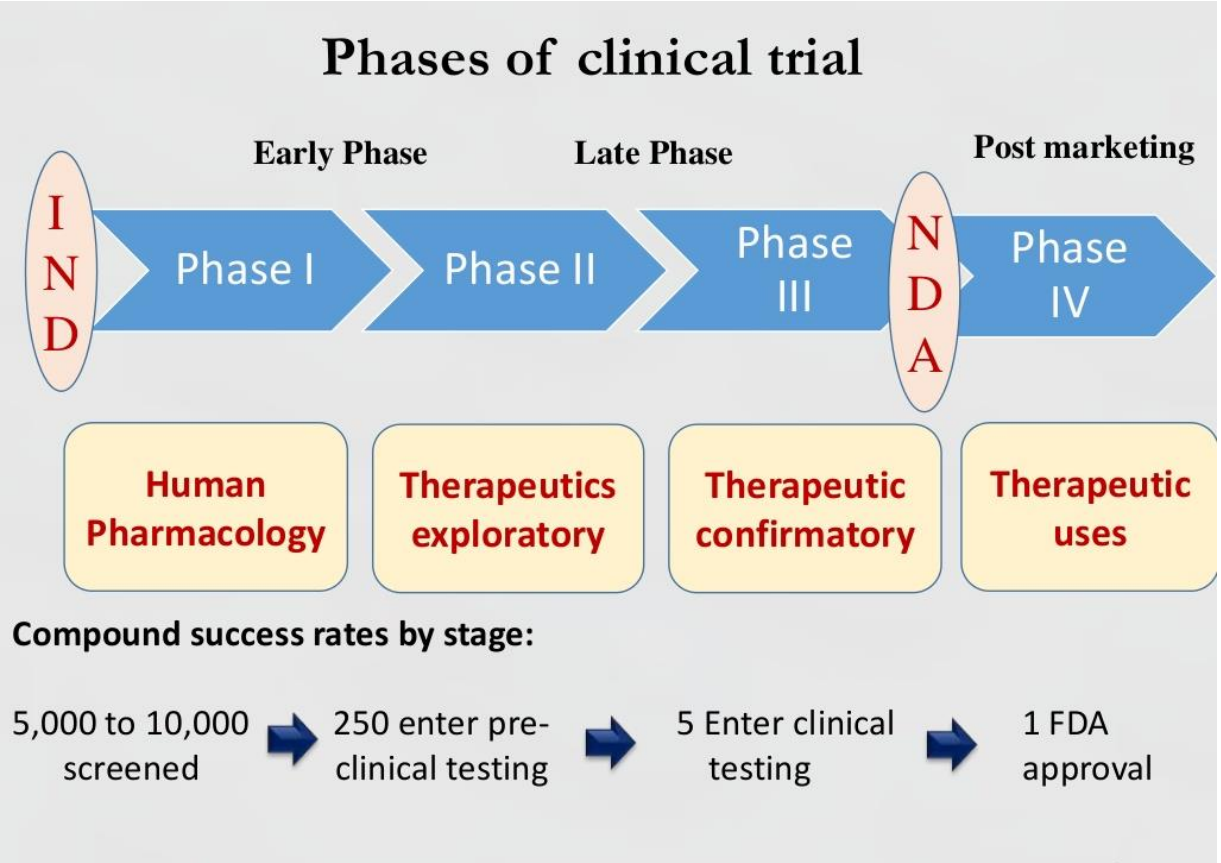
1. Clinical trials and the drug development programme
2. Variability in patient response to treatment
3. Reducing the noise

   Stratified sampling, Paired comparisons, Crossover trials
4. The placebo effect

   Examples, A clever idea: SPCD
5. And there's more ...

   When to stop a trial, Adaptive designs
6. Conclusions

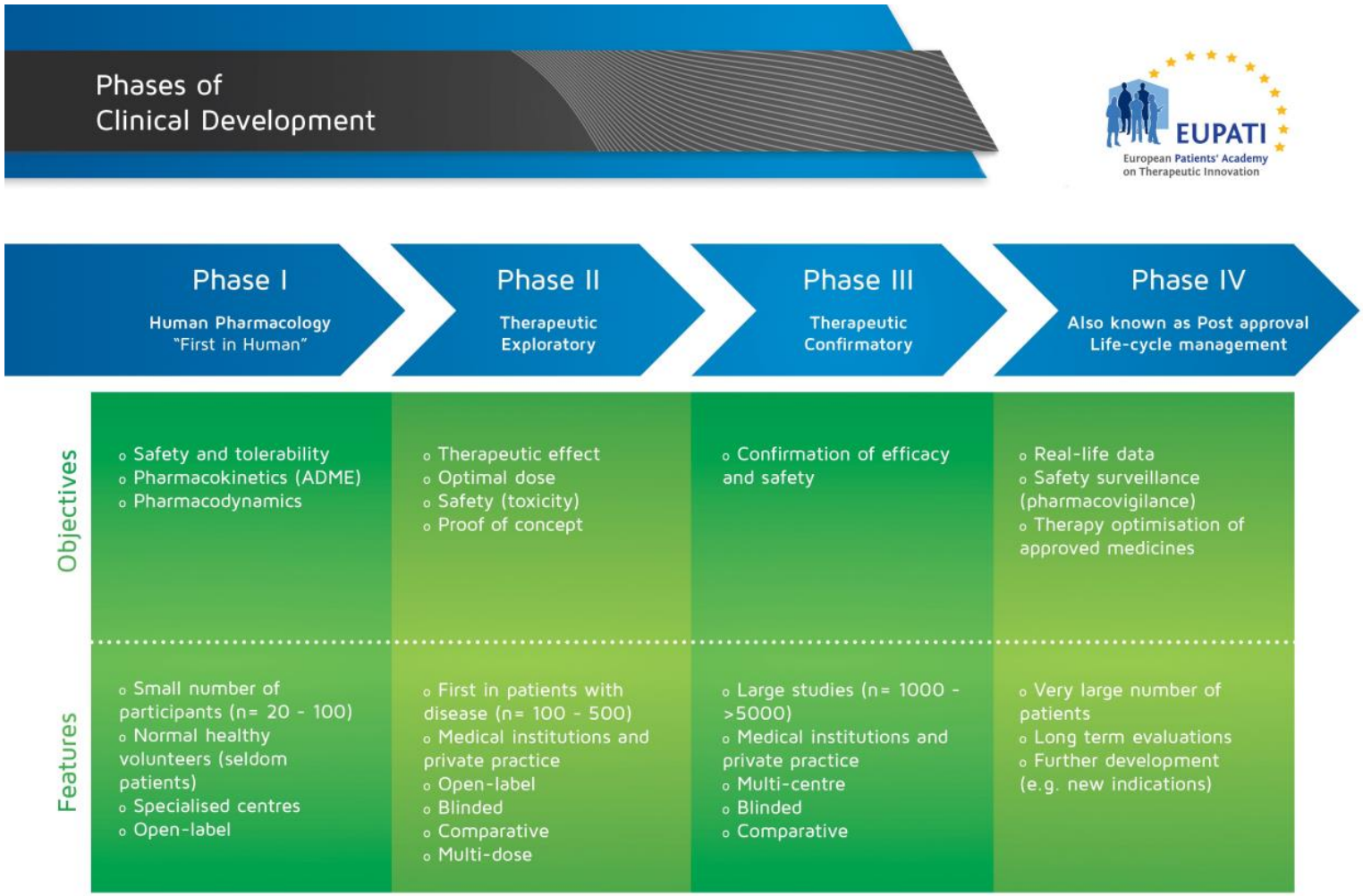
1. Clincial trials and the drug development programme



Courtesy of Dr Anup Petare

It can take 10 to 15 years for an Investigational New Drug (IND) to proceed to a New Drug Application (NDA) and receive approval.

Clinical trials and the drug development programme



Courtesy of European Patients' Academy on Therapeutic Innovation

Clinical trials play a key role in Phases I, II and III.

2. Variability in patient response to treatment

Consider a Phase III trial comparing the current "Standard of care" (Treatment A) with a new drug (Treatment B).

When a group of patients are given a particular drug, their responses to this treatment can vary considerably.

Some are more seriously ill,

Some may not take the medication as prescribed,

Genetic factors may influence the course of the disease.

We can recruit a large number of patients to allow such variation to "average out", revealing the true effect of the new treatment.

How should we allocate patients to Treatments A and B?

According to some systematic scheme? Or at random?

An example of a Phase III clinical trial

Example: Treatment for hypertension (High blood pressure)

Treatment A: Control, the standard drug in current use

Treatment B: New drug

Define the study's "primary endpoint" to be change in Systolic BP,

Initial SBP − SBP after 6 months

We decide to recruit 200 patients.

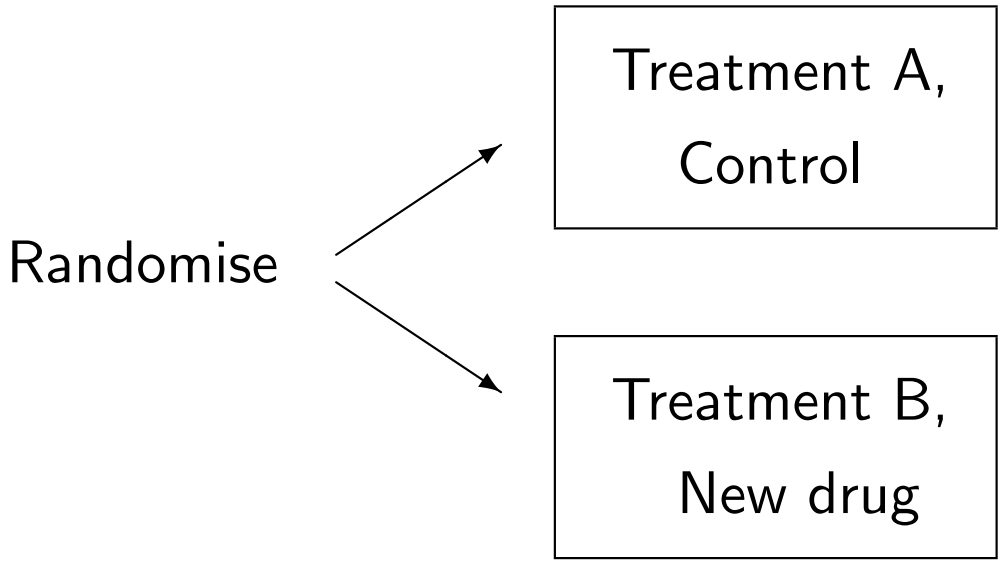
On admission to the study, each patient is allocated at random to Treatment A to Treatment B.

We record the prognostic variables:

Sex, Smoker/Non-smoker, Body Mass Index (BMI)

Example: Treatment for hypertension

Treatment allocations are generated randomly as each patient enters the trial.

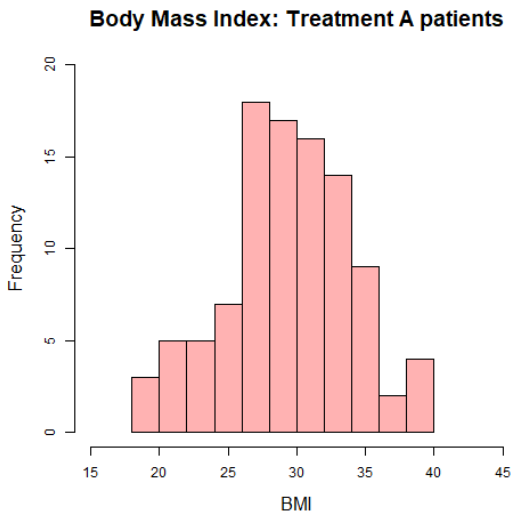


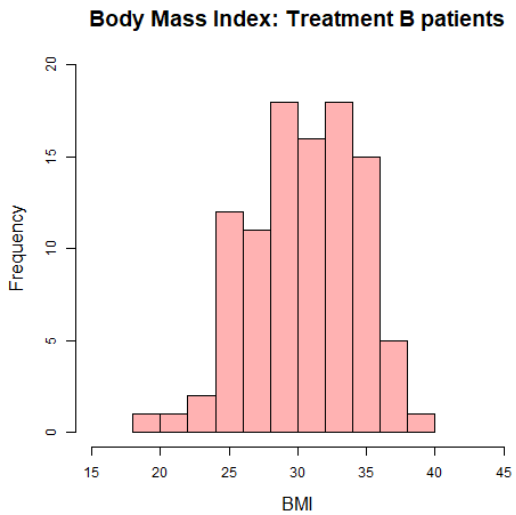
In a "double blind" trial, neither the patient nor the physician knows which drug is being administered.

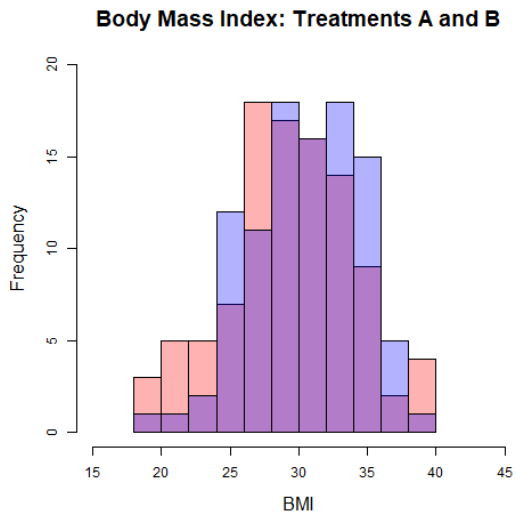
Blinding helps avoid any bias, intentional or unintentional, in the study results.

Example: Results of randomisation

Treatment	Male	Female	Non-Smoker	Smoker
A	51	49	68	32
B	56	44	62	38



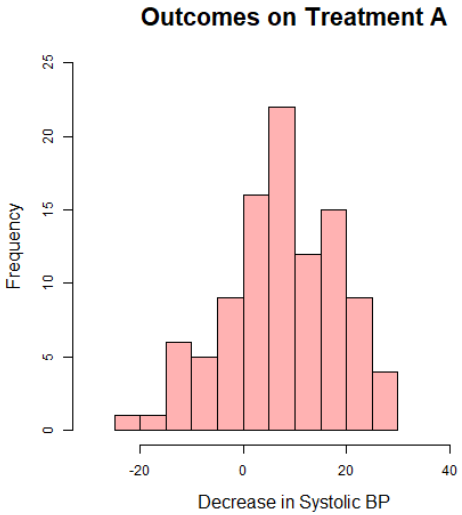


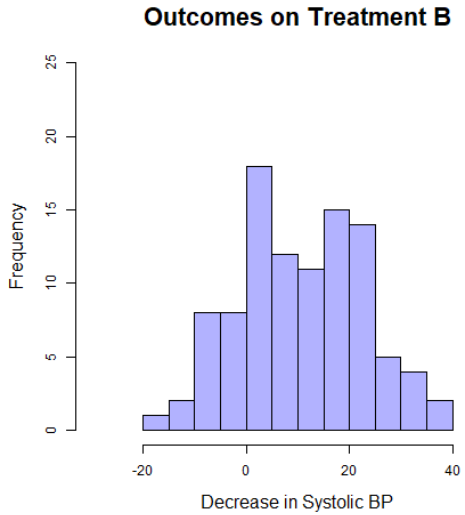


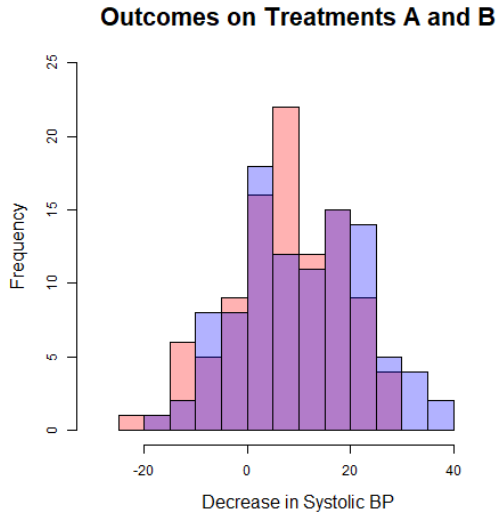
Randomisation automatically balances for variables we consider to be relevant — and for others we may be unaware of.

Example: Analysing the results

We can compare the reductions in SBP on Treatments A and B.







Mean decreases in SBP are

7.5 on Treatment A,

11.1 on Treatment B.

Does this difference represent a real treatment effect?

Example: A randomisation test for significance

Suppose there is no difference between treatments, so a given patient will respond in the same way to either Treatment A or B.

How likely is it that an observed difference between Treatments A and B should be as large as

$$\bar{X}_A - \bar{X}_B \;=\; 11.1 - 7.5 \;=\; 3.6\,?$$

We can answer this question by creating artificial data sets — where each patient keeps his or her observed response but we allocate treatment labels, A and B, at random.

The values produced for $\bar{X}_A - \bar{X}_B$ form the distribution of this statistic under the null hypothesis of "No treatment difference".

If our observed value of $\bar{X}_A - \bar{X}_B$ is found to be unlikely under this distribution, we conclude there really *is* a treatment effect.

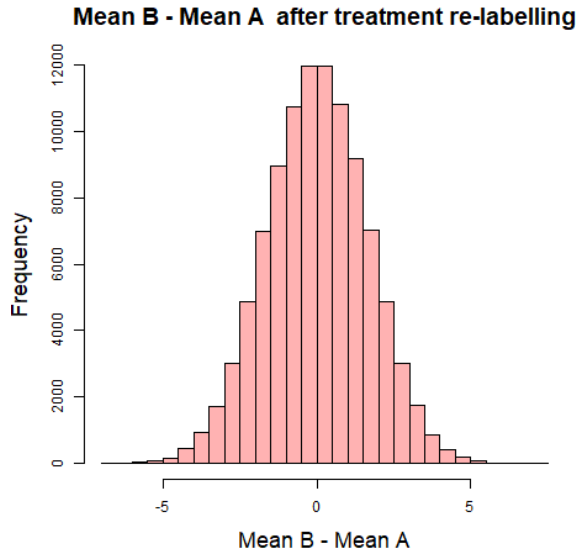
Example: A randomisation test for significance

One result of re-randomising treatment labels:

Patient number	Original treatment	Decrease in SBP	Re-randomised Treatment
1	A	10.0	A
2	A	3.8	B
3	A	12.7	A
4	A	10.9	A
...	...	...	...
101	B	21.4	B
102	B	24.6	A
103	B	−0.2	A
104	B	−1.4	B
...	...	...	...

Now: $\bar{X}_A - \bar{X}_B = 7.9 - 10.7 = -2.8.$

Example: A randomisation test for significance



I created 100,000 artificial data sets.

In 1,333 cases the data gave

$$\bar{X}_A - \bar{X}_B \geq 3.6$$

If each patient's response is the same under either drug,

$$P(\bar{X}_A - \bar{X}_B \geq 3.6) \approx 1333/10^5 = 0.013.$$

So, under the null hypothesis of no treatment difference, seeing such an extreme result in our data is unlikely.

Hence, we say we "Reject the null hypothesis at the (one-sided) significance level 0.013".

3. Reducing the noise

(a) Stratified sampling

We could modify the rule for randomising patients to Treatments A and B so that allocations are balanced with respect to to Sex, Smoker or Non-smoker, and High or Low BMI.

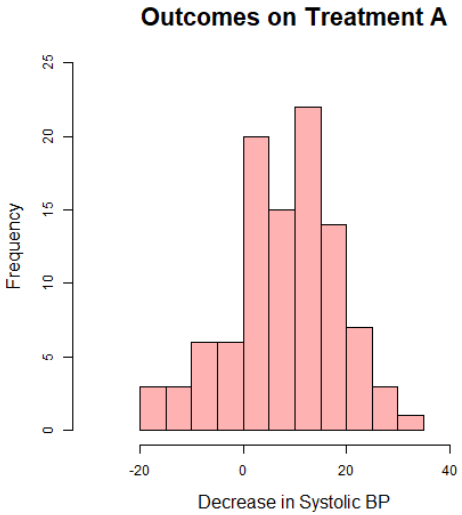
Suppose this leads to the following numbers of patients, with equal allocations to Treatments A and B, 15:15 or 10:10, in each cell.

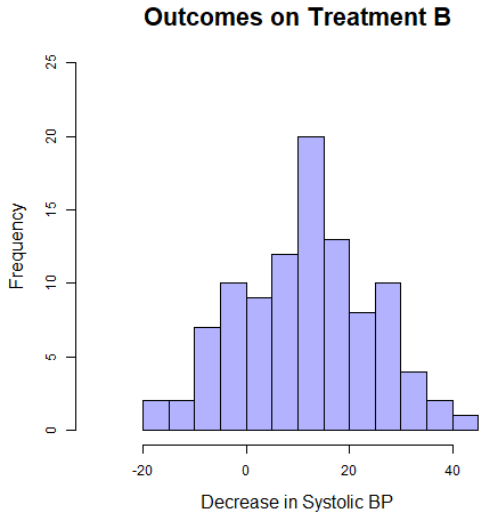
	Non-smoker		Smoker	
BMI	Male	Female	Male	Female
< 30	30	30	20	20
> 30	30	30	20	20

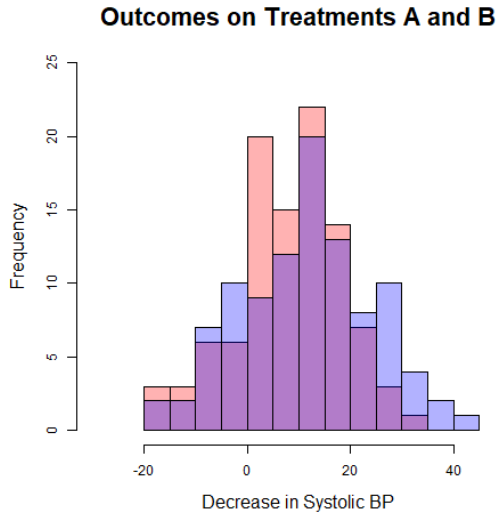
Now, our randomisation test should consider other randomisations with the same total numbers on Treatments A and B within each of the 8 cells.

Stratified sampling

Suppose we observe the following reductions in Systolic BP.







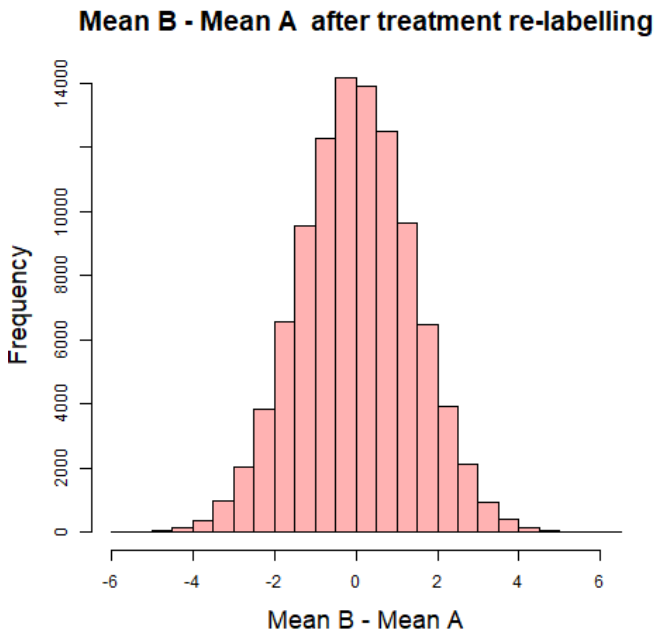
Mean decreases in SBP are

8.1 on Treatment A,

11.6 on Treatment B.

Does the difference $11.6 - 8.1 = 3.5$ show a real treatment effect?

Stratified sampling



Again, I generated 100,000 artificial data sets.

Values of $\bar{X}_A - \bar{X}_B$ were less variable this time, ranging from -4 to 4 rather than -5 to 5.

In 624 cases the data gave

$$\bar{X}_A - \bar{X}_B \geq 3.5$$

If each patient's response is the same under either drug,

$$P(\bar{X}_A - \bar{X}_B \geq 3.5) \approx 624/10^5 = 0.006.$$

So we "Reject the null hypothesis at significance level 0.006".

Stratification removes variability due to Sex, Smoker/Non-Smoker and High/Low BMI, giving a more powerful study.

Comments on randomisation tests

Randomisation tests have a place in the historical development of statistical methods.

R. A. Fisher and O. Kempthorne argued that randomisation is fundamental and model-based analyses are only acceptable because they give (approximately) the same answer as randomisation tests.

Nowadays, model-based analyses have become standard. They are easy to implement and allow more complex modelling of data.

But, there are lessons to learn from randomisation tests.

In design: Randomisation facilitates a direct comparison of treatments, dealing automatically with confounding factors.

In analysis: If special treatment allocation (randomisation) rules are applied, these must be respected in the data analysis.

(In many clinical trials, investigators use "Permuted Block Designs" then ignore this in their analysis.)

Reducing the noise

(b) A paired comparison

Consider a clinical trial investigating Treatments A and B for an eye condition, which involve applying a drop of medication to each eye.

Since each patient has two eyes, we can allocate Treatments A and B randomly to the Left eye and the Right eye.

In taking the difference in response between the two eyes of each patient, we remove the "patient effect".

Reducing the noise

(c) Crossover trials

It is rare that one can administer two treatments simultaneously to the same patient. However, we may be able to administer different treatments sequentially to the same patient.

A two-period crossover design

In designing and analysing a crossover design, we need to

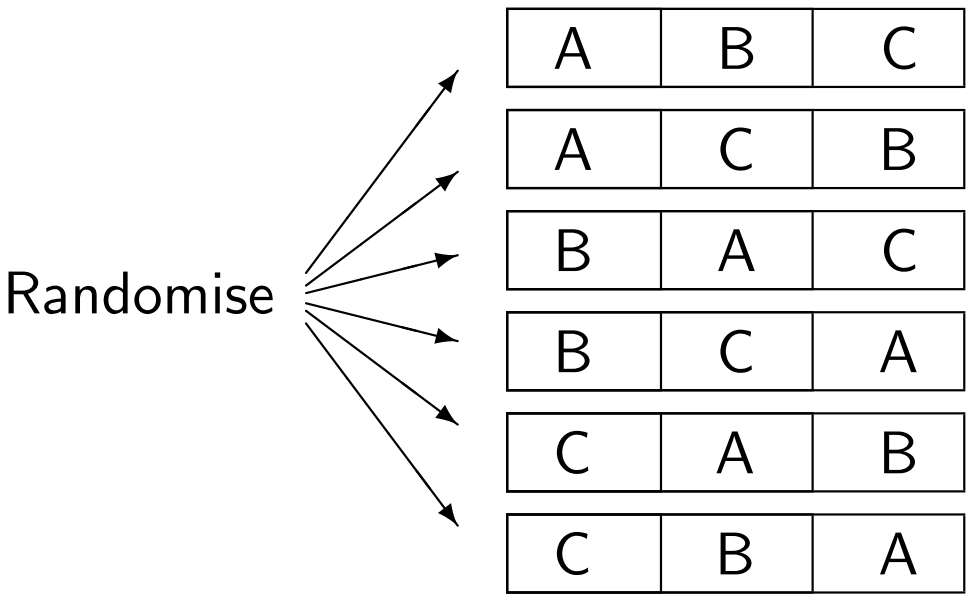
Allow time for the previous treatment to "wash out",

Consider possible "period effects" and "carryover effects".

The TRIMASTER study

TRIMASTER is a 3-period crossover trial in Type 2 Diabetes.

Each patient receives all 3 treatments in a random order.



Third line treatments

A: DPP4 Inhibitor,

B: SGLT2 Inhibitor,

C: Thiazolidinedione

This trial is currently under way: see

https://clinicaltrials.gov/ct2/show/NCT02653209

Aims of the TRIMASTER study

The TRIMASTER study aims to identify subgroups of patients who respond well to a particular treatment — leading to stratified treatment or "personalised medicine".

The patients

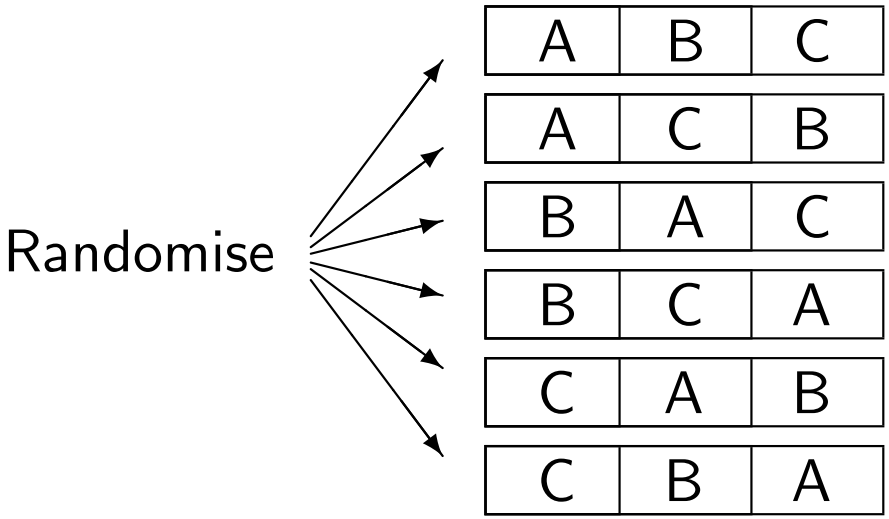
The trial will recruit 600 patients, aged 30-80. Patients will have been on two classes of therapy for at least 3 months, and have HbA1c $> 58\,mmol/mol$ (7.5%) (a high blood glucose level).

Outcomes

After each treatment period, glycaemic response (HbA1c) will be measured.

At the end of the study, patients will be given feedback on their response to each therapy. Then, each patient will be asked which treatment they would take long term and reasons for this choice.

The TRIMASTER study: Statistical analysis



The model for the observed responses should include:

- Patient effects,
- Drug effects,
- Period effects — or possibly a time trend for each patient.

Then we can look for subgroups of patients who respond well to a particular treatment.

4. The placebo effect

Placebo (*Latin: I shall please*)

A dummy medicine containing no active ingredients.

When a new drug is compared to a "control", we want to see the effect of the drug itself — rather than any other response that may arise from the patient's being involved in a clinical trial.

Patients know they are in a clinical trial — informed consent is required — but they should not know the treatment they receive.

To maintain this blinding to treatment, patients on the control arm are given a dummy treatment or "placebo".

More about the placebo effect

An example

In one study, people were given a placebo and told it was a stimulant.

After taking the pill, their pulse rate sped up, their blood pressure increased, and their reaction speeds improved.

When people were given the same pill and told it was to help them get to sleep, they experienced the opposite effects.

The power of suggestion

When patients expect side effects such as headaches, nausea, or drowsiness, they can show these reactions to an inert placebo.

Side effects associated with the active drug are seen to occur for patients in **both** treatment arms.

This is called the **Nocebo effect** (Nocebo, *Latin: I shall harm*).

More about the placebo effect

See the webpage 10 Crazy Facts About the Placebo Effect

A placebo still works even though you know it is a placebo

The colour of a placebo pill affects how well it works

Placebo effects have become more powerful over the years

Placebo surgery is effective in curing injuries

Placebo has an evil twin named Nocebo

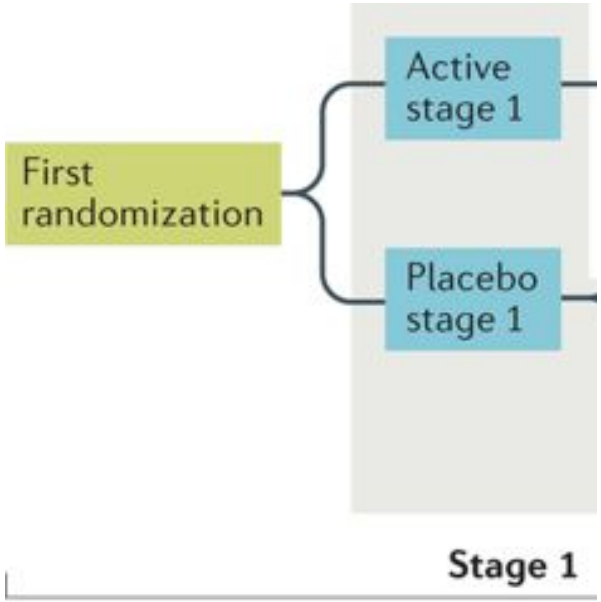
Placebo also occurs amongst dogs (and other animals)

You can placebo yourself into inebriation

A clever experimental design

In studies of anti-depressants, placebo response rates can be as high as 40%, making it hard to show that a new drug is effective.

The Sequential Parallel Comparison Design (SPCD) aims to identify patients who will not exhibit a placebo response and then to compare treatments within this group.

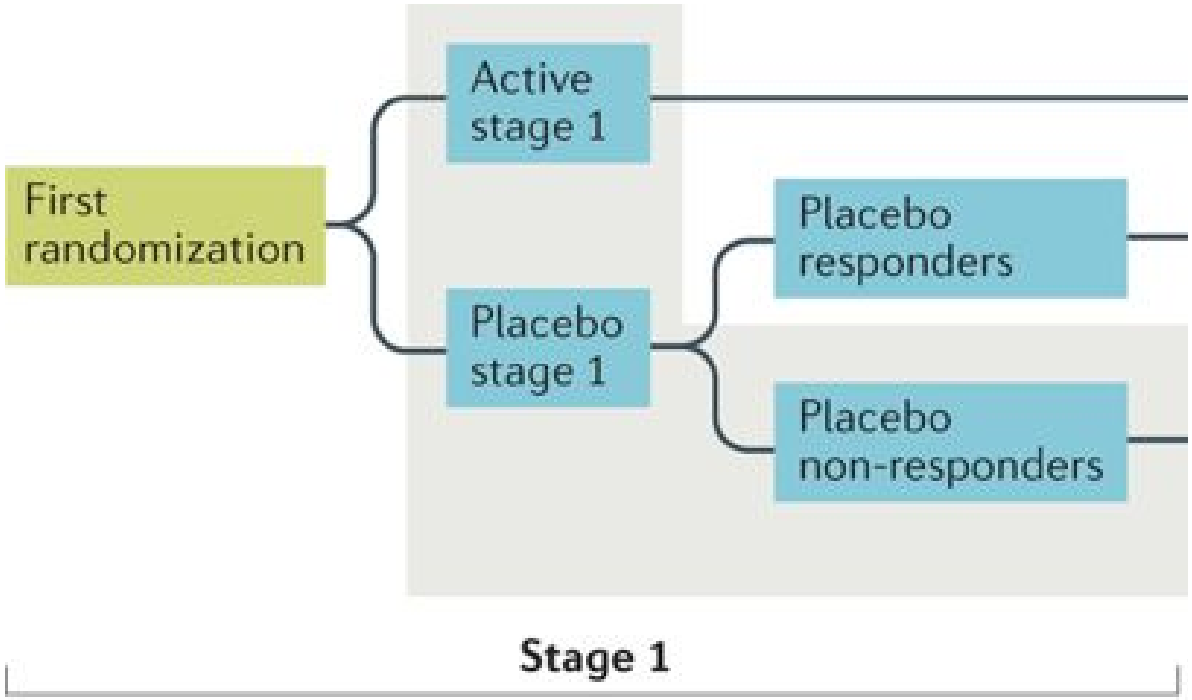


Courtesy of Nature Reviews

In Stage 1, the majority of patients are randomised to placebo.

The Sequential Parallel Comparison Design

SPCD, after Stage 1 results:



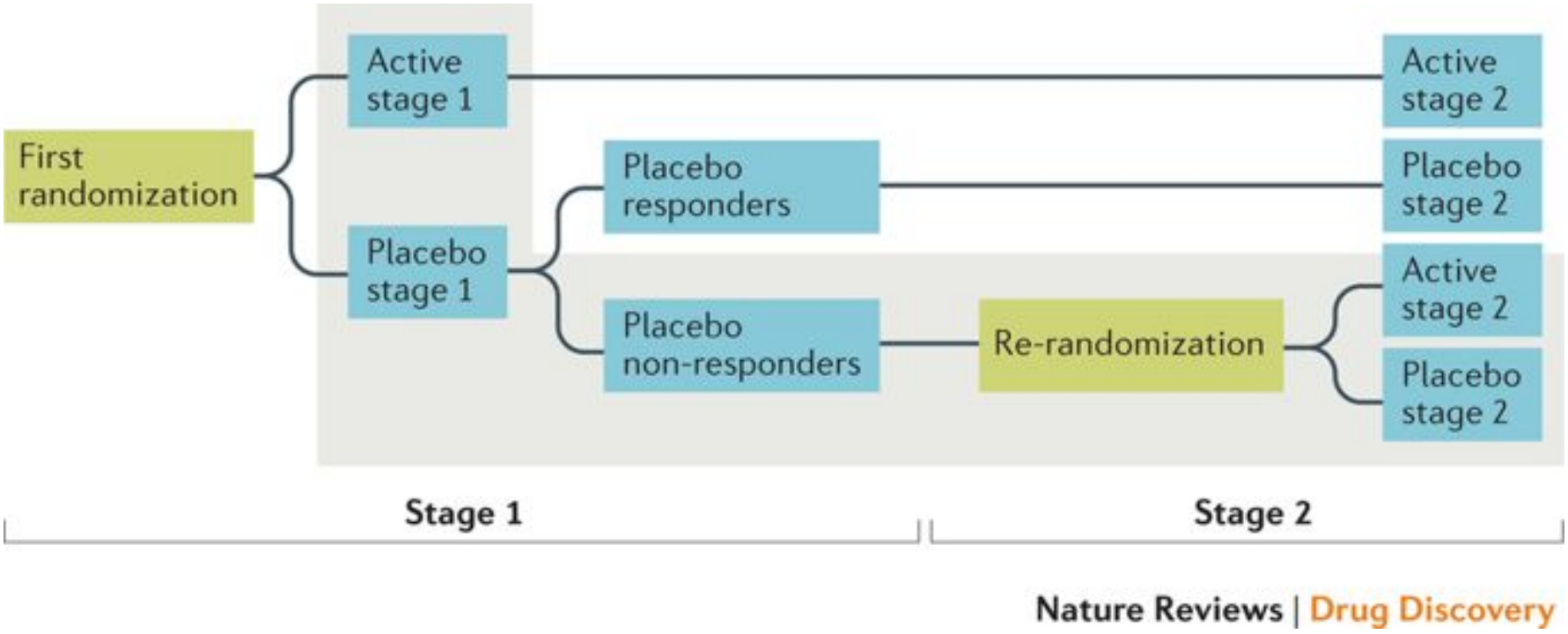
Courtesy of Nature Reviews

The results of Stage 1 define a group of patients who did not respond to placebo.

It is this group who will become the main focus of the study.

The Sequential Parallel Comparison Design

SPCD, after Stage 2 re-randomisation:



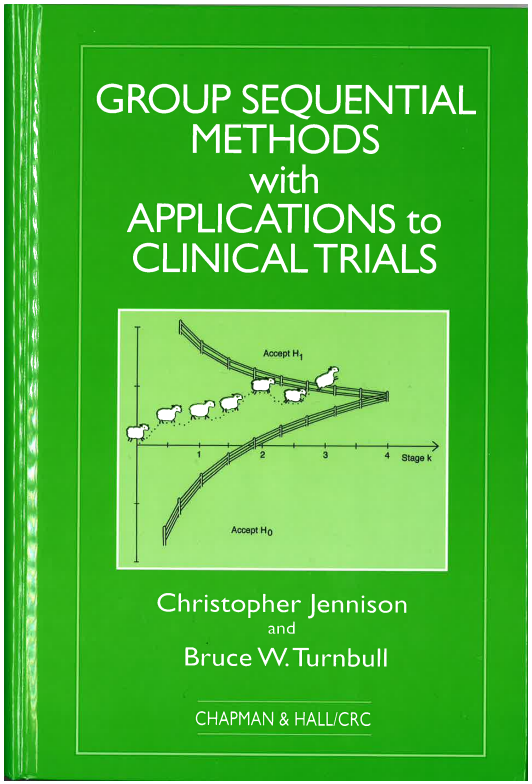
Data from groups in the grey box are pooled for the final analysis.

An SPCD was used in the trial of the anti-depressant ALKS 5461.

Placebo response rate was 26% in Stage 1 and 15% in Stage 2.

5. Group sequential and adaptive clinical trial designs

I have worked on statistical methods for clinical trials since my PhD research.



I have written many journal papers with Bruce Turnbull.

Our book has become a standard text on methods for monitoring clinical trials and stopping rules that allow a trial to be terminated early, for either positive or negative reasons.

In recent years, there has been great interest in "adaptive designs" which allow investigators to change a trial after it is under way.

Group sequential and adaptive clinical trial designs

It is not usual for statistical methodology to make headlines in the Wall Street Journal:

MARKETPLACE

THE WALL STREET JOURNAL.

TECHNOLOG

ent

The as

MONDAY, JULY 10, 2006 B1

FDA Signals It's Open to Drug Trials That Shift Midcourse

By Anna Wilde Mathews

CLINICAL TRIALS of medicines are traditionally performed in a "blinded" fashion so that the findings will remain secret until the studies are completed. But regulators and the pharmaceutical industry are increasingly interested in starting to use a very different model that lets studies change as they go along, based on early results.

Drug companies have begun to perform such adaptive trials for their new medicines, hoping for more efficient tests that could save millions of dollars. The Food and Drug Administration, meanwhile, is sending increasingly encouraging signs that it is open to considering the results of such trials. In a move that could lay the groundwork for greater future use of such studies, Scott Gottlieb, an FDA deputy commissioner, is set to announce today plans to develop regulatory guidelines for adaptive trials. The FDA has also put together an internal team to work with its drug-review divisions on the adaptive designs, which are statistically complex.

"We think it's time to start exploring the appropriate use of these designs in the appropriate situations," says Robert T. O'Neill, director of the FDA drug center's office of biostatistics. Over the past year, all of the FDA drug-review divisions have seen at least one adaptive trial submitted by companies, he says.

The most ambitious adaptive designs would represent a big change from traditional clinical-trial practices, and the idea has sparked controversy among researchers. Now, once trials are set in motion, they are supposed to be left largely untouched until they are finished and the drug company finds out the results. One exception: The studies often have an independent data-monitoring board that has the power to shut a trial down for ethical or safety reasons.

Adaptive trials have aspects that are "fundamentally different from what we currently do," says Michael Krams, who joined Wyeth in April as assistant vice president for adaptive trials. The results of an ongoing study are watched closely, and changes to the design occur as it continues, guided by a complex plan developed in advance, typically through computer simulations. If one treatment looks more effective, a greater proportion of patients may be funneled to it. If one group of patients appears to be benefiting more, the trial might start adding a larger share of that type of person.

A Trial Basis

Some ways that adaptive designs may allow clinical trials to be adjusted based on early results:

- Route a larger share of patients to the treatment that seems to work the best
- Drop treatments that don't appear to be effective
- Add more of the type of patients who seem to be reacting best to a particular treatment
- Merge two different phases of drug development into one trial

Source: WSJ research

Pharmaceutical companies hope such approaches hold the potential for major savings, though so far they are largely focused on early-stage trials. Advocates of adaptive designs say they can involve a reduction of 30% or more in the number of patients needed in a trial, and can save time as well. They also say that adaptive trials carry major benefits for patients, who have reduced odds of getting a less-effective treatment.

"It helps us pick the winners and losers faster," says Steve Ruberg, director of global medical information sciences at Eli Lilly & Co. The company has three ongoing early-stage adaptive drug trials in areas including oncology and diabetes, he says.

Bristol-Myers Squibb Co. is planning a migraine-drug trial that will use adaptive principles to help determine how much medicine to give. The study will start with 10 to 15 different doses, far more than

Please Turn to Page B8, Column 4

After a lot of work in converting some highly innovative proposals into practical methodology, the field of adaptive clinical trial design is taking shape.

(1) Seamless Phase II-III Trials

We noted earlier that drug development proceeds through a sequence of phases.

Perhaps such a rigid approach is not necessary.

Seamless design

- A clinical trial design that combines into a single trial objectives which are traditionally addressed in separate trials (*operationally* seamless)

Adaptive Seamless design

- A seamless trial in which the final analysis will use data from patients enrolled before and after the adaptation (*inferentially* seamless)

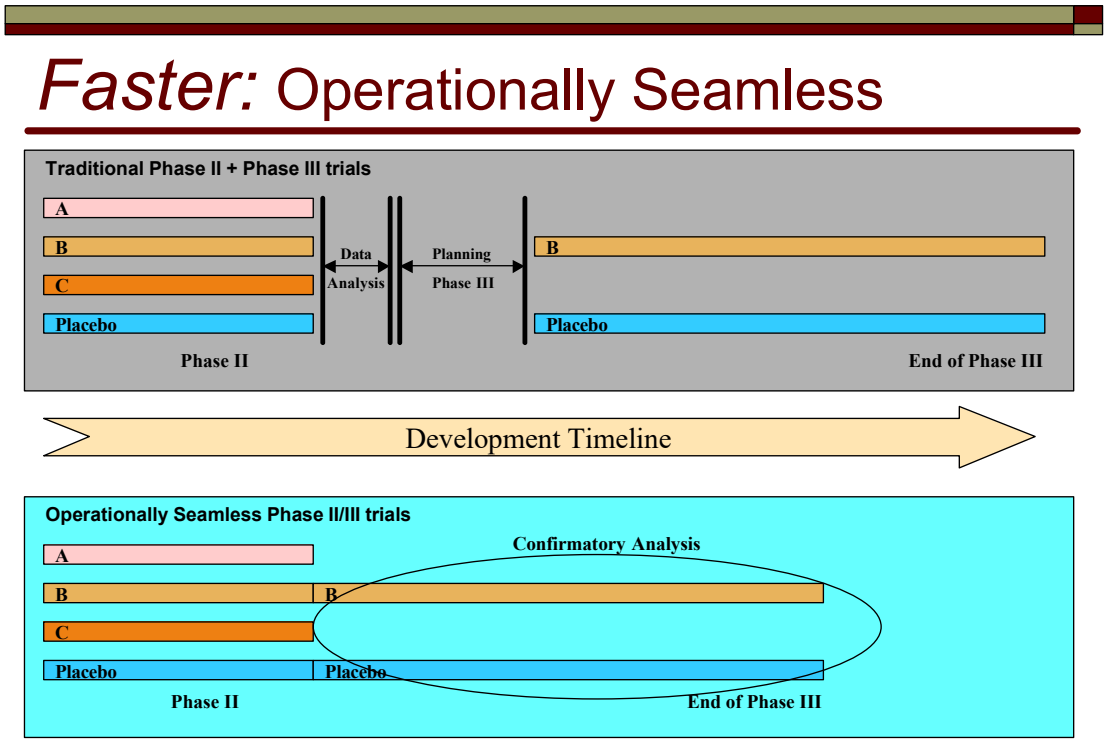
Courtesy of Dr Vlad Dragalin

What might be the advantages of merging two phases?

And how difficult would it be to do this?

Seamless Phase II-III Trials

"Operationally seamless" designs avoid delays between Phases.

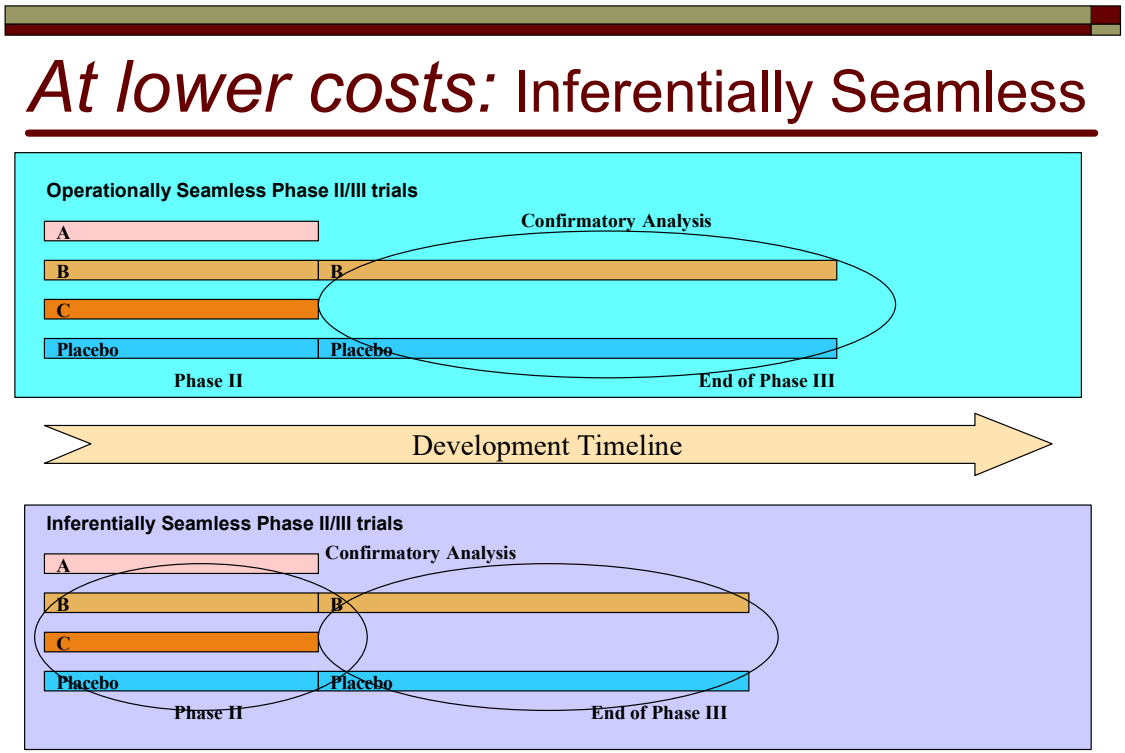


Courtesy of Dr Vlad Dragalin

Such designs require pre-planning so that rules are in place to guide the progression from the Phase II "treatment selection" stage to the "confirmatory" Phase III stage.

Seamless Phase II-III Trials

An "Inferentially seamless" design requires novel statistical methods to combine data from Phases II and III.



Courtesy of Dr Vlad Dragalin

Regulators have approved Seamless Phase II-III designs — but they impose strict conditions on how they are conducted.

(2) A Phase III survival study with treatment selection

A Phase III trial of cancer treatments comparing

Experimental Treatment 1: Intensive dosing

Experimental Treatment 2: Slower dosing

Control treatment

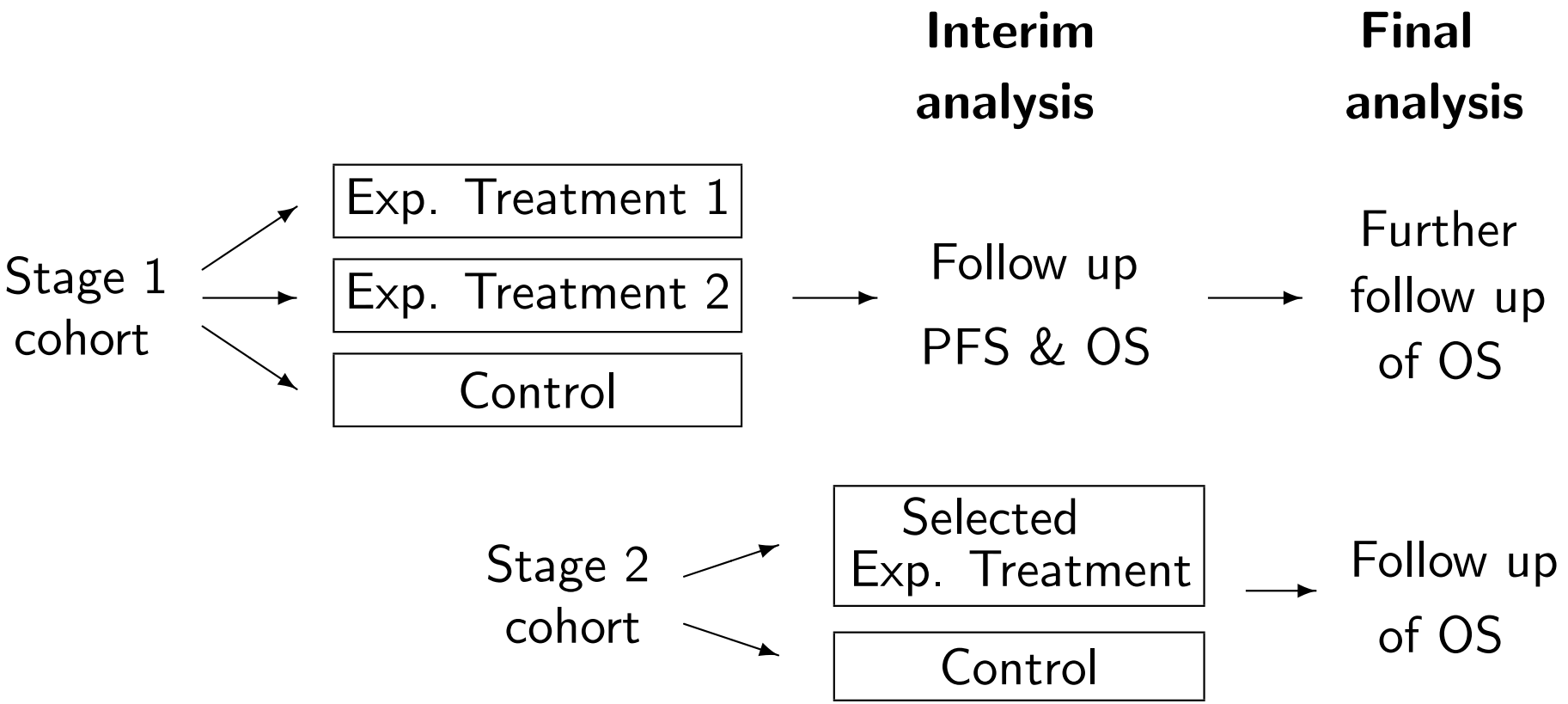
The primary endpoint is Overall Survival (OS).

At an interim analysis, information on OS, Progression Free Survival (PFS), PK measurements and safety will be used to choose between the two experimental treatments.

Note that PFS is useful here as it is more rapidly observed.

After the interim analysis, patients will only be recruited to the selected treatment and the control.

Overall plan of the trial



At the final analysis, we test the null hypothesis that Overall Survival is no better on the selected treatment than on the control.

Special methods are needed to handle data from the continued follow-up of Stage 1 patients.

Group sequential and adaptive clinical trial designs

There are other types of Phase III adaptive design:

- Multi-arm multi-stage designs,
- Designs with sample size "re-estimation",
- "Enrichment" designs that can focus, adaptively, on a subgroup of patients.

I am currently working with research students on

- Phase III trials with a survival endpoint where longitudinal data on a biomarker can help make an early stopping decision,
- Joint planning of Phase II and Phase III trials — for a single drug or for a portfolio of projects,
- Optimising the design of Phase I "First in Human" trials.

6. Conclusions

Experimental design plays a critical role in clinical trials.

There are practical issues:

- Randomisation of patients to treatment,
- Blinding of patients and physicians,
- Placebo drugs.

On a more mathematical note, statistical methodology

- Provides data analysis and inferences from trial data,
- Underpins innovative trial designs,
- Can help find effective new treatments sooner.