# Clinical Biostats series: Cross-over designs

Assoc. Prof. Cameron Hurst
cphurst@gmail.com

Faculty of Medicine,
Chulalongkorn University

5th July, 2560

# Preamble

Researchers in many clinical disciplines often want to investigate the efficacy of new therapies. Ideally, we do this using Randomized controlled trials (RCTs), a study design that tends to provide the highest level of evidence

Today, I will cover (albeit briefly) an RCT design that is often used in many clinical areas, but is certainly prevalent in particular areas: **Crossover RCTs**. Specifically I will cover:

- Rationale and design considerations
- Statistical methods for the analysis of Crossover RCTs
- Sample size considerations

Some of you may have covered these designs before, but I hope to elaborate on some issues that you may have not have appreciated when you first learnt about crossover studies.

# Yes.....but why????

But why do you need a biostatistician to come here and talk to you about this. Statistical methods associated with most RCTs are generally trivial, aren't they??

The first thing I will say is that the design, execution and analysis of data from crossover designs is a little more complicated than standard parallel designs.

**SO WHY DO WE PUT US THROUGH THIS PAIN? WHAT WOULD WE HOPE TO GAIN?**

**ANSWER: In a nutshell, Efficiency (MUCH smaller sample sizes needed)**

# What we cover today

1. Standard parallel designs
2. Crossover designs
3. Analysis of data from crossover designs
   - Methods
   - A worked example: Music therapy in in rehabilitation
4. Sample size for crossover designs
5. A quick exercise

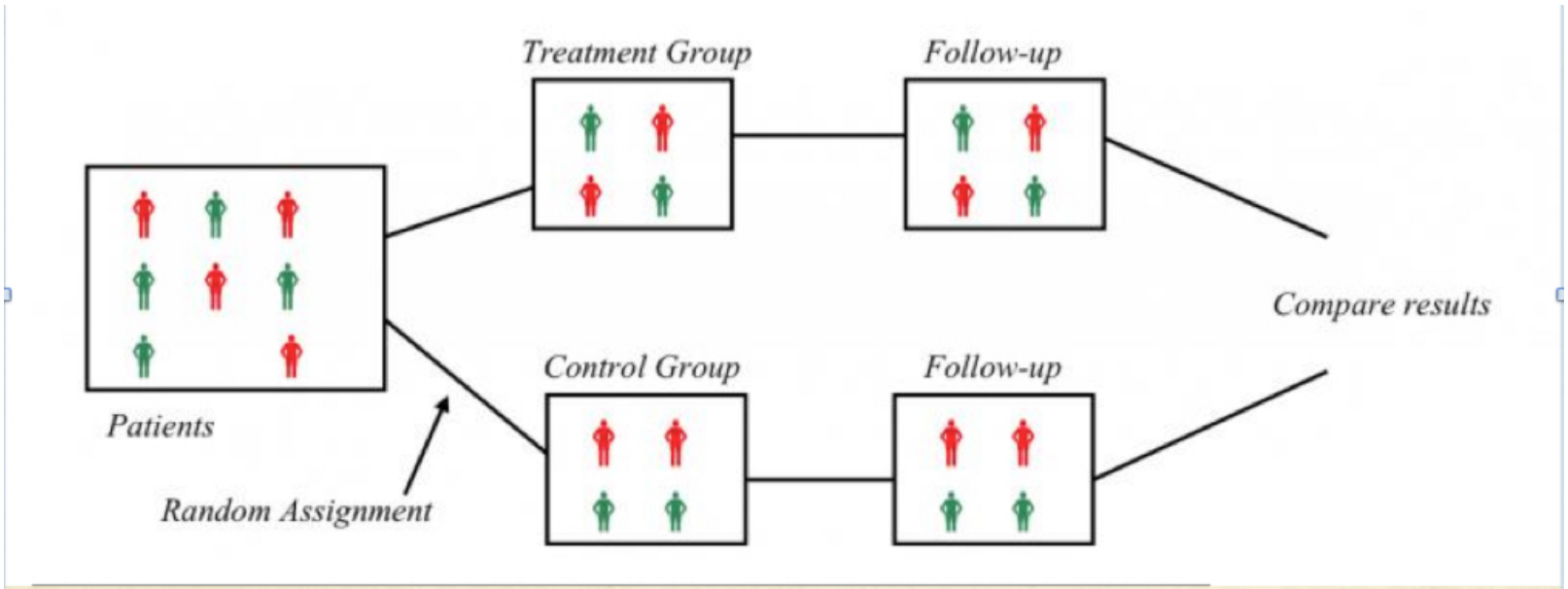
# A brief revision of RCTs

- Recall that a randomized controlled trial (RCT) is a study design that randomly assigns participants into a group (usually, an experimental group or a control group). As the study is conducted, the only expected difference between the groups in a randomized controlled trial (RCT) is the outcome variable being studied.
- Basically, RCTs considered to have the highest level of evidence of any of the 'primary data' study designs.
- However, not always feasible and/or ethical.
- Unlike **observational studies**, an RCT represents an **experiment**: the researcher directly **intervenes** in the patient experience.

# Parallel RCTs: The basic idea

The most common form of RCT is the standard two-arm parallel design:

Figure: Standard two-arm parallel RCT design



# Parallel RCTs

- A standard two-arm **parallel** RCT is one where (typically) one set of patients receives one treatment, and the remaining patients receive another treatment
- Patients are **randomized** into their treatment groups to try to balance any confounders: 'alternative explanations'
- Other mechanisms like **blinding** and **concealment** minimize other types of bias introduced by observers and/or patients knowing membership/treatment
- Importantly, in parallel RCTs different patients receive different treatments, implying that differences in patient response come from two sources:
  1. The treatment effect (what we are interested in)
  2. Between subject variation: Differential response to therapy(but **we hope** these differences 'average out')

# Sample size for Parallel RCTs

By using different patients in different arms, there is substantial within-group variation $\Rightarrow$ N has to be (comparatively) large For the two independent samples case, Continuous outcomes:

$$N_{pergroup} = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2\sigma^2}{MCD^2}$$

and for Binary outcomes:

$$N_{pergroup} = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2\sigma^2}{MCD^2}$$

**Parallel RCTs and within-group variation**

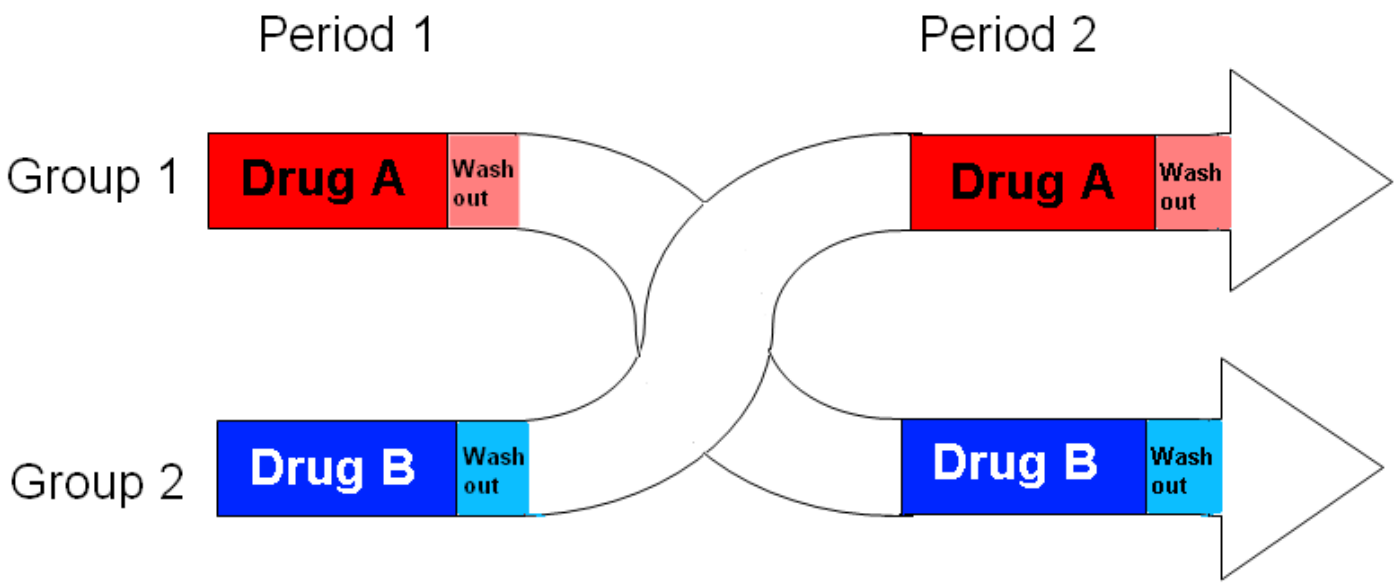
Within-group variation in parallel RCTs, $\sigma$, leads to quite large sample sizes being required to adequately power the analysis

# Cross-over RCTs

- An alternative RCT design is the **Cross-over RCT**
- This design has some major advantages (more later)
- But these advantages typically lead to being able to conduct comparative (or non-inferiority/equivalence) trials with considerably fewer patents
- Cross over RCTs are very efficient, but they do present some difficulties, and often not possible

# Cross-over designs: What's the big idea?

Figure: Basic 2 x 2 crossover design




In a 2 by 2 crossover design we have 2 treatments administered over two periods (of time)

# Cross-over designs: Problems and challenges

I have already mentioned that sample size in CO designs is MUCH smaller. Why don't we use them all the time? Answer: They also present challenges:

1. **More difficult** Often, there are lots of issues to solve when conducting a crossover RCT (especially design and analysis consideration). Some problems are even insurmountable.
2. **Some tattoos just don't wash off**. Many treatments have an ongoing effect. CO designs only work when a treatment 'washes out' $\Rightarrow$ No carryover: we can only consider treatments with an 'acute' effect for an ongoing condition. **E.G. What about a surgical intervention??**

# Cross-over designs: Problems and challenges

Challenges continues:

3. **Period vs carryover effects** The **Carryover effect** is the contamination of the second period treatment by the initial treatment (residual effects). But there may be additional 'time' effects in the **Period effect**. The period effect is the difference in results in the first period vs the second period. Is this the same? Not necessarily. E.G., **Period effect = Carryover effect + Disease progression + Summer effect**
   How do we seperate these effects? Generally, we can't

**CO designs: Costs vs benefits**

MAJOR benefits to a cross-over design (over a parallel design), but generally more difficult, and often, not even possible.

# Cross-over designs: What's the benefit?

If they are more complex (than parallel designs), why do we try as hard as we can to make them work?

The answer is quite simple. They have some MAJOR advantages: **MUCH smaller sample size required?**

**But what is it about CO-designs that make them so efficient?** The answer is that:

1. not only do we take repeated measures of the same patients (therefore better accounting for between-subject and within-subject variability)
2. but also, **EACH patient** experiences **EACH treatment**

This first source of variation is accounted for in any repeated measures design, but only in cross-over designs can we account for the second.

# Cross-over designs: Clarifying the benefits?

So, to simplify:

- We answer a 'stronger' question in CO designs: **Does THE patient do better on treatment A than treatment B?**
- Design may be more attractive to patients as EACH patient receives EACH treatment
- We remove the patient effect in this study design (indeed this is why CO designs are so much more efficient)

**UPSHOT: Under the assumption of no carryover effects, a CO design provides much more information (per patient) than a simple parallel RCT**

# Building blocks in CO designs

In CO designs there are three 'effects':

1. **Treatment effect:** Addresses our research question: Do the treatments have the same efficacy, or different? (Same as any other comparative trial)
2. **Sequence effect:** Does the sequence in which patients receive their treatments matter? i.e. Does AB = BA?
3. **Carryover effect:** Is their any residual effect of the first treatment, when a patients receives their second (or can we assume it has 'washed-out')

**Note:**

- Generally, we are only interested in the **Treatment** effect
- The **Sequence** and **Carryover** effects are just **nuisance effects** that we want to *rule out*

# Statistical analysis in crossover studies

Without going into to major detail (not within scope of this lecture), there are two main approaches to the analysis of data arising from crossover studies:

1. **Naive approaches:** These assume that there is NO period (inc. carryover) effect and breaks down the analyses into two separate (independent) components:(1) Analysis of the treatment effect (our main research objective); and (2) Analysis of the Sequence effect (which we would like to rule out)
2. **Model-based approach:** Specifically using the **mixed models**. These models include the three components that **should be** considered in a crossover study, and all in the same model: (1) The **treatment effect**; (2) the **sequence effect**; AND (3) the **period effect**

# A VERY brief intro to mixed models

Don't worry, I am not going to go into too much details about mixed models. But, I want you to note mixed models have:

- **Fixed effects**: Can be BETWEEN- or WITHIN-subject
- BUT they also include **random effects** which (in our case) tells us the identity of each patient, thereby allowing us to track them through time and treatment allocation (and take care of their 'non-independence').

## Mixed models:

- Mixed Models tracking of each patient allows analysis of data from longitudinal and other complex designs
- If you understand mixed models, you will be a MASTER OF THE (clinical research) UNIVERSE. They are, by far, the most important and useful models in all biostatistics.

# Mixed models for Crossover studies: The RHS

Simplest model (random intercept) would look like:

$$LHS = \beta_0 + \beta_{Treat} + \beta_{Seq} + \beta_{Period} + \nu_{Patient} + \epsilon$$

where: $\beta_{Treat}$ is the **treatment effect**: which we usually want to be significant ($\beta_{Treat} \neq 0$)
$\beta_{Seq}$ is the **sequence effect** (i.e. AB vs BA) which we want to non-significant ($\beta_{Seq} = 0$)
$\beta_{Period}$ is the **period effect** ($\sim$ carryover) which we want to be non-significant ($\beta_{Period} = 0$)
$\nu_{Patient}$ is a patient specific variation

## Assumptions

I have been simplistic in this model. If we believe patients differential response depends on treatment or period, we need more complex models. FUN FUN FUN.

# The LHS: Type of mixed model depends on the type of outcome

I will just mention two:

1. Linear mixed models (Continuous outcomes):
   $Y = \beta_0 + \beta_{Treat} + \beta_{Seq} + \beta_{Period} + \nu_{Patient} + \epsilon$
2. Binary logistic mixed effect model (Binary outcome)
   $Log(Odds) = \beta_0 + \beta_{Treat} + \beta_{Seq} + \beta_{Period} + \nu_{Patient} + \epsilon$

## LMMs vs Binary Logistic Mixed effect models

- The Linear mixed model differs from the binary logistic mixed effect model in the same way as linear regression differs from binary logistic regression.
- The Binary logistic effect model is a child of the Generalized Linear Mixed Model (GLMM) family

# Worked example: Improving efficacy of rehab.

## Rationale and RQ

We have patients who have undergone trauma and are in need of rehabilitation. The efficacy of rehabilitation therapy is highly dependent on patients engagement in their rehabilitation program. We believe that patients listening to music as they perform their exercises may lead to better rehabilitation outcomes.

**RQ: Does music therapy help with patient engagement in their rehabilitation therapy?**

We will use a crossover RCT to address this question.

# Worked example: Improving efficacy of rehab.

## Design: A simple 2 x 2 CO RCT

- Thirty (30) patients were equally allocated to one of two 'sequence' groups ($n_1 = n_2 = 15$)
- Two treatments were considered: A (No music); and B (Music) administered to patients while exercising
- The group 1 (AB) received the sequence of treatment (Period 1= (A) No music and Period 2 = (B) Music, whereas for tgroup 2 (BA) this sequence was reversed
- Maximum Heart rate (among many other outcomes) during the exercise session was our outcome of interest
- A superior treatment (no music or music) was deemed to be one that led to patients with a higher maximum heart rate (during their exercises)

# Worked example: Improving efficacy of rehab.

## Model

As heart rate is a continuous outcome, we used a Linear mixed model to conduct our analysis. Specifically, I will (here) use a random intercept LMM:

$$HR_{max} = \beta_0 + \beta_{Music} + \beta_{Seq} + \beta_{Period} + \nu_{PatientID} + \epsilon$$

# Results

| Source | Effect | L95 | U95 | p value |
|---|---|---|---|---|
| $\beta_0$ | 128.27 | 121.87 | 134.67 | $<0.05$ |
| Treatment B | 3.09 | 1.72 | 4.46 | $<0.05$ |
| SequenceBA | 1.71 | -7.25 | 10.67 | 0.355 |
| Period2 | -0.27 | -1.64 | 1.10 | 0.350 |

- Average $HR_{max}$ for the no music group in the first period was 128.27 (the referent)
- The "Music" effect led to a statistically significant increase in $HR_{max}$ (mean diff $= 3.09$, $95\% CI : 1.72, 4.46$, $p < 0.05$)
- Neither sequence or period could be shown to have a significant impact on $HR_{max}$

# Implications and questions:

I want let's consider some (potentially) difficult questions

## Something to think about:

1. Does the 'non-significance' of the sequence and period effect $\Rightarrow$ Everything fine?
2. Would a significant sequence and/or period effect mean (necessarily) that our results (treatment effect) aren't valid?
3. CO designs are very powerful. Do you think the 'statistically significant' increase (on average) of 3 units of $HR_{max}$ is really **SIGNIFICANT**
4. Could I have fit a better model (to reflect my suspicions about potential problems).....TOUGH ONE

# Sample size for CO designs: Continuous outcomes

Sample size calculation (continuous outcomes) for CO designs **looks** VERY simple:

$$N_{pergroup} = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma_m^2}{2\Delta_{MCD}^2}$$

Where:

1. $Z_{1-\alpha/2}$ is about our significance level ($\alpha = 0.05$);
2. $Z_{1-\beta}$ is about our power: $\beta = 0.1(90\%)$ or $\beta = 0.2(80\%)$;
3. $\sigma_m$ is our standard deviation; and
4. $\Delta_{MCD}$ represents our minimal **CLINICAL!!!!!!** difference

EASY!!!! ..... Isn't it???

# Sample size vs design: A quick look

Independant samples (i.e. standard parallel design):

$$N_{pergroup} = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2\sigma^2}{\Delta^2_{MCD}}$$

Paired samples (e.g. pre-post study):

$$N_{pergroup} = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2\sigma^2}{\Delta^2_{MCD}}$$

Crossover design:

$$N_{pergroup} = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2\sigma^2_m}{2\Delta^2_{MCD}}$$

**All else being equal**, crossover designs need $\frac{1}{4}$ as many patients as a standard parallel trial, and $\frac{1}{2}$ the number of patients as a standard paired sample (single follow-up) study

# BUT....is all else equal?

If all else was equal, we could work out the sample size of a CO trial simply by 'quartering' ($\frac{1}{4}$) the number of patients needed in a standard parallel trial, but now we come to the **NASTY part of cross-over sample size calculation,** $\sigma_m^2$

But what is $\sigma_m^2$?? Are you ready??

$$\sigma_m^2 = \sigma_{BT}^2 + \sigma_{BC}^2 - 2\rho\sigma_{BT}\sigma_{BC} + \sigma_{WT}^2 + \sigma_{WC}^2$$

where

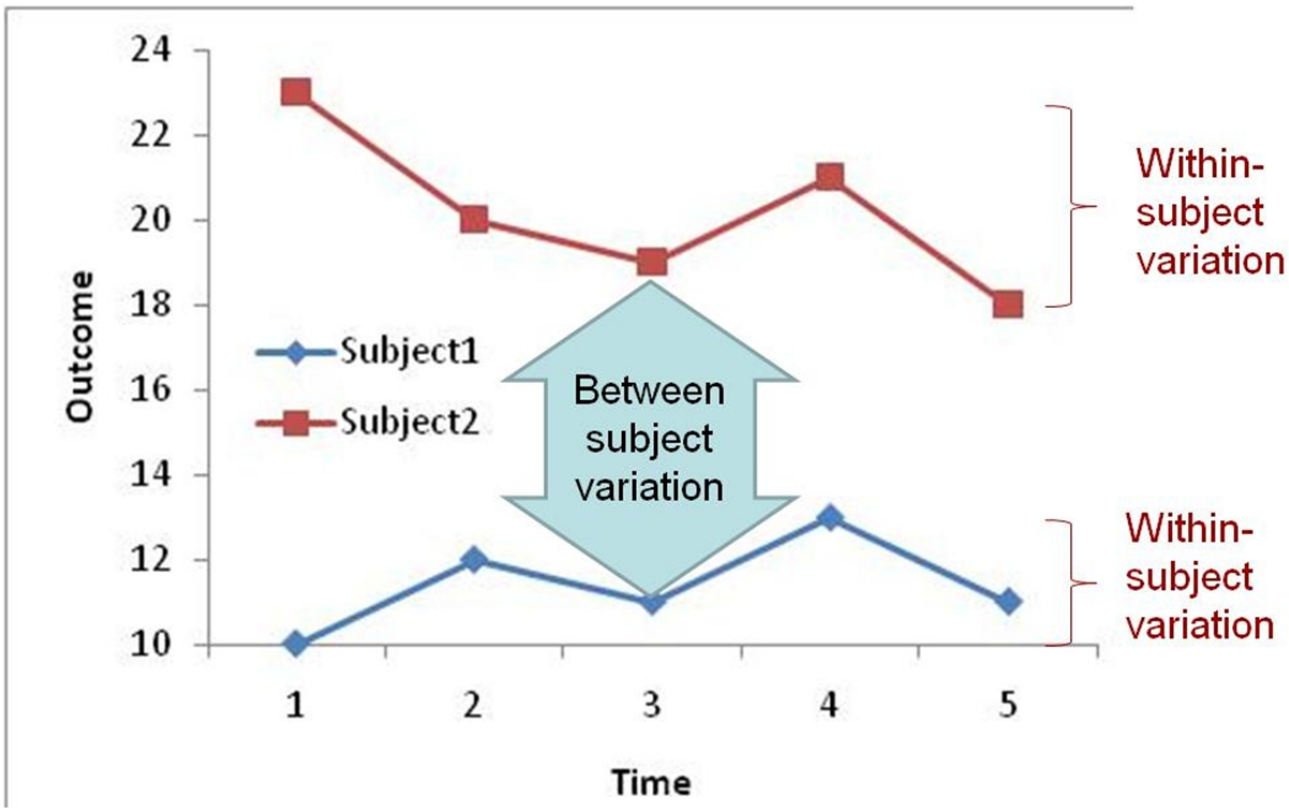
$\sigma_{BT}$ and $\sigma_{BC}$ are the between (B) subject standard deviations in the treatment (T) and control (C) groups, respectively

$\sigma_{WT}$ and $\sigma_{WC}$ are the within (W) subject standard deviations in the treatment (T) and control (C) groups, respectively

$\rho$ is the within-subject correlation

# Between vs Within subject variation

Figure: Within Vs Between variation for two individuals



# Where do we get $\sigma_{BT}$, $\sigma_{BC}$, $\sigma_{WT}$, $\sigma_{WC}$ and $\rho$?

Where do we get all of these values from? Only good pilot data will give us realistic values (papers rarely report them). In practice, we often don't know most of them so we 'guestimate' them. If we make some simplifying assumptions:

$\sigma_{BT} = \sigma_{BC}$: Differential patient response to treatment is independent of the type of treatment (BTW: same as $\sigma$ used in standard parallel trial sample size)
$\sigma_{WT} = \sigma_{WC}$: Patients repeated values within a treatment will vary approximately the same (regardless of the nature of the treatment),
then $\sigma_m^2$ simplifies to

$$\sigma_m^2 = 2(\sigma_B^2 - \rho\sigma_B^2 + \sigma_W^2) = 2((1-\rho)\sigma_B^2 + \sigma_W^2)$$

# Bringing it on home...

- CO trials have some major advantages over standard parallel trials.
  1. **Require MUCH smaller samples**
  2. **All patients receive all treatments**
  3. **Address a stronger question: Will patient do better on treatment A, then treatment B**
- But they also have some costs:
  1. **Only really work when treatment effect is 'acute' (no carryover) AND only when the disease 'survives' at approximately same level of 'severity'.**
  2. **Difficult to work out sample size**
  3. **More difficult to analyze**
- But if they are possible, the advantages FAR outweigh the disadvantages

# Exercise:

Let's quickly consider three research scenerios in which we are considering what RCT design we might use:

1. Two different methods for knee surgery
2. Two different antibiotics to treat lower respiratory tract infections in children
3. Two different therapies to alleviate nausea in cancer patients undergoing chemotherapy

# Example 1: Knee surgery

Here we will consider a functional outcome (after a standard period of recuperation) in patients undergoing two different procedures for knee surgery:

**Can we use a cross-over design?**

**If not, why not?**

# Example 2: LRTIs in children

Now we will want to compare the efficacy of a new antibiotic (against Amoxicillin) for treating lower respiratory tract infection in children. The outcome is successful infiltrate resolution at 2 weeks.

**Can we use a cross-over design?**

**If not, why not?**

# Example 3: Nausea in cancer patients

Finally, we want to consider two different drugs to alleviate nausea in patients prescribed a particular chemotherapy to treat a particular type of cancer. Both drug are to be used 'on-demand' to deal with nausea events. Both the number of nausea events and number of doses are noted. The outcome is about the drug alleviating the nausea symptoms.

**Can we use a cross-over design?**

**If not, why not?**

# Thank you

# The end