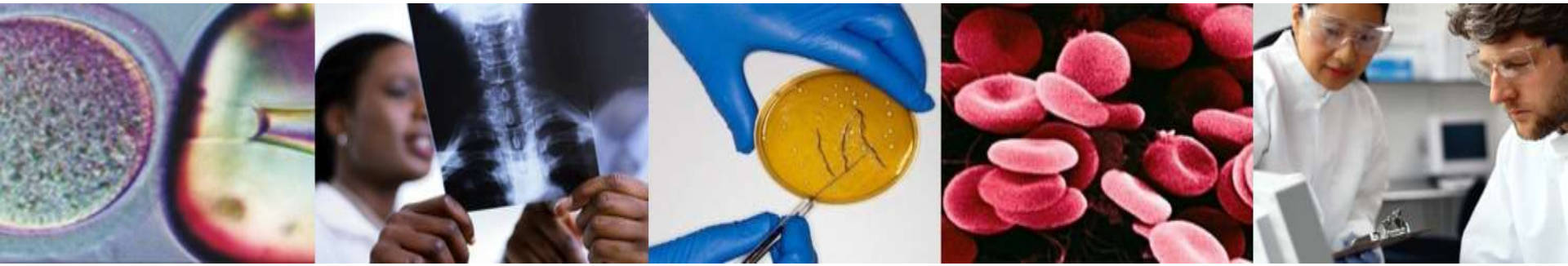
MRC | Biostatistics Unit



Multi-arm multi-stage designs: what do they offer wounds research?

James Wason

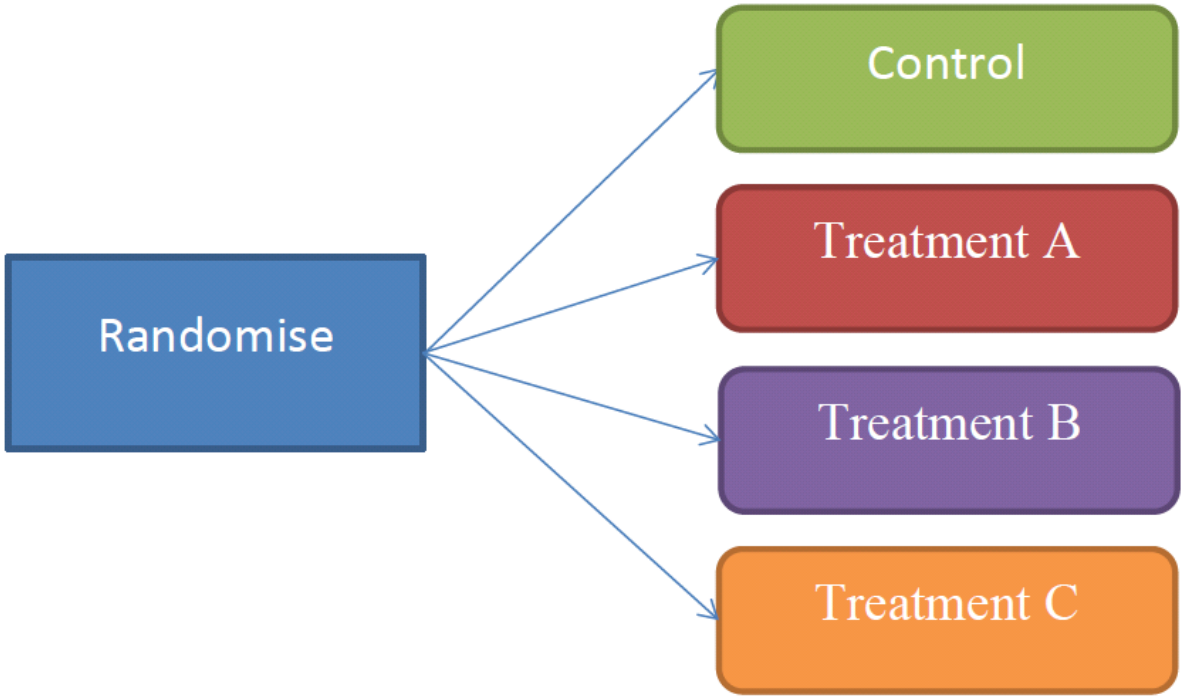
MRC Biostatistics unit

Multi-arm trials

- In many therapeutic areas there may be several possible treatments awaiting trials.
- Traditionally these would be tested one by one in separate randomised controlled trials.
- An alternative is one trial where several novel treatments are compared.

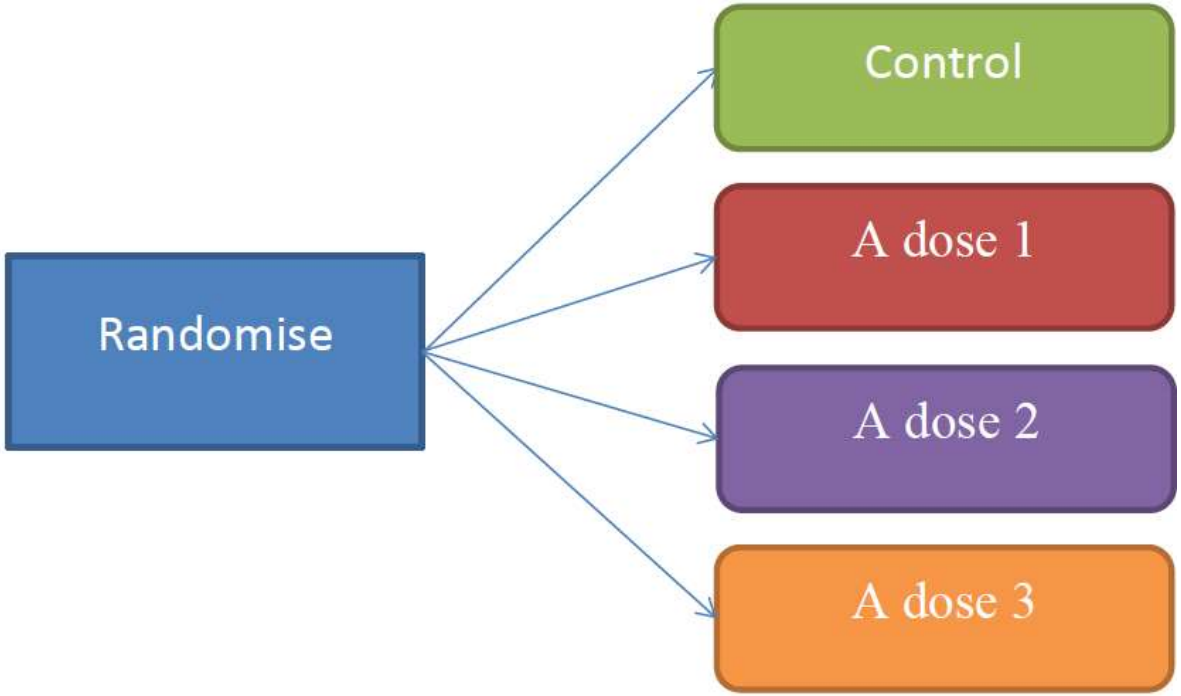
Multi-arm trials

- Multi-arm trials could be testing several distinct treatments against control:



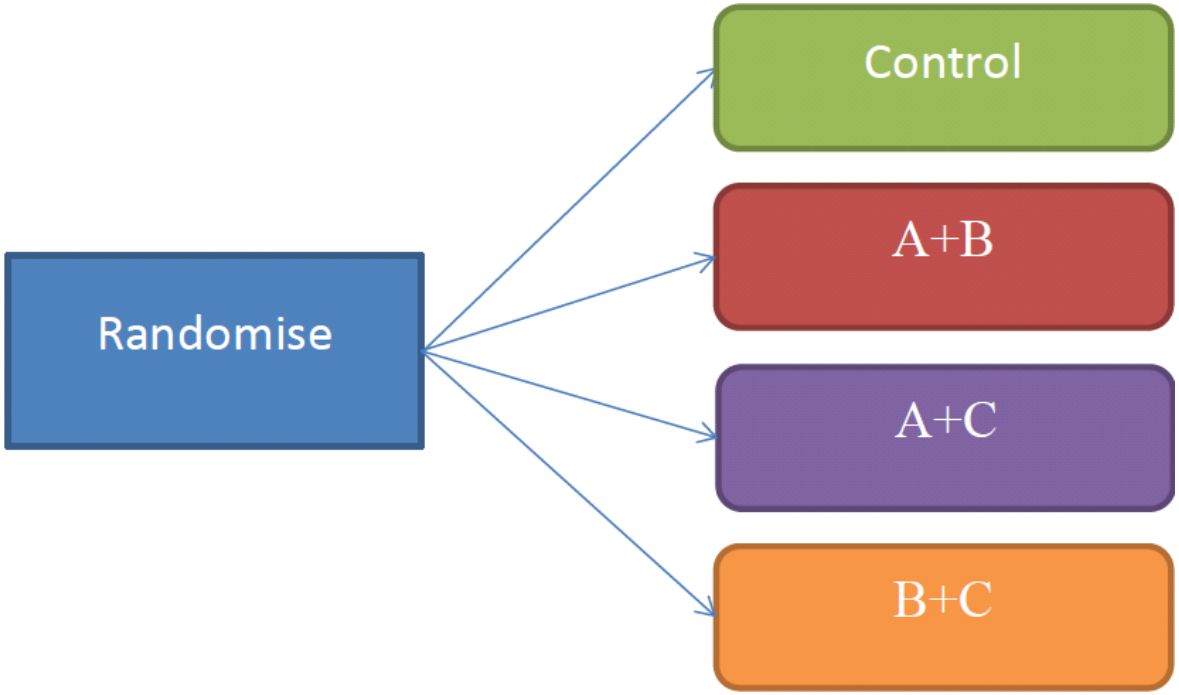
Multi-arm trials

- Or testing different doses or implementations of a treatment/strategy:



Multi-arm trials

- Or testing different combinations of treatments:



Advantages of multi-arm trials

- There are several major advantages of multi-arm trials:
  - The shared control group means that fewer patients are needed for the control treatment (low sample size needed compared to separate trials).
  - Greater number of treatments can be tested with limited number of patients.
  - Evidence that multi-arm trials are more popular with patients, as they are more likely to get a new treatment (higher recruitment rates, Parmar et al, Lancet 2014).

Multi-arm multi-stage trials

- Can also add interim analyses to a multi-arm trial (multi-arm multi-stage).
- At the interim analyses, modifications can be made based on the results of the trial so far.
- For example:
  - Ineffective treatments could be dropped;
  - The allocation to different arms could be changed;
  - New arms could be added in.
  - Trial could be stopped for efficacy if effective treatment found.
- Interim analyses add additional efficiency and make the trial more ethical.

Types of multi-arm multi-stage trials

- Several types of MAMS trials, depending on what happens at the interim analyses.

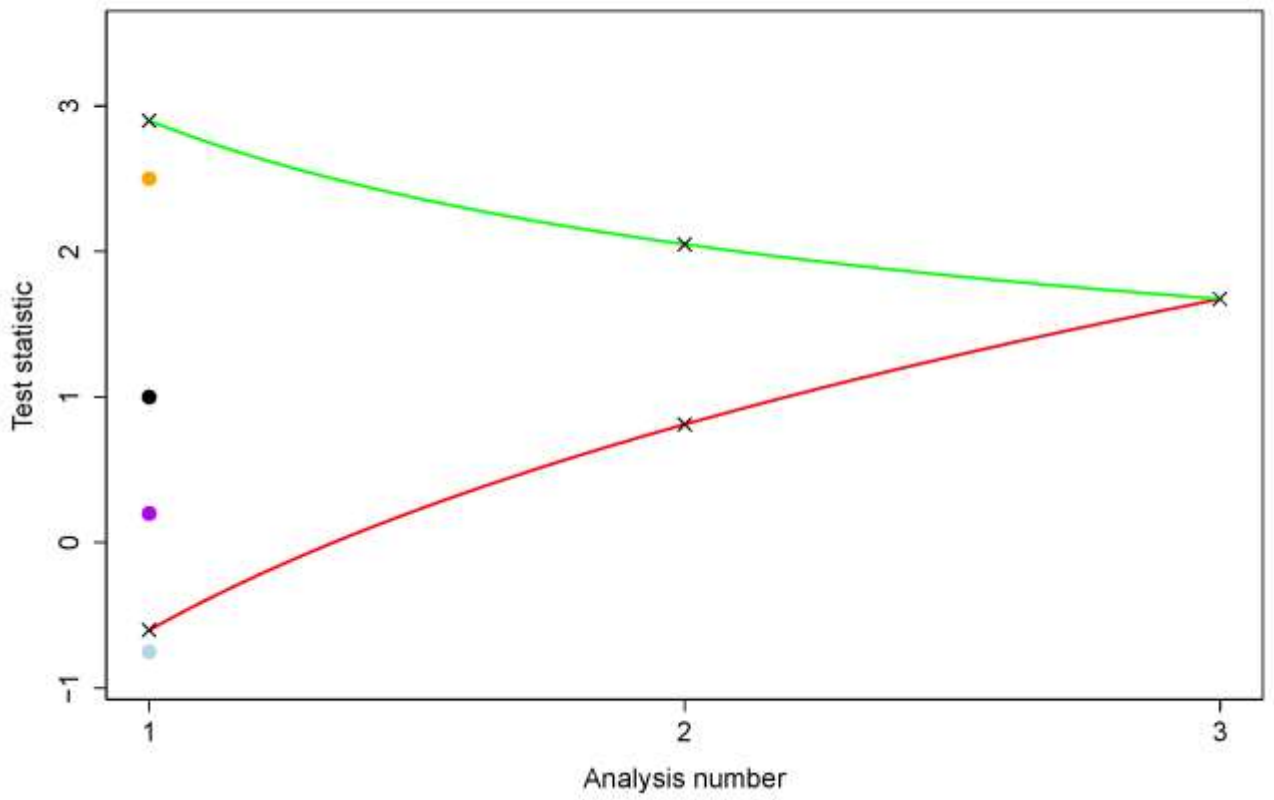
1. Group-sequential multi-arm multi-stage designs
2. Multi-stage drop-the-losers designs
3. Adaptive randomisation multi-arm trials.

- I will cover the first two in this talk.

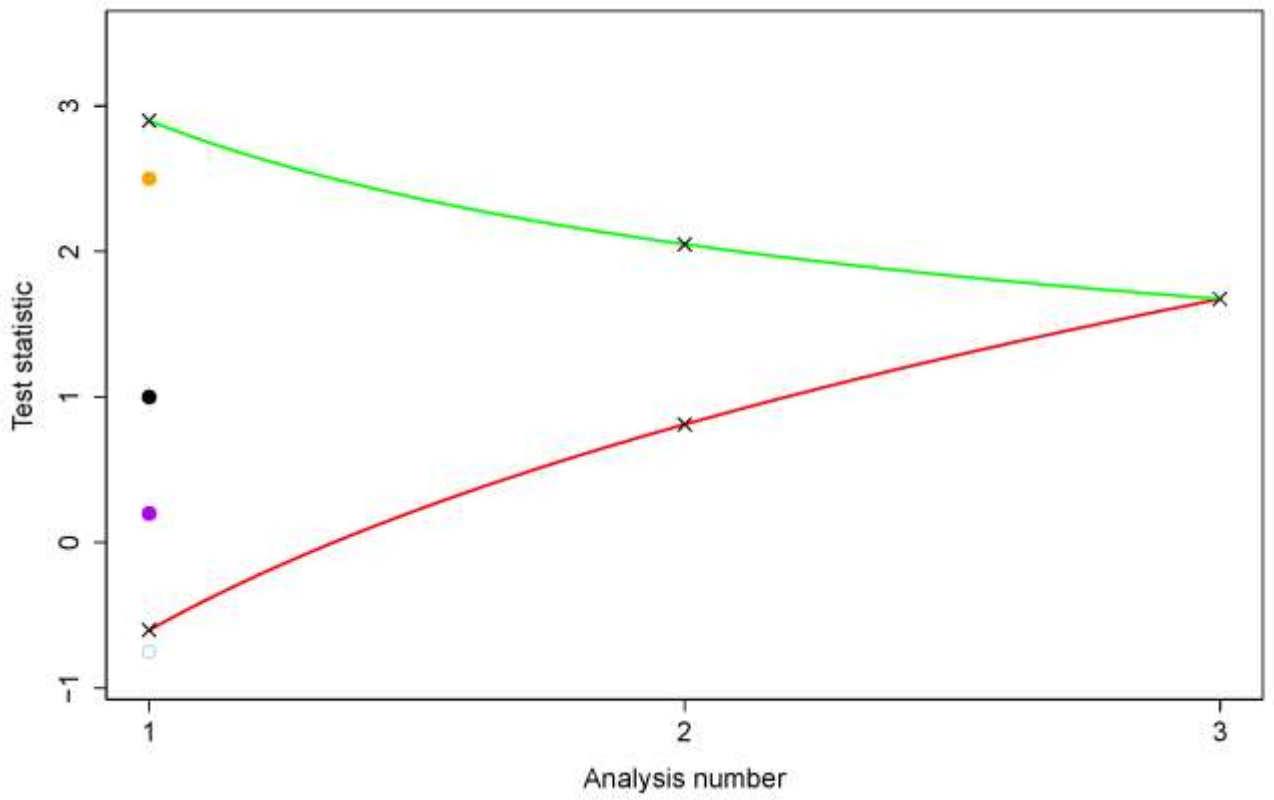
Group sequential MAMS trials

- This involves specifying futility and efficacy stopping boundaries.
- At each interim analysis, test statistics comparing each experimental treatment against control are calculated.
  - If below the futility boundary, the treatment is dropped for futility.
  - If above the efficacy boundary, treatment (or trial) stops with the conclusion that the treatment is effective.
- If early stopping for efficacy is not desirable, can just set efficacy boundaries to infinity.

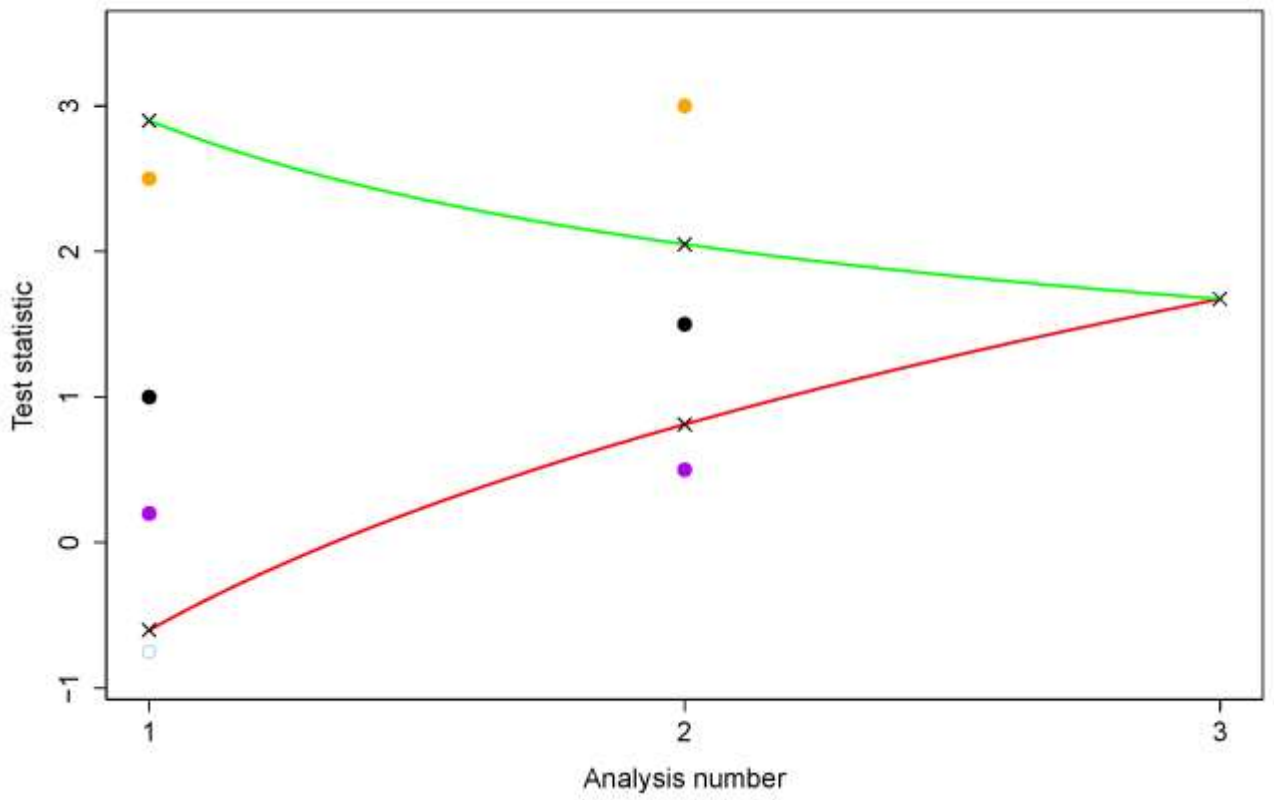
Group sequential MAMS trials



Group sequential MAMS trials



Group sequential MAMS trials



Group-sequential MAMS trials

- The sample size and stopping boundaries are chosen to control the type I error rate and power.
- These quantities are more complicated than in traditional two-arm trials.
- See Wason et al (Statistical Methods in Medical Research) for more details on powering a MAMS trial.

Efficiency of group-sequential MAMS trials

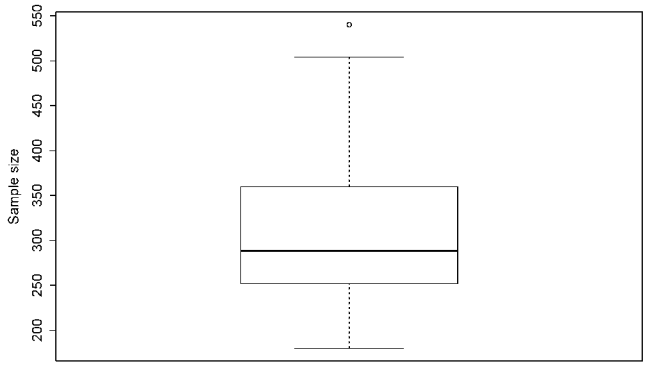
- The main advantage of having interim analyses is that the *expected sample size (ESS)* will be reduced.
- On average, group-sequential MAMS designs should be more efficient than multi-arm trials.
  - That is, the ESS will be lower than the total sample size needed by the multi-arm trial.
  - However, the maximum sample size (MSS) will be larger.
- Choice of stopping boundaries will make a big difference to the ESS and MSS.
- See Magirr et al, Wason and Jaki.

Endpoints

- Group-sequential MAMS designs have been developed for normal outcomes (previously mentioned papers), binary outcomes (Bratton et al.), time-to-event outcomes (Royston et al).
- Delay between recruitment and assessment of patients' outcomes is important: long delay means less efficiency gain from MAMS approach.
- Can also be applied with an intermediate endpoint at interim analyses that is observed more quickly (see MRC STAMPEDE trial).

Multi-stage drop-the-losers designs

- In the group-sequential MAMS design, the number of treatments at each stage is random and unknown in advance.
- Therefore the sample size used in the trial is also random.
- The sample size recruited in a trial is generally highly correlated with the cost and length of the trial.
- Boxplot showing sample size distribution in TAILoR trial (Magirr et al., four experimental treatments, three stages).



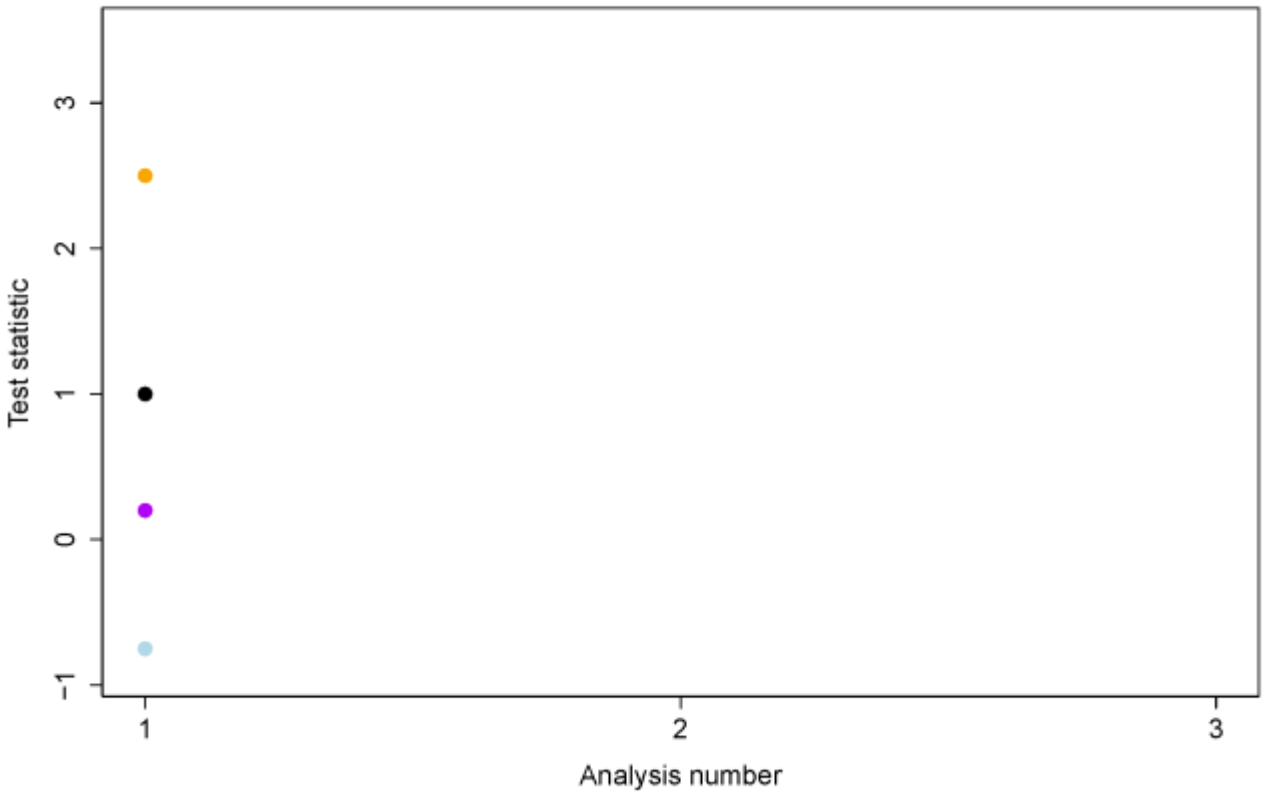
Multi-stage drop-the-losers designs

- Expected sample size = 293, but 5% of the time it will be >432.
- This creates problems – how much funding should such a trial apply for? Cannot be the expected sample size as often that won't be enough.
- Applying for maximum will mean the funder is committing a lot of money (although may get some back).
- Other issues with logistics, e.g. length of trial staff contracts.
- A MAMS design that has a fixed sample size would be a big advantage, even if it may lose some efficiency on average.

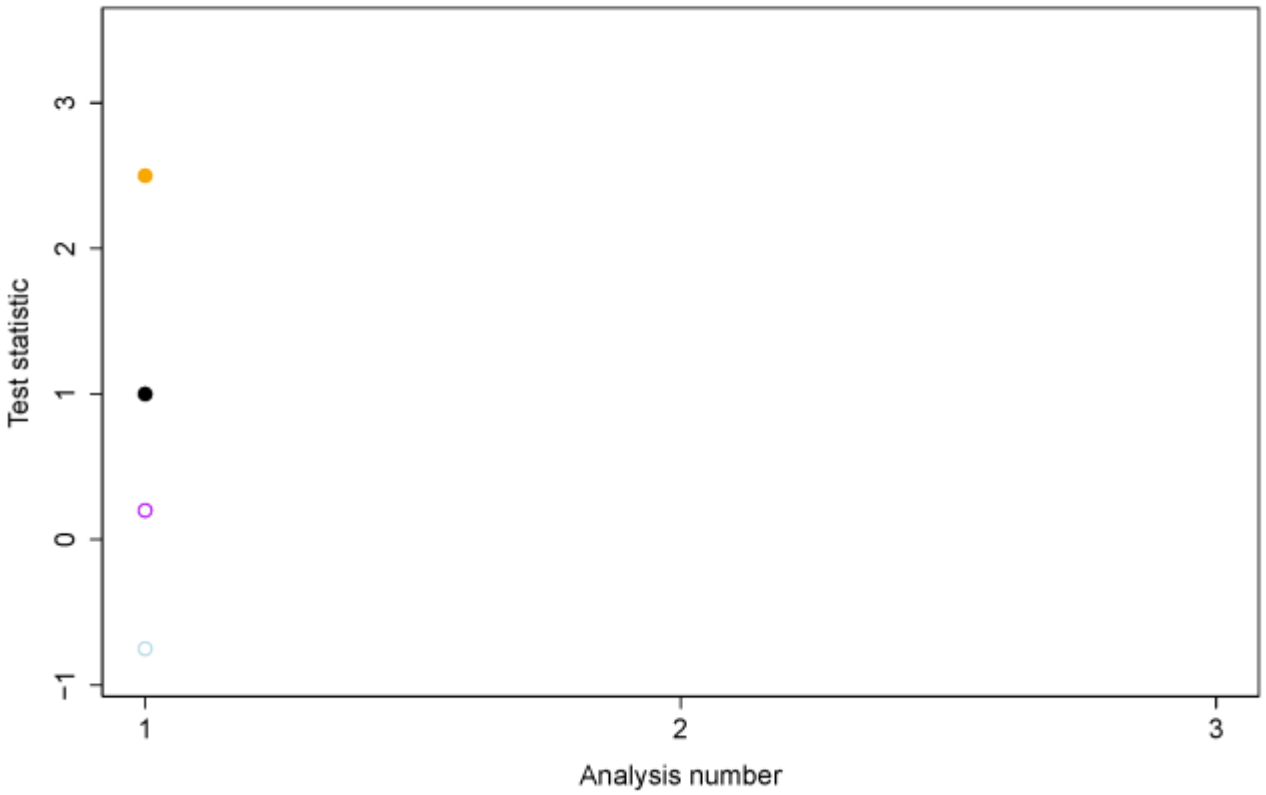
Multi-stage drop-the-losers designs

- Alternative design that has been well studied is the two-stage drop-the-losers design.
- Starts with a number of experimental treatments and a control treatment and a control treatment. At interim analysis, the best performing experimental treatment continues along with the control treatment.
- Can extend this to more than two stages – extra stage is worthwhile if there are four experimental treatments or more (Wason et al).

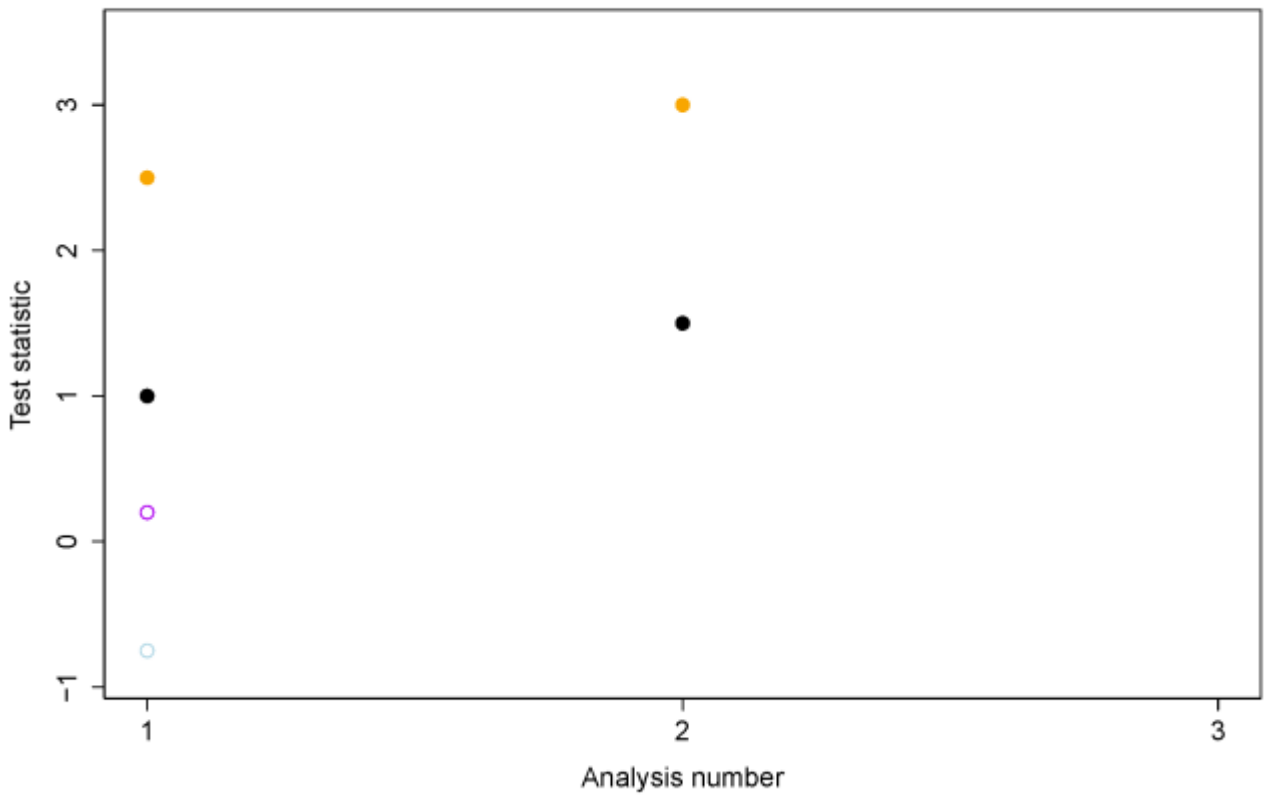
Multi-stage drop-the-losers designs



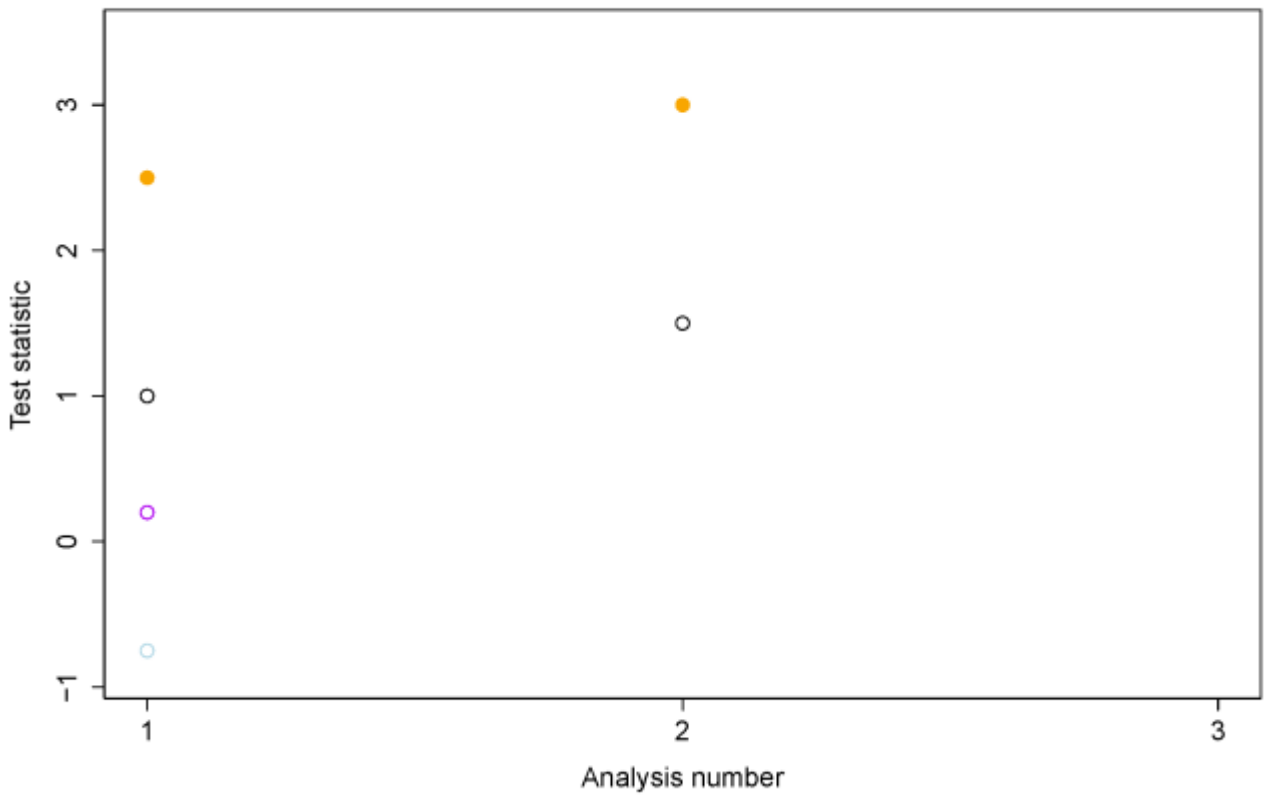
Multi-stage drop-the-losers designs



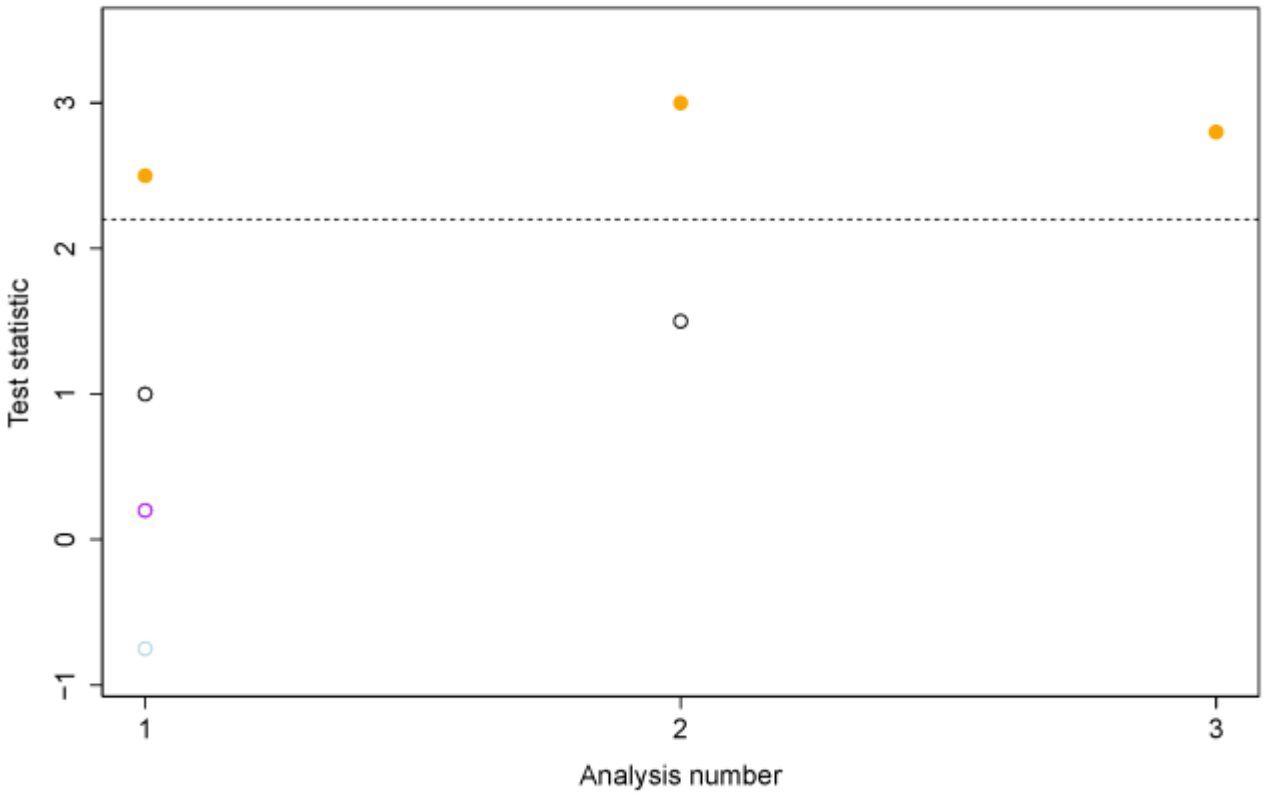
Multi-stage drop-the-losers designs



Multi-stage drop-the-losers designs



Multi-stage drop-the-losers designs



Multi-stage drop-the-losers designs

- The main disadvantage – cannot take more treatments forward than planned, even if two treatments have very similar efficacy.
- Can take fewer treatments forward than planned, or allow early stopping for futility – this will reduce the type I error rate and power.

Case study – MIDFUT trial

- Currently under consideration for funding.
- Phase II/III trial to compare four interventions vs treatment as usual for hard to heal diabetic foot ulcers.
- Phase II part of trial – aims to choose two experimental treatments to progress to phase III.
- Phase II endpoint: >= 50% reduction in wound area at 4 weeks post randomisation.
- Phase II sample size: 324 patients with 2:1:1:1:1 randomisation in favour of control (for efficiency).

Case study – MIDFUT trial

- At interim analysis, data on all patients assessed so far will be used to determine efficacy, safety, cost-effectiveness of the experimental treatments.
- Pre-specified progression criteria used to decide which experimental treatments progress.
  - Up to two can progress.
  - Have to show initial signs of improvement over control on the 4 week outcome (not necessarily significant improvement though).
  - Other criteria can also be used in decision.

Case study – MIDFUT trial

- For purposes of calculating the type I error rate, it is assumed that the two most efficacious treatments always progress – this will result in the highest type I error rate.
- Thus using other information in the decision will only lower the type I error rate.
- Possible for only one experimental treatment to progress (or none).

Case study – MIDFUT trial

- In phase III part of the trial, 336 patients will be randomised equally between control and remaining experimental treatments.
- Phase III endpoint: time to healing of reference DFU.
- Final test is done using two-sided p-value threshold of 0.02 – controls the total chance of making a type I error at 5%.
- Overall power for an effective treatment to progress at phase II and be recommended at phase III is 83.3%.

Case study – MIDFUT trial

- This is a drop-the-losers design, which was appealing for being able to specify a fixed sample size in advance (660 patients).
- Design has big advantage of allowing phase II patients to be included in the phase III analysis.
- Challenges: interim analysis must be done quickly so that number of patients recruited to dropped treatments is minimised.
- Four week endpoint at phase II is ideal for minimising this ‘overrun’.
- Many other issues of implementation of MAMS, but these have been successfully solved in other areas (Sydes et al).

Conclusion

- Multi-arm trials are very useful when multiple treatments are available for testing.
- Adding interim analyses (multi-arm multi-stage) adds further efficiency and allows ineffective treatments to be dropped early.
- Efficiency of MAMS will depend on delay between recruitment and the endpoint assessment of a patient.
- In wound treatment trials, if endpoints are assessed quickly, approach may be very useful, although introduces new challenges.

References

- Magirr D, Jaki T, and Whitehead J. A generalized Dunnett test for multiarm multistage clinical studies with treatment selection. Biometrika 99:494-501, 2012.
- Parmar M et al. More multiarm randomised trials of superiority are needed. The Lancet 384, 283-284, 2014.
- Sydes M et al. Issues in applying multi-arm multi-stage methodology to a clinical trial in prostate cancer: the MRC STAMPEDE trial. Trials, 10:39, 2009.
- Wason J and Jaki T. Optimal design of multi-arm multi-stage trials. Statistics in Medicine, 31:4269-4279 2012.
- Wason J, Magirr D, Law M, Jaki T. Some recommendations for multi-arm multi-stage trials. Statistical methods in medical research, 2012.
- Wason J, Stecher L, Mander A. Correcting for multiple testing in multi-arm trials: is it necessary and is it done? Trials, 2014.