An Alternative Design for Trials of Patients with Rare Pathogens:

Conducting trials with difficult to find cases

Aaron Dane

Disclosure Statement

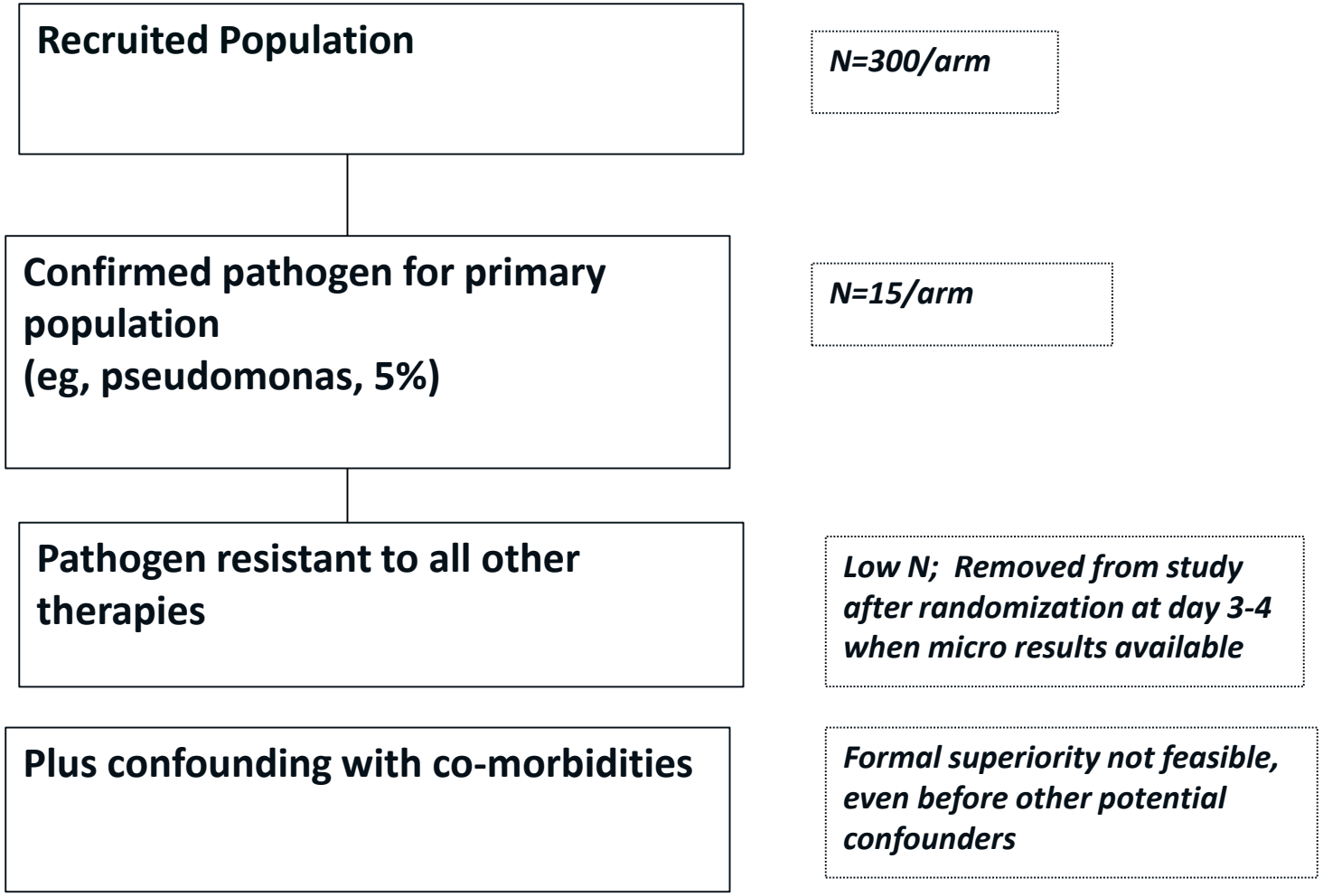
Acted as a statistical consultant for Achaogen, Allecra, Amicrobe, Amplyx, Cidara, ContraFect, Davolterra, Destiny, F2G, Geom, GSK, Gyroscope, Kymab, Mironid, Nabriva, Pfizer, Phico, Pled, Roche, Scynexis, Spero, TenNor, Transcrip, VenatoRx and Zavante

This presentation is the result of collaboration with Prof. Nigel Stallard (Warwick University, UK) and has been refined with input from Paul Newell and John Rex

Challenges of rare pathogen development
Why not just run superiority studies?

- Superiority is preferred when feasible as this resolves many of the points to consider with non-inferiority trials
- However, showing superiority on a clinical endpoint is not routinely feasible for anti-infectives
  - We cannot deliberately study ineffective comparators in seriously ill patients
  - Formal demonstration of superiority is challenging when patients are rare
  - There will often be at least one therapy with some efficacy (but may have toxicity)
- And it's also not desirable from a societal viewpoint
  - We want to always have a range of suitable antibiotics on hand to treat future resistance, not just those where we can demonstrate superiority right now

Why is superiority so difficult in an RCT?



We must consider this both within and beyond the AMR community

- Much of what we are developing is for tomorrow's patients
  - We want clinical superiority to be rare right now
- Despite this, we need a reliable and feasible way to assess new compounds
  - Rare pathogens are hard to find: enrolling even 100 patients can be very difficult
- We also need to consider the safety database
  - What is the minimal N on drug for safety in this setting?

What can you do with 100 patients?..

A Possible Approach to Trial Design for Rare Pathogens

Decision making in clinical trials

- What are we most interested in for any clinical trial?
  - To be confident we can show an effective treatment works
  - To be confident we will not approve ineffective treatments
- Can we look at this differently for rare pathogens?
  - These patients are hard to find during clinical trials
  - We still want drugs before resistant pathogens are more common
  - We have looked to draw on ideas used in the orphan drug area
- Even with smaller studies we need a framework for decision making
  - Sponsors need to understand what is required to assess study risk
  - Provides clarity regarding decision making criteria at the design stage

Aim

- To propose a framework for decision making and sample size for rare pathogen studies where feasibility is extremely challenging and when (sub-optimal) therapies are still available

- This is not about performing an interim analysis where we decide to continue to recruit more patients, but rather how to understand the risks with a smaller study

- This talk focuses on traditional frequentist statistics, but has also been considered within a Bayesian framework

Large v Small trials with rare pathogens

- Larger trials lead to higher power
  - But if the trial is too large (or takes too long) this deprives patients of an effective therapy and may mean it is not feasible to develop the drug
- Smaller trials may be more feasible
  - But if trial is too small we have larger chance of making the wrong decision
- Common theme: How to work with the only (small) dataset possible?
  - Can we show that there is a “sweet spot” for sample size?
  - There can be diminishing returns outside the “sweet spot”

Can we define a “sweet spot” to balance these questions

What are we aiming for?

- If TEST is worse than CONTROL…
  - Every patient randomized to TEST risks a worse outcome
  - If TEST is approved, this problem is perpetuated
  - *Mitigation: Within a small trial, avoid incorrect approval (Type 1 errors)*

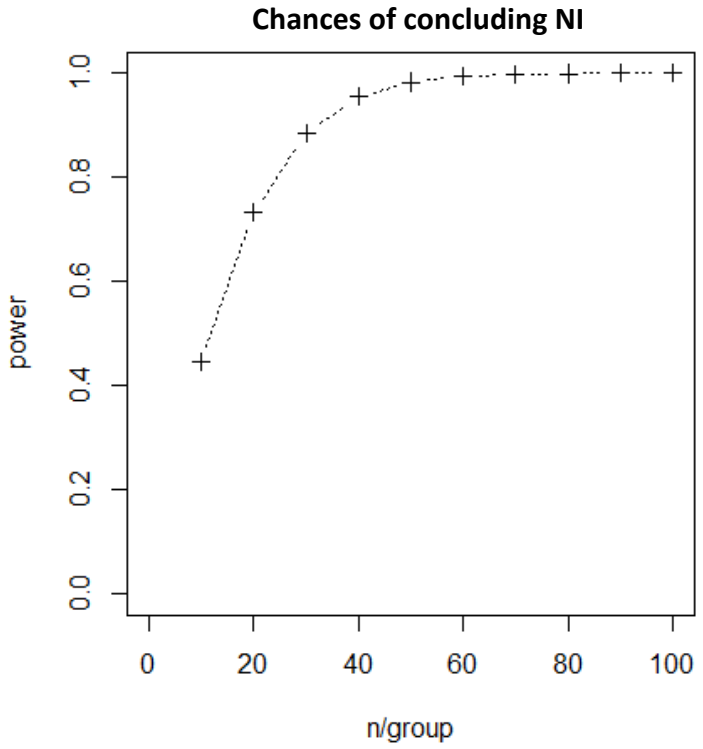
- If TEST is better than CONTROL…
  - Every patient randomized to CONTROL risks a worse outcome
  - If TEST is not approved, this problem is perpetuated
  - *Mitigation: Within a small trial, keep the power high*

- If TEST is similar to CONTROL…
  - We still want to make additional therapies available
  - *Mitigation: Within a small trial, keep the power high*

When we run the trial we do not know which of these situations is true so we must understand the type I error and power for a range of study sizes

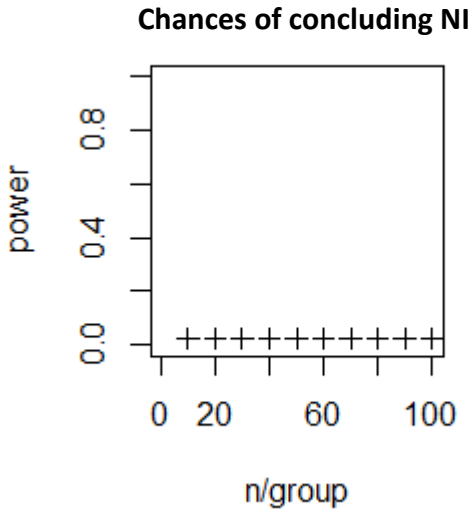
When Test is better than Control

- Imagine this NI trial of TEST vs. CONTROL
  - TEST's response rate is 60% (but you don't yet know this)
  - CONTROL's response rate is 40% (you do know this)
  - Select NI margin of 20% and use a 95% CI
  - The correct outcome is to conclude NI
- You run the trial
  - Probability to conclude NI (power) goes up with sample size
- Figure shows, with more than 40-50/group, things don't improve a lot
  - Power is reaching 90%+
  - You might want more for the safety database, but you can already show TEST is at least NI to control for efficacy



When Test is worse than Control

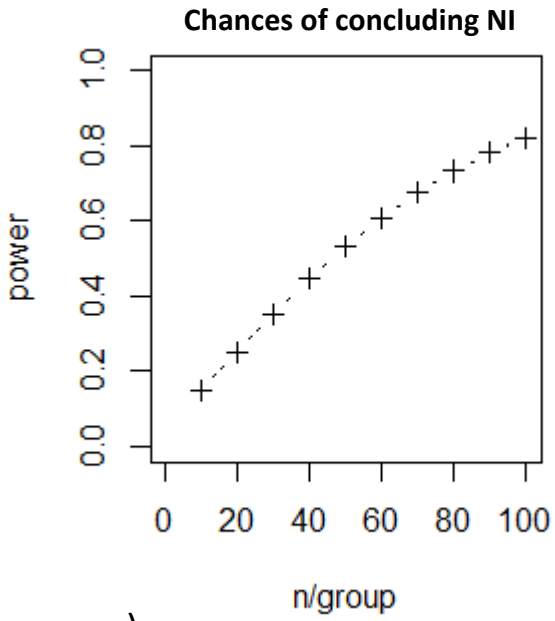
- Still an NI trial of Test v Control
  - TEST's response rate is 20% (but you don't yet know this)
  - CONTROL's response rate is 40%
  - This again uses a margin of 20% and a 95% CI
  - In this case we do not want to conclude NI
- You run the trial
  - The plot shows we are unlikely to conclude NI
- As shown at right, power is always very low
  - This is Type I error in this situation
  - This shows it's unlikely you'll conclude NI
  - Again, ~40-50/group is probably enough



We will rarely make the error of concluding NI, however big the study

When Test and Control are the same

- One last variation
  - TEST's response rate is 40% (but you don't yet know this)
  - CONTROL's response rate is 40%
  - Again use a margin of 20% and a 95% CI
  - Would we want to conclude NI in this case? (generally yes)



- You run the trial
  - Chances of concluding NI goes up with sample size, but requires larger N

- Power rises steadily with N
  - ~50/group has power around 50% (or 50/50 chance of success)
  - Around 80/group gives power of ~80%
  - It takes a long time to get to 90%+ power

A larger study would be needed for traditional levels of power and with 95% CIs. Studies are unlikely to be undertaken if they only have a 50/50 chance of success

What does it mean for patients after the trial?

- Suppose we have an overall population of 1,000 patients with a rare pathogen
  - We include 100 patients within a clinical trial (50 per arm)
  - Then we treat the remaining 900 patients outside of the trial (assume we use the new SOC after trial completes)

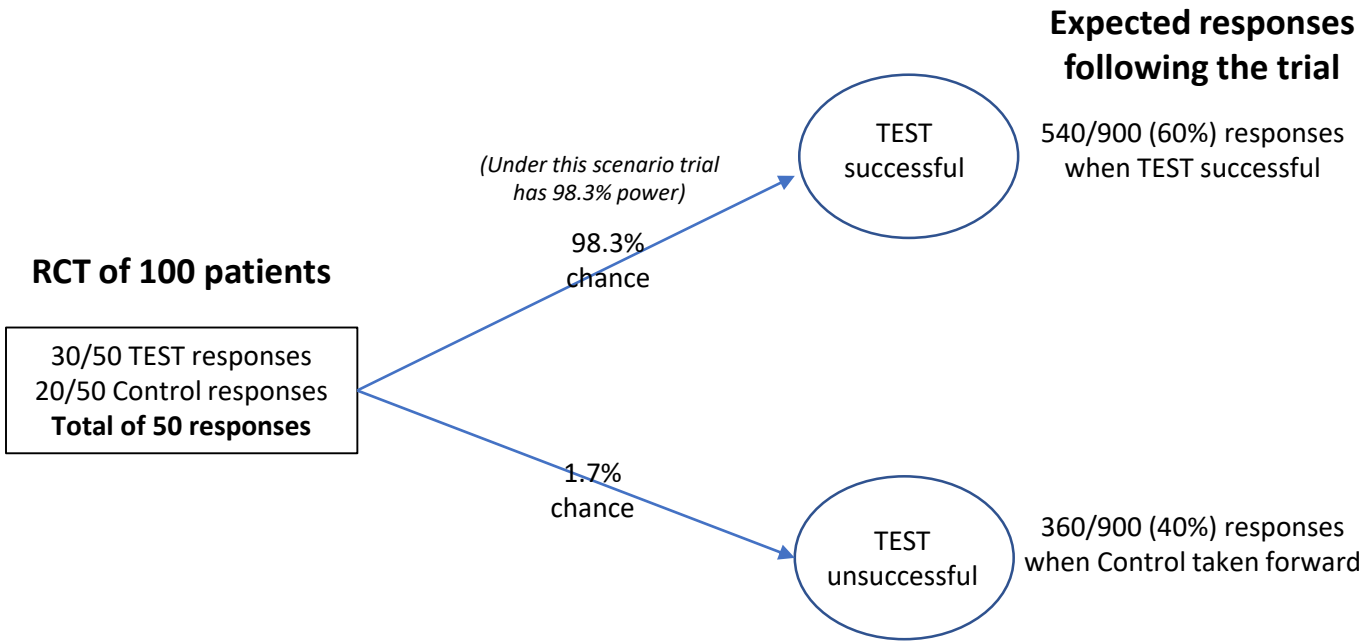
Patients included in trial	Remaining patients
50 patients on TEST	1,000 – 100 = 900 patients *(Overall population minus patients included in trial)*
50 patients on Control	

- From this we can calculate the total success rate across entire 1,000 subjects?
  - How many of 50 patients on TEST from trial would be a response
  - How many of 50 patients on Control from trial would be a response
  - How many of remaining 900 patients would be a response (depending upon whether TEST or Control is new SOC)

Bigger not always better with rare diseases (1)

Assuming 1,000 patients with rare pathogen and 100 patients in RCT

Consider scenario with TEST = 60%, Control=40%



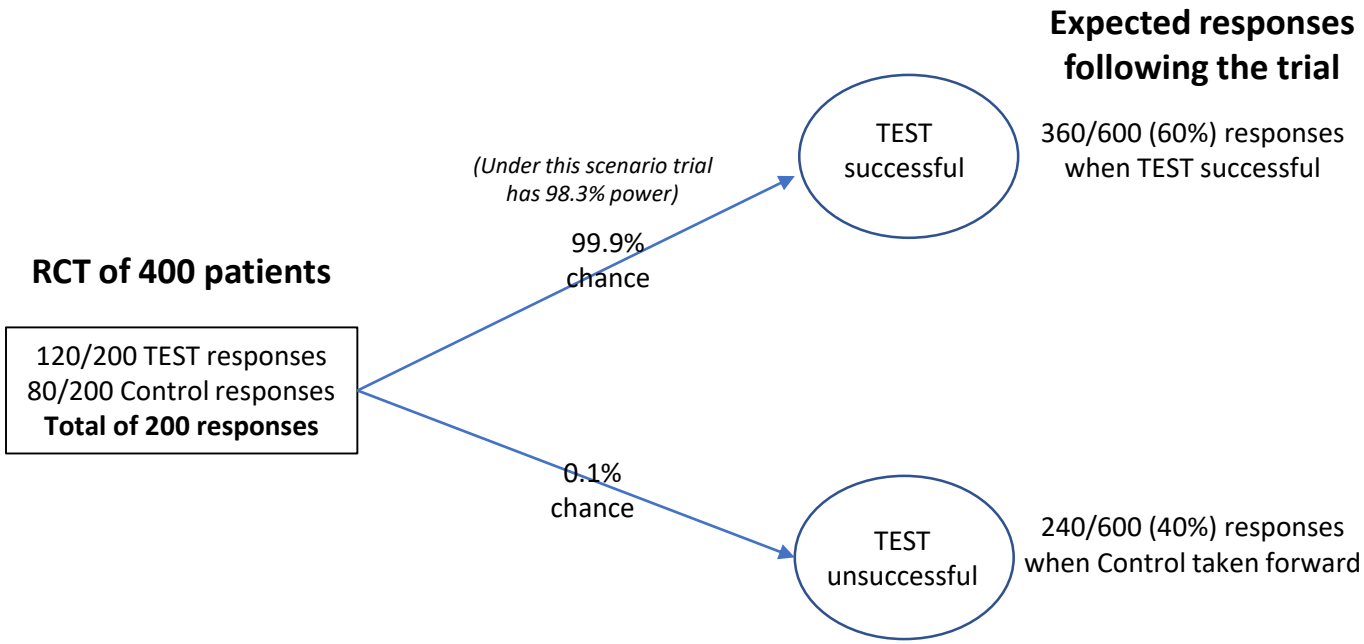
Expected number and proportion of successes in the overall population

50/arm: 30 + 20 + (0.983 x 540 + (0.017 x 360) = **587 of 1,000 patients**

Bigger not always better with rare diseases (2)

Now consider 1,000 patients with rare pathogen and 400 patient RCT

More patients are in RCT → fewer overall successes expected



Expected number and proportion of successes in the overall population

200/arm: 120 + 80 + (0.999 x 360 + (0.001 x 240) = **560 of 1,000 patients**

We want get as many patients as possible receiving the drug with 60% response

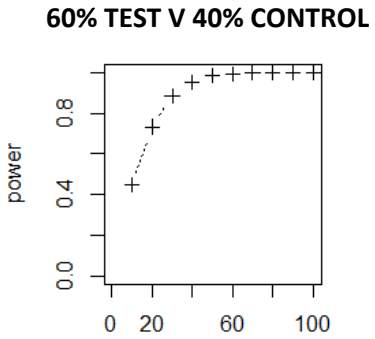
Recap: Finding the “sweet spot”

- We need to find a sample size where:
  - We have a good chance of success when effective
  - We have a low chance of approval when ineffective
  - We have a reasonable chance of success when similar
  - The expected number of patients benefitting is maximised

The following plots summarize this information…

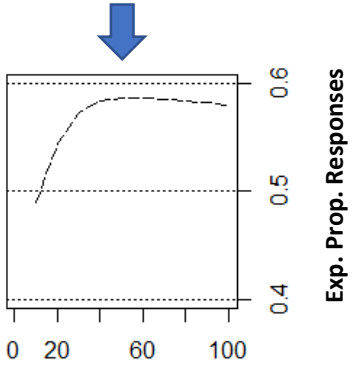
When TEST is 20% better 40/arm is enough

How often is TEST selected?



Trial with 40 patients gives reasonable power

Proportion of responses are expected in the overall population



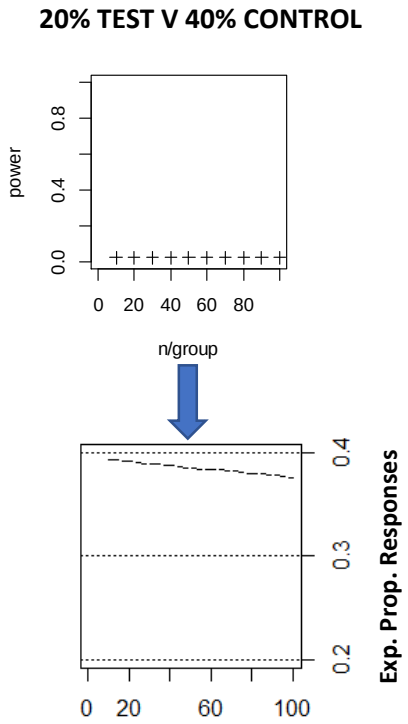
Trial with 40 patients also optimizes the expected successes in the overall population (based on response during and after the trial)

E (prob.success) = proportion of responses expected in the overall population (during and after the trial)

Incorrect approval is low when TEST is 20% worse

How often is TEST selected?

Proportion of responses are expected in the overall population



In this case TEST is worse than Control, so you would not want TEST to be selected

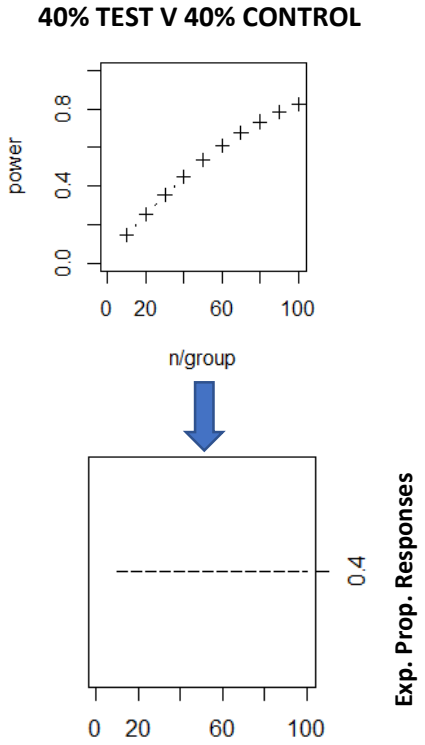
The probability of recommending TEST is low irrespective of N

The expected proportion of responses again falls with larger N given more patients are given (less effective) TEST than in a smaller study

Larger sample sizes needed when TEST is similar

How often is TEST selected?

Proportion of responses are expected in the overall population



Probability of recommending TEST steadily increases with larger N; a larger study would be required for approval in this case

As TEST and Control have the same response rate the expected proportion of responses will always stay the same (in this case 40%)

This occurs with an existing therapy with some efficacy, but we need more options due to emerging resistance or toxicity with currently available drugs

If we can only recruit 40-50 patients per arm for a limited population, we need to find a way of running a feasible trial

What is the “sweet spot”

- In this case ~40-80/arm has reasonable power
  - Assuming control has 40% efficacy…
  - 20% inferior products are detected with 40/arm
  - 20% superior products are detected with 40/arm
  - Products with similar efficacy need N nearer 80/arm which may be feasible
  - The number of patients benefitting overall often drops with larger study size
- How to provide clear criteria when it is only feasible to recruit 50 patients/arm?
- One approach is to consider using different confidence intervals (such as 80% CI) for areas of large unmet need

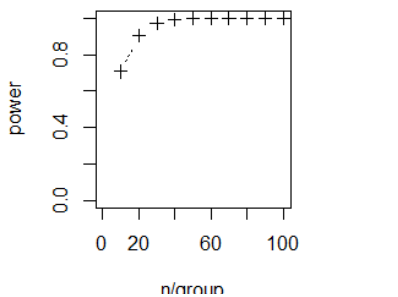
Using 80% CI shifts the risk profile

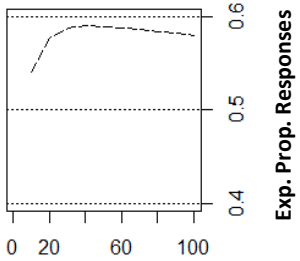
80% CIs & 20% NI margin

How often is TEST selected?

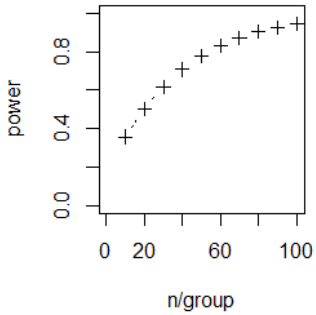
Proportion of responses are expected in the overall population

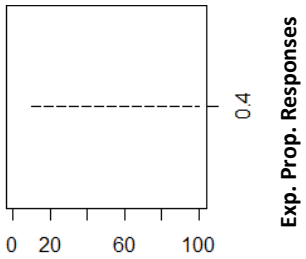
Test is better (60% v 40%)



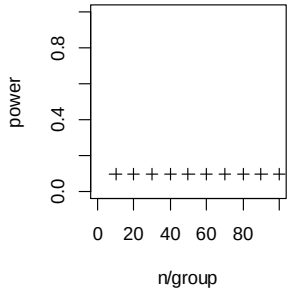


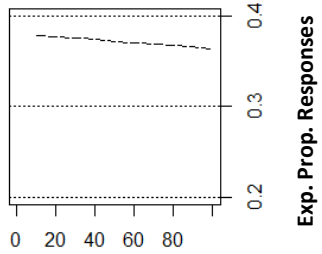
Test is similar (40% v 40%)





Test is worse (20% v 40%)





Observations:

- When TEST is 20% worse (20% v 40% response) risk of approval using 80% CI is higher (10%)
- When TEST is similar or better using 80% CI is better than using a 95% CI
- But of course we don't know which is the truth

Note: this is an example framework; the use of (say) 80% CI will depend on degree of unmet need, the potential benefit/risk improvement and the likelihood of Test agent is worse than Control

In summary

- This is a framework to display trade-offs when only a small trial is possible
  - What are reasonable false positive and false negative error rates
  - As a community we need to decide how to trade these risks when we cannot run large trials

- Data on 100-200 patients with rare pathogens can be very informative, but we need clear criteria for success that can be agreed ahead of trials
  - How to maximise our chances of approving a more effective new drug with (say) 100 patients?
  - How to limit the risk of approving a less effective new drug with (say) 100 patients?

- The example shows the risks when new drug is 20% better or 20% worse, but can be applied to other scenarios of better/worse response

Using plots of power, Type I error and overall number of patients benefitting from therapy can be used to agree success criteria in trials of rare pathogens

Back-up Slides

Expected Number of Responses

Looking at the two sample size scenarios side by side

Responses in trial			Responses following trial		
RCT sample size	**RCT success (TEST)**	**RCT sucess (Control)**	**Probability TEST is taken forward**	**If TEST taken forward**	**If Control taken forward**
50/arm	30/50	20/50	0.983	540/900	360/900
200/arm	120/200	80/200	0.999	360/600	240/600

<u>Expected number and proportion of successes in the overall population</u>

50/arm: 30 + 20 + (0.983 x 540 + (0.017 x 360) = 587 → (i.e. ~59% of all patients)

200/arm: 120 + 80 + (0.999 x 360 + (0.001 x 240) = 560 → (i.e. 56% of all patients)

TEST responses in RCT

Control responses in RCT

Expected responses after RCT is
Probability TEST taken forward x number responses if TEST taken forward plus
Probability Control taken forward x number responses if Control taken forward

The expected number of responses is important when understanding whether a patient may benefit from the approval of a new drug so is a key element of the following slides