THE EFFECTIVE STATISTICIAN ACADEMY

Mastering study design and strategy with Simulation

Presenters:

Kim Hacquoil
Chief Data Scientific Officer

Jamie Inshaw
Strategic Consulting Team Lead

Agenda Outline

Scene Setting: Challenges and Opportunities for simulation

Simulation: What, Why and When?

Simulation Framework and Plans

Where do you get the data from?

Regulators Viewpoint

Communication of Simulations

Efficient Simulations

Example

What is simulation within drug development setting?

Definition: Simulation

Simulation is the act or process of **imitating** the functioning or behaviour of a **real-world system** or process by means of another system or process, usually a computer program.

A simulation requires a **model** that represents the **key characteristics** or functions of the selected system or process.

A simulation can be used for **study, training, testing, or demonstrating** purposes.

Simulation – Wikipedia
Simulation - definition of simulation by The Free Dictionary
Simulation Definition & Meaning - Merriam-Webster
What does simulation mean? (definitions.net)
Simulation - Definition, Meaning & Synonyms | Vocabulary.com

Simple Example

A Phase II study comparing active treatment to placebo on a biomarker - How can we investigate the power curve using simulation?

Clinical Trial

Design: Randomised 1:1 two-arm trial

Primary endpoint: Biomarker at 12 weeks post randomisation (active vs placebo)
Assume Normally distributed with **true** treatment difference 4.5
Assume $\sigma^2 = 10^2$ **known**

Success: defined as treatment effect p-value

Sample Size

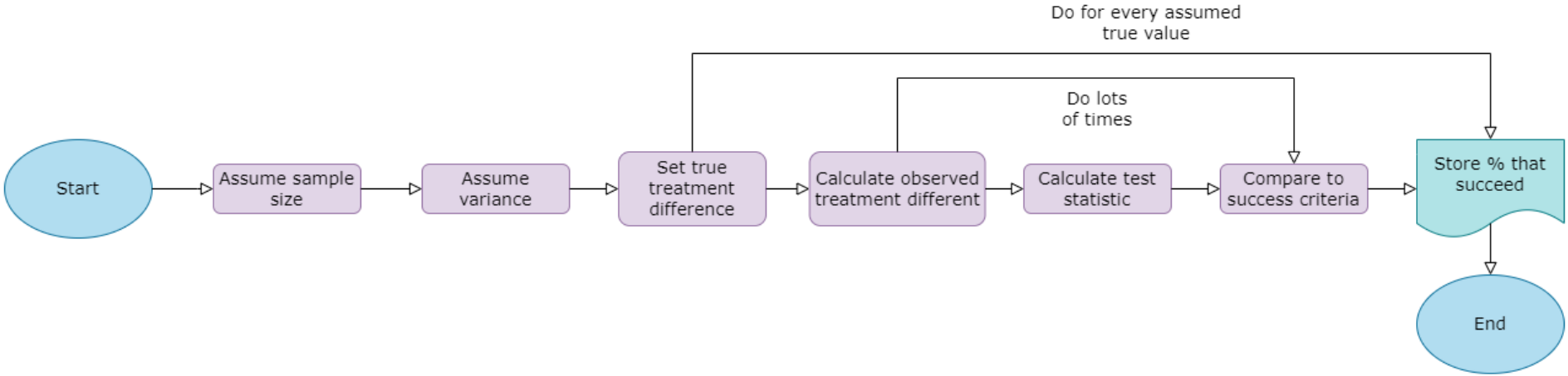
100 per arm

power of 90% and a level of significance of 5% (two sided), T-test

exploristics
INNOVATIVE ANALYTICS SOLUTIONS

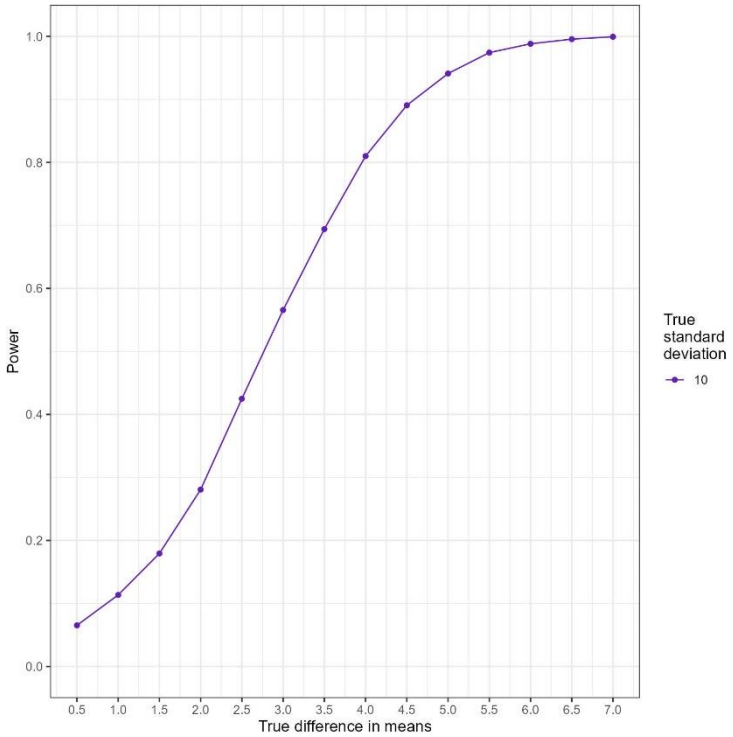
The Effective Statistician Academy

How might you approach using simulation instead?



How do you simulate this in R?

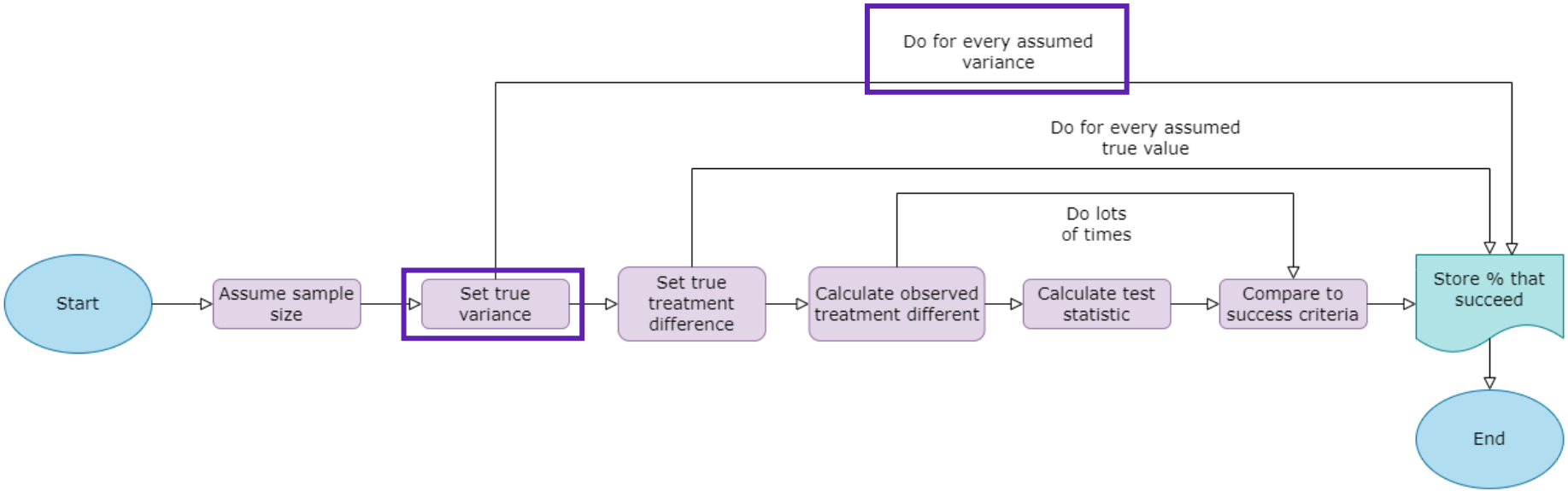
```r
gensums<-function(n,meandiff, sd, iteration){
  se<-sd*sqrt(2/(n/2))
  result<-rnorm(mean=meandiff, sd=se, n=1)
  z<-result/se
  p<-2*pnorm(-abs(z))
  out<-data.frame(mean=meandiff, se=se, p=p, iteration=iteration)
  return(out)
}

combinethem<-function(meandiff, sd,n){
  res<-do.call("rbind",lapply(c(1:5000),gensums, meandiff=meandiff,sd=sd, n=n))
  out<-data.frame(n=n,meandiff=meandiff, sd=sd, power=nrow(res[res$p<0.05,])/nrow(res))
  return(out)
}


power<-do.call("rbind",lapply(c(seq(0.5,7,0.5)),combinethem, sd=10, n=200))
```



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

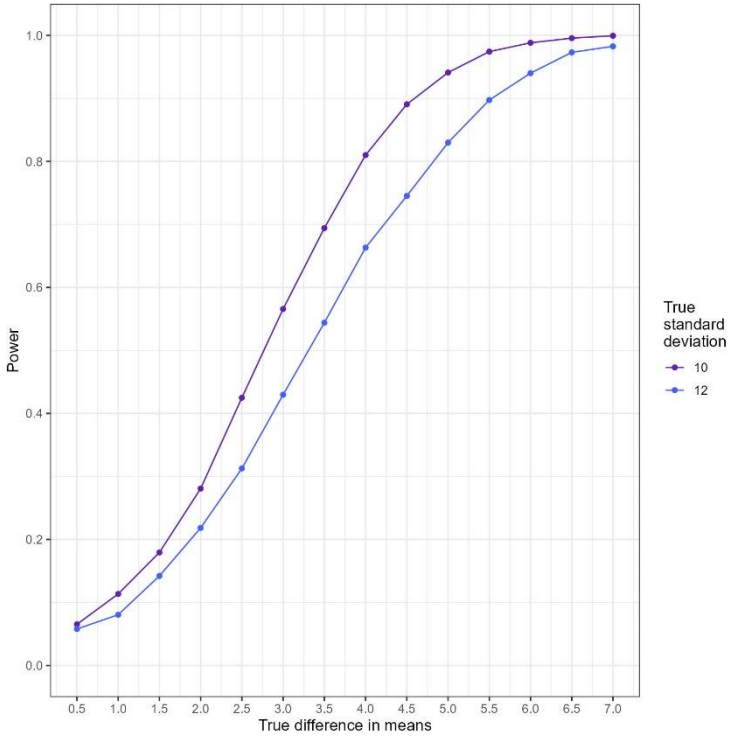
The Effective Statistician Academy

How might you extend for other unknowns using simulation?



How do you extend this in R to incorporate other unknowns?

```
gensums<-function(n,meandiff, sd, iteration){
  se<-sd*sqrt(2/(n/2))
  result<-rnorm(mean=meandiff, sd=se, n=1)
  z<-result/se
  p<-2*pnorm(-abs(z))
  out<-data.frame(mean=meandiff, se=se, p=p, iteration=iteration)
  return(out)
}

combinethem<-function(meandiff, sd,n){
  res<-do.call("rbind",lapply(c(1:5000),gensums, meandiff=meandiff,sd=sd, n=n))
  out<-data.frame(n=n,meandiff=meandiff, sd=sd, power=nrow(res[res$p<0.05,])/nrow(res)
  return(out)
}


power<-do.call("rbind",lapply(c(seq(0.5,7,0.5)),combinethem, sd=10, n=200))

powersdhigh<-do.call("rbind",lapply(c(seq(0.5,7,0.5)),combinethem, sd=12, n=200))
powerboth<-rbind(power,powersdhigh)
```



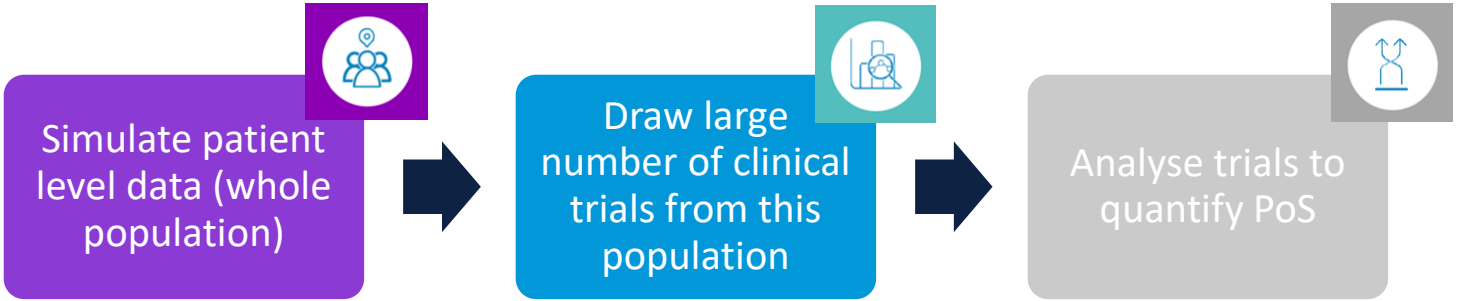
exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

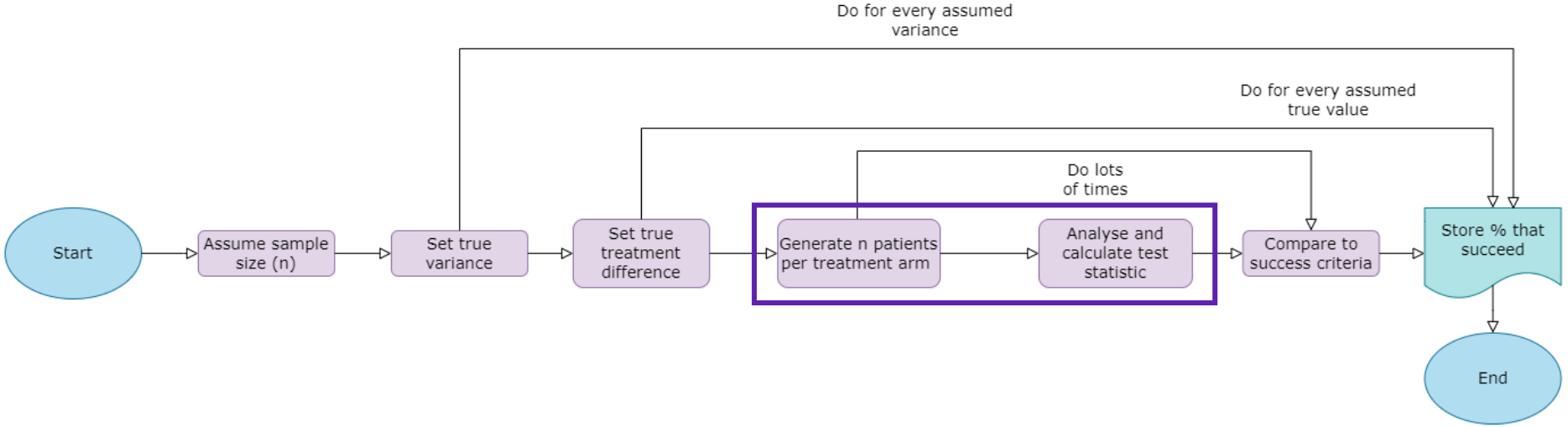
In silico clinical trial simulation

A Phase II study comparing active treatment to placebo on biomarker - How can we investigate the power curve using patient level simulation?

Patient-level simulation is conceptually straightforward

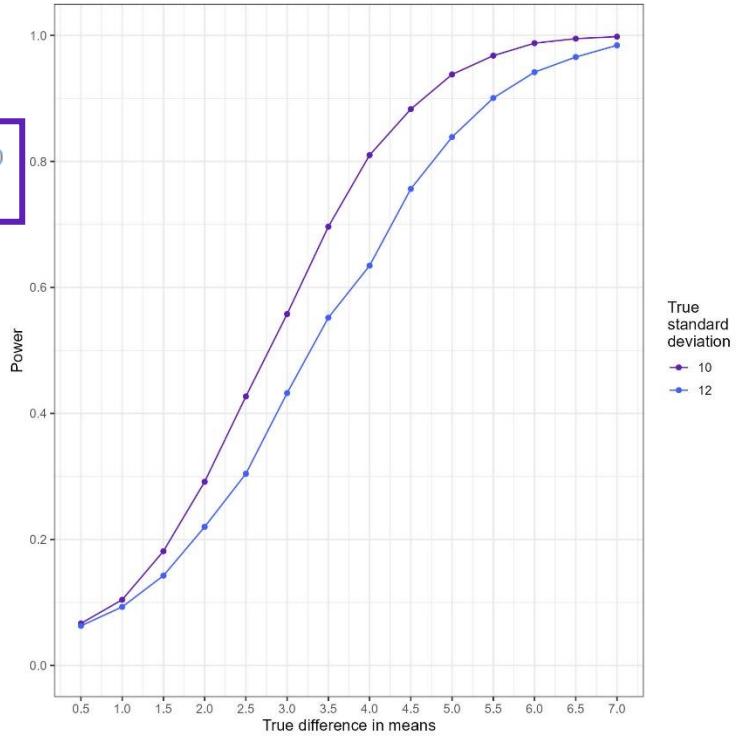


How might you amend the example for patient level simulation?



How do you simulate virtual Patients?

```
gensumsipd<-function(n,meandiff, sd, iteration){
  ipd<-data.frame(id=c(1:n), arm=c(rep(0,(n/2)),rep(1,(n/2))),
                  outcome=c(rnorm(n=n/2,mean=0,sd=sd),rnorm(n=n/2,mean=meandiff,sd=sd)))
  test<-t.test(ipd$outcome ~ ipd$arm)
  p<-test$p.value
  out<-data.frame(mean=meandiff, se=se, p=p, iteration=iteration)
  return(out)
}

combinethemipd<-function(meandiff, sd,n){
  res<-do.call("rbind",lapply(c(1:5000),gensumsipd, meandiff=meandiff,sd=sd, n=n))
  out<-data.frame(n=n,meandiff=meandiff, sd=sd, power=nrow(res[res$p<0.05,])/nrow(res))
  return(out)
}


power<-do.call("rbind",lapply(c(seq(0.5,7,0.5)),combinethemipd, sd=10, n=200))

powersdhigh<-do.call("rbind",lapply(c(seq(0.5,7,0.5)),combinethem, sd=12, n=200))
powerboth<-rbind(power,powersdhigh)
```



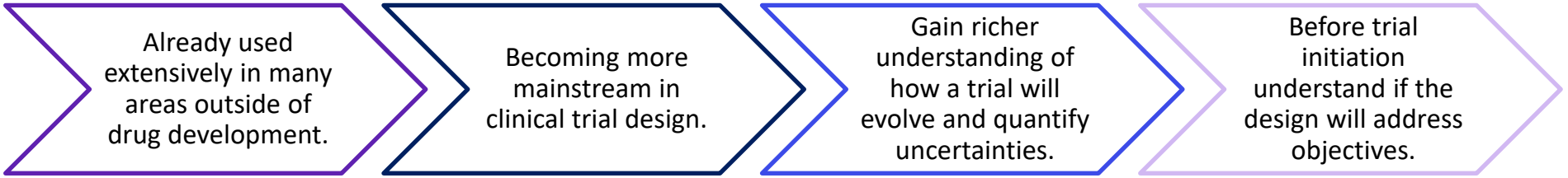
exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Simulation: What, Why and When?

Its easier to explore through simulation!

Why should we use simulation in clinical trial design?



Why do you use simulation in clinical trial design?

"to quantify some operating characteristic of a design that contains some features that **cannot be quantified analytically**"

"most reliable way to **explore different possible truths** and understand where the key risks lie in the study with regards to probability of success"

"allows us to **handle all the factors** we wish to consider in the same framework"

"**easier and therefore quicker** than deriving analytically challenging equations to achieve results that should be analogous"

"easiest way to **explore 'what if' scenarios**"

"**risks to the trial** success are often identified"

exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Assurance Example

A Phase III study comparing active treatment to placebo on clinical endpoint

Assurance in clinical trial design - O'Hagan - 2005 - Pharmaceutical Statistics - Wiley Online Library

Clinical Trial

Design: Randomised 1:1 two-arm trial, n patients per group

Primary endpoint: Clinical outcome at 6 months post randomisation (active vs placebo)

Assume Normally distributed $\bar{y}_A - \bar{y}_P \sim N(\theta, \tau^2)$, where $\tau^2 = \frac{2\sigma^2}{n}$

Assume we have a prior $p(\theta) \sim N(\mu, s^2)$

Assume σ^2 is known

Success: defined as treatment effect p-value <0.025 required in the primary endpoint (one-sided)

How do we calculate the Assurance for this?

Analytical Approach

Predictive distribution is:

$$p(\bar{y}_A - \bar{y}_P) \sim N(\mu, \tau^2 + s^2)$$

Therefore:

$$\text{Prob}(\bar{y}_A - \bar{y}_P > \tau \boldsymbol{Z}_{\boldsymbol{\alpha}}) = \Phi\left(\frac{-\tau Z_\alpha + \mu}{\sqrt{\tau^2 + S^2}}\right)$$

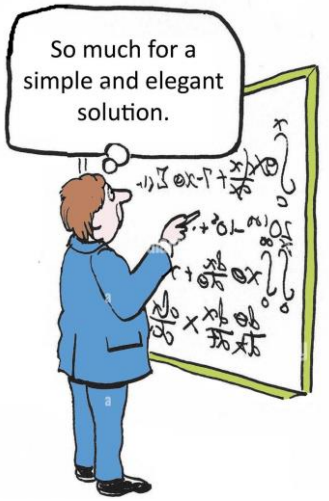
Simulation Approach

1. Generate $\tilde{\theta}$ from the prior distribution $p(\theta) \sim N(\mu, s^2)$
2. Generate a treatment effect value for $(\bar{y}_A - \bar{y}_P) \sim N(\tilde{\theta}, \tau^2)$
3. Calculate test statistic $\frac{\bar{y}_A - \bar{y}_P}{\tau}$
4. Do steps 1-3 lots of times and calculate the proportion of times the test statistic > Z_α

What happens if we don't want to assume σ^2 is known?

Analytical Approach

Gets more complicated!!!



Simulation Approach

1. Generate $\tilde{\theta}$ from the prior distribution $p(\theta) \sim N(\mu, s^2)$
2. Generate $\tilde{\sigma}^2$ from the prior distribution $p(\sigma^2)$
3. Generate an outcome variance value $\hat{\sigma}^2$ from $\frac{\tilde{\sigma}^2}{2(n-1)} \chi^2_{2(n-1)}$
4. Generate a treatment effect value for $(\bar{y}_A - \bar{y}_P) \sim N(\tilde{\theta}, \hat{\tau}^2)$
5. Calculate test statistic $\frac{\bar{y}_A - \bar{y}_P}{\hat{\tau}}$
6. Do steps 1-5 lots of times and calculate the proportion of times the test statistic > Z_α

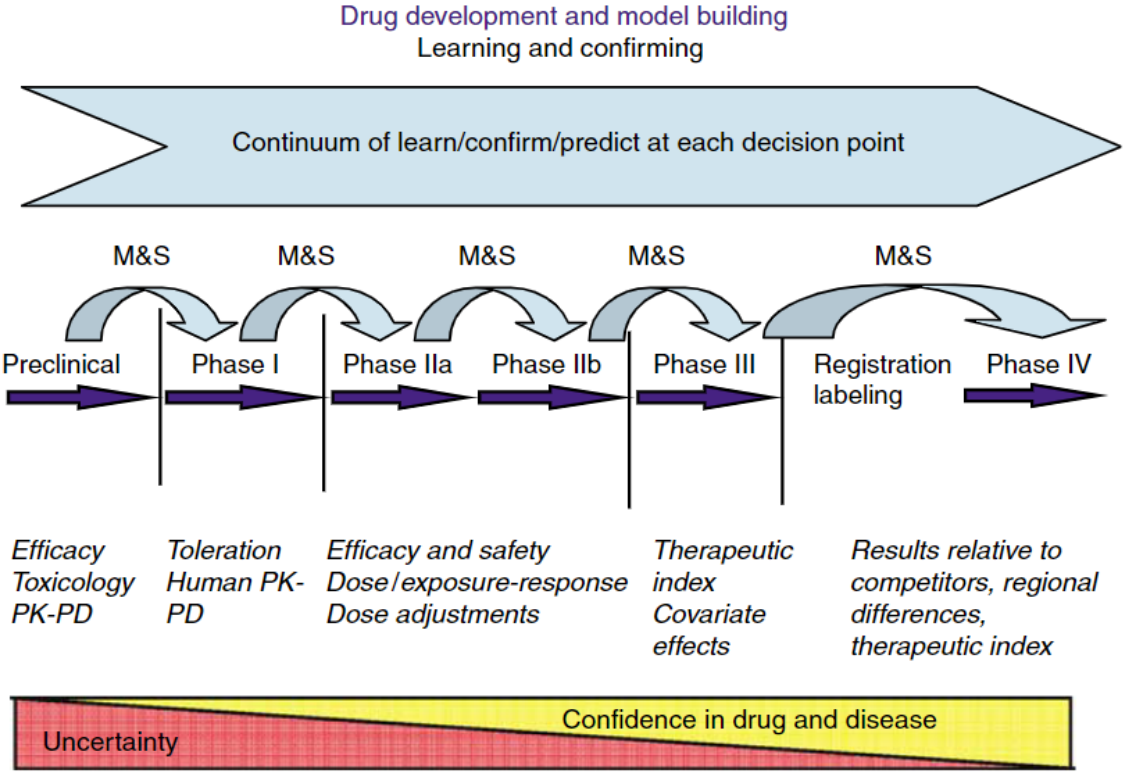
Summary of Assurance Example

Use analytical approach	Simulation or analytical approaches are appropriate	Simulation is the only solution

Simulation: What, Why and When?

The Why and the When are intrinsically linked

When do you use simulation in drug development?



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Different types of simulation in drug development

Translation or Extrapolation

- Between adults and paediatrics
- Dose response
- Between pre-clinical and clinical

Prediction

- Study outcomes and operating characteristics
- Events and recruitment
- Drug supplies and resources e.g. reduce chance of shortage

Testing and optimising

- Study outcomes
- Code
- Virtual twins
- Exploration of unknowns

Why might you simulate virtual patients?.... When theoretical approaches or simplifications don't work for you

Estimands and intercurrent events

Non-standard distributions

Repeated measures between observations for a patient

Dose response

Certain adaptive designs

Recruitment

Enrichment designs

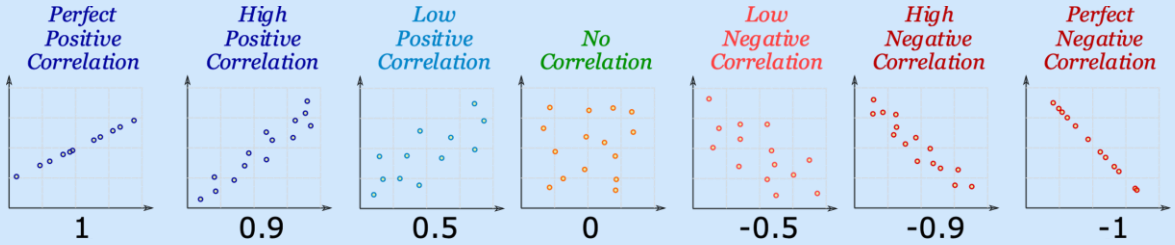
Correlations between endpoints

Time to event

exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Example: Benefits of simulation - Incorporating multiple endpoints and correlations into study design decision-making



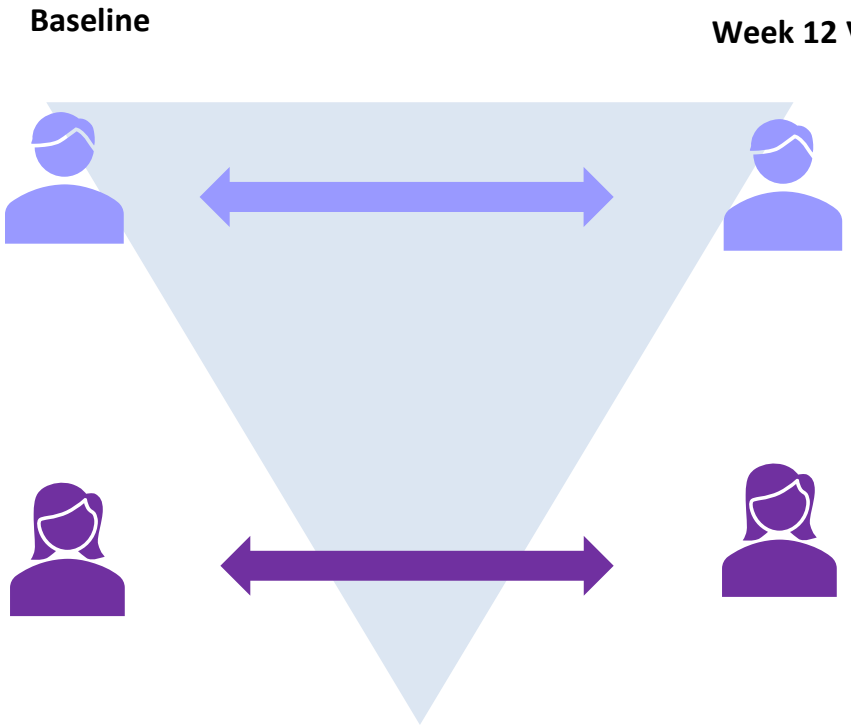
Clinical Trial Example 1

Randomised 1:1 two-arm trial

Primary endpoint:
Biomarker reduction: baseline and week 12 assessments

Success defined as treatment effect p-value <0.05 required in the primary endpoint

Correlation in clinical trials – within individuals



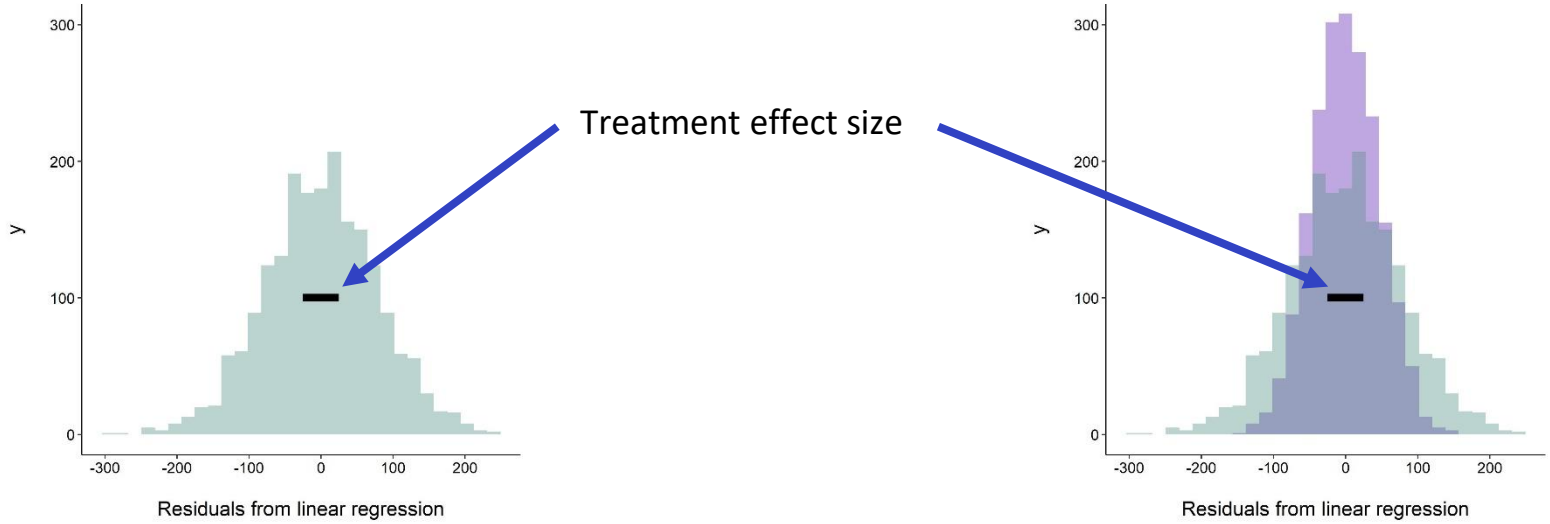
Participants with high baseline biomarker levels more likely to have a higher level at the week 12 visit.

Participants with low baseline biomarker levels more likely to have a lower level at the week 12 visit.

Correlation in clinical trials – within individuals

- Variance is large
- Treatment effect is small
- Difficult to identify treatment effect

- Following adjustment for baseline levels
- "Explains" a proportion of the variance
- Increases chance of identifying treatment effect



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

How does our probability of success change if we factor-in the correlation between the baseline and week 12 visit?

Fewer participants in this trial would be attractive on multiple levels, provided no loss in probability of success.

CEO

Can we increase our probability of success if we factor-in the expected correlation?

Statistician

CMO

Very reliable data that suggests baseline and week 12 visits will be correlated.

Patients

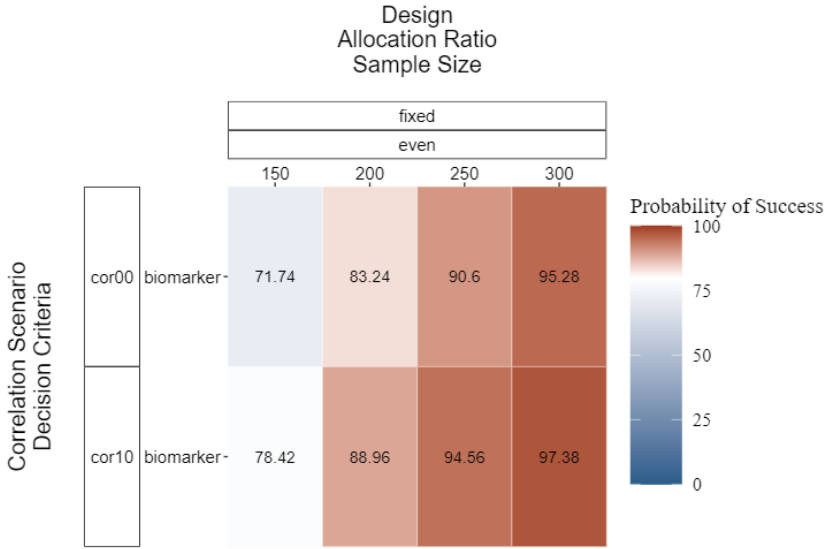
If it leads to faster approval, we get access to treatment sooner!

exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Clinical trial 1 – within individual correlation

The probability of observing p<0.05 for the biomarker endpoint increases in the scenario where there is a correlation between baseline and week 12 values of the biomarker.



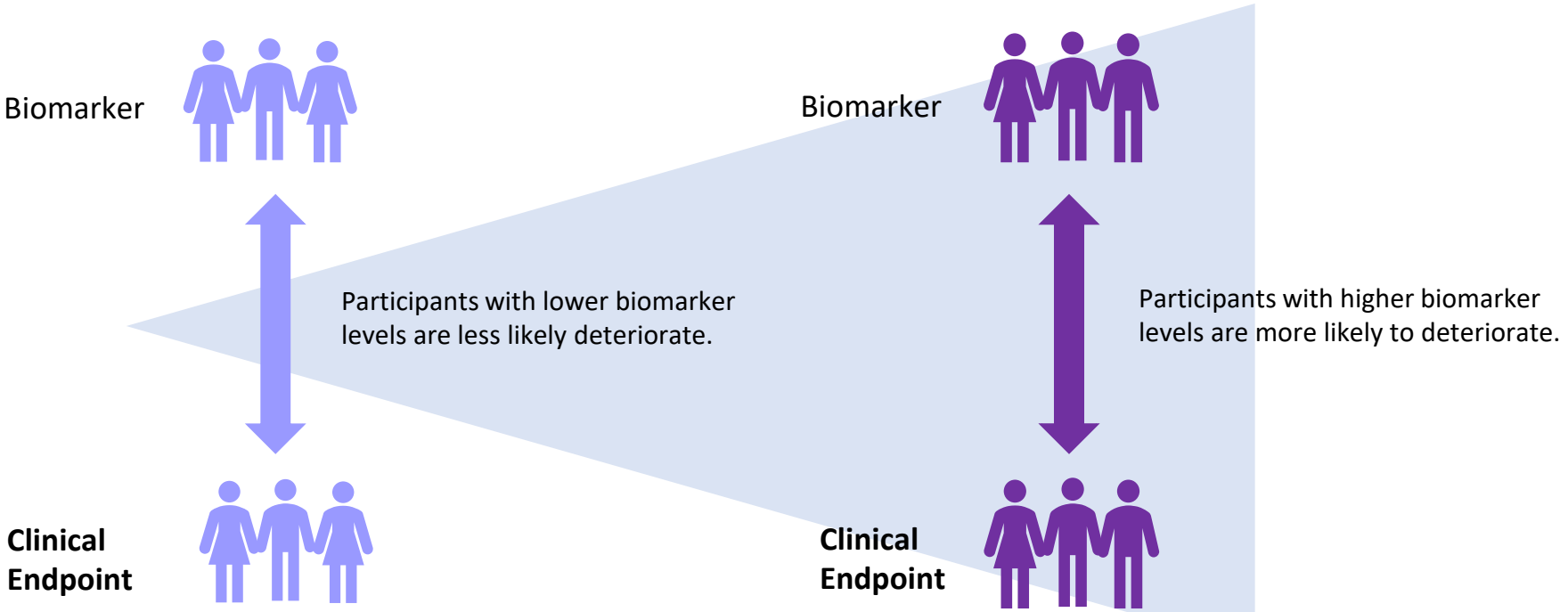
Clinical Trial Example 2

Randomised 1:1 two-arm trial

Co-primary:
1. Biomarker reduction
2. Reduction in a clinical deterioration

Success defined as treatment effect p-value <0.05 required for ***both*** biomarker reduction & clinical deterioration reduction

Correlation in clinical trials – between endpoints



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

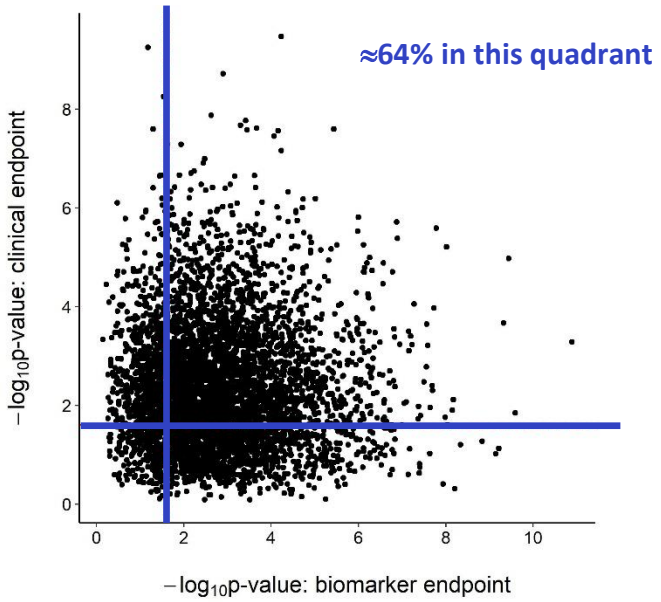
Correlation in clinical trials – between endpoints

Probability of "significant" biomarker p-value = **0.8**

Assuming **no correlation** between endpoints

Probability of "significant" clinical endpoint p-value = **0.8**

0.8 x 0.8 = 0.64



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

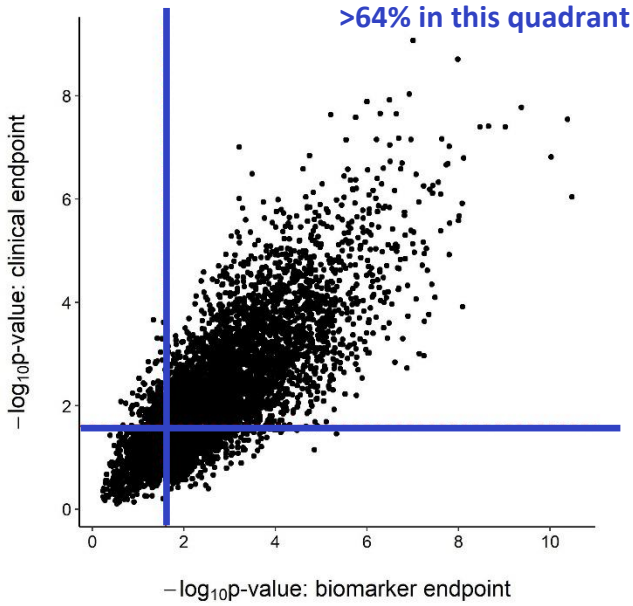
Correlation in clinical trials – between endpoints

Probability of "significant" biomarker p-value = **0.8**

Assuming **correlation** between endpoints

Probability of "significant" clinical endpoint p-value = **0.8**

≠0.8 x 0.8



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

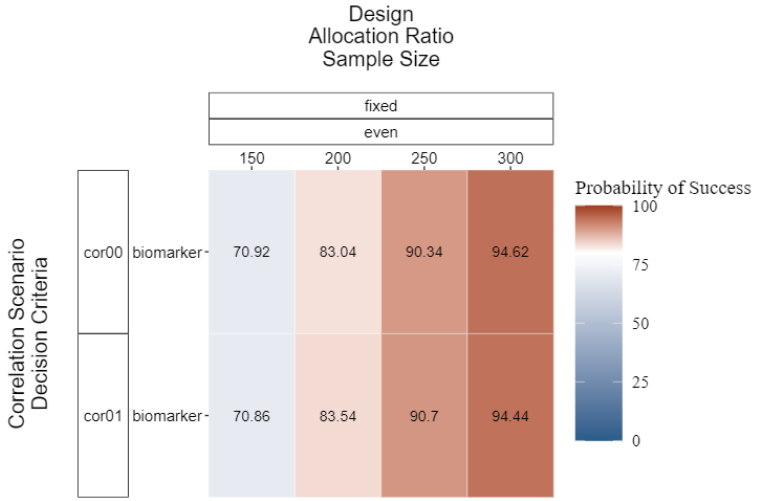
How does our probability of success change if we factor-in the correlation between the two endpoints?

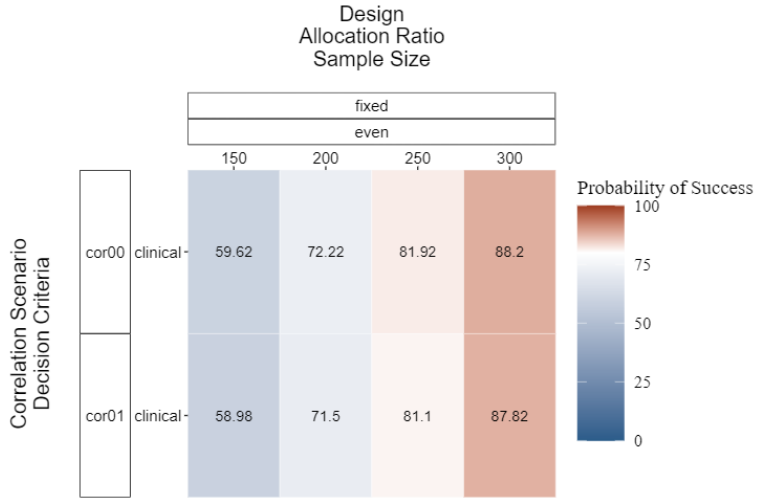
exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Clinical trial 2 – co-primary endpoints

The probability of observing p<0.05 for the biomarker and clinical endpoints is very similar regardless of the scenario.



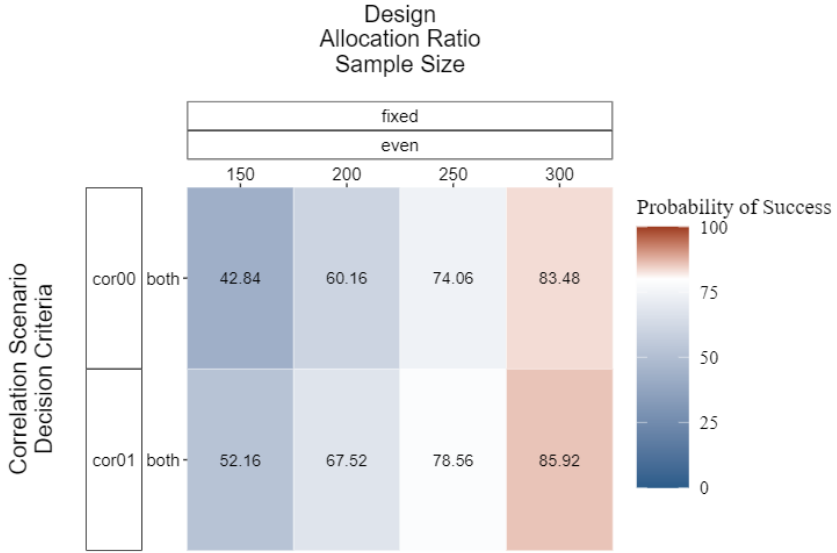


exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Clinical trial 2 – co-primary endpoints

However, correlation between endpoints means that the probability of observing both endpoints with $p<0.05$ increases in the scenarios where the correlation is 0.8 between endpoints.



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

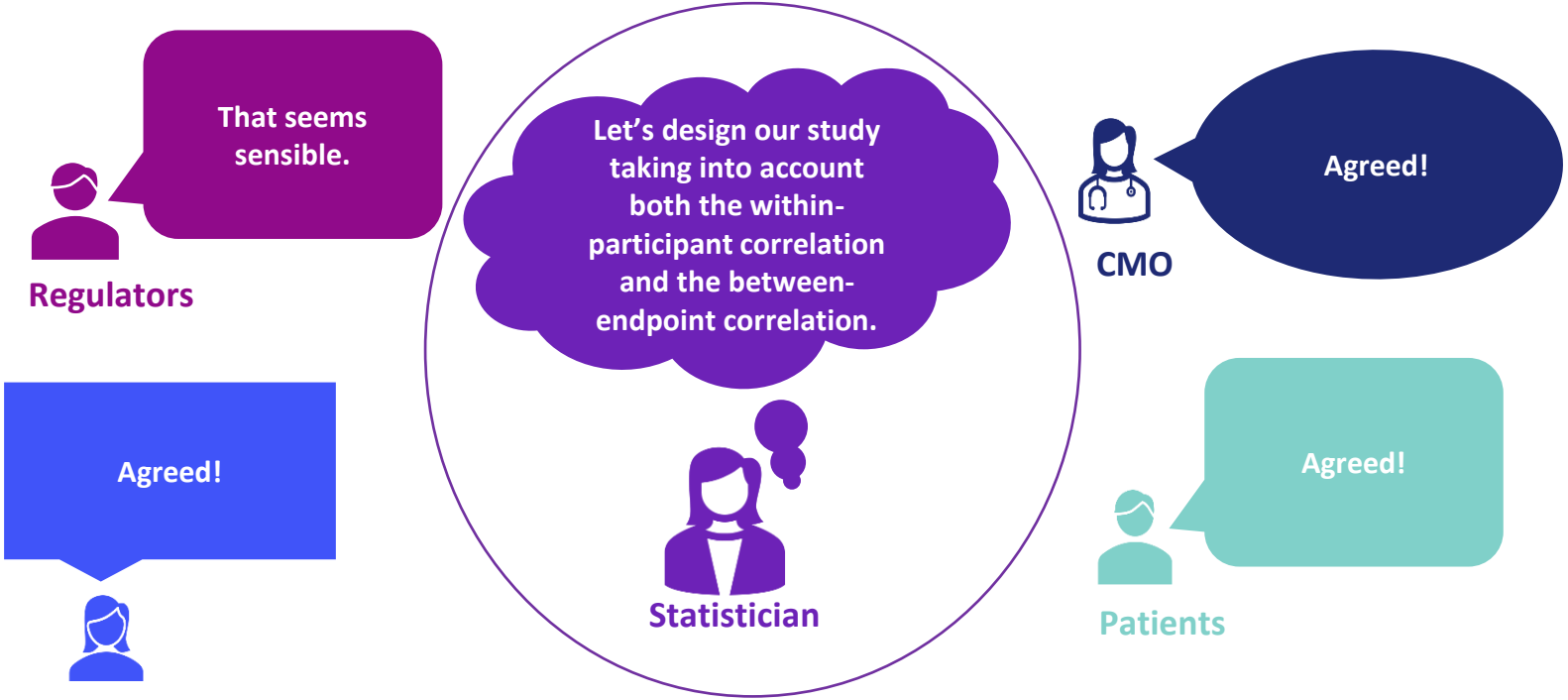
Clinical Trial Example 3

Randomised 1:1 two-arm trial

Co-primary:
1. Biomarker reduction: baseline and week 12 visit assessments
2. Reduction in a clinical deterioration

Success defined as treatment effect p-value <0.05
required for ***both*** biomarker reduction & clinical deterioration reduction

How does our probability of success change if we factor-in both the correlation within individual and between the two endpoints?



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

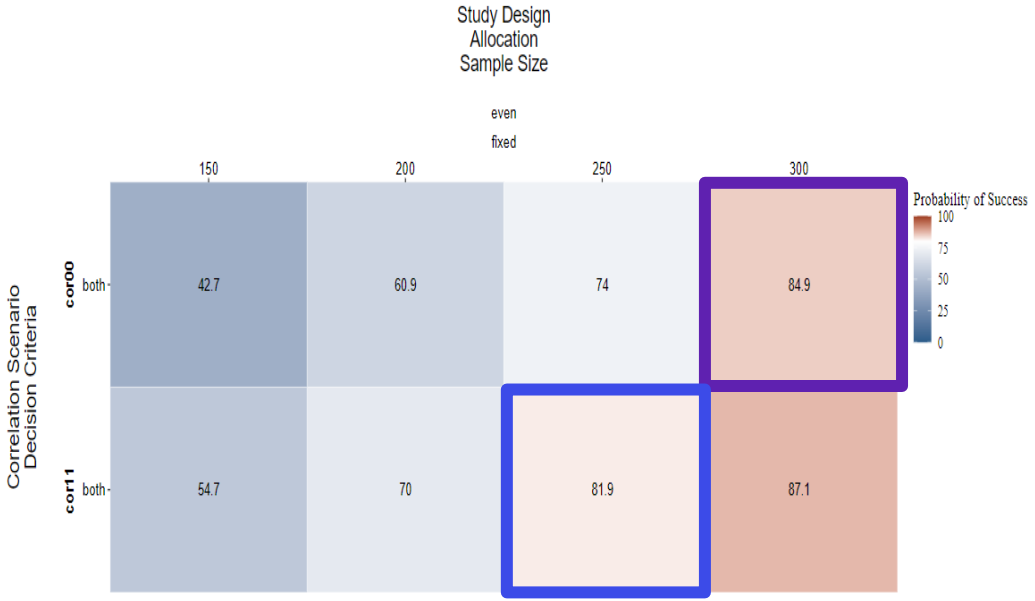
The Effective Statistician Academy

Clinical trial 3 – co-primary endpoints

Recruiting less participants without impacting the probability of success of the trial can save many thousands of pounds with no increased risk of study failure

In this study, approximately 300 participants would be required to get >80% probability of success if the correlations are ignored
(or assumed to be 0).

By factoring in:
i) the correlation between the baseline and week 12 measurement of the biomarker, and
ii) the correlation between the biomarker and the clinical endpoint,
We reduce the sample size requirement by approximately 50 participants to obtain the same probability of success.



exploristics
INNOVATIVE ANALYTICS SOLUTIONS

The Effective Statistician Academy

Summary of Examples

Clinical Trial 1:
within-participant correlation harnessed to increase probability of success.

Clinical Trial 2:
co-primary endpoint required by regulators. Between-endpoint correlation factored-in to design.

Clinical Trial 3:
within-participant correlation and between-endpoint correlation both factored-in to design to increase probability of success.

- Simulation of both the within-participant correlation and the between-endpoint correlation allowed us to:
  - Understand the probability of success for each endpoint separately (conditional on assumptions)
  - Understand the overall trial probability of success (conditional on requiring significance on both endpoints)
  - Reduce the requisite number of participants vs. traditional simple sample size calculations

Summary

- Simulation is the act of imitating real life and is used for study, training, testing or demonstrating purposes.
- Using simulation for clinical trial design often makes things easier and allows for better interpretations.
- Simulations are impactful throughout the whole drug development process.