Estimand Sample Sizes

Thomas Drury – GSK Statistics & Data Science Innovation

Estimand Sample Sizes

Acknowledgment and Disclaimer

- The following people were also part of the ongoing work and discussions highlighted in this talk:
  - Dan Bratton (GSK).
  - Sunita Rehal (GSK).
  - Fi Guillard (Veramed UK).

The talk reflects my own opinions and are not necessarily the views of GSK or the people acknowledged above

Estimand Sample Sizes

Introduction

- The estimands framework has made the target of inference more specific (clearer).
- Information on intercurrent events (IEs) required and how they are reflected in the effect (IE strategy).
- The release of ICH E9(R1) is positive step increasing awareness and discussions around:
  - The effect of interest – Which effects are well aligned to answer the main clinical question?
  - The conduct of trials – Awareness that data may need collecting post-IE.
  - The statistical analysis – Are the standard estimation methods good enough anymore?

Estimand Sample Sizes

Introduction

- However, the addendum also spells out trial design should be informed by the chosen estimand:

> *"The specification of appropriate estimands (see A.3.) will usually be the main determinant for aspects of trial design, conduct (see A.4.) and analysis (see A.5.)…"*

- So far, across industry the impact of IEs on study design and sample size seems a bit vague.
- Only found a single paper* looking at sample size and IE impact.
- Anecdotally have seen implicit assumptions:
  - Sample sizes based on effects and variability for **patients receiving the treatment**.
  - Sample sizes based on effects and variability which is "***some aggregate of everything***" pre- and post-IEs.

* Fang, Y., & Jin, M. (2021). Sample Size Calculation When Planning Clinical Trials with Intercurrent Events. Therapeutic Innovation and Regulatory Science, 55(4), 779–785.

Estimand Sample Sizes

Introduction

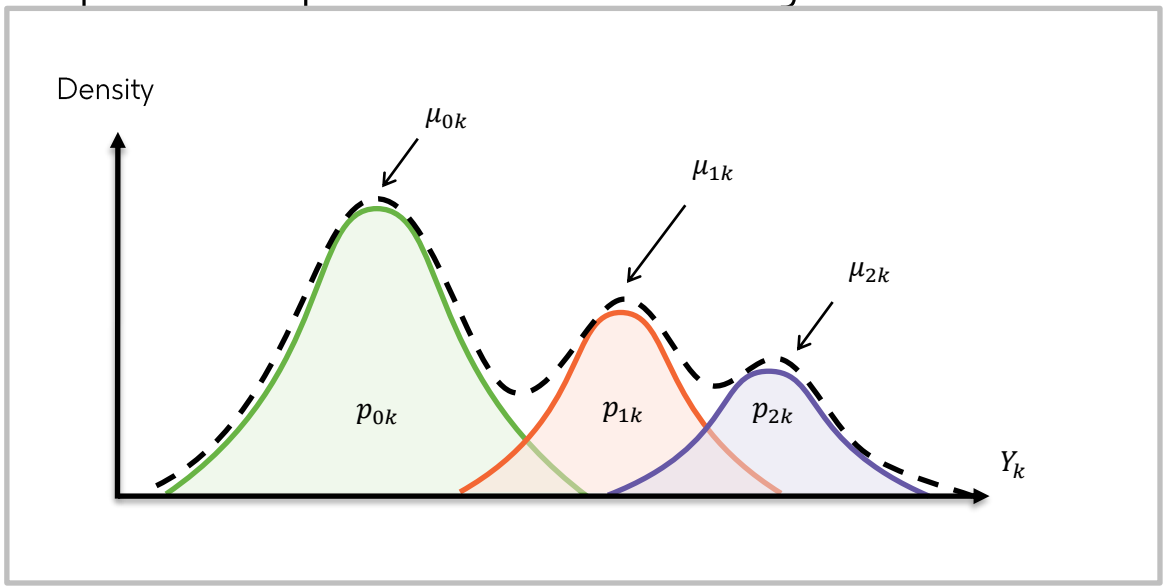
- It would be **better** to do **more thorough** assessments of **sample size** and power.
- One option is using **patient level simulation.**
  - Generally, **time consuming** to develop code and run large simulations.
  - Creates a **lot of work** for study statisticians.

Can we approximate power for estimands without patient level simulations?

Estimand Sample Sizes

The outcomes are a mixture

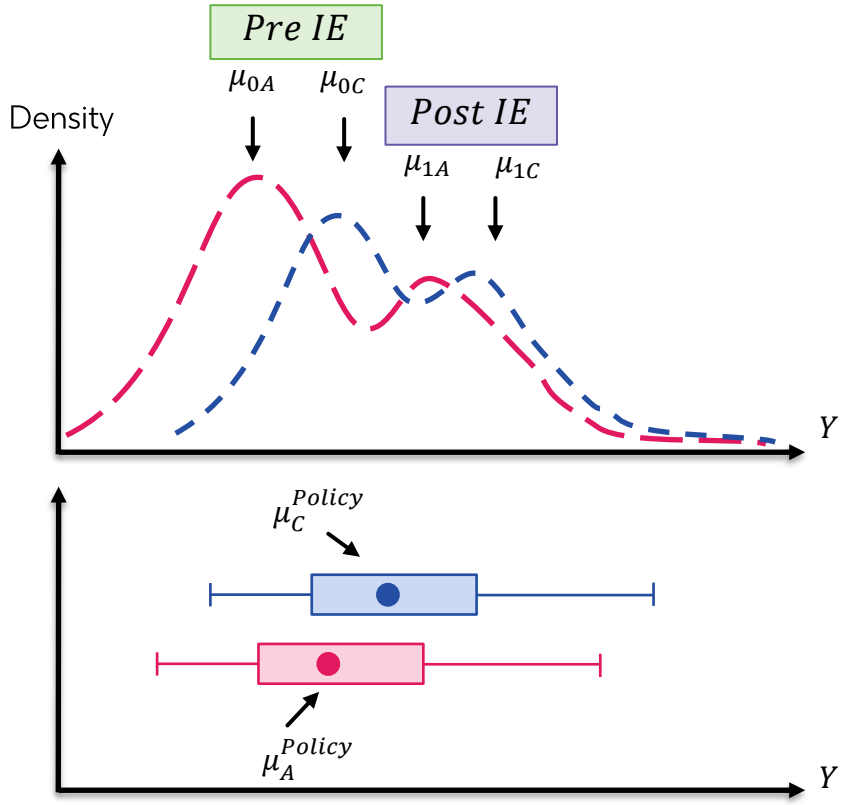
- Patients in group k at the primary timepoint are either still "**as randomized**" or in some other "**post-IE**" states S.
- The **outcome distribution** is then a **mixture** of patients in different "**states**".
- Can **use mixture theory** to help with the power calculation – e.g. for continuous outcomes:



Estimand Sample Sizes

Simple continuous example with a single IE

Single IE “Discontinuation from assigned treatment” using the treatment policy strategy.



- From finite mixture theory:

$$E\left[Y_k^{Policy}\right] = (1-p_k)\cdot\mu_{0k} + p_k\cdot\mu_{1k}$$
$$\mathrm{V}\left[Y_k^{Policy}\right] = (1-p_k)\cdot\sigma_{0k}^2 + p_k\cdot\sigma_{1k}^2 + (1-p_k)\cdot p_k\cdot(\mu_{0k}-\mu_{1k})^2$$

- Assuming independence between groups:

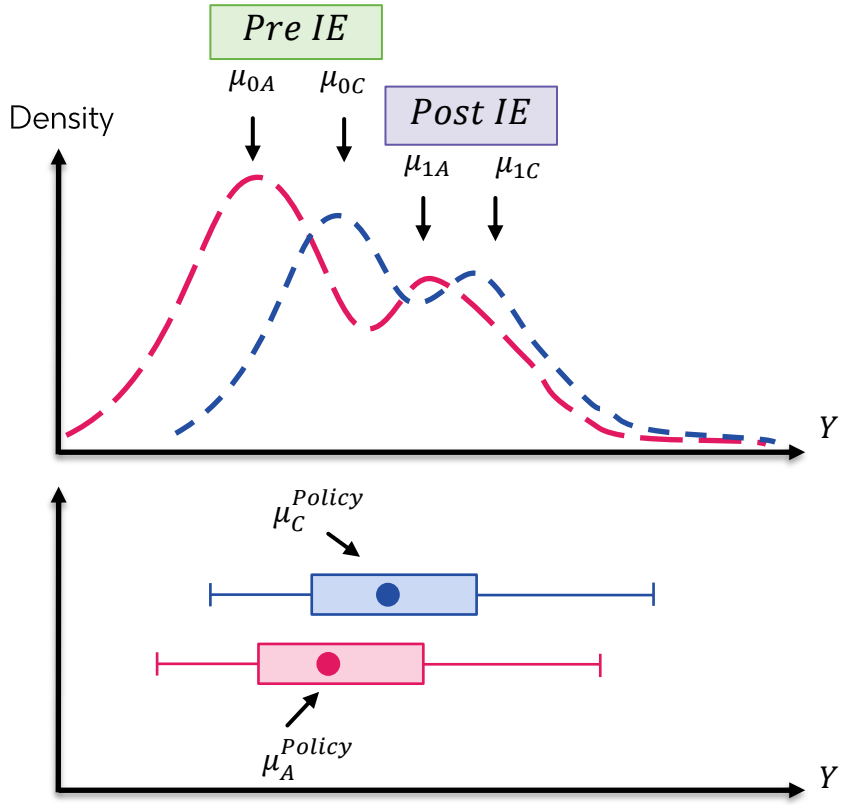
$$\Delta^{Policy} = E\left[Y_A^{Policy}\right] - E\left[Y_C^{Policy}\right]$$
$$\sigma^{Policy} = \sqrt{\mathrm{V}\left[Y_A^{Policy}\right] + \mathrm{V}\left[Y_C^{Policy}\right]}$$

$$N = \left(Z_{\alpha/2} + Z_\beta\right)^2 \cdot \left(\frac{\sigma^{Policy}}{\Delta^{Policy}}\right)^2$$

Estimand Sample Sizes

Simple continuous example with a single IE

Single IE “Discontinuation from assigned treatment” using the treatment policy strategy.



- From finite mixture theory:

$$E[Y_k^{Policy}] = (1 - p_k) \cdot \mu_{0k} + p_k \cdot \mu_{1k}$$
$$V[Y_k^{Policy}] = (1 - p_k) \cdot \sigma_{0k}^2 + p_k \cdot \sigma_{1k}^2 + (1 - p_k) \cdot p_k \cdot (\mu_{0k} - \mu_{1k})^2$$

- Assuming independence between groups:

$$\Delta^{Policy} = E[Y_A^{Policy}] - E[Y_C^{Policy}]$$
$$\sigma^{Policy} = \sqrt{V[Y_A^{Policy}] + V[Y_C^{Policy}]}$$

$$N = (Z_{\alpha/2} + Z_\beta)^2 \cdot \left(\frac{\sigma^{Policy}}{\Delta^{Policy}}\right)^2$$

- Requires information on a lot of quantities.
- Unlikely to have reliable information from the literature.

Estimand Sample Sizes

Create simple but reasonable assumptions

- If we are **happy** to make **reference-based assumptions** in our **analyses (e.g. Reference based MI)**.
- We can **do the same** in the **sample size calculation** to get a reasonable "rule of thumb".
- The **most common IE strategies** are generally:
  - **Treatment Policy** – Could assume mean shift to control (e.g. **Jump to Reference**) as likely conservative.
  - **Composite** – Could assume shift to agreed failure value (e.g. **zero change**, or **percentiles** or **SD from mean**).
  - **Hypothetical** – Could assume post IE outcome removed (**set to missing**) collapses to standard SS calc.

Estimand Sample Sizes

Key benefits to formula approximations for power calculations

- Allows **fast assessment of power** compared to patient level simulation.
- Allows **clear assessment** of the **impact of IEs** and **missing data**:
  - Chosen **IE strategy**.
  - Expected **IE rates**.
  - Expected **Missingness rates**.
- Allows **easier simulation of assurance (PoS)**.

Estimand Sample Sizes

Simple continuous example with treatment policy IE

Primary Estimand

Population	Patients with mild cognitive impairment due to Alzheimer's disease.
Endpoint / Variable	Change from baseline clinical dementia rating sum of boxes (CDR-SoB) at week 96 including any subsequent effects of treatment discontinuation.
Treatments	Assignment to Active or Control.
Summary Measure	Difference in mean.
Intercurrent Events	Treatment discontinuation (treatment policy).

Description:

Difference in mean change from baseline clinical dementia rating sum of boxes (CDR-SoB) at week 96 for patients with mild cognitive impairment due to Alzheimer's disease assigned to Active or Control including any subsequent effects of treatment discontinuation.

Estimand Sample Sizes

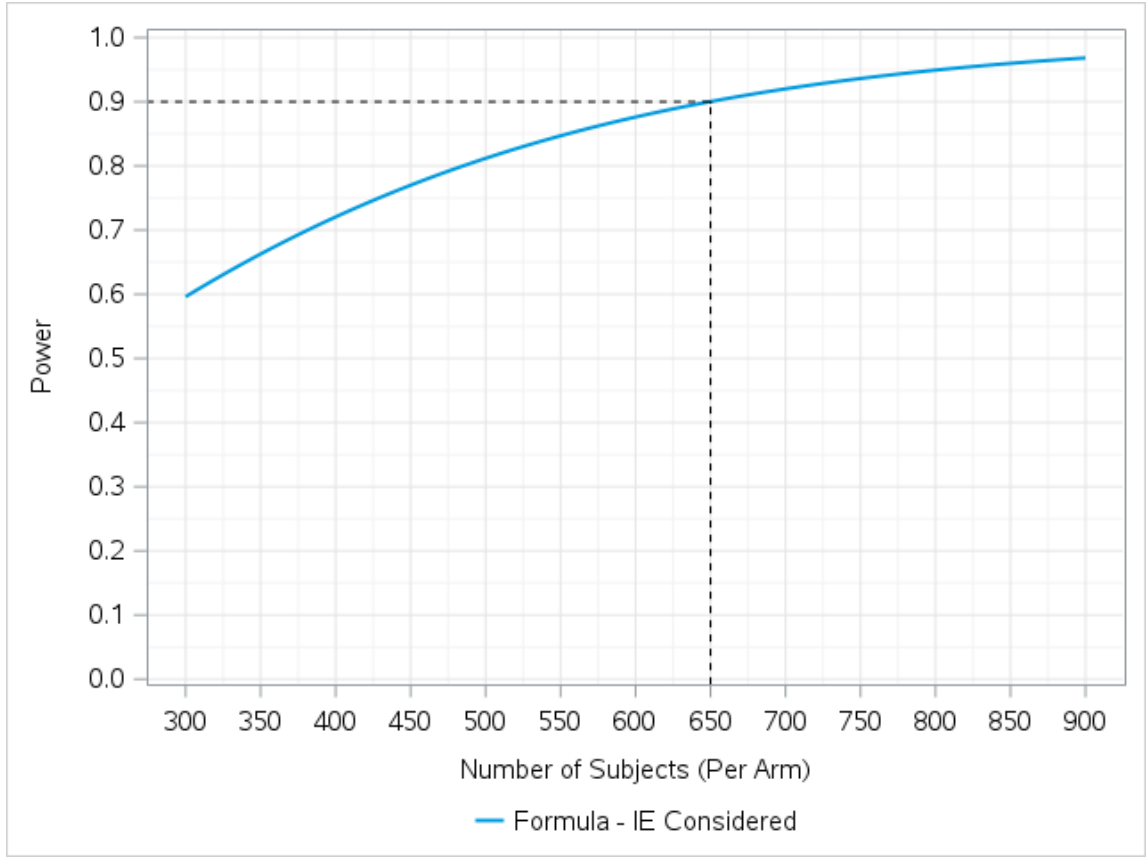
Simple continuous example with treatment policy IE

- Assumptions for power calculation:
  - When **control is taken for 96 weeks** the average change from baseline CDR-SoB is -2.0 (negative = worse).
  - When **active is taken for 96 weeks** the average change from baseline CDR-SoB is -1.4 (30% effect).
  - The **SD is** 2.8 for **both control and active when "on-treatment"**.
  - The rates of **treatment discontinuation** are 15% in each arm.
  - Any **treatment discontinuation** creates instant loss of effect (likely conservative).
  - The **off-treatment SD is** 3.0 for **both control and active** (different treatment options).
  - Aiming for 90% **power** with 5% **significance** (two-sided).

Estimand Sample Sizes

Simple continuous example with treatment policy IE

```
data power_formula1;

 mu0_c = 2.00; mu1_c = 2.00;   *** ON AND OFF CONTROL MEANS ***;
 mu0_a = 1.40; mu1_a = 2.00;   *** ON AND OFF ACTIVE MEANS  ***;
 sd0_c = 2.85; sd1_c = 2.85;   *** ON AND OFF CONTROL SD ***;
 sd0_a = 2.85; sd1_a = 2.85;   *** ON AND OFF ACTIVE SD  ***;
 p_c   = 0.15;                 *** PROPORTION OF IES IN CONTROL ARM ***;
 p_a   = 0.15;                 *** PROPORTION OF IES IN ACTIVE ARM ***;
 alpha = 0.05;                 *** SIGNIFICANCE LEVEL ***;

 m_policy_c = (1-p_c)*mu0_c + p_c*mu1_c;
 m_policy_a = (1-p_a)*mu0_a + p_a*mu1_a;

 v_policy_c = (1-p_c)*(sd0_c**2) + p_c*(sd1_c**2) + ((1-p_c)*p_c*(mu0_c - mu1_c)**2);
 v_policy_a = (1-p_a)*(sd0_a**2) + p_a*(sd1_a**2) + ((1-p_a)*p_a*(mu0_a - mu1_a)**2);

 do nsub = 300 to 900 by 1;
  delta = abs(m_policy_a - m_policy_c);
  se    = sqrt((v_policy_c/nsub) + (v_policy_a/nsub));
  power = probnorm((delta/se)-probit(1-alpha/2));
  output;
 end;

run;
```

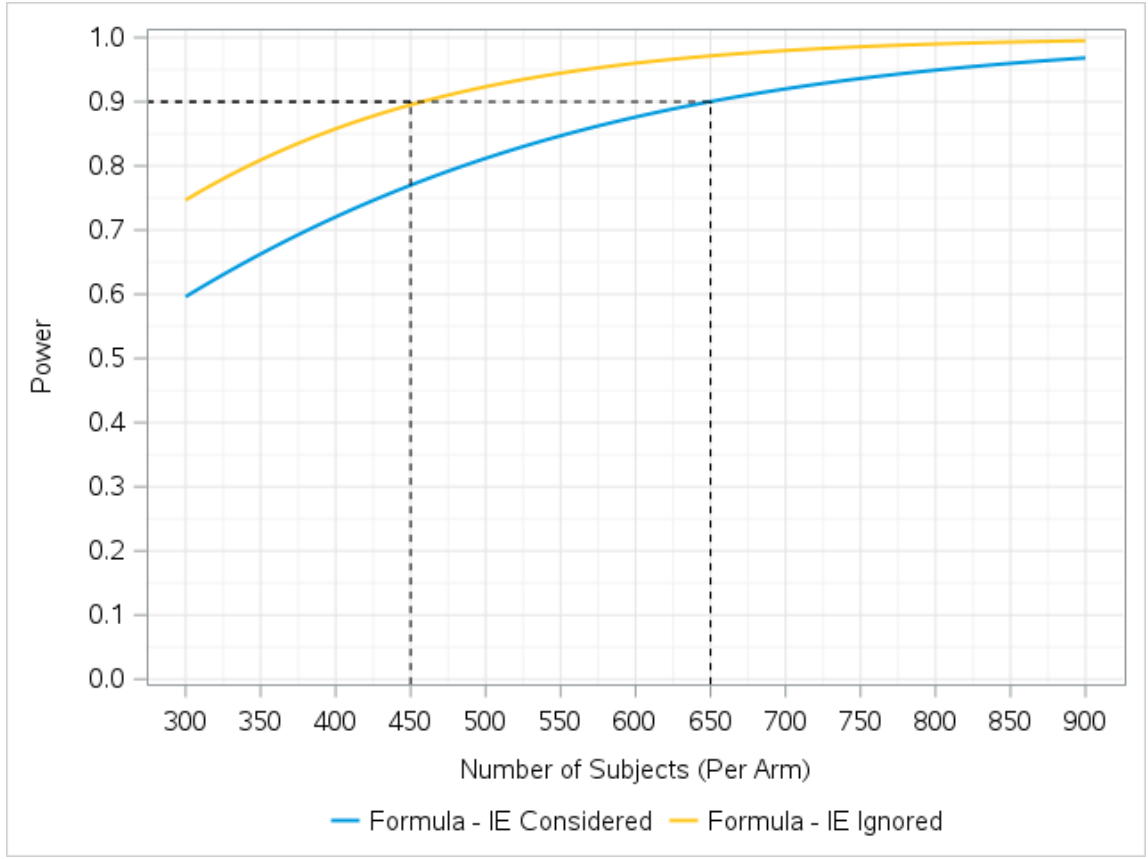


Estimand Sample Sizes

Simple continuous example with treatment policy IE

```
data power_formula1;

 mu0_c = 2.00; mu1_c = 2.00;   *** ON AND OFF CONTROL MEANS ***;
 mu0_a = 1.40; mu1_a = 2.00;   *** ON AND OFF ACTIVE MEANS  ***;
 sd0_c = 2.85; sd1_c = 2.85;   *** ON AND OFF CONTROL SD ***;
 sd0_a = 2.85; sd1_a = 2.85;   *** ON AND OFF ACTIVE SD  ***;
 p_c   = 0.15;                 *** PROPORTION OF IES IN CONTROL ARM ***;
 p_a   = 0.15;                 *** PROPORTION OF IES IN ACTIVE ARM ***;
 alpha = 0.05;                 *** SIGNIFICANCE LEVEL ***;

 m_policy_c = (1-p_c)*mu0_c + p_c*mu1_c;
 m_policy_a = (1-p_a)*mu0_a + p_a*mu1_a;

 v_policy_c = (1-p_c)*(sd0_c**2) + p_c*(sd1_c**2) + ((1-p_c)*p_c*(mu0_c - mu1_c)**2);
 v_policy_a = (1-p_a)*(sd0_a**2) + p_a*(sd1_a**2) + ((1-p_a)*p_a*(mu0_a - mu1_a)**2);

 do nsub = 300 to 900 by 1;
  delta = abs(m_policy_a - m_policy_c);
  se    = sqrt((v_policy_c/nsub) + (v_policy_a/nsub));
  power = probnorm((delta/se)-probit(1-alpha/2));
  output;
 end;

run;
```

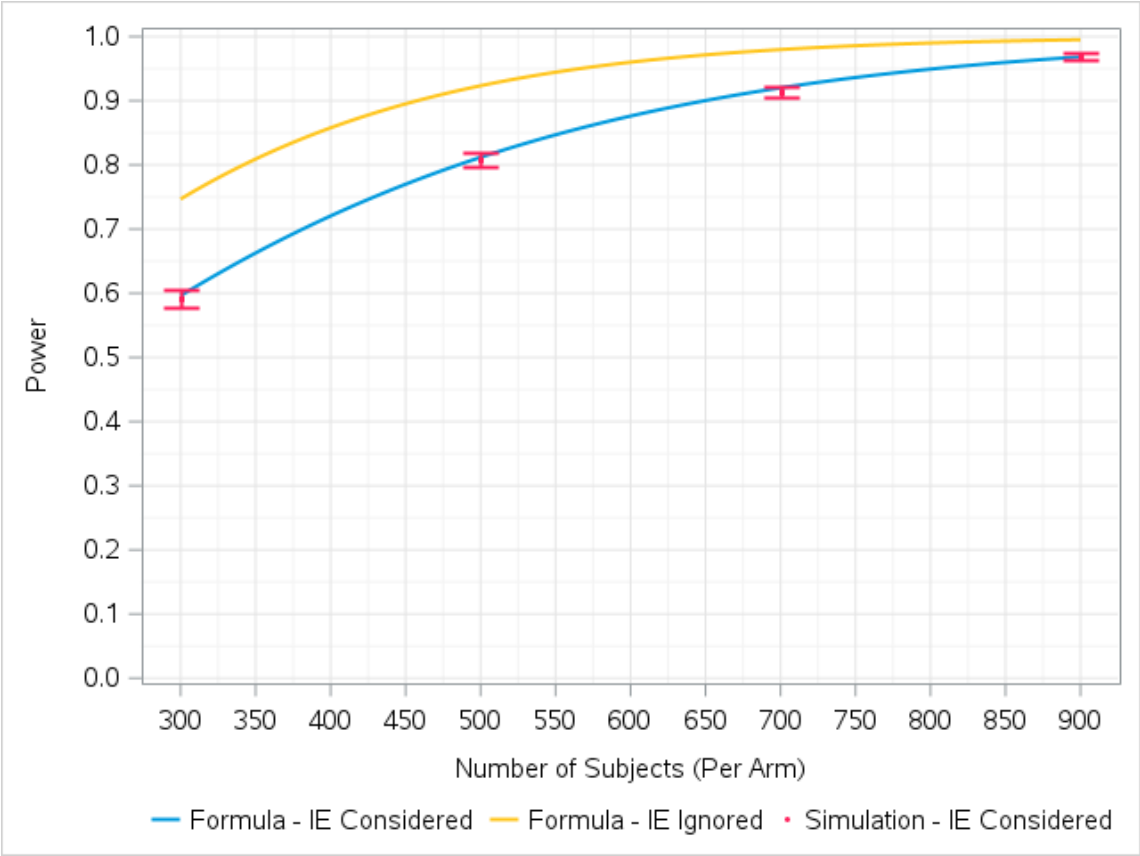
```
data power_formula2;

 mu_c = 2.0;     *** ON CONTROL MEANS ***;
 mu_a = 1.4;     *** ON ACTIVE MEANS  ***;
 sd_c = 2.8;     *** ON CONTROL SD ***;
 sd_a = 2.8;     *** ON ACTIVE SD  ***;
 alpha = 0.05;   *** SIGNIFICANCE LEVEL ***;

 do nsub = 300 to 900 by 1;
  delta = abs(mu_a - mu_c);
  se = sqrt(((sd_c**2)/nsub) + ((sd_a**2)/nsub));
  power = probnorm((delta/se)-probit(1-alpha/2));
  output;
 end;

run;
```



Estimand Sample Sizes

Simple continuous example with treatment policy IE

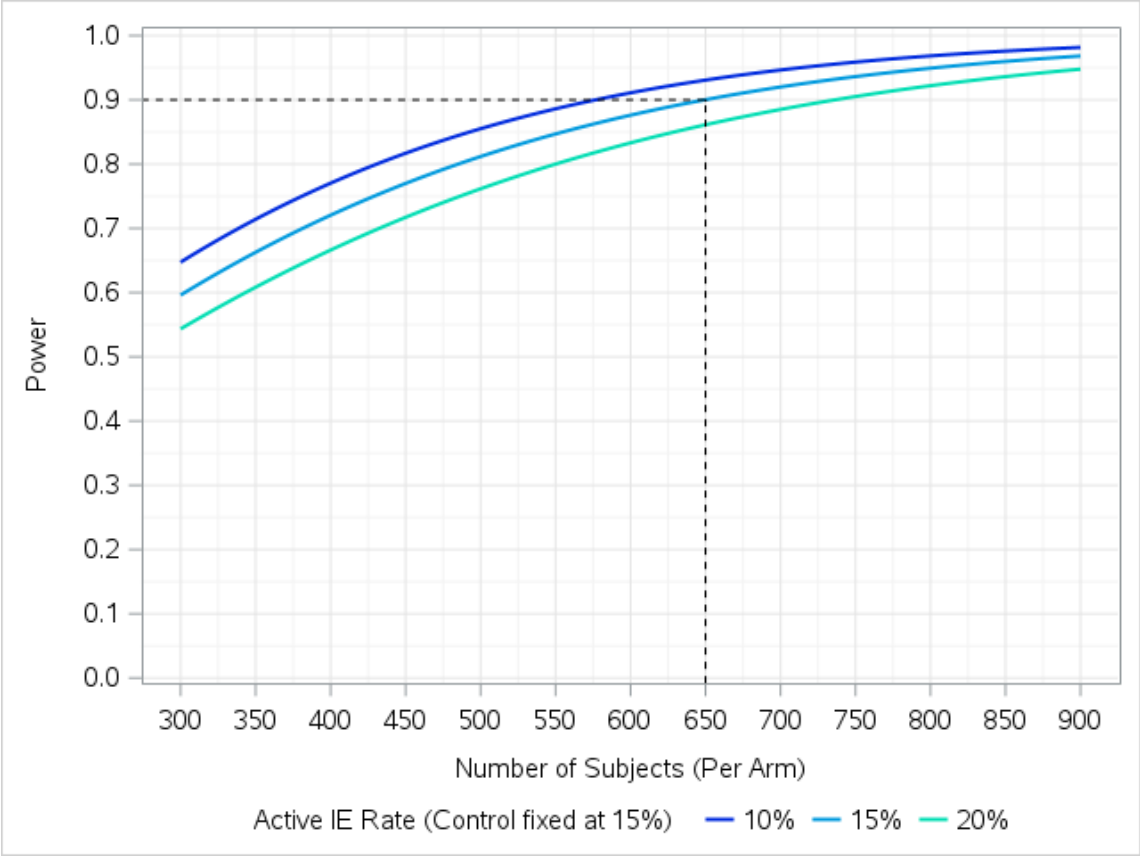
- Comparing to simulated power:
  - Approximate agreement (most 95% CIs overlap).
  - Formula ~ 1 second.
  - Simulation (5000 sims on 1 core) ~ 3 minutes.
  - Only 4 estimates → 90% power **between** 500-700.



Estimand Sample Sizes

Simple continuous example with treatment policy IE

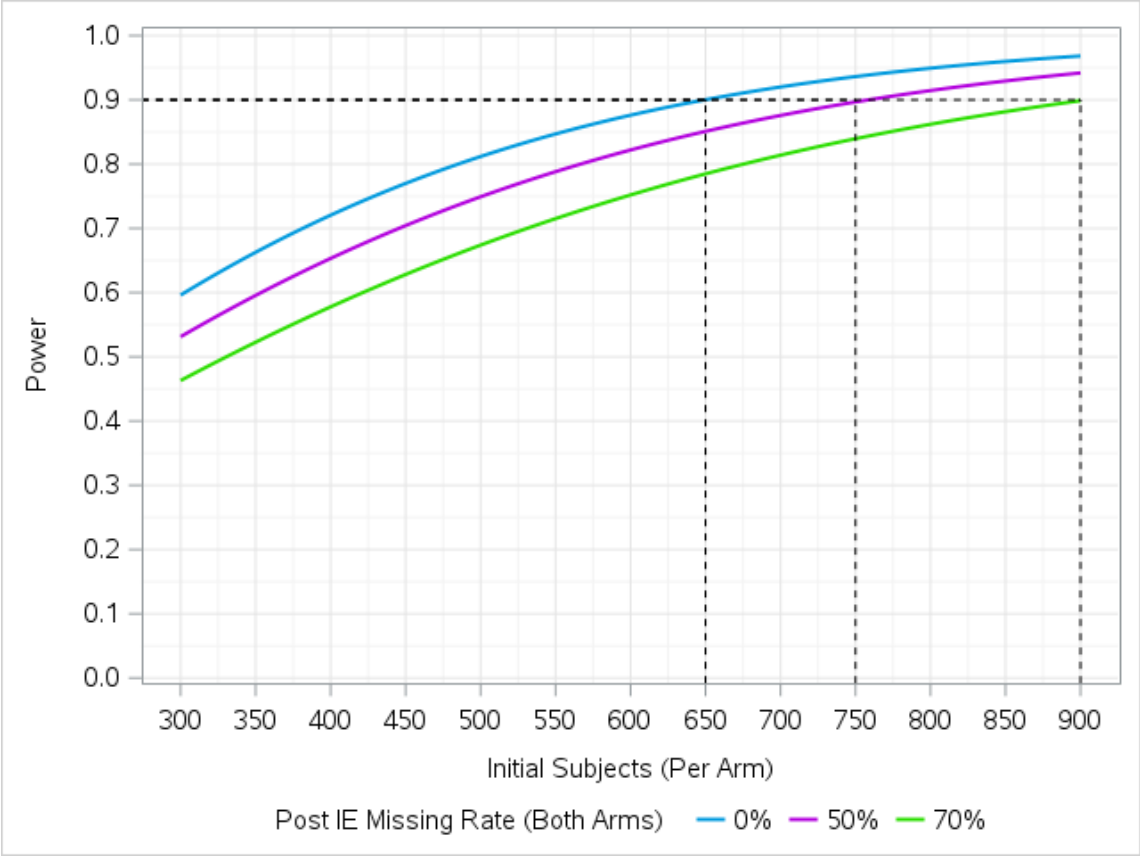
- Sample Size Sensitivity:
  - What if the IE rates are different?
  - E.g. assume active IE rates 10% or 20%.
  - Different IE rates can **improve** or **reduce** power.
  - Expected IE rates quite important.



Estimand Sample Sizes

Simple continuous example with treatment policy IE

- Sample Size Sensitivity – Missing Data:
  - Methods get more complex – but still possible*.
  - Power can depend on missing data assumptions.
  - Eg. assume missing data like post-IE data (RDMI).
  - Assume 50% or 70% missing data (post-IE).
  - Marginally ~ 7.5% and 10% overall missing data.



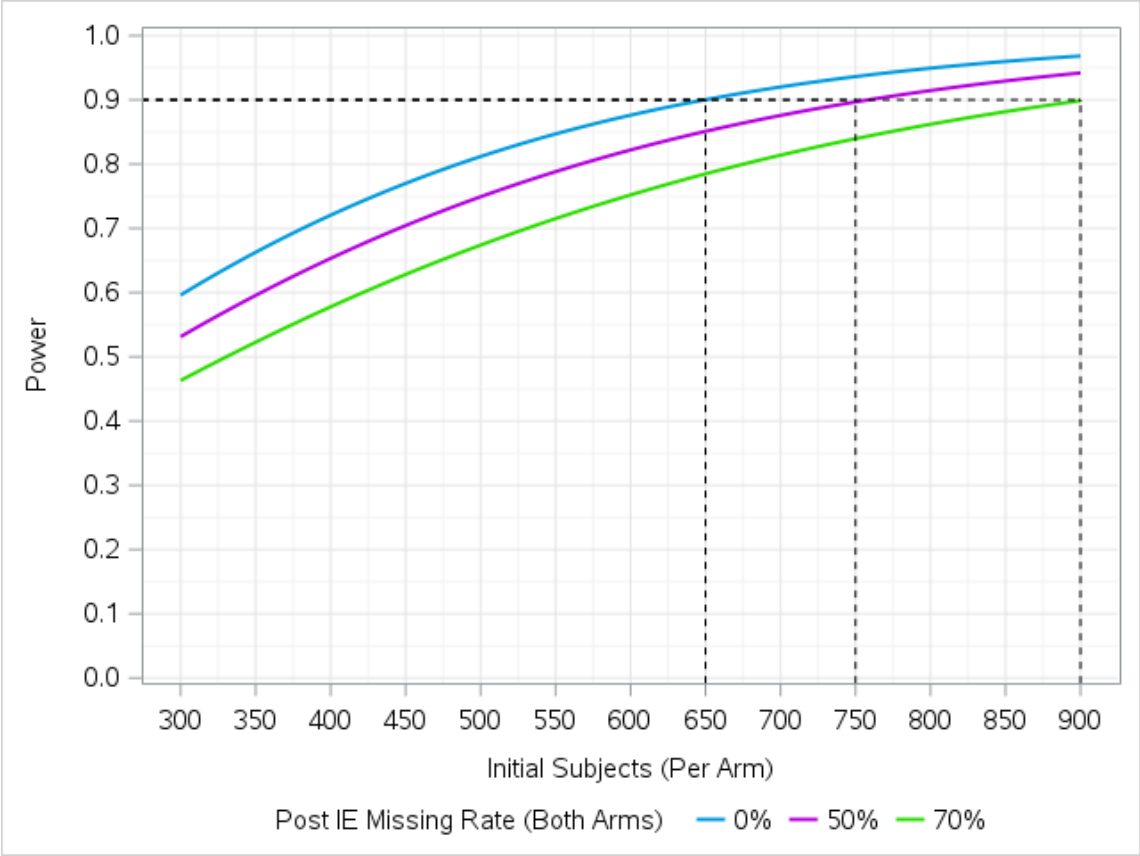
** Bell, J., Drury, T., Mütze, T., Pipper, C.B., Guizzaro, L., Mitroiu, M., Rantell, K.R., Wolbers, M. and Wright, D. (2025), Estimation Methods for Estimands Using the Treatment Policy Strategy; a Simulation Study Based on the PIONEER 1 Trial. Pharmaceutical Statistics, 24: e2472.* https://doi.org/10.1002/pst.2472

Estimand Sample Sizes

Simple continuous example with treatment policy IE

- Sample Size Sensitivity – Missing Data:
  - Methods get more complex – but still possible*.
  - Power can depend on missing data assumptions.
  - Eg. assume missing data like post-IE data (RDMI).
  - Assume 50% or 70% missing data (post-IE).
  - Marginally ~ 7.5% and 10% overall missing data.

The missing data rate can be **very** important.

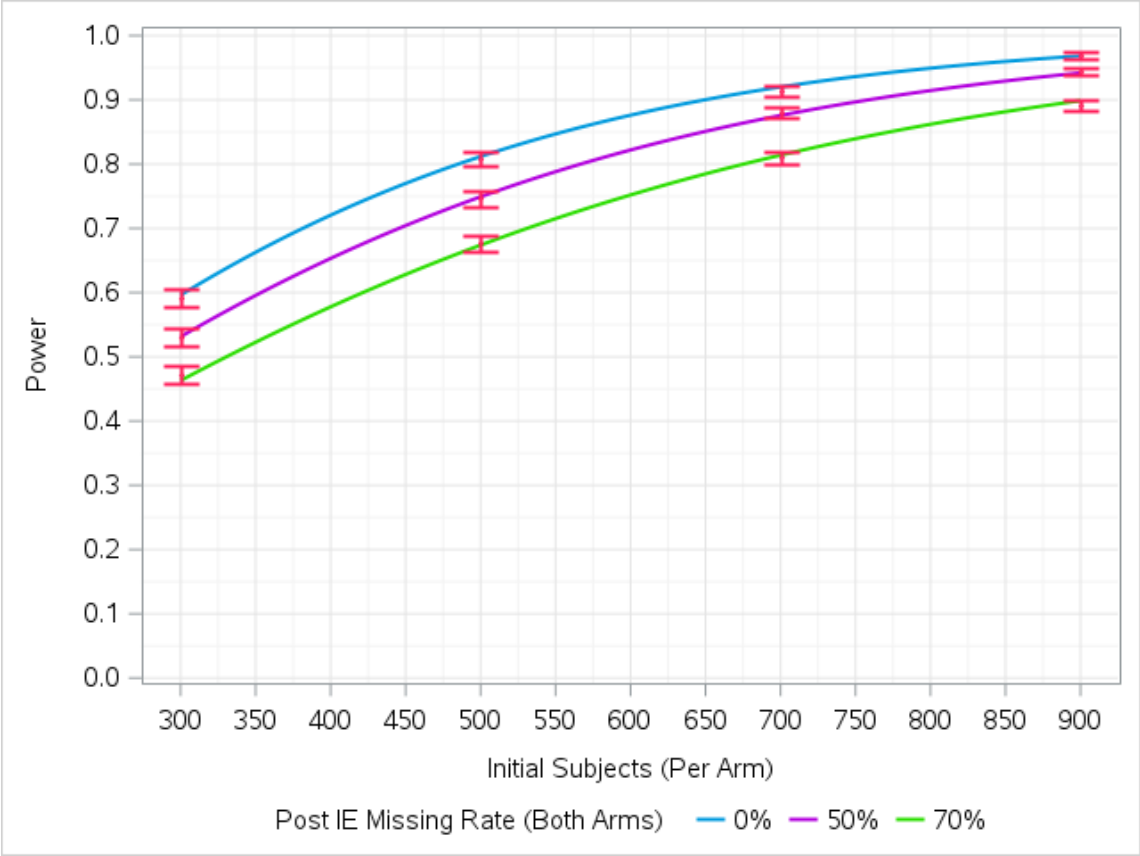


** Bell, J., Drury, T., Mütze, T., Pipper, C.B., Guizzaro, L., Mitroiu, M., Rantell, K.R., Wolbers, M. and Wright, D. (2025), Estimation Methods for Estimands Using the Treatment Policy Strategy; a Simulation Study Based on the PIONEER 1 Trial. Pharmaceutical Statistics, 24: e2472.* https://doi.org/10.1002/pst.2472

Estimand Sample Sizes

Simple continuous example with treatment policy IE

- Comparing to simulated power:
  - Approximate agreement (most 95% CIs overlap).
  - Formula ~ 1 second.
  - 5000 simulations on 1 core ~ 30 minutes.
  - Simulation requires multiple imputation.
  - Again, only 12 power estimates simulated.



Estimand Sample Sizes

Simulation validation work in progress – continuous outcomes

- **Stage 1**: Compare power for a **single IE vs. simulation** over large part of the **estimand "design space"**.
  - Effect sizes.
  - IE strategies, rates, and mechanisms.
  - Missing data.
- **Stage 2**: Develop and test methodology for **two IEs** (using same scenarios as stage 1).
- **Stage 3**: Develop **software tools** on Github to simplify work at the design stage.

Estimand Sample Sizes

Other ongoing simulation work

- Time to events – Hazard Ratios (PSI poster P045 – Dan Bratton).
- Binary outcomes – Risk Differences & Odds Ratios.
- Recurrent events – Rate Ratios for Negative Binomial.
- Comparing assurance calculations vs. subject-level simulations.

Estimand Sample Sizes

Summary

- The **design** of a clinical trial should be **aligned to the primary estimand**(s).
- The **rates of IEs and missing data** can have a **large impact** on the required sample size.
- Using **mixture theory** to account for IEs is **quicker and easier** than simulating patient level data.
- All results and code from our sample size comparison work will be opened sourced on GitHub*.

* *https://github.com/GSK-Biostatistics*

Estimand Sample Sizes

Summary

Thank you for listening!

Questions?