Clinical trial simulation: Planning with the OCTAVE framework and some computational principles

Kim May Lee

Department of Biostatistics and Health Informatics

King's College London, UK

16th April 2026

HRB-TMRN webinar

Outline of the talk

- Idea of clinical trial simulation (CST)
- OCTAVE framework for planning CTS
- The use of pseudocode
- Data generation approaches
- Validation of code
- Approximation approaches
- Graphical presentation and tools

Other topics covered in the tutorial paper

Statistics in Medicine

TUTORIAL IN BIOSTATISTICS | Open Access

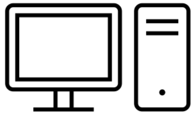
Clinical Trial Simulation: Planning With the OCTAVE Framework, Implementation and Validation Principles

Kim May Lee, Babak Choodari-Oskooei, Michael J. Grayling, Peter Jacko, Peter K. Kimani, Aritra Mukherjee, Philip Pallmann, Tom Parke, David S. Robertson, Ziyan Wang, Christina Yap, Thomas Jaki ... See fewer authors

First published: 16 March 2026 | https://doi.org/10.1002/sim.70449 | VIEW METRICS

- Components of clinical trials
- Simulation tasks management
- Presenting CTS result to stakeholders
- Reporting CTS in a grant application
- Suggestions on making code open-access

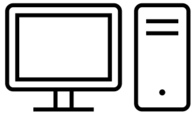
What is simulation



A simulation study uses a computer program to learn about different situations to understand what might happen, instead of doing a real-world experiment[1]

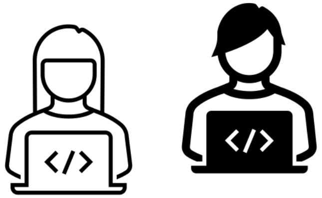
[1] Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P., 1989. Design and analysis of computer experiments. *Statistical science*, *4*(4), pp.409-423.

What is simulation



A simulation study uses a computer program to learn about different situations to understand what might happen, instead of doing a real-world experiment[1]

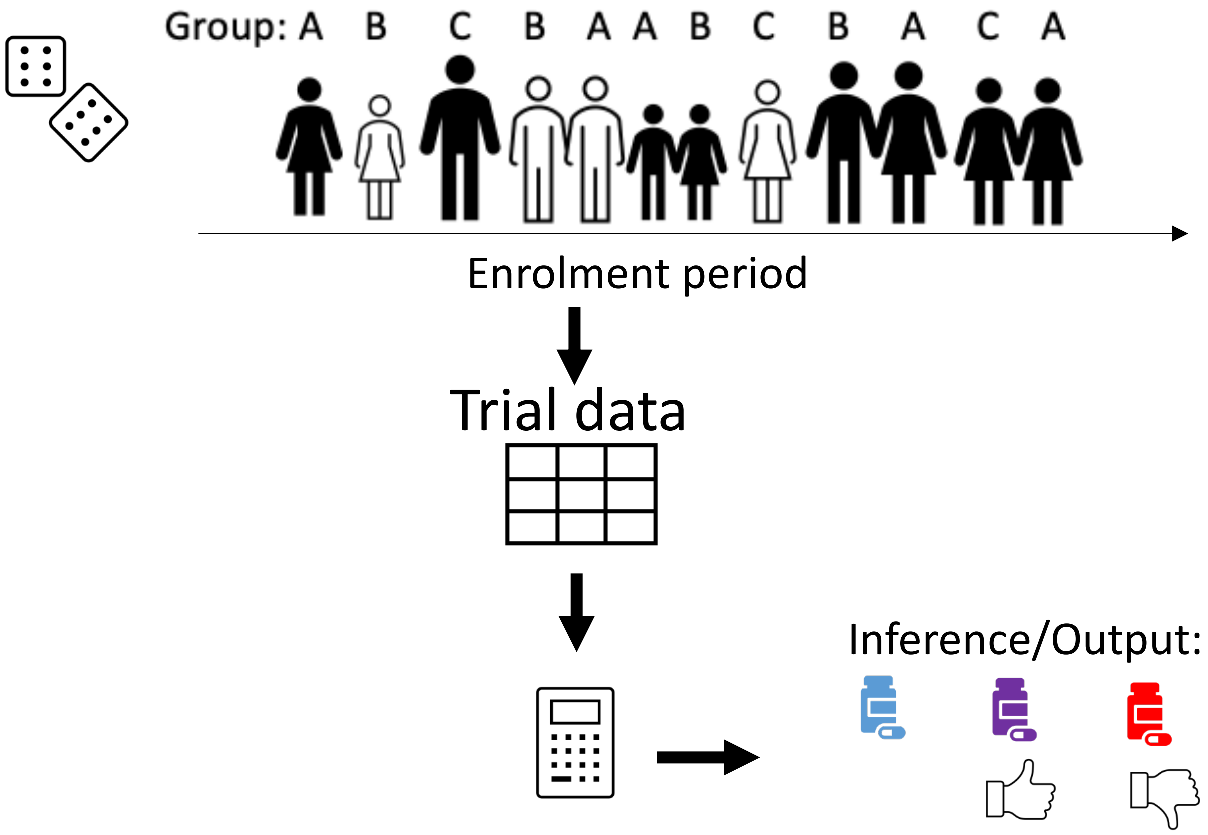
Clinical trial simulation uses computer models to explore trial scenarios and outcomes without running the real trial



[1] Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P., 1989. Design and analysis of computer experiments. *Statistical science*, *4*(4), pp.409-423.

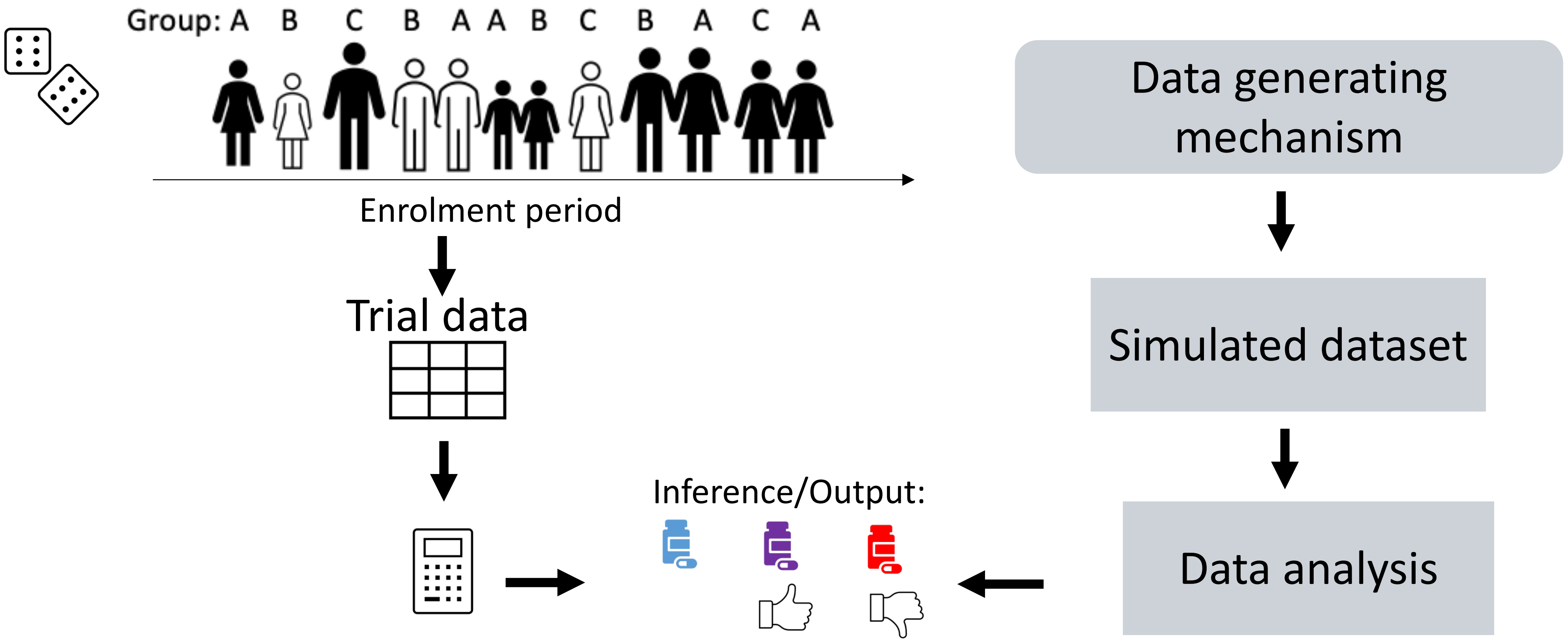
Clinical trial simulation can…

- Offer a preview of trial procedures and results before real-world implementation



Clinical trial simulation can…

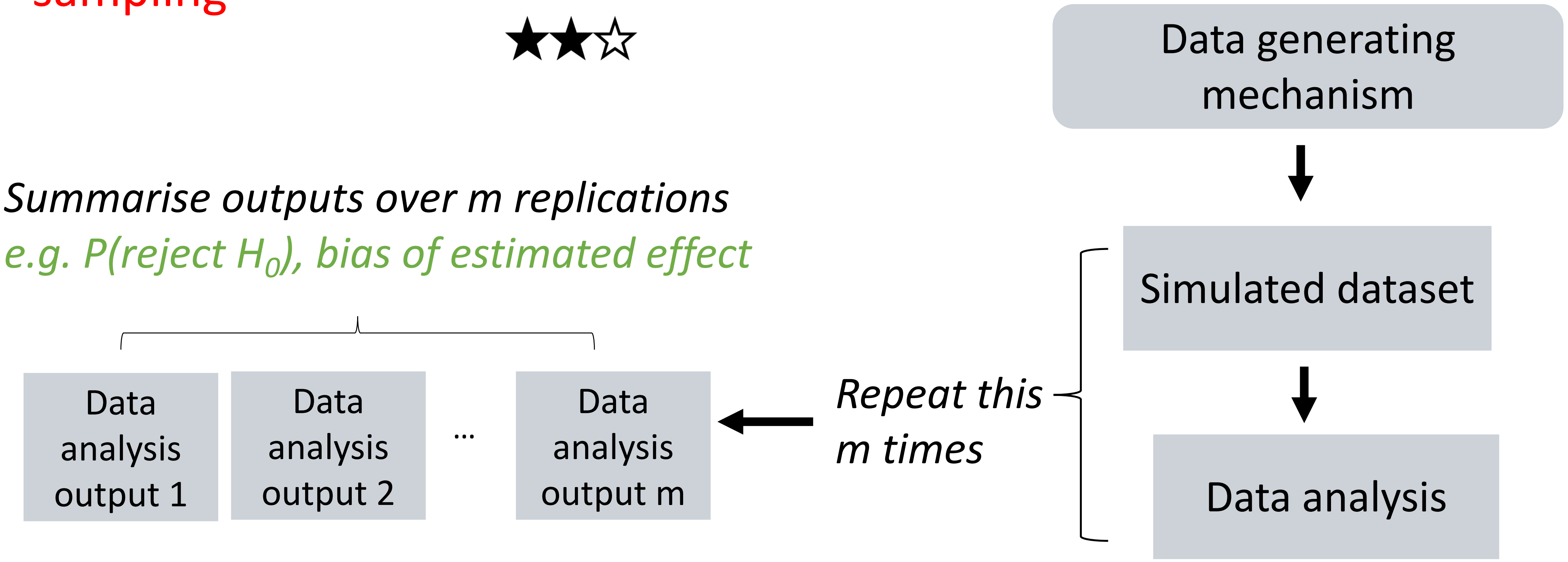
- Offer a preview of trial procedures and results before real-world implementation



Clinical trial simulation can…

- Evaluate the properties of a design and analysis strategy by repeated sampling

★★☆



Clinical trial simulation can…

- Identify the optimal design set-up when the whole process is repeated with different data generating mechanisms

Data generating mechanism

Simulated dataset

Data analysis

Repeat this m times

Data analysis output 1

Data analysis output 2

…

Data analysis output m

Summarise outputs over m replications
e.g. P(reject H_0), bias of estimated effect

Complex innovative clinical trial designs

Trial set-up with **unconventional design features**, where closed-form mathematical expressions for

- sample size calculation
- unbiased effect size estimation

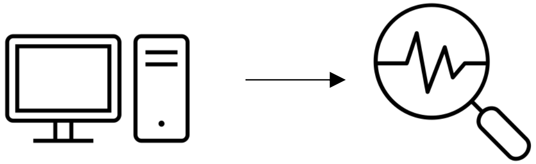
are **unavailable**

Complex innovative clinical trial designs

Trial set-up with **unconventional design features**, where closed-form mathematical expressions for

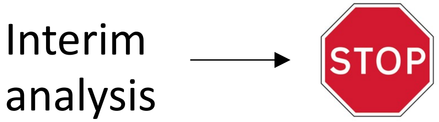
- sample size calculation
- unbiased effect size estimation

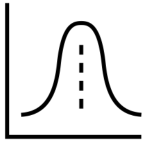
are **unavailable**



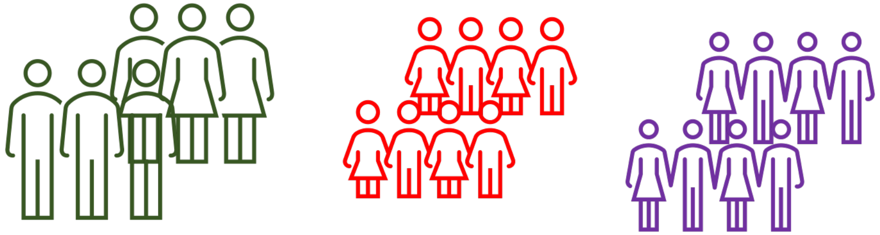
E.g., set-up consisting

- Rules that guide trial activities





- Complex rules for statistical decision making
- Multiple interventions and subpopulations



The OCTAVE framework

- **Objectives of conducting CTS**
- **Characteristics of underlying factors**
- **Trial set-up**
- **Analysis methods**
- Characteristics of design ← **Valuation** → Performance of analysis method
- Reporting ← **Evidence** → Reproducibility

The OCTAVE framework

Objectives of conducting CTS

Characteristics of underlying factors

Trial set-up

Define structure of data generating mechanism

Analysis methods

Characteristics of design

Valuation

Performance of analysis method

Reporting

Evidence

Reproducibility

Objectives of conducting CTS

- Underlying factors
- Trial set-up
- Analysis methods

Objectives of conducting CTS- example 1

- **Vary** — Underlying factors
- **Fixed** — Trial set-up
- **Fixed** — Analysis methods

Assess the sensitivity of **one** design and analysis strategy with respect to some underlying assumptions

**Evaluate the properties of a four-arm, two-stage design with a sequential t-test assuming different delay lengths in the endpoint*

* Wason JM, Brocklehurst P, Yap C. BMC medicine. 2019 Aug 2;17(1):152.

Objectives of conducting CTS- example 2

- Fixed — Underlying factors
- Vary — Trial set-up
- Fixed — Analysis methods

Compare the properties of **multiple** designs that use the **same** analysis methods under a given set of assumptions

******Investigate the performance of different randomization approaches in a platform trial with beta-binomial outcome model*

** Sirkis T, Jones B, Bowden J. BMC Medical Research Methodology. 2022 Aug 6;22(1):216.*

Underlying factors

- Related to trial design and/or conduct
- The *true* properties or characteristics are *unknown* to the investigators at the planning stage
- The properties or characteristics *cannot be altered* by human intervention.

Underlying factors

- Related to trial design and/or conduct
- The *true* properties or characteristics are *unknown* to the investigators at the planning stage
- The properties or characteristics *cannot be altered* by human intervention.

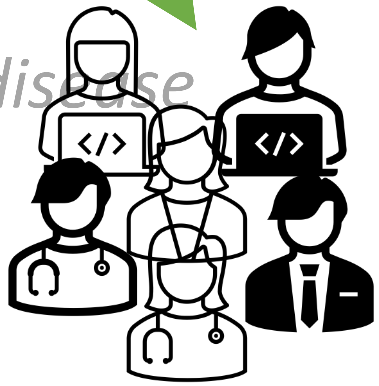
Examples: treatment effect of endpoint(s), variability in how participants respond to a treatment, prognostic covariates, disease progression, missing data types

Underlying factors

- Related to trial design and/or conduct
- The *true* properties or characteristics a
  investigators at the planning stage
- The properties or characteristics *cann* n
  intervention.

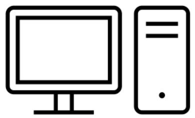
We can select the factors but we hardly know the true numerical representations

Examples: treatment effect of endpoint(s), variability in how participants respond to a treatment, prognostic covariates, disease progression, missing data types



Characteristics of underlying factors

Assumptions



Trial set-up

- A trial design is a structured plan that describes how a study will be conducted to address clinical research questions.

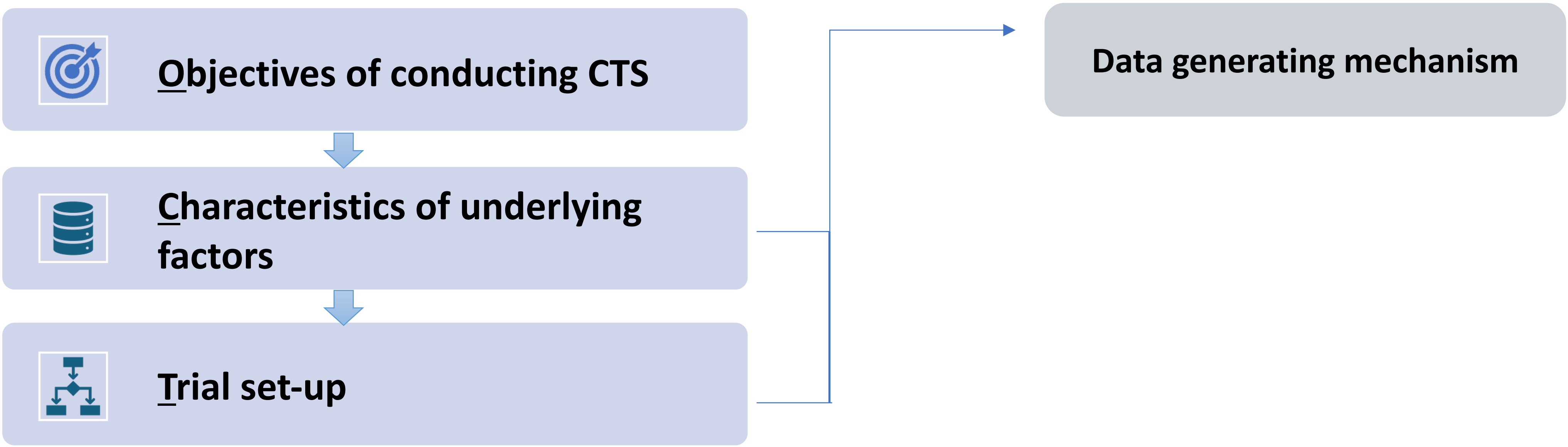
→ We refer to trial set-up as the statistical and numerical aspects of the plan

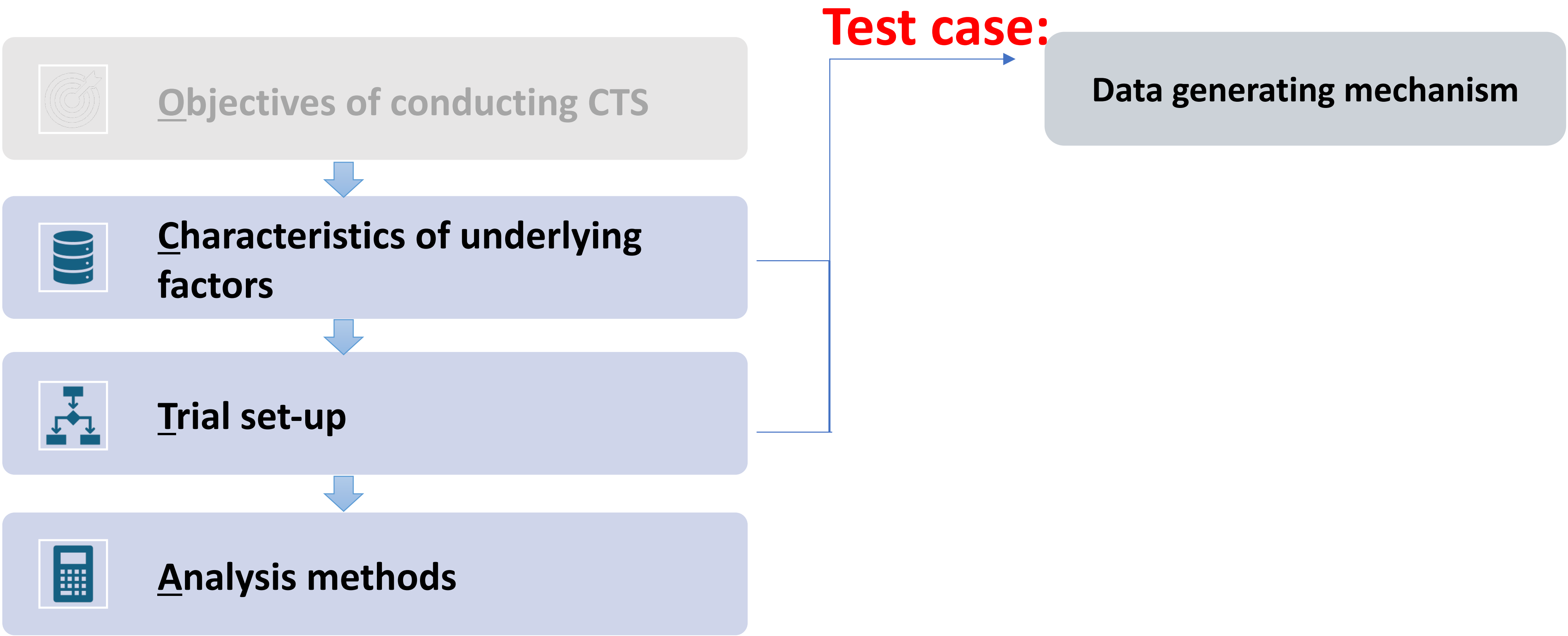
Trial set-up

- A trial design is a structured plan that describes how a study will be conducted to address clinical research questions.

→ We refer to trial set-up as the statistical and numerical aspects of the plan

Examples of trial set-up: number of interventions, randomization method, sample size per arm/per stage, number of interim analyses, rules for guiding trial activities and statistical decisions

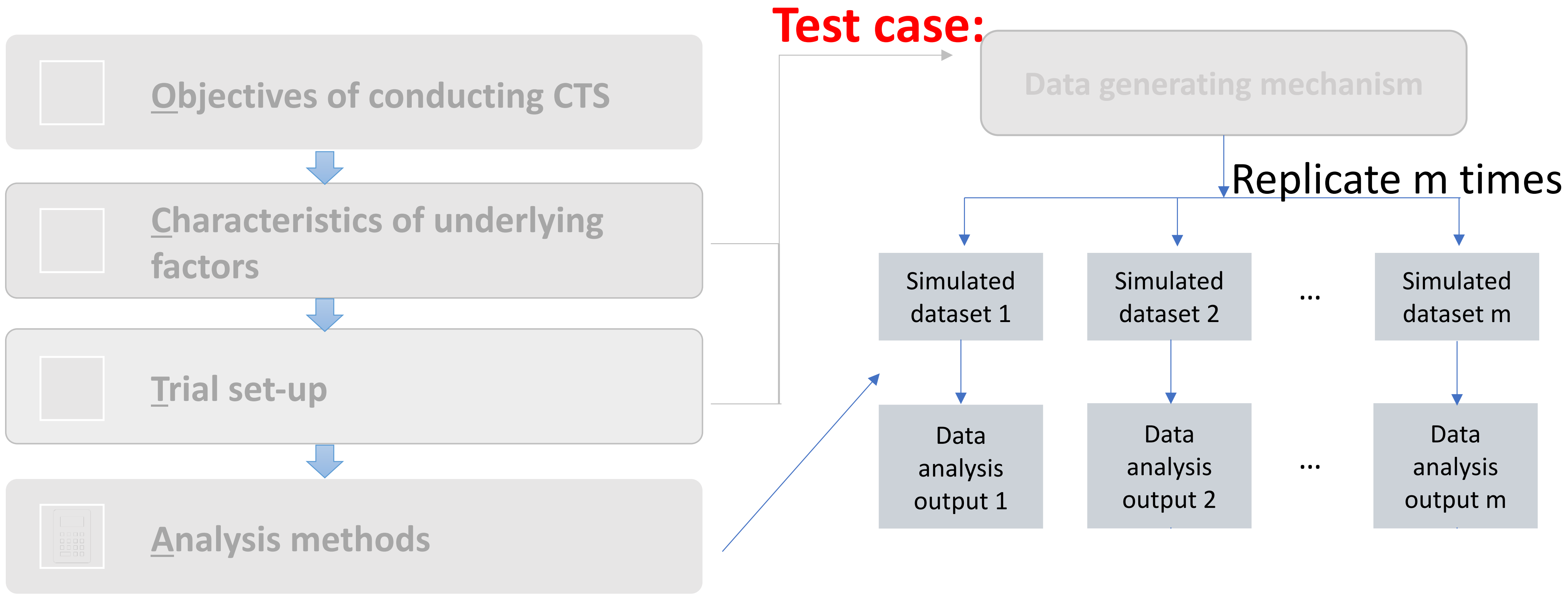

Objectives of conducting CTS
Characteristics of underlying factors
Trial set-up
Data generating mechanism



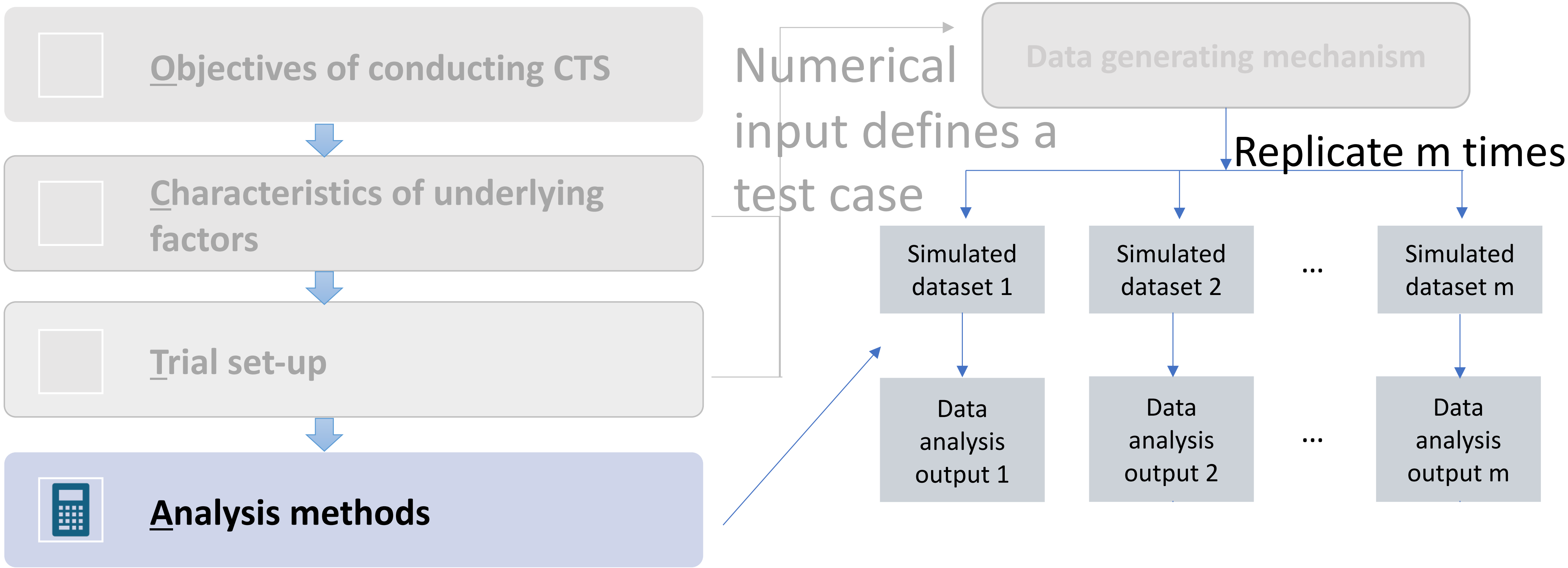
A **test case** comprises

- the underlying factors,
- the trial set-up, and
- specification of analysis model/ methods

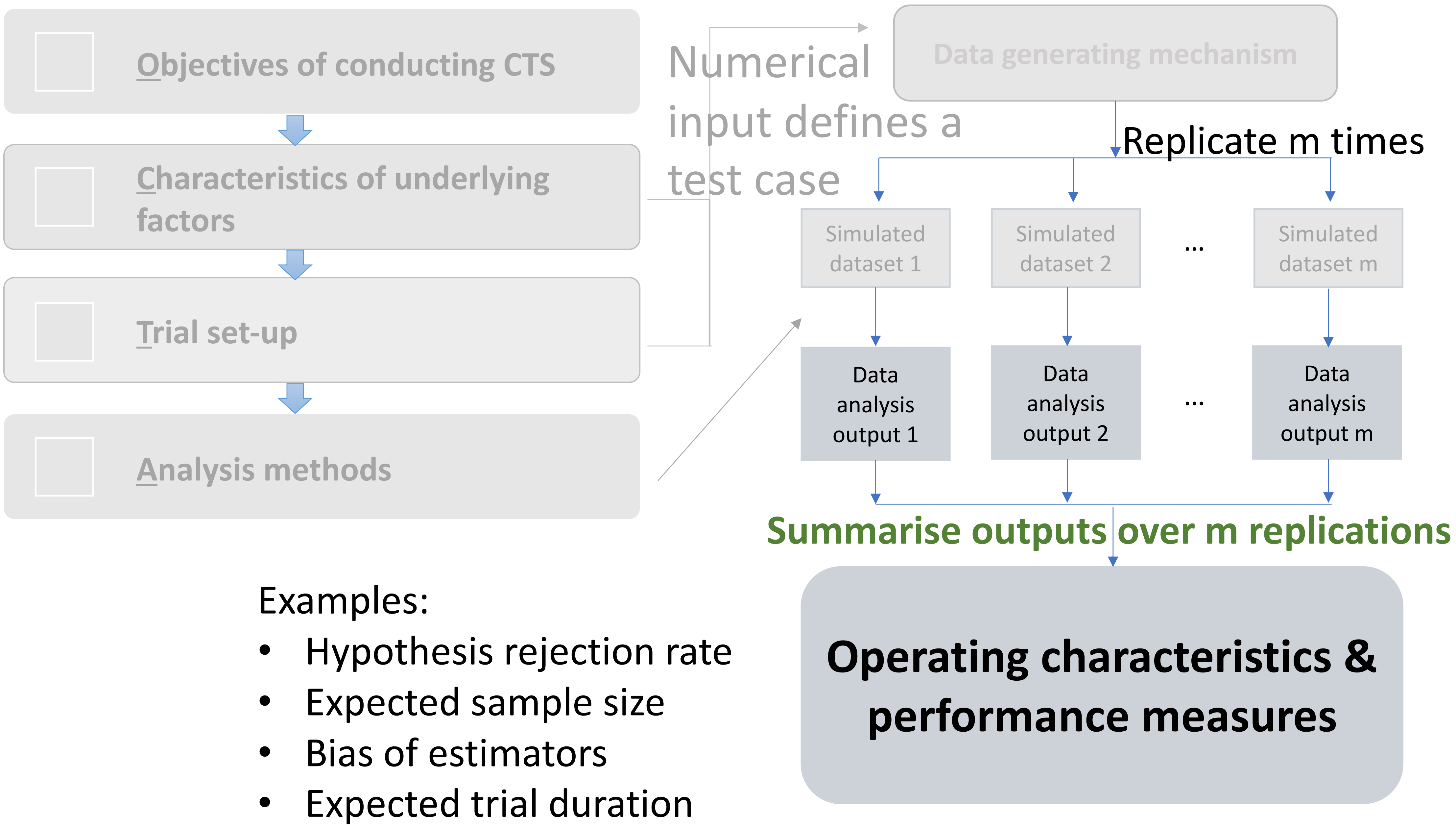
each set to concrete numerical inputs for this scenario

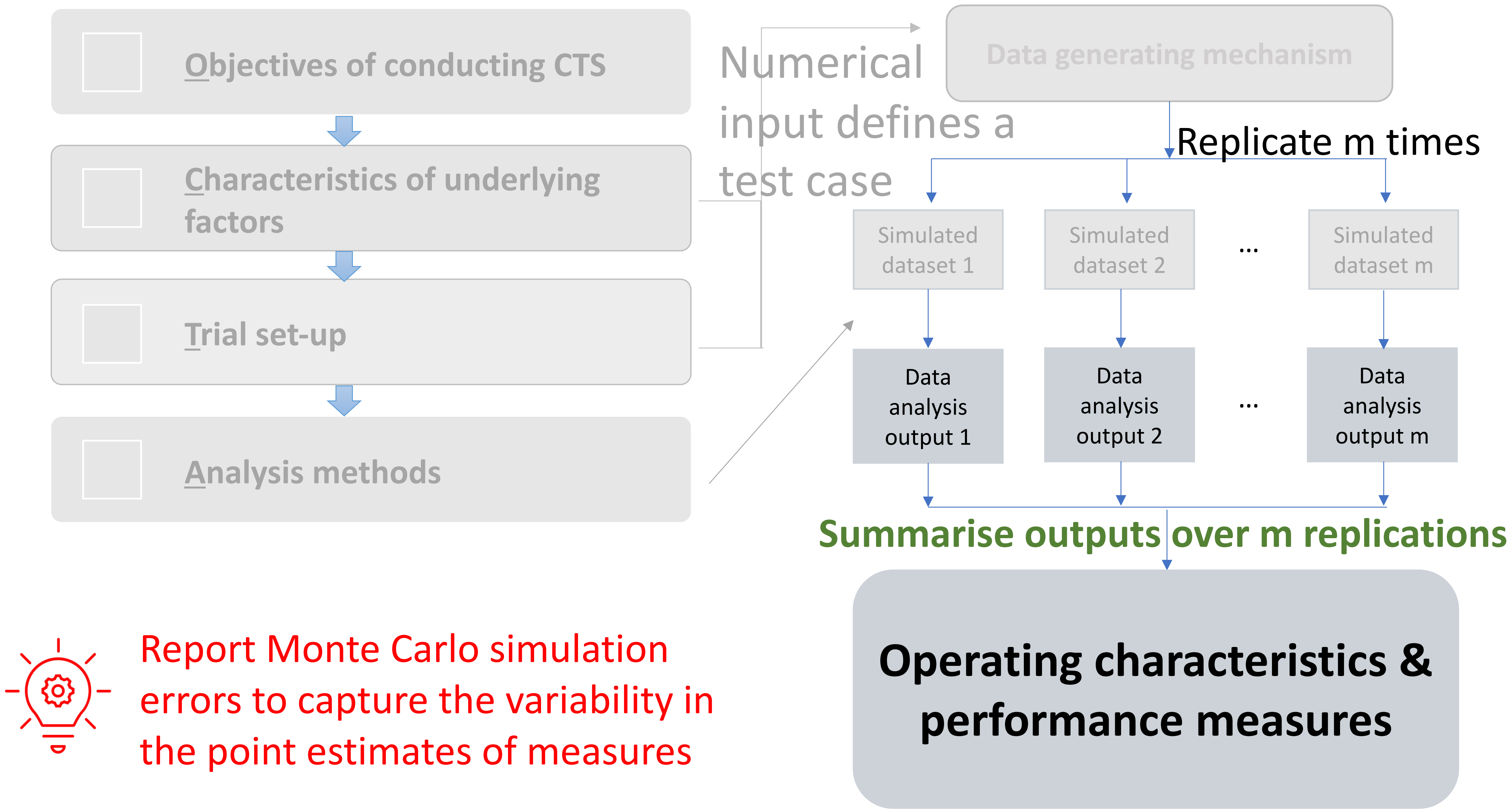


- A test case can be used to evaluate the replicated outputs *conditionally* given the *specified ancillary configuration*—without averaging across alternative configurations.

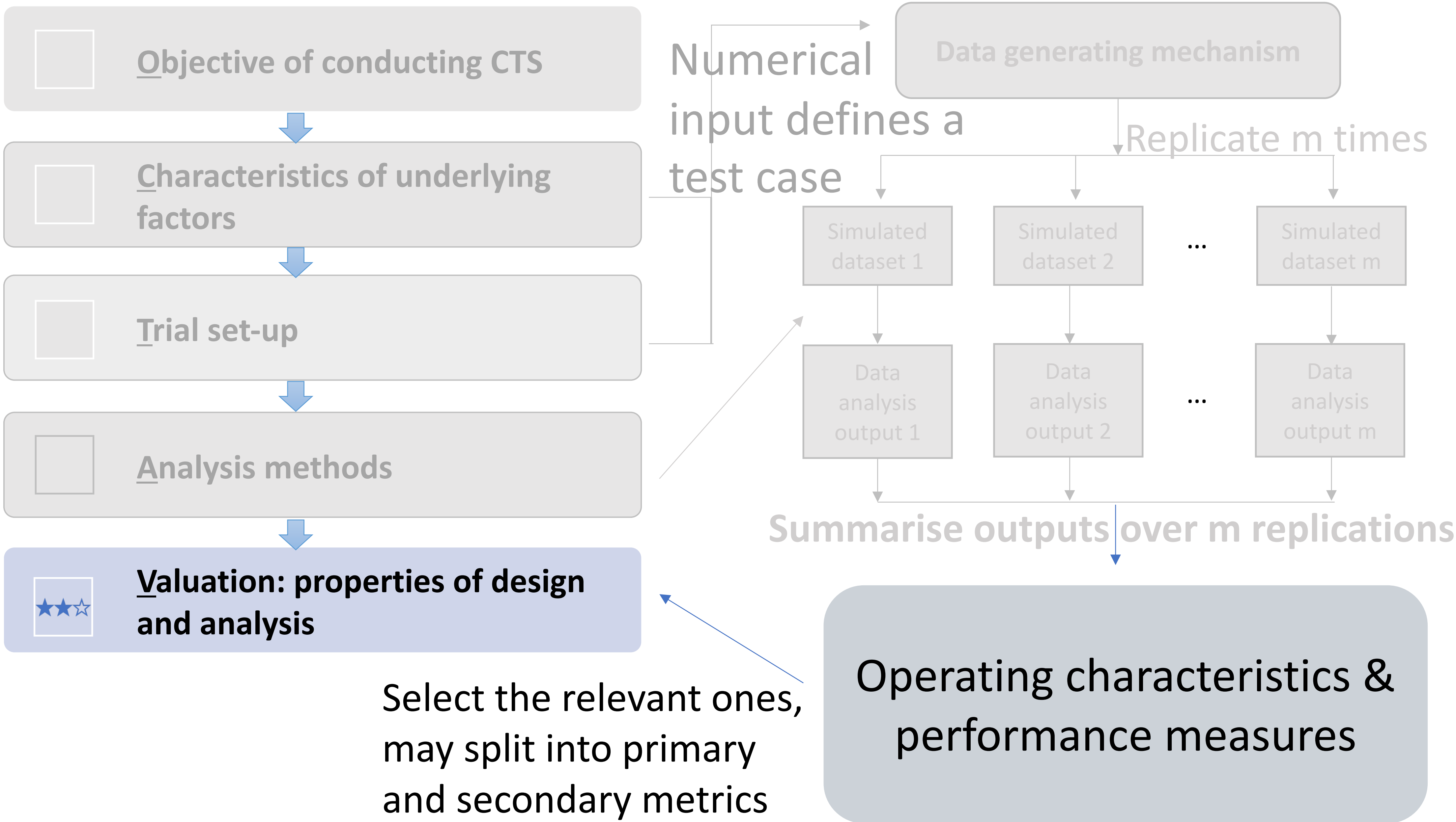


- Statistical techniques and procedures used to evaluate accumulating trial data.
- They rely on *underlying assumptions* regarding the data or the data generating process.


Objectives of conducting CTS
Characteristics of underlying factors
Trial set-up
Analysis methods
Numerical input defines a test case
Data generating mechanism
Replicate m times
Simulated dataset 1
Simulated dataset 2
...
Simulated dataset m
Data analysis output 1
Data analysis output 2
...
Data analysis output m
Summarise outputs over m replications
Operating characteristics & performance measures
Examples:
• Hypothesis rejection rate
• Expected sample size
• Bias of estimators
• Expected trial duration



Objectives of conducting CTS
Characteristics of underlying factors
Trial set-up
Analysis methods
Numerical input defines a test case
Data generating mechanism
Replicate m times
Simulated dataset 1
Simulated dataset 2
…
Simulated dataset m
Data analysis output 1
Data analysis output 2
…
Data analysis output m
Summarise outputs over m replications
Operating characteristics & performance measures
Report Monte Carlo simulation errors to capture the variability in the point estimates of measures


Objective of conducting CTS
Characteristics of underlying factors
Trial set-up
Analysis methods
Valuation: properties of design and analysis
Numerical input defines a test case
Data generating mechanism
Replicate m times
Simulated dataset 1
Simulated dataset 2
…
Simulated dataset m
Data analysis output 1
Data analysis output 2
…
Data analysis output m
Summarise outputs over m replications
Operating characteristics & performance measures
Select the relevant ones, may split into primary and secondary metrics

Example:

Is a particular Bayesian design worth implementing?

Primary metrics:

i) Predictive probability of success
ii) Expected sample size

Secondary metrics:

i) Frequentist type I error rate
ii) Mean squared error of estimated treatment effects

Select the relevant ones, may split into primary and secondary metrics

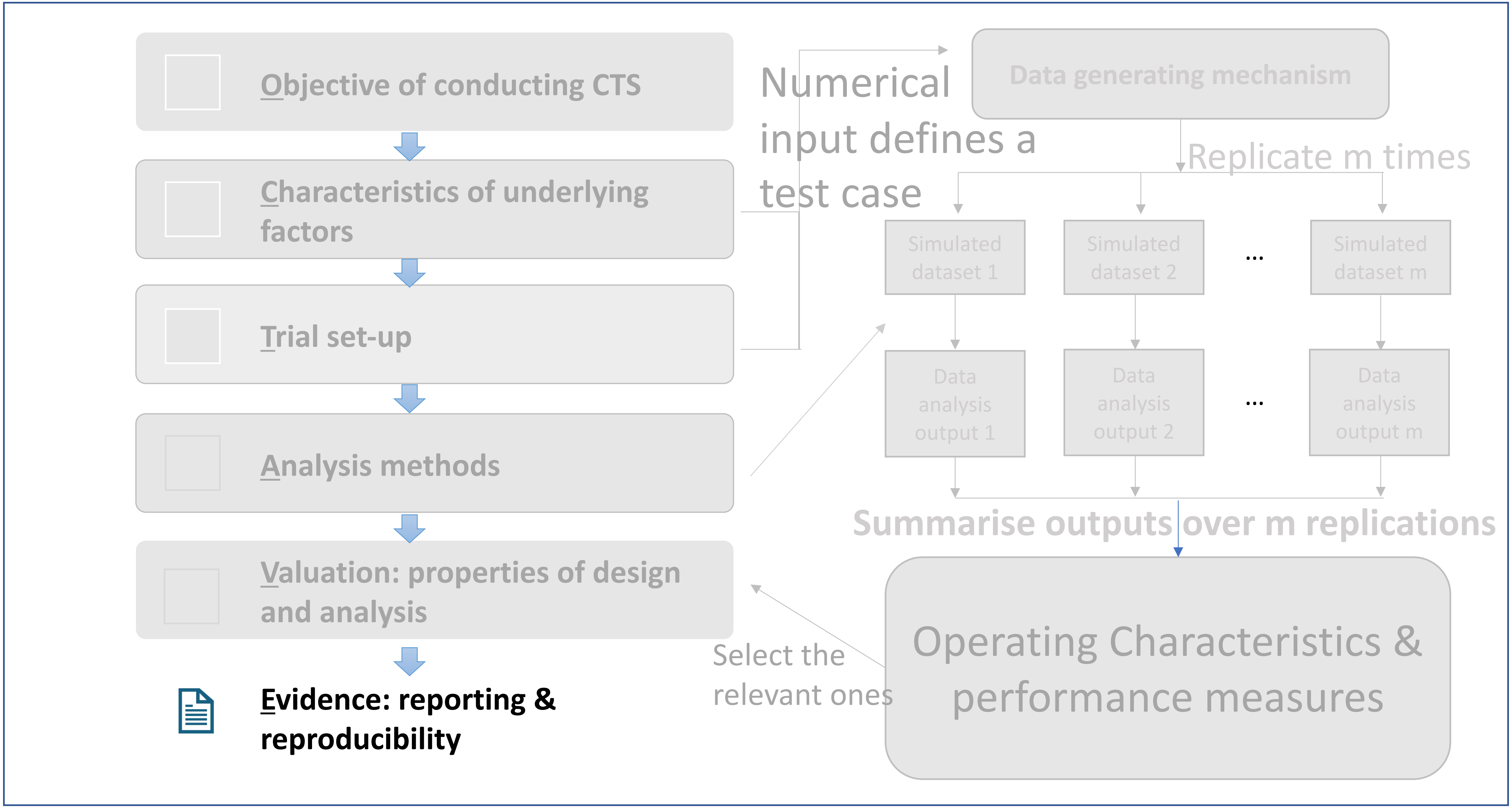
Operating characteristics & performance measures

Valuation of complex designs

Statistical benefits

Implementation complexity

- Operational complexity
  - E.g., the average time to execute interim analyses, staff hours required per participant
- Regulatory uncertainty
  - E.g., additional months to approval compared with conventional designs
- Lack of familiarity with novel designs
  - E.g., the number of staff training sessions needed before trial launch


Objective of conducting CTS
Characteristics of underlying factors
Trial set-up
Analysis methods
Valuation: properties of design and analysis
Evidence: reporting & reproducibility
Numerical input defines a test case
Data generating mechanism
Replicate m times
Simulated dataset 1
Simulated dataset 2
…
Simulated dataset m
Data analysis output 1
Data analysis output 2
…
Data analysis output m
Summarise outputs over m replications
Operating Characteristics & performance measures
Select the relevant ones

O
C
T
A
V
E

Clear documentation

- should explicitly state which components vary
- facilitates reproduction of the simulation results
- showing the output of single runs and repeated runs

  → confirms inclusion of the required analyses and summaries

  → help start the discussion and clarify what is happening

Objective of conducting CTS

Numerical

Data generating mechanism

relevant ones

performance measures

Evidence: reporting & reproducibility

Pseudocode for outlining simulation details

- Break the simulation process down into specific steps and decision-making algorithms
- Show the overall coding structure and flow

Identify logical flaws or missing decision points before coding

Algorithm 1 Pseudocode for the two-endpoint Bayesian two-stage design

for $j \in (1 : M_{\text{sim}})$ **do**
  $i \leftarrow j$
  Simulate accrual, outcomes and drop-out indicator for N_{max} subjects.
  $D_{\text{interim}} \leftarrow$ observed outcomes of N_{interim}
  # — *Interim analysis stage* —
  $PP_A \leftarrow P(\hat{\mu}_{\text{treat,A}} > \hat{\mu}_{\text{control,A}} \mid D_{\text{interim}})$
  $PP_B \leftarrow P(\hat{\mu}_{\text{treat,B}} > \hat{\mu}_{\text{control,B}} \mid D_{\text{interim}})$
  if $PP_A > \theta_{\text{eff}}$ **and** $PP_B > \theta_{\text{eff}}$ **then**
    Stop early $\leftarrow$ Efficacy
    Trial continues $\leftarrow$ FALSE
    $PP_{\text{primary}} \leftarrow PP_A$
    if $PP_{\text{primary}} > \theta_{\text{final}}$ **then**
      Trial outcome $\leftarrow$ Success
    else
      Trial outcome $\leftarrow$ Failure
    end if
    $i \leftarrow j + 1$
  else if $PP_A < \theta_{\text{fut}}$ **or** $PP_B < \theta_{\text{fut}}$ **then**
    Stop early $\leftarrow$ Futility
    Trial continues $\leftarrow$ FALSE
    $i \leftarrow j + 1$
  else
    Stop early $\leftarrow$ No
    Trial continues $\leftarrow$ True
  end if
  # — *Final analysis stage (if trial continues)* —
  if *Trial continues* is TRUE **then**
    $D_{\text{final}} \leftarrow$ Observed outcomes of N_{max}
    $PP_{\text{primary}} \leftarrow P(\hat{\mu}_{\text{treat,A}} > \hat{\mu}_{\text{control,A}} \mid D_{\text{final}})$
    if $PP_{\text{primary}} > \theta_{\text{final}}$ **then**
      Trial outcome $\leftarrow$ Success
    else
      Trial outcome $\leftarrow$ Failure
    end if
  end if
  $i \leftarrow j + 1$
end for
Calculate and return probability of success using *Trial Outcome*

Data generation approaches

Individual participant level

- Created dynamically or as a database in advance

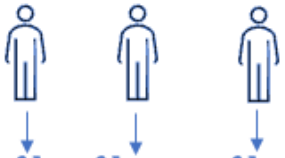
$y_{01}, y_{02}, \ldots, y_{0n} \sim N(\delta_0, \sigma^2)$ $y_{11}, y_{12}, \ldots, y_{1n} \sim N(\delta_1, \sigma^2)$

Data generation approaches

Individual participant level

- Created dynamically or as a database in advance
- Add complexity to code & computing storage
- Captures underlying correlation structures inherently

$y_{01}, y_{02}, \dots, y_{0n} \sim N(\delta_0, \sigma^2)$



$y_{11}, y_{12}, \dots, y_{1n} \sim N(\delta_1, \sigma^2)$

Data generation approaches

Individual participant level

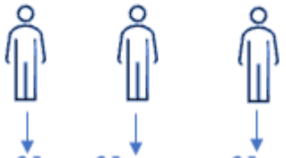
- Created dynamically or as a database in advance
- Add complexity to code & computing storage
- Captures underlying correlation structures inherently

VS

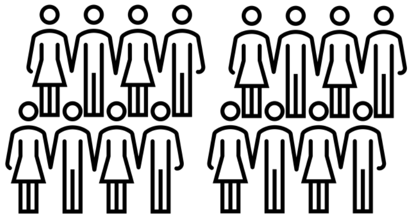
Aggregated level

- Explicit assumptions on data
- Less complex code, quicker to run & less storage
- Cannot explore aspects related to individual participants

$y_{01}, y_{02}, \dots, y_{0n} \sim N(\delta_0, \sigma^2)$



$y_{11}, y_{12}, \dots, y_{1n} \sim N(\delta_1, \sigma^2)$



$$Z \sim N(\theta, \frac{2\sigma^2}{n})$$

Validation of code/algorithm

- Unit testing to ensure each part functions correctly
  - E.g. test randomization & stopping rules in isolation
- Include checks that stop the code and provide informative error messages
- Statistical analysis validation
  - Existing methods: Compare to trusted software or known results
  - New methods: Hand-check results on small datasets or compare with similar approaches.
- Test extreme cases to assess the robustness of the algorithm to variations in input parameters

Validation of code/algorithm (cont.)

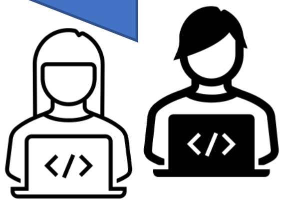
- Inspection of simulated data, e.g., check
  - simulated data reflect the intended treatment effects and trial parameters
  - interim analysis datasets are correctly extracted
  - trial progression reflects interim decisions
- Integration testing

  E.g., individual participant data generation → randomization → outcome generation → analysis according to decision rules

  For computation of performance measures

Approximation approaches

Useful when computing resources are limited



- Gaussian process and optimization approaches
  - ROSA approach[1]: choose an optimal set of test cases for sensitivity analysis with respect to the underlying assumptions
  - *adaptr* R package[2]: compute Bayesian multi-arm multi-stage designs with early stopping
- *Linear interpolation*
  - *gsbDesign* R package[3]: approximate operating characteristics of group-sequential Bayesian designs over a range of true treatment effects

[1] *Han et al. The American Statistician. 2024;78(1):76-87.*
[2] *Granholm et al. Journal of Open Source Software. 2022;7(72):4284.*
[3] *Gerber et al. Journal of Statistical Software. 2016;69:1-23.*

Graphical presentation and tools

- 2D plots
  - x-axis: varying components of CTS, y-axis: performance measures
  - Line curves connecting discrete points should be used cautiously
  - Truncation of the y-axis range may be employed to clarify trends, with all truncated values explicitly reported in the figure caption
  - Include confidence intervals computed using the Monte Carlo errors

Graphical presentation and tools

- 2D plots
  - x-axis: varying components of CTS, y-axis: performance measures
  - Line curves connecting discrete points should be used cautiously
  - Truncation of the y-axis range may be employed to clarify trends, with all truncated values explicitly reported in the figure caption
  - Include confidence intervals computed using the Monte Carlo errors
- Modern tools for automated visualizations e.g. AIRSHIP[1], INTEREST[2], simsum[3], and rsimsum[4] enable results to be displayed in a multi-dimensional view for identification of unique results (if there are any)

[1] *Meyer et al. SoftwareX. 2023;22:101347*
[2] *Gasparini et al. Journal of data science, statistics, and visualisation. 2021;1(4).*
[3] *White. The Stata Journal. 2010;10(3):369-85.*
[4] *Gasparini. Journal of Open Source Software. 2018;3(26):739.*

Thank you

- **Kim May Lee**
- kim.lee@kcl.ac.uk
- **Department of Biostatistics and Health Informatics, King's College London, UK**

Statistics in Medicine

TUTORIAL IN BIOSTATISTICS | Open Access

Clinical Trial Simulation: Planning With the OCTAVE Framework, Implementation and Validation Principles

Kim May Lee, Babak Choodari-Oskooei, Michael J. Grayling, Peter Jacko, Peter K. Kimani, Aritra Mukherjee, Philip Pallmann, Tom Parke, David S. Robertson, Ziyan Wang, Christina Yap, Thomas Jaki ... See fewer authors

First published: 16 March 2026 | https://doi.org/10.1002/sim.70449 | VIEW METRICS

KING'S College LONDON