

# Teaching Statistics with Simulations

Nicola Orsini

Karolinska Institutet

49th Winter Conference in Statistics  
Storhogna

March 11, 2026

- What is the main challenge?
- **DICE**
- How does it work?
- How to implement it?
- Final thoughts

# The challenge

- In many statistics courses students are given a dataset and asked to answer questions using statistical software.
- Students apply statistical methods to produce inferential results and write conclusions, even though the parameter value targeted by statistical inference is unknown.
- Pre-data elements such as the research question, study design, sample size justification, can be mentioned but often play a secondary role.

Therefore, teaching emphasizes **post-data analysis**, whereas in scientific investigations much of the reasoning occurs during **pre-data planning**.

# BIOMETRIKA.

---

## THE PROBABLE ERROR OF A MEAN.

By STUDENT.

### *Introduction.*

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

# Why simulations?

- To explore the consequences of a data-generating mechanism
- To illustrate the distinction between parameter values and sample estimates
- To connect statistical methods with study design and interpretation

## Viewpoint

# Rolling the DICE (Design, Interpret, Compute, Estimate): Interactive Learning of Biostatistics With Simulations

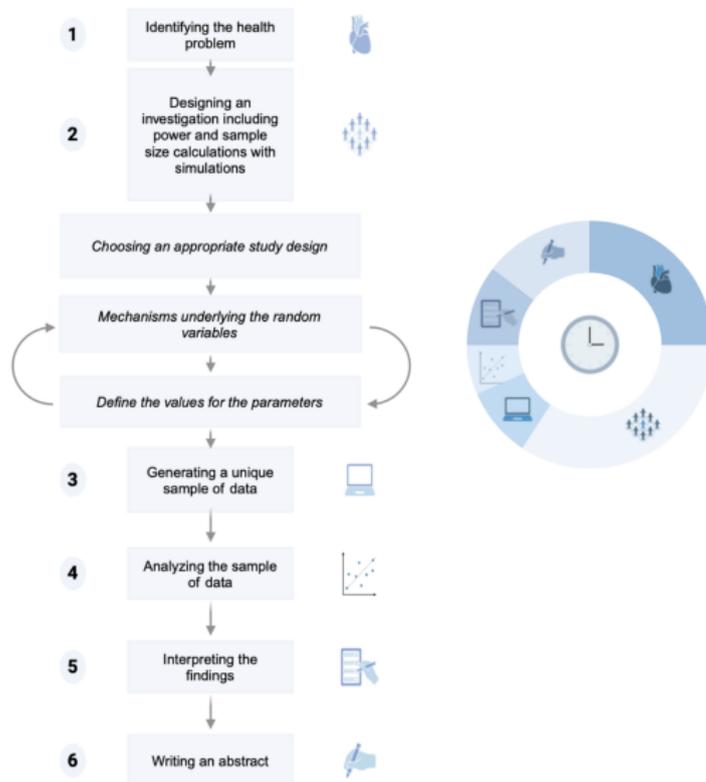
Robert Thiesmeier\*, MMSc; Nicola Orsini\*, PhD

Department of Global Public Health, Karolinska Institutet, Solna, Sweden

\*all authors contributed equally

- Journal Impact Factor: **12.6**
- Ranked **#1 of 86 journals** in “Education, Scientific Disciplines” by Clarivate Analytics.

# The DICE steps



# Implementation of DICE

- Prepare a tutorial with good motivating examples
- Explain the code needed to simulate one study
- Randomly assign students to small groups (4–5)
- Allow sufficient time for the group activity (2 hours)
- Set a deadline for submitting the abstract

## *The* NEW ENGLAND JOURNAL *of* MEDICINE

ESTABLISHED IN 1812

OCTOBER 8, 2020

VOL. 383 NO. 15

### Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure

#### ABSTRACT

##### **BACKGROUND**

Sodium–glucose cotransporter 2 (SGLT2) inhibitors reduce the risk of hospitalization for heart failure in patients regardless of the presence or absence of diabetes. More evidence is needed regarding the effects of these drugs in patients across the broad spectrum of heart failure, including those with a markedly reduced ejection fraction.

##### **METHODS**

In this double-blind trial, we randomly assigned 3730 patients with class II, III, or IV heart failure and an ejection fraction of 40% or less to receive empagliflozin (10 mg once daily) or placebo, in addition to recommended therapy. The primary outcome was a composite of cardiovascular death or hospitalization for worsening heart failure.

# What was the statistical plan?



## Trial Statistical Analysis Plan

**c12916255-05**

<b>BI Trial No.:</b>	1245.121
<b>Title:</b>	<p>A phase III randomised, double-blind trial to evaluate efficacy and safety of once daily empagliflozin 10 mg compared to placebo, in patients with chronic Heart Failure with reduced Ejection Fraction (HFrEF).</p> <p>EMPEROR-Reduced</p> <p>Including Protocol Amendment 1 &lt;1245.121&gt;-protocol-version-02 [c09098452-02], Protocol Amendment 2 &lt;1245.121&gt;-protocol-version-03 [c09098452-03], and Protocol Amendment 3 &lt;1245.121&gt;-protocol-version-04 [c09098452-04]</p>
<b>Investigational Product:</b>	Empagliflozin, BI 10773

# From statistical plan to data-generating mechanism

## Data-generating mechanism

$$x \sim \text{Bernoulli}(0.5)$$

$$\lambda = e^{\beta_0 + \beta_1 x}$$

$$U \sim U(0, 1)$$

$$T = -\frac{\log(U)}{\lambda}$$

$$D = I(T < 2)$$

$$T^* = \min(T, 2)$$

Observed data:  $(T^*, D, x)$

## R implementation

```
n <- 3360
x <- rbinom(n, 1, 0.5)

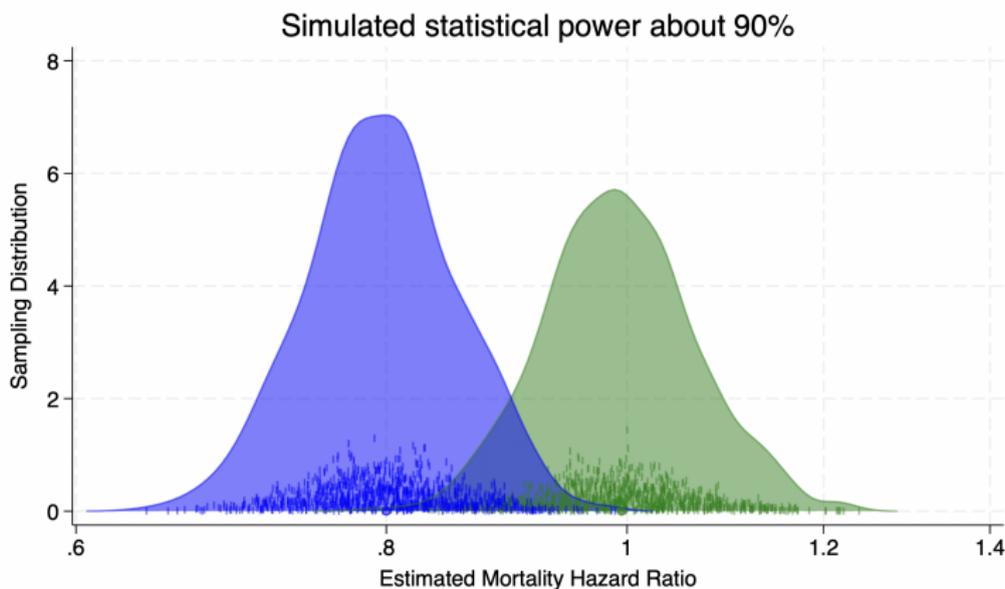
rate0 <- -log(0.85)/1
b0 <- log(rate0)
b1 <- log(0.80)

time <- -log(runif(n)) /
        exp(b0+b1*x)

follow_up <- 2
D <- (time < follow_up)

time[time > follow_up] <- follow_up
```

1,000 replications of trial ( $n=3,630$  and expected primary outcomes = 850)



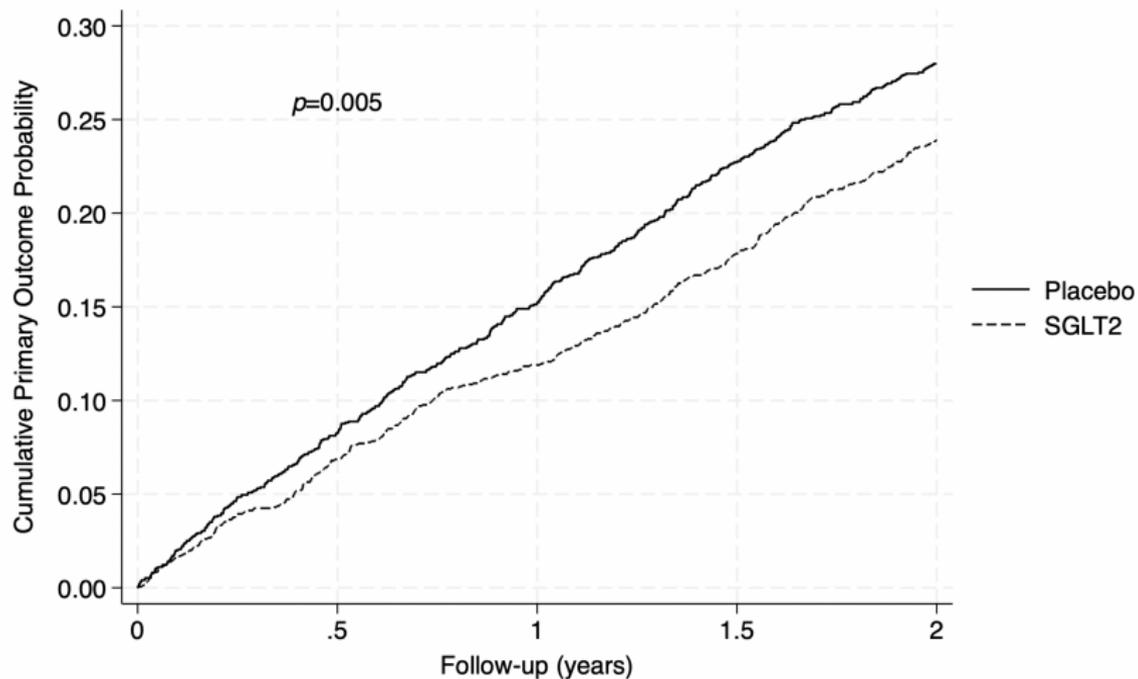
- Blue: sampling distribution under the alternative  $HR = 0.80$
- Green: sampling distribution under the null  $HR = 1$

# A unique trial

- In practice a scientific investigation observe only **one realization**
- The observed trial is one draw from the data-generating mechanism
- Fixing a seed (i.e. YMD) reproduces the same dataset.

```
set.seed(20230615)
```

# Data analysis: Kaplan-Meier curves



Parameter	Estimate	Std. error	$z$	$p$
$\beta_1$ (Treatment)	-0.192	0.068	-2.82	0.005

$$HR = e^{\hat{\beta}_1} = e^{-0.192} \approx 0.83$$

**Interpretation:** Compared with placebo, patients allocated to the SGLT2 inhibitor had an estimated 17% lower primary event rate (HR = 0.83, 95% CI 0.72, 0.94). A Wald-type two-sided test indicates strong incompatibility between this trial and the hypothesis of no treatment effect ( $z = -2.82$ ,  $p = 0.005$ ).

# Write a structured Abstract

## Abstract

**Title:** Effect of SGLT2 inhibitors on a composite outcome among patients with heart failure: a large randomized trial.

**Background:** Little is known about the effect of SGLT2 inhibitors on patients with known heart failure with reduced ejection fraction. The aim of this investigation was to examine the effect of an SGLT2 inhibitor on cardiovascular death or hospitalization for worsening heart failure.

**Methods:** A randomized trial with patients assigned to receive an SGLT2 inhibitor (10 mg once daily) or placebo. The primary outcome was time from treatment randomization until cardiovascular death or hospitalization or end of follow-up (2 years), whichever came first. A total of 3,360 adults with chronic heart failure were eligible to participate in the trial providing a power of 90% to detect a hazard ratio of 0.80.

**Results:** Compared with the placebo group, the primary event rate in the SGLT2 group was estimated to be 17% lower (Hazard Ratio = 0.83, 95% confidence interval [CI] 0.72 to 0.94;  $z = -2.82$ ,  $p = 0.005$ ).

**Conclusion:** In patients with heart failure with reduced ejection fraction, treatment with SGLT2 inhibitors reduced the rate of cardiovascular death or hospitalization for heart failure and supports their use as part of comprehensive therapy.

## Another trial from the same mechanism

Parameter	Estimate	Std. error	$z$	$p$
$\beta_1$ (Treatment)	-0.127	0.066	-1.92	0.055

$$HR = e^{\hat{\beta}_1} = e^{-0.127} \approx 0.88$$

**Interpretation:** Compared with placebo, patients allocated to the SGLT2 inhibitor had an estimated 12% lower primary event rate (HR = 0.88, 95% CI 0.77, 1.00). A Wald-type two-sided test **fails to reject the null hypothesis of no treatment effect** ( $z = -1.92$ ,  $p = 0.055$ ).

A well-designed and conducted trial does not guarantee statistical significance.

## Example: Observational study with a binary outcome

Consider three variables

- $Z$  : old age
- $X$  : treatment
- $Y$  : mortality

Older individuals are more likely to receive the treatment, and age also increases the risk of mortality regardless of treatment.



How to disentangle the effect of the treatment from the effect of age?

# Data-generating mechanism producing confounding

## Counfounder model

$$Z \sim \text{Bernoulli}(0.4)$$

## Treatment model

$$P(X = 1|Z) = \text{logit}^{-1}(\alpha_0 + \alpha_1 Z)$$

## Outcome model

$$P(Y = 1|X, Z) = \text{logit}^{-1}(\beta_0 + \beta_1 X + \beta_2 Z)$$

## R implementation

```
n <- ?
```

```
Z <- rbinom(n,1,0.4)
```

```
pX <- plogis(-1 + log(2)*Z)
```

```
X <- rbinom(n,1,pX)
```

```
pY <- plogis(-3 + log(0.8)*X + log(7)*Z)
```

```
Y <- rbinom(n,1,pY)
```

# Final thoughts

- **DICE** is a group-based activity for postgraduate health science students that covers the full cycle of a scientific investigation within a short time window.
- It integrates traditional lectures with individual and group-based exercises.
- It encourages students to stop treating data as simply given.
- It promotes coherent reasoning from research question to interpretation.
- It helps students recognize common misinterpretations of inferential results.
- It enables the generation of an unlimited number of teaching datasets.



- Orsini, N., Thiesmeier, R., & Båge, K. (2024). A Simulation-Based Approach to Teach Interaction Effects in Postgraduate Biostatistics Courses. *Journal of Statistics and Data Science Education*, 32(4), 395–404.
- Thiesmeier, R., Orsini, N. (2024). Rolling the DICE (Design, Interpret, Compute, Estimate): Interactive Learning of Biostatistics With Simulations. *JMIR Medical Education*, 10:e52679.
- Thiesmeier, R., Orsini, N. (2024). Teaching Statistics in Health Sciences: The Potential of Simulations in Public Health. *CHANCE*, 37(2), 34–39.