

Practical: Randomization inference

Qingyuan Zhao (Statistical Laboratory)*

July 7, 2026

Randomized controlled trials (RCTs) are widely regarded as the "gold standard" of establishing causality. The often forgotten component of the RCTs is that they can be objectively analyzed by randomization test. Haines and coauthors investigated the impact of disinvestment from weekend allied health services. We will use their dataset to explore the concept of randomization in the design and analysis of an experiment.

While you answer question 1, you may want to open an R session and install the following packages: `install.packages(c("readxl", "combinat"))`. (Raise your hand if you have trouble installing `readxl`.)

1. Skim through the abstract and read the section called "Design" in their article. Then answer the following questions: What is the name of the design of the experiment in this study? How was it carried out?
2. Download the patient-level data, then run the following code in R.

```
data <- readxl::read_excel("pmed.1002412.s002.xlsx")
data <- subset(data, hospital == "Dandenong" & study1 == 1)
```

The second line selects all patients involved in the first trial in the Dandenong hospital.

3. Unfortunately, this published dataset is not very well annotated. The columns `index_ward` and `sw_step` contain the identifiers for hospital ward and time step (in calendar month), respectively. In which order do you think the wards crossed over to no weekend health services? You may find the following R code useful.

```
table(data[, c("index_ward", "sw_step", "no_we_exposure")])
```

If you are a bit lost, the answer to this question is:

```
cross_over_realized <- c(3, 5, 7, 6, 2, 4)
```

Execute it in R, then use the following code to define the exposure and outcome of interest ("los" is short for length of stay).

```
data$exposure_status <- as.numeric(data$sw_step >= cross_over_realized[data$index_ward])
data$log_acute_los <- log(data$acute_los)
```

4. [Group discussion] Execute the following code in your R session. What do you observe about the two interval estimators of the treatment effect (of no weekend health services on log length of stay)?

```
confint(lm(log_acute_los ~ exposure_status, data))
confint(lm(log_acute_los ~ exposure_status + as.factor(index_ward), data))
```

5. [Group discussion] Next, we explore the randomization analysis of this dataset. Let $Y_i(0)$ denote the potential outcome of patient i when there is weekend health services, and $Y_i(1)$ denote the potential outcome when there is no weekend health services. The null hypothesis is that $Y_i(0) = Y_i(1)$ for all i . In defining potential outcome this way, what exposure function did we use? Is the exposure the same as the treatment being randomized in the experiment (crossover order)? When is this not a reasonable assumption?

*qyzhao@statslab.cam.ac.uk

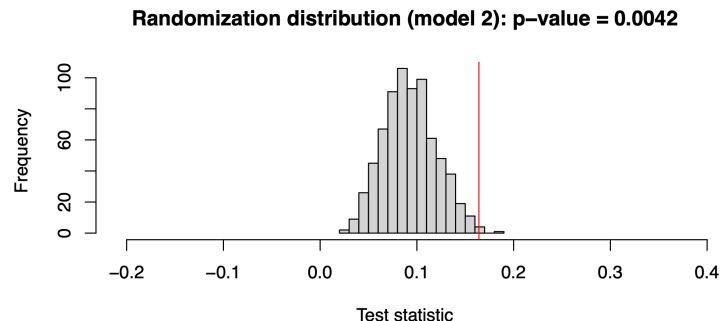
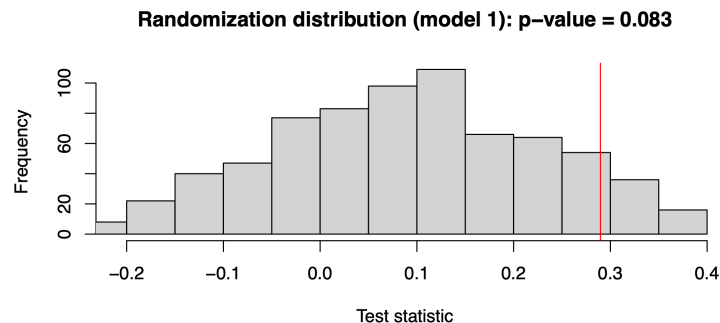
6. Read the following code, then execute it in your R session (it may take 30 seconds to finish). The expected output is included for your reference.

```
get_statistic <- function(index_ward,
                           sw_step,
                           log_acute_los,
                           cross_over) {
  exposure_status <- (sw_step >= cross_over[index_ward])
  c(lm(log_acute_los ~ exposure_status)$coef[2],
    lm(log_acute_los ~ exposure_status + as.factor(index_ward))$coef[2])
}

T_obs <- get_statistic(data$index_ward, data$sw_step,
                      data$log_acute_los, cross_over_realized)

T_random <- sapply(combinat::permn(2:7),
                  get_statistic,
                  index_ward = data$index_ward,
                  sw_step = data$sw_step,
                  log_acute_los = data$log_acute_los)

par(mfrow = c(2, 1))
for (m in 1:2) {
  hist(T_random[m, ], 20,
       main = paste0("Randomization distribution (model ", m, "): ",
                     "p-value = ", signif(mean(T_random[m, ] >= T_obs[m]), 2)),
       xlab = "Test statistic", xlim = range(T_random))
  abline(v = T_obs[m], col = "red")
}
```



7. [Group discussion] Explain to each other what the code above does and discuss the results. Here are a few questions you may want to think about:
- How do the two randomization tests compare with each other?
 - How do the randomization tests compare with the results from the corresponding normal linear models in question 4?
 - The randomization distribution of the second test statistic is clearly not centered at 0. Why?
 - How can you "invert" the randomization tests to obtain an interval estimator of the treatment effect?