Day 4, Practical 1, Hely's solution

Helene Charlotte Wiese Rytgaard

May 16, 2023

1. What is the overall research question considered in the paper?

They wish to investigate the causal effect of different levels of nutritional support on a clinical outcome.
= the probability of being discharged alive from the pediatric intensive care unit (PICU) by a given day under different feeding regimes.

2. What are the observed variables? How many longitudinal time-points are there and how are they defined?

Observed data O collected over 7 days of follow-up (each day constituting a time-point):

$$O = (Z_0, A_0, Y_1, M_1, Z_1, A_1, \dots, Y_T, M_T, Z_T, A_T, Y_{T+1}).$$

The ordering of O represent the assumed causal ordering among variables.
At each time-point/day t:

- $A_t \in \{0, 1\}$ represents the patient's feeding status.
  - "fed"/"not fed" on a giving day ("fed" = reaching a specific threshold of caloric intake).
- Z_t collects covariates/confounders. (Z_0 includes both baseline covariates and the first measurement of time-varying covariates).
  - baseline: age, sex, weight, height, RACHS-1 risk score, ...
  - follow-up: mechanical ventilation, renal replacement therapy, infection, ...
- $M_t \in \{0, 1\}$ indicate whether a patient has died by the end of day t.
- $Y_t \in \{0, 1\}$ indicate whether a patient has been discharged alive by the end of day t.

(After death/discharge, all variables are deterministically set to their last observed values).

3. Which variables affect each other? Are there any time-dependent confounders?

"Younger children and those with higher risk scores tends to be fed less aggressively and are also more likely to stay longer in the PICU [...]"

→ baseline confounders: age, RACHS-1 risk score.

"While all patients are mechanically ventilated at baseline, being taken off mechanical ventilation is a strong predictor of discharge in the next few days. Patients just taken off mechanical ventilation can, for safety reasons, be fed only by the parenteral (not the enteral) route, making it less likely that their caloric intake will reach the 20% threshold."

$\rightarrow$ time-varying confounder: mechanical ventilation.

Lack of adjustment for baseline confounders would bias effect of feeding towards a (more) beneficial effect. Lack of adjustment for time-varying confounder would bias effect of no feeding towards a (more) beneficial effect.

4. What are the hypothetical regimes that are considered? Static regimes? Dynamic? What are the corresponding causal parameters?

A longitudinal feeding regime is denoted $\bar{a} = (a_0, a_1, \dots, a_T)$, defining a feeding intervention for each day t before the end of follow-up where the patient is alive and not yet discharged. Note that:

- a *static regime* specifies a constant for each t;
- a *dynamic regime* specifies a decision rule for each t.

Different *static regimes* are considered:

- "Never feed" specifying $a = (0, 0, \dots, 0)$.
- "Always feed" specifying $a = (1, 1, \dots, 1)$.
- "Feed from day 3" specifying $a = (0, 0, 1 \dots, 1)$.
- "Feed by day 3" specifying $a = (A_0, A_1, 1 \dots, 1)$ (i.e., leaving A_0, A_1 as observed).

Dynamic regime:

- "Feed patient on each day they are not mechanically ventilated". (Recall, patients just taken off mechanical ventilation will be less likely to reach the threshold of caloric intake required to be "fed").

Corresponding *counterfactual outcomes*: Any treatment strategy is denoted d, and the outcome that would had been observed at time t under a strategy d is denoted Y_t^d.

The corresponding *causal parameter* of interest is the intervention-specific mean outcome, i.e., the expected discharge status by a selected time $t^* \in \{1, \dots, T+1\}$, under a strategy d:

$$\Psi_{d,t^*}(P) = \mathbb{E}_P[Y_{t^*}^d];$$

this is the absolute risk of discharge by day t^* if all subjects had followed the strategy d.

5. Go over the stated causal assumptions required to identify the causal parameters.
6. Sequential randomization assumption = *exchangeability*:

$$Y_{t^*}^d \perp\!\!\!\perp A_t \mid \bar{Z}_t, \bar{A}_{t-1} = d_t, \quad \text{for, } t \in \{0, 1 \dots, t^* - 1\} \text{ and strategy } d.$$

i.e., all covariates predictive of both feeding strategy and discharge by time t^* must be measured (and adjusted for).

- *Positivity*:

$$P(A_t = d_t \mid \bar{Z}_t, \bar{A}_{t-1} = d_{t-1}) > 0.$$

Patients must have a positive probability of following considered strategy d. Note that the probability of following a given strategy is quite different for each strategy. However, from the results section it seems that positivity was not an issue for any of the strategies (second column, p. 1375).

6. Results: The authors consider a 'naive' approach, an IPW estimator, a g-formula estimator and a TMLE estimator. What results do they find with each method? How do they differ in terms of bias and variance?

7. The 'naive estimates' (unadjusted absolute probability of discharge, calculated among those following the strategy) show a higher probability of being discharged if "never fed" compared to "fed by day 3", i.e., a beneficial effect of less caloric intake.

8. IPW, g-formula estimation and TMLE all reduces the probability of being discharged if "never fed" and by that the difference between "never fed" and "fed by day 3".

9. Wider confidence intervals for IPW than TMLE, as we would expect.

10. See also Figure 2: Very wide confidence intervals for the strategy "feed from day 1". This is directly linked to the lower probability of following this regime.