Choice of the estimand in presence of intercurrent events

Brice Ozenne

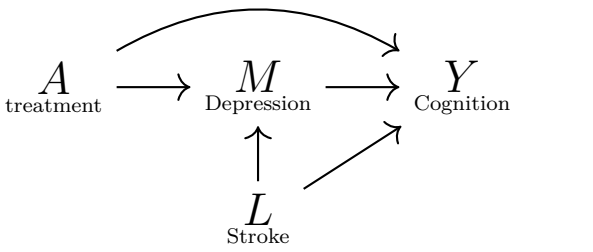
October 2, 2025

1 Setting

Random variables:

- A treatment (active/control)
- L common cause of depression and cognition, e.g. gene, stroke.
- M depression status (yes/no)
- Y cognitive score

Working DAG
(directed acyclic graph)



2 Potential outcome notation

- $Y^{a=1}$ cognitive score for a subject from the study population had she received the active treatment
- $Y^{a=0}$ cognitive score for a subject from the study population had she received the control treatment

The cognitive score had a subject from the active group received the active treatment is the observed cognitive score: $Y|A{=}1 = Y^{a=1}$. Similarly in the control group.

- $Y^{a=1,m=0}$ cognitive score for a subject from the study population had she received the active treatment and not experienced depression
- $Y^{a=1} \mid M^{a=1} = 1$ cognitive score for a subject from a subset of the study population, composed of those who would have experienced depression had they received the active treatment, had she received the active treatment.

3 Estimands

Among depressed strategy: compare, among those who are observed to experience depression, the expected cognitive score between the active and control group.

✔ Well defined sub-populations

⚠ prone to selection bias since a different subset of patients is used in each treatment arm, e.g., if the treatment is protective more frail people will be in the control group (as the corresponding patients in the active group did not experienced depression thanks to the treatment), leading to an unfair comparison w.r.t. cognition.

Formally: $\Psi_{\text{on}} = \mathbb{E}\left[Y^{a=1}|M^{a=1}=1\right] - \mathbb{E}\left[Y^{a=0}|M^{a=0}=1\right]$

A variant of this strategy compare the cognition score at end of follow-up or start of depression, which-ever comes first. This is not recommended as it mixes early and late direct effect in a proportion related to the indirect effect.

Principal stratum strategy: similar to among depressed but uses the same subset in each treatment arm to avoid selection bias. There are several principal stratum that can be used:

- $\Psi_{\text{always}} = \mathbb{E}\left[Y^{a=1}|M^{a=1}=M^{a=0}=1\right] - \mathbb{E}\left[Y^{a=0}|M^{a=1}=M^{a=0}=1\right]$: comparing cognition among subjects that would have experienced depression regardless to the treatment.
  ✔ Corresponds to a direct treatment effect on cognition.
  ⚠ Hypothetical sub-population that requires additional statistical assumptions to be identified.

- $\Psi_{\text{always if active}} = \mathbb{E}\left[Y^{a=1}|M^{a=1}=1\right] - \mathbb{E}\left[Y^{a=0}|M^{a=1}=1\right]$: comparing cognition among subjects that would have experienced depression under the active treatment.
  ✔ Well defined sub-population (depressed patients in the active group)
  ⚠ Mixes direct and indirect treatment effect since the control treatment may affects cognition through depression.
  A similar estimand can be defined under the control treatment.

$$
\begin{aligned}
&\mathbb{E}\left[Y^{a=1}|M^{a=1}=1\right] - \mathbb{E}\left[Y^{a=0}|M^{a=1}=1\right] = \mathbb{E}\left[Y^{1,M(1)}|M(1)=1\right] - \mathbb{E}\left[Y^{0,M(0)}|M(1)=1\right] \\
&= \mathbb{E}\left[Y^{1,M(1)}|M(1)=1\right] - \mathbb{E}\left[Y^{0,M(1)}|M(1)=1\right] + \mathbb{E}\left[Y^{0,M(1)}|M(1)=1\right] - \mathbb{E}\left[Y^{0,M(0)}|M(1)=1\right] \\
&= \underbrace{\mathbb{E}\left[Y^{1,1}|M(1)=1\right] - \mathbb{E}\left[Y^{0,1}|M(1)=1\right]}_{\text{direct effect}} + \underbrace{\mathbb{E}\left[Y^{0,1}|M(1)=1\right] - \mathbb{E}\left[Y^{0,M(0)}|M(1)=1\right]}_{\text{indirect effect (non-0 when M(0) } \neq \text{ M(1))}}
\end{aligned}
$$

Hypothetical strategy: comparing the cognition in the entire study population, active vs. control, had all patients experienced depression.

✔ Well defined population

⚠ Rely on modeling assumptions, e.g., the average cognition of those who did not experience depression had they had experienced depression is the same as those who did experience depression and belong to the same treatment group/covariates.

Formally: $\Psi_{\text{hypo}} = \mathbb{E}\left[Y^{a=1,m=1}\right] - \mathbb{E}\left[Y^{a=0,m=1}\right]$

Mediation strategy: comparing the cognition in the entire study population, active vs. control, had patients experienced depression as if they had all received the active treatment (natural direct effect).

- Formally: $\Psi_{\text{NDE}} = \mathbb{E}\left[Y^{a=1,m=M(1)}\right] - \mathbb{E}\left[Y^{a=0,m=M(1)}\right]$

✔ Well defined population

⚠ Also rely on modeling assumptions.