Course 3 · Week 3 — Survival II, causal inference, HTE

Cheatsheet — biostats_courses

R. Heller

Time-varying covariates / landmarking

- Counting-process data: split follow-up into intervals.
- Landmark analysis: define a landmark time t^*, condition on surviving to t^*, classify by status at t^*.

```
library(survival)
coxph(Surv(tstart, tstop, event) ~ x, data = long)
```

- Immortal-time bias: exposed group can only be exposed if they live long enough to be exposed ⟶ spurious survival benefit.

Competing risks

```
library(tidycmprsk); library(ggsurvfit)
cif <- cuminc(Surv(time, event_f) ~ x, data = df)
ggsurvfit::ggcuminc(cif, outcome = "event1")

# Fine-Gray subdistribution hazard for event1
crr(Surv(time, event_f) ~ x, data = df, failcode = "event1")
```

- Cause-specific HR answers "given you are event-free, what is the hazard for event k?"
- Subdistribution HR answers "what is the hazard for event k in the whole cohort, treating competing events as obstacles?"

DAGs

```
library(dagitty); library(ggdag)
dag <- dagitty("dag { X -> Y ; C -> X ; C -> Y }")
adjustmentSets(dag, exposure = "X", outcome = "Y")
impliedConditionalIndependencies(dag)
```

- Adjust for confounders, **not** colliders.
- Adjusting for a mediator estimates the direct effect only.

Propensity scores / IPTW

```
library(MatchIt); library(cobalt)
m <- matchit(treat ~ x1 + x2, data = df, method = "nearest")
love.plot(m)   # check balance: SMD < 0.1 is the goal
fit <- lm(y ~ treat, data = match.data(m))
```

IPTW: weight by $1/\hat{e}_i$ for treated, $1/(1-\hat{e}_i)$ for untreated. Stabilised weights stay closer to 1; trim extremes to avoid outliers.

G-methods, IV, DiD, RDD

Method	Identifying assumption
G-formula / IPTW	positivity + no unmeasured confounding
Instrumental variables	exclusion restriction + relevance

Method	Identifying assumption
Difference-in-differences	parallel trends
Regression discontinuity	continuity at the cutoff

Heterogeneous treatment effects

- Causal forest (`grf::causal_forest`) or meta-learners (S-, T-, X-).
- Report conditional ATEs with adjusted intervals; avoid spurious subgroup p-values.

Decision rule for Week 3

- Treatment assigned over time ⟶ time-varying Cox.
- Multiple competing events ⟶ cumulative incidence, Fine-Gray.
- Observational causal question ⟶ DAG first, choose identification strategy second.
- Want the whole distribution of treatment effect? ⟶ causal forest.

Common pitfalls

- Adjusting for a post-treatment variable (induces collider bias).
- Quoting HR from a Cox model that violates PH.
- Matching but never checking balance.
- Calling IV estimates “generalisable” when they are local to compliers.

Further reading

- Hernán & Robins, *Causal Inference: What If.*
- Therneau & Grambsch, *Modeling Survival Data.*