Causal Inference in R

Define the target trial, identify adjustment variables and estimate a transparent contrast

Rverse Analytics

A regression coefficient is not causal because covariates were added. State the intervention, comparator, eligible population, time zero, outcome, follow-up and causal contrast before selecting an adjustment method.

Target-trial checklist

Component	Example
Eligibility	adults initiating one of two therapies
Treatment strategies	initiate A vs initiate B at baseline
Assignment	emulate randomization conditional on measured confounders
Time zero	treatment initiation; eligibility assessed before it
Outcome	12-month hospitalization risk
Intercurrent events	switching, discontinuation, death strategy
Contrast	ATE risk difference and risk ratio

Avoid immortal-time bias by aligning eligibility, treatment assignment and follow-up start.

Draw assumptions before fitting

Variable role	Adjust?	Reason
Common cause of treatment and outcome	yes	confounder
Cause of outcome only	often yes	improves precision
Cause of treatment only	sometimes	may amplify bias/variance
Mediator after treatment	no for total effect	blocks part of effect
Collider caused by treatment/outcome causes	no	opens a noncausal path
Instrument	not as ordinary confounder	affects treatment, not outcome directly

Use `dagitty` when available to encode the DAG and derive sufficient adjustment sets:

```
library(dagitty)
g <- dagitty("dag { age -> treatment; age -> outcome;
  severity -> treatment; severity -> outcome; treatment -> outcome }")
adjustmentSets(g, exposure = "treatment", outcome = "outcome")
```

Inspect exchangeability and positivity

```
library(dplyr)

d <- d |>
  mutate(treatment = as.integer(treatment == "Treated"))

with(d, table(treatment, useNA = "ifany"))
d |>
  group_by(treatment) |>
  summarise(across(c(age, severity, baseline_risk),
    list(mean = ~mean(.x, na.rm = TRUE), sd = ~sd(.x, na.rm = TRUE))))

ps_fit <- glm(treatment ~ age + severity + baseline_risk + center,
              family = binomial, data = d)
d$ps <- predict(ps_fit, type = "response")
range(d$ps[d$treatment == 1]); range(d$ps[d$treatment == 0])
```

No treated or untreated observations in parts of covariate space means the target contrast is not identified there without extrapolation. Redefine the population or treatment strategy rather than hiding extreme weights.

IPTW for the average treatment effect

```
p_treated <- mean(d$treatment)
d <- d |>
```

```r
  mutate(
    sw = if_else(treatment == 1,
      p_treated / ps,
      (1 - p_treated) / (1 - ps))
  )

summary(d$sw)
quantile(d$sw, c(.01, .5, .99), na.rm = TRUE)

library(survey)
des <- svydesign(ids = ~1, weights = ~sw, data = d)

# Marginal risks and a population-average risk ratio
svyby(~outcome, ~treatment, des, svymean, vartype = "ci")
fit_rr <- svyglm(outcome ~ treatment, design = des,
                 family = quasibinomial(link = "log"))
exp(cbind(Estimate = coef(fit_rr), confint(fit_rr)))
```

For ATT weights, target treated participants explicitly. Do not switch ATE/ATT after seeing overlap; the estimand changes.

Balance, not propensity-model prediction

```r
# Manual weighted standardized mean difference for one continuous covariate
wmean <- function(x, w) weighted.mean(x, w, na.rm = TRUE)
wvar <- function(x, w) {
  ok <- complete.cases(x, w); x <- x[ok]; w <- w[ok]
  sum(w * (x - weighted.mean(x, w))^2) / sum(w)
}

smd_w <- function(x, a, w) {
  m1 <- wmean(x[a == 1], w[a == 1]); m0 <- wmean(x[a == 0], w[a == 0])
  s <- sqrt((wvar(x[a == 1], w[a == 1]) + wvar(x[a == 0], w[a == 0])) / 2)
  (m1 - m0) / s
}

smd_w(d$age, d$treatment, d$sw)

# Ergonomic alternatives when installed
# WeightIt::weightit(treatment ~ age + severity + baseline_risk, data = d)
# cobalt::bal.tab(weights_object, un = TRUE)
```

Show balance before and after weighting for every prespecified confounder, plus propensity overlap and weight distribution. A nonsignificant baseline test is not evidence of balance.

Standardization / g-computation

```r
out_fit <- glm(outcome ~ treatment + age + severity + baseline_risk + center,
               family = binomial, data = d)

d1 <- transform(d, treatment = 1)
d0 <- transform(d, treatment = 0)
risk1 <- mean(predict(out_fit, newdata = d1, type = "response"))
risk0 <- mean(predict(out_fit, newdata = d0, type = "response"))
c(risk_treated = risk1, risk_control = risk0,
  risk_difference = risk1 - risk0, risk_ratio = risk1 / risk0)
```

Use bootstrap or influence-function methods for uncertainty. Doubly robust estimators combine treatment and outcome models but still require consistency, positivity and no unmeasured confounding.

Sensitivity and reporting

- Assess alternate confounder sets justified by different plausible DAGs.
- Report trimming/truncation rules and how they change the target population.
- Use negative controls or quantitative bias analysis when credible.
- Treat E-values as one sensitivity summary, not proof against unmeasured confounding.
- Report the causal estimand, time zero, assumptions, balance, overlap, weights and absolute effects with 95% CIs.
