Clinical Trial Analysis in R

From estimand and analysis sets to adjusted effects and sensitivity analyses

Rverse Analytics

Start from the estimand, not the available rows. Define treatment, population, variable, intercurrent-event strategy and population-level summary before opening treatment assignments.

Analysis populations

Set	Operational definition	Role
Randomized / ITT	all randomized, analysed as assigned	primary superiority principle
Modified ITT	prespecified subset of randomized participants	define exclusions precisely
Per protocol	participants meeting prespecified adherence criteria	supportive; selection risk
Safety	received ≥1 dose, analysed as treated	harms and exposure

Never recreate analysis sets from analyst judgment. Derive them from versioned rules and preserve an exclusion audit trail.

Randomization and baseline table

```
library(dplyr)
library(gtsummary)

trial_d <- raw |>
  mutate(
    arm = factor(arm, levels = c("Control", "Treatment")),
    itt = randomized == 1,
    safety = received_dose == 1
  )

flow <- trial_d |>
  count(arm, randomized, received_dose, completed, name = "n")

tab1 <- trial_d |>
  filter(itt) |>
  select(arm, age, sex, baseline_score, strat_center) |>
  tbl_summary(by = arm, missing = "always") |>
  add_n() |>
  add_difference(test = everything() ~ "smd") |>
  bold_labels()
```

Baseline imbalance tests do not test whether randomization “worked.” Show descriptive balance and prespecified standardized differences; adjust for planned prognostic variables and randomization strata.

Continuous primary outcome: ANCOVA

```
itt_d <- filter(trial_d, itt)

fit <- lm(week12_score ~ arm + baseline_score + strat_center,
          data = itt_d, na.action = na.exclude)

estimate <- emmeans::emmeans(fit, ~ arm) |>
  pairs(reverse = TRUE, adjust = "none") |>
  summary(infer = TRUE)

broom::tidy(fit, conf.int = TRUE)
```

Use an adjusted final-value model when that is the prespecified estimand. Change scores are not automatically superior and may be less efficient.

Binary and time-to-event outcomes

```
# Logistic estimand: adjusted odds ratio
fit_bin <- glm(response ~ arm + baseline_risk + strat_center,
               family = binomial, data = itt_d)
```

```r
gtsummary::tbl_regression(fit_bin, exponentiate = TRUE)

# Marginal standardized risks and risk difference
std <- marginaleffects::avg_predictions(fit_bin, by = "arm", type = "response")
marginaleffects::avg_comparisons(fit_bin, variables = "arm", type = "response")

# Time-to-event
fit_cox <- survival::coxph(
  survival::Surv(time, event) ~ arm + strat_center,
  data = itt_d, ties = "efron", x = TRUE
)
survival::cox.zph(fit_cox)
```

Pair relative effects with absolute effects at a clinically meaningful horizon whenever possible.

Missing outcomes and intercurrent events

Event	Strategy example	Interpretation
Treatment discontinuation	treatment policy	outcome regardless of discontinuation
Rescue medication	hypothetical or composite	depends on explicit scientific question
Death before measurement	composite / while-alive / principal stratum	cannot be treated as ordinary missing score
Loss to follow-up	MI under MAR + MNAR sensitivity	uncertainty about unseen outcome

```r
# Multiple imputation within the ITT set
imp <- mice::mice(select(itt_d, week12_score, arm, baseline_score,
                         strat_center, adherence, week6_score),
                  m = 40, seed = 20260711, printFlag = FALSE)
fit_mi <- with(imp, lm(week12_score ~ arm + baseline_score + strat_center))
summary(mice::pool(fit_mi), conf.int = TRUE)
```

Subgroups and multiplicity

```r
# One interaction model; do not compare subgroup p-values
fit_sub <- lm(week12_score ~ arm * sex + baseline_score + strat_center,
              data = itt_d)
anova(update(fit_sub, . ~ . - arm:sex), fit_sub)

# Prespecified family adjustment
p.adjust(c(p_primary, p_key_secondary_1, p_key_secondary_2), method = "holm")
```

Subgroup results require interaction estimates and CIs. Forest plots should distinguish prespecified from exploratory subgroups and show sparse cells.

Reproducibility and reporting

- Freeze the statistical analysis plan and shells before unblinding.
- Validate derivations, treatment coding, visit windows and analysis populations independently.
- Report participant flow, protocol deviations, missing outcomes, harms and denominators.
- Present primary estimate, 95% CI, exact *p*, effect scale and clinical margin.
- Include prespecified sensitivity analyses and explain disagreements with the primary result.

Rverse Analytics · rverseanalytics.com · Original reference sheet