Introduction

Slides

Chapter slides here. (To convert html to pdf, press E → Print → Destination: Save to pdf)

R code

```
#########################################################################
# This code generates the numerical results in  chapter 1       ##
#########################################################################

# load the survival package
library(survival)
# install.packages("Wcompo")
library(Wcompo) # for weighted total events
library(rmt) # for hfaction data
library(tidyverse) # for data wrangling

# load hfaction data
data(hfaction)
head(hfaction)

# convert status=1 for death, 2=hospitalization
hfaction <- hfaction |>
  mutate(
    status = case_when(
      status == 1 ~ 2,
      status == 2 ~ 1,
      status == 0 ~ 0
    )
  )

head(hfaction)
```

```r
# count unique patients in each arm
hfaction |>
  group_by(trt_ab) |>
  distinct(patid) |>
  count(trt_ab)

# TFE: take the first event per patient id
hfaction_TFE <- hfaction |>
  arrange(patid, time) |>
  group_by(patid) |>
  slice_head() |>
  ungroup()

# Mortality analysis ------------------------------------------------------

## get mortality data
hfaction_D <- hfaction |>
  filter(status != 2) # remove hospitalization records

## Cox model for death against trt_ab
obj_D <- coxph(Surv(time, status) ~ trt_ab, data = hfaction_D)
summary(obj_D)
#> n= 426, number of events= 93
#> coef exp(coef) se(coef)      z Pr(>|z|)
#> trt_ab -0.3973    0.6721   0.2129 -1.866   0.0621 .

# TFE analysis ------------------------------------------------------------

## how many of first events are death (1) or hosp (2)
hfaction_TFE |>
  count(status)

# Cox model for TFE against trt_ab
obj_TFE <- coxph(Surv(time, status > 0) ~ trt_ab, data = hfaction_TFE)
summary(obj_TFE)


# Mortality vs TFE --------------------------------------------------------

library(ggsurvfit)
library(patchwork)
```

```r
pD <- survfit2(Surv(time, status) ~ trt_ab, data = hfaction_D) |>
  ggsurvfit(linewidth = 1) +
  scale_ggsurvfit() +
  scale_color_discrete(labels = c("Usual care", "Training")) +
  scale_x_continuous("Time (years)", limits = c(0, 4)) +
  labs(y = "Overall survival")

pTFE <- survfit2(Surv(time, status > 0) ~ trt_ab, data = hfaction_TFE) |>
  ggsurvfit(linewidth = 1) +
  scale_ggsurvfit() +
  scale_color_discrete(labels = c("Usual care", "Training")) +
  scale_x_continuous("Time (years)", limits = c(0, 4)) +
  labs(y = "Hospitalization-free survival")


pD + pTFE + plot_layout(guides = "collect") &
  theme(
    legend.position = "top",
    legend.text = element_text(size = 12)
    )

# ggsave("images/intro_hfaction_unis.png", width = 8, height = 4.5)

# Total events (proportional mean) ----------------------------------------
## fit proportional means model with death = 2 x hosp
obj_ML <- CompoML(hfaction$patid, hfaction$time, hfaction$status,
                  hfaction$trt_ab, w = c(2, 1))
## summary results
obj_ML

## plot model-based mean functions
plot(obj_ML, 0, ylim= c(0, 5), xlim= c(0, 4), xlab="Time (years)",
     col = "red", lwd = 2)
plot(obj_ML, 1, add = TRUE, col = "blue", lwd=2)
legend(0, 5, col = c("red", "blue"), c("Usual care","Training"), lwd = 2)
```

Motivating Examples

In trials where patients may experience both fatal and nonfatal events, time-to-first-event analyses can obscure clinically meaningful treatment effects. For instance, in a colon cancer

trial of 619 patients, relapse-free survival was used as the primary endpoint. Although the treatment significantly delayed relapse (log-rank $p < 0.001$), 89% of deaths occurred after relapse and were ignored in the analysis.

Similarly, in the HF-ACTION trial of heart failure patients, a composite of death and hospitalization failed to show a statistically significant benefit (log-rank $p = 0.10$), even though 88% of patients died and 69% had recurrent hospitalizations. By equating hospitalization with death and counting only the first event, the analysis diluted the observed mortality effect.

These examples highlight the limitations of standard composite endpoints and motivate approaches that account for the clinical hierarchy and full event history.

Composite Endpoints and Guidelines

Composite endpoints increase statistical power by aggregating different event types, reducing the need for multiple testing and improving trial efficiency. However, their validity hinges on appropriate construction and interpretation.

Regulatory guidelines provide direction:

- **ICH E9** recommends using a single primary variable or a composite when individual outcomes cannot capture the treatment effect.
- **FDA** allows time-to-first or total-event analysis but emphasizes component-wise interpretation.
- **European HTA bodies** advise combining outcomes of similar severity and always including mortality when relevant, especially if it censors other outcomes.

These guidelines support the use of composite endpoints while cautioning against misinterpretation.

Traditional Composite Methods

Let $\mathcal{H}^*(t)$ denote the cumulative history of death and K nonfatal events up to time t.

Time to First Event

The standard approach defines:

$$N^*_{\text{TFE}}(t) = I\left\{N^*_D(t) + \sum_{k=1}^{K} N^*_k(t) \geq 1\right\},$$

where $N_D^*(t)$ and $N_k^*(t)$ are counting processes for death and nonfatal event k, respectively. This method reduces all outcome complexity to a single event time, analyzed via Kaplan–Meier, log-rank, or Cox models.

Weighted Total Events

To incorporate more information, the total number of events can be weighted:

$$N_{\mathrm{R}}^*(t) = w_D N_D^*(t) + \sum_{k=1}^{K} w_k N_k^*(t),$$

with w_D, w_k representing clinical importance. Estimation is performed under a proportional means model:

$$E\{N_{\mathrm{R}}^*(t) \mid Z\} = \exp(\beta^{\mathrm{T}} Z)\mu_0(t).$$

The `Wcompo` R package provides tools for this approach via the `CompoML()` function.

HF-ACTION Revisited

In the HF-ACTION trial, separate analyses illustrate the limitations of standard methods:

- **Cox model for death** showed a 32.8% risk reduction (HR = 0.672, $p = 0.062$).
- **Cox model for TFE** (death or first hospitalization) showed only a 16.2% reduction (HR = 0.838, $p = 0.111$).
- **Weighted total events** (death weighted twice) estimated a 14.3% reduction in event burden, but the p-value was still nonsignificant.

These results demonstrate how frequent nonfatal events can mask treatment benefits on mortality, particularly in analyses that do not prioritize outcome severity or account for survival time.

Win Ratio and Hierarchical Approaches

The **win ratio (WR)** offers a solution by comparing patient pairs across treatment arms. Each pair is evaluated hierarchically:

1. Longer survival indicates a win.
2. If both survive, fewer or later nonfatal events confer a win.
3. If tied, functional outcomes may be compared.

The win ratio is defined as:

$$\widehat{\text{WR}} = \frac{\#\text{ of treatment wins}}{\#\text{ of control wins}}.$$

Alternative metrics include:

- **Net benefit** (PIF): wins minus losses.
- **Win odds** (WO): accounts for ties:

$$\text{WO} = \frac{\text{wins} + \frac{1}{2}\text{ties}}{\text{losses} + \frac{1}{2}\text{ties}}.$$

These measures align with clinical priorities and are increasingly used in practice.

Estimand Challenges

A critical limitation of the win ratio is its dependency on censoring. Because patients may be censored at different times, the win/loss status of a pair can change over time. The observed win ratio thus reflects an average over the censoring distribution, not a fixed population quantity.

This raises concerns for interpretation and estimation:

- **Testing** remains valid under the null hypothesis, especially if the treatment consistently outperforms control.
- **Estimation**, however, is biased unless a fixed follow-up time is defined or a model constrains the win ratio to be constant over time.

The ICH E9 (R1) addendum emphasizes that a clearly defined **estimand** is essential for interpreting treatment effects. For the win ratio, this requires careful attention to follow-up duration and censoring.

Toward Model-Based Estimands

Limitations of traditional and empirical approaches motivate model-based alternatives: - The **restricted mean time in favor (RMT-IF)** quantifies the net amount of time one group remains in a more favorable state. - The **proportional win-fractions (PW) model** expresses the win ratio as a function of covariates, enabling estimation of treatment effects under a semiparametric framework.

Both methods respect clinical hierarchies and avoid dependence on arbitrary censoring patterns. They form the foundation of the methods introduced in subsequent chapters.

Example R code

```
########################################
# 1. Proportional means model using Wcompo
########################################
library(Wcompo)
obj <- CompoML(id, time, status, Z, w = c(2, 1))
summary(obj)
plot(obj)
```