EFSPI REGULATORY STATISTICS WORKSHOP 2026 · SHORT TOPIC 4

Composite Strategies for Continuous Endpoints: *Failure Values or Failed Question?*

CONTRIBUTORS

Baldur Magnuson (UCB)* **Marcel Wolbers (Bayer)*** James Bell (Elderbrook Solutions) Christian Bressen Pipper (Novo Nordisk) Mickaël De Backer (UCB) Thomas Drury (GSK) Oliver N. Keene (KeeneONStatistics) Vivian Lanius (Boehringer Ingelheim) Marian Mitroiu (Biogen) Tobias Mütze (Novartis) Khadija Rerhou Rantell (MHRA) Bohdana Ratitch (Bayer) David Wright (AstraZeneca)

* *Presenting*

Clinical Context & Problem Statement

Clinical context

- **Continuous endpoint** measured over time

e.g. forced vital capacity in pulmonary disease

- Multimorbid population at risk of a **terminal intercurrent event (IE)**

e.g. 5-10% expected deaths

- **Composite strategy** selected for the IE

Analogy:

  - *TTE: progression in score or death*
  - *Continuous: longitudinal score + death*

Similar information, similar question

The problem

- **Operationalization** requires:
  - A **failure value** when experiencing the IE
  - A **summary measure**

- The conventional summary measure for continuous endpoints is the **mean difference** but:
  - IE has **no natural location on endpoint scale**
  - Choice of failure value critically **affects interpretability, analysis methods, and power**

FOR THE PANEL

Question 1: Mean Differences & Failure Values

What regulatory considerations should guide the choice of failure value? Merits of — and preferences among — these options:

a)

Pre-defined fixed value

A single explicit unfavorable value on the endpoint scale, used for all IE patients (e.g. the worst possible value).

b)

Population-level value

A single implicit value from the control-arm distribution amongst subjects without the IE (e.g. 5% quantile or mean − 2 SD) **estimated** from trial data

c)

Sample-level value

E.g. worst observed value or multiple imputation from an unfavorable range.

FOR THE PANEL

Questions 2 & 3 — Beyond the Failure Value

2 Alternative summary measures

The mean difference has a long history, but alternative summary measures (e.g., the difference in medians or Hodges-Lehmann pseudo-medians) may be more faithful to the actual amount of information at disposal: ‘the terminal event is worse’.

In which circumstances does the panel recommend the use of alternative summary measures?

Does the panel prefer a specific alternative summary measure?

3 Alternative estimands (time permitting)

Could the panel comment on the merits of reformulating the clinical research question and, consequently, choosing alternative primary or supplementary estimands?