Developing a simulation-based decision framework for interpretation of interim survival data in oncology trials

Andrew Mills

Content Overview

- Problem statement
- Approach
  - Simulation
  - Decision framework
- Outcomes
  - Heatmaps and interpretations
  - General guidance

Collaborators

AMGEN

AstraZeneca

Bristol Myers Squibb

Crossbow therapeutics

Daiichi-Sankyo

EMD SERONO

Fred Hutch Cancer Center

Genentech
A Member of the Roche Group

GSK

Johnson&Johnson Innovative Medicine

Lilly

MERCK

NMDgroup
CLARITY – DATA & ANALYTICS AT WORK FOR YOU

NOVARTIS

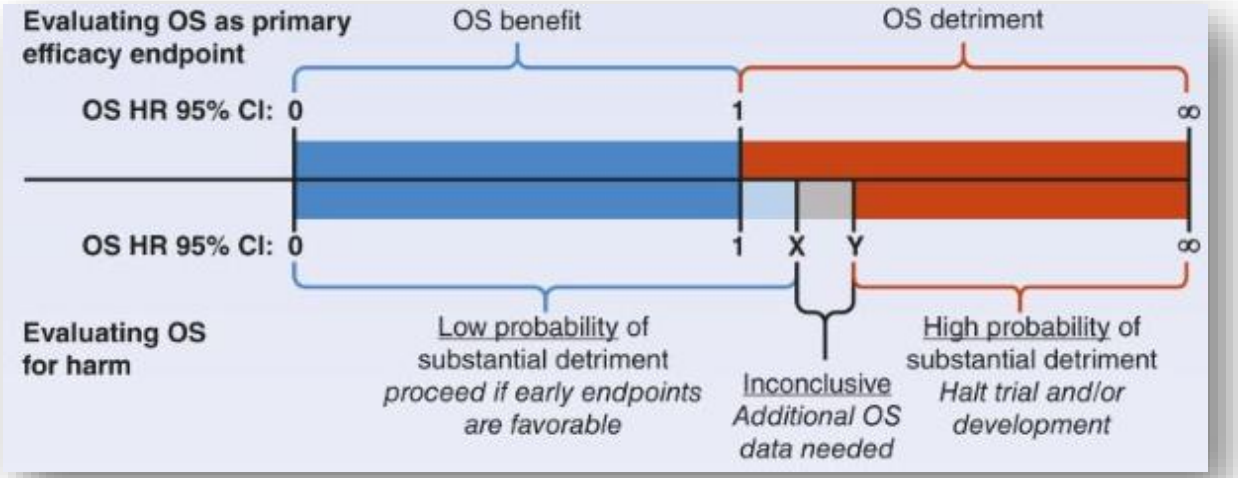
Pfizer

FDA

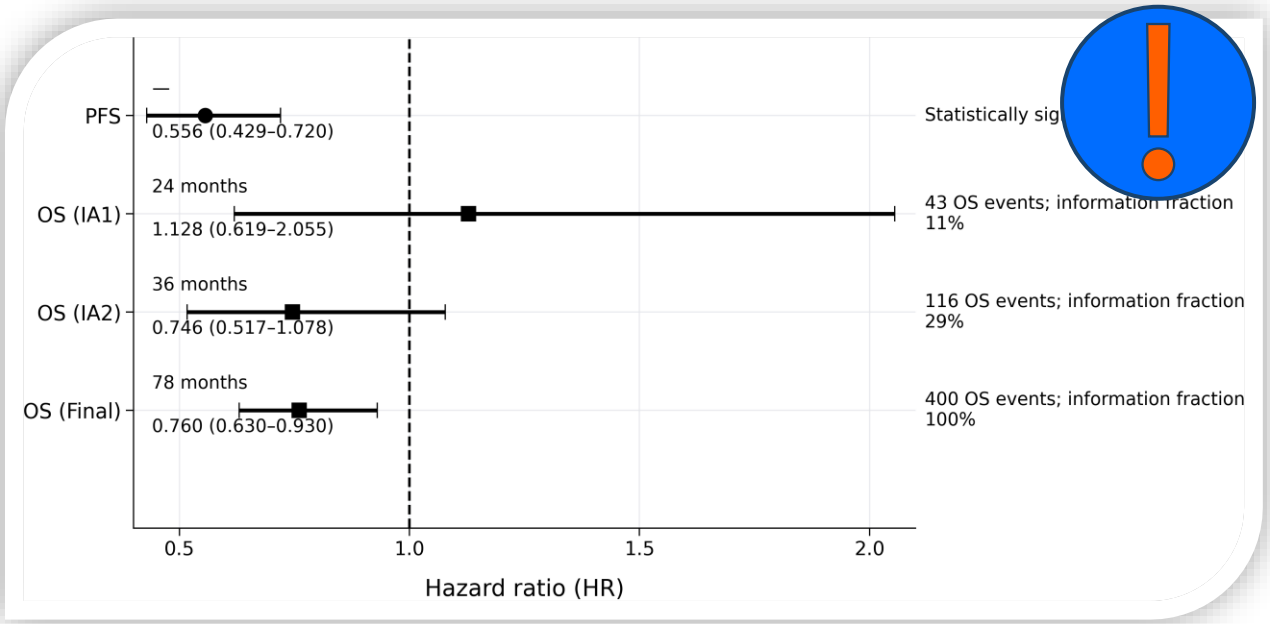
Problem Statement

Problem Statement

- **Interim OS is often reviewed at the time of a surrogate endpoint readout**
  - Evaluating OS for harm
  - Interpretation is challenging due to often immature data
  - Greater variability, greater chance of erroneous conclusions
- **Additional factors adding to the complexity of interpretations**
  - Non-proportional hazards
  - Treatment crossover



Rodriguez LR, Gormley NJ, Lu R, Amatya AK, Demetri GD, Flaherty KT, et al. Improving collection and analysis of overall survival data. Clin Cancer Res. 2024;30(18):3974–3982. doi:10.1158/1078-0432.CCR-24-0919



MONALEESA-2 Trial (Ribociclib + Letrozole in HR+/HER2- Metastatic Breast Cancer)

Problem Statement

Approaches to Assessment of Overall Survival in Oncology Clinical Trials

Guidance for Industry

DRAFT GUIDANCE

U.S. Food and Drug Administration. Approaches to assessment of overall survival in oncology clinical trials: Draft guidance for industry. Draft guidance. August 2025. Available from: https://www.fda.gov

2. *Assessment of Harm*

- To adequately inform regulatory decision-making, any trial should be designed a priori to adequately capture the number of events needed to rule out harm based on a specified threshold(s). Sponsors should specify and justify in the SAP a threshold or a range of thresholds to indicate the potential for harm for the overall survival summary measure.

  - If the sponsor anticipates that overall survival data will be immature or there will be high uncertainty in the overall survival summary measures at the time of the proposed final analysis, sponsors should conduct additional calculations or simulations to assess if it is likely that harm may be ruled out with additional follow-up time (which may or may not be feasible). Various methodologies can be used to calculate the probability of ruling out harm based on observing hypothetical future data. Sponsors should specify in the SAP approaches to rule out harm under a variety of scenarios, justification of methods, and justification of assumptions (including the assumption of observing additional events over a specified time period). Sponsors should include a variety of assumed scenarios for future data. However, the uncertainty in these methods generally increases for longer looks into the future, and evaluation of harm cannot solely be based on hypothetical future data.

Approach

Approach

1. Define the anticipated survival patterns (active vs. control)
2. Define plausible alternative scenarios (sensitivity analyses)
3. Determine Decision Criteria for identifying harm at Final OS Analysis
4. Evaluate Operating Characteristics at Final Analysis under anticipated and alternative scenarios
5. Evaluate Interim Operating Characteristics using decision rules defined at the Final Analysis

Simulation

Decision

Simulation – CSE Framework

Censoring at Data Cut

Loss to follow-up

Realistic simulated data

ADaM-like structure

Correlated simulated variables

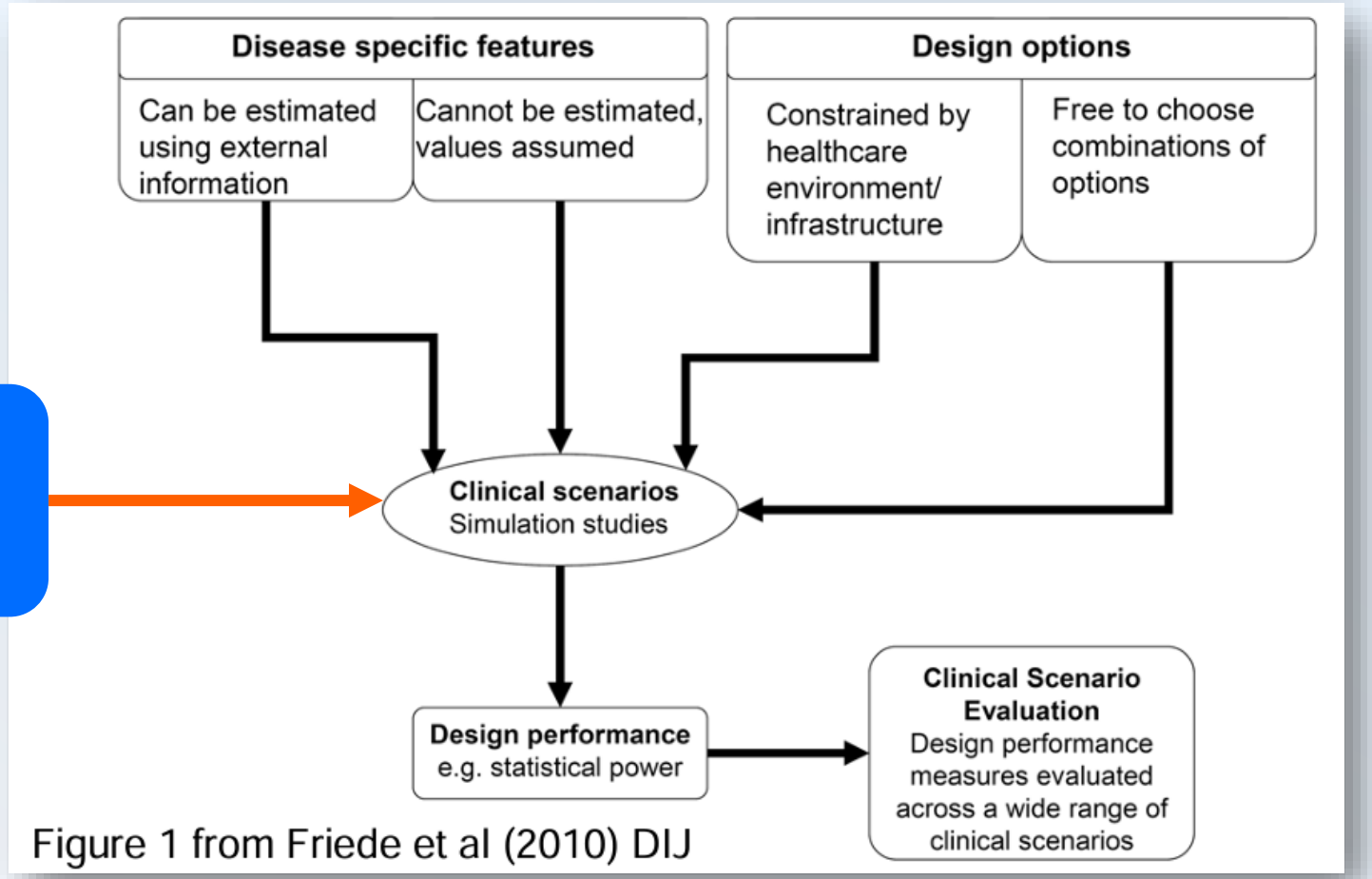
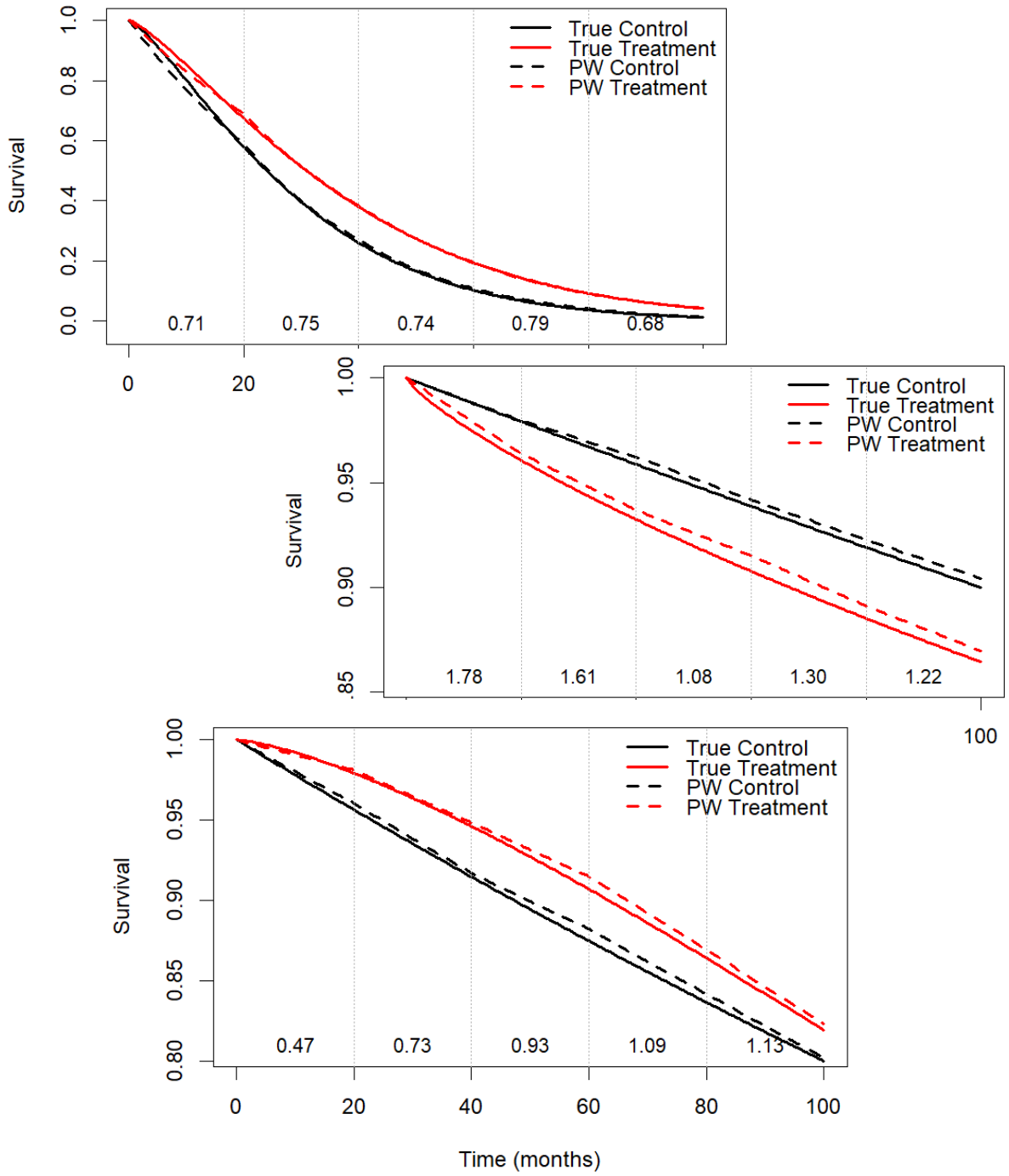


Figure 1 from Friede et al (2010) DIJ

Simulation

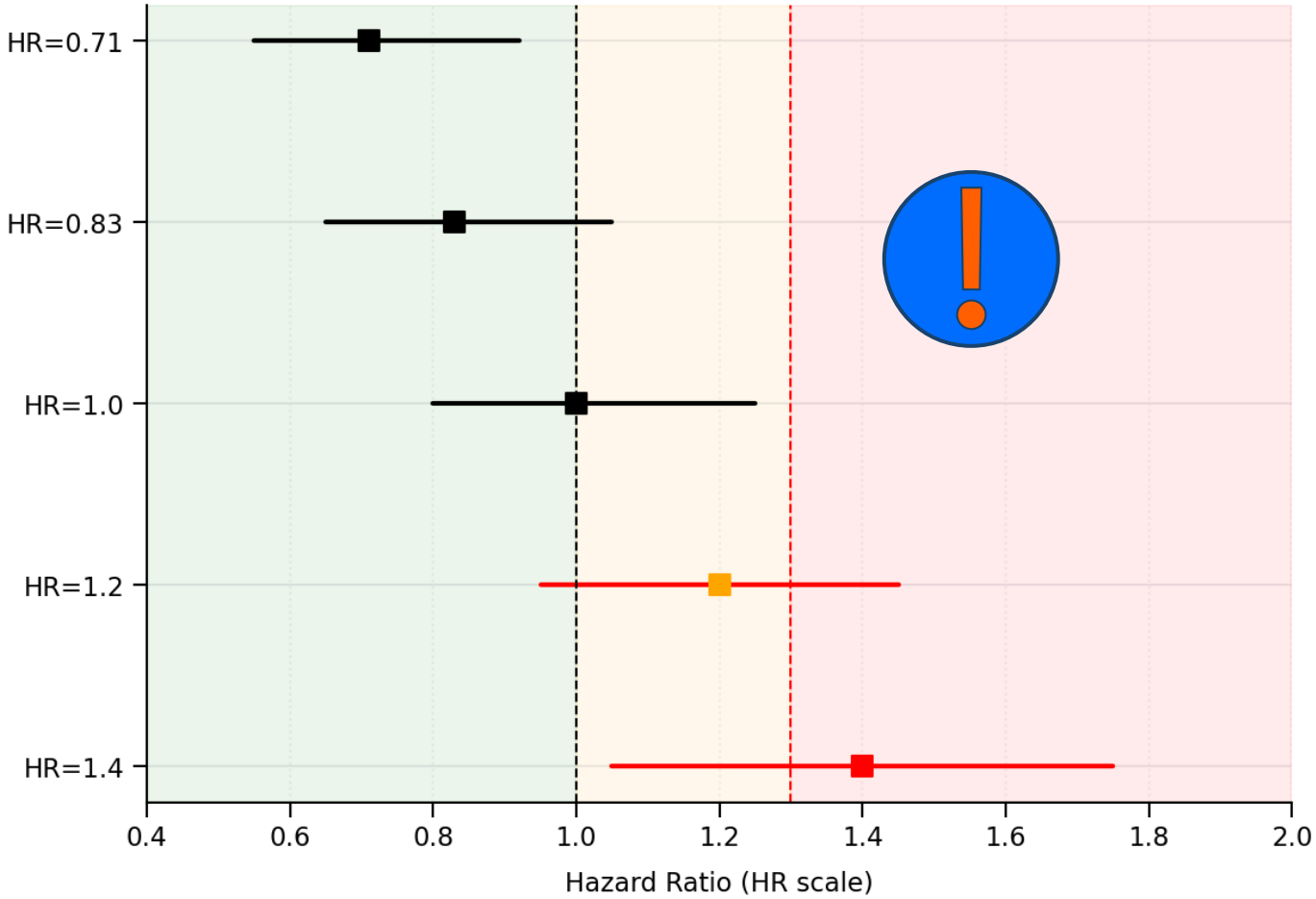
- **Proportional hazards**
  - Strong benefit (HR = 0.71)
  - Modest benefit (HR = 0.83)
  - No effect (HR = 1.0)
  - Modest harm (HR = 1.2)
  - Strong harm (HR = 1.4)
- **Non-proportional hazards**
  - Diminishing benefit (HR* = 0.71, 0.83)
  - Delayed benefit (HR* = 0.71, 0.83)
  - Early harm (HR* = 1.4)



* Target average HR, due to non-constant hazard rates

Decision Framework – Safe or Harm

- **Truth**:
  - User assigned / justified
- **Final OS analysis**:
  - HR Point estimate > 1.0
  - HR Point estimate > 1.3
  - HR Upper Confidence Limit > 1.3
- **Interim OS analysis**:
  - HR Upper Confidence Limit > 1.1 – 1.8

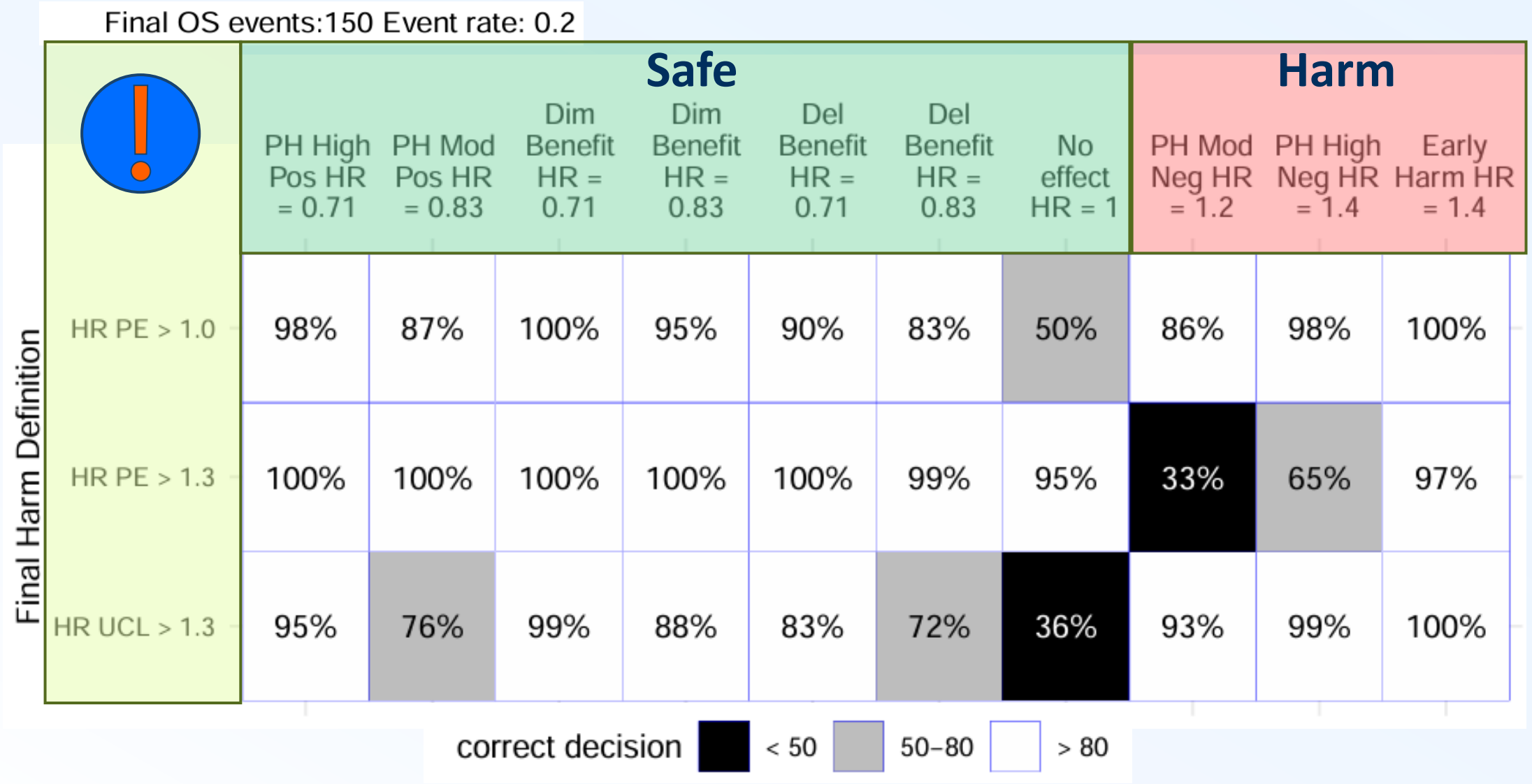


Decision Framework

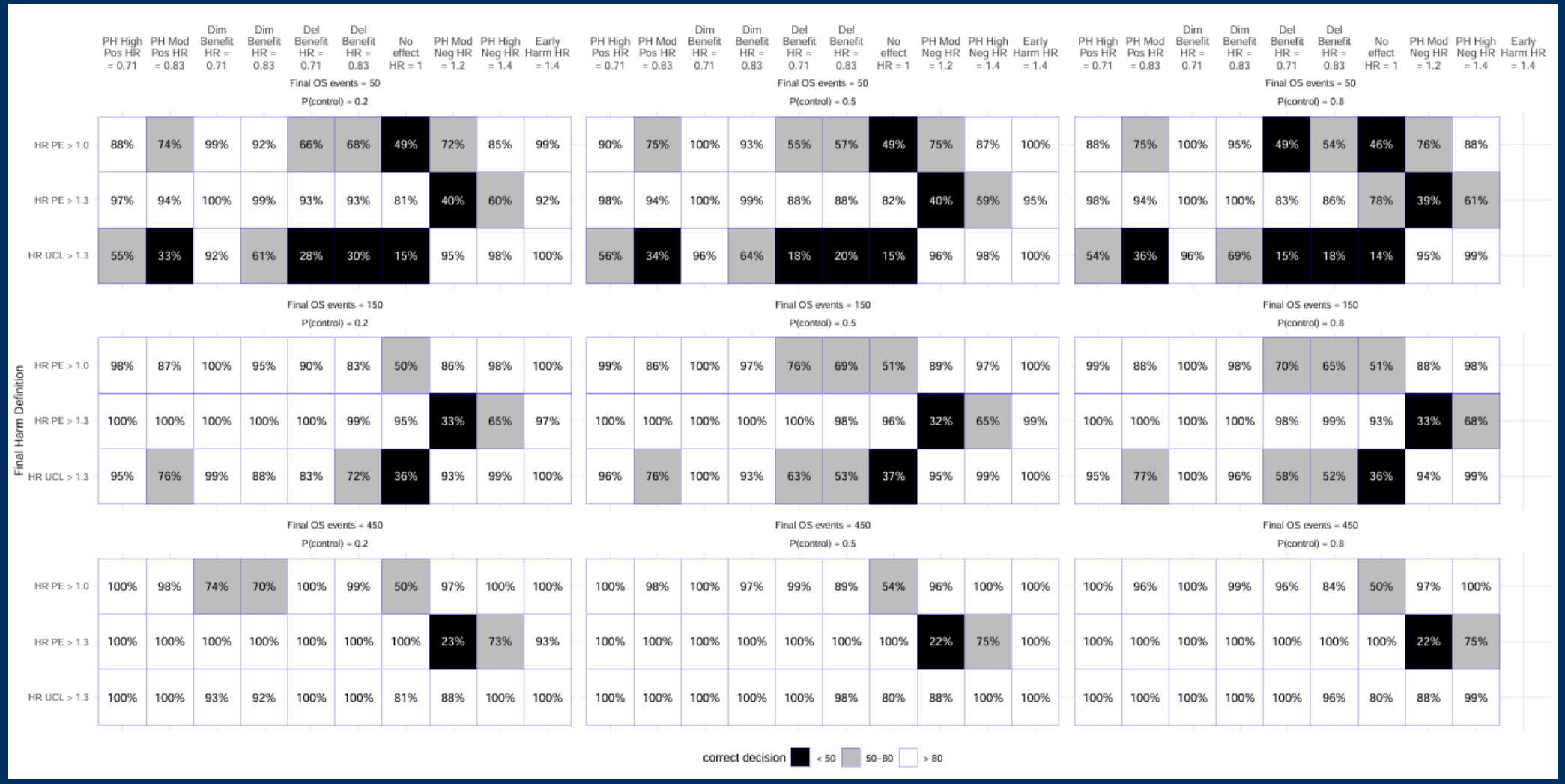
- **Establish the reliability of the final OS analysis**
  - Primary operating characteristic: Percent *correct* decision
    - Compare Final OS result (Safe / Harm) to Truth (Safe / Harm)
- **Evaluate interim OS performance**
  - Primary operating characteristic: Percent *consistent* decision
    - Compare Interim OS result (Safe / Harm) to Final OS result (Safe / Harm)
  - Secondary operating characteristics
    - Sponsor risk (Harm at Interim, Safe at Final)
    - Regulator risk (Safe at Interim, Harm at Final)
- **Assess a range of survival patterns**

Outcomes

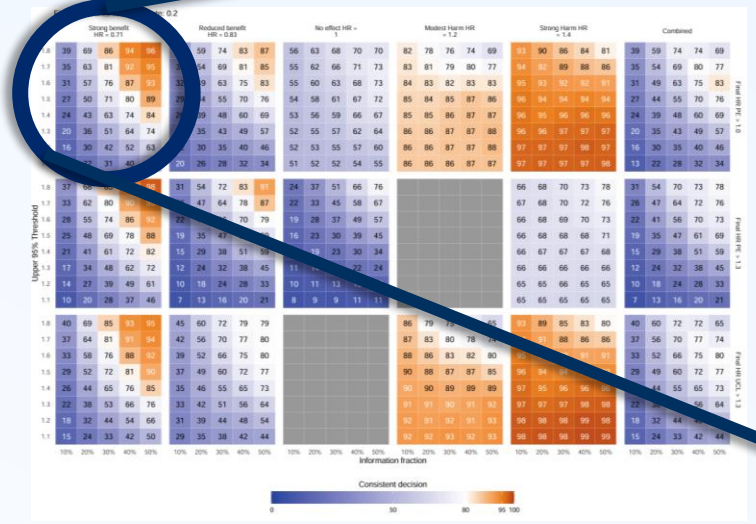
Heatmaps and Interpretation

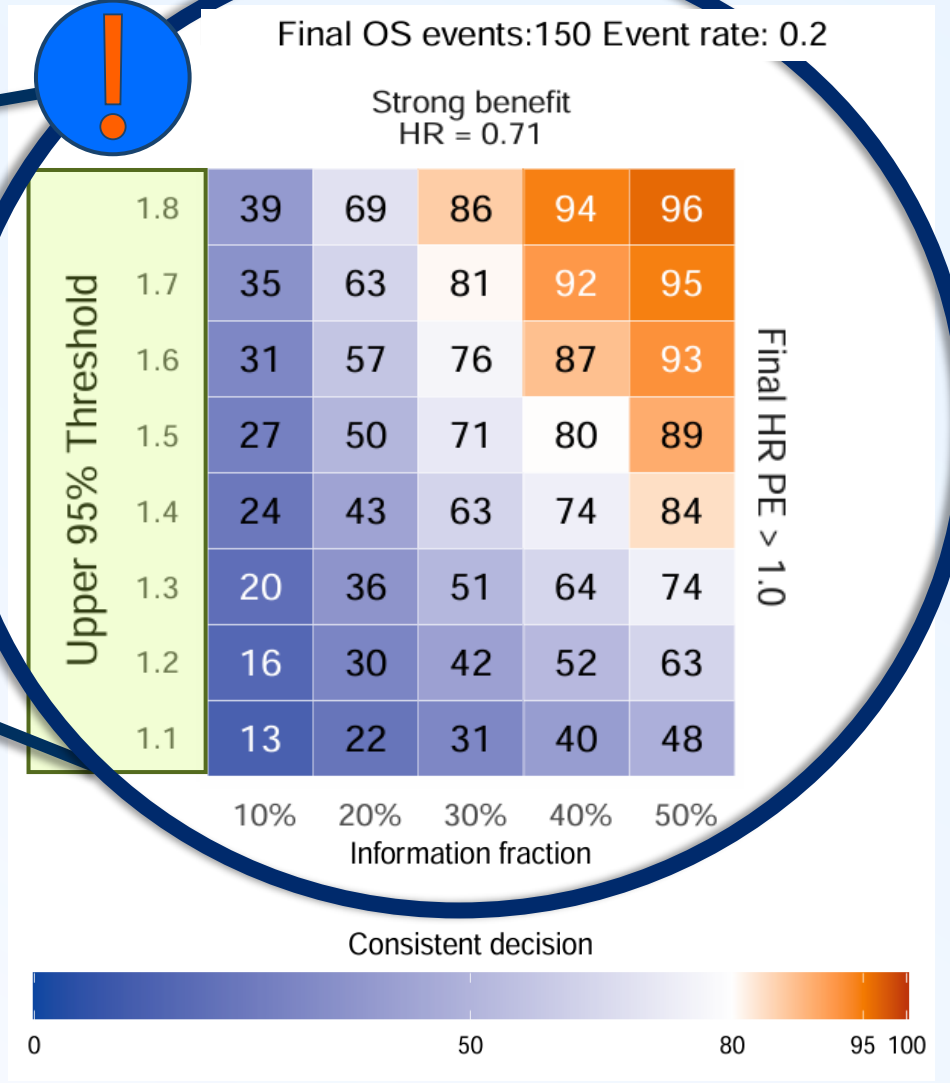


Heatmaps and Interpretation



Heatmaps and Interpretation





General Guidance

Future publication will aim is to determine

- Viable Final OS harm definition paths :
  - Harm threshold and OS Event number combinations
- Areas of consistent interim-to-final conclusions
  - And areas that present distinct challenges likely to warrant separate consideration
- Viable interim UCL thresholds offering the best balance across explored scenarios
  - Including an observed minimum number of interim OS events needed before conclusions are reasonably reliable

Concluding Remarks

- Ensure planned interim timings are adequately assessed through simulation (or alternative)
- Ensure harm thresholds are evaluated, justifiable and communicated with key stakeholders / regulators
- Upcoming publication
- Simulated data and Decision framework R code will be made publicly available

Thank You for Listening!

andmil02@mmsholdings.com