FRIENDS of CANCER RESEARCH

Interpreting Interim Overall Survival in Oncology Trials

Project Overview and Key Findings

Névine Zariffa

NMD Group, LLC

Intermediate Endpoints Create Earlier Decision Points

FRIENDS of CANCER RESEARCH

Why intermediate endpoints are increasingly used

- Intermediate endpoints such as PFS, DFS, and ORR are increasingly used to support earlier assessment of treatment benefit, as efficacy gains and expanding subsequent therapy options make mature OS difficult to obtain within decision-relevant timeframes.
- Waiting for fully powered OS readouts as the sole means of assessing new therapies is often impractical given these timelines.
- In this paradigm, interim OS can be used to contextualize intermediate endpoint findings and inform benefit-risk assessments, including evaluation of potential harm.

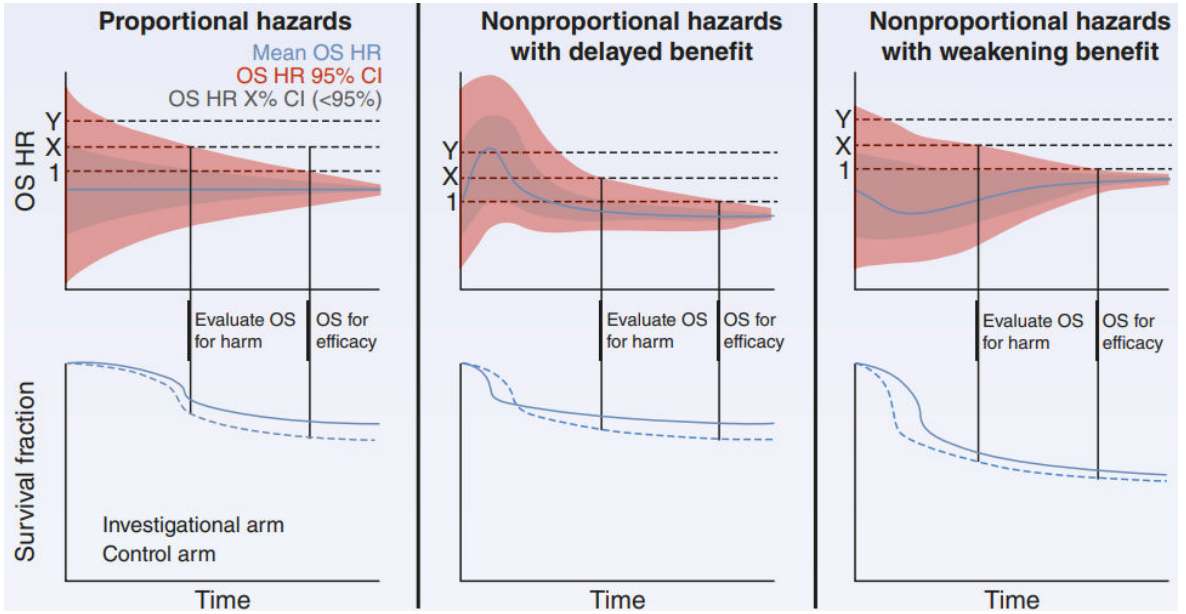
Progression Free Survival, PFS; Disease Free Survival, DFS; Objective Response Rate, ORR; Overall Survival, OS

Interim OS Interpretation Depends on Timing and Survival Pattern

FRIENDS of CANCER RESEARCH

Why Interim OS Can Be Difficult to Interpret

- Interim OS may be based on few events and limited follow-up.
- Early estimates can be unstable and sensitive to timing.
- Non-proportional hazards can make early OS data poorly representative of the eventual OS effect.
- Crossover and subsequent therapies can further complicate attribution.

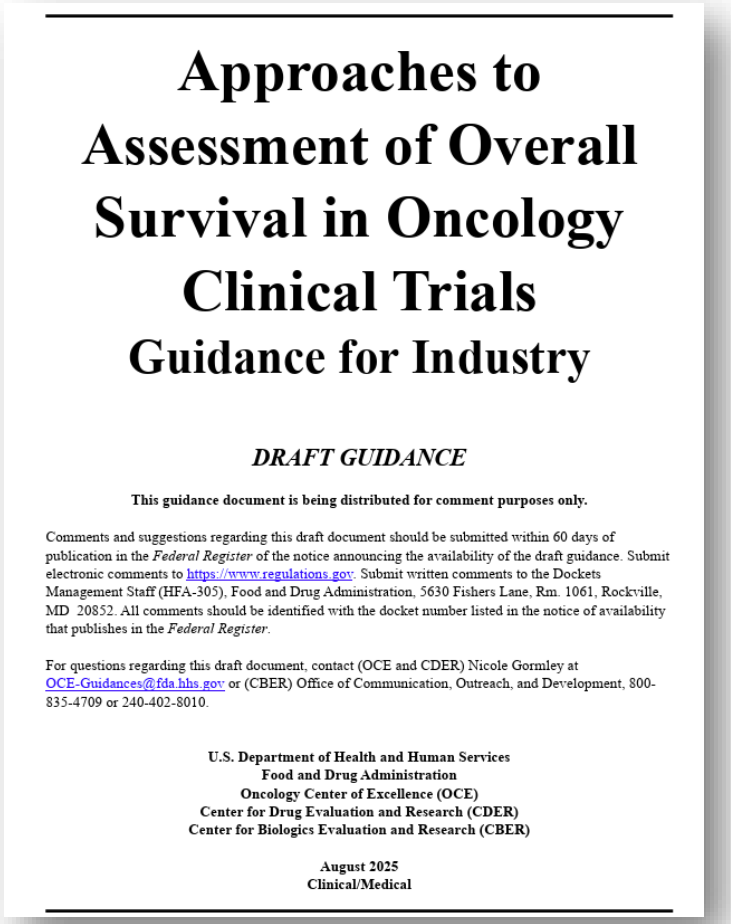


Rodriguez et al, CCR, 2024

Question: How can we best analyze and interpret interim OS?

Regulatory and Scientific Need for A Structured Approach

FRIENDS of CANCER RESEARCH



Approaches to Assessment of Overall Survival in Oncology Clinical Trials

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (OCE and CDER) Nicole Gormley at OCE-Guidances@fda.hhs.gov or (CBER) Office of Communication, Outreach, and Development, 800-835-4709 or 240-402-8010.

U.S. Department of Health and Human Services
Food and Drug Administration
Oncology Center of Excellence (OCE)
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

August 2025
Clinical/Medical

- Pre-specify how OS will be assessed as a safety endpoint.
- Define timing, number of events, information fraction, and harm thresholds.
- Justify thresholds and assumptions prospectively.
- Use simulations or other calculations when OS is immature or uncertain.

From White Paper Concept to Simulation Project

FRIENDS of CANCER RESEARCH

White Paper proposed simulation framework for interim OS interpretation

Working group established to conduct a simulation study

FRIENDS of CANCER RESEARCH

IMAS The difference is in the data.®

AMGEN

AstraZeneca

Bristol Myers Squibb

Crossbow therapeutics

Daiichi-Sankyo

EMD SERONO

Fred Hutch Cancer Center

Genentech A Member of the Roche Group

MERCK

NMDgroup CLARITY – DATA & ANALYTICS AT WORK FOR YOU

NOVARTIS

Pfizer

FDA

GSK

Johnson&Johnson Innovative Medicine

Lilly

Methodology: A Comprehensive, Consistent Scaffold

FRIENDS of CANCER RESEARCH

Design

1. Define the anticipated survival patterns (active vs. control)
2. Define plausible alternative scenarios (sensitivity analyses)
3. Determine candidate decision criteria for identifying harm at the final analysis

Evaluation

4. Evaluate operating characteristics at final analysis under anticipated and alternative scenarios
5. Evaluate interim operating characteristics relative to the final analysis

Simulation Design

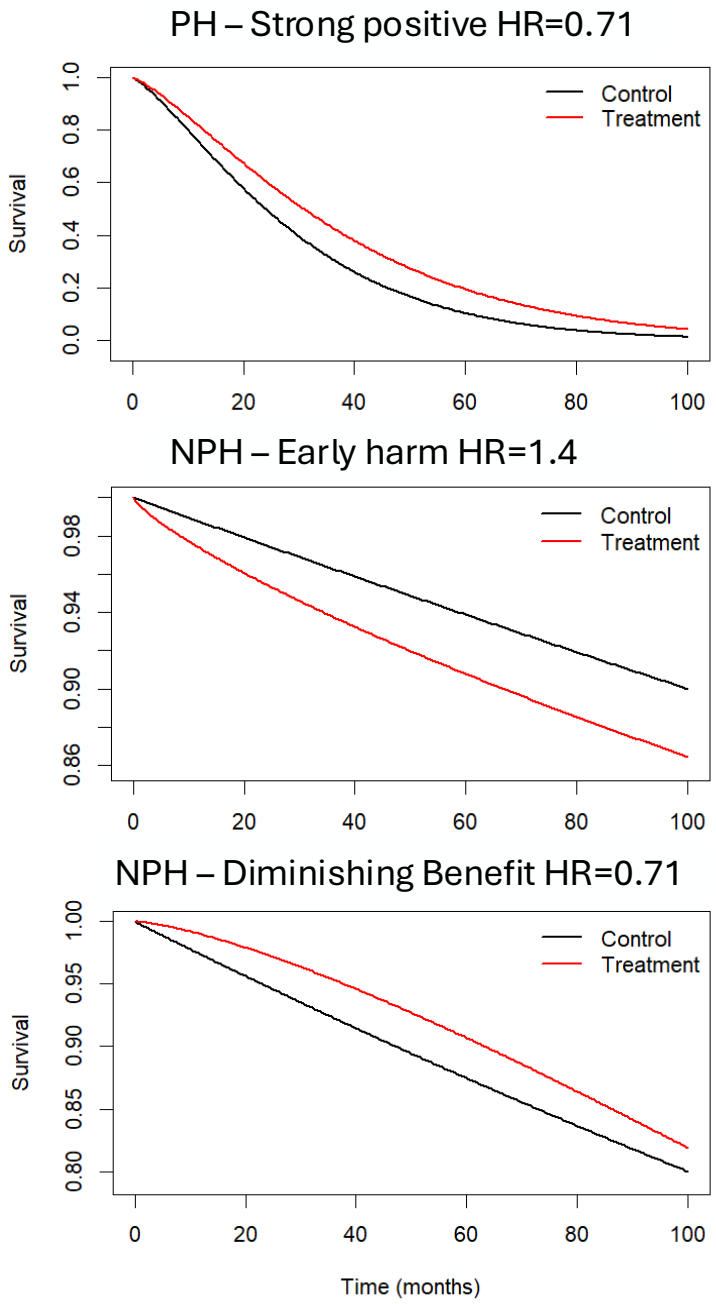
Proportional Hazards

- 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 hazard ratio (HR), due to non-constant HRs



Decision Framework: Defining and Evaluating "Safe" vs "Harm"

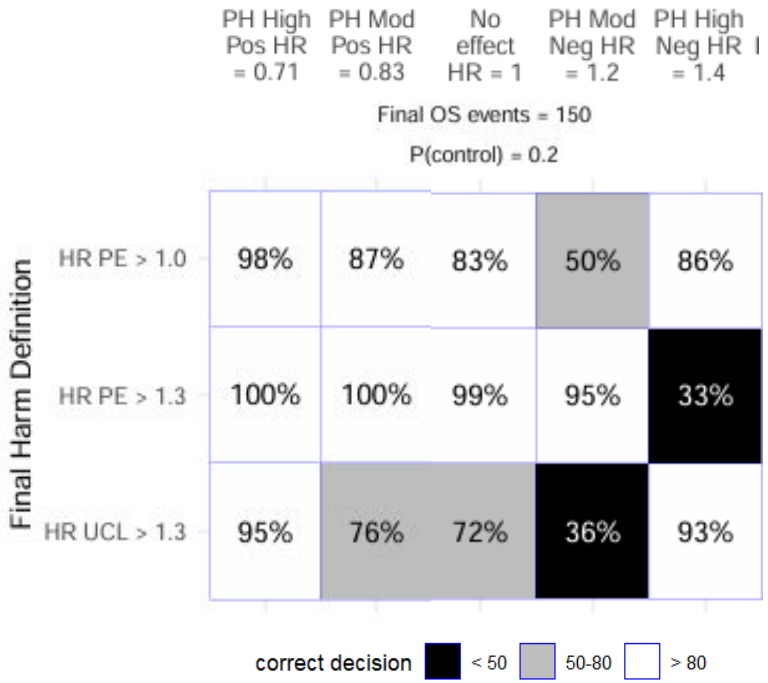
FRIENDS of CANCER RESEARCH

1. Define the decision rules for each scenario:

Final OS Analysis	Interim OS Analysis
• HR point estimate (PE) > 1.0 • HR PE > 1.3 • HR Upper Confidence Limit (UCL) > 1.3	• HR UCL thresholds evaluated from 1.1–1.8 across 10%–50% information fraction

2. Evaluate operating characteristics

Evaluating Candidate Final OS Harm Definitions



Interpreting the Final OS Table

- Rows: candidate final OS harm definitions
- Columns: plausible underlying truths
- Cells: % of simulations with the correct final decision

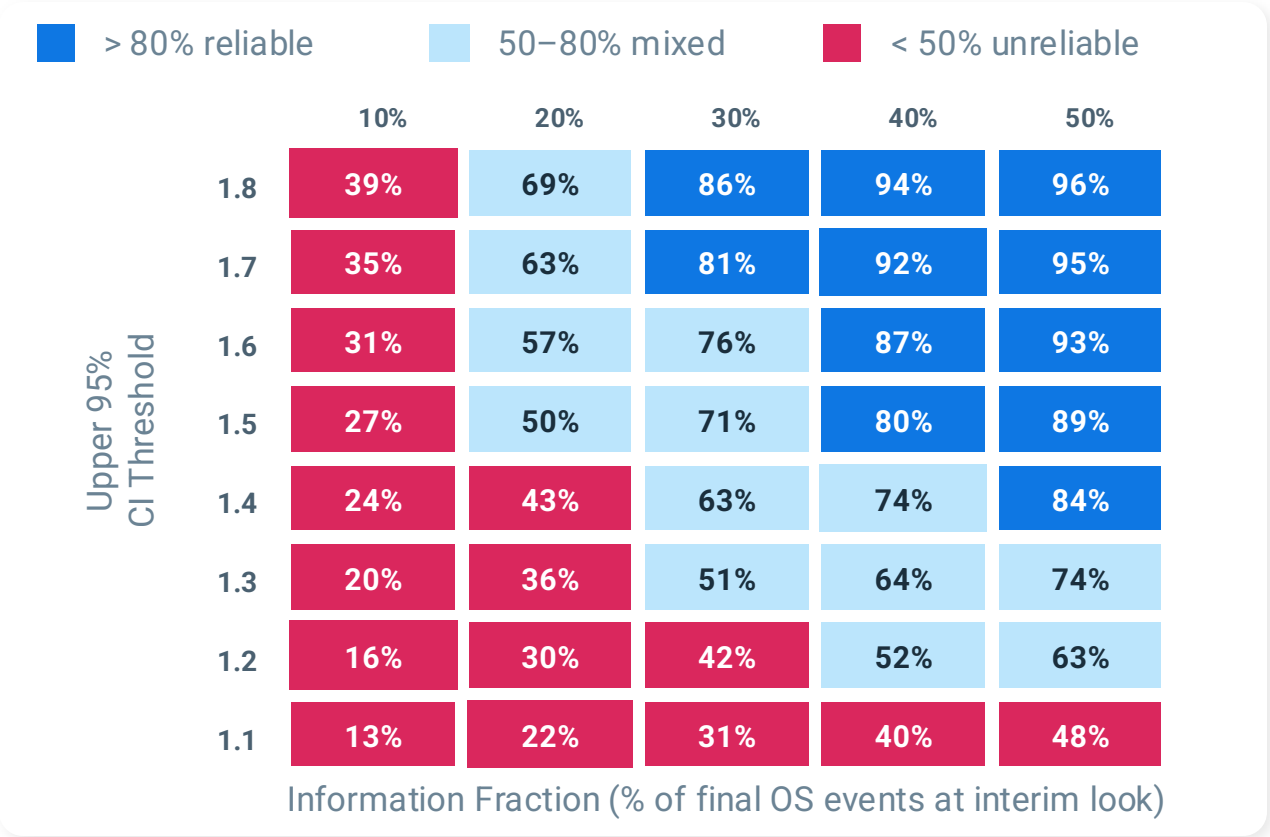
Preliminary Interpretation

- HR PE > 1.0 and HR UCL > 1.3 performed reasonably well overall
- HR PE > 1.3 was less robust overall, though performance varied by scenario

Evaluating Interim Performance Across Thresholds

FRIENDS of CANCER RESEARCH

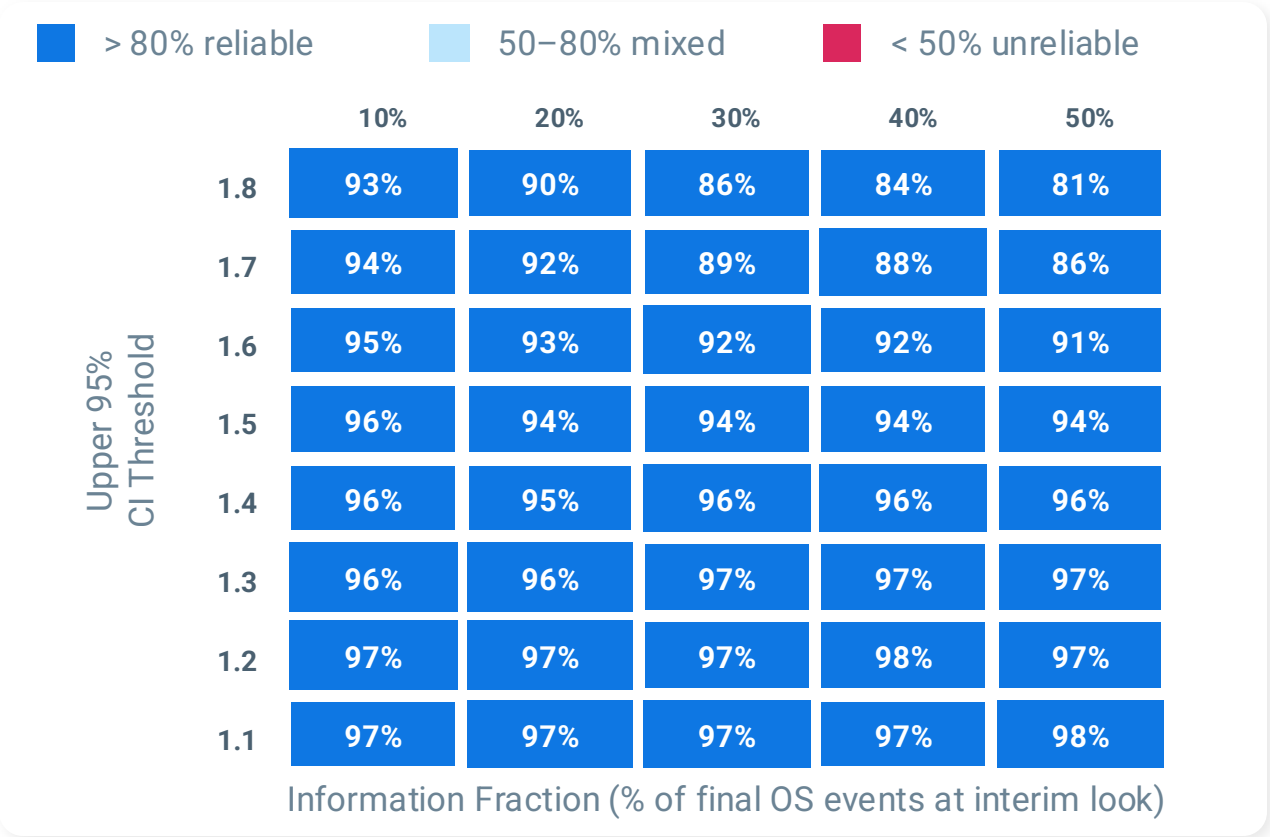
Strong benefit scenario (HR = 0.71) | Final OS = 150 events | Event Rate = 0.2



- Very early interim looks are generally unreliable
- Reliability improves with increasing information fraction
- Threshold choice matter; lower thresholds tend to overcall harm, higher thresholds may miss it

Evaluating Interim Performance Across Thresholds

Strong Harm Scenario (HR = 1.4) | Final OS = 150 events | Event Rate = 0.2

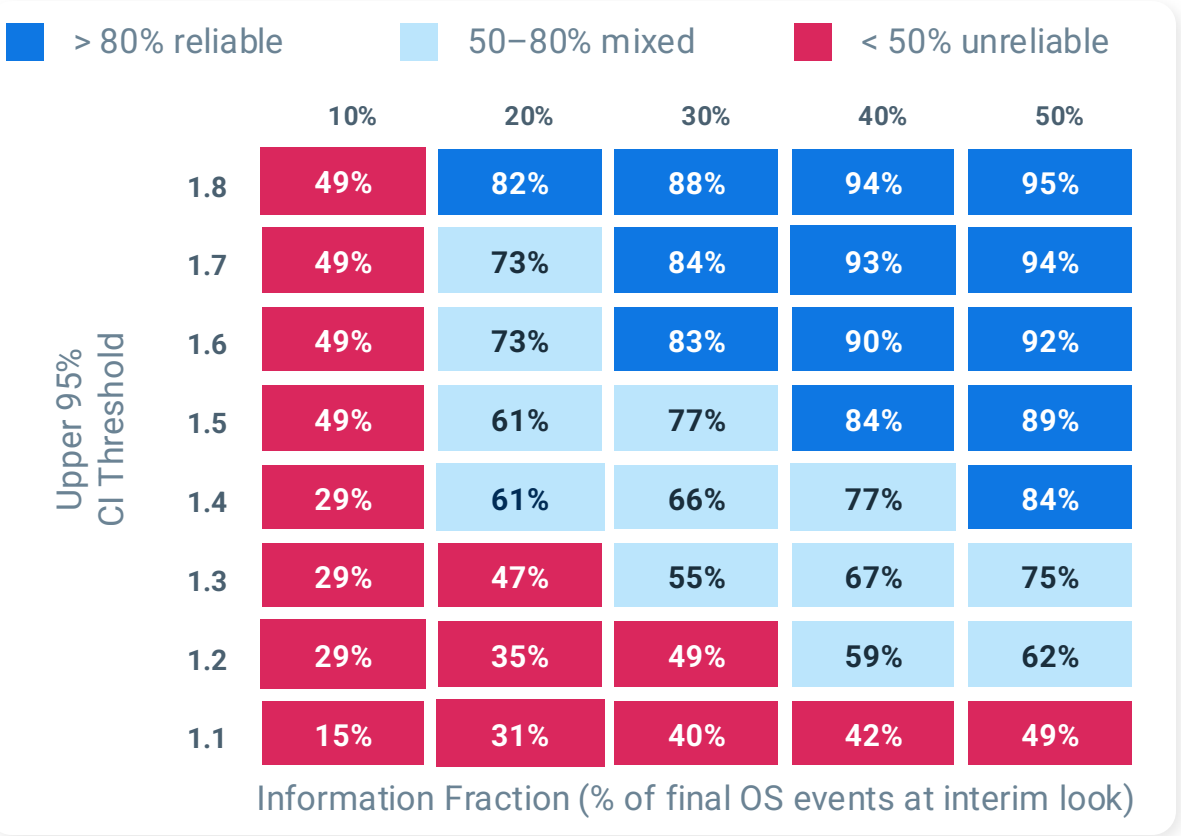


- When true harm is present and proportional, interim OS reliably detects it regardless of threshold or information fraction
- Performance is consistently high (>80%) across all threshold and timing combinations, suggesting the decision framework is robust under clear harm scenarios

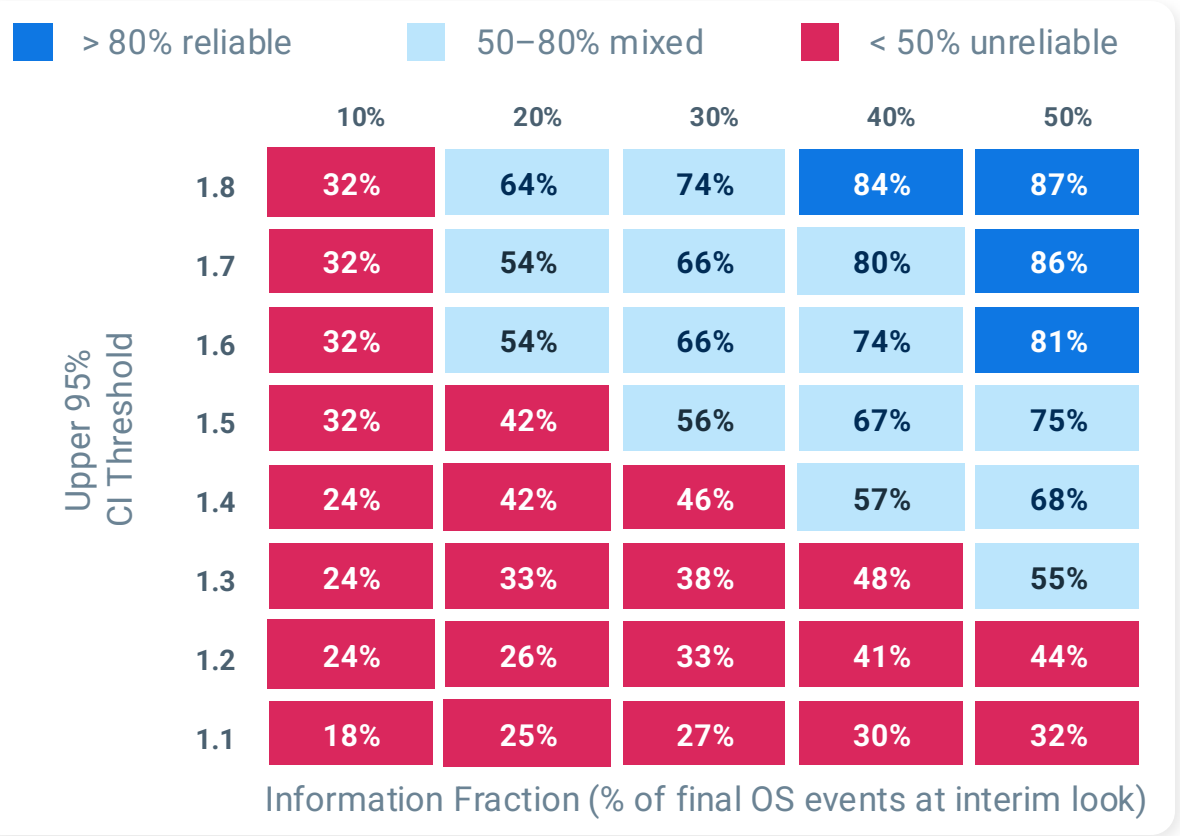
Evaluating Interim Performance Across Thresholds

FRIENDS of CANCER RESEARCH

Diminishing benefit scenario (NPH HR = 0.83) | Final OS = 150 events | Event Rate = 0.2



Delayed benefit scenario (NPH HR = 0.83) | Final OS = 150 events | Event Rate = 0.2



What the Analysis Tells Us

Final harm definitions behave differently across plausible truths

Some definitions perform well in benefit or harm scenarios but less well under null or modest harm scenarios.

Timing and interim threshold choice shape the balance of risks

Very early interim looks can be unstable regardless of threshold. Lower interim thresholds may overinterpret harm; higher thresholds may miss emerging harm signals.

Scenario context matters

Interpretation should account for the expected survival pattern, event count, follow-up, crossover, and whether non-proportional hazards are plausible.

What this Means in Practice

- **Prospectively plan for interim OS monitoring**
  - Recognize early in trial design whether interim OS is appropriate given the indication, expected survival pattern, event count, and decision context.
- **Simulate planned interim timings and thresholds**
  - Evaluate expected event counts, information fractions, and plausible alternative survival scenarios before results are observed.
- **Pre-specify and justify the framework**
  - Define harm thresholds, operating characteristics, assumptions, and interpretation plans in the protocol/SAP.
- **Use interim OS as part of the totality of evidence**
  - Interpret alongside early efficacy, safety, follow-up, crossover, and clinical context.