Session 3: Post-hoc Analyses of Overall Survival Endpoints

MODERATOR:
Jonathon Vallejo, PhD
U.S. Food & Drug Administration

INTRO:
Qing Xu, PhD
U.S. Food & Drug Administration

OVERVIEW OF RECOMMENDATIONS SESSION CHAIR:
Steven Snapinn, PhD, FASA
Seattle-Quilcene Biostatistics LLC

ADDITIONAL PANELISTS:
• Wenjuan Gu, PhD, FDA
• Bindu Kanapuru, MD, FDA
• Paz Vellanki, MD, PhD, FDA
• Grzegorz Nowakowski, MD, Mayo Clinic
• Qian Shi, PhD, Mayo Clinic
• Emmanuel Zuber, PhD, Novartis Pharma AG
Session III: Post-hoc Analyses of Overall Survival Endpoint

FDA-AACR-ASA Workshop
July 18, 2023
Qing Xu Ph.D
Acknowledgement

• Lisa Rodriguez
• Jonathon Vallejo
• Yuan-li Shen
• Nicole Gormley
Outline

• Scope of OS analysis plans
• Limitations of post-hoc OS analyses
• Post-hoc assessments of efficacy, safety and potential for harm with examples
• Discussion
Scope of OS Statistical Analysis Plans

Overall Survival Analysis

Pre-planned Analysis
- Appropriate sample size calculation
  - Type I error control
- Inference can be made

No pre-planned analysis
- No appropriate sample size calculation, no Type I error control
  - Session III
- The results will be considered descriptive in nature

Not a pre-planned analysis
- No formal hypothesis testing
- The results will be considered descriptive in nature
Limitations of Non-prespecified OS Analysis

- OS may not be formally tested
- Concern of insufficient sample size for robust benefit-risk assessment
  - Lack of power for OS efficacy assessment
  - Potential insufficient information at the time of PFS efficacy analysis
  - High level of uncertainty
- Concern of lack of Type I error control for efficacy evaluation
- Concern of insufficient follow-up time (i.e., “immature OS”)
- Concern of impact of confounding factors
  - Subsequent therapy or crossover, competing risks, informative censoring, etc.
- Challenging to interpret OS results
- Limited ability to draw definitive conclusions
Post-hoc Efficacy Assessment of OS
Post-hoc Efficacy Assessment of OS

- Results are considered as descriptive
- Usual treatment effect assessment methods are used
  - e.g., stratified Log-rank test, Stratified Cox model
- Additional exploratory analyses are included
  - Assess non-proportional hazard
  - Assess the impact of informative censoring
  - Assess the impact of intercurrent event
  - Other descriptive analyses can be helpful for interpretation
    - e.g., Competing risks, complex statistical model and so on
DUO Trial ODAC: Unplanned Post-hoc Analysis Performed for Impact of Crossover on OS

**Primary endpoint:**
- PFS per IRC

**Key secondary endpoint:**
- ORR per IRC, OS

<table>
<thead>
<tr>
<th>Method</th>
<th>Overall Survival HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>5-year OS - ITT</td>
<td>1.09 (0.79, 1.51)</td>
</tr>
<tr>
<td><strong>Model-Based Survival Analyses</strong></td>
<td></td>
</tr>
<tr>
<td>MSM(^1)</td>
<td>1.06 (0.72, 1.59)</td>
</tr>
<tr>
<td>RPFTM(^2)</td>
<td>1.22 (0.88, 1.67)</td>
</tr>
</tbody>
</table>

- Results are consistent with the primary OS analysis
- Analyses supportive of potential OS detriment

https://www.fda.gov/advisory-committees/advisory-committee-calendar

\(^1\)MSM-IPTW: Marginal structural model
\(^2\)RPFTM: Rank preserving failure time
Post-hoc Safety Assessment of OS
Post-hoc Safety Assessment of OS

• **OS as a safety endpoint**
  – Descriptive analyses
    • e.g., landmark approaches, event rates, HR with it 95% CI
  – No sufficient power regardless pre-planned or post hoc
    • Can the trial rule out the “harm” based on the study design with insufficient OS information?
    • Can the “harm” be defined based on different level of uncertainties?
    • Whether seeking safety signals alone is sufficient for safety evaluation?
  – What information is available to understand any potentially detrimental OS effect?
    • Narratives, causality of adverse events, disease context, degree of uncertainties, both clinical and statistical considerations and judgements?
  – If the trial is ongoing, how should the next OS analysis time point be selected?
Concerning Potential OS Detriments-Level of Uncertainties

### PI3k ODAC

<table>
<thead>
<tr>
<th>Study</th>
<th>Primary Endpoint</th>
<th>OS as Secondary Endpoint?</th>
<th>Detailed OS Plan?</th>
<th>Deaths PI3Ki arm</th>
<th>Deaths Control arm</th>
<th>Hazard Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>312-0123</td>
<td>PFS</td>
<td>Yes</td>
<td>No</td>
<td>8% (12/157)</td>
<td>3% (4/154)</td>
<td>3.34 (1.08, 10.39)</td>
</tr>
<tr>
<td>313-0124</td>
<td>PFS</td>
<td>Yes</td>
<td>No</td>
<td>5% (10/191)</td>
<td>1% (1/95)</td>
<td>4.74 (0.6, 37.12)</td>
</tr>
<tr>
<td>313-0125</td>
<td>PFS</td>
<td>Yes</td>
<td>No</td>
<td>8% (27/320)</td>
<td>6% (9/155)</td>
<td>1.51 (0.71, 3.23)</td>
</tr>
<tr>
<td>DUO</td>
<td>PFS</td>
<td>Yes</td>
<td>No</td>
<td>50% (80/160)</td>
<td>44% (70/159)</td>
<td>1.09 (0.79, 1.51)</td>
</tr>
<tr>
<td>CHRONOS-3</td>
<td>PFS</td>
<td>Yes</td>
<td>No</td>
<td>18% (56/307)</td>
<td>21% (32/151)</td>
<td>0.87 (0.57, 1.35)</td>
</tr>
</tbody>
</table>

While OS information is early, the same pattern indicating the potential for harm has been observed across multiple trials for one class of drugs.
The Stability of the estimate improves with a greater number of events observed.

Level of Uncertainties - Insufficient Information Could Result in Random High or Random Low Bias

Zhang 2012

Later times indicate more observed events.
Further Assessment of Safety for Potential Harm with Uncertainties
Potential Methods for Assessment of “Harm” with Insufficient Information

• Probability with assumption-based approaches
  • Predictive Probability
  • Bayesian Probability
  • Conditional probability
  • Proposal from session II presentation

• The concept of the methods involves utilizing the observed information to predict the outcomes of future data under different assumptions
Challenges of Probability Approaches for Assessment of “Harm”

Challenge of defining marginal OS benefit HR: 0.6, 0.7...

Challenge of defining degree of uncertainties OS IF*: 30%, 40%....

Results are highly sensitive to the assumptions

Challenge of defining OS harm boundary HR>1, 1.1...

Take into account disease backgrounds and treatment landscapes

Interpretation of probability(>0.3...) can vary based on different individuals

Keynotes

- It is often unclear what assumptions are appropriate to make
- The methods do not account for the shapes of the future survival curves
- Given the limitations presented, we are open to discussing the use of these methods for confirmative trials.

*IF: Information Fraction
Example: Post-Hoc Assessment of Harm using Predictive Approach
Polatuzumab Vedotin ODAC

Assumption: HR=0.8; target death event=631;
Observed OS event=131, Observed HR=0.94, *IF=21%

<table>
<thead>
<tr>
<th>Estimated follow-up</th>
<th>2 years after final OS analysis</th>
<th>5 year after final OS analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projected # of events (Estimated IF)</td>
<td>196 (31%)</td>
<td>279 (44%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probability of Point Estimate of HR &gt;1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumed HR=0.9</td>
</tr>
<tr>
<td>Assumed HR=0.94</td>
</tr>
<tr>
<td>HR=1.00</td>
</tr>
</tbody>
</table>

Results are sensitive to the assumptions

Interpretation is challenging
- Any concern regarding HR=0.94 when IF~20%?
- How high of a probability is considered too high?
- What is the comparator product?
  - Has it established an OS benefit?
- What will be the survival shapes for the future data?

*IF: information fraction

https://www.fda.gov/advisory-committees/advisory-committee-calendar
Summary

• The Agency considers OS as both an efficacy and safety endpoint
• Results of the post-hoc assessment are considered descriptive
• Challenges in the interpretation of OS estimates arise when there is insufficient OS information
  – The totality of the evidence should be considered for the OS safety evaluation
  – Capturing relevant information is important for post-hoc analysis at design stage.
Session 3: Post-hoc Analyses of Overall Survival Endpoints

Steven Snapinn, PhD, FASA
Seattle-Quilcene Biostatistics LLC, Seattle WA
Disclosure Information

Overall Survival in Oncology Clinical Trials
July 18, 2023 | Bethesda, Maryland

Steven Snapinn, PhD

I have the following relevant financial relationships to disclose:

As a consulting biostatistician, I have consulting contracts with multiple pharmaceutical companies, including companies that are developing oncology treatments for FDA approval.
Session 3: Post-hoc Analyses of Overall Survival Endpoints

MODERATOR:
Jonathon Vallejo, PhD
U.S. Food & Drug Administration

INTRO:
Qing Xu, PhD
U.S. Food & Drug Administration

OVERVIEW OF RECOMMENDATIONS
SESSION CHAIR:
Steven Snapinn, PhD, FASA
Seattle-Quilcene Biostatistics LLC

ADDITIONAL PANELISTS:
- Wenjuan Gu, PhD, FDA
- Bindu Kanapuru, MD, FDA
- Paz Vellanki, MD, PhD, FDA
- Grzegorz Nowakowski, MD, Mayo Clinic
- Qian Shi, PhD, Mayo Clinic
- Emmanuel Zuber, PhD, Novartis Pharma AG
Session 3 Objectives

- To discuss analysis methods to assess for potential harm when no pre-specified assessments of OS were planned, and the interpretation of such post-hoc analyses.
- To assess what would be acceptable degrees of uncertainty in post-hoc assessments of OS, considering the amount of follow-up information available and disease epidemiology.
Quantitative Approach to Incorporate Belief in Surrogacy

Hypothetical PFS Results

Modeled on PFS Results for Keytruda in Patients with cHL (KEYNOTE-204)

Table 62: Efficacy Results in Patients with cHL in KEYNOTE-204

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>KEYTRUDA 200 mg every 3 weeks n=151</th>
<th>Brentuximab Vedotin 1.8 mg/kg every 3 weeks n=153</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients with event (%)</td>
<td>81 (54%)</td>
<td>88 (58%)</td>
</tr>
<tr>
<td>Median in months (95% CI)*</td>
<td>13.2 (10.9, 19.4)</td>
<td>8.3 (5.7, 8.8)</td>
</tr>
<tr>
<td>Hazard ratio * (95% CI)</td>
<td>0.65 (0.48, 0.88)</td>
<td></td>
</tr>
<tr>
<td>p-Value †</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td>Objective Response Rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORR * (95% CI)</td>
<td>66% (57, 73)</td>
<td>54% (46, 62)</td>
</tr>
<tr>
<td>Complete response</td>
<td>25%</td>
<td>24%</td>
</tr>
<tr>
<td>Partial response</td>
<td>41%</td>
<td>30%</td>
</tr>
<tr>
<td>Duration of Response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median in months (range) *</td>
<td>20.7 (0.0+, 33.2+)</td>
<td>13.8 (0.0+, 33.9+)</td>
</tr>
</tbody>
</table>

* Based on Kaplan-Meier estimates.
† Based on the stratified Cox proportional hazard model.
‡ Based on stratified log-rank test. One-sided p-Value, with a prespecified boundary of 0.0043.
§ Difference in ORR is not statistically significant.
+ Denotes a censored value.
Quantitative Approach to Incorporate Belief in Surrogacy

PFS Results Lead to a Belief in an OS Benefit

- Potential Model for OS HR Based on PFS Result:
  - $\log{\text{OS-HR}} \mid \log{\text{PFS-HR}} \sim N(2/3, 0.25^2)$
Quantitative Approach to Incorporate Belief in Surrogacy

Approval Decisions Can Be Based on Believed OS Benefit

FDA-AACR-ASA WORKSHOP: OVERALL SURVIVAL IN ONCOLOGY CLINICAL TRIALS
Quantitative Approach to Incorporate Belief in Surrogacy

Model Allows Incorporation of Observed OS Data

- Suppose 6 deaths occurred
  - 5 Deaths in Active Group
  - 1 Death in Control Group
Quantitative Approach to Incorporate Belief in Surrogacy

PFS Result + Observed OS Data Leads to Revised Probabilities

FDA-AACR-ASA WORKSHOP: OVERALL SURVIVAL IN ONCOLOGY CLINICAL TRIALS
When is OS NOT a pre-specified endpoint?

- **Pre-specification**
  - Enables confirmatory conclusions to be drawn
  - Ensures robust data collection/cleaning

- **Mortality data should always be evaluated, but there does not always need to be a pre-specified testing strategy**
  - e.g., when the number of expected deaths is small (low power)

- **Need to specify the intended clinical benefit of the treatment**
  - When the intended clinical benefit is something other than OS (e.g., QoL)
    - OS analyses required to rule out harm
    - Tumor response and PFS are not typically clinical benefits
  - When OS is the intended clinical benefit
    - Need to demonstrate a benefit on a likely surrogate (e.g., PFS)
    - OS analyses required to ensure consistency with surrogate
Post-hoc Questions

What analyses of OS should be conducted post-hoc, when there were no pre-specified analyses to assess for harm?

- Need to include
  - A thorough description of mortality data, including narratives and graphical displays
  - Various measures of uncertainty on the treatment effect
  - Summaries of relevant additional information (internal and external)

- Need to recognize that, with small numbers, the effect estimates are unreliable

- Various specific measures of treatment effect on OS may be considered
  - Measures that consider the full survival distributions (e.g., the hazard ratio, the difference in restricted mean survival) have advantages
Post-hoc Questions

Should harm be defined similarly to how it is defined when it is a pre-specified endpoint, or are separate thresholds or approaches needed for post-hoc assessments of harm?

- Clinical trials provide estimates of the true treatment effect
  - True treatment effect: The definition of harm depends on clinical context, not on observed data
  - Treatment effect estimate: The definition of harm needs to take uncertainty into account
- Threshold for harm should be evaluated in the context of the evidence for benefit on other measures (e.g., PFS, adverse events, QoL)
  - The specific comparator matters (active vs placebo/observation)
  - Should be driven by clinically relevant considerations
  - Could develop a grid of thresholds
In the setting of indolent diseases or any settings in which the OS information is early (i.e., the survival data may be considered “immature”), what amount of uncertainty in the OS information would be acceptable?

### Considerations

- What level of information/certainty could reasonably be expected from a feasible drug development plan in this disease setting?
- What has been understood on the observed death events and likely mortality process?
  - Non-proportion hazards complicates this assessment
- The acceptability of uncertainty regarding an OS harm should be appreciated in the context of a benefit/risk discussion
- What is the type of decision for which this uncertainty is assessed (e.g., conditional vs. full approval vs research/non-regulatory contexts)
In what scenarios would obtaining reliable post-hoc evaluations of OS be infeasible or impossible? If continued follow-up is infeasible or impossible to adequately assess OS for a given trial, what analyses would be helpful (e.g., use of simulated data)?

- Intercurrent events make reliable evaluations of OS problematic
  - When the intercurrent event represents a real-world scenario, such as a switch to standard of care in nonresponders, the treatment policy estimand is recommended
  - In other situations, such as cross-over from the control regimen to the experimental regimen following disease progression, the treatment policy estimand can be problematic

- The use of simulation to investigate diverse assumption (e.g., correlation between survival and other endpoints) can be helpful
  - This could include “tipping point” analyses (e.g., determine the minimum degree of surrogacy that would be required for the PFS benefit to translate into an OS benefit)
Other Considerations

- Use of data beyond date of death (e.g., cause of death, other therapies) to inform a credible post-hoc evaluation of OS
- Recommended methods for assessing the impact of intercurrent events
- When and how should external information be leveraged in assessing OS
- Measures of treatment benefit
- Quantitative Bayesian approach combine OS data with surrogate (PFS) data
Measure of Treatment Benefit

- The hazard ratio is commonly used, but has been criticized and is undefined in the case of nonproportional hazards.

- Other potential measures include:
  - Generalized hazard ratio
  - Difference in medians
  - Difference in landmark survival
  - Difference in restricted mean survival time
Recent paper recommends the use of the generalized hazard difference

- Also known as the difference in exposure-adjusted subject incidence rates
- Inverse is the NYNT: Number of patient-years of treatment to result in one fewer event

Unlike other measures, accounts for two distinct dimensions of benefit

Reference: Treatment effect measures under nonproportional hazards. Snapinn, Jiang, Ke; Pharmaceutical Statistics 2023
Session 3: Post-hoc Analyses of Overall Survival Endpoints

MODERATOR:
Jonathon Vallejo, PhD
U.S. Food & Drug Administration

INTRO:
Qing Xu, PhD
U.S. Food & Drug Administration

OVERVIEW OF RECOMMENDATIONS
SESSION CHAIR:
Steven Snapinn, PhD, FASA
Seattle-Quilcene Biostatistics LLC

ADDITIONAL PANELISTS:
• Wenjuan Gu, PhD, FDA
• Bindu Kanapuru, MD, FDA
• Paz Vellanki, MD, PhD, FDA
• Grzegorz Nowakowski, MD, Mayo Clinic
• Qian Shi, PhD, Mayo Clinic
• Emmanuel Zuber, PhD, Novartis Pharma AG
1. When is OS NOT a pre-specified endpoint?

2. What analyses of OS should be conducted post-hoc, when there were no pre-specified analyses to assess for harm?

3. Should harm be defined similarly to how it is defined when it is a pre-specified endpoint, or are separate thresholds or approaches needed for post-hoc assessments of harm?

4. In the setting of indolent diseases or any settings in which the OS information is early (i.e., the survival data may be considered “immature”), what amount of uncertainty in the OS information would be acceptable?

5. In what scenarios would obtaining reliable post-hoc evaluations of OS be infeasible or impossible? If continued follow-up is infeasible or impossible to adequately assess OS for a given trial, what analyses would be helpful (e.g., use of simulated data)?