

Session 4: Subgroup Considerations

MODERATOR:

Ruixiao Lu, PhD

Alumis Inc./American Statistical Association

INTRO:

Anup Amatya, PhD

U.S. Food & Drug Administration

OVERVIEW OF RECOMMENDATIONS

SESSION CHAIR:

Keith Flaherty, MD, FAACR

MGH Cancer Center

ADDITIONAL PANELISTS:

- **Harpreet Singh, MD, FDA**
- **Jianjin Xu, PhD, FDA**
- **Elizabeth Garrett-Mayer, PhD, ASCO**
- **Marjorie C. Green, MD, Merck**
- **Mary Redman, PhD, Fred Hutchinson Cancer Center**

Subgroup analysis of Overall Survival in benefit-risk assessment

Anup Amatya, Ph.D.

Lead Mathematical Statistician, FDA



Disclosure Information

- No conflicts of interest to disclose.

Key points

- ❖ Subgroup analysis of OS is important but challenging
- ❖ Pre-specification of expectations and analyses plan increases credibility and provide opportunities for better data collection
- ❖ In absence of pre-specification, interpretation of subgroup results will be largely driven by the context of trial design, disease setting and scientific rationale.

If trial demonstrates benefit only in subgroup (e.g., biomarker positive), FDA may approve narrower indication than overall enrolled population

Subgroup analysis is important

- ❖ Clinical trial population is heterogeneous
- ❖ Overall analysis masks extent of benefit-risk in subgroups of patients.
 - Benefit limited to certain patient population (e.g., a biomarker positive patients)
 - Higher toxicity in certain patient population (e.g., a biomarker negative patients)

Subgroup analysis is important to examine consistency of observed benefit, confirm expected benefit/harm, and uncover clinical clues in subsets of patients studied in a clinical trial.

Subgroup analysis is challenging

❖ **Small sample size**

- Increases variability in treatment effect estimates
- Increases chances of imbalance in baseline characteristics
- Reduces power to conduct formal statistical testing

❖ **Multiplicity**

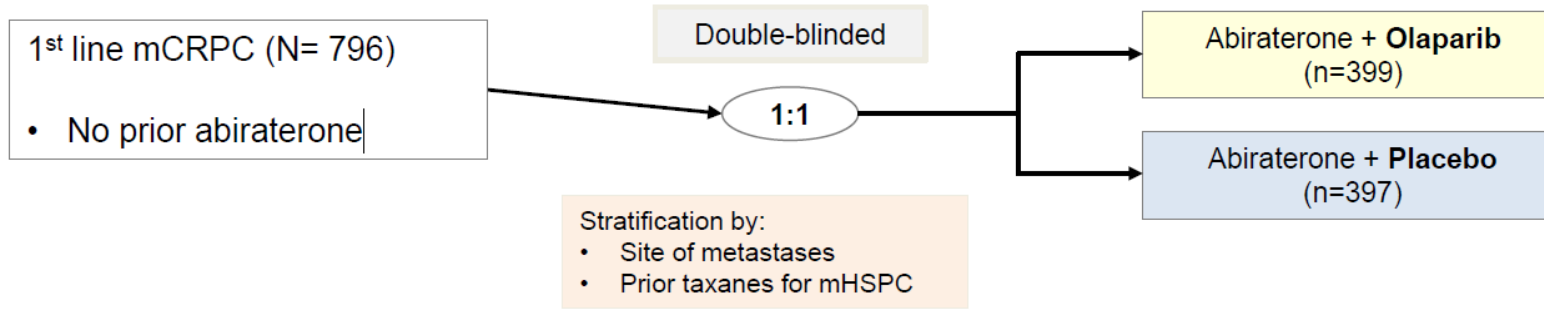
- Increases the chances of observing heterogeneous treatment effect
- Increases the chances of discordant results across different endpoints

❖ **Lack of biological rationale**

- Diminishes credibility

Challenging to determine how concerning is the observed differences in the subgroups

PROpel study design



Primary endpoint: rPFS by investigator
Key secondary endpoints: OS

- No stratification by BRCA status
- Post-hoc analysis by BRCA status

| | All patients (ITT) | BRCAM (11%) | Undetermined BRCA status (35%) | Non-BRCAM (54%) |
|-------------------|--------------------|--------------------------|--------------------------------|--------------------------|
| n (O vs P) | | 47, 38 | 138, 146 | 214, 213 |
| rPFS, HR (95% CI) | 0.67 (0.56, 0.81) | 0.24 (0.12, 0.46) | 0.66 (0.46, 0.94) | 0.85 (0.66, 1.11) |
| OS, HR (95% CI) | 0.81 (0.67, 1.00) | 0.30 (0.15, 0.60) | 0.73 (0.52, 1.03) | 1.06 (0.81, 1.39) |

- Efficacy in overall population is largely attributed to the effects of *BRCAM*
- Consistency across two trials (PROpel and Study 8) and ovarian cancer trials
- Inadequate retrospective assessment of *BRCAM* status.
- No stratification by *BRCAM* status.
- No pre-specified formal analysis by *BRCAM* status.

Subgroup analyses

- ❖ **When there is** a persuasive statistical evidence of benefit in overall population
 - To identify the subgroup(s) that benefit the most
 - To further explore subgroups that did not appear to have any benefit
 - May be used to restrict indication in cases of limited efficacy and/or potential OS detriment especially when external data raises concern for harm ²
 - ❑ e.g., Eribulin indication limited to liposarcoma histology in Soft tissue sarcomas
 - ❑ e.g., Atezolizumab indication limited to the PD-L1+ Triple-Negative Breast Cancer

These analyses should be interpreted with caution

Considerations for interpretation

(of subgroup OS results)



- ❖ Biological rationale
- ❖ Study design: head-to-head vs. add-on
- ❖ Therapy: non-targeted vs targeted, monotherapy vs. combination
- ❖ Disease: Indolent vs. aggressive
- ❖ Endpoints: tumor response vs. clinical benefit (feel, function, survive)
- ❖ Maturity of the data and evidence from other endpoints
- ❖ Evidence of benefit/harm from other trials
- ❖ New knowledge of predictive value and prevalence of the biomarker-negative population
- ❖ Unmet need and availability of data in biomarker-negative population
- ❖ Additional post-hoc analyses

Regulatory Considerations

- ❖ Grant broad approval to overall study population
 - Include subgroup results in the labeling?
 - Limitation of use on the labeling?
 - Issue post-marketing requirement?
- ❖ Restrict approval to a subgroup (e.g., Biomarker+ subgroup)
 - Include a complementary subgroup (e.g., biomarker- subgroup) results in the labeling?
- ❖ Complete Response (i.e., not approve for marketing)

Subgroup analyses

- ⚠ **When there is no** persuasive statistical evidence of benefit in overall population
 - Identify one or more subgroups with some apparent benefit
 - Not suitable for efficacy claim
 - Verification in additional trial is needed

Can not be used to “rescue” failed trial with efficacy in a subgroup

Trial design considerations

for subgroup analysis

- ❖ Key subgroups of interest based on biological or mechanistic rationale.
- ❖ Objective of the analysis, whether is it for efficacy or safety assessment.
- ❖ Expected treatment effect (or lack thereof) in each subgroups
- ❖ Expected contribution of the subgroups to the overall treatment effect.
- ❖ Sample size required to conduct formal hypothesis testing in the subgroups.
- ❖ Expected enrollment in subgroups and their complementary subgroups.
- ❖ Feasibility of stratification for prognostic or predictive subgroups.
- ❖ Analysis plan including statistical methods, measure of uncertainty, and targeted thresholds.

Pre-specification improves interpretability and credibility

“Merely listing a factor for future subgroup analysis is not regarded as pre-specification of particular interest in that subgroup.”

—Guideline on the investigation of subgroups in confirmatory clinical trials EMA/CHMP/539146/2013

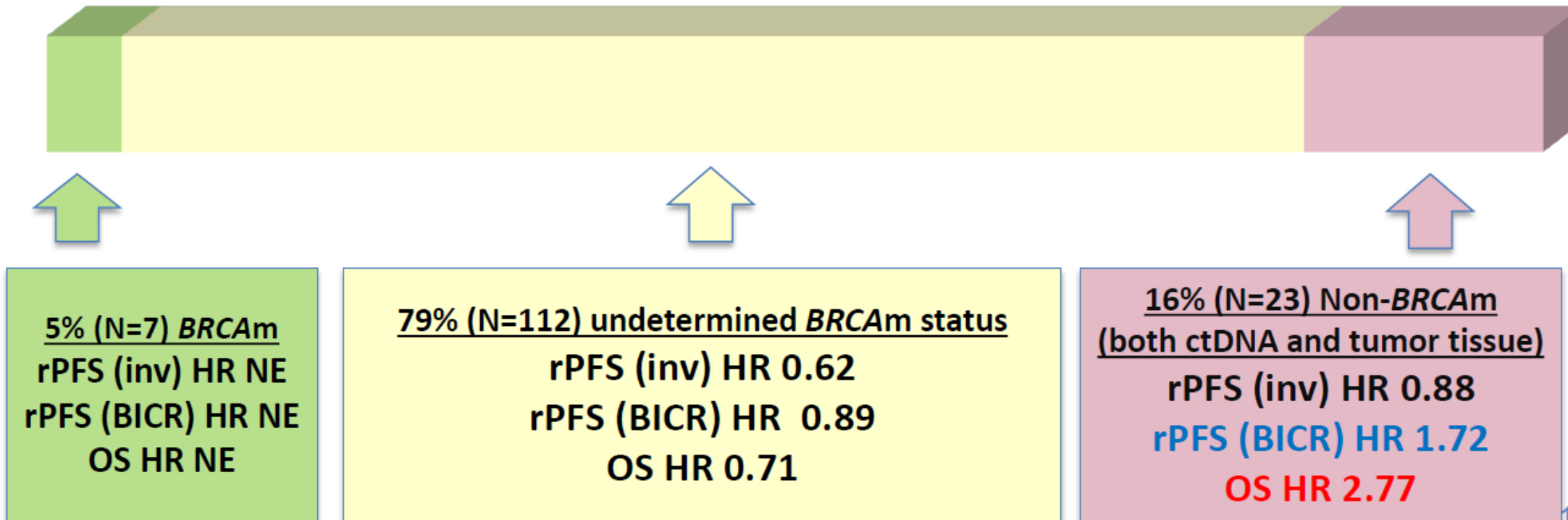
Summary

- ❖ Subgroup analyses of OS are important but challenging to interpret—even more so when based on early analyses of immature data.
- ❖ Pre-specification of OS evaluation plan, rationale, and expectations ensures availability and increases credibility of subgroup findings.
- ❖ Post-hoc subgroup analyses of a failed trial are considered appropriate only for hypothesis generation, not for efficacy claim.
- ❖ Benefit-risk evaluation in the subgroups relies on statistical considerations as well as on the pathophysiological rationale and other practical considerations.



Final Study 8 Results: Similar to PROpel

ITT- rPFS (inv) HR 0.65 (0.44, 0.97); OS HR 0.91 (0.60, 1.38)



NE = Not Evaluable

FDA-AACR-ASA Workshop

OVERALL SURVIVAL IN ONCOLOGY CLINICAL TRIALS

July 18, 2023 | 8 a.m.-5 p.m. | Bethesda, Maryland



Session 4: Subgroup Considerations

Keith T. Flaherty, MD, FAACR

Massachusetts General Hospital Cancer Center, Boston, MA

Join the conversation with
#AACRSciencePolicy



Disclosure Information

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



Keith Flaherty

I have the following relevant financial relationships to disclose:

Employee of: Massachusetts General Hospital

Consultant for: Clovis Oncology, Strata Oncology, Kinnate, and Scorpion Therapeutics, PIC Therapeutics, Apricity, C-Reveal, Tvardi, ALX Oncology, xCures, Monopteros, Vibliome, Soley Therapeutics, Alterome, Immagine, and intrECate, Nextech, Takeda, Novartis, Transcode Therapeutics, and Roche/Genentech

Speaker's Bureau for: N/A

Grant/Research support from: N/A

Stockholder in: Clovis Oncology, Strata Oncology, Kinnate, and Scorpion Therapeutics, PIC Therapeutics, Apricity, C-Reveal, Tvardi, ALX Oncology, xCures, Monopteros, Vibliome, Soley Therapeutics, Alterome, Immagine, and intrECate, Nextech, and Transcode Therapeutics

Honoraria from: N/A

- and -

My additional financial relationship disclosures are: N/A

Session 4: Subgroup Considerations

MODERATOR:

Ruixiao Lu, PhD

Alumis Inc./American Statistical Association

INTRO:

Anup Amatya, PhD

U.S. Food & Drug Administration

OVERVIEW OF RECOMMENDATIONS

SESSION CHAIR:

Keith Flaherty, MD, FAACR

MGH Cancer Center

ADDITIONAL PANELISTS:

- **Harpreet Singh, MD, FDA**
- **Jianjin Xu, PhD, FDA**
- **Elizabeth Garrett-Mayer, PhD, ASCO**
- **Marjorie C. Green, MD, Merck**
- **Mary Redman, PhD, Fred Hutchinson
Cancer Center**

Session Objectives

- To discuss the role of subgroups in relation to the primary ITT analyses and the interpretation of OS subgroup results, regardless of Type I error control.

Discussion Questions

1. How should trial design aspects be pre-specified in order to adequately plan for evaluation of treatment effect within and across subgroups, including sample size and stratification?
2. How should OS estimates be prioritized to assess treatment effect, specifically considering estimates in the overall, pre-specified subgroups, and post-hoc subgroups? And what role do estimates based on Bayesian approaches have, such as shrinkage estimates?
3. If there is a priori information regarding a prognostic or predictive factor, which analyses should be pre-specified in the SAP? Which should be prioritized to evaluate treatment effect?
4. Should there be analyses pre-specified to assess potential harm in pre-defined subgroups?
5. Can the potential for OS harm be assessed in subgroups if this was not a pre-specified analysis, even though the subpopulations may have been pre-defined?
6. What are the considerations for the weight of evidence that should be placed on the OS subgroup results if they are based on post-hoc subgroup analyses?

Subgroup Considerations

- 1. How should trial design aspects be pre-specified in order to adequately plan for evaluation of treatment effect within and across subgroups, including sample size and stratification?**
 - Oncology trials often enroll heterogeneous populations consisting of multiple subpopulations based on either histology, prognostic or predictive molecular characteristics
 - OS integrates efficacy and safety; in cancers with multiple lines of available therapy, toxicity can preclude pursuit of subsequent therapies.
 - Well-validated/known subgroup analysis (ideally stratified, if feasible) should be pre-specified; may require longer follow-up
 - particularly for populations with biological/mechanistic rationale (efficacy) or that may be more vulnerable to adverse effects (e.g., elderly or poor performance status)
 - Objective regarding OS (efficacy or safety) should be pre-specified; power calculation and sample size estimate should be performed if formal hypothesis testing will be conducted for key subgroups.
 - Lower threshold to take non-significant evidence of harm seriously when therapy is associated with life-threatening toxicity

Subgroup Considerations

- 2. How should OS estimates be prioritized to assess treatment effect, specifically considering estimates in the overall, pre-specified subgroups, and post-hoc subgroups? And what role do estimates based on Bayesian approaches have, such as shrinkage estimates?**
 - Biomarker negative subgroups can be quite small and underpowered for OS comparisons with biomarker positive subgroups. The results interpretation based on small subgroup or post-hoc analysis should be cautious. Several studies may be needed to confirm if there is any detrimental effect in OS.
 - If there are (new) subgroup related questions/features that arise during a study, but before analysis is started/database is locked, consideration should be given to modifying the SAP
 - Adaptive design features may also be considered to allow additional enrichment or continued follow up of subset of patients
 - The shrinkage estimate approach assumes there is no a priori expectation of differential OS, which is unlikely to hold for biologically distinct subgroups that may be expected to have differential sensitivity
 - Alternatively, a threshold hazard ratio could be pre-specified to ensure a high likelihood of observing hazard ratio greater than the threshold HR in a subgroup, if the investigational treatment is truly detrimental in a subgroup.

Subgroup Considerations

- 3. If there is a priori information regarding a prognostic or predictive factor, which analyses should be pre-specified in the SAP? Which should be prioritized?**
- If there is no preliminary efficacy data in the biomarker negative subgroup, a study could be designed for enrollment with biomarker positive patients only, a study may also be designed for enrollment of separate cohorts based on biomarker status with early futility analysis
 - Alpha-controlled formal testing of primary endpoint for subgroups based on biomarker status is preferred
 - Pre-specify the magnitude of improvement in primary and key secondary endpoints that would be considered clinically meaningful in both biomarker positive and biomarker negative subgroups
 - A pre-specified power analysis for OS with subgroup assessment could be helpful to decide the timeline of the maturity of the OS data and may facilitate the decision making even after finishing the study for potential longer follow-up for more mature OS data in a specific subgroup.

Subgroup Considerations

4. Should there be analyses pre-specified to assess potential harm in pre-defined subgroups?

- Well validated/routine subgroups (age, gender, routine biomarkers, etc) should carry more stringency to conduct such analysis
- Interpretation of results based on small subgroups should use caution, and detrimental effect of OS may need more than one study
- If death rate is much higher in the treatment group, there is concern of potential harm even without pre-specification
- If the study is in the maintenance setting (when therapy would not be given per SoC) or it's an add-on design where the investigational drug is adding toxicity, in order to mitigate the concern of potential harm, pre-defined subgroup OS analysis should be specified as an interim analysis if early OS analysis is feasible

Subgroup Considerations

- 5. Can the potential for OS harm be assessed in subgroups if this was not a pre-specified analysis, even though the subpopulations may have been pre-defined?**
- For hypothesis driven assessment with less prior clinical evidence, assessment of toxic deaths, DLTs, or severe AEs may be better than OS if the sample size/number of events are low
 - Unlike efficacy signal, safety signal in presence of considerable uncertainty is taken seriously
 - Post-hoc OS subgroup analysis should be conducted for exploratory purposes to remove concern of potential harm if deemed necessary from a clinical perspective or with accumulating new information regardless of pre-specification
 - If suggestion of detriment is observed across multiple studies, it is reasonable to assess potential harm if subpopulations are not pre-defined
 - A single trial should not be considered to provide definitive evidence of harm in a retrospective subset analysis

Subgroup Considerations

- 6. What are the considerations for the weight of evidence that should be placed on the OS subgroup results if they are based on post-hoc subgroup analyses?**
- In general, post-hoc subgroup analysis has lots of limitations and should be viewed as hypothesis generation. However, the investigators/sponsors should evaluate if there's repeated evidence of (harmful) OS effect in the same subgroup from other trials
 - There should be a clear delineation of whether post-hoc subgroup analyses were undertaken for efficacy or safety assessment
 - OS subgroup results from the post-hoc analysis of a failed trial is not adequate to offer substantial evidence of effectiveness
 - Observed safety signal, such as higher rate of death in treatment arm or hazard ratio crossing null value, observed in non-protocol specified analyses of biologically plausible subgroups should raise concern, particularly when combined with evidence from prior trials and corroborating evidence of higher toxicity, may necessitate halting the trial or restricting indication for the safety reasons

Session 4: Subgroup Considerations

MODERATOR:

Ruixiao Lu, PhD

Alumis Inc./American Statistical Association

INTRO:

Anup Amatya, PhD

U.S. Food & Drug Administration

OVERVIEW OF RECOMMENDATIONS

SESSION CHAIR:

Keith Flaherty, MD, FAACR

MGH Cancer Center

ADDITIONAL PANELISTS:

- **Harpreet Singh, MD, FDA**
- **Jianjin Xu, PhD, FDA**
- **Elizabeth Garrett-Mayer, PhD, ASCO**
- **Marjorie C. Green, MD, Merck**
- **Mary Redman, PhD, Fred Hutchinson
Cancer Center**

Session 4 Questions

1. How should trial design aspects be pre-specified in order to adequately plan for evaluation of treatment effect within and across subgroups, including sample size and stratification?
2. How should OS estimates be prioritized to assess treatment effect, specifically considering estimates in the overall, pre-specified subgroups, and post-hoc subgroups? And what role do estimates based on Bayesian approaches have, such as shrinkage estimates?
3. If there is a priori information regarding a prognostic or predictive factor, which analyses should be pre-specified in the SAP? Which should be prioritized to evaluate treatment effect?
4. Should there be analyses pre-specified to assess potential harm in pre-defined subgroups?
5. Can the potential for OS harm be assessed in subgroups if this was not a pre-specified analysis, even though the subpopulations may have been pre-defined?
6. What are the considerations for the weight of evidence that should be placed on the OS subgroup results if they are based on post-hoc subgroup analyses?