Comprehensive Guide: Subgroup Analyses in Clinical Trials

Regulatory Perspectives, Case Examples, and Practical Recommendations

Clinical Research Guidance Document

2025-10-15

Table of contents

# 1. Executive Summary

This document provides comprehensive guidance on conducting, interpreting, and reporting subgroup analyses in confirmatory clinical trials, with particular emphasis on oncology. It synthesizes FDA regulatory perspectives, international guidelines, and real-world case examples to provide practical recommendations for researchers, sponsors, and regulatory reviewers.

|  |
| --- |
| Key Point |
| Subgroup analyses are essential for personalized medicine but require careful planning, rigorous methods, and cautious interpretation to avoid misleading conclusions. |

# 2. Part 1: Foundational Principles

## 2.1 Purpose of Subgroup Analyses

Subgroup analyses assess treatment effects in specific patient subsets within a clinical trial, typically defined by:

* **Demographics:** Age, sex, race, ethnicity
* **Clinical characteristics:** Disease severity, stage, histology, comorbidities
* **Molecular markers:** PD-L1 expression, HRD status, genetic alterations
* **Geographic factors:** Region, study site
* **Baseline risk factors**

### 2.1.1 Primary Goals

1. Identify heterogeneity in treatment effects across patient populations
2. Support regulatory decision-making and labeling
3. Inform clinical practice and individualized treatment decisions
4. Generate hypotheses for future research

## 2.2 Three Categories of Subgroup Analyses

### 2.2.1 Inferential Subgroup Analyses

* Pre-specified with adequate statistical power and alpha control
* Intended to establish efficacy in a specific subgroup
* Requires prospective planning for sample size and multiplicity control
* Used when differential effects are anticipated based on strong biological rationale

### 2.2.2 Supportive Subgroup Analyses

* Pre-specified but not powered for formal hypothesis testing
* Descriptively assess consistency of treatment effects across subgroups
* Conducted after demonstrating significant effect in total population
* Help ensure broad applicability of findings

### 2.2.3 Exploratory Subgroup Analyses

* May or may not be pre-specified
* Hypothesis-generating rather than confirmatory
* Provide insights into treatment mechanisms and biological characteristics
* Should be clearly labeled and interpreted with caution

# 3. Part 2: Planning and Design

## 3.1 Pre-specification Requirements

### 3.1.1 Protocol and Statistical Analysis Plan Must Include:

1. Clear definition of each subgroup with scientific/clinical rationale
2. Specification of subgroup variables and cut-points
3. Statistical methods including interaction tests
4. Plan for handling multiplicity
5. Distinction between inferential, supportive, and exploratory analyses

### 3.1.2 Best Practices

* Limit subgroups to those with strong justification
* Base subgroup selection on biological plausibility or prior evidence
* Avoid data-driven post hoc subgroup definitions
* Document decision-making process transparently

## 3.2 Sample Size and Power Considerations

### 3.2.1 Key Challenges

* Trials are typically powered for overall population, not subgroups
* Time-to-event analyses depend on number of events, not just sample size
* Subgroups often lack adequate power for reliable inference
* Single pivotal trials don’t allow replication of subgroup findings

### 3.2.2 Strategies

1. If subgroup is critical, design trial with adequate power for that subgroup
2. Consider enrichment strategies when appropriate
3. Use hierarchical testing to split alpha between populations
4. Plan for adequate representation in key subgroups

## 3.3 Statistical Methodology

### 3.3.1 Interaction Testing

* Use formal tests for treatment-by-subgroup interaction
* Don’t rely solely on within-subgroup comparisons
* Include interaction terms in regression models
* Report interaction p-values alongside subgroup-specific estimates

### 3.3.2 Multiplicity Control

* Multiple testing increases Type I error risk
* Consider Bonferroni or other appropriate adjustments
* Use hierarchical testing when evaluating multiple subgroups
* Balance statistical rigor with clinical relevance

### 3.3.3 Graphical Displays

* Use forest plots to visualize effects across subgroups
* Show confidence intervals for each subgroup estimate
* Display interaction p-values
* Include consistency assessments

# 4. Part 3: FDA Case Examples and Regulatory Decisions

## 4.1 Approvals in Total ITT Population Despite Subgroup Concerns

### 4.1.1 Case 1: Nivolumab (CHECKMATE-057)

**Indication:** Second-line metastatic non-squamous NSCLC

**Trial Results:**

* ITT population: HR 0.73 (95% CI: 0.60-0.89), p <0.002
* PD-L1 <1% subgroup: HR 0.90 (95% CI: 0.66-1.24)
* ~10% objective response rate in PD-L1 <1% patients

**FDA Decision:** Approved for entire study population

**Rationale:**

* Exploratory subgroup analysis with limited power
* Missing PD-L1 data (22%)
* No evidence of harm in PD-L1 <1% subgroup
* Meaningful responses observed despite modest HR
* Subgroup details included in labeling for informed decision-making

|  |
| --- |
| Key Lesson |
| Lack of statistical significance in subgroup doesn’t preclude approval when no harm is evident and some patients benefit. |

### 4.1.2 Case 2: Pembrolizumab (KEYNOTE-042)

**Indication:** First-line advanced NSCLC with PD-L1 ≥1%

**Trial Results:**

* ITT (TPS ≥1%): HR 0.81 (95% CI: 0.71-0.93), median OS improvement 4.6 months
* TPS 1-49% subgroup: HR 0.92 (95% CI: 0.77-1.11), median OS 13.4 vs 12.1 months
* TPS ≥50% subgroup: Greatest benefit

**FDA Decision:** Approved for all patients with TPS ≥1%

**Rationale:**

* Trial not powered for TPS 1-49% subgroup analysis
* Active control (chemotherapy) as comparator
* No detrimental effect on OS in lower expression subgroup
* Uncertainty exists but doesn’t warrant restriction

|  |
| --- |
| Key Lesson |
| Modest effects in exploratory subgroups with active controls may support broad approval when no harm demonstrated. |

### 4.1.3 Case 3: Margetuximab (SOPHIA)

**Indication:** Later-line HER2+ metastatic breast cancer

**Trial Results:**

* ITT: HR 0.76 (95% CI: 0.59-0.98) for PFS
* CD16A F/F or F/V: HR 0.68 (95% CI: 0.52-0.90)
* CD16A V/V: HR 1.78 (95% CI: 0.87-3.62) - trend favoring trastuzumab

**FDA Decision:** Approved for entire study population with post-marketing commitment

**Rationale:**

* No clear mechanistic basis for heterogeneity by CD16A allotypes
* Small V/V subgroup (69 of 536 patients)
* Imbalanced patient characteristics within V/V subgroup
* Post-marketing studies required for further characterization

|  |
| --- |
| Key Lesson |
| Exploratory subgroup findings require mechanistic justification and adequate sample size to restrict indication. |

## 4.2 Approvals Restricted to Subgroups

### 4.2.1 Case 4: Olaparib + Bevacizumab (PAOLA-1)

**Indication:** First-line maintenance for HRD+ advanced ovarian cancer

**Trial Results:**

* ITT: HR 0.59 (95% CI: 0.49-0.72), p <0.0001
* HRD-positive (66% of ITT): HR 0.33 (95% CI: 0.25-0.45)
* HRD-negative (34% of ITT): No PFS benefit, OS numerically favored control

**FDA Decision:** Restricted to HRD-positive population only

**Rationale:**

* Strong biological plausibility (PARP inhibitor mechanism in HRD tumors)
* Lack of efficacy evidence in HRD-negative patients
* Potential for harm suggested by OS trend
* Consistent data across multiple PARP inhibitors
* Well-defined biomarker with validated test

|  |
| --- |
| Key Lesson |
| Restriction warranted when clear mechanistic rationale, lack of benefit, and potential harm in complementary subgroup. |

### 4.2.2 Case 5: Eribulin (Study 309)

**Indication:** Unresectable or metastatic liposarcoma

**Trial Results:**

* ITT (liposarcoma 32%, leiomyosarcoma 68%): HR 0.75 (95% CI: 0.61-0.94)
* Liposarcoma subgroup: HR 0.51 (95% CI: 0.35-0.75)
* Leiomyosarcoma subgroup: HR 0.90 (95% CI: 0.69-1.18)

**FDA Decision:** Approved only for liposarcoma

**Rationale:**

* Pre-specified exploratory subgroup analysis showed strong differential effect
* Biological differences between sarcoma subtypes
* Consistent differential effects across secondary endpoints
* Effect in larger leiomyosarcoma subgroup insufficient

|  |
| --- |
| Key Lesson |
| Substantial evidence may be limited to subgroup when biological differences support heterogeneity. |

### 4.2.3 Case 6: Atezolizumab + Paclitaxel (IMpassion130)

**Indication:** PD-L1+ triple-negative breast cancer

**Trial Results:**

* ITT: HR 0.79 (95% CI: 0.68-0.92), median PFS 7.0 vs 5.5 months
* PD-L1+: HR 0.60 (95% CI: 0.48-0.77), median PFS 7.4 vs 4.8 months, OS HR 0.71

**FDA Decision:** Accelerated approval limited to PD-L1+ population

**Rationale:**

* Trial designed with formal testing in both ITT and PD-L1+ populations
* ITT benefit magnitude modest (1.5 months) creating uncertainty
* PD-L1+ subgroup showed greater PFS benefit (2.6 months)
* Favorable OS trend in PD-L1+ subgroup
* Alpha split justified focused approval

|  |
| --- |
| Key Lesson |
| Pre-specified formal subgroup testing enables regulatory decision to focus on subgroup with more robust benefit. |

# 5. Part 4: Interpretation Framework

## 5.1 Factors Supporting Credibility of Subgroup Findings

### 5.1.1 Strong Evidence

* Pre-specification in protocol with clear rationale
* Adequate statistical power
* Biological/mechanistic plausibility
* Consistency across multiple endpoints
* Consistency with external evidence
* Formal interaction test showing significant difference
* Large effect size in subgroup

### 5.1.2 Weak Evidence

* Post hoc, exploratory analysis
* Small sample size
* One of many subgroups examined
* No biological rationale
* Inconsistent across endpoints
* Isolated finding not replicated
* Non-significant interaction test

## 5.2 Red Flags for Over-interpretation

### 5.2.1 Beware of:

* Ecological bias (aggregate-level associations)
* Chance findings from multiple comparisons
* Small subgroups with wide confidence intervals
* Lack of biological plausibility
* Selective reporting of positive findings
* Confounding within subgroups
* Departure from pre-specified analysis plan

## 5.3 Totality of Evidence Approach

### 5.3.1 Regulatory Decision-Making Considers:

1. Statistical evidence and interaction tests
2. Biological plausibility and mechanism of action
3. Preclinical and pharmacologic support
4. Clinical experience with similar agents
5. Consistency across efficacy endpoints (PFS, OS, ORR)
6. Safety profile in different subgroups
7. Medical need and available treatment alternatives
8. Quality and validation of biomarker tests
9. Practical feasibility of patient selection

# 6. Part 5: Practical Recommendations

## 6.1 For Trial Sponsors and Investigators

### 6.1.1 During Trial Design

1. Engage statisticians and regulatory experts early
2. Identify subgroups based on strong scientific rationale
3. Limit number of subgroups to avoid multiplicity
4. Power trial adequately if subgroup analysis is critical
5. Pre-specify all planned subgroup analyses in protocol
6. Plan for biomarker assessment with validated assays
7. Ensure adequate representation in key subgroups
8. Consider enrichment strategies when appropriate

### 6.1.2 During Trial Conduct

1. Collect subgroup data systematically and completely
2. Minimize missing biomarker or covariate data
3. Maintain blinding to subgroup analyses until appropriate
4. Follow pre-specified analysis plan

### 6.1.3 During Analysis and Reporting

1. Conduct pre-specified analyses as planned
2. Use interaction tests appropriately
3. Apply multiplicity adjustments when warranted
4. Clearly distinguish pre-specified vs exploratory analyses
5. Report all subgroup analyses transparently (positive and negative)
6. Use forest plots and graphical displays
7. Interpret findings in context of overall trial results
8. Acknowledge limitations and uncertainties

## 6.2 For Regulatory Reviewers

### 6.2.1 Assessment Framework

1. Verify pre-specification and scientific justification
2. Evaluate adequacy of sample size and power
3. Review statistical methods and multiplicity control
4. Assess biological plausibility of findings
5. Consider consistency across endpoints
6. Evaluate complementary subgroup data
7. Review safety across subgroups
8. Consider available treatment alternatives
9. Assess medical need and benefit-risk balance
10. Determine appropriate scope of indication

### 6.2.2 Labeling Considerations

1. Include relevant subgroup data to inform practice
2. Present both favorable and unfavorable findings
3. Use clear language about strength of evidence
4. Distinguish formal vs exploratory analyses
5. Provide information for individualized decisions

## 6.3 For Clinicians and Healthcare Providers

### 6.3.1 Using Subgroup Information

1. Understand distinction between pre-specified and exploratory analyses
2. Consider biological plausibility of subgroup effects
3. Recognize limitations of small subgroup samples
4. Look for consistency across multiple studies
5. Integrate subgroup data with other clinical factors
6. Discuss uncertainties with patients
7. Consider individual patient characteristics and preferences
8. Use validated biomarker tests when available
9. Participate in post-marketing studies when appropriate
10. Report real-world outcomes to enhance evidence base

# 7. Part 6: Reporting Standards

## 7.1 CONSORT Extension for Subgroup Analyses

### 7.1.1 Required Elements

* Rationale and pre-specification status for each subgroup
* Definition of subgroup variables and cut-points
* Statistical methods including interaction tests
* Effect estimates with confidence intervals
* Interaction p-values
* Forest plots showing all subgroups
* Distinction between pre-specified and exploratory analyses
* Complete reporting (not selective)

## 7.2 Transparency Checklist

### 7.2.1 Protocol/SAP

* Subgroups defined with rationale
* Analysis methods specified
* Multiplicity approach described
* Power considerations addressed

### 7.2.2 Reporting

* All planned analyses reported
* Pre-specified vs exploratory clearly labeled
* Interaction tests included
* Forest plots provided
* Negative findings reported
* Limitations acknowledged
* Interpretation appropriately cautious

# 8. Part 7: Common Pitfalls and Solutions

## 8.1 Pitfall 1: Data-Driven Subgroup Selection

**Problem:** Choosing subgroups after seeing data increases false positives

**Solution:** Pre-specify all subgroups in protocol with clear rationale

## 8.2 Pitfall 2: Multiple Testing Without Adjustment

**Problem:** Examining many subgroups inflates Type I error

**Solution:** Apply appropriate multiplicity adjustments or label as exploratory

## 8.3 Pitfall 3: Within-Group Comparisons Only

**Problem:** Comparing p-values between subgroups is misleading

**Solution:** Use formal interaction tests

## 8.4 Pitfall 4: Overinterpretation of Small Subgroups

**Problem:** Small samples yield unreliable estimates with wide CIs

**Solution:** Acknowledge uncertainty; seek replication

## 8.5 Pitfall 5: Ignoring Biological Plausibility

**Problem:** Spurious findings lack mechanistic basis

**Solution:** Integrate biological understanding into interpretation

## 8.6 Pitfall 6: Selective Reporting

**Problem:** Publishing only positive subgroup findings distorts evidence

**Solution:** Report all pre-specified analyses regardless of results

## 8.7 Pitfall 7: Overriding Main Results

**Problem:** Focusing on subgroup contradicting overall findings

**Solution:** Require strong, consistent, credible evidence before overriding main trial conclusions

# 9. Part 8: Key References and Guidelines

## 9.1 Regulatory Guidance

* FDA: Developing Targeted Therapies in Low-Frequency Molecular Subsets
* FDA: Enrichment Strategies for Clinical Trials
* FDA: Inclusion of Older Adults in Cancer Clinical Trials
* ICH E9: Statistical Principles for Clinical Trials
* 21 CFR 314.50: Effectiveness data by subgroups

## 9.2 Methodological Standards

* CONSORT Extension for Subgroup Analyses
* Amatya et al. (2021): Subgroup Analyses in Oncology Trials. *Clinical Cancer Research*, 27(21), 5753-5756
* BMJ Open (2019): Scoping Review of Guidance on Subgroup Effects

# 10. Summary: Key Takeaways

## 10.1 For Successful Subgroup Analyses

1. **Pre-specify** subgroups with clear scientific rationale
2. **Limit** the number to those clinically meaningful
3. **Power adequately** if subgroup analysis is critical
4. **Use interaction tests** to assess heterogeneity
5. **Control multiplicity** to manage false positive risk
6. **Interpret cautiously** in context of totality of evidence
7. **Report transparently** all pre-specified and exploratory analyses
8. **Consider biology** alongside statistics
9. **Balance access and safety** in regulatory decisions
10. **Communicate uncertainties** to support informed decisions

|  |
| --- |
| Bottom Line |
| Subgroup analyses are essential tools for personalized medicine and regulatory decision-making in oncology. Success requires careful planning, rigorous statistical methods, biological plausibility, transparent reporting, and cautious interpretation. The FDA takes a comprehensive, case-by-case approach considering statistical evidence, mechanistic understanding, clinical meaningfulness, and medical need to ensure appropriate patient populations benefit from new therapies. |

# 11. Appendix: Example Subgroup Analysis Table

Warning: package 'knitr' was built under R version 4.3.3

Example Subgroup Analysis Results

| Subgroup | N | Hazard Ratio (95% CI) | Interaction p-value |
| --- | --- | --- | --- |
| Overall | 500 | 0.75 (0.60, 0.93) | - |
| Age <65 | 300 | 0.78 (0.61, 1.00) | 0.45 |
| Age ≥65 | 200 | 0.70 (0.50, 0.98) |  |
| Male | 250 | 0.80 (0.62, 1.03) | 0.60 |
| Female | 250 | 0.70 (0.52, 0.94) |  |
| PD-L1 Positive | 150 | 0.60 (0.40, 0.90) | 0.10 |
| PD-L1 Negative | 350 | 0.85 (0.65, 1.10) |  |

# 12. Key Abbreviations

* **CI**: Confidence Interval
* **CPS**: Combined Positive Score
* **FDA**: Food and Drug Administration
* **HRD**: Homologous Recombination Deficiency
* **HR**: Hazard Ratio
* **ITT**: Intention-to-Treat
* **NSCLC**: Non-Small Cell Lung Cancer
* **ORR**: Objective Response Rate
* **OS**: Overall Survival
* **PFS**: Progression-Free Survival
* **SAP**: Statistical Analysis Plan
* **TPS**: Tumor Proportion Score

**Document Version:** 1.0  
**Last Updated:** October 15, 2025