Comprehensive Guide: Subgroup Analyses in Clinical Trials

Regulatory Perspectives, Case Examples, and Practical Recommendations

Larry Leon, Alison Pedley, …

# 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 emphasis on oncology and general therapeutic areas. It synthesizes FDA and EMA 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 and regulatory decision-making but require careful planning, rigorous methods, and cautious interpretation to avoid misleading conclusions. Both FDA and EMA emphasize the importance of pre-specification, biological plausibility, and replication. |

# 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, country
* **Baseline risk factors and prognostic scores**

### 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
5. Verify consistency of overall trial conclusions across key subgroups

## 2.2 Key Regulatory Concepts

### 2.2.1 From EMA Guideline (EMA/CHMP/539146/2013)

**Heterogeneity:** The extent of differences within the target patient population in factors that are prognostic for outcome or predictive of treatment effects. The more heterogeneous the population, the more important the investigation of treatment effects in well-defined subgroups.

**Consistency:** The extent to which estimated treatment effects in relevant subgroups assure that the overall treatment effect applies to the breadth of the trial population.

**Credibility:** The extent to which subgroup findings can be concluded as being well substantiated and relied upon for decision making. Depends on pre-specification, biological plausibility, and replication.

**Biological Plausibility:** The extent to which a differential effect between subgroups might be predicted based on clinical, pharmacological, and mechanistic considerations.

**Replication:** Whether an effect is seen in multiple independent data sources.

## 2.3 Three Categories of Subgroup Analyses

### 2.3.1 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

### 2.3.2 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

### 2.3.3 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
6. Categorization of subgroups as “key” or “exploratory”

### 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
* Consider factors used for stratification as key subgroups

## 3.2 EMA Strategy for Subgroup Selection

### 3.2.1 Three Categories of Factors:

**Category 1:** Strong reason to expect inconsistent response across levels - **separate trials usually needed**

**Category 2:** Reason to consider prognostic or some biological plausibility for inconsistent response: - Factors used to stratify randomization - Key demographic factors including genomic factors - Factors related to mechanism of action/pharmacology - Stage, severity, or phenotype of disease - Use of concomitant medications - Region, country, or center (where relevant)

**Category 3:** Good argumentation for consistent response OR absence of rationale to determine plausibility

## 3.3 Sample Size and Power Considerations

### 3.3.1 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

### 3.3.2 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
5. Consider whether sufficient evidence will be generated in key subgroups

## 3.4 Statistical Methodology

### 3.4.1 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
* **EMA recommendation:** Use nominal significance levels >5% for signal generation

### 3.4.2 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
* **Note:** Adjustment may be counter-intuitive for exploratory analyses meant to avoid overlooking untoward effects

### 3.4.3 Graphical Displays

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

### 3.4.4 EMA Recommendations on Scale

* Interactions are scale and model dependent
* Explore first on scale where endpoint is commonly analyzed
* Present supplementary analyses on complementary scale if inconsistency observed
* Even if relative effects similar, absolute effects may differ with clinical importance
* Consider benefit-risk on absolute scale, especially when toxicity present

# 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: EMA Assessment Scenarios

## 5.1 Scenario 1: Overall Statistically Persuasive Evidence

**Situation:** Clinical data are overall statistically persuasive with therapeutic efficacy demonstrated in primary analysis population.

**Goal:** Verify that conclusions of therapeutic efficacy and safety apply consistently across subgroups.

### 5.1.1 Assessment Approach:

1. Explore consistency using forest plots and interaction tests
2. If well-planned assessment shows no inconsistency → investigation complete
3. If inconsistency observed in key subgroup → finding may be credible if:
   * **Biological plausibility** for inconsistent effect in expected direction, OR
   * **Replication** across multiple data sources
   * Further supported by nominally significant interaction tests
   * Evidence across different endpoints

### 5.1.2 Decision Framework (See Annex flowchart):

* **Exploratory subgroups:** Generally disregard unless extreme or replicated
* **Key subgroups:** Pursue if biological plausibility OR replication exists
* **Not available to replicate:** Apply precautionary principle if possibly credible

## 5.2 Scenario 2: Borderline or Unconvincing Efficacy/Risk-Benefit

**Situation:** Clinical data statistically persuasive but therapeutic efficacy or risk-benefit borderline or unconvincing across breadth of population.

**Sub-scenarios:**

1. Benefit statistically but not clinically persuasive across breadth
2. Benefit persuasive but risks prevent positive risk-benefit across breadth
3. Benefit persuasive but risks prevent positive risk-benefit in subset

### 5.2.1 Minimum Criteria for Credible Subgroup (ALL required):

1. **Well-defined subgroup:** External evidence it is clinically relevant entity
   * Usually expected to be considered at planning (stratification or key subgroup)
2. **Mechanistic rationale:** Pharmacological or plausible explanation why drug could have different efficacy/risk-benefit in subgroup and complement
3. **Larger treatment effect in subgroup:** Statistical evidence should meet same standards as would be expected for full population
4. **Replication:** Findings from other relevant trials (internal or external)
   * **Challenge for single pivotal study:** Biological plausibility and data must be exceptionally strong
5. **Risk-benefit in subgroup:** Carefully inspected with relevance of safety data from ITT population considered
6. **Baseline balance:** Close inspection required; adjustment for imbalances may be needed

### 5.2.2 Decision Framework (See Annex flowchart):

**If subgroup identified a priori:** - Check directional consistency - If YES and statistically significant → Check replication - If replicated → **CREDIBLE** - If not available → **POSSIBLY CREDIBLE** (high-risk decision)

**If subgroup not identified a priori:** - Requires clinically and statistically extreme evidence - PLUS compelling retrospective explanation for plausibility - May be credible but **HIGH-RISK DECISION**

## 5.3 Scenario 3: Failed Primary Analysis

**Situation:** Clinical data fail to establish statistically persuasive evidence in primary analysis population.

**EMA Position:** From formal statistical point of view, no further confirmatory conclusions possible.

### 5.3.1 Exceptional Circumstances for Consideration:

May pursue approval without additional studies ONLY if:

1. Clinical setting where trials not feasible to repeat, OR
2. Trials of considerable size where even subpopulations have considerable randomized evidence

### 5.3.2 Requirements (Beyond Scenario 2 criteria):

* **ALL** Scenario 2 minimum criteria must be fulfilled
* **PLUS:** Clear rationale why properly planned trial failed despite drug being efficacious
* **AND:** Why additional prospective studies unfeasible or unwarranted

**EMA Warning:** This would be regarded as inadequate in most instances.

# 6. Part 5: Interpretation Framework

## 6.1 Factors Supporting Credibility of Subgroup Findings

### 6.1.1 Strong Evidence (EMA/FDA):

* 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
* Directional consistency with prior expectations
* Replication across trials

### 6.1.2 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
* Baseline imbalances within subgroup

## 6.2 Red Flags for Over-interpretation

### 6.2.1 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
* Selection bias from data-driven subgroup identification

## 6.3 EMA Guidance on Interaction Tests

|  |
| --- |
| Limited Role of Interaction Tests |
| Currently available tests lack power (sensitivity) to detect clinically important inconsistency while losing specificity if assessed at increased Type I error levels. The role is therefore **limited to signal generation only**.  **Recommendation:** If conducted, use nominal significance levels **greater than 5%** for signal generation purposes. |

### 6.3.1 Limitations:

* Power further reduced if subgroup and complement differ in size
* Interaction tests alone inadequate for decision making
* Must assess estimated treatment effects and clinical relevance
* Scale and model dependent (additive vs. multiplicative)

## 6.4 Totality of Evidence Approach

### 6.4.1 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
10. Baseline covariate imbalances and need for adjustment

# 7. Part 6: Practical Recommendations

## 7.1 For Trial Sponsors and Investigators

### 7.1.1 During Trial Design

1. Engage statisticians and regulatory experts early
2. Identify subgroups based on strong scientific rationale
3. Categorize subgroups as Category 1, 2, or 3 per EMA framework
4. Limit number of subgroups to avoid multiplicity
5. Power trial adequately if subgroup analysis is critical
6. Pre-specify all planned subgroup analyses in protocol
7. Plan for biomarker assessment with validated assays
8. Ensure adequate representation in key subgroups
9. Consider enrichment strategies when appropriate
10. Include country/region as stratification factor in multi-regional trials
11. Discuss expected heterogeneity in target population
12. Document why any sub-populations excluded from trial

### 7.1.2 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
5. Consider blind review to update planning based on emerging evidence

### 7.1.3 During Analysis and Reporting

1. Conduct pre-specified analyses as planned
2. Use interaction tests appropriately (signal generation)
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. Always present complementary subgroups together
8. Interpret findings in context of overall trial results
9. Check for baseline imbalances; adjust if needed
10. Acknowledge limitations and uncertainties
11. Consider analyses on alternative scales if interaction scale-dependent

## 7.2 For Regulatory Reviewers

### 7.2.1 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. Check for baseline imbalances within subgroups
9. Consider available treatment alternatives
10. Assess medical need and benefit-risk balance
11. Determine appropriate scope of indication
12. Apply appropriate assessment scenario (1, 2, or 3)
13. Use EMA decision flowcharts for systematic assessment

### 7.2.2 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. Consider Section 4.1, 4.3, 4.4, or 5.1 as appropriate
7. Justify in assessment report why subgroup findings credible/not credible

## 7.3 For Clinicians and Healthcare Providers

### 7.3.1 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
11. Be aware of scale dependence of effects (relative vs. absolute)
12. Consider absolute benefit in context of toxicity for risk-benefit

# 8. Part 7: Reporting Standards

## 8.1 CONSORT Extension for Subgroup Analyses

### 8.1.1 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)
* Presentation of complementary subgroups

## 8.2 EMA Reporting Recommendations

### 8.2.1 In Assessment Reports:

* Reflect sponsor discussion of heterogeneity
* Address completeness and balance of discussion
* Present consistency results (forest plots)
* Discuss any inconsistency for credibility
* Explicitly state if inconsistent results not credible
* Provide extensive justification for Scenarios 2 and 3

### 8.2.2 In SmPC (Summary of Product Characteristics):

* **Section 4.1 (Indication):** Restriction if efficacy not established
* **Section 4.3 (Contraindications):** If risk-benefit negative
* **Section 4.4 (Warnings):** If important uncertainty in key subgroup
* **Section 5.1 (Pharmacodynamics):** Subgroup efficacy data
* Not usually necessary to highlight consistency
* Present ITT results first if population restricted to subgroup
* Explain briefly why subgroup findings credible (if can be communicated succinctly)

## 8.3 Transparency Checklist

### 8.3.1 Protocol/SAP

* Subgroups defined with rationale
* Categorization as key vs. exploratory
* Analysis methods specified
* Multiplicity approach described
* Power considerations addressed
* Stratification factors identified
* Expected heterogeneity discussed

### 8.3.2 Reporting

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

# 9. Part 8: Common Pitfalls and Solutions

## 9.1 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; categorize as key or exploratory

## 9.2 Pitfall 2: Multiple Testing Without Adjustment

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

**Solution:** Apply appropriate multiplicity adjustments OR label as exploratory; recognize that adjustment counter-intuitive for safety signal detection

## 9.3 Pitfall 3: Within-Group Comparisons Only

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

**Solution:** Use formal interaction tests; remember they have limited power

## 9.4 Pitfall 4: Overinterpretation of Small Subgroups

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

**Solution:** Acknowledge uncertainty; seek replication; consider power limitations

## 9.5 Pitfall 5: Ignoring Biological Plausibility

**Problem:** Spurious findings lack mechanistic basis

**Solution:** Integrate biological understanding into interpretation; weight plausibility heavily

## 9.6 Pitfall 6: Selective Reporting

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

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

## 9.7 Pitfall 7: Overriding Main Results

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

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

## 9.8 Pitfall 8: Ignoring Scale Dependence

**Problem:** Interactions may differ on relative vs. absolute scale

**Solution:** Analyze on common scale first; present supplementary analyses on alternative scale; consider clinical relevance

## 9.9 Pitfall 9: Assuming Consistency Without Investigation

**Problem:** Not investigating whether treatment effect consistent across subgroups

**Solution:** Strategy that assumes consistency without investigation is insufficient per EMA

## 9.10 Pitfall 10: Over-reliance on Interaction Tests

**Problem:** Non-significant interaction test used to dismiss clinically important differences

**Solution:** Remember tests lack power; assess estimated effects and clinical relevance

# 10. Part 9: Key References and Guidelines

## 10.1 Regulatory Guidance

### 10.1.1 FDA:

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

### 10.1.2 EMA:

* **Guideline on the investigation of subgroups in confirmatory clinical trials (EMA/CHMP/539146/2013)** - Effective August 1, 2019
* Points to consider on multiplicity issues in clinical trials (CPMP/EWP/908/99)
* Points to consider on adjustment for baseline covariates (CHMP/EMA/295050/2013)

### 10.1.3 ICH:

* ICH E9: Statistical Principles for Clinical Trials (CPMP/ICH/363/96)
* ICH E5: Ethnic Factors in the Acceptability of Foreign Clinical Data
* ICH E17: Multi-Regional Clinical Trials (EMA/CHMP/ICH/453276/2016)

## 10.2 Methodological Standards

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

# 11. Summary: Key Takeaways

## 11.1 For Successful Subgroup Analyses

### 11.1.1 Planning:

1. **Pre-specify** subgroups with clear scientific rationale
2. **Categorize** as key (Category 2) or exploratory (Category 3)
3. **Limit** the number to those clinically meaningful
4. **Power adequately** if subgroup analysis is critical
5. **Document** expected heterogeneity in target population

### 11.1.2 Analysis:

1. **Use interaction tests** for signal generation (not definitive)
2. **Control multiplicity** appropriately
3. **Consider scale** dependence of interactions
4. **Adjust** for baseline imbalances if needed
5. **Present** forest plots with complementary subgroups

### 11.1.3 Interpretation:

1. **Interpret cautiously** in context of totality of evidence
2. **Weight** biological plausibility and replication heavily
3. **Apply** appropriate assessment scenario framework
4. **Consider** both relative and absolute effects for risk-benefit
5. **Distinguish** between key and exploratory subgroups

### 11.1.4 Reporting:

1. **Report transparently** all pre-specified and exploratory analyses
2. **Communicate** uncertainties clearly
3. **Justify** credibility of subgroup findings
4. **Include** relevant information in labeling
5. **Use** EMA/FDA frameworks for systematic assessment

|  |
| --- |
| Bottom Line |
| Subgroup analyses are essential tools for personalized medicine and regulatory decision-making. Success requires:   * **Careful planning** with pre-specification and categorization * **Rigorous statistical methods** recognizing their limitations * **Biological plausibility** and mechanistic understanding * **Transparent reporting** of all analyses * **Cautious interpretation** considering totality of evidence * **Regulatory alignment** with EMA and FDA frameworks   Both FDA and EMA take comprehensive, case-by-case approaches considering statistical evidence, mechanistic understanding, clinical meaningfulness, and medical need to ensure appropriate patient populations benefit from new therapies. |

# 12. Appendix A: EMA Assessment Flowcharts

## 12.1 Flowchart 1: Scenario 1 (Consistency Assessment)

*(See EMA Guideline Annex 1, Page 19)*

**Purpose:** Establishing credibility when considering consistency with overall statistically persuasive evidence

**Key Decision Points:**

1. Consider heterogeneity and biological plausibility
2. Classify as “key subgroup” (some plausibility) or “exploratory subgroup” (no obvious plausibility)
3. Is differential or inconsistent effect observed?
4. If key subgroup: Is effect directionally consistent with prior expectations?
5. Is effect replicated across trials?

**Outcomes:** - **CREDIBLE:** Plausibility + replication - **POSSIBLY CREDIBLE:** Plausibility OR replication (if data not available) - **NOT CREDIBLE:** Neither plausibility nor replication

## 12.2 Flowchart 2: Scenario 2 (Borderline Efficacy/Risk-Benefit)

*(See EMA Guideline Annex 2, Page 20)*

**Purpose:** Establishing credibility to find subgroup with clinically relevant efficacy or improved risk-benefit

**Key Decision Points:**

1. Consider heterogeneity and biological plausibility
2. Was subgroup identified and discussed a priori?
3. If YES: Is effect directionally consistent? Is evidence statistically significant? Is effect replicated?
4. If NO: Is there extreme evidence + retrospective compelling explanation?

**Outcomes:** - **CREDIBLE:** A priori + consistent + significant + replicated - **POSSIBLY CREDIBLE:** A priori + significant but not replicated (HIGH-RISK) - **CREDIBLE but HIGH-RISK:** Not a priori but extreme evidence + explanation + replicated - **NOT CREDIBLE:** Insufficient evidence on multiple criteria

# 13. Appendix B: Example Subgroup Analysis Table

Example Subgroup Analysis Results with Treatment-by-Subgroup Interaction Assessment

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

**Interpretation Notes:**

* Overall effect statistically significant and clinically meaningful
* Age subgroups show consistent direction; interaction p=0.45 (not significant)
* Sex subgroups show consistent direction; interaction p=0.60 (not significant)
* PD-L1 subgroups show larger effect in positive group; interaction p=0.10 (borderline)
* All confidence intervals overlap, suggesting consistency
* PD-L1 finding warrants further investigation given biological plausibility

# 14. Key Abbreviations

* **CHMP**: Committee for Medicinal Products for Human Use (EMA)
* **CI**: Confidence Interval
* **CPS**: Combined Positive Score
* **EMA**: European Medicines Agency
* **FDA**: Food and Drug Administration
* **HRD**: Homologous Recombination Deficiency
* **HR**: Hazard Ratio
* **ICH**: International Council for Harmonisation
* **ITT**: Intention-to-Treat
* **MAA**: Marketing Authorisation Application
* **NSCLC**: Non-Small Cell Lung Cancer
* **ORR**: Objective Response Rate
* **OS**: Overall Survival
* **PFS**: Progression-Free Survival
* **SAP**: Statistical Analysis Plan
* **SmPC**: Summary of Product Characteristics
* **TPS**: Tumor Proportion Score

**Document Version:** 2.0  
**Last Updated:** October 15, 2025  
**Includes:** FDA Guidance, EMA CHMP Guideline (EMA/CHMP/539146/2013), ICH Guidelines, Case Examples