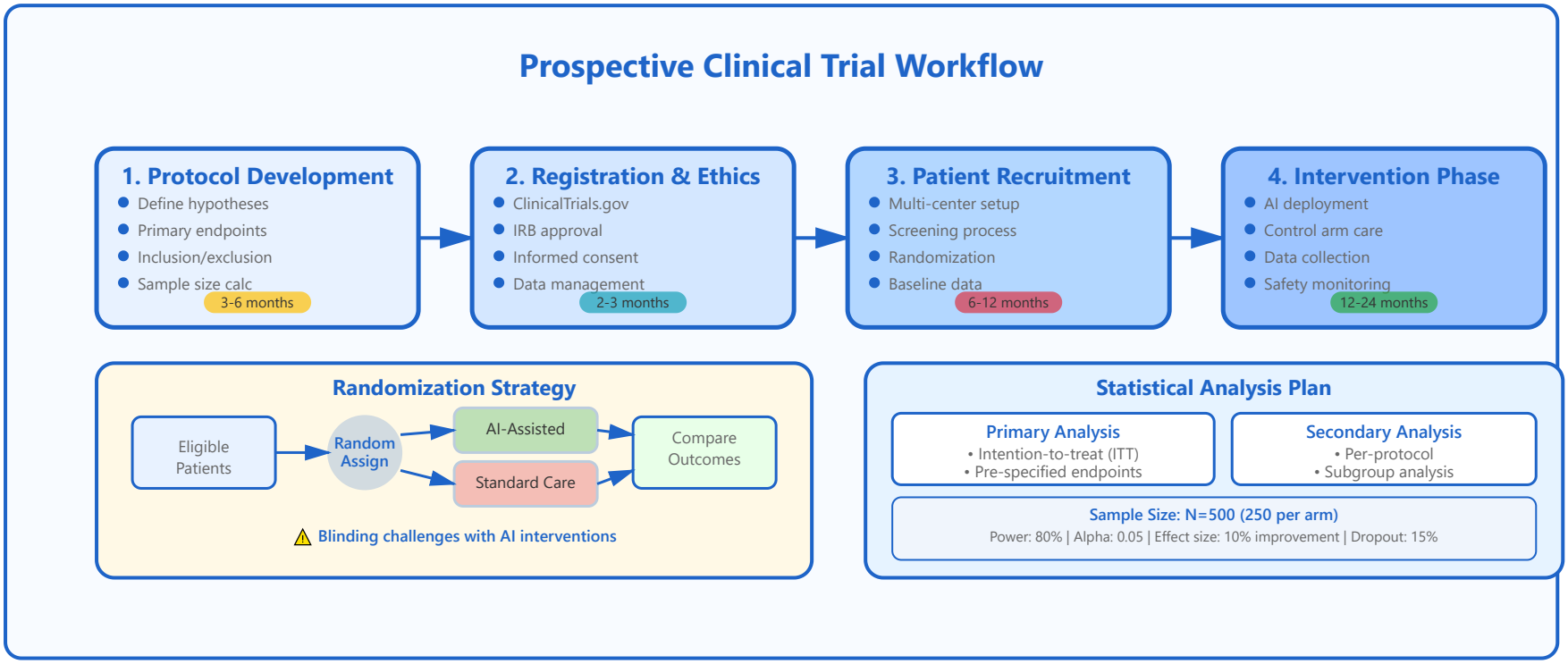
Prospective Trials: Comprehensive Guide



Trial Protocols

Pre-specified hypotheses and endpoints. Registered in clinicaltrials.gov

Endpoint Selection

Diagnostic accuracy vs clinical outcomes. Hard outcomes (mortality) vs surrogate markers

Sample Size

Power analysis for adequate statistical power. Account for prevalence and effect size

Randomization

AI-assisted vs standard of care. Cluster randomization by site to avoid contamination

Analysis Plans

Pre-specified statistical analysis. ITT vs per-protocol analysis

Detailed Explanations & Examples

1. Trial Protocols

A trial protocol is a comprehensive document that serves as the blueprint for conducting a prospective clinical trial. It defines the study objectives, design, methodology, statistical considerations, and organization of the trial. For AI/ML medical devices, protocols must be particularly rigorous as they establish the scientific validity and regulatory compliance framework.

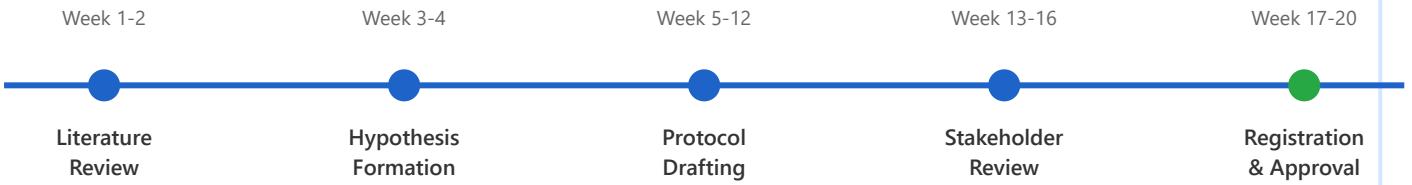
Registration requirement: All clinical trials must be registered in publicly accessible databases (e.g., ClinicalTrials.gov) before patient enrollment begins. This promotes transparency, prevents selective reporting, and allows the scientific community to track ongoing research. The registration must include key information such as the primary hypothesis, study design, eligibility criteria, interventions, and primary outcomes.

Key Components of Trial Protocols:

- **Study Objectives:** Clear statement of primary and secondary hypotheses
- **Study Design:** Randomized controlled trial (RCT), parallel-group, crossover, or adaptive design
- **Eligibility Criteria:** Inclusion and exclusion criteria defining the target population

- **Intervention Description:** Detailed specification of the AI system and control intervention
- **Primary Endpoints:** Main outcomes to be measured (must be clinically relevant)
- **Data Management:** Procedures for data collection, quality control, and security
- **Ethical Considerations:** Informed consent process and patient protection measures

Protocol Development Timeline



🔍 Example: AI-Assisted Lung Cancer Detection Trial

Study Title: "Prospective Randomized Trial of AI-Assisted CT Interpretation for Early Detection of Lung Nodules"

Primary Hypothesis: AI assistance will increase the detection rate of lung nodules ≥6mm compared to standard radiologist interpretation alone.

Design: Multi-center, parallel-group RCT with 1:1 randomization

Registration: NCT12345678 (ClinicalTrials.gov)

Primary Endpoint: Sensitivity for detecting nodules ≥6mm confirmed by follow-up CT at 3 months

2. Endpoint Selection

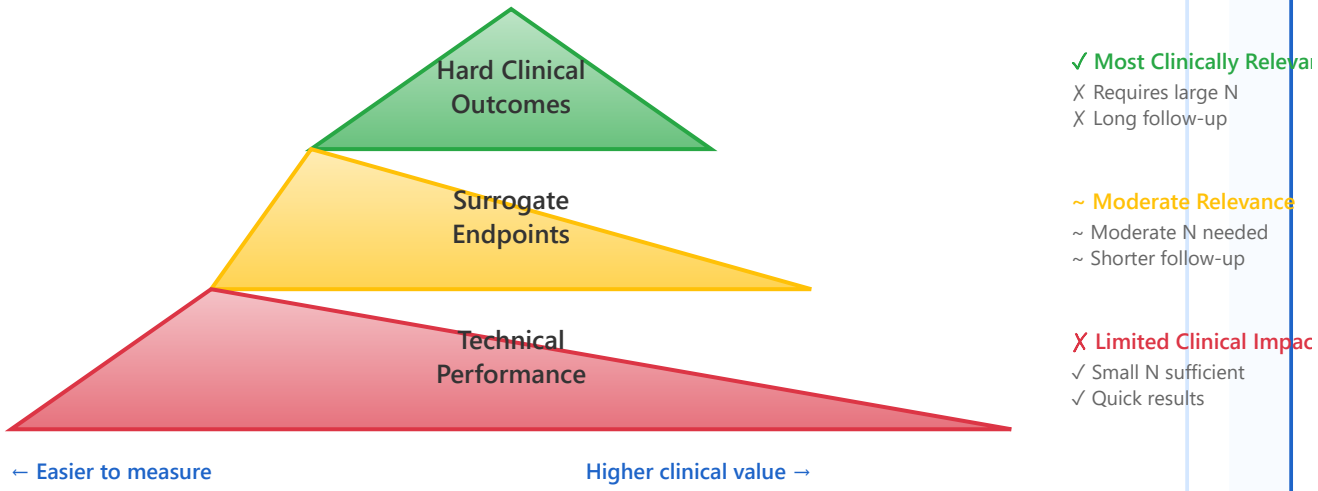
Endpoint selection is one of the most critical decisions in trial design. The choice between different types of endpoints can dramatically affect the trial's feasibility, duration, cost, and ultimately its clinical impact. For AI medical devices, there is often a tension between using technical performance metrics (e.g., diagnostic accuracy) versus patient-centered clinical outcomes (e.g., mortality reduction).

Hierarchy of Endpoints: Clinical outcomes such as mortality, morbidity, and quality of life are considered the gold standard but require larger sample sizes and longer follow-up. Surrogate endpoints (e.g., detection rates, time to diagnosis) are more feasible but require validation that improvements in these measures translate to clinical benefits.

Types of Endpoints:

- **Hard Clinical Outcomes:** Death, myocardial infarction, stroke, hospital admission
- **Surrogate Markers:** Detection sensitivity, specificity, time to diagnosis, treatment initiation
- **Process Measures:** Reading time, diagnostic confidence, inter-reader agreement
- **Patient-Reported Outcomes:** Quality of life, satisfaction, anxiety levels
- **Economic Outcomes:** Cost-effectiveness, resource utilization, length of stay

Endpoint Hierarchy and Trade-offs



🔍 Example: Diabetic Retinopathy AI Screening

Primary Endpoint Options:

Option A (Technical): Sensitivity and specificity for detecting referable diabetic retinopathy
→ Fast, requires N=200-500, results in 6 months

Option B (Surrogate): Proportion of patients receiving timely ophthalmology referral
→ Moderate duration, requires N=1,000-2,000, results in 12-18 months

Option C (Hard Outcome): Incidence of vision loss at 2 years
→ Long duration, requires N=5,000-10,000, results in 30+ months

Selected Approach: Primary endpoint = referral rates (surrogate), with vision outcomes as long-term secondary endpoint

3. Sample Size Calculation

Sample size determination is crucial for ensuring adequate statistical power to detect clinically meaningful differences while avoiding unnecessarily large and costly trials. The calculation depends on the expected effect size, baseline event rates, desired statistical power (typically 80-90%), and significance level (typically $\alpha=0.05$). For AI trials, special considerations include the prevalence of the target condition and expected improvement over standard care.

Power Analysis Components: The statistical power of a trial represents the probability of detecting a true effect when it exists. Underpowered studies risk missing real benefits (Type II error), while overpowered studies waste resources. Sample size calculations must also account for expected dropout rates, protocol violations, and crossover between study arms.

Sample Size Calculation Factors:

- **Effect Size:** Minimum clinically important difference (MCID) to detect
- **Statistical Power:** Typically 80% ($\beta=0.20$) or 90% ($\beta=0.10$)
- **Significance Level:** Usually $\alpha=0.05$ (two-sided) or $\alpha=0.025$ (one-sided)
- **Disease Prevalence:** Proportion of population with target condition
- **Baseline Performance:** Expected performance of control group
- **Dropout Rate:** Expected loss to follow-up (typically 10-20%)
- **Study Design:** Parallel vs crossover, clustering effects

Sample Size Calculation Example

Sample Size Formula (Two Proportions)

$$n = 2 \times [Z_{\alpha/2} + Z_{\beta}]^2 \times \bar{p}(1-\bar{p}) / (p_1 - p_2)^2$$

where $\bar{p} = (p_1 + p_2) / 2$

Example Calculation

Given Parameters:

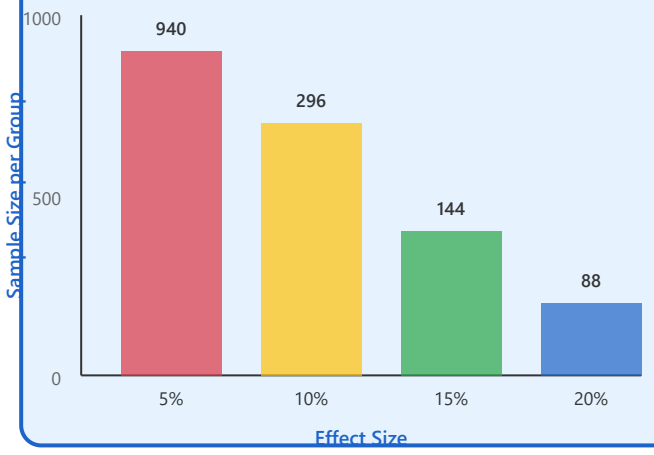
- Control sensitivity (p_2) = 75%
- AI sensitivity (p_1) = 85%
- Effect size = 10% improvement
- Power (1-β) = 80%
- Alpha (α) = 0.05 (two-sided)
- Dropout rate = 15%

Calculation:

$Z_{\alpha/2}$ = 1.96 (for α=0.05)
Z_{β} = 0.84 (for power=80%)
$\bar{p} = (0.85 + 0.75) / 2 = 0.80$

$n = 2 \times (1.96 + 0.84)^2 \times 0.80 \times 0.20 / (0.85 - 0.75)^2$
$n \approx 251$ per group
Adjusted for dropout: 251 / 0.85

Effect Size Impact



🔍 Example: AI-Based Stroke Detection

Clinical Scenario: Emergency department AI system for rapid large vessel occlusion (LVO) detection on non-contrast CT

Parameters:

- Expected LVO prevalence in suspected stroke patients: 15%
- Standard radiologist sensitivity: 82%
- Targeted AI-assisted sensitivity: 92% (10% absolute improvement)
- Power: 90%, Alpha: 0.05 (two-sided)
- Expected dropout/protocol violations: 10%

Calculation Result:

- Required sample with LVO: 350 patients
- Total patients to screen: 350 / 0.15 = 2,334 patients

• Adjusted for dropout: 2,334 / 0.90 = 2,593 patients
• Per site (10 centers): ~260 patients over 12 months

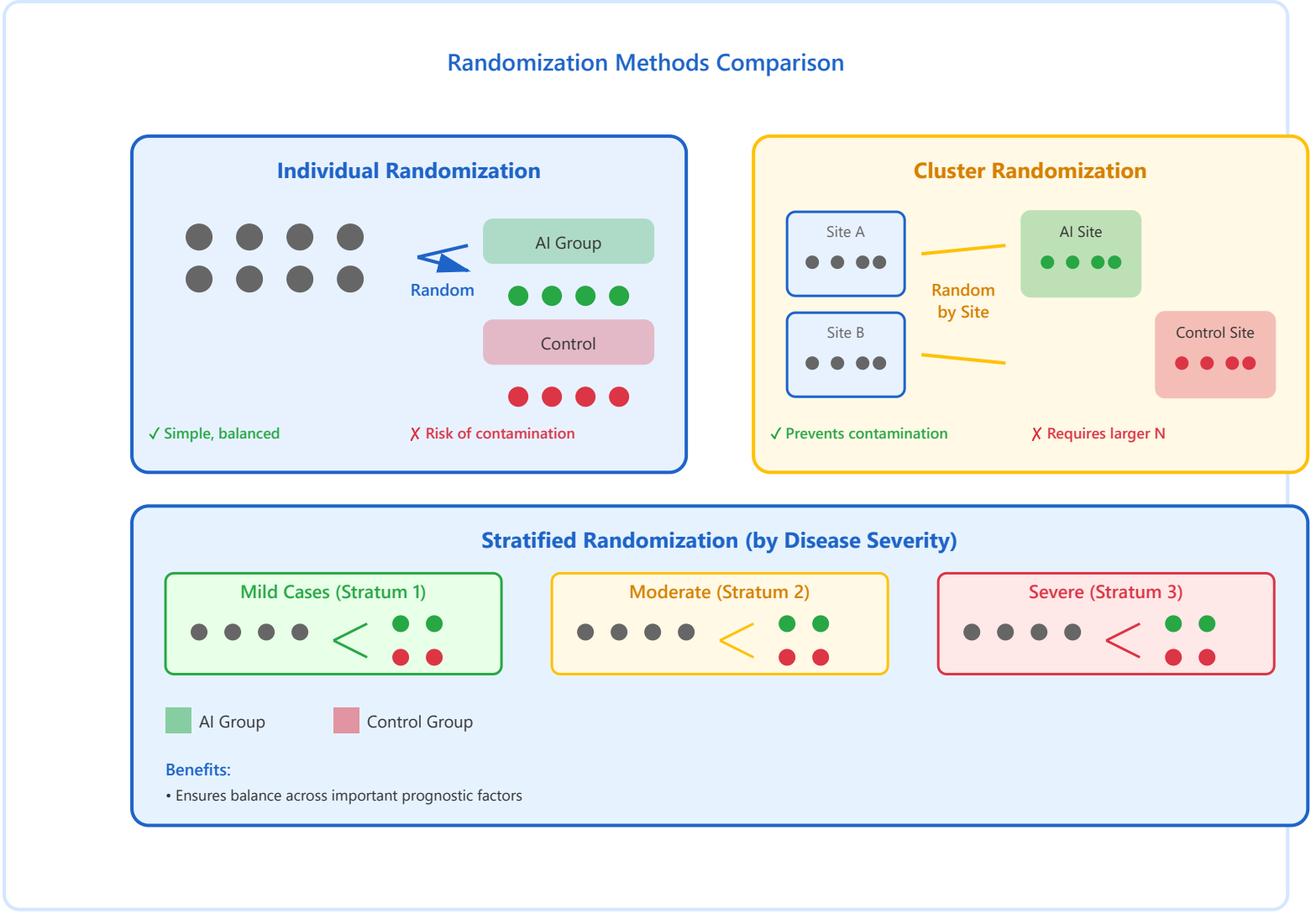
4. Randomization Strategies

Randomization is the cornerstone of prospective trials, ensuring that treatment groups are comparable at baseline and that observed differences can be attributed to the intervention rather than confounding factors. For AI medical device trials, proper randomization methods help mitigate selection bias, ensure balanced groups, and maintain the scientific validity of results. However, AI interventions present unique challenges for randomization and blinding.

Randomization Levels: Randomization can occur at different levels - patient level, provider level, or site level. For AI trials, cluster randomization (by site or provider) is often preferred to prevent contamination, where knowledge of AI recommendations might influence the control group's management. Each approach has distinct advantages and statistical considerations.

Randomization Methods:

- **Simple Randomization:** Pure random assignment (like coin flip), may result in imbalanced groups
- **Block Randomization:** Ensures balance over blocks of patients (e.g., blocks of 4 or 6)
- **Stratified Randomization:** Separate randomization within important subgroups (e.g., by disease severity)
- **Cluster Randomization:** Randomize entire sites or providers to avoid contamination
- **Adaptive Randomization:** Adjust allocation ratios based on accumulating data
- **Minimization:** Balance multiple prognostic factors simultaneously



🔍 Example: Multi-Center Sepsis Prediction AI Trial

Study Design: Cluster-randomized trial with stratification

Randomization Approach:

1. **Unit of randomization:** ICU wards (cluster randomization)
2. **Stratification factors:**
   - Hospital size (Small <200 beds, Medium 200-400, Large >400)
   - Academic vs community hospital
   - Baseline sepsis mortality rate

3. **Allocation:** 20 ICUs total, 10 randomized to AI-assisted care, 10 to standard care
4. **Blinding:** Outcome assessors blinded to group assignment
5. **Crossover consideration:** None - contamination risk too high

Rationale: Cluster randomization prevents contamination where clinicians aware of AI predictions might alter care in control patients. Stratification ensures balance across hospital characteristics that affect baseline sepsis outcomes.

5. Statistical Analysis Plans

A comprehensive Statistical Analysis Plan (SAP) must be developed and finalized before any outcome data is examined. This pre-specification is essential to maintain the scientific integrity of the trial and prevent data dredging or p-hacking. The SAP defines all analysis methods, handling of missing data, sensitivity analyses, and subgroup analyses. For AI trials, the analysis plan must address both the primary efficacy question and important safety and implementation considerations.

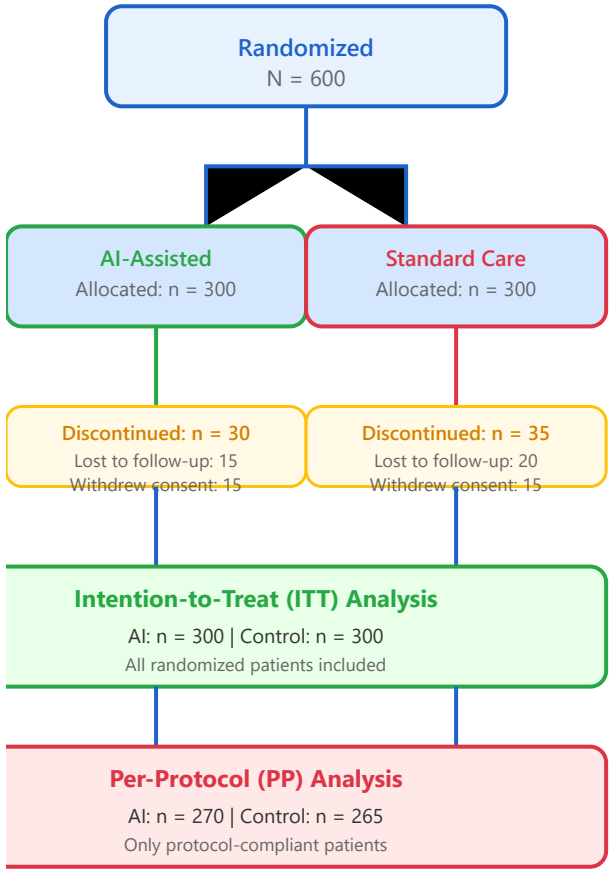
Analysis Populations: The Intention-to-Treat (ITT) principle requires analyzing all randomized patients in their assigned groups regardless of protocol adherence. This preserves randomization benefits and reflects real-world effectiveness. Per-protocol analysis (only patients who completed the protocol) addresses efficacy under ideal conditions but risks bias. Both analyses provide complementary information about the intervention's impact.

Key Components of Analysis Plans:

- **Primary Analysis:** Pre-specified statistical test for primary endpoint (e.g., chi-square, t-test, survival analysis)
- **Intention-to-Treat (ITT):** All randomized patients analyzed in assigned groups
- **Per-Protocol (PP):** Only patients completing protocol per specification
- **Missing Data:** Methods for handling dropouts and missing values (e.g., multiple imputation)

- **Sensitivity Analysis:** Testing robustness of results under different assumptions
- **Subgroup Analysis:** Pre-specified exploratory analyses in patient subsets
- **Interim Analysis:** Planned data looks during trial with adjusted significance levels
- **Multiple Testing:** Adjustment for multiple comparisons (e.g., Bonferroni correction)

Analysis Populations and Flow



Statistical Methods

Primary Endpoint Analysis

Endpoint: 30-day mortality (binary)
Method: Chi-square test (or logistic regression)
Population: ITT (primary), PP (sensitivity)

Secondary Endpoints

- Length of stay: t-test or Mann-Whitney U
- Time to event: Kaplan-Meier, log-rank test
- Adverse events: Fisher's exact test
- Cost-effectiveness: Bootstrap methods

Missing Data Handling

Primary: Multiple imputation (m=50)
Sensitivity 1: Complete case analysis
Sensitivity 2: Worst-case scenario

Pre-specified Subgroups

- Age: <65 vs ≥65 years
- Disease severity: APACHE II score
- Test for interaction with treatment

Multiplicity: Bonferroni correction ($\alpha = 0.05/3 = 0.0167$)

🔍 Example: Cardiac Arrest Prediction Algorithm

Statistical Analysis Plan Summary:

Primary Analysis:
• Endpoint: Incidence of in-hospital cardiac arrest (binary)
• Method: Generalized estimating equations (GEE) accounting for clustering by hospital
• Population: ITT including all randomized patients
• Significance: Two-sided $\alpha = 0.05$

Secondary Analyses:
• Time to cardiac arrest: Cox proportional hazards model
• ICU transfer rates: Poisson regression
• Rapid response team activations: Negative binomial regression

Missing Data:
• Primary outcome has 100% ascertainment (administrative data)
• Secondary outcomes: Multiple imputation using chained equations

Sensitivity Analyses:
1. Per-protocol analysis excluding sites with <80% AI adherence
2. As-treated analysis based on actual AI exposure
3. Analysis excluding COVID-19 pandemic period

Subgroup Analyses:
• Hospital size (tertiles)
• Baseline cardiac arrest rate (above/below median)
• Academic vs community hospitals
• Interaction tests for all subgroups with Bonferroni correction