Proposal for a Multi-Arm Biomarker RCT

Executive Summary

April 1, 2025

Motivation

- **Goal:** Design a multi-arm RCT to validate treatment-predictive biomarkers in psychiatry.

Motivation

- **Goal:** Design a multi-arm RCT to validate treatment-predictive biomarkers in psychiatry.
- **(1) Efficiency:** Validating known biomarker candidates is often more statistically efficient than discovering them *de novo*.

Motivation

- **Goal:** Design a multi-arm RCT to validate treatment-predictive biomarkers in psychiatry.
- **(1) Efficiency:** Validating known biomarker candidates is often more statistically efficient than discovering them *de novo*.
- **(2) Future Data:** Yields a rich dataset suitable for *post-hoc* development of complex predictive models (which treatment for whom?).

Model & Assumptions: Setup

- **Trial Structure:** K distinct experimental treatment arms.

Model & Assumptions: Setup

- **Trial Structure:** K distinct experimental treatment arms.
- **Key Simplification:** Powering based on **one treatment, one biomarker** assumption.

Model & Assumptions: Setup

- **Trial Structure:** K distinct experimental treatment arms.
- **Key Simplification:** Powering based on **one treatment, one biomarker** assumption.
  - Each arm k is paired prospectively with a unique biomarker X_k.

Model & Assumptions: Setup

- **Trial Structure:** K distinct experimental treatment arms.
- **Key Simplification:** Powering based on **one treatment, one biomarker** assumption.
  - Each arm k is paired prospectively with a unique biomarker X_k.
  - Aligns with literature reporting & simplifies power calculations.

Model & Assumptions: Setup

- **Trial Structure:** K distinct experimental treatment arms.
- **Key Simplification:** Powering based on **one treatment, one biomarker** assumption.
  - Each arm k is paired prospectively with a unique biomarker X_k.
  - Aligns with literature reporting & simplifies power calculations.
  - (Acknowledges reality: one biomarker might affect multiple treatments).

Model & Assumptions: Linear Model

- **Goal:** Power the trial to detect significance of parameters in:

Model & Assumptions: Linear Model

- **Goal:** Power the trial to detect significance of parameters in:

$$Y_i = \beta_0 + \sum_{k=1}^{K} [\beta_{1k} T_{ik} + \beta_{2k} X_{ik} + \beta_{3k} (T_{ik} \cdot X_{ik})] + \text{error}_i$$

Model & Assumptions: Linear Model

- **Goal:** Power the trial to detect significance of parameters in:

$$Y_i = \beta_0 + \sum_{k=1}^{K} [\beta_{1k} T_{ik} + \beta_{2k} X_{ik} + \beta_{3k} (T_{ik} \cdot X_{ik})] + \text{error}_i$$

- $T_{ik} = 1$ if patient i in arm k, 0 otherwise.

Model & Assumptions: Linear Model

- **Goal:** Power the trial to detect significance of parameters in:

$$Y_i = \beta_0 + \sum_{k=1}^{K} [\beta_{1k} T_{ik} + \beta_{2k} X_{ik} + \beta_{3k} (T_{ik} \cdot X_{ik})] + \text{error}_i$$

- $T_{ik} = 1$ if patient i in arm k, 0 otherwise.
- $X_{ik} =$ Biomarker value for patient i associated with arm k.

Model & Assumptions: Coefficient Interpretation

- β_0 : Average outcome for the reference treatment group.

Model & Assumptions: Coefficient Interpretation

- β_0 : Average outcome for the reference treatment group.
- β_{1k} : Main effect of treatment k vs. reference (traditional RCT target).

Model & Assumptions: Coefficient Interpretation

- β_0 : Average outcome for the reference treatment group.
- β_{1k} : Main effect of treatment k vs. reference (traditional RCT target).
- β_{2k} : General association of biomarker X_k with outcome (prognostic effect; nuisance/confounder control).

Model & Assumptions: Coefficient Interpretation

- β_0 : Average outcome for the reference treatment group.
- β_{1k} : Main effect of treatment k vs. reference (traditional RCT target).
- β_{2k} : General association of biomarker X_k with outcome (prognostic effect; nuisance/confounder control).
- $\beta_{\mathbf{3k}}$: **Interaction effect** (Primary Target!) - How biomarker X_k modulates treatment k's effect. Goal is accurate, significant estimation.

Power Estimates: Basis

- **Source:** Simulations based on plausible effect sizes (β_{3k}) derived from literature.

Power Estimates: Basis

- **Source:** Simulations based on plausible effect sizes (β_{3k}) derived from literature.
- **Examples Modeled (K=4):**

Power Estimates: Basis

- **Source:** Simulations based on plausible effect sizes (β_{3k}) derived from literature.
- **Examples Modeled (K=4):**
  - iAPF (EEG) for rTMS (e.g., $\beta_3 \approx 0.2 - 0.3$)

Power Estimates: Basis

- **Source:** Simulations based on plausible effect sizes (β_{3k}) derived from literature.
- **Examples Modeled (K=4):**
  - iAPF (EEG) for rTMS (e.g., $\beta_3 \approx 0.2 - 0.3$)
  - AUD History (EHR) for Ketamine (e.g., $\beta_3 \approx 0.6$ or higher from OR)

Power Estimates: Basis

- **Source:** Simulations based on plausible effect sizes (β_{3k}) derived from literature.
- **Examples Modeled (K=4):**
  - iAPF (EEG) for rTMS (e.g., $\beta_3 \approx 0.2 - 0.3$)
  - AUD History (EHR) for Ketamine (e.g., $\beta_3 \approx 0.6$ or higher from OR)
  - Inflammatory Markers (Blood) for ECT (e.g., $\beta_3 \approx 0.2$)

Power Estimates: Basis

- **Source:** Simulations based on plausible effect sizes (β_{3k}) derived from literature.
- **Examples Modeled (K=4):**
  - iAPF (EEG) for rTMS (e.g., $\beta_3 \approx 0.2 - 0.3$)
  - AUD History (EHR) for Ketamine (e.g., $\beta_3 \approx 0.6$ or higher from OR)
  - Inflammatory Markers (Blood) for ECT (e.g., $\beta_3 \approx 0.2$)
  - Speech Latency (Voice) for Novel Agents (e.g., $\beta_3 \approx 0.3$)

Power Estimates: Basis

- **Source:** Simulations based on plausible effect sizes (β_{3k}) derived from literature.
- **Examples Modeled (K=4):**
  - iAPF (EEG) for rTMS (e.g., $\beta_3 \approx 0.2 - 0.3$)
  - AUD History (EHR) for Ketamine (e.g., $\beta_3 \approx 0.6$ or higher from OR)
  - Inflammatory Markers (Blood) for ECT (e.g., $\beta_3 \approx 0.2$)
  - Speech Latency (Voice) for Novel Agents (e.g., $\beta_3 \approx 0.3$)
- **Caveats:** Literature estimates are rough (differing populations, disease criteria, outcome measures, reported stats). Use with caution.

Power Estimates: Definitions Control

- **Statistical Power:** Probability of correctly detecting a true effect (interaction $\beta_{3k} \neq 0$) when it exists. Target often 80%.

Power Estimates: Definitions Control

- **Statistical Power:** Probability of correctly detecting a true effect (interaction $\beta_{3k} \neq 0$) when it exists. Target often 80%.
- **Family-Wise Error Rate (FWER):** Probability of making *at least one* false positive discovery (claiming $\beta_{3k} \neq 0$ when it's truly 0) across all K tests.

Power Estimates: Definitions Control

- **Statistical Power:** Probability of correctly detecting a true effect (interaction $\beta_{3k} \neq 0$) when it exists. Target often 80%.
- **Family-Wise Error Rate (FWER):** Probability of making *at least one* false positive discovery (claiming $\beta_{3k} \neq 0$ when it's truly 0) across all K tests.
- **Control Method:** Used Holm's procedure (conservative) to control FWER at 5%.

Power Estimates: Simulation Results

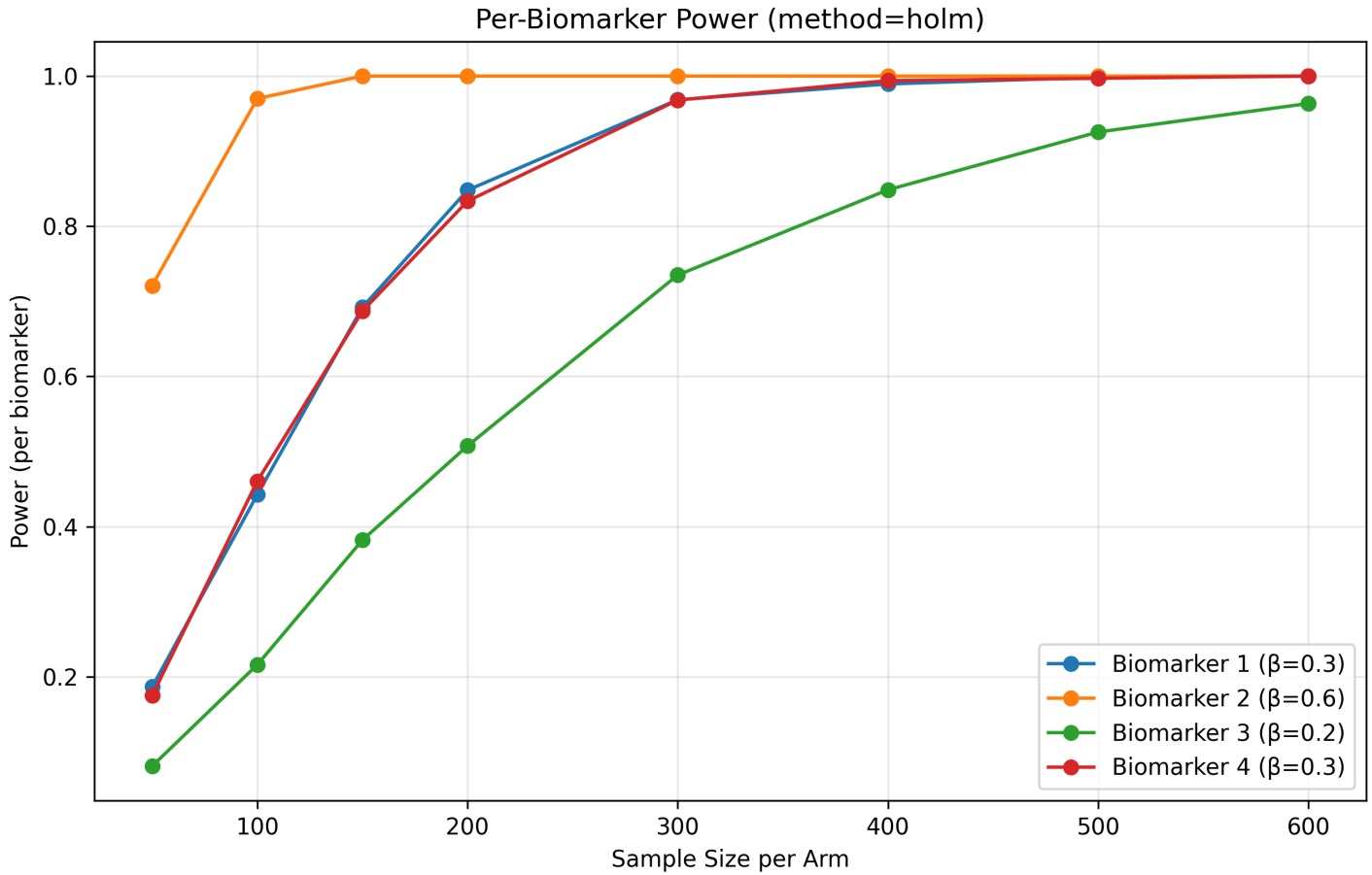

Figure: Power per biomarker vs. sample size *per arm* (K=4, Holm FWER control).

Power Estimates: Simulation Results

- **Observation 1:** Required sample size varies greatly by interaction strength (β_{3k}).

Power Estimates: Simulation Results

- **Observation 1:** Required sample size varies greatly by interaction strength (β_{3k}).
- **Example (80% Power):**

Power Estimates: Simulation Results

- **Observation 1:** Required sample size varies greatly by interaction strength (β_{3k}).
- **Example (80% Power):**
  - Strong ($\beta_{3k} \approx 0.6$): 50 subjects/arm

Power Estimates: Simulation Results

- **Observation 1:** Required sample size varies greatly by interaction strength (β_{3k}).
- **Example (80% Power):**
  - Strong ($\beta_{3k} \approx 0.6$): 50 subjects/arm
  - Weak ($\beta_{3k} \approx 0.2$): 350 subjects/arm

Power Estimates: Simulation Results

- **Observation 1:** Required sample size varies greatly by interaction strength (β_{3k}).
- **Example (80% Power):**
  - Strong ($\beta_{3k} \approx 0.6$): 50 subjects/arm
  - Weak ($\beta_{3k} \approx 0.2$): 350 subjects/arm
- **Observation 2:** If goal is to validate *all* K biomarkers, trial must be sized for the *weakest* interaction (350/arm -¿ Total N $= \mathbf{1400}$).

Considerations Future Directions

- **Biomarker Correlation:** Assumed biomarkers are correlated (via latent factor U).

Considerations Future Directions

- **Biomarker Correlation:** Assumed biomarkers are correlated (via latent factor U).
  - Higher correlation generally *decreases* power, potentially requiring larger sample sizes than shown if correlation is strong. (Estimating this pre-trial is valuable).

Considerations Future Directions

- **Biomarker Correlation:** Assumed biomarkers are correlated (via latent factor U).
  - Higher correlation generally *decreases* power, potentially requiring larger sample sizes than shown if correlation is strong. (Estimating this pre-trial is valuable).
- **Adaptive Designs?:** Could potentially *reduce* sample size (e.g., drop futile arms/biomarkers early, enrich promising ones).

Considerations Future Directions

- **Biomarker Correlation:** Assumed biomarkers are correlated (via latent factor U).
  - Higher correlation generally *decreases* power, potentially requiring larger sample sizes than shown if correlation is strong. (Estimating this pre-trial is valuable).
- **Adaptive Designs?:** Could potentially *reduce* sample size (e.g., drop futile arms/biomarkers early, enrich promising ones).
  - Trade-off: Increased operational and statistical complexity.

Considerations Future Directions

- **Biomarker Correlation:** Assumed biomarkers are correlated (via latent factor U).
  - Higher correlation generally *decreases* power, potentially requiring larger sample sizes than shown if correlation is strong. (Estimating this pre-trial is valuable).
- **Adaptive Designs?:** Could potentially *reduce* sample size (e.g., drop futile arms/biomarkers early, enrich promising ones).
  - Trade-off: Increased operational and statistical complexity.
- **Forthcoming: Predictive Modeling Simulations:** Planning extended simulations focusing on building *predictive models* using the full feature set (many biomarkers, instruments) to generate patient-specific treatment profiles (see separate plan).

End