CHOOSING PRO STATISTICAL ENDPOINTS TO MAXIMIZE SUCCESSFUL OUTCOMES

Dennis D. Gagnon, MA MABE
Senior Director, Strategic Consulting
Thomson Reuters (Healthcare) Inc.

Margaret Rothman, Ph.D.
Senior Director, Worldwide Market Access, Patient-Reported Outcomes
Janssen Pharmaceutical Companies of Johnson & Johnson

Purpose

PRO researchers are frequently called upon to recommend statistical endpoints for clinical trials. The importance of the relationship between PRO statistical endpoints used in clinical trials, study design and research hypotheses is frequently overlooked.

This workshop will focus on understanding this relationship to improve trial outcomes.

Agenda

- Understand the question
- Failure vs. Success: Setting the Stage
- Considerations in the Choice of PRO Statistical Endpoints
- Two Case Studies
- Interactive Session

Understand the Question

Begin with the end in mind

- What PRO questionnaire should we put in this trial?
- What may be reasonably expected from this drug, e.g., decreased pain (change from severe to mild pain), improved mobility (change from walk across room to walk one block)?
  - early understanding of expected effects and specificity are important in selecting endpoints
- When will first effects be seen?
- Are expected effects likely to be detected by patients?
  - Where does the patient's perception fit into the hierarch of outcomes
- What existing measures are available to assess the expected effects?
- How well do items and domains in the scale match the expected effects?

Mean Change or Responder?

A Thought Experiment

- Assume:
  - An RCT with 200 subjects, randomized 1:1, includes a PRO measure
  - We know a priori that the PRO Minimum Important Difference (MID) = 3.0, and the Minimum Important Change (MIC) = 5.0
  - Active arm: 75 subjects achieve change of 5.5, 25 achieve change of 4.5
  - Control arm: 25 subjects achieve change of 5.5, 75 achieve change of 4.5
- Results for two different PRO statistical endpoints:
  - Mean change: 5.25 for active, 4.75 for control (Difference: 0.50, p <0.0001) Failure
  - Responder rate: 75% for active, 25% for control (Difference: 50%, p<0.001) Success

Cumulative Distribution of Change

- To be fair, in the artificial example, there really isn't much difference between groups
- The US FDA has suggested displaying the change scores as a cumulative distribution by treatment groups
- This allows for a comparison across groups for a variety of response definitions
- However, this type of display still assumes the validity and usefulness of change scores
- Maybe change scores are not appropriate?
Some Examples Where Change Scores May Not be Appropriate

- Where we expect substantial attrition:
  - Recurrence prevention study
  - Late stage oncology
- Where final follow-up scores are hypothesized to be similar between groups:
  - Tolerability study
  - Or

Considerations in the Choice of PRO Statistical Endpoints

PRO Statistical Endpoints Should Be

- **Compatible:**
  - with study design (time-to-event, fixed duration, etc.)
  - research objectives (consistent with endpoint model)
- **Testable:**
  - One hypothesis statement, independent from other study endpoints
- **Meaningful:**
  - intrinsic meaning from the clinical and/or patient’s perspective (obvious treatment benefit); plain language
- **Feasible:**
  - Sufficient statistical power; likely to show a difference
- **Valuable:**
  - Demonstrates a clear medical need

Examples of Compatibility: Primary & PRO Endpoints
(Some Suggested Possibilities)

- **Oncology:**
  - Progression Free Survival; Time to PRO Decrement
  - Proportion of Tumor Responders; Proportion of PRO Responders
- **Pain:**
  - Change in PAIN Visual Analog Scale (VAS); Change in PRO
  - Mean Follow-up VAS; Area Under the PRO Curve (AUC)
  - Time to pain improvement; Time to PRO Improvement
- **Virology:**
  - Adverse Event Comparison; Worst PRO Score During Follow-up
  - Microbiologic Cure Rate; Proportion of Subjects with Symptom Reduction
- **Immunology:**
  - Proportion of Psoriasis Responders; Proportion of PRO Responders
  - Time to Symptom Reduction; Time to PRO Improvement

Case Study 1:
**Everolimus for Metastatic Renal Cell Carcinoma**

- A phase III trial of everolimus over placebo in patients with metastatic renal cell carcinoma (mRCC).
- 416 subjects randomized 2:1 to everolimus 10 mg/d (n=277) or placebo (n=139).
- Progression-free survival (PFS; primary endpoint) and safety assessed to the end of double-blind treatment.
- Secondary endpoints included:
  - Objective response rate
  - Overall survival
  - Functional Assessment of Cancer Therapy-Kidney Symptom Index (FKSI)

Case Study 1: Everolimus for Metastatic Renal Cell Carcinoma

- The attrition in this study is typical for late stage oncology.
- A PRO change score from baseline to last available assessment doesn’t make sense with this attrition rate.
- To capture the PRO treatment benefit:
  - Median time to definitive deterioration of the FKSI was tested between treatment groups using a Kaplan-Meier estimator.
  - Can’t tell from the Motzer publication if this was specified a priori or conducted post hoc.
- PRO statistical endpoint mirrors primary endpoint (PFS).


Case Study 2: Paliperidone Palmitate in Schizophrenia

- Relapse prevention and maintenance of social functioning are important treatment objectives in the long-term management of schizophrenia.
- Social functioning can be measured by the Personal and Social Performance (PSP) scale.
- Phase III studies were conducted testing paliperidone palmitate (Janssen, Division of Ortho-McNeil-Janssen Pharmaceuticals, Inc.) in preventing relapse in schizophrenia.


Redefine PSP statistical endpoint as time to significant decrement in the PSP:
- Mirrors primary statistical endpoint (time to relapse)
- Based upon previous work showing that a 10-point decrease in PSP is an important decrement
- Analysis based upon a Kaplan-Meier survival curve
- Time-to-PSP decrement analysis conducted post hoc to demonstrate treatment benefit of paliperidone palmitate.

Case Study 2: Paliperidone Palmitate in Schizophrenia

Exercise: Which PRO Endpoints?
Assume: We hypothesize that the mean PRO scores over 72 weeks will look as below for the intervention and control group. Which PRO statistical endpoint should we use?

Exercise:
Match Endpoint to Study Design/Component

Conclusions
Conclusions

• There are many possible PRO statistical endpoints to choose from
• The choice is context dependent
• We want to choose a PRO statistical endpoint that is:
  - **Compatible** (with study design and objectives)
  - **Testable** (with a single statistical test)
  - **Meaningful** (presents a clear benefit in plain language)
  - **Feasible** (should result in positive findings if true effect is present)
  - **Valuable** (demonstrates a clear medical need)