Statistical Considerations in Clinical Research Studies

Karan P. Singh, PhD, FMSSANZ, FASA
Professor and Chair
Department of Epidemiology and Biostatistics

No financial disclosures

No conflicts of interest
Outline

- Introducing Clinical Trials
- The Business of Oncology Medication Clinical Trials
- Biostatistician’s Role in Oncology Medication Clinical Trials
- Data Management & General Data Flow Diagram
- Primary Data Collection
- Example-Randomized Trial
- Analysis of Secondary Data that Supports the Primary Topic of Interest
- Qs & As

A Major Piece of the Puzzle: Clinical Data Management

- It usually takes 80%-90% of resources for a study

- Measurement error and information bias
  - Reduce statistical power
  - Inaccurate conclusion

- Data sharing and reproducibility
Clinical Data Management

Clinical data management consists of various activities involving the handling of data or inform that is outlined in the protocol to be collected/analyzed.

A multidisciplinary research team includes:
• Research Nurses
• Clinical Data Managers
• Investigators
• Biostatisticians
• Database Programmers

What Are Clinical Trials?

“Clinical trials are research studies that explore whether a medical strategy, treatment, or device is safe and effective for humans.”

“Studies follow strict scientific standards. These standards protect patients and help produce reliable study results.”

National Heart, Lung and Blood Institute (NHLBI) of NIH
http://www.nhlbi.nih.gov/studies/clinicaltrials
### Phases of Clinical Trials

Clinical trials are conducted in phases, each phase answers a separate research question.

**Pre-clinical studies** - Whether a drug candidate has scientific merit for further development as an investigational new drug.

**Phase 0 Trials** - Also called human microdosing studies-speeding up the development of promising drugs or imaging agents.

**Phase I** - A small group of people to evaluate its safety.

**Phase II** - A larger group of people to evaluate its efficacy, not any therapeutic effect whatsoever.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>III</td>
<td>Large groups of people to evaluate its effectiveness by comparing with commonly used treatments.</td>
</tr>
<tr>
<td>IV</td>
<td>Gathering information on the drug's effect in various populations and any side effects associated with long-term use.</td>
</tr>
<tr>
<td>V Trials</td>
<td>Translational research/comparative effectiveness research and community-based research used to signify the integration of a new clinical treatment into widespread public health practice.</td>
</tr>
</tbody>
</table>

**Source:** U.S. National Library of Medicine, NIH
Phase I Trial Designs

Non-statistically based designs: Algorithmic designs, 3+3 cohort
- Tends to underestimate the maximum tolerated dose (MTD)
- Problem in incorporating explicit targeted dose-limiting toxicity (DLT) rate
- Challenge to extend to more complex settings

Nonparametric designs: Biased-coin design (BCD)
- Identify MTD better than algorithmic designs
- Challenge to extend it to more complex settings

Phase I Trial Designs

Parametric Designs: Continual Reassessment Method (CRM)
- Relocate an estimate of the MTD after treating each patient and assessing for toxicity.
- The final estimate of the MTD occurs after reaching a pre-specified # of patients.
- Escalation with overdose control (EWOC)
- Computationally intensive
Phase I Trial Designs

Parametric Designs: Continual Reassessment Method (CRM)

- Easily extendable to more complex settings
- We need more resources

Some References

- Moller, S, An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses. Statistics in Medicine, 14:911-922, 1995.

Adaptive Trial Design

This design allows flexibility to redesign the trial based on the results of interim analysis.

- Redesigning the trial saves time and reduces number of patients
- Incorporate practical considerations
- Appraisal doesn’t undermine the integrity and validity of the trial

“An adaptive design for a medical device clinical study is defined as a clinical trial design that allows for prospectively planned modifications based on accumulating study data without undermining the trial’s integrity and validity.”


Adaptive Trial Design

Examples of adaptations to a trial based on interim analysis
• Early stopping due to efficiency or futility
• Sample size re-estimation
• Adaptive randomization
• Dropping inferior treatment groups

Three commonly used statistical methods based on test statistic
• Sum of stage-wise p-values (MSP)
• Product of stage-wise p-values (PSP)
• Weighted inverse normal of stage-wise p-values (WINP)

There are good books on clinical trials authored by Pocock, Freedman, Furberg and DeMet, Piantadosi, Chow and Liu, Green, et.al., Everitt and Pickles, Mark Chang and others.

Biostatistician’s Role

Consultant: Inactive role
• Need a sample size calculation hours before submission

Example of a biostatistician’s inactive role:
Biostatistician’s Role: Design Stage

Collaborator: Active role
- Get involved early in planning stage
- Consider all aims/end points of the study
- Choose most efficient design that might be suitable (pre-post, Cross-over, factorial, CRM, two-stage, non-inferiority, equivalence, Stepped-Wedge …)

Biostatistician’s Role: Design Stage

- Sample size considerations
- Interim monitoring plan

Specific method used depends on
- The Specific aim(s)/objective(s)
- The study design, including the planned number of measurements per ‘subject’
- The outcome(s) and predictor(s)
Biostatistician’s Role: Design Stage

- The proposed statistical analysis plan
  - Need to consider: Accrual / Enrollment
    - (response rate for questionnaires)
    - Drop-outs and missing data
- Budgetary, time, and resource constraints
- Requires you to make assumptions
- Assume specific effect size (variability), power, etc

Measurement issues

- Is there measurement error that should be considered?
- Multiple endpoints (e.g. tumor response & Time to death)
- Account for type I and type II errors
- Stratifications or adjustments are included if necessary
- Plans for missing data
Biostatistician’s Role: Design Stage

- 40+ pages of survey plus other assessments every 3 months during the project.

The biostatistician involved in the project should ask:
- How long is the project?
- How much funding is available or to be applied for?
While we wait...

- Primary data collection takes time to plan, groom, and verify before it is ready for analysis.

- While data is being processed, secondary data can be used for analyses that support the primary topic of interest.

Secondary Data Analysis

- There is no single “best” dataset for secondary data analysis.

- Combining datasets is sometimes necessary and requires thorough investigation and creative thinking.
Issues with combining data

- What can I link with?
- What datasets are available?
- What are the contents?
- How can I be creative with these constraints?
- How can I relate it to my research?

What can I link with?

- A variable that can link the data sets
  - Aggregated variables

- The more common variables include:
  - Region names
  - Federal Information Processing Standards codes for county identifiers
What datasets are available?
Survey Related

- National Health and Nutrition Examination Survey
- National Health Interview Survey
- National Health Care Surveys
- National Vital Statistics System
- Behavioral Risk Factor Surveillance System
- Youth Risk Behavior Surveillance System

Strengths & Limitations

➢ **Combining data introduces Strengths**
  • Quantitative and Qualitative data in the same dataset
  • Additional variables to include in your correlative analyses

➢ **Limitations**
  • Additional variables may not be available for all observations in the merging dataset
  • Data is collected in an inconsistent manner
The Tough Questions

**RESEARCHER**
- How can I relate it to my research?
- Manageable research questions?

**BIOSTATISTICIAN**
- How can I be creative within the data constraints?
- Can data be combined appropriately?

Dynamic Data Example:
Medicare Administrative Claims Data

- Persistent updated data analysis from multiple data sources
  - Medical Concierge
  - 12 Sites & Tumor Registry
  - Centers for Medicare & Medicaid Services data
- Each data source has moving components
  - Each source’s data is updated at different rates (quarterly, 6 months, annual)
### Take Home Messages

#### RESEARCHER
- Active engagement and communication to formulate & shape research questions.
- Recognize the limitations and potential of the combined data.

#### BIOSTATISTICIAN
- Understand the dataset and contents
- Feedback for collaborators on the data analysis plan

---

Thanks for your patience and time!