Clinical Data Transparency and Disclosure: Statistical Perspectives

Marcia Levenstein, DSc
Pfizer Inc.
(Based on joint work with Demissie Alemayehu, PhD)
Abstract

The increased demand for clinical trial data transparency and disclosure poses considerable challenges and opportunities from a statistical perspective. A central issue is the need to guarantee protection of patient and innovation, while ensuring access to patient-level data. This paper offers options to navigate this dilemma, with emphasis on the role of good clinical and statistical practices as proven safeguards for transparency, the importance of adopting best practices for reporting of data from secondary analyses, and the need for optimal collaboration among stakeholders to facilitate data sharing.
Objectives

• Review positions of alternative stakeholders in statistical contexts
• Underscore roles of Good Clinical and Statistical Practices as proven safeguards for transparency
• Highlight statistical issues with privacy, disclosure and patient-level data sharing
• Assess infrastructural issues in data sharing
• Outline requirements for balanced reporting and dissemination of results
Introduction

- EMA draft Policy on publication and access to clinical-trial data: Key elements
  - Enabling public scrutiny and secondary analysis of clinical trials
  - Protection of personal data
  - Respect for the boundaries of patients' informed consent
  - Protection of commercially confidential information
  - Ensuring future investment in bio-pharmaceutical research and development
  - Addressing the consequences of inappropriate secondary data analysis
  - Protecting the Agency's and the European Commission's deliberations and decision-making process
  - Ensuring that transparency is a two-way street


- Sponsor, EFPIA-PhRMA positions
  - Protection of patient privacy
  - Ensuring integrity of scientific research and regulatory approval systems
  - Promotion of investment in biomedical research
Introduction (cont.)

- Potential Reasons for Data Sharing
  - Replication of sponsor findings
  - Confirmatory or exploratory analysis outside of protocol objectives
  - Aggregate data analysis (pooled/meta-analysis)
  - To advance science, e.g. study design

- Statistical Implications
  - Requirements for replication of sponsor findings
  - Data quality considerations
  - Requirements for reporting of secondary analysis results and potential impacts on public health
    - Pitfalls of post-hoc analyses
    - Understanding what constitutes the totality of evidence
  - Conceptual issues with patient-level vs. aggregate data analysis
  - Infrastructural issues with data sharing.
Good Clinical Practice (ICH E6) vs. Transparency

• Adherence to GCP a primary requirement for transparency

• Key elements of GCP (ICH E6) relative to transparency:
  
  – “Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective task(s)” (Sec 2.8).
  
  – “All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification” (Sec 2.10).
  
  – “The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s)” (Sec 2.11).
  
  – Clinical Trial/Study Reports
    
    “Whether the trial is completed or prematurely terminated, the sponsor should ensure that the clinical trial reports are prepared and provided to the regulatory agency(ies) as required by the applicable regulatory requirement(s)” (Sec 5.22).
Good Statistical Practice vs. Transparency

• Good statistical practice key to ensuring transparency
  – Pre-specification of hypotheses
    • Issues of multiplicity
    • Rationale for study size and power
  – Pre-specification of analysis plan
  – Quality standards for data
  – Programming QC and QA
  – Interpretation of results
Good Statistical Practice vs. Transparency (cont.)

• Analysis plan formulation and execution
V. DATA ANALYSIS CONSIDERATIONS

5.1 Prespecification of the Analysis

5.2 Analysis Sets

5.2.1 Full Analysis Set

5.2.2 Per Protocol Set

5.2.3 Roles of the Different Analysis Sets

5.3 Missing Values and Outliers

5.4 Data Transformation

5.5 Estimation, Confidence Intervals and Hypothesis Testing

5.6 Adjustment of Significance and Confidence Levels

5.7 Subgroups, Interactions and Covariates

5.8 Integrity of Data and Computer Software Validity
Good Statistical Practice vs. Transparency (cont.)

• Quality and reproducibility:
  – Data management plan which defines data handling rules, edit checks, dictionaries
  – Monitoring plan and ongoing data reviews
  – Data management processes including formal database release, audit trails
  – Prospective Statistical Analysis Plan
  – Programming QC/QA plan & execution (SDLC)
    Requirements/specs, design, implementation, testing/validation, deployment
### Statistical Considerations with Data Sharing

#### Case 1. Data sharing to replicate sponsor findings

- **Access requester**
  - Thorough familiarity with original research objectives, study protocol, data standards and analytical strategies
  - Adequate infrastructure to access and analyze data
  - Availability of programming and statistical resources
  - Willingness to review findings with sponsor

- **Sponsor**
  - Sponsor to provide controlled access to available relevant documents: protocol, annotated CRFs, SAP, data conventions, CSRs
  - Facilitate access to raw and derived data
  - Personnel to provide support for ad hoc requests
  - Review of results performed by external party
Case 2. Data sharing for confirmatory or exploratory analysis outside scope of protocol objectives

- **Confirmatory analysis**
  - Goal: Confirm pre-specified hypothesis
  - Methodology: Inferential statistics, dependent on model assumptions.
  - Pre-specification of analytical plan essential
  - Caution in interpreting results, since analysis is motivated by preconceived conjecture or prior findings

- **Exploratory analysis**
  - Goal: generation of hypotheses
  - Methodology: Mostly descriptive and flexible approaches to examine data without preconceptions
  - Promotes deeper understanding of processes
  - Major concern: False positives
Case 3. Data sharing for pooled/meta-analysis

- Patient level vs. aggregate data
- Quality
  - Publication bias
  - “Garbage in garbage out” phenomenon
- Heterogeneity
  - “Apples and oranges” phenomenon
  - Combinability of data
- Methodological issues
  - Degree of uniformity approaches
  - Issue of handling heterogeneity
  - Covariate issues:
    - Availability of relevant covariates across studies
    - Difficulties when only aggregate data are available on some of studies
• Personal data privacy
  — De-identification
    • HIPAA standard – well-defined and accepted, but lose information
    • Statistical standard – provides more detailed data, but lacks published methodologies
  — Re-identification
    • Recent demonstrations of re-identification of genetic data
    • Systematic review of re-identification attacks found ≈ 25% success overall and ≈ 35% success for health data
— Potential drawbacks:
  • Limits ability to link clinical trials data with other data sources
  • Limits ability to perform covariate adjusted analyses
Infrastructural Issues in Data Sharing

• Data standards, dictionaries
  – Need to be provided, best if standardized

• Data warehouses
  – Simplifies data aggregation
  – Best if conformed structure

• Portals vs. data transfers
  – Portals simplify access, provide better security, facilitate close management of data (https://www.clinicalstudydatarequest.com/Default.aspx)

• Accessibility of documents
  – Readily accessible of information from sponsor: protocol, SAP, data structure & dictionaries, results, algorithms
Reporting and Dissemination of Results

- Adherence to “fair balance” principles
- Requirements by sponsors to ensure transparency


Guideline for Industry

Structure and Content of Clinical Study Reports

July 1996
ICH E3
Guidelines for reporting data from secondary analyses
  – NEJM guidance for subgroup analyses
Guidelines for Reporting Subgroup Analysis.

In the Abstract:
Present subgroup results in the Abstract only if the subgroup analyses were based on a primary study outcome, if they were prespecified, and if they were interpreted in light of the totality of prespecified subgroup analyses undertaken.

In the Methods section:
Indicate the number of prespecified subgroup analyses that were performed and the number of prespecified subgroup analyses that are reported. Distinguish a specific subgroup analysis of special interest, such as that in the article by Sacks et al.,\textsuperscript{8} from the multiple subgroup analyses typically done to assess the consistency of a treatment effect among various patient characteristics, such as those in the article by Jackson et al.\textsuperscript{9} For each reported analysis, indicate the end point that was assessed and the statistical method that was used to assess the heterogeneity of treatment differences.
Indicate the number of post hoc subgroup analyses that were performed and the number of post hoc subgroup analyses that are reported. For each reported analysis, indicate the end point that was assessed and the statistical method used to assess the heterogeneity of treatment differences. Detailed descriptions may require a supplementary appendix.
Indicate the potential effect on type I errors (false positives) due to multiple subgroup analyses and how this effect is addressed. If formal adjustments for multiplicity were used, describe them; if no formal adjustment was made, indicate the magnitude of the problem informally, as done by Jackson et al.\textsuperscript{9}

In the Results section:
When possible, base analyses of the heterogeneity of treatment effects on tests for interaction, and present them along with effect estimates (including confidence intervals) within each level of each baseline covariate analyzed. A forest plot\textsuperscript{21,22} is an effective method for presenting this information.

In the Discussion section:
Avoid overinterpretation of subgroup differences. Be properly cautious in appraising their credibility, acknowledge the limitations, and provide supporting or contradictory data from other studies, if any.
Secondary analysis:
• Report results in light of available body of knowledge and vis-à-vis original study objectives

• Association is not the same as causation
• Apply Hill’s Criteria
  • Strength of Association
    ➢ Clinical vs. statistical significance
  • Assessment of Consistency
    ➢ Review similar reports for drug or drug class
  • Biological Plausibility
    ➢ Known potential biologic basis to suggest a causal link
• Etc.
“Talking about unsettled science is a much more complicated communications challenge than simply disseminating hard conclusions.”

“The goal should be to effectively communicate findings to patients, healthcare providers, and regulatory agencies to enable informed decision-making.”

Gottlieb, 2006
Concluding Remarks

• Need for access to clinical data should balance legitimate scientific objectives vs. regulatory, logistical and innovation concerns

• Good clinical and statistical principles proven guarantee for transparency

• Best practices for reporting trial results with fair balance further enhance credibility
  – Caution required for reporting data from secondary analyses

• Collaboration among various stake holders essential for effective infrastructure for data sharing
Back-up Slides
Privacy and Confidentiality

“Every precaution must be taken to protect the privacy of research subjects and the confidentiality of their personal information” (Sec 24).

Research Registration and Publication and Dissemination of Results

“Every research study involving human subjects must be registered in a publicly accessible database before recruitment of the first subject” (Sec 35).

“Researchers, authors, sponsors, editors and publishers all have ethical obligations with regard to the publication and dissemination of the results of research. Researchers have a duty to make publicly available the results of their research on human subjects and are accountable for the completeness and accuracy of their reports. All parties should adhere to accepted guidelines for ethical reporting. Negative and inconclusive as well as positive results must be published or otherwise made publicly available. Sources of funding, institutional affiliations and conflicts of interest must be declared in the publication. Reports of research not in accordance with the principles of this Declaration should not be accepted for publication” (Sec 36).
Efficacy Evaluation

11.1 Data Sets Analyzed
11.2 Demographic and Other Baseline Characteristics
11.3 Measurements of Treatment Compliance
11.4 Efficacy Results and Tabulations of Individual Patient Data
  11.4.1 Analysis of Efficacy
  11.4.2 Statistical/Analytical Issues
    11.4.2.1 Adjustments for Covariates
    11.4.2.2 Handling of Dropouts or Missing Data
    11.4.2.3 Interim Analyses and Data Monitoring
    11.4.2.4 Multicenter Studies
    11.4.2.5 Multiple Comparisons/Multiplicity
    11.4.2.6 Use of an “Efficacy Subset” of Patients
    11.4.2.7 Active-Control Studies Intended to Show Equivalence
    11.4.2.8 Examination of Subgroups
11.4.3 Tabulation of Individual Response Data
11.4.4 Drug Dose, Drug Concentration, and Relationships to Response
11.4.5 Drug-Drug and Drug-Disease Interactions
11.4.6 By-Patient Displays
11.4.7 Efficacy Conclusions

Key Elements of E3: Efficacy
SAFETY EVALUATION

12.1 Extent of Exposure

12.2 Adverse Events
   12.2.1 Brief Summary of Adverse Events
   12.2.2 Display of Adverse Events
   12.2.3 Analysis of Adverse Events
   12.2.4 Listing of Adverse Events by Patient

12.3 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events
   12.3.1 Listing of Deaths, Other Serious Adverse Events, and Other Significant Adverse Events
      12.3.1.1 Deaths
      12.3.1.2 Other Serious Adverse Events
      12.3.1.3 Other Significant Adverse Events
   12.3.2 Narratives of Deaths, Other Serious Adverse Events, and Certain Other Significant Adverse Events
   12.3.3 Analysis and Discussion of Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

Clinical Laboratory Evaluation

12.4.1 Listing of Individual Laboratory Measurements by Patient (Appendix 16.2.8) and Each Abnormal Laboratory Value (see section 14.3.4)

12.4.2 Evaluation of Each Laboratory Parameter
   12.4.2.1 Laboratory Values Over Time
   12.4.2.2 Individual Patient Changes
   12.4.2.3 Individual Clinically Significant Abnormalities

Vital Signs, Physical Findings, and Other Observations Related to Safety

Safety Conclusions