A CDER Reviewer's Experience with COVID-19 Programs

Cesar Torres

October 26, 2022

29th Annual Biopharmaceutical Applied Statistics Symposium

DISCLAIMER

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

OVERVIEW

1 Challenges

2 Successes

3 Learning Opportunities

1

CHALLENGES

Standard for Approval (NDAs/BLAs)

- Examples of substantial evidence described in FDA draft guidance
  - Two adequate and well-controlled clinical investigations
  - One adequate and well-controlled clinical investigation plus confirmatory evidence
  - Reliance of FDA's previous finding of effectiveness of an approved drug when scientifically justified and legally permissible

Standard for Emergency Use Authorization

- For FDA to issue an EUA, the chemical, biological, radiological, or nuclear agent(s) referred to in the HHS Secretary's EUA declaration must be capable of causing a serious or life-threatening disease or condition.
- COVID-19 PHE: Based on the totality of scientific evidence available, including data from adequate and well-controlled clinical trials, if available, it is reasonable to believe that
  - The product *may be effective* in diagnosing, treating, or preventing such disease or condition caused by SARS-CoV-2, or a serious or life-threatening disease or condition caused by an FDA-regulated product that is used to diagnose, treat, or prevent a disease or condition caused by SARS-CoV-2; and
  - The known and potential benefits of the product, when used to diagnose, treat, or prevent such disease or condition, outweigh the known and potential risks of the product.
- There is no adequate, approved and available alternative to the product for diagnosing, treating, or preventing the disease or condition.

Complex and Innovative Designs (CIDs)

- Examples of elements
  - Adaptations to patient population, treatment arm selection, randomization ratio, endpoint
  - Formal incorporation of prior information
  - Use of posterior probability to determine trial success criteria
- Operating characteristics (e.g., type I error rate, treatment effect point estimate bias, coverage of 95% interval for point estimate) typically cannot be evaluated using analytical methods
  - Simulation likely necessary
- Process of evaluating CIDs is iterative

Proof-of-Concept (POC) Studies

- Conventional development program
  - Conduct 1+ Phase 2 POC studies to collect preliminary evidence, inform design of future studies
  - Conduct 1+ Phase 3 pivotal studies
- Argument CDER sometimes received: Due to urgent nature of the pandemic, there is no time to conduct POC studies before conducting pivotal studies

Starting at Pivotal Studies

- Designs of pivotal studies might not be adequately informed if POC studies are not first conducted. Examples:
  - Incorrect dosing regimens investigated
  - Underpowered primary analysis
  - Wrong primary endpoint

Unplanned Adaptations

- Unplanned adaptations sometimes led to difficulties in interpretation of analysis results. Example:
  - Changes made after DMC reviewed comparative interim results
    - Even if results from external data sources motivated adaptations, concerns remained

Interpretability

- Primary analysis results not statistically significant, but other analyses had p-values lower than nominal threshold (e.g., two-sided $\alpha = 0.05$). Examples:
  - Subgroup selection
  - Endpoint selection
    - E.g., mortality
  - Covariate selection
- Non-statistically significant ('failed') primary analysis generally leads to above additional analyses to be viewed in a hypothesis-generating light, regardless of whether they are prespecified

Pragmatic Study Designs

- Potential advantages
  - Possibility of substantially larger sample sizes
  - Ability to quickly start randomizing and enrolling patients
- Potential challenges
  - Range of data collected may be limited
  - Proposals often open-label

Data Reliability

- Data reliability/quality issues
  - Hospitals overwhelmed with influx of patients
- Example: Data collected do not always include information such as concomitant medications received by patients

Estimands

- An estimand has five components[1]
  - Treatment regimens to be compared
  - Population
  - Variable (i.e., outcome measure, endpoint)
  - Handling of intercurrent events
  - Population-level summary contrast

[1]https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical

Clinical Questions of Potential Interest

- Compared to control, how does investigational product (IP) perform, when patients and treating physicians do not know the randomized treatment assignment?
  - Double-blind (DB) design addresses this in unbiased manner
- Compared to control, how does IP perform, when patients and treating physicians know whether IP is being received?
  - Open-label (OL) design addresses this in unbiased manner

Relevance of OL Clinical Question

- Might be high if knowledge of IP receipt is not impactful
  - For objective endpoints (e.g., mortality), is it true that knowledge of IP receipt does not impact endpoint determination?
  - Knowledge of IP receipt can influence treating physician decisions
  - Knowledge of IP receipt can influence patient behavior
- If knowledge of treatment receipt is sufficiently influential, DB clinical question may be more relevant
  - Depending on setting, DB design may not be feasible

2

SUCCESSES

Studies Were Concurrently Controlled

- Proposals submitted with designs being uncontrolled, or control arm was external/historical
- Non-concurrent data to include in comparison might be limited due to the following changing over time during pandemic
  - Availability of resources
  - Standard of care
  - Availability of other treatments
  - Prevalence of COVID-19 variant(s)
- Due to importance of minimizing confounding/bias in safety and efficacy evaluations, FDA strongly recommended randomized, concurrent enrollment into a control arm
- Study designs typically double-blind, further minimizing bias

Key Efficacy Endpoints Were Clinically Relevant

- Some proposals included use of primary endpoints that were not direct measures of how a patient feels, functions, or survives
  - Biomarkers proposed for quicker evaluations
  - Difficulty anticipated for translating benefit on biomarker to benefit on clinical outcomes
- With feedback from CDER, study primary endpoints typically clinically relevant

Sample Sizes Considered Mortality Evaluations

- From the beginning, mortality considered an important endpoint for COVID-19 inpatient population
- Other outcomes chosen for primary endpoint, likely due to lower sample sizes needed for adequate powering
- FDA feedback conveyed that regardless of the primary endpoint, mortality evaluations would be important
  - Sponsors considered precision in mortality evaluations when determining sample sizes at the study level or program level

Futility Determinations Were Implemented

- Some programs proceeded to larger pivotal studies without POC studies being conducted first
- Concern: If underlying truth is the IP is not effective or is harmful, then continuing a study to the final analysis led to
  - A waste of resources
  - Preventing patients from entering trials with more promising treatments
- Solution: Conduct interim analyses that allow for possibility of stopping the study early for futility

Futility Determination Rules

- Goal: The probability of determining futility at the interim analysis is
  - Low, if the treatment is truly effective
  - Moderate, if the treatment is truly ineffective
  - High, if the treatment is truly harmful

Changes to Handling Intercurrent Events

- The standard approach to handling death eventually became to use a composite variable strategy, using sufficiently unfavorable value
- For inpatient populations, treatment policy used to handle other intercurrent events (e.g., treatment discontinuation, protocol violations)

High Degree of Collaboration

- Quick publication of Guidance documents for industry
- Quick turnaround times from both FDA and sponsor side

Longer Follow-up for COVID-19 Inpatients

- Initially, patients were followed for roughly 28 days, to correspond with timing of primary endpoint
- Based on FDA feedback, protocols started including provisions to ascertain vital status at Day 60
  - Allowed for comparisons between treatment arms with respect to 60-day mortality

3

LEARNING OPPORTUNITIES

Endpoint Selection

- What considerations are taken into account?
  - Degree to which there is need for modeling assumptions
  - Statistical efficiency
  - Clinical interpretability

Recovering from Moderate-Severe Infection

- Endpoint often recommended by Office of Immunology and Inflammation: Proportion alive and free of respiratory failure at Day 28
  - Difference in probabilities readily interpretable by clinicians, which facilitates benefit-risk evaluations
  - “Alive” and “free of respiratory failure” are both favorable
- Endpoint often prespecified by sponsors: Time to recovery through Day 28
  - Hypothesis tests using time-to-event (TTE) statistical methodology can be more powerful than those using binary endpoint statistical methodology
  - “Recovery” is favorable, while “death” is unfavorable
  - Recovered patients can be re-hospitalized before Day 28

Hazard Ratio

- Hazard function: $\lim_{h \to 0} \frac{P(t \leq T < t+h)}{P(T \geq t)}$
- Hazard ratio (HR): $\frac{\lim_{h \to 0} \frac{P(t \leq T_{IP} < t+h)}{P(T_{IP} \geq t)}}{\lim_{j \to 0} \frac{P(t \leq T_{ctrl} < t+j)}{P(T_{ctrl} \geq t)}}$
- Difference in probabilities: $P(T_{IP} \leq t^*) - P(T_{ctrl} \leq t^*)$
  - Select t^* using hypothesis-generating data from 1+ POC studies
- In some settings, difference in probabilities more clinically interpretable than hazard ratio

Potential Future Research: Alternative TTE Analyses

- Comparing restricted mean survival times
  - Still limited in interpretability
- Parametric modeling (requires further research regarding robustness to violations of parametric assumptions):
  - Accelerated failure time model, with summary contrast being ratio of means or ratio of some quantile (e.g., median)
  - Arm-specific parametric modeling, with summary contrast being difference in unrestricted means

Utility of Conventional Methodology

Being familiar with a particular statistical approach is **not** the same as having a good understanding of it

Odds Ratio

- Odds: $\frac{p}{1-p}$
- Odds ratio (OR): $\frac{\frac{p_{IP}}{1-p_{IP}}}{\frac{p_{ctrl}}{1-p_{ctrl}}}$
- Difference in probabilities: $p_{IP} - p_{ctrl}$
- Difference in probabilities more clinically interpretable than odds ratio

Odds Ratio

- Examples of statements regarding OR
  - *“There is [isn’t] evidence to suggest that the OR is equal to 1”*
  - *“An OR of 5 indicates a greater risk than an OR of 4”*

Odds Ratio

- Rank the following in terms of the RD:
  - $p_c = 0.4, OR = 2$
  - RD ≈ 0.171
  - $p_c = 0.3, OR = 3$
  - RD ≈ 0.263
  - $p_c = 0.2, OR = 4$
  - RD ≈ 0.300
  - $p_c = 0.1, OR = 5$
  - RD ≈ 0.257

Odds Ratio

- Order these all, in terms of the RR:
  - $p_c = 0.2, OR = 1.128$
  - $p_c = 0.4, OR = 1.179$
  - $p_c = 0.6, OR = 1.294$
  - $p_c = 0.8, OR = 1.833$
- They all have RR of ≈ 1.1
- At least somewhat difficult to gauge clinical significance of a specific OR

Utility of Conventional Methodology

Being familiar with a particular statistical approach is **not** the same as having a good understanding of it

Alternative Binary Endpoint Analyses

- Use logistic regression model to derive difference in probabilities[1]

[1]Ge, Miaomiao, et al. "Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences." *Drug information journal: DIJ/Drug Information Association* 45.4 (2011): 481-493.

Ordinal Endpoints

Consider the following ordinal endpoint

1. Not hospitalized
2. Hospitalized, not requiring oxygen support
3. Hospitalized, requiring oxygen support
4. Hospitalized, on mechanical ventilation
5. Hospitalized, requiring organ support
6. Dead

How Best to Analyze Ordinal Endpoints?

- Proportional odds model commonly proposed
  - Summary contrast: common odds ratio (COR), which is the odds ratio when dichotomizing the ordinal endpoint, under the proportional odds assumption[1]
- If the odds ratio from a logistic regression model has limited clinical interpretability, then so does the COR from the proportional odds model

[1]The assumption that the odds ratio is the same regardless of whatever dichotomization is used.

Alternative Analyses for Ordinal Endpoints

- Change from baseline (e.g., using an ANCOVA model)
  - Still limited in interpretability
- Partial credit strategy[1]

[1]Evans, Scott R., and Dean Follmann. "Using outcomes to analyze patients rather than patients to analyze outcomes: a step toward pragmatism in benefit: risk evaluation." *Statistics in biopharmaceutical research* 8.4 (2016): 386-393.

Partial Credit Strategy (PCS)

- Basic idea: Assign a numeric score to each value on the ordinal scale
  - Give ordinal value of 1 (not hospitalized) a score of 100
  - Give ordinal value of 6 (dead) a score of 0
  - Determine scores for each of the remaining ordinal values
- Ordinal scale has been mapped to a utility function!
  - Can compare average utility between treatment arms, using a difference in mean utility

Further Research Needed on PCS

- How to best elicit feedback from clinicians in a structured manner to determine appropriate scores for remaining ordinal values?
- What are the operating characteristics (e.g., asymptotic distribution of test statistic, asymptotic bias of treatment effect point estimate, asymptotic coverage of corresponding 95% interval) of inferential procedures when using the PCS?

Key Takeaways

- There have been challenges and successes in the evaluation of candidate treatments for COVID-19 infection
- Public health is best served when prioritizing clinical considerations (e.g., clinical interpretability)
- Looking forward, statistical research should focus on methods that will lead to clinically interpretable results


FDA
U.S. FOOD & DRUG
ADMINISTRATION