A (Constructive/Provocative) Critique of the ICH E9 Addendum

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

April 18, 2018

Disclosures

- Regularly consult with pharmaceutical and device companies
    - DSMB's
    - Mock advisory panels
    - Data analysis
    - Study design
    - Training
    - Litigation
- Served on the National Academy of Sciences (NAS) Panel that issued the report "The Prevention and Treatment of Missing Data in Clinical Trials"
- Funding from NIH, PCORI and FDA for the development of statistical methods and associated software.

Acknowledgements

- Miguel Hernan
- Johns Hopkins Causal Inference Group
- Nicholas King - McGill
- Jonathan Darrow - Harvard
- Jeremy Sugarman - Johns Hopkins
- Raja Srinivasa - John s Hopkins
- Frank Bretz - Novartis
- Rob Hemmings - EMEA
- Cartoons from the Internet

Quick Response to Devan and Tom

- Big applause for "What is the question?"
- "Provocative/constructive" aimed at influencing the revision
- "Treatment policy" is a really bad term
- Principal stratification is scientifically interesting but just too "assumption-laden" to be "primary"

ICH and E9 Addendum

- "... bring[s] together regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of drug registration"
- Issues guidelines
- No academic input in drafting E9 Addendum
- Comment period

EMEA/PMDA

Estimand

- A population parameter that quantifies the *effect* of treatment relative to control.
- NAS Report
  - Target of inference in a randomized clinical trial
  - *Causally* interpretable
  - Motivated the ICH Addendum
- ICH Addendum
  - Avoids the word "causal" ("C-word")
  - But uses the language of causal inference
  - *"What would have happened to the same subjects under different treatment conditions?"*
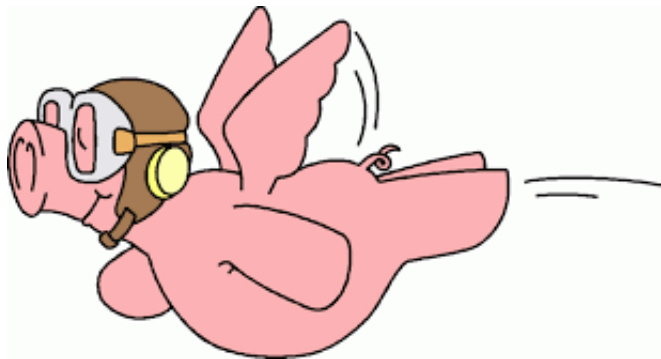
Causal Inference

- Mathematical language for
    - defining causal estimands
    - formulating <u>identification</u> assumptions
- Extension of standard mathematical language of statistics
- Typically formulated in terms of potential outcomes, where each individual is defined to have outcomes under treatment and under control.
- Estimand is a contrast of a feature (e.g., mean) of the population distributions of the two potential outcomes.

Intention to Treat (ITT)

- Participants are analyzed in their assigned groups regardless of actual receipt of their assigned treatment
- Has historically been favored because it alleviates concerns about baseline confounding (see ICH E9).
- Estimates the effect of *assignment* to treatment versus *assignment* to control to an intervention rather than that of actually adhering to it.
- The magnitude of the effect depends on the degree to which participants adhered to their assigned treatments.
- Can miss effects, positive or negative.

ICH E9 Addendum

- Motivated by the deficiencies of ITT, the ICH E9 Addendum introduces a framework that allows consideration of a variety of alternative measures of treatment effects (i.e., estimands) to be estimated from data collected from randomized clinical trials.

- The Addendum defines four attributes of an estimand:
  1. Population
  2. Endpoint
  3. Effect measure
  4. Approach to handling "intercurrent events" (e.g., non-adherence, competing risks).

Treatment Strategies

- Most drug studies involve treatments that must be sustained over time.
- The Addendum would be better served by focusing on clinical relevant treatment strategies, defined so that all patients have the potential to be adherent.
- A patient who does not take treatment due to contra-indications should not be considered non-adherent.
- A treatment strategy might take the form:
  - "do not take treatment"
  - "take treatment until contra-indications occur"
- It might also involve flexible dosing and specify allowable adjuvant/concomitant therapies

Estimand

A better set of attributes for an estimand would be:

1. Target Population: entire population with disease or a subgroup of this population based on a principal stratum
2. Treatment Strategy: Precisely defined in the study protocol
3. Outcome: Post-randomization events can be factored in provided that clinical relevance is maintained
   1. ITT: outcome regardless of adherence to the treatment strategy
   2. Composite: defined to include the occurrence of key post-randomization event(s)
   3. Counterfactual: outcome under full adherence to the treatment strategy
   4. While Adherent: outcome during adherence to the treatment strategy
4. Effect Measure: as defined in the Addendum

Principal Stratification

- Subgroup of the population based on potential outcomes.

  - Compliers - patients who comply with their assigned treatment, whatever it be.
  - Treatment Compliers - patients who comply when assigned treatment
  - Survivors - patients who survive regardless of their assignment treatment

- Non-identifiable subgroup of the diseased population.
- Requires extremely strong assumptions for identification

Approval Based on Non-Identifiable Subgroup

- Potentially exposing a subset of patients to an ineffective treatment.
- Already an issue when approval is based on an effect based on the overall population.
- If treatment doesn't have serious side effects, then there shouldn't be ethical concerns.
- This already occurs in enrichment trials.
- How does one write the treatment label?

Approval Based on Principal Stratification

- Assumptions are too strong to merit use as primary estimand.
- Lowers the level of evidence.
- Why not just conduct an enrichment study?

1   INDICATIONS AND USAGE
OPANA ER is indicated for the relief of moderate to severe pain in patients requiring continuous, around-the-clock opioid treatment for an extended period of time.

Limitations of Usage
OPANA ER is not intended for use as an as needed analgesic.

OPANA ER is not indicated for pain in the immediate post-operative period if the pain is mild, or not expected to persist for an extended period of time.

OPANA ER is only indicated for post-operative use if the patient is already receiving the drug prior to surgery or if the post-operative pain is expected to be moderate or severe and persist for an extended period of time. Physicians should individualize treatment, moving from parenteral to oral analgesics as appropriate. (See American Pain Society guidelines).

ICH Addendum and Sensitivity Analysis

- Does not go far enough
- Suggests that it is sufficient to explore the sensitivity of inference under a few assumptions
- Rigorous sensitivity analysis should be conducted - tipping point analysis.

My Recommendation

- Design studies with clinically relevant treatment strategies where *all* patients have the potential to adhere.
- If a patient fails to adhere, then incorporate into the outcome.
- Carefully monitoring of patient adherence.
- The benefits of randomization can be largely preserved.
- Missing data can be minimized.
- igorous sensitivity analysis to missing data assumptions can be employed.
- Put onus on sponsor to maximize adherence and minimize missing data.

Case Study: Pain

- Double-blind, placebo-controlled, randomized trial to evaluate the effectiveness of a new drug (at a fixed dosage level) for patients with chronic pain.
- Typically considered unethical to prohibit the use of rescue medications.
- Allowing the use of rescue medications as needed, however, can make it difficult to detect whether the new drug yields a beneficial effect.
- In the extreme, suppose all patients on placebo initiate rescue medications, while those on new drug do not. The ITT effect on pain is then a comparison of the new drug vs. rescue medications.

Case Study: Pain

- "therapeutic window can vary considerably among patients exposed to the same medication due to variability in absorption, metabolism and physiologic distribution of analgesic medications resulting from variability in age, sex, genetics and body weight and interactions between drug and disease"(Dworkin et al., 2010).
- This can lead to issues of tolerability and lack of efficacy, possibility leading to study withdrawals.

Case Study: Pain

- These issues can be addressed using the ideas of Dworkin et al. (2010) and Farrar, Dworkin and Max (2006).
- Studies can be designed with a structured approach to concomitant analgesic treatment use:

    "A third approach would therefore allow use of only a limited number of specific analgesic medications (e.g., those approved by regulatory agencies or with generally accepted evidence of efficacy) at stable dosages throughout the trial and to specific maximum dosages for each. " (Dworkin et al., 2010)

Case Study: Pain

- Studies can be designed using flexible dosage strategies:

  " ... flexible dosing to effect and tolerability may be advantageous compared with using fixed dosage that can exceed the therapeutic window and cause withdrawals due to adverse events in some individuals, or undershoot the therapeutic window and cause withdrawals due to lack of efficacy in others. Moreover, flexible dosing reflects clinical practice because the dosages subjects receive are adjusted on the basis of both effectiveness and tolerability." (Dworkin et al, 2010)

- The structured approach to concomitant medication use plus flexible dosing forms a well-defined treatment strategy, which will have to be monitored carefully.

Case Study: Pain

- Patients who cannot adhere to the strategy can be considered failures.
- A rank-based analysis can be employed where these individuals are assigned the worst rank.
- Treatments can be then compared with respect to the cumulative distribution function approach of Farrar, Dworkin and Max (2006) and Permutt and Li (2017) - this approach only depends on the ability to rank outcomes.

Conclusion - ICH E9 Addendum

- "moving the ball forward" or "moving the goal posts"?
- I am concerned that it could lead to the regulatory approval of drugs/devices based on weaker evidence.
- I believe that many existing problems with clinical trials can be largely solved through better design and execution.