Some Thoughts on Implementation of E9(R1)

Thomas Permutt
Associate Director for Statistical Science and Policy
Office of Biostatistics
Office of Translational Sciences
Center for Drug Evaluation and Research
Regulatory Context

• E9(R1) is authoritative (means what it says)
  – But not yet definitive (will be revised)
• FDA policy will conform to E9(R1) as revised
  – Possible regional guidance
  – Not yet under way
  – Formal process for dissemination
• No new research to be presented
• So what is the purpose of this talk?
  – Points to consider
Comparisons

• **Causal estimand** means treatment effect

• **Treatment effect** means comparison of outcomes under different treatments
  – In the same subjects (Rubin causal model) or ...
  – In comparable subjects

• Kinds of comparison
  – External comparisons
  – Nonrandomized comparisons
  – Randomized comparisons
Kinds of Comparison

• External comparisons
  – To something not in study
  – Very, very hard

• Nonrandomized comparisons
  – Between not necessarily comparable groups
  – Very hard

• Randomized comparisons
  – Between necessarily (on average) comparable groups
  – Easy, but not always exactly what is wanted
Kinds of Comparison

• External comparisons
  – Very, very hard, but sometimes needed

• Nonrandomized comparisons
  – Very hard

• Randomized comparisons
  – Easy, but not always exactly what is wanted
  – But there may be more choices than are apparent

• Nonrandomized comparisons
  – An easy one!
External Comparisons

- **Historical controls**
  - Because the randomized study is impractical or unethical
  - Rely on constancy of outcome

- **Noninferiority—Putative placebo**
  - Because the randomized study is impractical or unethical
  - Rely on constancy of effect (of comparator vs. placebo)

- **If no rescue (“hypothetical”)**
If No Rescue

• Not in patients not needing rescue
• Rather, in a study (world?) without rescue
  – Is this relevant?
    • Sometimes
    • Compare to noninferiority
• What constancy assumption is needed?
  – Quasi-internal comparison (before rescue)
  – What sensitivity analysis is needed?
• Widely known methods may not be enough
Nonrandomized Comparisons

• Goldilocks epidemiology
• Adjusting for confounders
• What to adjust to
Goldilocks Epidemiology

MCAR
Goldilocks Epidemiology

MNAR

MCAR
Goldilocks Epidemiology

Flowchart: MCAR (Missing Completely at Random), MAR (Missing at Random), MNAR (Missing Not at Random), OK (as a reference point).
Goldilocks Epidemiology

MCAR

MAR

MNAR

OK

true

sensitivity
Goldilocks Epidemiology

MCAR

MAR

OK

not OK

true

MNAR
Nonrandomized Comparisons

• Are hard
• Are necessary in epidemiology
  – May not be necessary in randomized trials even with dropouts
  – Because this is not missing data
• Adjust for everything in sight
• Then worry about what’s not in sight
• No unmeasured confounders means
  – Measured, and
  – Correctly modeled
Nonrandomized Comparisons

• Can use MAR techniques
• Can’t just hypothesize MAR
What to Adjust to

outcome

LSmean

LSmean

adjusted effect

\( \gamma_c \)

\( \gamma_c \)

mean

confounders
What to Adjust to

mean

not mean

multiple deltas then average

It doesn’t matter!
Average Treatment Effect, Plug-In Estimator

• Fit curves for treatment groups
• Plug in some observed values of confounders
• Calculate deltas
• Average deltas
Which Observed Values of Confounders to Adjust to?

• Matters some when treatment effects vary with confounders
• Differently weighted averages of valid treatment effects
• But not adjusting gives confounded effect
What to Adjust to?

• Effect among the treated
  – Population that would adhere to test drug, whether or not they would adhere to control
  – Other populations possible
    • If all patients adhered
  – See SBR paper

• Important thing is to adjust
  – Because it’s nonrandomized comparison
  – Doesn’t matter very much to what
  – Do not accept an informal interpretation of “if all patients adhered”
    • Sounds more like intent-to-treat, but it isn’t (nonrandomized)
    • Different kind of hypothetical
Randomized Comparisons

- Classic—ITT
- Modified
Intent to Treat

• Don’t redefine it
  – Get all primary outcome data
  – But missingness is inevitable, so ...
  – Just use what you have without exclusions
  – Serious follow-up and adjustment

• Minimize missing data
  – Minimize loss to follow-up
  – Do not minimize nonadherence
Modified Randomized Comparisons

• While on treatment
• Dropout as failure
While on Treatment

• Last observation
  – Not carried forward

• Average observation
  – ≈ MMRM main effect
  – Not MAR (wrt simple model)
  – Not de jure
    • if that exists, ...
    • MMRM still doesn’t estimate it ...
      – Except under MAR
      – Maybe even under MAR (see SBR paper)
While on Treatment

• Fine statistically
  – Unbiased (but ...)
  – Preserves Type I error (but ...)

• Meaningful estimand/hypothesis?
  – Yes—Discontinuation due to cure
  – No—Discontinuation due to failure (toxicity/lack of efficacy)
  • Especially when some failures are given good scores and others bad scores
  – Measure benefit!
Dropout as Failure

• Need not dilute treatment effect
• Need not lose much information
  – Possible gain in information compared to handling as “missing” (but see Cro et al.)
• Trimmed mean, rank methods
• Compare to survival
  – Don’t treat survival as “missing” time to event
  – Don’t impute time of death for surviving patients!
  – Dropout considered as failure is censoring/competing risk, not missingness
Easy Nonrandomized Comparison

• Nonrandomized comparisons are hard
  – Because “no measured confounders” is a very strong assumption
• But there might be an easy one
  – Be suspicious!
Easy Nonrandomized Comparison

• Suppose:
  – ITT effect (all patients followed and counted) is 5
  – Half the patients on active drug adhere
  – It doesn’t matter whether you adhere to placebo
    • Not: placebo nonadherents are like placebo adherents
  – You have to adhere to active drug for it to work

• Then the effect in adherents is diluted by the non-effect in nonadherents

• How much is it diluted?
How Much Dilution?

• ITT effect = 
  \( (\text{proportion of adherents})(\text{effect in adherents}) + (\text{proportion of nonadherents})(\text{effect in nonadherents}) \) 
  = \( (0.5)(y) + (0.5)0 = 5 \)

• \( y = 10 \)
Two Questions

• Why follow after discontinuation if interest is in adherent subjects?

• How can nonrandomized comparisons ever be easy?
Nonrandomized Comparison
Hard Because ...

- Need to adjust for confounders
- What is the best confounder to adjust for?
Nonrandomized Comparison
Hard Because ...

- Need to adjust for confounders
- What is the best confounder to adjust for?
  - Good predictor of (placebo) outcome
- But there is perfect predictor of placebo outcome in dropouts!
  - If they are followed
Estimands/Effects/Comparisons

• External comparisons
  – Very, very hard, but sometimes needed

• Nonrandomized comparisons
  – Very hard

• Randomized comparisons
  – Easy, but not always exactly what is wanted
  – But there may be more choices than are apparent

• Nonrandomized comparisons
  – An easy one!
References


• Permutt T. Effects in Adherent Subjects. Statistics in Biopharmaceutical Research. 2018 (what to adjust to, what MMRM estimates)