A meta-analytic framework for decision making and error control in clinical trials with external control arms

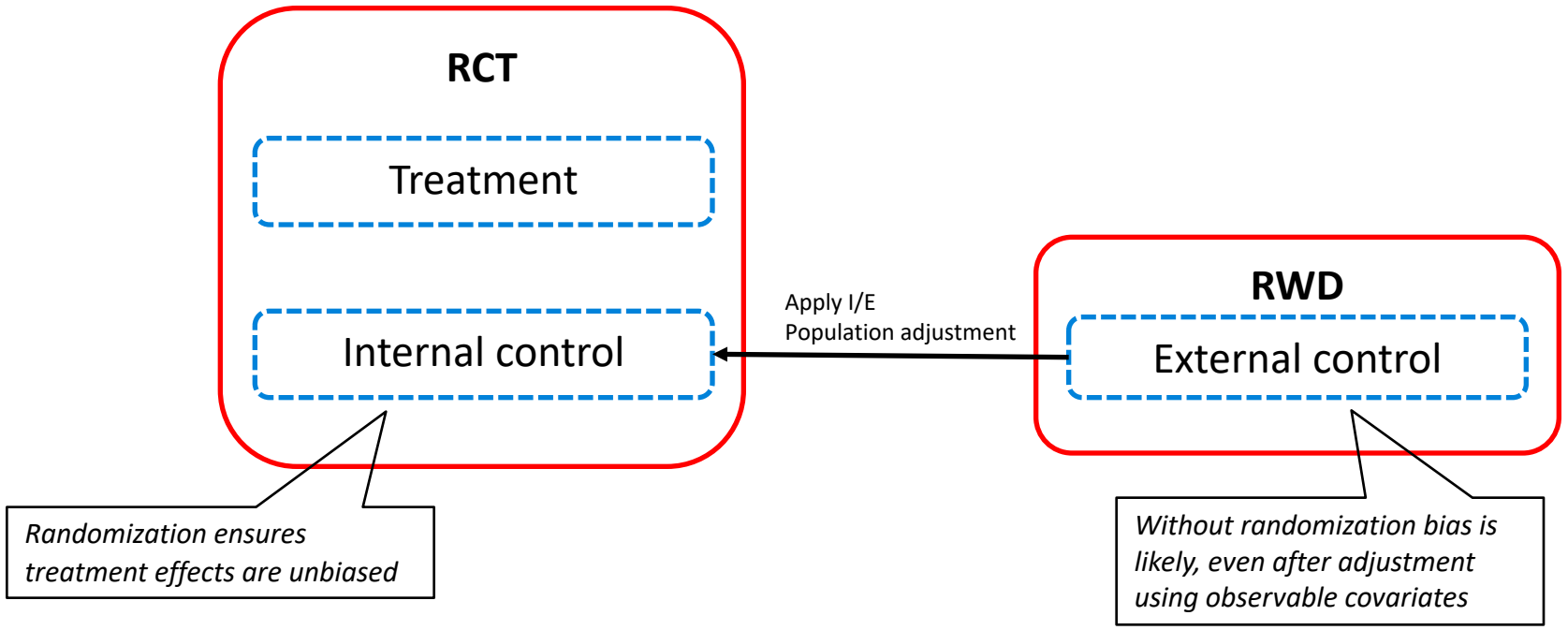
Devin Incerti | Roche & Genentech

June 23, 2021 | Leveraging Observational Data with Machine Learning

Acknowledgements

This work reflects the efforts of our core research team: Chris Harbron, Michael Bretscher, and Ray Lin

Using external controls to estimate treatment effects



An external control arm is used to augment (hybrid control design) or replace (external control design) the internal control arm

RCT = randomized clinical trial; RWD = real-world data

How are external controls used for decision making?

- *Randomized controlled trials* are the gold standard for unbiased estimation of treatment effects
- However, *single arm* trials are used in some contexts and high-quality external control studies can improve decision making in these contexts
- Single-arm trials are primarily used for regulatory decision making when RCTs are either unethical or infeasible
  - High unmet need
  - Scarcity of patients
  - Randomization is unethical
- Single-arm trials are common in early phase trials and external controls can inform internal decision making

A problem with external control analyses: trial patients may differ from RWD patients in ways that cannot be adjusted for

Study specific biases

Systematic biases

The process of being in a trial

- Patient selection
- Site selection
- Higher levels of attention

Measurement biases

- Some variables more likely to be captured in trials than clinical practice (e.g., ECOG)
- Trial data collected more frequently
- Variables measured in different ways (e.g., PFS)

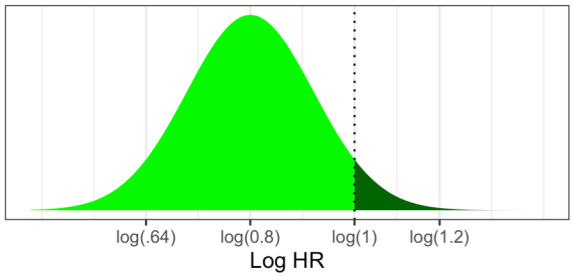
Unmeasured and unknown biases

- Unobserved heterogeneity is large in medicine and likely correlated with data source
- Difficult to bound the size of unmeasured confounding

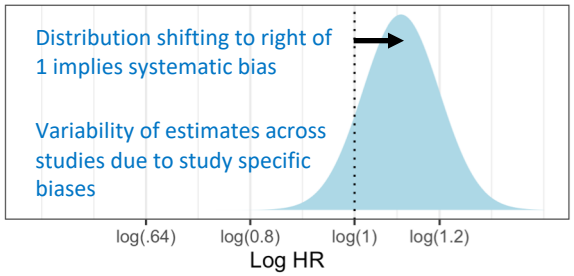
Using historical data to incorporate additional bias and uncertainty from non-randomized data

Can we adjust for systematic and study specific biases?

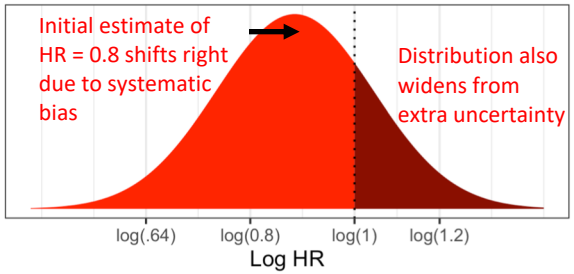
1. ”Naively” estimate log HR in comparison of treatment to external control in **new** single-arm study



2. Use **reference studies** to estimate log HR in comparison of external to internal control:



Adjust (1) using (2) to incorporate additional bias and variability from non-randomized design



A meta-analytic approach to compare internal and external control arms from historical studies

Aim is to adjust (log) HRs, λ, from a new study[1]

$$\lambda^{new}_{TRTvIC} = \lambda^{new}_{TRTvEC} - \lambda^{new}_{ICvEC} \qquad \text{(Equation 1)}$$

Typical estimate — Adjustment

Model & estimation

Use meta-analytic model for distribution of λ_{ICvEC}. Estimate from a set of reference studies[2]

$$\hat{\lambda}_{ICvEC,j} \sim N(\lambda_{ICvEC,j}, V_j)$$

Point estimate — "Truth" — Standard error

$$\lambda_{ICvEC,j} \sim N(\mu, \sigma^2)$$

Systematic bias — Between study variability

Prediction

A prediction for the new study is then made as follows:

1. Draw λ^{new}_{TrtvEC} from the posterior

$$\hat{\lambda}^{new}_{TRTvEC} \sim N(\lambda^{new}_{TRTvEC}, V^{new})$$

2. Draw λ^{new}_{ICvEC}

$$\lambda^{new}_{ICvEC} \sim N(\mu, \sigma^2)$$

3. Use Equation 1 above

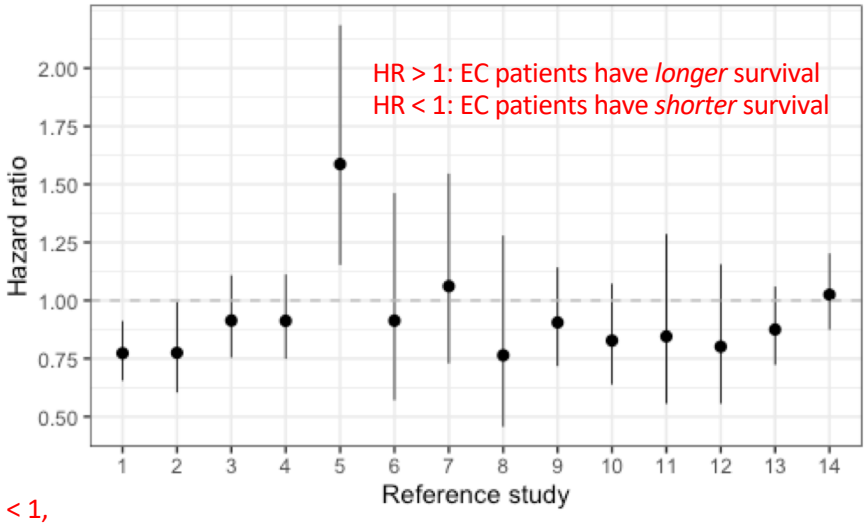
[1]Relationship will hold exactly for true log HRs under proportional hazards

[2]Use Bayesian model with non-informative hyperpriors for estimation

Model estimation using 14 “reference studies” for treatment of advanced non-small cell lung cancer

- Use 14 reference studies based on randomized phase II/III clinical trials for advanced NSCLC
- Each trial had treatment arm and internal control (IC) arm; external controls (ECs) built from Flatiron Health
- For each study, applied I/E based on trial and used propensity score methods to estimate HRs comparing the IC to the EC[1,2]
- Prespecified primary analysis used IPTW-ATT weights and removed EC patients with propensity scores below the 1st percentile or above the 99th percentile

Hazard ratios (internal vs. external control)



Overall bias suggests HR < 1, implying EC patients tend to have shorter survival

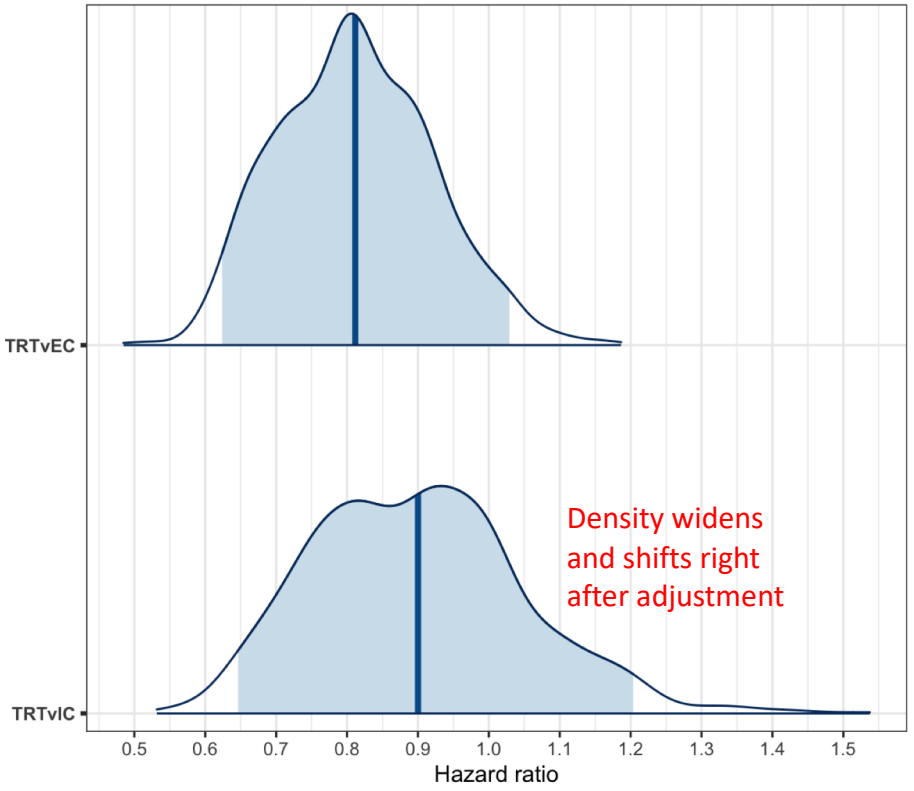
Parameter estimates for meta-analytic model

	Mean	SD	2.5%	97.5%
μ	-0.100	0.042	-0.181	-0.018
σ^2	0.082	0.056	0.021	0.313

[1]Covariates were age, sex, race, smoking status, histology, cancer stage, histology, and time since diagnosis
[2]Other propensity score methods included IPTW-ATT weighting (without trimming), 1:1 neareast neighbor matching (with/without caliper) and 1:1 genetic matching (with/without caliper)

Example prediction for a hypothetical new study

- Consider a hypothetical new study with an estimated log HR of 0.80 and standard error equal to the median from the reference studies (0.136)
- The adjustment shifts the density to the right and increases the width of the uncertainty interval
- Results in a HR that is less "optimistic"



	Mean	SD	2.5%	97.5%
TRTvEC	0.80	0.11	0.62	1.03
TRTvIC	0.90	0.15	0.65	1.22

Note: Shaded region in plot denotes 95% credible intervals and vertical lines are medians; table reports summaries of the posterior distribution of the HRs

Challenges and considerations

- How can we identify historical reference studies?
  - How similar should these studies be to the new trial?
  - What characteristics should govern selection criteria (e.g., phase, population characteristics, line of therapy, etc.)?
- How many reference studies are needed to reliably estimate the meta-analytic model?
- What if hazard ratios are not proportional for some of the reference studies (or for the new trial)?
- Is it more difficult to create prespecified analysis plans when using this methodology?
  - Requires identification of trials
  - If a relevant study has not already been performed, will also require specification of a consistent methodology to estimate log HRs (internal control vs. external control) across reference studies

Conclusions

- Randomized controlled trials remain the gold standard, but single-arm studies are useful for both regulatory and internal decision making in certain contexts
- External controls are needed to evaluate efficacy in single-arm trials, but they create additional bias and variability due to their non-randomized design
- We developed a meta-analytic methodology that use historical reference trials to adjust for this bias and variability
- May increase acceptance of external control analyses and can improve decision making
- Hybrid designs that dynamically borrow information from external controls based on their compatibility with an internal control arm are an alternative option for addressing unmeasured sources of bias that can balance type I and type II error

Doing now what patients need next