External Control Arm Methods in Practice

Emma Mackay
Scientific Advisor - Statistics
Inka Health

Presentation to the Xu Lab at UHN

Sept. 10, 2024

Disclosures

- Advisor at Inka Health, an HEOR analytics consulting company
- No conflicts of interest related to this presentation
- Opinions expressed are my own

Presentation Roadmap

- Part I: Introduction to external control arms (ECA)
  - What is an ECA?
  - Potential Outcomes Framework
  - Design principles
- Part II: Implementation challenges
  - Operationalizing inclusion/exclusion criteria for ECAs
  - Practical challenges of adjusting for confounders
- Part III: Introduction to Bayesian borrowing and hybrid approaches

What is an ECA?

- In some rare disease settings it may be impractical or infeasible to run a randomized controlled trial (RCT)
- Instead of an RCT, idea is to compare outcomes in a single-arm trial vs. an appropriately constructed external control arm (ECA)
  - May use real-world data (RWD) or historical trials to construct ECA
  - Care needs to be taken to mitigate risk of bias in absence of randomization
- Regulatory and health technology assessment (HTA) bodies have indicated potential receptiveness to the use of ECAs where RCTs are infeasible[1][2][3]

[1] US FDA. Guidance for Industry. E 10 Choice of Control Group and Related Issues in Clinical Trials. 2001. link here

[2] US FDA. Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological.... Draft Guidance. 2023. link here

[3] UK NICE. NICE real-world evidence framework. 2022. link here

- Want to assess the causal effect of being treated vs. untreated / SoC
- Define the treatment indicator $A_i \in \{0, 1\}$ for patient i
- Consider patient's potential outcomes Y_{1i} and Y_{0i} if they were treated or untreated, respectively–only observe one of these in practice and the other is counterfactual, i.e. the observed outcome Y_i is:

$$Y_i = A_i Y_{1i} + (1 - A_i) Y_{0i}$$

- We would like to estimate the average treatment effect, $ATE = \mathrm{E}[Y_1 - Y_0]$

Assumptions

- We will need a few assumptions to hold:
  1. Exposure precedes outcome temporally
  2. Positivity assumption: each patient must have non-zero probability of appearing in both the treated and untreated groups
  3. Stable unit treatment value assumption (SUTVA): patient's potential outcomes are not affected by the exposure status of other patients
  4. (Conditional) ignorability assumption: $\{Y_0, Y_1\} \perp A$ (or $\{Y_0, Y_1\}|X \perp A$) where X are covariates measured at baseline

Assumptions

- We will need a few assumptions to hold:
  1. Exposure precedes outcome temporally
  2. Positivity assumption: each patient must have non-zero probability of appearing in both the treated and untreated groups
  3. Stable unit treatment value assumption (SUTVA): patient's potential outcomes are not affected by the exposure status of other patients
  4. (Conditional) ignorability assumption: $\{Y_0, Y_1\} \perp A$ (or $\{Y_0, Y_1\}|X \perp A$) where X are covariates measured at baseline

- In a randomized controlled trial (RCT) setting we can use careful design to ensure that these assumptions are likely to be satisfied so that

$$\mathrm{E}[Y|A=1] - \mathrm{E}[Y|A=0] = \mathrm{E}[Y_1|A=1] - \mathrm{E}[Y_0|A=0] = \mathrm{E}[Y_1 - Y_0] = ATE$$

Assumptions

- We will need a few assumptions to hold:
  1. Exposure precedes outcome temporally
  2. Positivity assumption: each patient must have non-zero probability of appearing in both the treated and untreated groups
  3. Stable unit treatment value assumption (SUTVA): patient's potential outcomes are not affected by the exposure status of other patients
  4. (Conditional) ignorability assumption: $\{Y_0, Y_1\} \perp A$ (or $\{Y_0, Y_1\}|X \perp A$) where X are covariates measured at baseline

- In a randomized controlled trial (RCT) setting we can use careful design to ensure that these assumptions are likely to be satisfied so that

$$E[Y|A = 1] - E[Y|A = 0] = E[Y_1|A = 1] - E[Y_0|A = 0] = E[Y_1 - Y_0] = ATE$$

- However, in observational study settings it is much more difficult to ensure that these conditions are met–especially for assumptions (2) and (4)
  - Additionally, we will want to use the weaker version of (4) and adjust for X

Target Trial Emulation Principles

- Considerable planning goes into developing an RCT protocol and statistical analysis plan, including specifying:
  - Eligibility criteria and how they will be assessed
  - Measurement of patient patient characteristics at baseline
  - Randomization to treatment and specifying doses and any blinding procedures
  - Definition of how endpoints will be measured
  - Procedures for collecting repeated measures and documenting adverse events
  - Pre-specification of analysis plan
- When conducting an ECA analysis we lack the ability to tightly control all of these factors through prospective design and data collection

Target Trial Emulation Principles

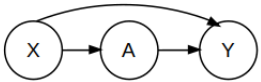
- Target trial emulation[4] is the practice of emulating as closely as possible aspects of an ideal RCT when conducting an observational study, including:
  - A well-defined period at which eligibility is assessed (to avoid immortal time bias)
  - Applying harmonized eligibility criteria across data sources
  - Careful specification of an index date (time zero) and comparable definitions of endpoints across data sources
  - Compatible definitions of baseline measures and a well-defined baseline period
  - Definition of treatment regimes (e.g. cancer lines of therapy)
  - Use methods to adjust for potential confounders in the absence of randomization
- In practice this is an iterative process due to the challenges that come from working with disparate data sources–especially real-world data (RWD)

[4] Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. American journal of epidemiology. 2016.

Motivation for Inverse Probability of Treatment Weighting (IPTW)

- For simplicity, suppose we have two types of patients in our dataset–those who are high-risk at baseline (red) and low-risk (blue)

- Naive comparison of outcomes by treatment group would underestimate the treatment effect since poor prognosis patients are overrepresented in the treated group
- Can represent this scenario using a directed acyclic graph (DAG):



- Goal is to re-weight patients in the treated and untreated groups such that baseline characteristics are balanced in the weighted pseudo-population, breaking the link between X and A

- We first need to compute the probability of treatment ($A_i = 1$) for each patient given their baseline characteristics (X): $\Pr(A_i = 1 | X_i = x)$
- We call $e(x) = \Pr(A_i = 1 | X_i = x)$ a propensity score

	Treated			Untreated		
Patients						
$e(x)$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

- Next we need to compute the IPTW weights
- We give each patient i a weight, w_i, which is inversely proportional to their probability of being in their treatment group given their baseline characteristics:

$$w_i = 1[A_i = 1]\frac{1}{e(x)} + \Big(1 - 1[A_i = 1]\Big)\frac{1}{1 - e(x)}$$

	Treated			Untreated		
Patients						
$e(x)$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
Weights	$\frac{3}{2}$	$\frac{3}{2}$	3	3	$\frac{3}{2}$	$\frac{3}{2}$

- Lastly, we apply these weights to each patient to yield a weighted pseudo-population

- Note that:
  - The number of high-risk and low-risk patients are balanced between treatment groups in the pseudo-population–A and X are no longer associated
  - Notice The pseudo-population has twice as many patients in each group–this is because the average weight is 2

- With the link broken between X and A we can now compare outcomes in our treated vs. untreated pseudo-populations–e.g. a difference in weighted means or via a weighted (generalized) regression model
- This regression model is called an 'outcome regression model'
- However, we need to be careful to account for the weighting appropriately when computing standard errors and conducting hypothesis tests. Two common approaches:
  - Robust / sandwich covariance estimators
  - Bootstrapping

- Remember that a patient with a weight of 2 is still only contributing 1 patient's worth of information–check how weights are being handled!
- It is the relative rather than absolute weights that are important
  - Can rescale weights by a factor that does not depend on X and still have a valid pseudo-population
  - A common choice is to use stabilized weights:

$$w_i = 1[A_i = 1]\frac{\Pr(A_i = 1)}{e(x)} + \Big(1 - 1[A_i = 1]\Big)\frac{1 - \Pr(A_i = 1)}{1 - e(x)}$$

- Useful formula for effective sample size (ESS): $ESS = \frac{\left[\sum_{i=1}^{n} w_i\right]^2}{\sum_{i=1}^{n} w_i^2}$
  - Note that if the n weights are all equal, $ESS = n$, and if one patient has nearly all of the weight, $ESS \to 1$

Estimating Propensity Scores

- Non-parametric estimation of the propensity score may be impractical or infeasible
- Solution: estimate the propensity score using a parametric model
- Basic approach using logistic regression:
  - Fit logistic regression of A on X (including transformations of underlying covariates)

$$A_i \sim \text{Bernoulli}(p_i) \qquad \forall\ i = 1, ..., n$$
$$\text{logit}(p_i) = \mathbf{x_i}'\beta$$

  - Compute fitted values $e_i(\mathbf{x_i}) = \widehat{p_i}$ on the inverse-logit (0-1) scale
- Can think of the propensity score as a dimensionality reduction method

Evaluating Our Constructed ECA

- Application of IPTW (or other methods like matching) does not guarantee that balance will be achieved between treatment groups
- After constructing our pseudo-population we should always
  - Perform diagnostics–are there (near) violations of positivity?
  - Assess balance in patient characteristics and possible omitted variables bias
  - Evaluate ESS
- Even with these checks it is difficult in practice to assess whether the joint distributions of baseline characteristics are comparable between treatment groups

- Weights used so far were for ATE estimands–reflecting effect in the combined treated + untreated population
  - ATE pseudo-population had 50-50 high-risk to low-risk split for both treated and untreated (rather than $\frac{2}{3}$ high-risk in treated group or $\frac{1}{3}$ in untreated)

- Can also get average treatment effect on the treated (ATT) if we use slightly modified weights:

$$w_i = 1[A_i = 1] + \Big(1 - 1[A_i = 1]\Big)\frac{e(x)}{1 - e(x)}$$

- Other weighting approaches can be used for other estimands too

Estimands and Effect Modification

- Why do we care about these different estimands?
- Recall high-risk proportion was $\frac{2}{3}$ in treated and $\frac{1}{3}$ untreated groups
- Suppose the treatment effect relative to no treatment is smaller for high-risk patients
  - Implies that treatment effect estimate we are aiming at will vary depending on the proportion of high-risk patients
  - Risk status an 'effect modifier' for the outcome
- It's clear our exposed and unexposed groups reflect different patient populations. What is the target population of interest where we are considering rolling out the treatment?

Outcome Regression Methods

- An alternative to IPTW would be to fit an outcome regression model which controls baseline confounders
- A few limitations to this approach:
  - Unlike IPTW, cannot assess whether we are able to adequately adjust for imbalances in observed covariates, nor can we do this in a manner where the outcome data is held out
  - Is misspecification more likely to be an issue in the outcome regression model or propensity score regression model?
  - Outcome regression produces conditional estimates in contrast to marginal estimates like ATE and ATT (and ITT from RCTs)
- However, it's possible to extend this approach to compute marginal estimates (see marginal structural model (MSM) literature)

What About Matching Methods? (1 / 3)

- More popular among clinicians due to ease of interpretability compared to weighted pseudo-populations but, compared to IPTW:
  - often more challenging to implement
  - often less efficient at preserving effective sample size
  - additional challenges working with estimands

- Idea: take each patient in the treated group and pair them up with a similar patient in the untreated group
  - Comparing outcomes between treated and untreated in the matched population will give us an ATT estimate if we find a match for all exposed patients
  - We ignore unmatched patients from the unexposed group.

What About Matching Methods? (2 / 3)

- Let's return to our earlier example:

- For the treated group, we could:
  - match each of the two red patients with a duplicate of the one red patient in the untreated group, and
  - match the blue patient with one of the two blue patients in the untreated group

- Note that:
  - No way to conduct 1:1 matching here and no unique solution
  - When computing standard errors we should account for both any weighting/duplication as well as the pairing using cluster-robust standard errors (or we could bootstrap)

What About Matching Methods? (3 / 3)

- Exact matching is going to be infeasible in practice
- Solution: could match based on propensity scores
  - To avoid bad matches we often apply a caliper to ensure propensity scores are sufficiently similar between matches
- Many other matching approachs–e.g. see R MatchIt package documentation[5])

[5] Ho D, Imai K, King G, Stuart E. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software. 2011. doi:10.18637/jss.v042.i08.

Covariate Selection

- Selection of covariates to adjust for should be based on reviews of the literature and clinical input–variable selection should not be done using the analysis dataset
- Ideally we would include all baseline covariates which might be expected to be associated with the outcome
- In practice when conducting ECAs we often have extremely limited sample sizes which limits the potential complexity of our propensity score model and outcome regression specifications

What Could Go Wrong?

- ECAs are usually used when it is impractical or infeasible to include a concurrent control arm. This usually means:
  - Limited sample sizes in single arm trial data and immature survival data
  - Limited sample sizes in external data sources–e.g. few patients may be screened for rare cancer mutations in real-world settings
  - Difficult to operationalize eligibility criteria and define time zero for survival analysis in RWD
  - Non-overlap in patient populations (positivity violation)

- I will focus on examples from oncology settings

Difficulties Achieving Balance: Case Study

- Relapsed/recurrent multiple myeloma (rrMM) is treated with various cocktails of drugs in successive lines of therapy
- Head-to-head trials are unavailable for all combination therapies
- He et al.[6] used individual patient data (IPD) for 3 clinical trials incorporating arms for daratumumab + pomalidomide + dexamethasone (D-Pd), bortezomib + dexamethasone (Vd), and daratumumab + Vd (D-Vd) to compare D-Pd vs. Vd and D-Vd.
- Due to substantial differences in baseline characteristics across the trials and very limited sample sizes after harmonizing eligibility criteria, matching and weighting methods yielded poor balance and limited effective sample sizes

[6] He J, Berringer H, Heeg B, Ruan H, Kampfenkel T, Dwarakanathan HR, Johnston S, Mendes J, Lam A, Bathija S, Mackay EK. Indirect Treatment Comparison of Daratumumab, Pomalidomide, and Dexamethasone Versus Standard of Care in Patients with Difficult-to-Treat Relapsed/Refractory Multiple Myeloma. Advances in Therapy. 2022.

Difficulties Achieving Balance: Case Study

- Our solution was to use a cardinality matching method which uses computational methods to find the largest equal-sized subsets of treatment and control patients for which minimum balance criteria were met
- Cardinality matching yielded much better balance and precision relative to propensity score matching and stabilized IPTW.

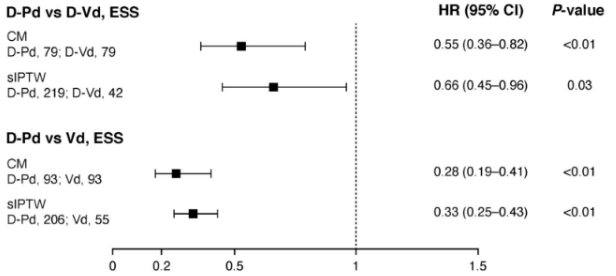
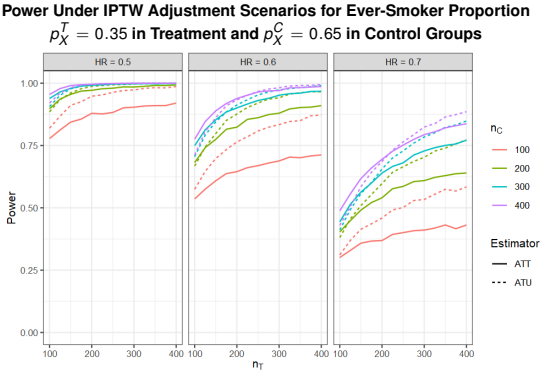


Fig. 5 Progression-free survival. Values less than 1 favor D-Pd. *CI* confidence interval, *CM* cardinality matching, *D-Pd* daratumumab, pomalidomide, and dexamethasone, *D-Vd* daratumumab, bortezomib, and dexamethasone, *ESS* effective sample size, *HR* hazard ratio, *sIPTW* stabilized inverse probability of treatment weighting, *Vd* bortezomib and dexamethasone

When Sample Sizes are Limited, Every Decision Matters

- Is misspecification more of a concern for the outcome regression model or propensity score model?
- Even choice of estimand has implications for precision / power[7]:
  - Figure (right) shows how power varies for an IPTW-adjusted Cox proportional hazards survival model with various hazard ratios (HR) and treatment and control group sample sizes with adjustment for a single unbalance binary covariate



[7] Mackay E, Springford A. P32 Power Implications of Estimator Choice in Synthetic Control Arm Analyses: Results from a Simulation Comparing Average Treatment Effects on the Treated and Untreated Under Propensity Score Weighting [Abstract]. Value in Health. 2022.

Recap So Far

- ECAs allow for indirect treatment comparisons in the absence of RCT data
- Care needed in identification of external data source (e.g. RWD or historical control arm)
  - Similar patient population (e.g. similar eligibility criteria for historical control, and similar baseline characteristics)
- Appropriate application of methods for adjusting for observed differences in potential confounders
  - Could control for confounders using parametric model (outcome regression)
  - Could construct ECA with similar baseline characteristics using matching or inverse probability of treatment weighting (IPTW) methods

What about residual heterogeneity across populations / data sources?

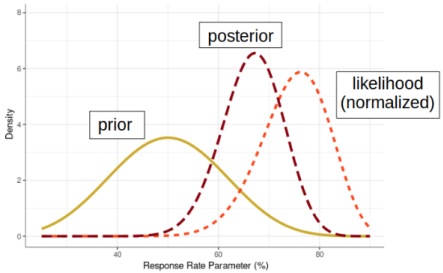
- Can we down-weight the information contribution of the external data when there is evidence of meaningful differences in (post-adjustment) outcomes between data sources? $\rightarrow$ “dynamic borrowing’
  - E.g. borrow less information from an external control to supplement a small concurrent control arm in an RCT when outcomes differ meaningfully
  - Or when estimating, say, the disease control rate for the control arm for a fully-externally controlled trial, heterogeneity in outcomes across external data sources should be reflected in the precision of estimate of the disease control rate
- A (very) brief introduction to Bayesian borrowing approaches to follow

Why Use a Bayesian Approach for Rare Diseases?

- Provides a principled framework for incorporating external information:
  - Start with our prior (which can be informed by external data)
  - Update our beliefs after observing new data
- Conducive to sequential 'Bayesian updating'
- Posterior inference allows us to quantify the amount of evidence in favour of a conclusion and allows for more nuanced decision rules
- See Mackay & Springford (2023) for additional discussion[8]

$$p(\theta \mid \text{Data}) = \frac{p(\text{Data} \mid \theta) \cdot p(\theta)}{p(\text{Data})}$$

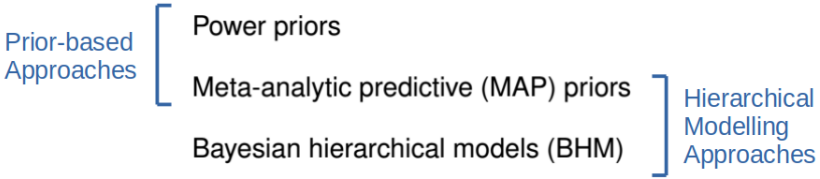
posterior; likelihood (contribution of new data); prior; normalizing constant



[8] Mackay EK, Springford A. Evaluating treatments in rare indications warrants a Bayesian approach. Frontiers in Pharmacology. 2023.

Methods Roadmap

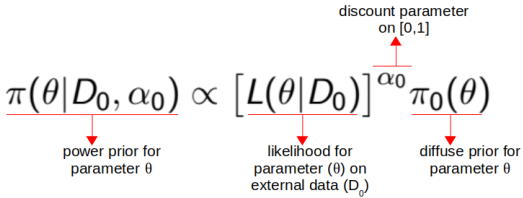
- At least 3 broad approaches to Bayesian borrowing:



- Typically external data is only available for standard of care / control treatments
  - Could use external data to form a prior for a parameter in a concurrent control arm or as a stand-in for a non-existent control arm
- Without loss of generality, examples will focus on binary response endpoints (will use external data to inform estimates of the control treatment response rate)

Power Priors

- Power priors[9,10]
  - Down-weight the external data by means of a discount parameter, $\alpha_0 \in [0, 1]$



  - As $\alpha_0 \to 0$ we ignore the external data (no pooling)
  - As $\alpha_0 \to 1$ we give it full weight (full pooling)
  - How to choose α_0?: 'tipping point' approach[11], target effective sample size for borrowing[12], dynamic borrowing based on consistency between data sources[5,6]

[9] Ibrahim JG, Chen MH. Power prior distributions for regression models. Stat. Sci. 2000.

[10] Ibrahim JG, Chen MH, Gwon Y, Chen F. The power prior: Theory and applications. Statistics Med. 2015.

[11] Best N, ... Assessing efficacy in important subgroups in confirmatory trials: An example using Bayesian dynamic borrowing. Pharm. Stat. 2021.

[12] Richeldi L, ..., Maher TM. Trial of a preferential phosphodiesterase 4B inhibitor for idiopathic pulmonary fibrosis. NEJM. 2022.

- Demonstration of dynamic borrowing for binary response endpoint using a beta-binomial model with normalized power prior[13]
  - Observed response rate is $p = y/n$ in current data and $p_0 = y_0/n_0$ in historical data with $n = n_0 = 100$ patients
  - We use a Beta$(1, 1)$ prior for θ and use either a Beta$(1, 1)$ prior or fixed value for α_0 and report posterior means and 95% CrIs for the response rate for two scenarios: $(p, p_0) = (0.4, 0.2)$ and $p = p_0 = 0.4$

Scenario	**Prior for α_0**	**Response Rate (θ)**
Inconsistent ($p = 0.4, p_0 = 0.2$)	Beta$(1, 1)$	0.368 [0.275, 0.468]
Inconsistent ($p = 0.4, p_0 = 0.2$)	Full pooling ($\alpha_0 = 1$)	0.301 [0.241, 0.367]
Consistent ($p = p_0 = 0.4$)	Beta$(1, 1)$	0.401 [0.325, 0.480]
Consistent ($p = p_0 = 0.4$)	No pooling ($\alpha_0 = 0$)	0.401 [0.309, 0.498]

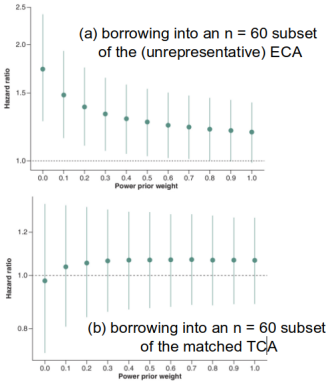
- The normalized power prior is able to improve precision and also partially mitigates bias when data sources are incompatible
- However, we also show that the amount of borrowing can be sensitive to choice of prior for α_0

[13] Mackay EK, Springford A. Impact of Hyperprior Choice for Bayesian Dynamic Borrowing via a Normalized Power Prior. JSM Proceedings. Alexandria, VA: American Statistical Association. 2023 Oct. https://doi.org/10.5281/zenodo.10001953

Static Borrowing using a Power Prior: An Example

- Example from Struebing et al. (2024)[14]
- Goal was to replicate the results of an RCT by
  - Constructing an external control arm (ECA) from real-world data
  - And augmenting limited sample sizes in the ECA by borrowing from a historical control arm
- ECA analysis failed to replicate RCT results for chemotherapy with or without cetuximab in first-line (1L) non-small cell lung cancer (NSCLC)
- Bayesian borrowing was conducted using a static power prior with a sliding scale of fixed borrowing weights (tipping point approach)

Posterior medians and 95% credible intervals for the hazard ratio for different Bayesian borrowing weights

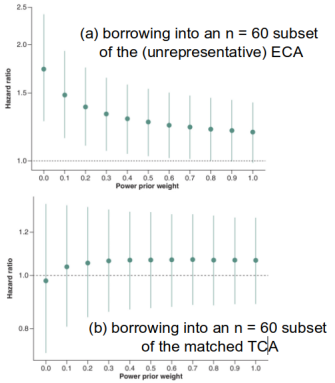


[14] Struebing A, McKibbon C, Ruan H, Mackay E, Dennis N, Velummailum R, He P, Tanaka Y, Xiong Y, Springford A, Rosenlund M. Augmenting external control arms using Bayesian borrowing: a case study in first-line non-small cell lung cancer. JCER. 2024.

Static Borrowing using a Power Prior: An Example

- Borrowing from the historical control was able to
  - Mitigate bias from the real-world ECA
  - Improve precision (especially when supplementing a small ECA of n = 60 patients)
- Also demonstrated approach for an 'ideal scenario' where the trial control arm (TCA) in the RCT was used as a 'hypothetical ECA'
- Takeaway: worth considering whether Bayesian borrowing approaches can be used to provide a structured means for incorporating additional external data sources (including aggregate data) beyond our ECA alone

Posterior medians and 95% credible intervals for the hazard ratio for different Bayesian borrowing weights



[14] Struebing A, McKibbon C, Ruan H, Mackay E, Dennis N, Velummailum R, He P, Tanaka Y, Xiong Y, Springford A, Rosenlund M. Augmenting external control arms using Bayesian borrowing: a case study in first-line non-small cell lung cancer. JCER. 2024.

Conclusion

- External control arm (ECA) methods provide a potentially powerful tool in settings where it is difficult or infeasible to conduct a well-powered RCT
- Nonetheless, care has to be taken in the design of ECA analyses to minimize risk of bias to the extent feasible
- ECA methods and Bayesian borrowing designs are currently very active areas of research and we are starting to see some early uptake of these methods among regulators and payers

References

- Best N, ... Assessing efficacy in important subgroups in confirmatory trials: An example using Bayesian dynamic borrowing. Pharm. Stat. 2021.
- He J, Berringer H, Heeg B, Ruan H, Kampfenkel T, Dwarakanathan HR, Johnston S, Mendes J, Lam A, Bathija S, Mackay EK. Indirect Treatment Comparison of Daratumumab, Pomalidomide, and Dexamethasone Versus Standard of Care in Patients with Difficult-to-Treat Relapsed/Refractory Multiple Myeloma. Advances in Therapy. 2022.
- Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. American journal of epidemiology. 2016.
- Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC. 2020. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.
- Ho D, Imai K, King G, Stuart E. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software. 2011. doi:10.18637/jss.v042.i08.
- Ibrahim JG, Chen MH. Power prior distributions for regression models. Stat. Sci. 2000.
- Ibrahim JG, Chen MH, Gwon Y, Chen F. The power prior: Theory and applications. Statistics Med. 2015.
- Mackay EK, Springford A. Evaluating treatments in rare indications warrants a Bayesian approach. Frontiers in Pharmacology. 2023.
- Mackay E, Springford A. P32 Power Implications of Estimator Choice in Synthetic Control Arm Analyses: Results from a Simulation Comparing Average Treatment Effects on the Treated and Untreated Under Propensity Score Weighting [Abstract]. Value in Health. 2022.
- Mackay EK, Springford A. Impact of Hyperprior Choice for Bayesian Dynamic Borrowing via a Normalized Power Prior. JSM Proceedings. Alexandria, VA: American Statistical Association. 2023 Oct. https://doi.org/10.5281/zenodo.10001953
- Richeldi L, ..., Maher TM. Trial of a preferential phosphodiesterase 4B inhibitor for idiopathic pulmonary fibrosis. NEJM. 2022.
- Struebing A, McKibbon C, Ruan H, Mackay E, Dennis N, Velummailum R, He P, Tanaka Y, Xiong Y, Springford A, Rosenlund M. Augmenting external control arms using Bayesian borrowing: a case study in first-line non-small cell lung cancer. JCER. 2024.
- US FDA. Guidance for Industry. E 10 Choice of Control Group and Related Issues in Clinical Trials. 2001. https://www.fda.gov/media/71349/download
- US FDA. Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products Guidance for Industry. Draft Guidance. 2023. https://www.fda.gov/media/164960/download
- UK NICE. NICE real-world evidence framework. 2022. https://www.nice.org.uk/corporate/ecd9/resources/nice-realworld-evidence-framework-pdf-1124020816837
- Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. Statistical methods in medical research. 2012.

Thank You!

Contact: Emma Mackay,
emma@inka.health