Boosting the performances of clinical trials from external data or related algorithms/models: conditions to respect

Yohann Foucher

Séminaire plateforme EUCLID FCRIN, Bordeaux
29 novembre 2024

Plan

Why augmenting randomized clinical trials (RCTs) ?

- RCTs represent the reference protocol for treatment evaluations.
- For large sample size, the comparability of the arms is guaranteed. **BUT** :
  - **Selection bias** : the characteristics of the studied patients are not representative of the targeted population.
  - **Near confounders** : some prognostic factors may be unbalanced because the sample-to-sample fluctuation. Examples of median of the sample sizes from systematic reviews :
    - 144 patients for RCTs in surgery. [1]
    - 75 patients for RCTs in intensive care units. [2]
  - **Etc.**

1. Robinson et al. Characteristics of Randomized Clinical Trials in Surgery From 2008 to 2020 : A Systematic Review. JAMA Network Open. 2021.
2. Anthon et al. Overall bias and sample sizes were unchanged in ICU trials over time : a meta epidemiological study. Journal of Clinical Epidemiology. 2019.

Methods for augmenting RCT with external data

- Matching with propensity scores : $Pr(RCT\ versus\ External \mid X)$.
- Estimating algorithms/models for predicting contractual outcomes : $f(Y \mid X)$.
- Simulating digital twins i.e. both baseline characteristics and outcomes : $f(Y, X)$.
- Etc.

Problem #1 : It introduces bias : selection bias, confounders, etc.

- **What are the conditions for ensuring unbiased results** from RCT augmented by models/algorithms from external data ?

Plan

The conditions for acceptability of an external control group [3]

A1. The group must have received a precisely defined treatment which must be the same in the RCT.

A2. The group must have been part of a recent study which contained the same requirements for patient eligibility.

A3. The methods of treatment evaluation must be the same.

A4. The distributions of important patient characteristics in the group should be comparable with those in the new trial.

A5. The external must have been collected in the same organization with largely the same clinical investigators.

A6. There must be no other indications leading one to expect differing results between the randomized and external controls.

3. Pocock. The combination of randomized and historical controls in clinical trials. Journal of Chronic Diseases. 1976.

The conditions for acceptability of an external control group [4]

C1. The respect for positivity between the external and RCT controls.
- It includes A2, A4 and A5.

C2. The absence of unmeasured confounders, i.e. determinants of both the outcome and the probability of inclusion in the RCT compared to the external group.
- It includes A6.

C3. The absence of a direct effect on the outcome of being included in the RCT.
- It includes A1 and A3.

4. Dang et al. A Cross-Validated Targeted Maximum Likelihood Estimator for Data-Adaptive Experiment Selection Applied to the Augmentation of RCT Control Arms with External Data. arXiv :2210.05802v1

How to respect the condition C1 (positivity) ?

1. Use external data but from a recent period.
2. Use external data from the same centers, at least for a large part.
3. Check the inclusion criteria in both the RCT and external study : the targeted population of the RCT must be embedded.
4. Compare the characteristics of two samples of controls to validate the positivity, [5] and restrict the inclusion criteria until no issue is identified.

5. Danelian G, Foucher Y, Léger M, Le Borgne F. and Chatton A. Identification of in-sample positivity violations using regression trees : The PoRT algorithm. Journal of Causal Inference (2023).

How to respect the condition C2 (absence of unmeasured confounders) ?

1. Draw a causal diagram with the inclusion in the RCT versus external study as an element.
2. Propose a complete list of the prognostic factors of the studied outcomes.
3. Propose a complete list of the determinants of the inclusions in the RCT versus external study.
4. Make sure that the variables in the two lists #2 and #3 will be available in both the RCT and the external study with the same methods of collection.

How to respect the condition C3 (no direct effect of being in the RCT) ?

- One can limit/prevent such a direct effect by :
  1. Check the studied treatments are identical in the two studies (doses, administration, etc.).
  2. Use the same definition of the collected outcomes.
  3. Ensure comparable monitoring in the two studies.
- Control patients in the RCT are required for a data-driven check of C3.
- An RCT based on a control arm completely-based with external data or digital twins does not allow for validating C3.

Problem #2 : these checks do not convince on the respect of C3

- **How to validate C3 by using the RCT data ?**

Plan

Consider the following usual RCT for superiority with k controls for 1 patient in the experimental arm

- A binary principal outcome Y with π_0 and π_1 the two expected proportions in the control and experimental arms.
- N_{1s} and N_{0s} the required sample sizes for a bilateral test with targeted type-I and type-II errors, α and β respectively.
- $N_{0s} = kN_{1s}$ and $N_{1s} = (\pi_1 - \pi_0)^2(\pi_0(1-\pi_0)/k + \pi_1(1-\pi_1))/(z_{\alpha/2} + z_\beta)^2$

Example

- For $k = 1$, $\alpha = 0.05$, $\beta = 0.20$, $\pi_0 = 0.50$ and $\pi_1 = 0.40$,
- the sample sizes are $N_{0s} = N_{1s} = 385$.

Consider the following method to augment the RCT with external data (one method among others)

- A model M (for instance a logistic regression) is estimated from the external data to predict the outcome $\hat{\pi}_{0i}$ given the characteristics X_i of a patient i with the control treatment.
- There is no direct effect of being in the RCT (condition C3) if M is well-calibrated for the outcome prediction at the inclusion in the control group of the RCT.
- A calibration measure can be O/E :
  - O is the number of observed events : $\sum_i Y_i$ among the control patients.
  - E is the number of expected events : $\sum_i \hat{\pi}_{0i}$ among the control patients.
- The value $O/E = 1$ indicates a perfect calibration.

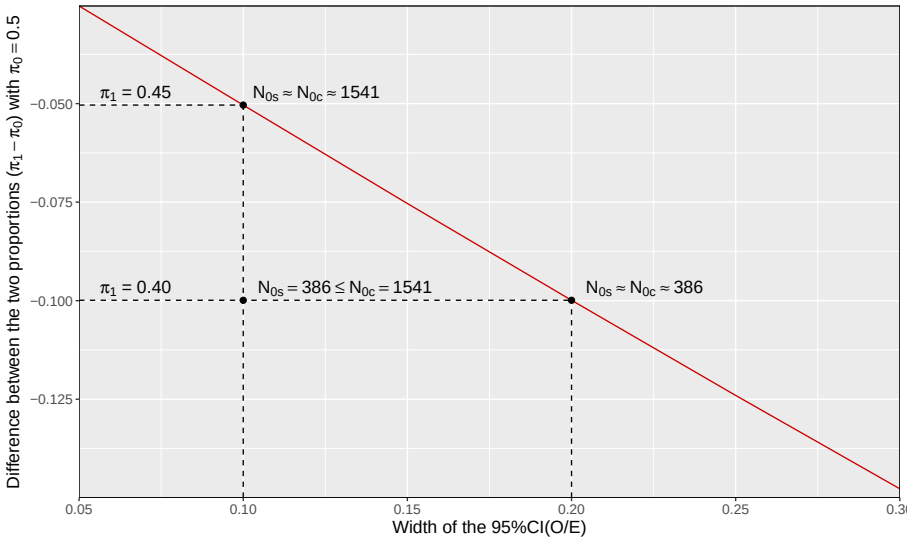
Minimal sample size N_{0c} to evaluate the calibration of the model M

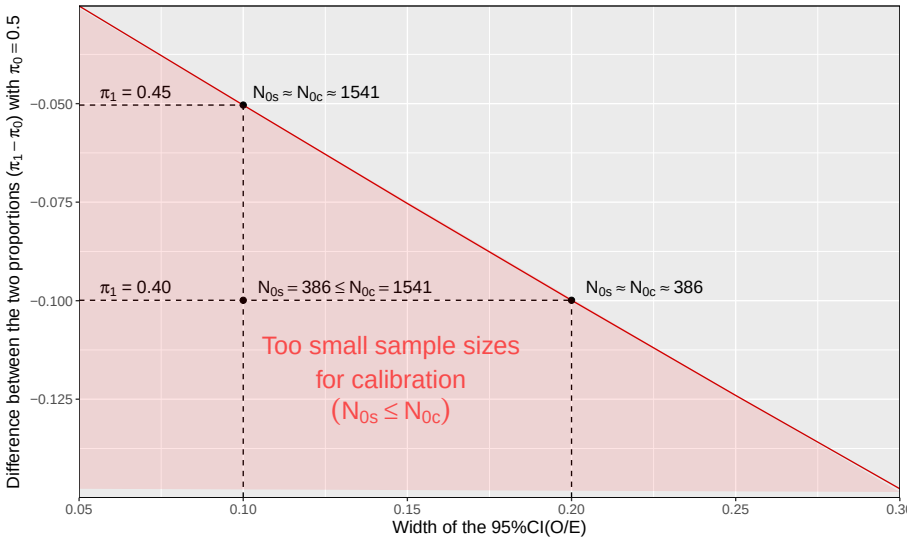
- By using the delta method :[6] $SE(\log(O/E) \approx \sqrt{(1-\pi_0)/(\pi_0 N_{0c})}$.
- Depending on w, the targeted width of confidence interval of the O/E, one can compute $SE_w(\log(O/E)$.
- One can deduce that $N_{0c} = (1-\pi_0)/(\pi_0 SE_w(\log(O/E)^2)$

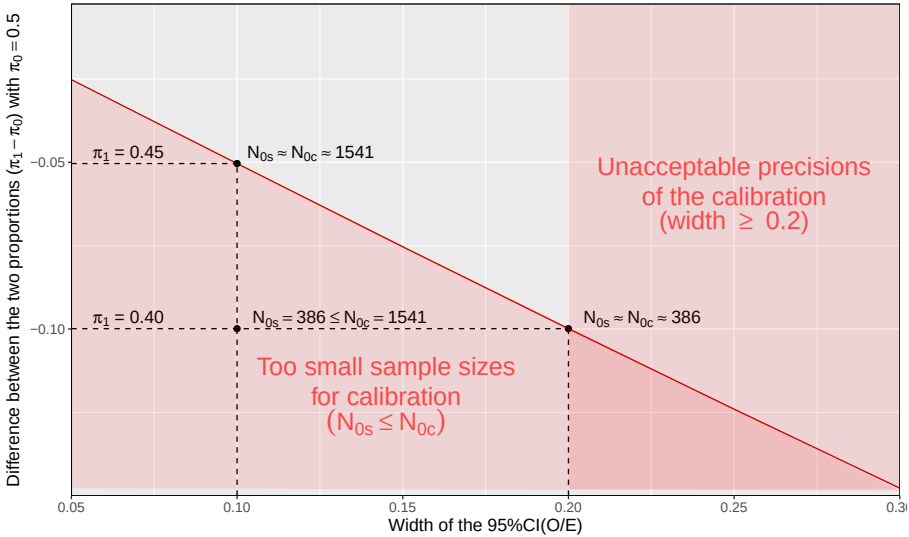
Example (continued, with $k = 1$, $\alpha = 0.05$, $\beta = 0.20$, $\pi_0 = 0.50$ and $\pi_1 = 0.40$)

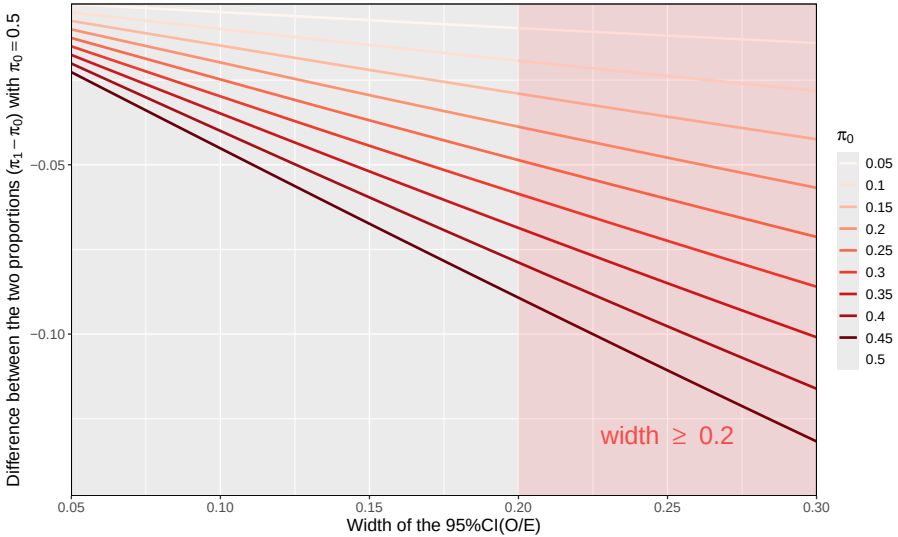
- For $\alpha = 0.05$ and $w = 0.10$ (i.e., $SE_w(\log(O/E) \approx 0.25$) then $N_{0c} = 1541$.
  - $N_{0c} > 385$: we cannot check the respect of C3.
- For $\alpha = 0.05$ and $w = 0.20$ (i.e., $SE_w(\log(O/E) \approx 0.51$,) then $N_{0c} = 386$.
  - $N_{0c} \approx 385$: we can check the respect of C3.

6. Debray et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res. 2019 ;28 :2768-2786.


Difference between the two proportions ($\pi_1 - \pi_0$) with $\pi_0 = 0.5$
−0.050
−0.075
−0.100
−0.125
$\pi_1 = 0.45$
$N_{0s} \approx N_{0c} \approx 1541$
$\pi_1 = 0.40$
$N_{0s} = 386 \leq N_{0c} = 1541$
$N_{0s} \approx N_{0c} \approx 386$
0.05
0.10
0.15
0.20
0.25
Width of the 95%CI(O/E)


Difference between the two proportions ($\pi_1 - \pi_0$) with $\pi_0 = 0.5$
−0.050
−0.075
−0.100
−0.125
$\pi_1 = 0.45$
$N_{0s} \approx N_{0c} \approx 1541$
$\pi_1 = 0.40$
$N_{0s} = 386 \leq N_{0c} = 1541$
$N_{0s} \approx N_{0c} \approx 386$
Too small sample sizes
for calibration
($N_{0s} \leq N_{0c}$)
0.05
0.10
0.15
0.20
0.25
Width of the 95%CI(O/E)


Difference between the two proportions ($\pi_1 - \pi_0$) with $\pi_0 = 0.5$
−0.05
−0.10
$\pi_1 = 0.45$
$\pi_1 = 0.40$
$N_{0s} \approx N_{0c} \approx 1541$
$N_{0s} = 386 \leq N_{0c} = 1541$
$N_{0s} \approx N_{0c} \approx 386$
Unacceptable precisions of the calibration (width $\geq$ 0.2)
Too small sample sizes for calibration ($N_{0s} \leq N_{0c}$)
0.05
0.10
0.15
0.20
0.25
Width of the 95%CI(O/E)


Difference between the two proportions ($\pi_1 - \pi_0$) with $\pi_0 = 0.5$
−0.05
−0.10
0.05
0.10
0.15
0.20
0.25
0.30
Width of the 95%CI(O/E)
width ≥ 0.2
π_0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Plan

Summary of the key points

- The augmention of RCT with models/algorithms for counterfactual predictions or even digital twins is possible under conditions C1 to C3.
- The use low-quality external data, such as retrospective registries or medico-administrative databases, should be used with caution
- The same multi-centric consortium should be involved in the the collection of external and RCT data.
- The RCTs with important sample sizes appear more adapted.
- The full external/virtual control arm do not allow us to validate the condition C3.

Number of perspectives

- The validation of the condition C3 should be performed before the end of the inclusions in the RCT to avoid non-conclusive results. It necessitates :
  - To increase the sample size of individuals prospectively included in the control arm.
  - To propose an adaptative design to achieve a regular RCT if the violation of C3.
- $k > 1$ including a mix of real and virtual patients will increase the performances of the approach.
- The use of sequential analyses to re-estimate the required sample sizes for calibration and superiority evaluations.
- Etc.