Approaches to selecting 'time zero' in external control arms with multiple potential entry points: a simulation study of eight approaches

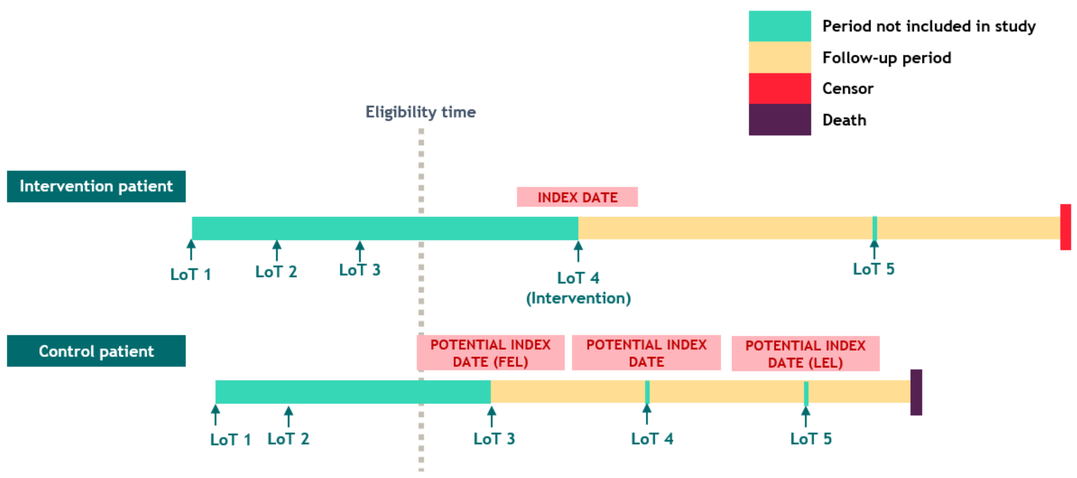
Hatswell AJ[1,2*], Deighton K[1], Thornton Snider J[3], Brookhart MA[4], Faghmous I[3], Patel AR[3]

[1]Delta Hat, Nottingham, UK; [2]Department of Statistical Science, University College London, UK; [3]Kite Pharma Inc, Santa Monica, CA, USA ; [4]NoviSci, Durham, NC, USA

Background

- When including data from an external control arm to estimate comparative effectiveness, there is a choice of when to set 'time zero', the point at which a patient would be eligible/enrolled in a contemporary study and from which outcomes are measured.
- No comprehensive list of methods, criteria for choosing, nor guidance on appropriate methods exist on selecting the most appropriate time zero.
- Figure 1 illustrates an intervention patients can enter at Line of Therapy 3+ (in this example they entered at LoT4), with control patients tracked through lines 3, 4 and 5 – all of which would have been eligible entry points for the study of the intervention.

Figure 1: Stylized example of multiple eligible entry points



- The issue is particularly prevalent in oncology, where remissions typically become shorter with each successive line of treatment
- The motivation for this simulation was the ZUMA-5 study (Ghione et al., 2022), where patients were treated at a late line, when compared with real world data (with complete patient histories) using a traditional 'first line in' approach there was a large mismatch between the studies
- The aim was to investigate the different approaches available, and understand relative merits of each approach – whilst ensuring a biased method was not selected for analysis of ZUMA-5.

Methods

The approaches identified and/or created for use in the analysis are presented below, and numbered in brackets:

- **First Line In (1)** - take the first line after inclusion criteria are met
  - Leads to an overrepresentation of earlier lines in real world data, relative to general prevalence, and particularly compared to trials
  - Example: Avelumab in Merkel cell carcinoma
- **Last Line In (2)** – take the last line available for patients in the data
  - We had a suspicion this would be biased – it is based on knowledge of the future (i.e., the line ends in either censoring or death)
  - Example: Blinatumomab in ALL (Rambaldi 2020)
- **Random line (3)**
  - Suggested by Hernán & Robins (2021), though we did not identify any applications of the method
- **Use all lines**
  - Robust variance estimation is required to account for correlation at the patient level and can use all available data, rather than sampling in some way
  - This could be done censoring OS at the next line (4), or without (5)
- **Aim to match the groups**
  - We could aim to pick the lines that balance the overall distribution between groups. We implemented this using mean absolute error (6), and mean squared error (7)
  - We can use propensity score matching with all lines available, to match each control patient to an intervention patient (with line included), and then remove them from the sample – so they are matched only once (8)

The simulation setup assumed patients deteriorate between lines, and receive the intervention at a late line – as clinical trials are typically conducted after licensed options are exhausted. With control patients treated (on average) at an earlier line, this created a bias against the intervention in a naïve comparison.

50,000 patient lifetimes were simulated for each intervention and control dataset to give a 'true' results, with 1000 control and 750 intervention patients then sampled and analysed using each of the 8 methods to derive a control dataset, which was then balanced using propensity scoring. This process was repeated 5000 times for each scenario, with sensitivity analyses adjusting key simulation parameters.

Results

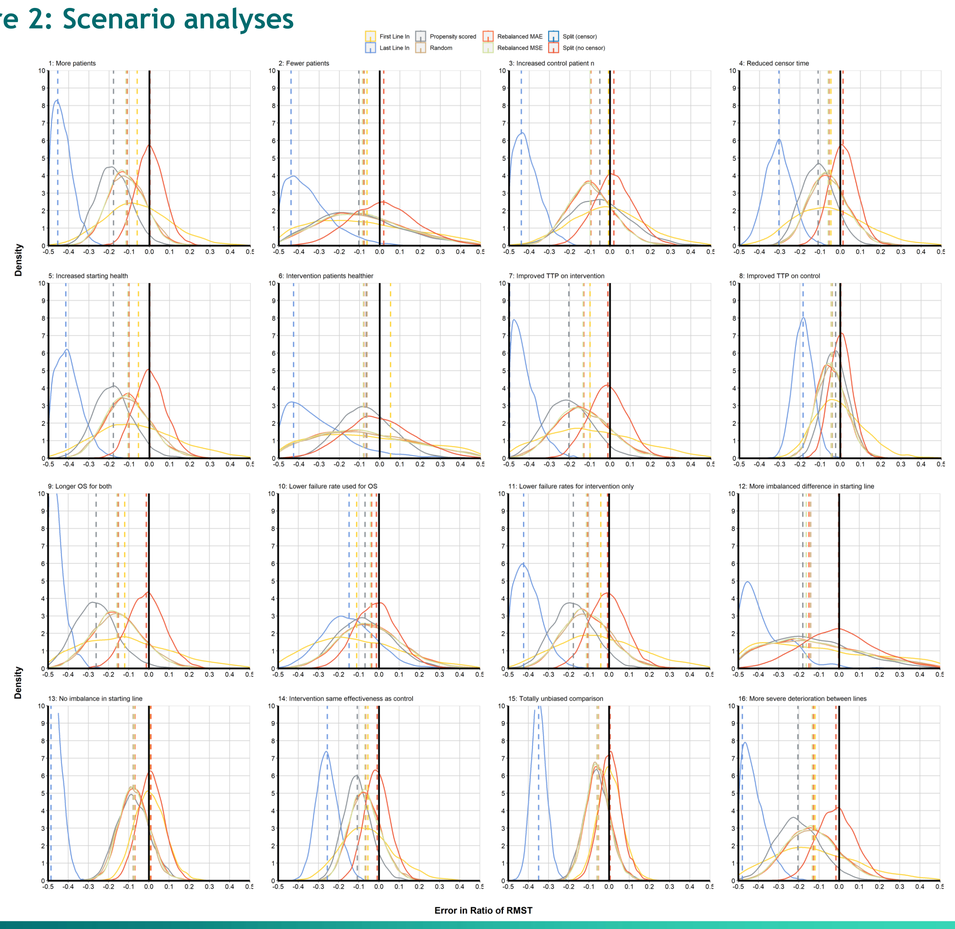
- Of the eight methods, five (random line [3], all lines [5], matched based on MAE [6], MSE [7], or propensity scores[8]) showed good performance in accounting for differences between the line at which patients were included. This can be seen in Table 1 through the bias in the Cox HR to the true values, as well as the high coverage probabilities of these methods.
- All lines (with censoring) cannot be used for survival outcomes as it performed poorly with extreme bias in all cases. Last Line In cannot be recommended, as line with the findings of Suissa (2021), we found it to be biased by deflating the control arm outcomes due to the inherent bias in the approach.
- First line In [1] was seen to be statistically inefficient (though not biased) in some scenarios, by leading to a poor overlap with the intervention study – this was particularly apparent in sensitivity analyses where patients were classed as treatment naïve vs experienced.

Table 1: Base case results: Cox PH model

Method	CoxHR Mean	Cox ME	Cox Bias	Cox Bias MCSE	CoxCov. prob	Cox %<hazard	Cox% stat concord.
PFS							
True result	0.533	0	0	0	100	0	100
1 First Line In	0.555	0.022	0.022	0.003	94	46.8	93.3
2 Last Line In	0.845	0.312	0.312	0.003	0.7	0	38.2
3 Random	0.597	0.064	0.064	0.002	85.3	16	99.7
4 All lines (censored)	0.543	0.01	0.01	0.001	95.7	43.3	100
5 All lines	0.543	0.01	0.01	0.001	95.7	43.3	100
6 Matched (MAE)	0.596	0.063	0.063	0.002	85	16	99.8
7 Matched (MSE)	0.600	0.067	0.067	0.002	83.4	14.6	99.7
8 Propensity scored	0.62	0.087	0.087	0.002	70.5	7.9	99.7
OS							
True result	0.628	0	0	0	100	0	100
1 First Line In	0.689	0.061	0.061	0.005	93.8	37.4	52
2 Last Line In	0.489	-0.139	-0.139	0.002	37.2	99.1	100
3 Random	0.616	-0.012	-0.012	0.002	95.5	59.0	96.8
4 All lines (censored)	1.083	0.455	0.455	0.004	0.6	0	0.6
5 All lines	0.666	0.037	0.037	0.002	93.7	30.3	97
6 Matched (MAE)	0.607	-0.021	-0.021	0.002	94.7	65.1	98.6
7 Matched (MSE)	0.608	-0.021	-0.021	0.002	95.4	62.7	98.3
8 Propensity scored	0.671	0.042	0.042	0.003	90.3	32.9	90.7

- Varying simulation parameters, patient characteristics, or intervention effectiveness gave similar results. This includes scenarios 13:15 where the bias inherent in the simulation is removed piecewise.
- The scenario results show a density plot of the error in the ratio of Restricted Mean Survival Times from the true result. The ideal result is a tight distribution centred around zero.

Figure 2: Scenario analyses



Conclusions

- Multiple methods are available for selecting the most appropriate time zero from an external control arm. Based on the simulation we demonstrate that some methods perform poorly under some or all circumstances, with several viable methods remaining – which in this application perform similarly, with no clear 'best' method.
- In selecting between the viable methods, analysts should consider the context of their analysis, and justify the approach selected – potentially even testing more than one approach.

References

Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183(8):758–64.
Suissa S. Single-arm trials with historical controls: study designs to avoid time-related biases. Epidemiol. 2021;32(1): 94–100. doi:10.1097/EDE.0000000000001267
Ghione, P., Palomba, M.L., Patel, A.R., Bobillo, S., Deighton, K., Jacobson, C.A., Nahas, M., Hatswell, A.J., Jung, A.S., Kanters, S., Snider, J.T., Neelapu, S.S., Ribeiro, M.T., Brookhart, M.A., Ghesquieres, H., Radford, J., Gribben, J.G., 2022. Comparative effectiveness of ZUMA-5 (axi-cel) vs SCHOLAR-5 external control in relapsed/refractory follicular lymphoma. Blood 140, 851–860. https://doi.org/10.1182/blood.2021014375
Rambaldi A, Ribera J-M, Kantarjian HM, et al. Blinatumomab compared with standard of care for the treatment of adult patients with relapsed/refractory Philadelphia chromosome–positive B-precursor acute lymphoblastic leukemia. Cancer. 2020;126:304–10.
Gokbuget N, Kelsh M, Chia V, et al. Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. Blood Cancer J. 2016;6:e473.

Full paper 'Approaches to Selecting "Time Zero" in External Control Arms with Multiple Potential Entry Points' published in Medical Decision Making

ahatswell@deltahat.com