Abstract ID #130198
Acceptance Code: CO90

Augmenting Synthetic Control Arms Using Bayesian Borrowing: A Case Study in First-line Non-small Cell Lung Cancer

Alessandria Struebing[1], Chelsea McKibbon[2], Haoyao Ruan[2], Emma Mackay[2], Natalie Dennis[1], Russanthy Velummailum[2], Philip He[3], Yoko Tanaka[3], Yan Xiong[3], Aaron Springford[2], Mats Rosenlund[1,4]

[1]Daiichi Sankyo Europe, Munich, Germany; [2]Cytel Inc., Toronto, ON, Canada; [3]Daiichi Sankyo, Inc., Basking Ridge, NJ, US; [4]Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden

Background

- Synthetic control arms (SCA) are an alternative method to providing comparative effectiveness estimates when control arm data from randomized controlled trials (RCT) are not available. SCAs are often constructed from real-world data (RWD) or historical control (HC) arms and incorporate statistical methods such as propensity score matching or weighting, in the absence of randomization, to limit the influence of confounding factors.[1]
- Bayesian borrowing (BB) methods use data from external sources to bolster limited sample sizes and can increase the precision of estimates in SCA analyses when implemented carefully.
- Existing guidance and precedent on the use of BB methods have been limited to the regulatory space; additional research is necessary to understand cases for future use in providing decision-grade evidence for health technology assessment (HTA).[2-5]

Objective

To assess the performance of an augmented SCA constructed from RWD by applying BB methods to historical clinical trial data in first-line (1L) advanced or metastatic non-small cell lung cancer (a/m NSCLC) to improve precision of overall survival (OS) estimates.

Methods

SCAs

- ConcertAI Patient360™ (RWD cohort) was used to construct an SCA for the without-bevacizumab cohort for an RCT in 1L advanced NSCLC comparing chemotherapy with (trial treatment arm [TTA]) or without cetuximab (trial control arm [TCA])[6] (individual patient data [IPD] provided by the National Cancer Institute via Project Data Sphere).
- Trial eligibility criteria were applied to the RWD, and patient characteristics (e.g., age, sex, race, Eastern Cooperative Oncology Group [ECOG] Performance Status [PS]) were matched between the TTA and RWD cohort using a cardinality matching (CM) approach to construct a matched subset of patients with a similar distribution of baseline characteristics (the "matched TTA" and "SCA"). CM was also used to construct a matched subset of the TCA (the "matched TCA").
- For the SCA, OS was defined as the time from 1L treatment initiation to death. Patients were censored at date of last follow-up or last activity on record (i.e., visit, treatment).
- Hazard ratios (HR) for OS and 95% confidence intervals (CI) were estimated using Cox proportional hazards (PH) models.
- Performance of the SCA was evaluated based on whether the SCA was able to replicate the intention-to-treat OS effect estimate for the TTA vs TCA in the matched subpopulation.
- SCA analysis was conducted using the R statistical programming language (version 4.2).

Augmentation via BB

- The SCA was augmented through BB from an HC[7] (pseudo-IPD obtained via digitized[8] published Kaplan-Meier curves [KM]).
- BB was conducted using a static power prior[9] under a Weibull PH parameterization with borrowing weights from 0.0 to 1.0, in increments of 0.1.
- Four BB scenarios were considered: 1) borrowing into the SCA; 2) borrowing in the matched TCA (mimics a hypothetic "ideal SCA" scenario); 3a) borrowing into a small-sample subset of the SCA; and 3b) borrowing in a subset of the matched TCA. Subsets in scenario 3a and 3b were selected based on a random sample without replacement.
- HR estimates for OS were summarized using posterior medians and 95% credible intervals (CrI), which captures the central 95% of the posterior probability for the HR.
- BB analysis was conducted using the R statistical programming language (version 4.2) and Stan probabilistic programming language (version 2.26.1) via the *rstan* R package.

Results

Patient characteristics

- Before matching, 365 and 374 patients were available in the TTA and TCA, respectively, and 181 patients from the RWD cohort met all inclusion/exclusion criteria. Post-CM, 178 patients were included in each cohort, and 531 patients were available in the HC arm.
- Post-CM, absolute standardized mean difference between the TCA/TTA and SCA was less than 0.1 for all variables considered.
- Patient characteristics were similar after matching. (**Table 1**)
- The HC had different baseline characteristics in general, notably in age, sex, and smoking history, and included patients with stage IIIB disease and ECOG PS score of 2. (**Table 1**)

Table 1. Baseline characteristics for matched cohorts

	Matched				Matched subset	
	TTA (n=178)	TCA (n=178)	SCA (n=178)	HC (n=531)	TCA (n=60)	SCA (n=60)
Median age (range)	66.5 (29.5–83)	64.6 (34.6–82.6)	66.0 (24–86)	60.8 (24.9–84.7)	65.0 (34.6–80)	63.5 (40–84)
Sex - n (%)						
Male	112 (62.9)	104 (58.4)	107 (60.1)	397 (74.8)	34 (56.7)	37 (61.7)
Female	66 (37.1)	74 (41.6)	71 (39.9)	134 (25.2)	26 (43.3)	23 (38.3)
ECOG - n (%)						
0	59 (33.1)	54 (30.3)	51 (28.7)	113 (21.3)	20 (33.3)	25 (41.7)
1	119 (66.9)	124 (69.7)	127 (71.3)	416 (78.3)	40 (66.7)	35 (58.3)
2	0	0	0	2 (0.4)*	0	0
Stage - n (%)						
IIIB	0	0	0	110 (20.7)*	0	0
IV	178 (100)	178 (100)	178 (100)	421 (79.3)	60 (100)	60 (100)
Histology - n (%)						
Squamous	97 (54.5)	93 (52.2)	89 (50.0)	221 (41.6)	26 (43.3)	27 (45.0)
Non-squamous	81 (45.5)	85 (47.8)	89 (50.0)	310 (58.4)	34 (56.7)	33 55.0)
Smoking history - n (%)						
Former	82 (46.1)	80 (44.9)	84 (47.2)	148 (27.9)	25 (41.7)	28 (46.7)
Current	70 (39.3)	78 (43.8)	64 (36.0)	234 (44.1)	27 (45.0)	20 (33.3)
Never	26 (14.6)	20 (11.2)	30 (16.9)	144 (27.1)	8 (13.3)	12 (20.0)

*As pseudo-IPD was utilized, we were not able to implement key eligibility criteria such as the removal of stage IIIB and ECOG=2 patients
Abbreviations: ECOG, Eastern Cooperative Oncology Group; HC, historical control; SCA, synthetic control arm; TCA, trial control arm; TTA, trial treatment arm

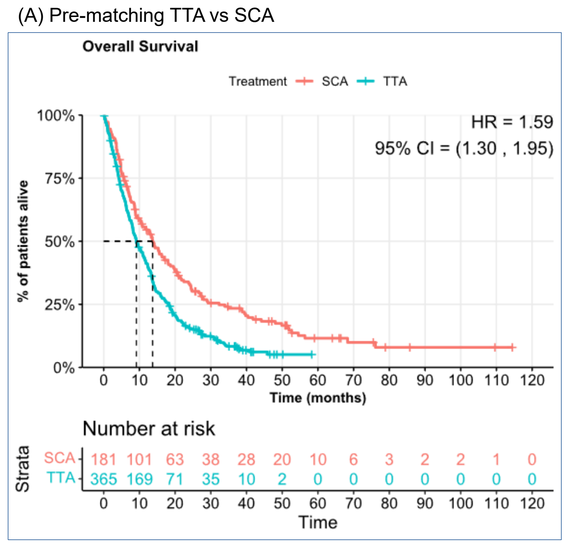
Results (cont.)

SCAs

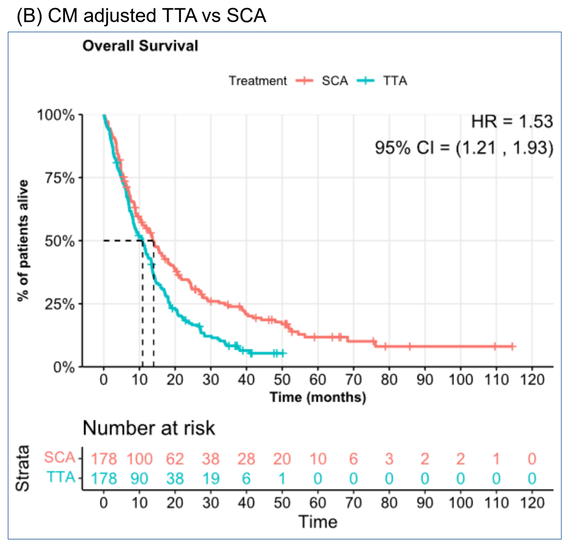
- HRs for OS before and after CM adjustment favored the SCA compared with the TCA (before matching: HR=1.59 (95% CI: 1.30, 1.95, p<0.001) and after matching: HR=1.53 (95% CI: 1.21, 1.93, p<0.001)). (**Figure 1**)
- The SCA analysis yielded much higher OS HR estimates than those obtained in the matched population of the RCT (0.91 [95% CI: 0.73,1.13]).
  - For comparison, the OS HR in the trial for the without-bevacizumab cohort was reported as 0.90 (95% CI: 0.78, 1.05, p=0.19).
- The SCA demonstrated the longest median OS, followed by the HC, the matched TTA, and the matched TCA. (**Figure 2**)
- Sensitivity analyses (eligibility period for the RWD cohort, number and type of subsequent therapies, epidermal growth factor receptor FISH status, non-random missingness in ECOG PS score) did not provide evidence of a key specific driver of the observed discrepancies.

Figure 1. KM curves of OS of TTA and SCA

(A) Pre-matching TTA vs SCA

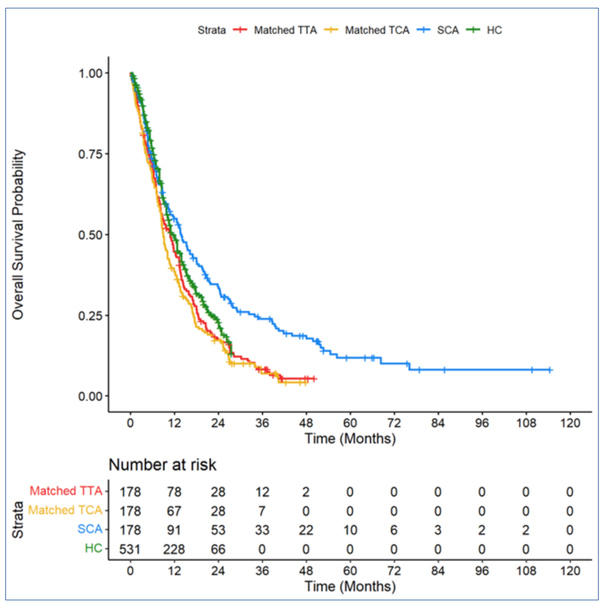


(B) CM adjusted TTA vs SCA



Abbreviations: CI, confidence interval; HR, hazard ratio; SCA, synthetic control arm; TTA, trial treatment arm

Figure 2. KM curves of OS for the matched TTA and TCA, SCA, and HC



Abbreviations: HC, historical control; SCA, synthetic control arm; TCA, trial control arm; TTA, trial treatment arm

BB

- In scenario 1, as borrowing weights increased from 0 to 1, the HR decreased from 1.67 (95% CrI: 1.33, 2.10)* to 1.30 (95% CrI: 1.08, 1.54), illustrating how the point estimate and CrI were pulled closer to those of the RCT. The HR estimates are very sensitive to the borrowing weight put on the HC. (**Figure 3A**).
- In scenario 2, the HR estimates increased from 0.91 (95% CrI: 0.73, 1.13) to 1.03 (95% CrI: 0.86, 1.22)—with the point estimate only slightly above 1—and the width of the 95% CrIs decreased slightly from 0.40 to 0.37 due to the similarity of outcomes and augmentation of the TCA via BB as borrowing weights increased from 0 to 1. (**Figure 3B**)

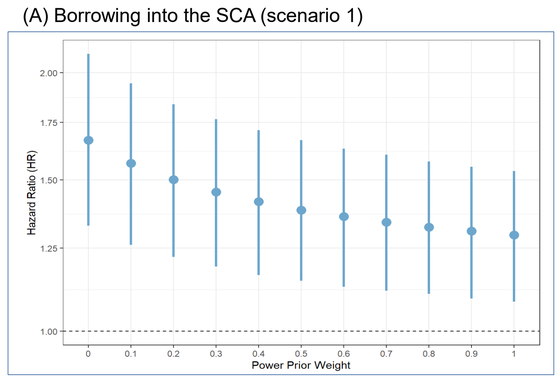
*Estimates differ slightly from the original SCA estimates since a parametric Weibull PH model is used instead of a Cox PH model.

BB (cont.)

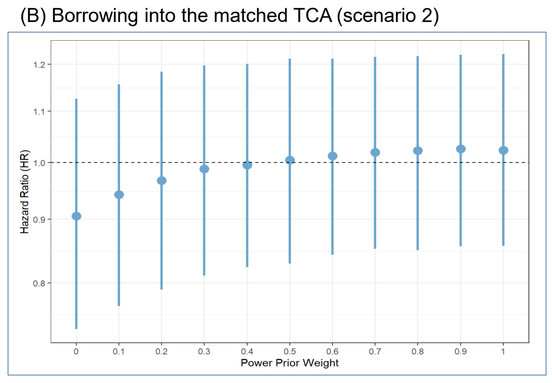
- In scenario 3a, the HR decreased from 1.73 (95% CrI: 1.27, 2.40) to 1.19 (95% CrI: 0.99, 1.42) and the 95% CrI width decreased from 1.13 to 0.43 due to augmentation of the limited sample size of the SCA as borrowing weights increased from 0 to 1. (**Figure 3C**)
- In scenario 3b, the HR increased from 0.98 (95% CrI: 0.72, 1.35) to 1.07 (95% CrI: 0.89, 1.23), but the width of the 95% CrI interval narrowed from 0.63 to 0.39 as borrowing weights increased from 0 to 1, demonstrating the potential of BB to improve precision when the SCA is well constructed and when sample sizes are small. (**Figure 3D**)

Figure 3. Posterior medians and 95% CrIs for the HR for different BB weights

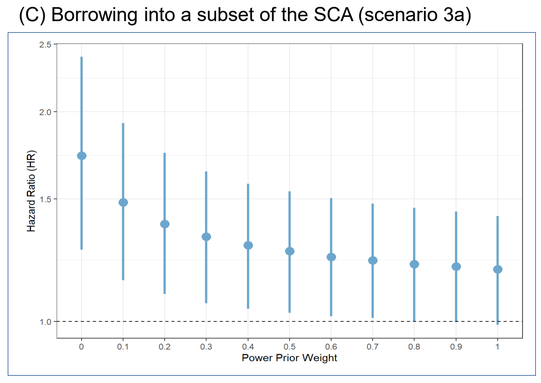
(A) Borrowing into the SCA (scenario 1)



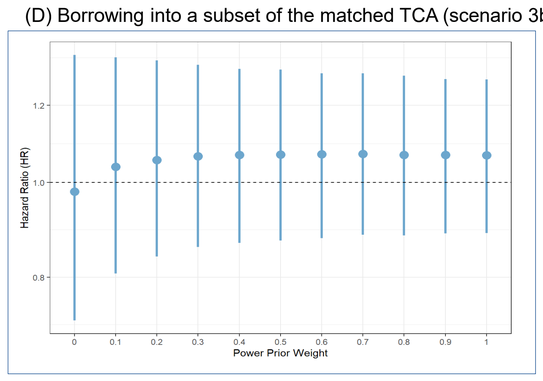
(B) Borrowing into the matched TCA (scenario 2)



(C) Borrowing into a subset of the SCA (scenario 3a)



(D) Borrowing into a subset of the matched TCA (scenario 3b)



Limitations

- In the construction of the SCA, a time restriction was not applied to the RWD cohort to ensure enough eligible patients were available for analyses; therefore, changes in standard of care for a/m NSCLC may have occurred.
- Characteristics related to disease severity such as laboratory values, biomarkers, comorbidity burden, and disease stage at diagnosis were not considered due to limited data available from historical trials.
- CM produces a treatment effect estimate in a matched subset of patients rather than in the overall trial population since the method finds the largest subset of matched patients for which covariate balance criteria are met.
- The HC borrowing source had some differences in patient characteristics at baseline. In particular, the trial had a greater proportion of stage III patients.

Conclusions

- The SCA using RWD was unable to successfully replicate the OS treatment effect estimate from the matched population of the RCT, possibly due to unmeasured confounding, differences in time periods and follow-up, and differences in subsequent therapy.
- By applying BB, we demonstrated how the incorporation of HCs can be applied to improve the precision of estimates and inform risk of bias assessments when outcomes and patient characteristics differ between data sources (e.g., SCA and appropriate HC), providing a valuable tool for consideration in comparative effectiveness for HTA.
- Our approach demonstrates the ability to produce more reliable measures of comparative effectiveness to increase confidence in choice of therapy for HTA applications.

References

1. Thorlund et al. Clin Epidemiol. 2020;12:457-467; 2. US Food and Drug Administration (FDA). 2021. https://www.fda.gov/media/127912/download; 3. European Medicines Agencies (EMA). 2006. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-trials-small-populations_en.pdf; 4. EMA. 2013. https://www.ema.europa.eu/en/documents/scientific-guideline/concept-paper-extrapolation-efficacy-safety-medicine-development_en.pdf; 5. US FDA. 2010. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials; 6. Herbst et al. The Lancet Oncology. 2018;19(1):101-114; 7. Socinski et al. J Clin Oncol. 2012;30(17):2055-2062; 8. Guyot et al. BMC Med Res Methodol. 2012;12:9; 9. Chen MH, Ibrahim JG. Statistical Science. 2000;15(1):46-60

Disclosures

This study was funded by Daiichi Sankyo Europe. A Struebing, ND and MR are employed by Daiichi Sankyo Europe. PH, YT, and YX are employed by Daiichi Sankyo Inc. CM, HR, EM, RV and A Springford are employees of Cytel Inc. This presentation was prepared using data from Dataset nci-data-470_NCT00946712 from the NCTN/NCORP Data Archive of the National Cancer Institute's (NCI's) National Clinical Trials Network (NCTN) obtained from www.projectdatasphere.org, which is maintained by Project Data Sphere. Data were originally collected from clinical trial NCT number NCT00946712. All analyses and conclusions in this presentation are the sole responsibility of the authors and do not necessarily reflect the opinions or views of the clinical trial investigators, the NCTN, the NCORP, the NCI, or Project Data Sphere.