# The Impact of Censoring Assumptions on STC-Adjusted Simulated Time-to-Event Data

**Elizabeth Vinand,[1] Andre Verhoek,[2] ***

[1]Cytel, Inc., London, UK [2]Cytel, Inc., Rotterdam, Netherlands. *At the time of the study.

## Background

- Unanchored simulated treatment comparisons (STC) estimate relative treatment effects in the absence of head-to-head trials.
- STCs adjust for population differences between studies by modeling the relationship between patient characteristics and outcomes using individual patient data (IPD) from one trial to estimate outcomes in another trial with aggregated data.
- When using STCs for survival outcomes, regression models predict events (e.g., disease progression or death); however, these do not inherently account for censoring, which occurs when patient observation ends (e.g., loss to follow-up or end of the trial).
- With no formal guidelines, researchers must simulate censoring based on assumptions to avoid bias in the STC arm.[1]
- Given that different censoring assumptions can greatly impact survival outcomes, this research aims to explore how varying these assumptions influences the estimates generated in STC analyses.

## Objective

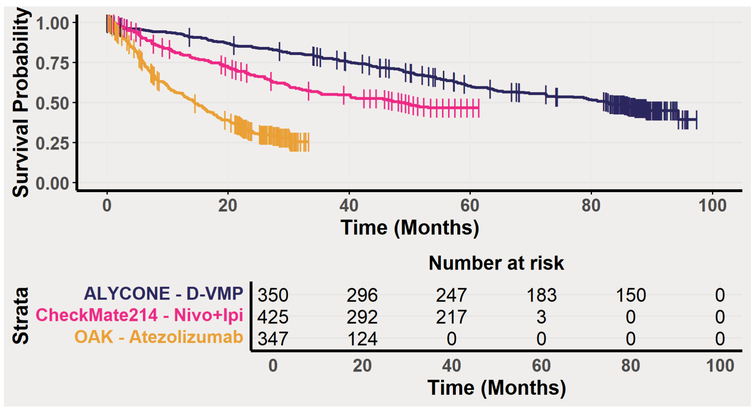
Explore how different censoring assumptions affect survival outcomes in unanchored STCs and identify the optimal censoring assumption that minimizes errors in STC analyses.

## Methods

### Data

- The active treatment arms from three phase 3 oncology trial datasets were used, focusing on overall survival (OS) outcomes.
- The first dataset was from the PD-L1 >1% subpopulation of the OAK trial of atezolizumab vs a control arm in previously treated non-small-cell lung cancer.[2,3] The second dataset was from the intermediate/poor-risk disease subgroup of the CheckMate 214 trial, of nivolumab plus ipilimumab vs a control arm in advanced renal-cell carcinoma.[4] The third dataset was from the ALYCONE trial of daratumumab in combination with bortezomib, melphalan, and prednisone (D-VMP) was compared to a control arm in newly diagnosed stem-cell transplantation ineligible multiple myeloma.[5] Only the active arm of each dataset was used to perform the STC.

**Figure 1. Kaplan Meier survival curves for the active treatment arms of each dataset**




Abbreviations: D-VMP, daratumumab in combination with bortezomib, melphalan, and prednisone; Nivo+Ipi, nivolumab plus ipilimumab.

### STC settings

- STCs were performed using digitized pseudo-IPD from the active arm of the trials (Fig. 1).
- Dummy baseline mean age and sex were created to ensure the variables adjusted for in the STC were non-significant treatment effect modifiers and would not influence predicted survival, enabling a direct comparison between the STC-predicted and unadjusted survival curves to identify the most accurate censoring assumption.
- Therefore, in this setting, the active arm is both the pivotal (unadjusted) and the competitor arm (STC-adjusted).
- To emphasize the adjustment process, extreme baseline differences between the pivotal and competitor arm were applied, with mean age set to 60.1 years in the pivotal arm vs 20 years in the competitor, and 50% vs 98% females, respectively.

### Censoring methods

Censoring scenarios assessed for impact on outcomes:

**1. Censoring at the end of the trial period** (Fig. 2A): all censoring is applied at the last observed event.

**2. Evenly spread censoring across the trial period** (Fig. 2B): censoring is distributed evenly over the entire trial period.

**3. Per-trial censoring based on occurrence probability** (Fig. 2C): censoring times simulated by sampling with replacement from the original Kaplan-Meier survival curve.

**4. Censoring centered around the trial period median** (Fig. 2D): censoring centered around the median event time using a Normal distribution with 20% of the median as standard deviation.

**5. Predominantly early censoring (fat-tailed distribution)** (Fig. 2E): censoring times sampled from a Pareto distribution, resulting in mostly early censoring occurrences.

**6. Predominantly late censoring (fat-tailed distribution)** (Fig. 2F): censoring times sampled from a Pareto distribution, configured to produce mostly late censoring.

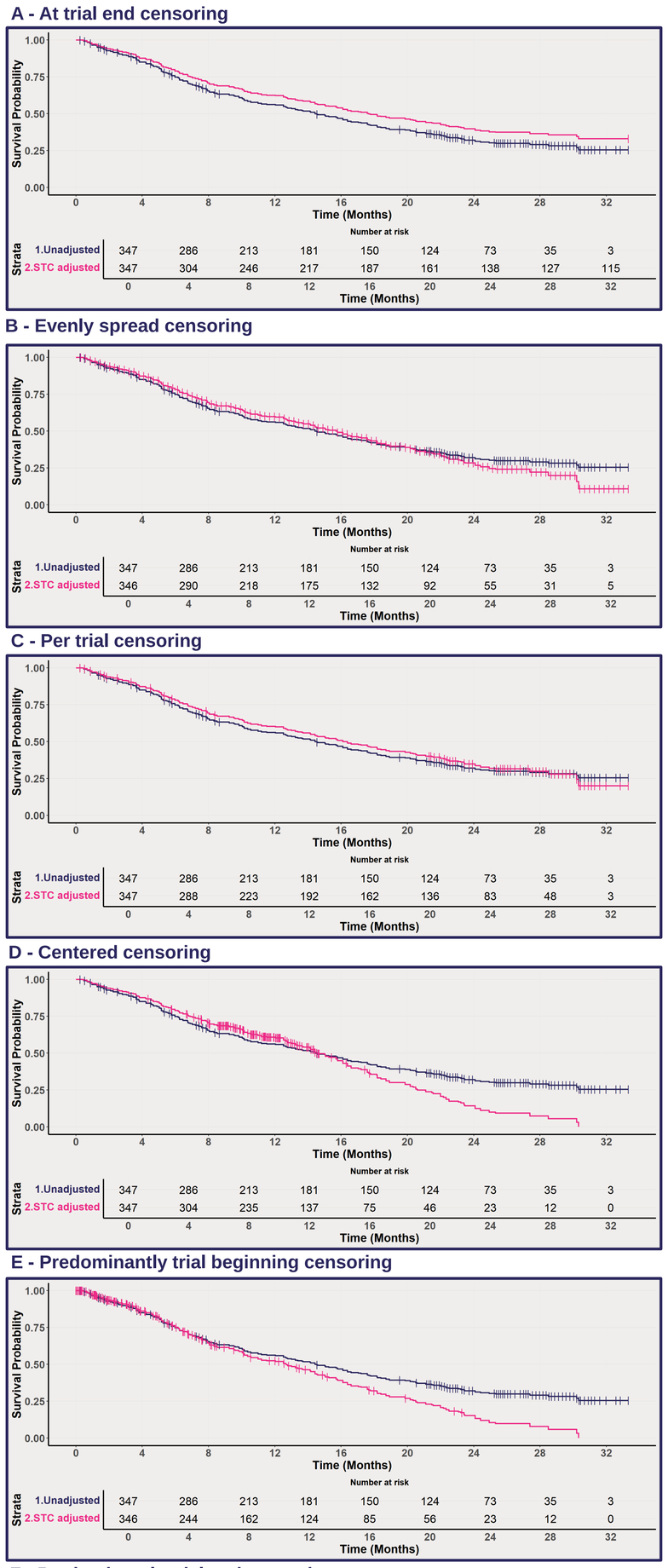
## Methods (continued)

### Benchmarking and evaluation

- We compared unadjusted and STC-adjusted hazard ratios (HR), median OS, and restricted mean survival time (RMST) at trial follow-up and at 10 years.
- These metrics allowed us to assess how effectively each censoring approach captured survival dynamics and influenced the accuracy of the STCs.

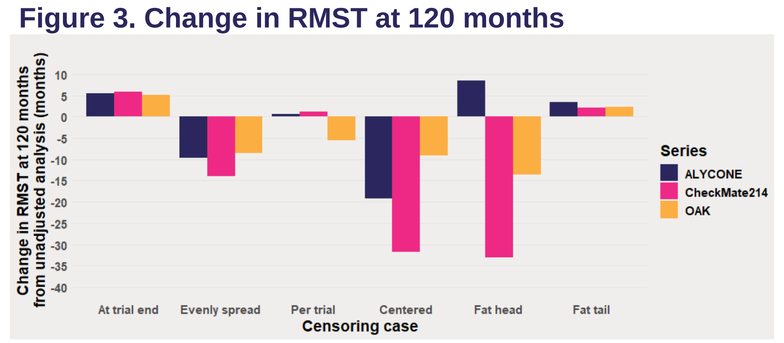
## Results

**Figure 2. Censoring scenario settings: the OAK trial**




Abbreviations: STC, simulated treatment comparison.

**Figure 3. Change in RMST at 120 months**




Abbreviations: RMST, Restricted Mean Survival Time.

## Results

- Across all three trials, similar patterns emerged for the impact of different censoring assumptions on survival outcomes.
- The **per-trial censoring** approach consistently stayed closest to the unadjusted benchmark results, both in terms of HRs and distribution choice, followed by the two scenarios involving censoring at the end of the trial (scenarios 1 and 6); other scenarios showed considerable variance (Tables 1–3).
- Scenario 5 performed well in the OAK trial but deviated substantially in the CheckMate 214 and ALYCONE trials, highlighting inconsistencies across datasets.
- Notably, centered censoring (scenario 4) did not work well in any of the datasets, consistently leading to distorted results.
- The impact on **HRs** was considerable, ranging from 0.81 to 2.09 across trials (Tables 1–3), underscoring the importance of censoring assumptions in shaping the results.
- The choice of best-fitting distribution was also affected, with most scenarios resulting in a different distribution vs the unadjusted trial. Again, per-trial censoring remained closest to the unadjusted distribution choices, offering the most consistent and reliable results.
- For **long-term extrapolation**, substantial differences were observed across all trials. Different assumptions notably influenced the optimal distribution fit, with the per-trial censoring approach yielding the most accurate long-term predictions for RMST at 10 years, closely followed by late censoring (Fig. 3).

**Table 1. Results from OAK trial**

| Censoring case | HR | Median survival (months) | Best fitting distribution | RMST trial period (months) | RMST at 10 years (months) |
|---|---|---|---|---|---|
| **Unadjusted** | **-** | **14.41** | **Log-normal** | **17.04** | **27.23** |
| At trial end | 0.81 | 17.66 | Log-normal | 19.02 | 32.42 |
| Evenly spread | 1.08 | 15.68 | Weibull | 16.35 | 18.75 |
| Per trial | 0.95 | 16.10 | Gamma | 17.59 | 21.71 |
| Centered | 1.38 | 14.58 | Gen Gamma | 18.17 | 18.17 |
| Trial beginning | 1.53 | 12.68 | Gompertz | 13.32 | 13.69 |
| Trial end | 0.87 | 17.57 | Log-logistic | 18.43 | 29.59 |

Abbreviations: Gen Gamma, generalized gamma; HR, hazard ratio; RMST, restricted mean survival time.

**Table 2. Results from CheckMate214 trial**

| Censoring case | HR | Median survival (months) | Best fitting distribution | RMST trial period (months) | RMST at 10 years (months) |
|---|---|---|---|---|---|
| **Unadjusted** | **-** | **48.3** | **Log-normal** | **40.15** | **61.07** |
| At trial end | 0.88 | NA | Gen Gamma | 42.09 | 66.95 |
| Evenly spread | 1.27 | 38.08 | Gamma | 36.71 | 47.10 |
| Per trial | 0.94 | 51.06 | Log-normal | 41.09 | 62.33 |
| Centered | 1.48 | 35.14 | Gen Gamma | 29.25 | 29.25 |
| Trial beginning | 2.39 | 24.42 | Weibull | 26.21 | 28.03 |
| Trial end | 0.93 | 51.06 | Log-normal | 41.42 | 63.11 |

Abbreviations: Gen Gamma, generalized gamma; HR, hazard ratio; NA, not applicable; RMST, restricted mean survival time.

**Table 3. Results from ALYCONE trial**

| Censoring case | HR | Median survival (months) | Best fitting distribution | RMST trial period (months) | RMST at 10 years (months) |
|---|---|---|---|---|---|
| **Unadjusted** | **-** | **82.96** | **Gompertz** | **68.12** | **76.13** |
| At trial end | 0.81 | 94.39 | Gompertz | 72.12 | 81.69 |
| Evenly spread | 1.32 | 72.54 | Gompertz | 62.64 | 66.49 |
| Per trial | 0.95 | 87.26 | Gompertz | 69.39 | 76.81 |
| Centered | 1.34 | 74.05 | Gen Gamma | 56.58 | 56.85 |
| Trial beginning | 2.27 | 53.42 | Gen Gamma | 84.52 | 84.52 |
| Trial end | 0.84 | 94.39 | Gompertz | 71.09 | 79.79 |

Abbreviations: Gen Gamma, generalized gamma; HR, hazard ratio; RMST, restricted mean survival time.

## Conclusions

- This study highlights the critical impact of censoring assumptions on survival outcomes in STCs.
- Across all three oncology trials, **the per-trial censoring approach** consistently provided the most accurate estimates, aligning closely with unadjusted results in terms of HRs, survival distributions, and long-term extrapolations.
- Censoring at the end of the trial (scenarios 1 and 6) also performed reasonably well, but other assumptions, especially centered censoring (scenario 4), introduced considerable variability and often led to inaccurate results.

### References

**1.** Lagakos et al. Biometrics. 1979:139-56. **2.** Fehrenbacher et al. J. Thorac. Oncol. 2018;13(8):1156–70. **3.** Rittmeyer et al. Lancet. 2017;389(10066):255–65. **4.** Albiges et al. ESMO open 5.6. 2020:e001079. **5.** Mateos et al. Blood. 2022. 140(Supplement 1):10157-9.

Presented at ISPOR Europe | Barcelona, Spain | November 17–20, 2024