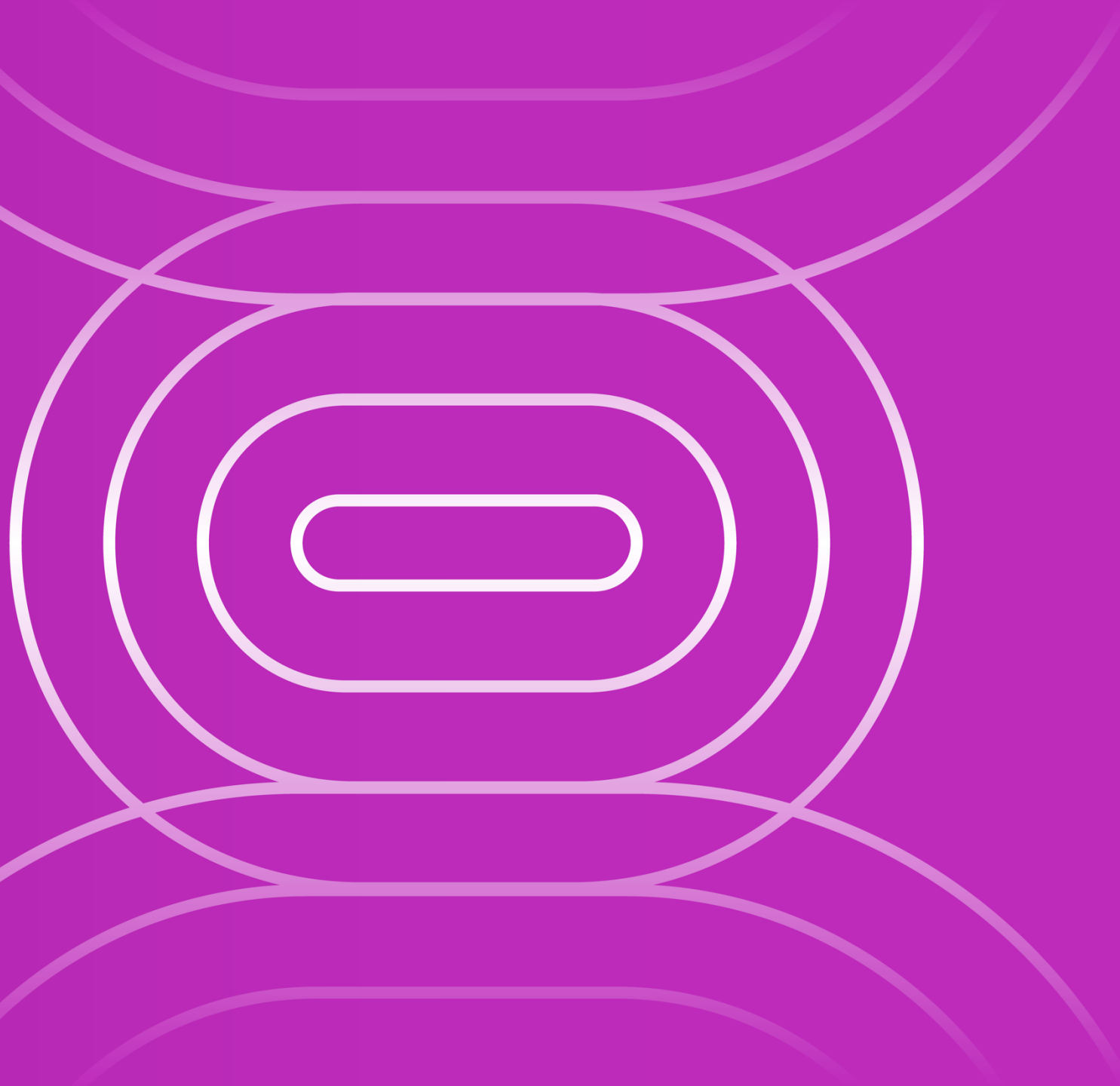


# PHUSE US Connect 2026



# Index Date Selection in Externally Controlled Arms: Methods, Impact, and Practical Considerations

Lorraine Fang, Lixiao Chen, Kelvin Lin, Kevin Yang, Jenny Han  
Marc De Benedetti, Lihua Yue

Bristol Myers Squibb, Lawrence Township, NJ, USA

March, 2026

# Section I

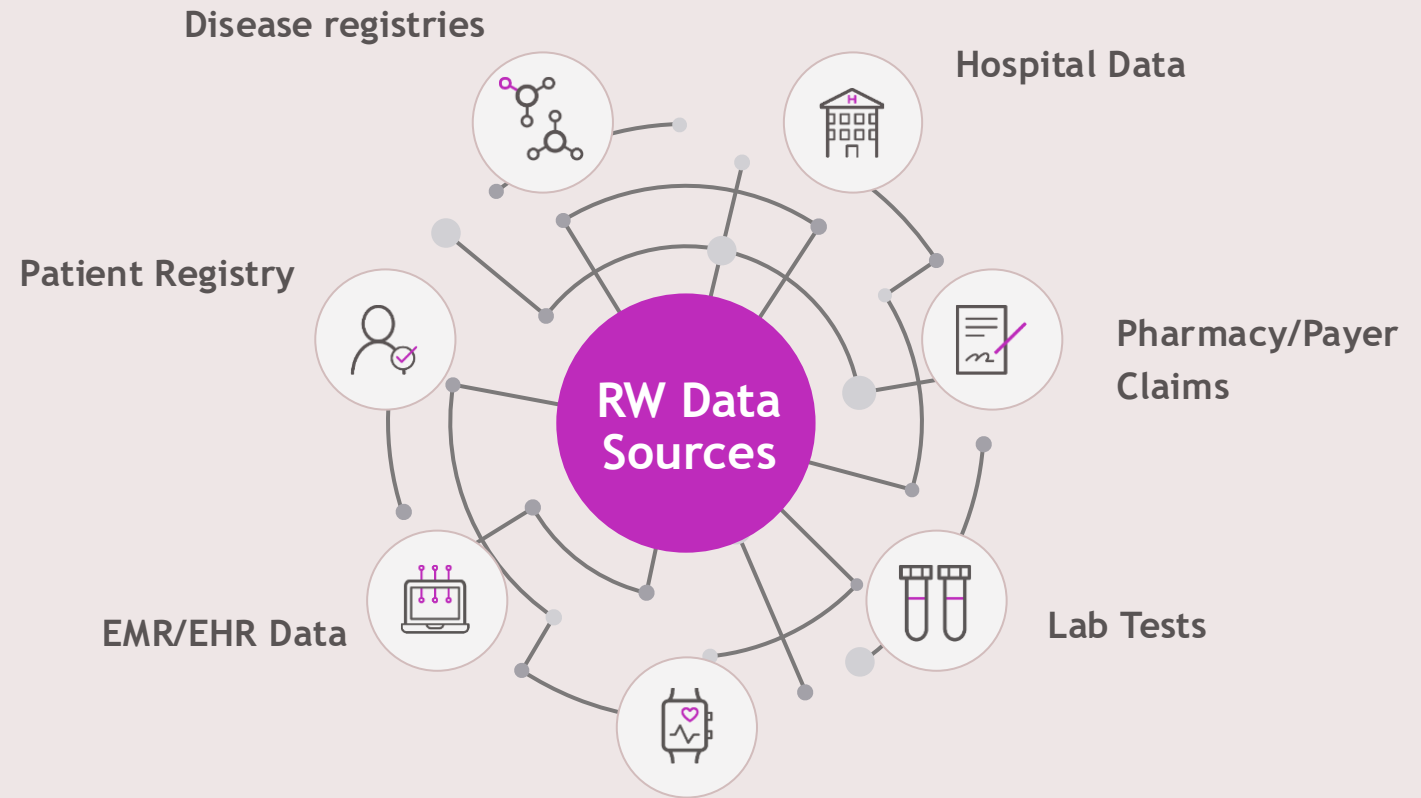
## Background & Methodological Overview

# Potential Use of External/Synthetic Control Arm

- Single-arm clinical trial study
  - No control group, no context to interpret outcomes
- Why might a clinical study only have one arm? For example:
  - Rare diseases – difficult to obtain large recruitment
  - Diseases with highly predictable courses and mortality
  - Situations when the effect of treatment is substantial (ethically better to give all subjects new treatment)
- How do we perform efficacy analyses on single-arm studies?
  - Use external/historical controls to artificially create a control arm to compare

# Data Sources for ECAs

- Data vendors act as an intermediary for sponsors
  - They pool data sources
- Common data vendors that have collated from these sources:
  - Cota
  - GRN
  - IQVIA
  - Flatiron
  - TriNetX



# Notable Challenges with using RWD

- Clinical trials: data collection is standardized and planned
- RWD is not standardized – issues that arise
  - Value accuracy, conformance, plausibility, consistency
  - Completeness/missing values
  - Data availability (e.g. variables may not be recorded in EMR's routinely)
  - Transparency of data processing/ data quality
  - RWD is often retrospective – we have subject's complete history
  - “Fairly” choosing an index date for these subjects?

# Propensity Score (PS) methodologies

The **propensity score** is the conditional probability that a subject belongs to their assigned treatment group given the observed covariates

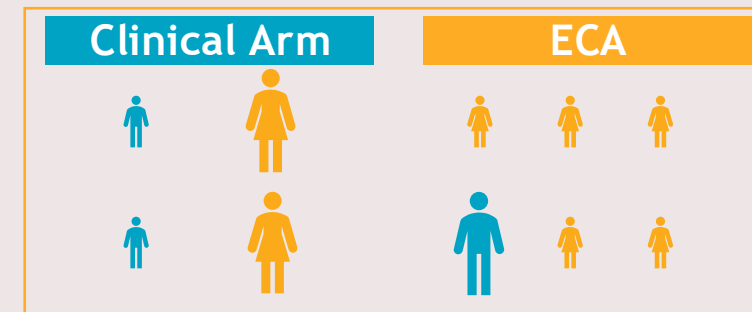
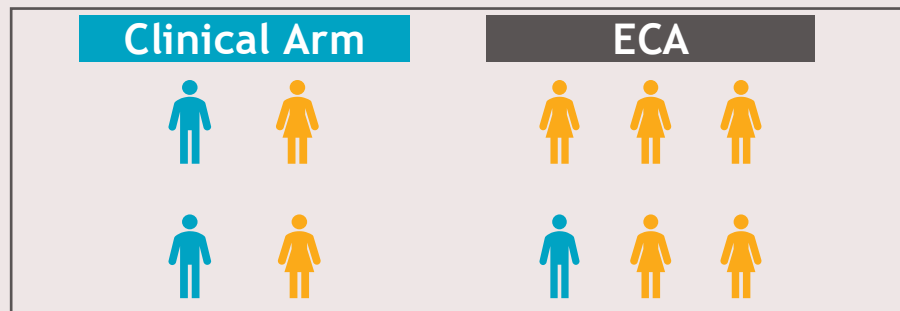
- The identified covariates must be associated with both the treatment group as well as the outcome
  - Example: age is commonly a covariate that is associated with both treatment group as well as outcome (older patients may be given different treatment plan due to their age compared to younger patients)
- These are the factors clinical trial physicians try to randomize when creating each arm in an RCT
- A subject's propensity score is a measure of the assignment bias introduced by their medical history.
- RWE studies use “**Propensity Score Methodologies**” to help **account for and correct these biases** (mimicking a RCT as best we can).

# Inverse Probability of Treatment Weighting

The propensity scores do not immediately allow us to properly compare between treated and non-treated subjects.

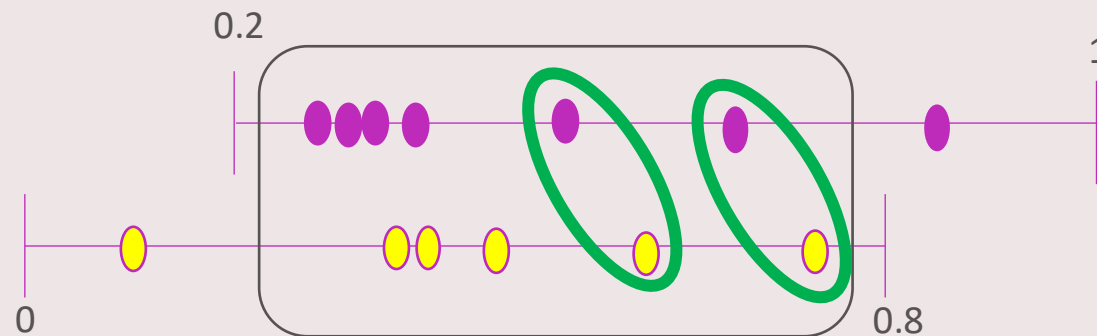
It's how the PS value is used that allows us to achieve balance between the two arms.

- Main idea behind these **weights** are to **scale each observation** in the input data set according to their PS.



# PS Matching

- Uses the PS value to “match” subjects in the other arm
- This represents a comparison between subjects that have a **similar set of covariates** – allowing of a meaningful comparison of outcomes.



# Section II

## Case Study

# Outline

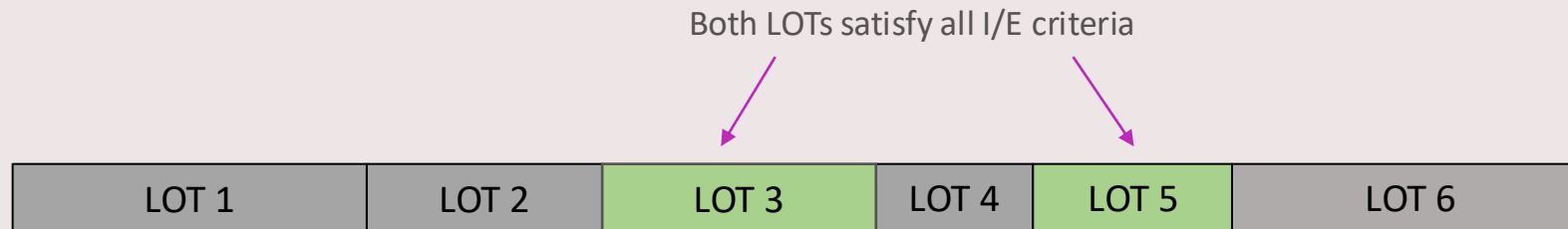
- Methods
  - Data Sources
  - Indexing Methods
  - Cohort Definitions
  - Statistical Methods
- Results
  - Numerical results
  - Discussion
- Summary

# Methods: Data Sources

- Active Arm (N = 257) – taken from the 3L+ LBCL study
  - TRANSCEND had PET-positive R/R diffuse large B-cell lymphoma (DLBCL) after exposure to an anthracycline and rituximab or other CD20-targeting agents, treated with Breyanzi
  - Index Date: Breyanzi infusion date
- External Control Arm (N=varying values)
  - NDS-NHL-001 study was a global, noninterventional, retrospective, observational study using multiple sources selecting patients with R/R LBCL
  - Inclusion/exclusion criteria mirrors TRANSCEND
  - There were N = 606 patients who met broad eligibility criteria, named ICC
  - Remaining eligibility was assessed at each line

# Context and Motivation

Clinical trial subjects have a very clear index date. However, the index date must be chosen retrospectively when using RW data.

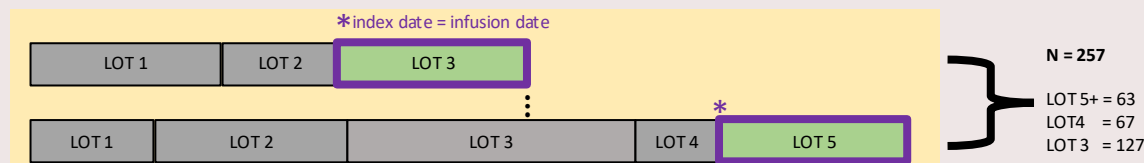


This project explores the performance and **impact** of multiple **index date** selection methods and their impact on efficacy results **using real patient data**

# Indexing Methods

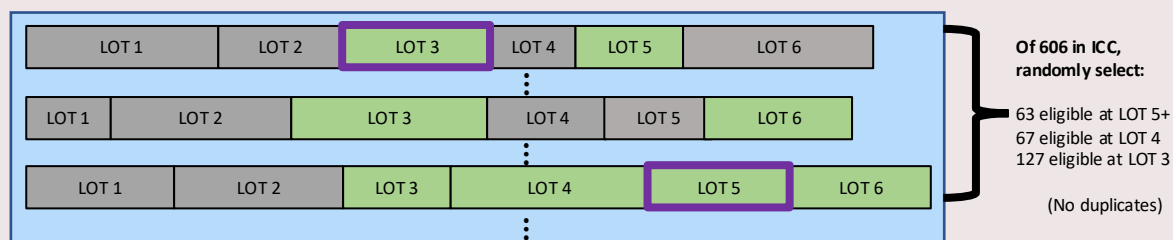
## 0. Clinical Subjects

Clinical subjects are indexed at date of treatment (Breyanzi infusion)



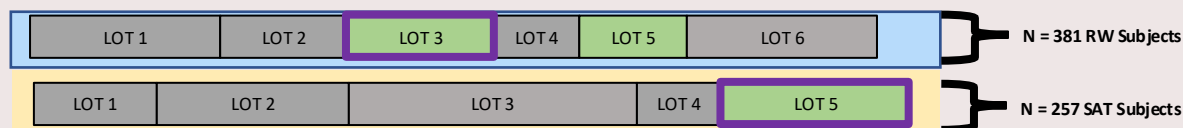
## 1. Stratified Random Line (SRL), N = 257

**Novel method:** Randomly select RW patients (without duplication) indexed to match the observed clinical LOT distribution



## 2. First Eligible Line (FEL), N = 381

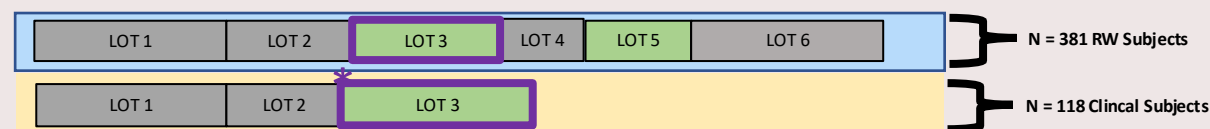
Include RW patients indexed at a time point when they first meet trial (TRANSCEND) eligibility criteria



# Indexing Methods

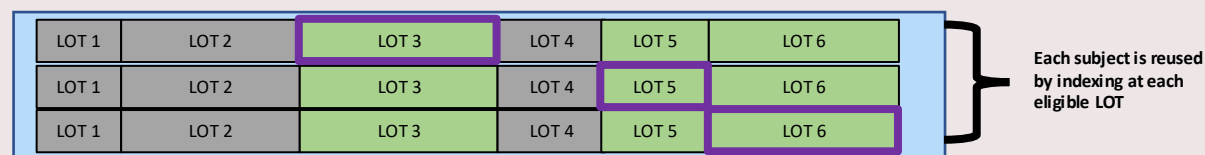
## 3. Restricted to 3<sup>rd</sup> Line (3L), $N_{\text{Clinical}} = 118$

Select patients from both arms who are first eligible at the start of 3<sup>rd</sup> line. Note: same RW cohort as FEL



## 4. All eligible Lines (AEL), $N = 653$

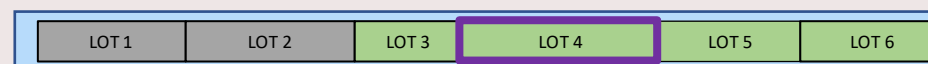
Select all eligible RW patients, reusing and indexing patients at all eligible LOTs.  $N=653$  with 443 unique patients



## 5. Random Eligible Line (REL)

For each RW patient, assign index date based on random selection of eligible LOT

Note: Similar to SRL, however, REL uses all patients and resulting LOT distribution may align with clinical



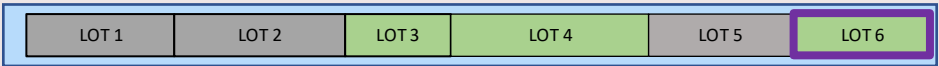
Legend

- = Clinical Trial
- = RWD comparator
- = Ineligible LOT
- = eligible LOT
- = Selected Index LOT

# Indexing Methods

## 6. Last Eligible Line (LEL), N = 443

Index RW patients at their most recent eligible LOT



## 7. Second-Last Eligible Line (SEL), N = 443

If there is more than one eligible line, index RW subjects at their second-last eligible line



# Indexing Methods: Summary

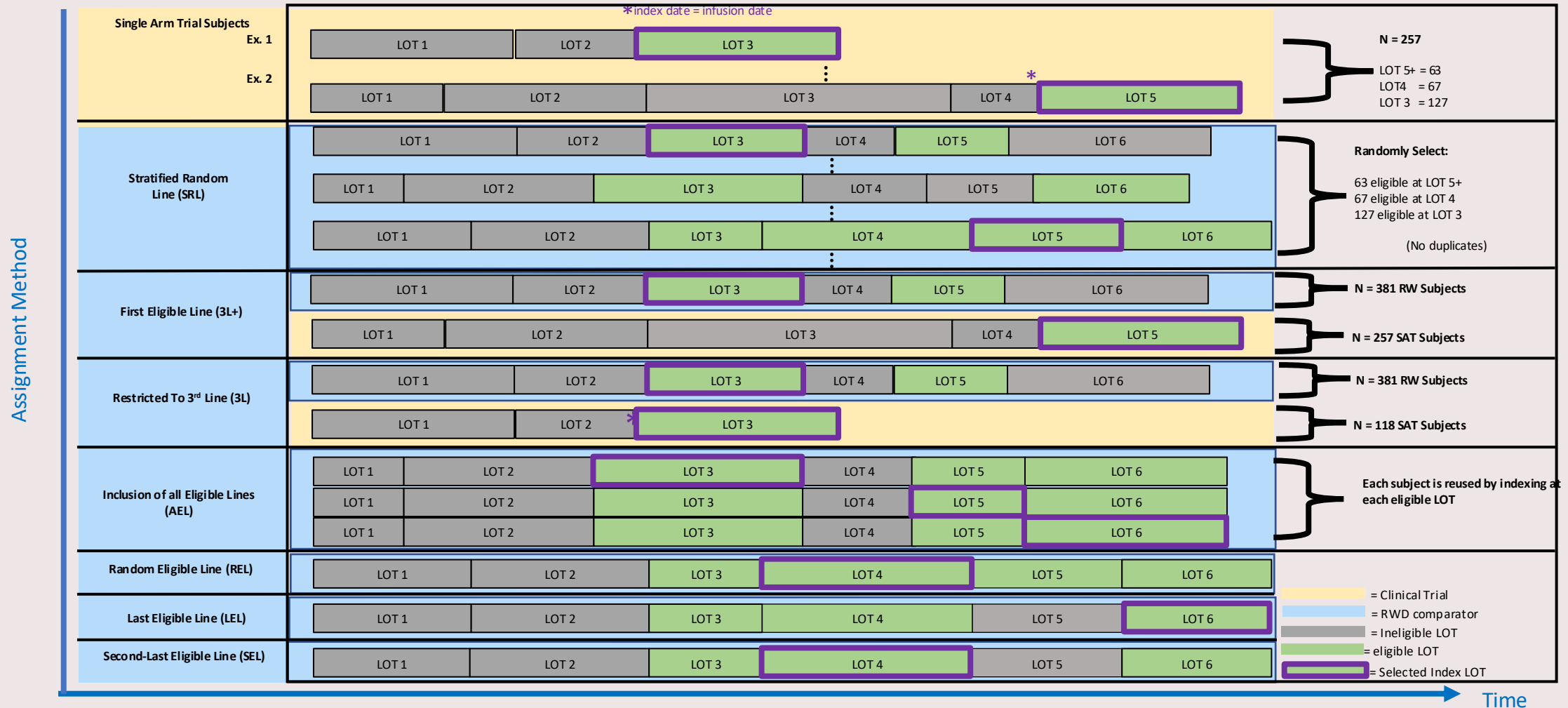
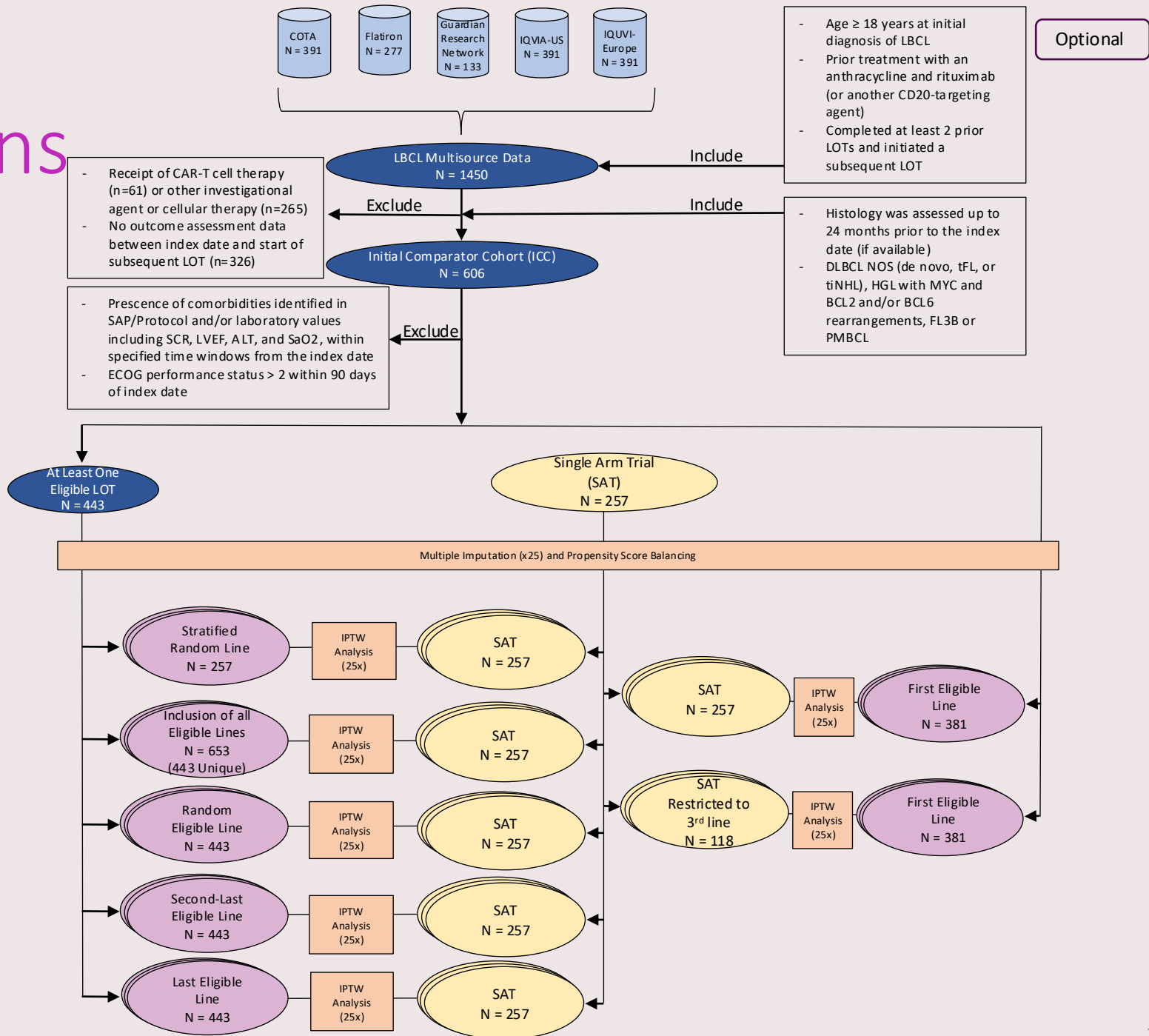


Figure 1. Visualizations of the seven indexing methods

# Cohort Definitions



# Statistical Methods

- **Balance** between the two arms was obtained using propensity score (PS) stabilized inverse probability treatment weighting (**sIPTW**)
- The following 12 prognostic factors were used in the propensity score (PS) model for each of the seven indexing methods:

- |                                                 |                                                    |
|-------------------------------------------------|----------------------------------------------------|
| 1. Age                                          | 7. Best response to any prior therapy              |
| 2. Gender                                       | 8. Prior HSCT (hematopoietic stem cell transplant) |
| 3. Time since diagnosis                         | 9. Chemotherapy -<br>refractory/sensitive          |
| 4. Number of Prior LoT                          | 10. Bulky Disease                                  |
| 5. Number of Prior LoT per year since diagnosis | 11. Extranodal Disease                             |
| 6. R/R to last therapy                          | 12. Disease Stage                                  |

Factors were identified based on available demographic and baseline disease characteristics in TRANSCEND, with additional factors included based on literature review and medical review.

- 25 independent multiple imputations were performed to handle missing RW data
- A SMD > 0.2 was used to assess balance of all prognostic factors

# Statistical Methods: End points

- Crude and sIPTW-adjusted overall response rate (ORR), complete response rate (CRR), overall survival (OS), and progression-free survival (PFS) were calculated
- Generalized linear models/Cox proportional hazard models with sIPTW weights were used in the adjusted analysis to estimate the relative risk (RR) and hazards ratio (HR) for each outcome of interest.

# Numerical Results: Demographics

•All RW subjects in FEL and 3L (ECA<sub>2,3</sub>) have 2 prior LOTs

- impossible to adjust in the PS model

•The following variables were unbalanced before sIPTW adjustment

- Number of Prior LOTs (except in SRL and 3L)
  - R/R to last therapy
- Prior HSCT
  - Bulky Disease

• All were balanced after sIPTW

Table 1. Summary of Selected Baseline Characteristics

|                               | SRL ECA <sub>1</sub><br>(N = 257) |       | FEL ECA <sub>2,3</sub><br>(N = 443) |       | AEL ECA <sub>4</sub><br>(N = 653) |       | REL ECA <sub>5</sub><br>(N = 443) |       | LEL ECA <sub>6</sub><br>(N = 443) |        | SEL ECA <sub>7</sub><br>(N = 443) |       | TRANSCEND<br>(N = 257) |       | 3L<br>(N = 118) |       |
|-------------------------------|-----------------------------------|-------|-------------------------------------|-------|-----------------------------------|-------|-----------------------------------|-------|-----------------------------------|--------|-----------------------------------|-------|------------------------|-------|-----------------|-------|
| Age, in years (μ)             | 61                                |       | 60.8                                |       | 60.8                              |       | 61.2                              |       | 61.3                              |        | 61.1                              |       | 60.2                   |       | 59.4            |       |
| Gender: Male (n, %)           | 95                                | 37%   | 143                                 | 37.5% | 228                               | 34.9% | 165                               | 37.2% | 165                               | 37.2%  | 165                               | 37.2% | 88                     | 34.2% | 36              | 30.5% |
| Months since diagnosis (μ)    | 26                                |       | 22                                  |       | 27.1                              |       | 28.7                              |       | 29.8                              |        | 28.1                              |       | 31.3                   |       | 21.1            |       |
| Number of Prior LoTs (μ, σ)   | 2.8                               | 0.82  | 2                                   | 0     | 2.7                               | 1.04  | 2.5                               | 0.81  | 2.7                               | 1.02   | 2.3                               | 0.73  | 2.9                    | 1.35  | 2               | 0     |
| Prior LoTs/Year (μ, σ)        | 2.2                               | 1.38  | 2.0                                 | 1.22  | 2.1                               | 1.31  | 2.0                               | 1.29  | 2.1                               | 1.31   | 2.0                               | 1.24  | 2.2                    | 1.64  | 2.2             | 1.44  |
| Refractory to Last Therapy    | 238                               | 92.6% | 341                                 | 89.5% | 592                               | 90.7% | 407                               | 91.9% | 407                               | 91.9%  | 397                               | 89.6% | 204                    | 79.4% | 94              | 79.7% |
| Best Response ≥ PR (n, %)     | 175                               | 68.1% | 229                                 | 60.1% | 422                               | 64.6% | 283                               | 63.9% | 187                               | 64.8%  | 282                               | 63.6% | 222                    | 86.4% | 99              | 83.9% |
| Had prior HSCT (n, %)         | 45                                | 17.5% | 42                                  | 11.0% | 117                               | 17.9% | 64                                | 14.4% | 72                                | 16.3%  | 64                                | 14.4% | 87                     | 33.9% | 32              | 27.1% |
| Chemosensitive (n, %)         | 60                                | 23.3% | 95                                  | 24.9% | 169                               | 25.9% | 117                               | 26.4% | 109                               | 24.6%  | 121                               | 27.3% | 86                     | 33.5% | 39              | 33.1% |
| Has Bulky Disease (n, %)      | 51                                | 19.8  | 80                                  | 21.0  | 131                               | 20.1  | 85                                | 19.2  | 84                                | 19.6.0 | 54                                | 19.0  | 29                     | 11.5  | 10              | 8.5   |
| Has Extranodal Disease (n, %) | 154                               | 59.9% | 512                                 | 56.4% | 401                               | 61.4% | 265                               | 59.8% | 276                               | 62.3%  | 263                               | 59.4% | 134                    | 52.1% | 64              | 54.2% |
| Disease Stage                 |                                   |       |                                     |       |                                   |       |                                   |       |                                   |        |                                   |       |                        |       |                 |       |
| I/II                          | 48                                | 18.7% | 73                                  | 19.2% | 106                               | 16.2% | 74                                | 16.7% | 72                                | 16.3%  | 77                                | 17.4% | 69                     | 26.8% | 30              | 25.4% |
| III/IV                        | 144                               | 56.0% | 200                                 | 52.5% | 374                               | 57.3% | 243                               | 54.9% | 245                               | 55.3%  | 241                               | 54.4% | 185                    | 72.0% | 87              | 73.7% |
| Mean Follow-Up Time (months)  |                                   |       |                                     |       |                                   |       |                                   |       |                                   |        |                                   |       |                        |       |                 |       |
| All Subjects                  | 8.6                               |       | 9.9                                 |       | 9.2                               |       | 9                                 |       | 8.3                               |        | 9.5                               |       | 14.5                   |       | 15.1            |       |
| Surviving Subjects (mo, %)    | 14.9                              | 33.5% | 16.3                                | 31.8% | 15.2                              | 34.8% | 15.1                              | 35.4% | 14.3                              | 35.0   | 15.5                              | 35.4% | 21.2                   | 52.1% | 21.3            | 55.9% |

# Numerical Results: Adjusted Outcomes

Table 2. Summary of Adjusted Efficacy Outcomes for all Endpoints

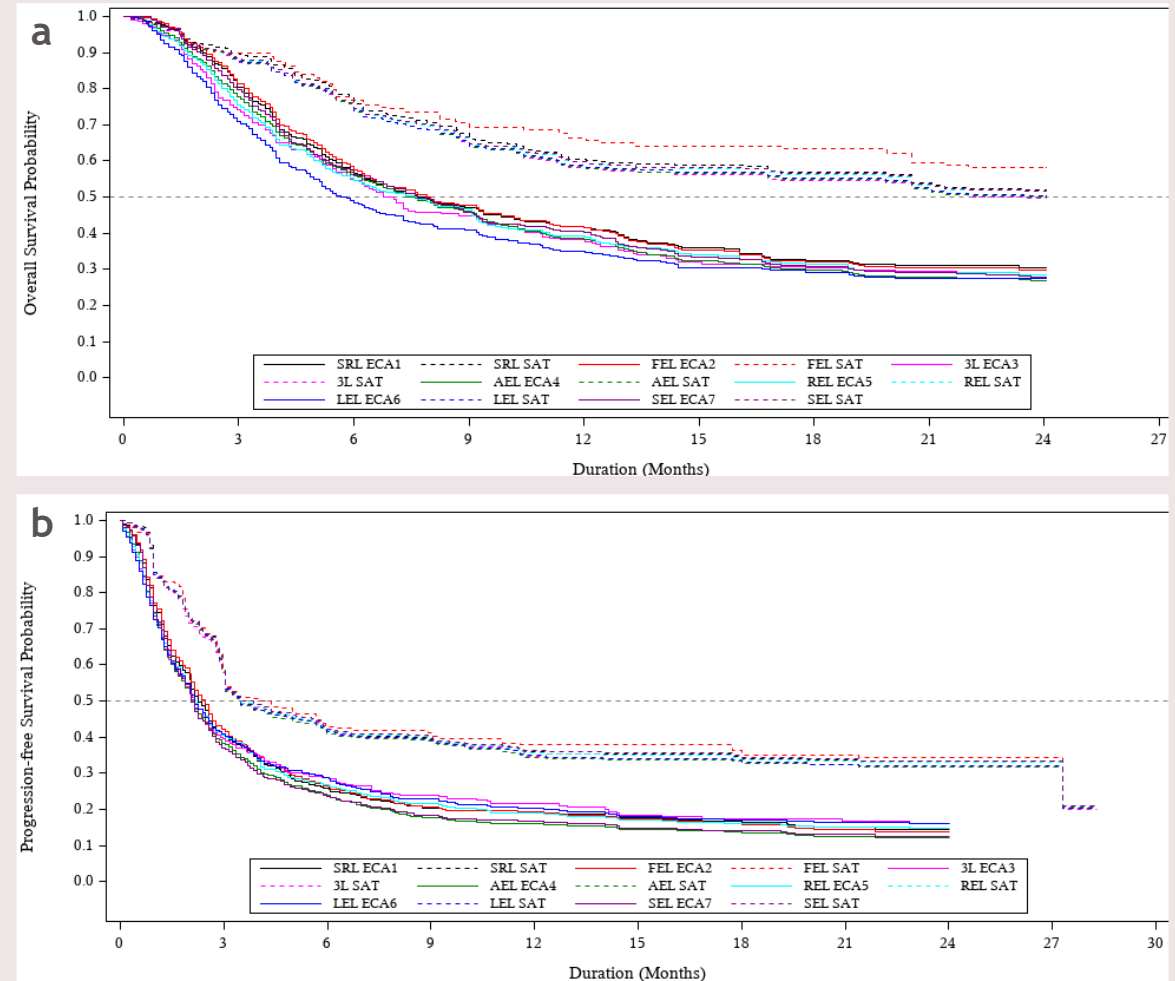
|                        | SRL ECA <sub>1</sub><br>(n=257) | SAT<br>(n=257) | FEL ECA <sub>2</sub><br>(n=381) | SAT<br>(n=257) | 3L ECA <sub>3</sub><br>(n=381) | SAT3L<br>(n=118) | AEL ECA <sub>4</sub><br>(n=653) | SAT<br>(n=257) | REL ECA <sub>5</sub><br>(n=443) | SAT<br>(n=257) | LEL ECA <sub>6</sub><br>(n=443) | SAT<br>(n=257) | SEL ECA <sub>7</sub><br>(n=443) | SAT<br>(n=257) |
|------------------------|---------------------------------|----------------|---------------------------------|----------------|--------------------------------|------------------|---------------------------------|----------------|---------------------------------|----------------|---------------------------------|----------------|---------------------------------|----------------|
| <b>ORR, %</b>          | 38.8                            | 73.8           | 38.9                            | 74.7           | 39.6                           | 76.1             | 34.3                            | 74.2           | 39.9                            | 74.7           | 40.2                            | 74.7           | 35.8                            | 74.5           |
| <b>RR (95% CI)</b>     | 1.9 (1.6-2.3)                   |                | 1.9 (1.6-2.3)                   |                | 1.9 (1.6-2.3)                  |                  | 2.2 (1.9-2.5)                   |                | 1.9 (1.6-2.2)                   |                | 1.9 (1.6-2.1)                   |                | 2.1 (1.8-2.4)                   |                |
| <b>p-value</b>         | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                       |                  | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                |
| <b>CR rate, %</b>      | 24.1                            | 50.1           | 20.4                            | 49.9           | 20.7                           | 52               | 18.3                            | 49.6           | 21.4                            | 50             | 23.9                            | 50             | 18.9                            | 50.7           |
| <b>RR (95% CI)</b>     | 2.1 (1.6-2.8)                   |                | 2.4 (1.9-3.2)                   |                | 2.5 (1.9-3.4)                  |                  | 2.7 (2.2-3.4)                   |                | 2.3 (1.8-3.0)                   |                | 2.1 (1.7-2.6)                   |                | 2.7 (2.1, 3.4)                  |                |
| <b>p-value</b>         | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                       |                  | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                |
| <b>Median OS (mo)</b>  | 6.8                             | 23.5           | 7.9                             | NR             | 8                              | NR               | 7.6                             | NR             | 7.6                             | NR             | 5.7                             | NR             | 7.7                             | NR             |
| <b>HR (95% CI)</b>     | 0.52 (0.40-0.68)                |                | 0.53 (0.41-0.69)                |                | 0.45 (0.31-0.65)               |                  | 0.52 (0.41-0.66)                |                | 0.51 (0.40-0.65)                |                | 0.48 (0.38-0.61)                |                | 0.51 (0.40-0.66)                |                |
| <b>p-value</b>         | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                       |                  | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                |
| <b>Median PFS (mo)</b> | 2.2                             | 3.5            | 2.3                             | 3.5            | 2.5                            | 4.4              | 2.2                             | 3.5            | 2.2                             | 3.5            | 2.2                             | 3.5            | 2.1                             | 3.9            |
| <b>HR (95% CI)</b>     | 0.60 (0.48-0.75)                |                | 0.58 (0.46-0.72)                |                | 0.57 (0.42-0.77)               |                  | 0.55 (0.45-0.66)                |                | 0.58 (0.47-0.71)                |                | 0.60 (0.49-0.73)                |                | 0.53 (0.50-0.77)                |                |
| <b>p-value</b>         | < 0.0001                        |                | < 0.0001                        |                | 0.0003                         |                  | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                | < 0.0001                        |                |

- Adjusted RR and HR values are all close, and highly significant in favour of Breyanzi
- Due to high treatment effect of Breyanzi, it was difficult to obtain deep insight
- Slight numerical differences can still be seen between the seven methods

# Numerical Results: Adjusted Outcomes

- Each indexing method shows a strong difference between ECA and TRANSCEND
- None of the ECA and TRANSCEND curves intercept, and the proportionality assumption appears to be valid

Figure 3. Adjusted OS (a) and PFS (b) for sIPTW



# Discussion

- Overall Survival

- LEL had the **lowest** median **OS** and second-lowest HR:

Median OS = 5.7 months, HR = 0.48 (0.38-0.61)

- 3L had the **longest** median **OS**, but the **lowest HR**:

Median OS = 8.0 months, HR = 0.45 (0.31-0.65)

- Other indexing methods had **comparable** median **OS** and **HR** values

- Progression-Free Survival

- Each method produced **similar median PFS** values: 2.1 – 2.5 months

- SRL and LEL had the **highest HR**: 0.6, CI's (0.48-0.75) and (0.49-73), respectively

- AEL and SEL performed most consistently across all 4 end points

- SRL and AEL were most effective in achieving balance

# Discussion: Noteworthy Pros/Cons from Findings

|     | Pros                                                                                                                                                                                              | Cons                                                                                                                                                                                                                             |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SRL | <ul style="list-style-type: none"> <li>No obvious bias introduced</li> <li>Results in exceptional balance</li> <li>Only method that will ensure balance of important prognostic factor</li> </ul> | <ul style="list-style-type: none"> <li>Requires large RW dataset</li> <li>Requires more RW subjects in each strata than observed in the clinical distribution</li> <li>Reduces statistical power by limiting N of ECA</li> </ul> |
| FEL |                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>Risk of being <b>biased towards</b> improved survival outcomes in <b>RW</b></li> </ul>                                                                                                    |
| 3L  | <ul style="list-style-type: none"> <li>Ensures uniform value for important highly prognostic factor (prior LOT)</li> <li>Helps to eliminate any potential selection bias</li> </ul>               | <ul style="list-style-type: none"> <li>Risks reducing sample size of analysis and thus reduces statistical power</li> </ul>                                                                                                      |
| AEL | <ul style="list-style-type: none"> <li>Most suitable in situations with small sample sizes</li> <li>Many of the other benefits of SRL</li> </ul>                                                  | <ul style="list-style-type: none"> <li>No standard accepted method to correct for with-in patient correlation that is introduced</li> </ul>                                                                                      |
| REL | <ul style="list-style-type: none"> <li>See SRL, however, cannot ensure balance</li> <li>Does not share the cons of SRL</li> </ul>                                                                 |                                                                                                                                                                                                                                  |
| LEL |                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>Potential to introduce bias (ex. survivor bias)</li> </ul>                                                                                                                                |
| SEL |                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>See LEL</li> </ul>                                                                                                                                                                        |

# Summary

- Time-dependent end points can be difficult to interpret in studies that use RWD
- Small body of work on this topic
  - still a new area of exploration
  - the **only study** that uses **patient-level real data** (RW and Clinical)
- Seven different indexing methods were used to explore OS, PFS, ORR, CRR
- Large treatment effect of Breyanzi made it **challenging to obtain deep insight** into differences between indexing methods
  - However, some **strengths/weaknesses** of each method were able to be **explored**
- **There is no one best method**
  - Each study has unique conditions and requirements
  - An **indexing method** needs to be **carefully chosen** during study design that best aligns with the identified strengths and weaknesses

*Thank You*

Lorraine.fang@bms.com