Supplementary Appendix

Table of Contents

1. Data collection procedures by source .................................................................................. 3
2. Eligibility criteria for SCHOLAR-5 (Sub-cohort A) ......................................................... 5
3. Variable Definitions .............................................................................................................. 5
4. Propensity score methods ..................................................................................................... 6
5. Missing Values ....................................................................................................................... 9
6. Multiple imputation ............................................................................................................. 11
7. Sensitivity analyses .............................................................................................................. 12
8. Index treatment selection .................................................................................................... 13

References ................................................................................................................................ 14

Supplementary Tables and figures ........................................................................................ 15

Table S1: Hierarchy of variables for propensity score specification ..................................... 15
Table S2: Karnofsky Status Conversion to ECOG Status ..................................................... 16
Table S3: Imputation Rules for Partial or Missing Start Dates ............................................. 17
Table S4: Data values for tumour bulk, response to last line of therapy, and time since last treatment, and relapsed/refractory status before and after multiple imputation ............ 18
Table S5: Demographic and clinical characteristics ............................................................. 19
Table S6: Randomly selected index treatment patterns for the final SCHOLAR-5 sample for comparative effectiveness before and after weighting .......................................................... 21
Table S7: Randomly selected index treatment patterns for the final SCHOLAR-5 sample for comparative effectiveness before and after weighting by continent .................................... 22
Table S8: Estimated proportions of overall response rate and complete response among the sensitivity analyses ................................................................................................................ 23
Table S9: Estimated proportions of overall response rate and complete response among the subgroup analyses ............................................................................................................. 24
Figure S2: Results from sensitivity analysis .......................................................................... 25
Figure S3: Results from subgroup analyses .......................................................................... 26
Figure S4: Duration of response Kaplan-Meier curve in the 3L+ population ....................... 27
Figure S5: Kaplan-Meier plots comparing ZUMA-5 to SCHOLAR-5 among patients that are 4th line of therapy or higher for A. Progression-free survival; B. Overall survival; C. Time-to-next-treatments; and D. Duration of response. ................................................................. 28
Figure S6: Kaplan-Meier plots comparing ZUMA-5 to SCHOLAR-5 among POD24 patients for A. Progression-free survival; B. Overall survival; C. Time-to-next-treatments; and D. Duration of response ................................................................................................................ 29
Figure S7: Kaplan-Meier plots comparing ZUMA-5 to SCHOLAR-5 among patients with prior exposure to stem cell transplants for A. Progression-free survival; B. Overall survival; C. Time-to-next-treatments; and D. Duration of response .................................................... 30
1. Data collection procedures by source

1. Memorial Sloan Kettering Comprehensive Cancer Center is one of the largest and the oldest Cancer Centers in the world, and it is ranked as the second most important Cancer Center in the United States. The lymphoma program at MSKCC includes more than 20 oncologists focusing exclusively on lymphoma, and a portfolio of more than 100 clinical trials dedicated to lymphoma. The data collection period extended through to 14 September 2020.

2. The Department of Hematology of Hospices Civils de Lyon (HCL) at Lyon Sud Hospital is one of the largest French and European haematological centers especially for the management of lymphoma patients. A specific clinical research team conducted more than 100 clinical trials specifically for lymphoid malignancies. The department is an active member of the Lymphoma Study Association (LYSA). The data collection period extended through to 22 July 2020.

3. The Barts Cancer Institute (BCI) was created in 2003, and brought together some of the most eminent cancer research teams in London to the Historic St. Bartholomew’s (Barts) Hospital, the oldest hospital in England and the Barts and The London School of Medicine and Dentistry, Queen Mary University of London and is a Cancer Research UK Centre of excellence. BCI forms part of the Cancer Research UK City of London (CoL) Centre, which is a world class hub for cancer biotherapeutics, together with our partners from three other of the central London Cancer Research UK centres: University College London, King’s College London, and The Francis Crick Institute. The data collection period extended through to 17 August 2020.

4. The Christie is a large Comprehensive Cancer Centre in the northwest of England receiving more than 14,000 new patient referrals annually. With the University of Manchester and Cancer Research UK the Christie forms the Manchester Cancer Research Centre (MCRC) and is also a partner in the Manchester Academic Health Science Centre. The Lymphoma Group has a large clinical trial and translational program and a research focused approach to patient care. The data collection period extended through to 13 October 2020.

5. The Vall d’Hebron University Hospital (VHUH) is the second largest hospital in Spain and it covers all medical and surgical specialties. It has more than 1400 beds and treats around 1,200,000 patients per year. Established in 2006, the Vall d’Hebron Institute of Oncology (VHIO) is a leading comprehensive cancer center of excellence where its scientists and research physicians work together as multidisciplinary teams to both accelerate and advance personalized and targeted therapies against cancer. The clinical research unit has conducted more than 400 clinical trials during the last year in oncological and haematological malignancies. The data collection period extended through to 4 September 2020.

6. IPO Porto is the largest Comprehensive Cancer Center in Portugal. Every year it treats around 40,000 patients, 10,000 of whom are new patients, in 11 integrated practice units. Its Clinical Research Unit has conducted more than 80 clinical trials in hematologic malignancies. IPO Porto Research Center also comprises two research units dedicated to real world evidence studies – Management, Outcomes Research and Economics in Healthcare (MOREHealth) Group and Cancer Epidemiology Group. The data collection period extended through to 17 July 2020.

7. Vanderbilt-Ingram Cancer Center is a leader in the prevention, diagnosis and treatment of cancer. The center's world-renowned team of experts provides an integrated, personalized and patient-centered approach to cancer care, including treatment, research, support, education and outreach. From a wide variety of wellness programs to a leading REACH for Survivorship Clinic, patients find support from diagnosis through survivorship. Vanderbilt-Ingram is a National Cancer Institute-designated Comprehensive Cancer Center, one of just two centers in Tennessee and 51
in the country to earn this highest distinction and ranks in the top 10 nationwide for cancer research grant support. The data collection period extended through to 17 December 2020.

Sub-cohort A Clinical sites 1-6
Data from 6 sites across the US, UK, France, and Spain were collected from electronic medical records. For eligible patients, data were accessed and extracted by appropriately trained analysts or research fellows from the different participating sites. Site selection was based on availability and completeness of data for variables of interest, as well as sufficient patient numbers given agreed inclusion/exclusion criteria. A common data model (CDM) was developed for this study and used to ensure consist variable names and definitions when extracting data.

Sub-cohort A Clinical site 7: Vanderbilt Medical Centre
Data for the VUMC SD component of the SCHOLAR-5 cohort came from electronic medical records collected through a wholly owned subsidiary of VUMC, Nashville Biosciences. Data from consented patients are de-identified under HIPAA Safe Harbor standards, including removal of identifying fields, manual review of clinical notes, use of global research identifiers, and time-shifting of index date. The study CDM was used to guide manual review. This manual review was performed by a physician trained in the use of the CDM. The most recent, as well as the relevant, prior hematology notes were identified and used to obtain clinical data. Relevant imaging reports and laboratory measurements were also reviewed and extracted based on the requirements of the CDM. Sub-cohort A patient level key variables included demographics, clinical characteristics (Table S5), therapeutic regimens received, and death or censoring dates. Patient-level line of treatment variables extracted included time varying baseline characteristics, best overall response for each line of therapy received, progression date and treatment start and end dates.

DELTA Trial cohort (Sub-cohort B)
The DELTA Trial Cohort included a subset of patients from the DELTA Phase 2 clinical trial of idelalisib sponsored by Gilead.¹ The long-term follow-up of the DELTA study collected data on subsequent lines of treatment after idelalisib including best response. In order to limit over-representation of idelalisib in the SCHOLAR-5 cohort, only the immediate follow-up line of therapy (LoT) after the interventional LoT was included. Patients from the DELTA trial that met the inclusion criteria for ZUMA-5 were identified by: 1) Assessing if they had sufficient prior LoT after re-aligning LoT with the ZUMA-5 definition (e.g., not counting anti-CD20 monotherapy); 2) Ensuring patients were relapsed or refractory; and 3) Ensuring patients had follicular lymphoma and had not transformed. Differences in key definitions across the two studies were identified and addressed. Sub-cohort B key variables included in this analysis were baseline clinical characteristics, best overall response, date of next treatment initiation and survival time.

ZUMA-5
Study KTE-C19-105 (ZUMA-5) was a Phase 2, multicenter, single arm, open-label study evaluating the efficacy of axicabtagene ciloleucel in patients with r/r B-cell iNHL of FL or MZL histological subtypes. ² Patients proceeded through the following study phases: screening, enrolment/leukapheresis, conditioning chemotherapy, investigational product (IP) treatment, post-treatment assessment, and long-term follow-up (LTFU). Data were collected through medical history, physical exams, and blood draws. Procedures also included baseline electrocardiograms, echocardiograms, PET-CT, possible bone marrow aspirate, and biopsy. Only patients with 18 months follow-up were included in the main analysis, resulting in a sample of 86 patients. The full group of 124 were included in a sensitivity analysis (see Section 7).

Disease response and progression assessments
Responses in Sub-cohort A and B used a variety of methods to assess response including computed
tomography (CT) scans and Cheson criteria. In ZUMA-5 tumor response and progression was evaluated using positron emission tomography (PET) scans and Lugano criteria.

2. Eligibility criteria for SCHOLAR-5 (Sub-cohort A)

   1. Overall inclusion criteria for the SCHOLAR-5 cohort were: Patients aged ≥18 years;
   2. Patients with histologically confirmed diagnosis of iNHL, with histological subtype limited to FL Grade 1, Grade 2, or Grade 3a or MZL nodal/extranodal based on criteria established by the World Health Organization (WHO) 2016 classification (data from patients with MZL were omitted at the analysis stage);
   3. Patients with r/r disease (i.e., r/r iNHL) starting third or higher line of therapy on or after 23rd July 2014 (exact date differed according to individual cohort component protocols). Prior line of therapy with anti-CD20 monotherapy did not count as line of therapy for eligibility.

Patient level Exclusion criteria for the SCHOLAR-5 cohort were:

   1. Transformed FL;
   2. FL Histological Grade 3b;
   3. Prior anti-CD19 CAR T-cell therapy or other genetically modified T-cell therapy;
   4. Eligible within 12 months before the last updated version of the database (site specific)

3. Variable Definitions

ECOG
The measure of ECOG used as a covariate was an augmented ECOG, meaning that when ECOG was not reported and the Karnofsky’s index of performance status was available, ECOG was derived using this score. The methods of imputation used for ECOG are detailed in Section 5 of the section on handling of missing values.

FLIPI
The FLIPI score ranges from 0 to 5 and consists of the five sub-scores for Stage, lactate dehydrogenase (LDH), hemoglobin (HB), age group and number of involved nodal sites. Each sub-score is scored with a score of either 0 or 1, with a score = 1 per criterion if
- Stage = III-IV
- LDH > upper limit of normal (ULN)
- HB <=12 g/dl
- Age > 60 years
- > 4 nodal sites

When FLIPI was not provided explicitly and all of the sub-scores were available, the overall score was derived from its definition.

Previous LoT
The number of previous LoT was assigned according to the number of previous eligible LoT. Eligible LoT differed from LoT assignment from some of the data sources. As such, in all sub-cohorts LoT were reviewed and LoT numbering re-assigned. Based on ZUMA-5 trial criteria, anti-CD20 monoclonal antibody monotherapy, radiotherapy on its own, surgery on its own and watch and wait were all ineligible as a line of therapy and not counted towards the prior lines of therapy.
**Relapsed versus refractory**

Refractory disease was defined as progressing (defined as PD) during or within 6 months after completion of the most recent prior treatment. Relapsed disease was defined as progressing after CR, PR or SD > 6 months after completion of the most recent prior treatment. Based on these definitions, as set in the SAP, some patients may have progressed and not be identified as relapsed or refractory. For example, a patient does not have a date of completion for the prior treatment. Someone in his or her last line of therapy, was assumed to still be on treatment and was deemed refractory. Cases where the exact classification of whether progressive disease constituted relapsed or refractory disease were not excluded.

**POD24**

For sub-cohort A, POD24 was defined as patients having progressed within 24 months after initiation of first-line anti-CD20 chemotherapy combination therapy. Only patients with a first line of therapy that included an anti-CD20 combined chemotherapy were eligible to be evaluated as POD24. Switching therapy within 24 months was not sufficient to be considered POD24. For Sub-cohort B, the definition was solely based on switching treatment within 24 months of initiating first-line chemoimmunotherapy because progression in first LoT was not collected. Defining POD24 based on switching treatments should capture all but a few patients meeting the definition above, but should also identify patients that do not meet the definition (e.g., a patient switching treatment for another reason than progression). As such, there is expected to be over-reporting of POD24 in Sub-cohort B and thus an under-correction for the imbalance. Such a bias will be in favor of SCHOLAR-5 rather than ZUMA-5.

4. Propensity score methods

**Propensity score methods overview**

Propensity score methods are used to provide unbiased estimates of comparative effectiveness. This involves the calculation of propensity scores for each patient in the pooled analysis set using a logistic regression predicting treatment assignment according to prognostic patient characteristics. Accordingly, a higher propensity score value represents a higher probability of a patient being included in the treated cohort based on their prognostic characteristics.

The application of formal weighting or matching methodologies using the derived propensity scores allows for baseline characteristics to be balanced between cohorts prior to the estimation of comparative effectiveness. This is superior to naïve comparisons between cohorts as the potential for systematic bias is reduced (Rosenbaum and Rubin, 1983), as demonstrated across many disease areas and in previous comparative effectiveness analysis of axi-cel in patients with refractory aggressive NHL (Neelapu et al., 2017).

**Propensity score weighting**

Using SMR weighting, treated patients (i.e., patients in ZUMA-5) are given a weight of one while weights for control patients (i.e., patients in SCHOLAR-5) are defined as the ratio of the estimated propensity score to one minus the estimated propensity score. This calculation is displayed below, where the weight, \(W_i\), assigned in the SMR method for each individual, \(i\), based on propensity score, \(\pi\), is:

\[
W_i = \frac{\pi_i}{1 - \pi_i}
\]

Thus, Standardized mortality ratio (SMR) weighting reweights the control patients to be representative of the treated population, while retaining the trial outcomes for ZUMA-5 to improve interpretability of the
results. A robust variance estimator is required when utilizing this approach, similar to the variance estimator used in the generalized estimating equation methodology, which results in confidence intervals that are conservative (i.e., have a slightly greater than nominal coverage). ³

The use of propensity score weighting (as opposed to matching) provides the advantage that all patients are included in the analysis. To ensure that baseline characteristics are balanced between the two cohorts, SMR weighting results in the provision of higher weights for patients in SCHOLAR-5 most similar to the ZUMA-5 population and small or negligible weights for those who exhibit minimal overlap with the ZUMA-5 population. This functions to increase the influence of comparable patients, while reducing (or effectively removing) the influence of patients who do not present with similar baseline characteristics.

**Propensity score matching**

Propensity score matching was performed as a sensitivity analysis using the procedure of 1:1 matching, which matches each patient in the treated (ZUMA-5) group with the control (SCHOLAR-5) patient exhibiting the nearest propensity score (this is also known colloquially as ‘greedy’ matching). Consequently, if all patients from the smallest group (i.e., treated or control) are matched, then the sample size for the patients included in subsequent analysis becomes double the sample size of the smallest group of patients. If an appropriate match is not available, for example due to a lack of overlap in propensity score values between groups, then cases are discarded and the matched sample size for analysis is reduced accordingly. PS matching was selected as a sensitivity analysis, rather than being used as the primary comparative analysis, in order to maintain the totality of the data for analysis.

For PS matching, a caliper width of 0.2 times the standard deviation of the propensity score was used⁴, and ‘random’ order, with a prespecified ‘seed’. PS matching was performed using the package ‘MatchIt’⁵ within R.

**Propensity score weighting versus matching**

Propensity score weighting provides the advantage that all patients are included in the analysis and retains the trial outcomes for the treated group to improve interpretability of the results, and

Alternatively, propensity score matching selects one control patient to be matched with each treated patient (1:1 matching) based on having the most similar propensity score values. Patients are either included or excluded from the analysis according to whether a match could be achieved, rather than incorporating weighted data as with the propensity score weighting approach. However, achieving a match for every patient within the pre-specified threshold is unlikely, which results in a reduced sample size for the analysis.

**Propensity score specification**

To apply propensity scores to each observation, clinical advice was obtained by the investigator team (prior to data availability) regarding which patient characteristics should be balanced between groups at baseline (providing that data allowed), as well as the hierarchical order of importance based on the prognostic value of each variable. Based on this pre-specified clinical input, Table S1 specifies covariates as low, medium or high priority, with the medium and high priority variables ranked according to the level of importance.

Although FLIPI score was initially part of the pre-specified list of covariates for the propensity score model, the large amount of missing data for this variable (69.5% for SCHOLAR-5) precluded incorporation into the model. In addition, further clinical feedback highlighted that FLIPI was most prognostic when assessed at disease diagnosis rather than at later lines of therapy. FLIPI score at diagnosis was not available for the ZUMA-5 cohort, which further prevented inclusion of this variable in
the propensity score model. Bone marrow involvement was also part of the pre-specified list of covariates for the propensity score model but was ultimately excluded due to large amounts of missing data (65.9% for SCHOLAR-5).

In order to maintain the validity of the covariates, any variable exhibiting >40% missing data in either dataset was not eligible for multiple imputation and subsequent inclusion in the propensity score model (see S6 or further details regarding multiple imputation).

The primary propensity score model was constructed with the requirement that the standardized mean difference (SMD) between groups for characteristics included in the model were <0.10, and that the p-value for tests of significant difference for these variables were >0.10. Once this was achieved, the resulting model was presented to clinical experts to confirm an appropriate balance between groups (including characteristics not included in the propensity score model).

To achieve an appropriate balance of baseline characteristics, it was pre-specified that variables would be refactored in a stepwise manner due to sparse counts and in reverse order of importance should the above requirements not be met with each implementation of the propensity score specification. As a result of SMD values >0.1, the following refactoring was required on subsequent implementations to achieve an SMD <0.1 for all variables included in the model:

- First implementation: no refactoring of variables.
- Second implementation: prior SCT refactored to a binary variable (yes/no) due to sparse counts.

The balance of baseline variables with each propensity score specification is presented in Table S4, as is the balance achieved with the final propensity score specification (second implementation). The successful balance of baseline variables after refactoring meant that variables did not need to be removed from the propensity score specification; however, the need to modify the propensity score from the initial implementation precluded the addition of low priority variables, as pre-specified in the statistical analysis plan. The primary comparative analysis propensity score specification was used for each sensitivity and subgroup analyses, with the baseline variables assessed for any statistical or clinically meaningful difference between groups.

Once weights had been calculated (whether propensity score, or matching based), these were used throughout the presented results.

**Assessing balance between groups**
The standardized mean difference is widely accepted as the most appropriate method for assessing the balance of patient characteristics between groups after the application of propensity score methods. This has the advantage of not being influenced by sample size and allowing for the comparison of the relative balance of variables measured in different units. An SMD less than 0.1 is considered a conservative estimate to indicate a negligible difference in the mean or prevalence of a covariate between treatment groups. This was used in combination with the requirement for a p-value for tests of significant difference for patient characteristics of >0.1. Once this was achieved in the present study, the resulting model was presented to clinical experts to confirm an appropriate balance between groups for all prognostically relevant patient characteristics.
5. Missing Values

Missing Data on Covariates

Multiple imputation (MI) is superior to single imputation in reducing statistical bias by building random error into the imputation. MI prevents under-estimation of the degree of variability in the data, which affects the estimate of the standard error.

Missing data were imputed using ‘mice’ package in R under the assumption that the missing values were missing at random (MAR). To support the assumption that imputed data are MAR, Little’s test was performed prior to imputation to test that the data were not missing completely at random (MCAR). Finally, to increase the accuracy of imputation, MI was not performed for a variable if the missing data exceeded 40% of the total for either the SCHOLAR-5 or ZUMA-5 data.

The methods specified in the MI procedure were for “arbitrary” patterns of missingness using the imputation method “FCS” (Full Conditional Specification). The seed were fixed to 123. The FCS method used a separate conditional distribution for each imputed variable (the form of each conditional distribution was assigned according to the imputed variable). The resulting dataset with 100 imputed values per missing covariate were stored and used for all corresponding causal inference analyses. Descriptive statistics and the number of missing data are presented for the model covariates in a separate table.

ECOG Performance Missing Data

The Karnofsky’s index of performance status (KPS) was converted to ECOG status 0-4 when ECOG was not available or missing. The ECOG 0-4 grade is summarized in Table S2. If the ECOG value was missing for the 6-months period before the index therapy start date and could not be taken from the Karnofsky status, it was checked whether the value right before and after the period was available, identical and within the range of 0-1, in which case the ECOG value was set to this stable pre/post value. The identical approach was taken for the Karnofsky status being classifiable as either 100% (ECOG=0) or 80-90% (ECOG=1). If the ECOG score could not be derived this way but was > 1 at the last measurement before the index date, the patient was excluded from any line of treatment analysis which occurred later than the ECOG measurement date.

Partial Dates

The following partial dates were imputed as per Table S3:

- Adverse event (AE) start dates
- Medication start dates (including LoT start dates)
- Clinical and laboratory dates:
  - Gene expression assessment dates
  - Laboratory characteristics assessment dates
  - Medical history/Comorbidity diagnosis dates

Additionally, for classifying prior, concomitant and post medications according to the treatment exposure start and end dates, the treatment end dates were imputed the following way:

1) If year and month are available but day were missing, the date was set to the last day of the month.
2) If year was available but day and month were missing, the date was set to December 31.
The LoT end date was defined differently to the treatment exposure end date described above and was always defined as starting date of the next LoT minus one day, while treatment exposure itself could end before the end of the LoT. For the last LoT, no end date was derived.

Imputation rule for partial or missing event dates for time-to-event variables (OS, PFS, TTNT, DoR):

1) If year and month were available but day was missing, the date was set to the last day of the month.
2) If the month was also missing or the date was completely missing, the time-to-event was not calculated.

Imputation rule for partial or missing censoring dates for time-to-event variables (OS, PFS, TTNT, DoR):

1) For partial or missing censoring dates the analogous rule applied, with the censoring date needed to have at least the month and year available, else the last available (imputed) date before the missing censoring date was used.

Imputation rules for partial or missing start dates for time-to-event variables (OS, PFS, TTNT, DoR):

1) If the start day for the calculation was missing, this day was set to the 1st day of the month.
2) If the month was also missing or the date was completely missing, the time-to-event was not calculated.

These rules led to conservative time-to-event outcomes for comparison, due to missing data being imputed for the comparator data and imputing either the most advantageous dates for the available treatment options in the real-world setting.

Missing days for age calculations were set to the 15th of the month, and missing days and months for the birth day were set to the 30th of June of the year.

**FLIPI Score**
If only one sub-score was missing, but the overall FLIPI score was available, the missing sub-score was derived and used for analysis.

**POD24**
As described in section 3, POD24 was not available for Subcohort B, and so was defined as patients switching treatment within 24 months.
6. Statistical tests

A two-sided 95% confidence interval (CI) was utilized and all tests were performed on the 5% alpha level (two-sided). Differences between ZUMA-5 and SCHOLAR-5 for continuous variables were assessed using linear regression modelling, whereas categorical variables were compared between groups using logistic regression models. Rates for time-to-event variables and associated 95% confidence intervals (CI) were summarized at 3-monthly intervals using the weighted KM method, while the relative difference in hazard of the outcome between groups was estimated using a Cox proportional hazards model. The absolute risk reduction (ARR) in probability of an event occurring in ZUMA-5 compared with SCHOLAR-5 at each 3 months interval was reported, with 95% CI calculated as:

\[
ARR \pm 1.96 \sqrt{(SE(S_{ZUMA-5})^2 + SE(S_{SCHOLAR-5})^2)
\]

Whereby \(SE(S_{ZUMA-5})^2\) and \(SE(S_{SCHOLAR-5})^2\) represent the standard errors of the weighted survival curves for ZUMA-5 and SCHOLAR-5 at each respective timepoint.

Robust variance estimation similar to that used in the generalized estimating equation methodology (i.e., robust sandwich variance) was used for all analyses performed after propensity scoring. This resulted in confidence intervals that are conservative (i.e., have a slightly greater than nominal coverage).

7. Multiple imputation

To enable inclusion of the most prognostic baseline variables in the propensity score specification, multiple imputation was applied for variables with missing data that were specified as part of the propensity score model. In accordance with the propensity score specification, only high and medium importance variables with <40% missing data in either dataset were eligible for multiple imputation. To assess the missing data mechanism, Little’s test of missing completely at random (MCAR) was performed, which provided a significance value of \(p < 0.001\). This rejection of the null hypothesis suggests that the data for multiple imputation were not MCAR and that the assumption for the data being missing at random (MAR) and therefore estimable from the remaining available data is supported. 10

Multiple imputation provides an effective approach to handle missing covariate data within prognostic modelling studies, as it can properly account for the missing data uncertainty. The multiply imputed datasets are each analyzed using standard prognostic modelling techniques to obtain the estimates of interest, before the estimates from each imputed dataset are then combined into one overall estimate and variance using Rubin’s rules.11 This approach captures both the within and between imputation variability, which results in conservative confidence intervals for difference testing.

Missing data was imputed using full conditional specification with the ‘mice’ package8 in R. This data supports that axi-cel represents a significant improvement in treatment options for patients with r/r FL (R Core Team, 2020). A prespecified ‘seed’ was used for reproducibility and to ensure that outcomes could not be altered through manipulation of the randomly generated multiply imputed datasets. In accordance with guidance from the academic literature, 100 multiply imputed datasets were generated for analysis and aggregation.12
Only four propensity score variables required multiple imputation due to missing data. The distributions for these variables before and after multiple imputation are provided in Table S4. This demonstrates the imputation of two data points for the time since last treatment and relapsed versus refractory status variables in the SCHOLAR-5 cohort, 12 data points for the response to last line of therapy variable in the ZUMA-5 cohort, and 52 data points for the tumor bulk variable in the SCHOLAR-5 cohort. The distribution of each variable remained consistent before and after multiple imputation. To test the impact of multiple imputation on the study findings, a sensitivity analysis without the use of multiple imputation was performed, as detailed in below.

8. Sensitivity analyses

Exclusion of DELTA trial patients from the SCHOLAR-5 cohort
Although the index treatment for patients from the DELTA trial was for the LoT received after completion of idelalisib treatment (i.e., the focus of the clinical trial), this group of patients could still be deemed to differ from the remainder of the SCHOLAR-5 cohort due to their involvement in the clinical trial. Consequently, a sensitivity analysis was performed with the DELTA patients removed from the SCHOLAR-5 cohort.

Stricter application of ZUMA-5 eligibility criteria
Restricting the SCHOLAR-5 cohort to patients with non-missing ECOG score and being within the age range of ZUMA-5 patients (34-79 years of age) was pre-specified as a 'modified effectiveness analysis set (mEAS)'. These additional restrictions ensured that the two cohorts were tightly aligned for ECOG and age when performing the analysis.

Analyses without multiple imputation
To test the impact of multiple imputation of missing covariate data on the study findings, a sensitivity analysis without the use of multiple imputation was performed. In accordance with the pre-specification of this analysis, propensity score variables with >5% missing data for either cohort were removed from the propensity score model. This resulted in the removal of tumor bulk from the model, as 36.4% of patients in SCHOLAR-5 had missing data for this variable. Response to last line of treatment was also removed from the model, as 14% of ZUMA-5 patients had missing data for this variable. The remaining variables imputed during the primary comparative analysis were time since last treatment and relapsed versus refractory status, which were missing for 2/143 (1.4%) patients in SCHOLAR-5; consequently, this variable remained in the model with the patients removed for complete case analysis.

Inclusion of all patients in the ZUMA-5 trial in the safety analysis set
To understand the impact of restricting the ZUMA-5 cohort to the inferential analysis set in the primary comparative analysis, sensitivity analysis was performed with the inclusion of all patients from the ZUMA-5 trial who received a dose of axicabtagene ciloleucel within the safety analysis set. Note that having a minimum of 18 months of follow-up was not required in the safety analysis set.

Simplified PS
To understand the effect of reducing the number of variables in the propensity score model, a simplified model was carried out, including only the variables that were pre-specified as high priority (see supplemental Table 1).
9. **Index treatment selection**

The recruitment of ZUMA-5 occurred at 19 clinical sites and started in June 2017 with database lock in September 2020.\(^{13}\) Conversely, SCHOLAR-5 was based on EMR data from seven clinical sites from July 2014 to October 2020. The difference in patient sampling between ZUMA-5 and SCHOLAR-5 lead to an imbalance in the number of prior LoT with SCHOLAR-5 having a lower mean number of prior LoT. To ensure the most appropriate balance of prior lines of therapy between SCHOLAR-5 and ZUMA-5 when performing comparative analysis, the index date for patients in Sub-cohort A was randomly selected across all available eligible LoT (Figure 1). This approach is advocated as a method to avoid bias (e.g., immortal time bias) when selecting an index date for patients who received multiple eligible lines of therapy.\(^{14,15}\)
References

1. Gopal AK, Kahl BS, de Vos S, et al. PI3Kδ inhibition by idelalisib in patients with relapsed indolent lymphoma. *The New England Journal of Medicine*. 2014;370(11):1008-1018.
2. Jacobson C, Chavez JC, Sehgal AR, et al. Primary Analysis of Zuma-5: A Phase 2 Study of Axicabtagene Ciloleucel (Axi-Cel) in Patients with Relapsed/Refractory (R/R) Indolent Non-Hodgkin Lymphoma (iNHL). *Blood*. 2020;136(Supplement 1):40-41.
3. Brookhart MA, Wyss R, Layton JB, Sturmer T. Propensity score methods for confounding control in nonexperimental research. *Circ Cardiovasc Qual Outcomes*. 2013;6(5):604-611.
4. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*. 2011;10(2):150-161.
5. Ho D, Imai K, King G, Stuart EA. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. 2011. 2011;42(8):28.
6. Jakobsen JC, Gluud C, Wetherslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology*. 2017;17.
7. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*. 2011;46(3):399-424.
8. Buuren Sv, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45(1):1-67.
9. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American Journal of Clinical Oncology*. 1982;5(6):649-655.
10. Little RJA. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*. 1988;83(404):1198-1202.
11. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*. 2009;9(1):57.
12. Graham JW, Olchowski AE, Gilreath TD. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prevention Science*. 2007;8(3):206-213.
13. National Library of Medicine. A Phase 2 Multicenter Study of Axicabtagene Ciloleucel in Subjects With Relapsed/Refractory Indolent Non-Hodgkin Lymphoma (ZUMA-5). In: Health Nlo ed; 2021.
14. 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):758-764.
15. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of Clinical Epidemiology*. 2016;79:70-75.
Supplementary Tables and figures

Table S1: Hierarchy of variables for propensity score specification

| Variable                                                                 | Categories/Units                  | Priority (rank) |
|-------------------------------------------------------------------------|----------------------------------|-----------------|
| Progressed disease within 24 months after initiation of first line anti-CD20 chemo combination therapy (POD24) | Yes, No                          | High            |
| Number of prior lines of therapy                                        | 2, 3, 4, 5+                      | High            |
| Relapsed vs. refractory at index date (Refractory was defined as a progression (defined as PD) during or within 6 months after completion of the most recent prior treatment. Relapsed was defined as progression after complete response (CR), partial response (PR) or stable disease (SD) > 6 months after completion of the most recent prior treatment) | Relapsed, Refractory            | High            |
| Prior SCT                                                               | Autologous, Allogenic, Both, None| High            |
| Tumour bulk (diameter of largest lesion)                               | Centimetres                      | Medium          |
| Time from last treatment                                                | Months                           | Medium          |
| Best response to last line of therapy                                  | CR, PR, SD, PD                   | Medium          |
| Age                                                                     | Years                            | Medium          |
| Prior anti-CD20 + alkylating agent                                     | Yes, No                          | Medium          |
| Region                                                                  | US, Europe                       | Low             |
| ECOG (augmented by the Karnofsky index if missing)                     | 0, 1                             | Low             |
| Sex                                                                     | Male, Female                      | Low             |
| Race                                                                    | White, Not White                 | Low             |
| Prior Phosphoinositide 3-kinase (PI3K) inhibitor                       | Yes, No                          | Low             |
| Prior Bruton Tyrosine Kinase (BTKI) inhibitor                           | Yes, No                          | Low             |
| Prior anti-CD20 single agent                                           | Yes, No                          | Low             |
| Prior alkylating agent                                                 | Yes, No                          | Low             |
| Prior Lenalidomide                                                     | Yes                              | Low             |
Variable hierarchy was validated by a team of clinical experts, and prespecified in the statistical analysis plan.

Table S2: Karnofsky Status Conversion to ECOG Status

| Karnofsky status                                                                 | Karnofsky grade | ECOG grade | ECOG status |
|---------------------------------------------------------------------------------|-----------------|------------|-------------|
| Normal no complaints; no evidence of disease                                    | 100%            | 0          | Fully active, able to carry on all pre-disease performance without restriction |
| Able to carry on normal activity; minor signs or symptoms of disease            | 90%             | 1          | Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work |
| Normal activity with effort; some signs or symptoms                             | 80%             | 1          | Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work |
| Care for self. Unable to carry on normal activity or to do active work          | 70%             | 2          | Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours |
| Requires occasional assistance, but able to care for most of his needs         | 60%             | 2          | Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours |
| Requires considerable assistance and frequent medical care                      | 50%             | 3          | Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours |
| Disabled. Requires special care and assistance                                  | 40%             | 3          | Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours |
| Severely disabled. Hospitalization indicated though death non-imminent         | 30%             | 4          | Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair |
| Very sick. Hospitalization necessary. Active supportive treatment necessary     | 20%             | 4          | Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair |
| Moribund                                                                        | 10%             | 4          | Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair |
**Table S3: Imputation Rules for Partial or Missing Start Dates**

| Start Date | Complete: **yyyymmdd** | Partial: **yyyyymm** | Partial: **yyyy** | Missing |
|------------|------------------------|---------------------|------------------|---------|
|            | < Study Day 0          | ≥ Study Day 0       | < Study Day 0    | ≥ Study Day 0 |
|            | yyyymmdd               | yyyymm              | yyyy             | yyyy    |
| Partial   |                        |                     |                  |         |
| yyyymm     | 1                      | 2                   | 1                | 1       |
| ≠ Study    |                        |                     | n/a              |         |
| Day 0      | 2                      | 2                   | 2                | 2       |
| Partial   |                        |                     |                  |         |
| yyyy      | 1                      | 3                   | 1                | 1       |
| ≠ Study    |                        |                     | n/a              |         |
| Day 0      | 3                      | 3                   | 3                | 3       |
| Missing    | 4                      | 1                   | 4                | 1       |

1 = impute the date of study day 0
2 = impute the first of the month
3 = impute January 1 of the year
4 = impute January 1 of the stop year
Note: if the start date imputation leads to a start date that is after the stop date, then the start date was not imputed.

Study Day 0 was defined as the day the patient received the first axicabtagene ciloleucel infusion for IAS analysis set. For FAS this was defined as the start date of the corresponding line of treatment.
Table S4: Data values for tumour bulk, response to last line of therapy, and time since last treatment, and relapsed/refractory status before and after multiple imputation

|                          | Before multiple imputation | After multiple imputation |
|--------------------------|----------------------------|----------------------------|
|                          | SCHOLAR-5 (n=143) | ZUMA-5 (n=86) | SCHOLAR-5 (n=143) | ZUMA-5 (n=86) |
| **Tumor bulk**           |                          |                          |                          |                          |
| Mean                     | 4.97                      | 5.20                      | 4.91                      | 5.2                      |
| SD                       | 2.67                      | 2.94                      | 2.69                      | 2.94                      |
| Median                   | 4.00                      | 4.35                      | 4.16                      | 4.35                      |
| Q1                       | 2.75                      | 3.27                      | 2.75                      | 3.27                      |
| Q3                       | 6.55                      | 6.43                      | 6.5                       | 6.43                      |
| Min                      | 1.40                      | 1.60                      | 1.4                       | 1.6                       |
| Max                      | 14.00                     | 16.70                     | 14.54                     | 16.7                      |
| N                        | 143                       | 86                        | 143                       | 86                        |
| Missing                  | 52 (36.4%)                | 0                         | 0                         | 0                         |
| **Time since last treatment (months)** |                          |                          |                          |                          |
| Mean                     | 18.46                     | 8.44                      | 18.25                     | 8.44                      |
| SD                       | 27.87                     | 11.68                     | 27.73                     | 11.68                     |
| Median                   | 6.97                      | 3.53                      | 6.76                      | 3.53                      |
| Q1                       | 1.18                      | 1.77                      | 1.16                      | 1.77                      |
| Q3                       | 22.93                     | 9.01                      | 22.66                     | 9.01                      |
| Min                      | 1.18                      | 0                         | 0                         | 0                         |
| Max                      | 172.93                    | 65.84                     | 172.93                    | 65.84                     |
| N                        | 143                       | 86                        | 143                       | 86                        |
| Missing                  | 2 (1.4%)                  | 0                         | 0                         | 0                         |
| **Response to prior line of therapy** |                          |                          |                          |                          |
| Complete response         | 41 (28.7%)                | 21 (28.4%)                | 41 (28.7%)                | 23.01 (26.8%)             |
| Partial response          | 49 (34.3%)                | 16 (21.6%)                | 49 (34.3%)                | 19.34 (22.5%)             |
| Stable disease            | 22 (15.4%)                | 21 (28.4%)                | 22 (15.4%)                | 24.15 (28.1%)             |
| Progressive disease       | 31 (21.7%)                | 16 (21.6%)                | 31 (21.7%)                | 19.50 (22.7%)             |
| Missing                  | 0 (0%)                    | 12 (14.0%)                | 0 (0%)                    | 0 (0%)                    |
| **Relapsed/refractory status** |                          |                          |                          |                          |
| Relapsed                 | 56 (39.7%)                | 23 (26.7%)                | 56 (39.4%)                | 23 (26.7%)                |
| Refractory               | 85 (60.3%)                | 63 (73.3%)                | 87 (60.6%)                | 63 (73.3%)                |
| Missing                  | 2 (1.4%)                  | 0 (0%)                    | 0 (0%)                    | 0 (0%)                    |
Table S5: Demographic and clinical characteristics.

|                      | ZUMA-5 | SCHOLAR-5 | SMD | Weighted SCHOLAR-5 | Weighted SMD | Weighted SCHOLAR-5 Excluding DELTA | Weighted SMD Excluding DELTA |
|----------------------|--------|-----------|-----|--------------------|--------------|----------------------------------|-------------------------------|
| Median age (range) * | 62 (34–79) | 64 (36-89) | 0.32 | 61.2 (36-89) | 0.036 | 62 (36-89) | \                          |
| Sex – no. (%)        |        |           |     |                    |              |                                  |                               |
| Male                 | 48 (55.8) | 81 (56.6) | 0.017 | 53 (61.9) | 0.123 | 32 (57.6) | 0.035                       |
| Female               | 38 (44.2) | 62 (43.4) | 32 (38.1) | 24 (42.4) | \                        | \                         |                               |
| Race – no. (%)       |        |           |     |                    |              |                                  |                               |
| Asian                | 2 (2.3) | 4 (4.7) | 0.18 | 1 (2.5) | 0.049 | 1 (3.6) | 0.28                       |
| Black or African American | 3 (3.5) | 1 (1.2) | 2 (2.9) | 2 (5.8) | \                        | \                         |                               |
| White                | 80 (93.0) | 78 (91.8) | 54 (93) | 27 (90.6) | \                        | \                         |                               |
| Other                | 1 (1.2) | 2 (2.4) | 1 (1.6) | 0 (0) | \                        | \                         |                               |
| Missing              | 0 | 58 | 27 (31.5) | 27 | \                        | \                         |                               |
| Region – no. (%)     |        |           |     |                    |              |                                  |                               |
| US                   | 82 (95.3) | 63 (44.1) | 1.35 | 41 (48.6) | 1.22 | 21 (37.6) | 1.55                       |
| Europe               | 4 (4.7) | 80 (55.9) | 44 (51.4) | 35 (62.4) | \                        | \                         |                               |
| Median size of largest nodal mass (IQR) – cm * | 4.4 (3.3 – 6.4) | 4.2 (2.8 – 6.5) | 0.102 | 4.0 (2.9 – 6.3) | 0.094 | \                        | \                         |                               |
| Follicular lymphoma subtype – no. (%) |        |           |     |                    |              |                                  |                               |
| Grade 1              | 20 (23.3) | 56 (42.4) | 0.42 | 30 (37.3) | 0.54 | 27 (51.3) | 0.511                    |
| Grade 2              | 43 (50.0) | 61 (46.2) | 42 (52.6) | 21 (40.7) | \                        | \                         |                               |
| Grade 3a             | 23 (26.7) | 15 (11.4) | 8 (10.1) | 4 (8.1) | \                        | \                         |                               |
| Missing              | 0 | 11 | 5 | 3 (6.1) | \                        | \                         |                               |
| Disease stage – no. (%) |        |           |     |                    |              |                                  |                               |
| I                    | 2 (2.3) | 4 (6.2) | 1.13 | 1 (4.6) | 1.29 | 1 (6) | 1.303                    |
| II                   | 9 (10.5) | 2 (3.1) | 0 (1.3) | 0 (1.7) | \                        | \                         |                               |
| III                  | 35 (40.7) | 17 (26.2) | 8 (27.0) | 6 (25) | \                        | \                         |                               |
| IV                   | 40 (46.5) | 42 (64.6) | 20 (67.1) | 16 (67.3) | \                        | \                         |                               |
| Missing              | 0 | 78 | 55 (64.9) | 33 | \                        | \                         |                               |
| Number of nodal sites – no. (%) |        |           |     |                    |              |                                  |                               |
| 1                    | 16 (22.5) | 14 (15.1) | 0.14 | 8 (14.1) | 0.12 | 5 (15.8) | 0.619                    |
| 2                    | 12 (16.9) | 17 (18.3) | 13 (21.5) | 5 (15.8) | \                        | \                         |                               |
| 3                    | 7 (9.9) | 9 (9.7) | 7 (10.9) | 1 (1.8) | \                        | \                         |                               |
| >4                   | 36 (50.7) | 53 (57.0) | 32 (53.6) | 21 (66.6) | \                        | \                         |                               |
| Missing              | 15 | 50 | 25 (29.4) | 25 (43.9) | \                        | \                         |                               |
| Bone marrow involvement – no. (%) |        |           |     |                    |              |                                  |                               |
| Yes                  | 20 (23.3) | 23 (43.4) | 0.44 | 14 (48.1) | 0.54 | 11 (49.4) | 0.565                    |
| No                   | 66 (76.7) | 30 (56.6) | 15 (51.9) | 11 (50.6) | \                        | \                         |                               |
| Missing              | 0 | 90 | 56 (65.9) | 34 (60.4) | \                        | \                         |                               |
| GELF high tumor burden – no. (%) |        |           |     |                    |              |                                  |                               |
| Yes                  | 44 (51.2) | 3 (17.6) | 0.75 | 1 (12.2) | 0.92 | 1 (12.2) | \                        |
| No                   | 42 (48.8) | 14 (82.4) | 7 (87.8) | 7 (87.8) | \                        | \                         |                               |
| Missing              | 0 | 126 | 77 | 48 | \                        | \                         |                               |
| FLIPI – no. (%)      |        |           |     |                    |              |                                  |                               |
| 0                    | 3 (3.5) | 2 (4) | 0 (0.4) | 0 (0.6) | 1 | |
| 1                    | 10 (11.6) | 4 (8) | 2 (9.5) | 2 (13.5) | \                        | \                         |                               |
| 2                    | 33 (38.4) | 11 (22) | 4 (17.4) | 2 (13) | \                        | \                         |                               |
| 3                    | 25 (29.1) | 19 (38) | 7 (32.8) | 7 (46.6) | \                        | \                         |                               |
| 4                    | 12 (14.0) | 10 (20) | 6 (28.1) | 2 (9.6) | \                        | \                         |                               |
|                          | ZUMA-5 | SCHOLAR-5 | SMD | Weighted SCHOLAR-5 | Weighted SMD | Weighted SCHOLAR-5 Excluding DELTA | Weighted SMD Excluding DELTA |
|--------------------------|--------|-----------|-----|--------------------|--------------|----------------------------------|-----------------------------|
| Median number of prior lines of therapy (range)* | 3 (2-9) | 2 (2-8) | 0.53 | 3 (2-8) | 0.047 | 3 (2-7) | 0.369 |
| Median time since last treatment (IQR) – months * | 3.5 (1.8 – 9.0) | 6.8 (1.2 – 22.7) | 0.46 | 2.3 (0.7-8.0) | 0.056 |
| Response to prior line of therapy – no. (%) * | Complete response | 23 (26.8) | 41 (28.7) | 0.50 | 19 (22.8) | 0.073 | 17 (31.2) | 0.145 |
| Partial response | 19.34 (22.5) | 49 (34.3) | 19 (22.4) | 10 (18.7) |
| Stable disease | 24.15 (28.1) | 22 (15.4) | 26 (31.2) | 14 (25.0) |
| Progressive disease | 19.5 (22.7) | 31 (21.7) | 20 (23.5) | 14 (25.1) |
| Relapsed/refractory to prior line of therapy – no. (%)* | Relapsed | 23 (26.7) | 56 (39.4) | 0.27 | 20 (23.4) | 0.077 | 20 (35.4) | 0.189 |
| Refractory | 63 (73.3) | 87 (60.6) | 65 (76.6) | 36 (64.6) |
| POD24 – no. (%)* | Yes | 49 (57.0) | 51 (35.7) | 0.44 | 47 (55.9) | 0.022 | 28 (49.5) | 0.15 |
| No | 37 (43.0) | 92 (64.3) | 37 (44.1) | 28 (50.5) |
| Prior stem cell transplant – no. (%)* | Yes | 21 (24.4) | 31 (21.7) | 0.065 | 24 (28.0) | 0.08 | 17 (29.9) | 0.123 |
| No | 65 (75.6) | 112 (78.3) | 61 (72.0) | 39 (70.1) |
| Prior anti-CD20 + alkylator combination therapy * | Yes | 86 (100) | 128 (89.5) | 0.48 | 85 (100) | 0 | 56 (100) | 0 |
| No | 0 | 15 (10.5) | 0 | 0 |
| Prior stem cell transplant – no. (%) | Autologous | 21 (24.4) | 28 (19.6) | 0.20 | 20 (23.6) | 0.34 | 13 (23.2%) | 0.447 |
| Allogenic | 0 (0) | 2 (1.4) | 3 (3.9) | 3 (5.9%) |
| Other | 0 (0) | 1 (0.7) | 0 (0.6) | 0 (0.8%) |
| None | 65 (75.6) | 112 (78.3) | 61 (72.0) | 39 (70.1%) |
| Median time since diagnosis (IQR) – months | 59.86 (35.1 – 96.6) | 84.79 (53.0.8.1 – 130.5) | 0.39 | 64.6 (41.0 – 115.8) | 0.10 | 74.5 (47.0 – 126.5) |
| ECOG – no. (%) | 0 | 51 (59.3) | 39 (33.1) | 0.55 | 21 (29.0) | 0.64 | 13 (30.7%) | 0.6 |
| 1 | 35 (40.7) | 79 (66.9) | 51 (71.0) | 30 (69.3%) |
| Missing | 0 | 25 | 13 | 12 (22.1%) |

* Variables used in propensity score weighting. Variables are shown for SCHOLAR-5 before and after propensity score weighting. The standardized mean difference between SCHOLAR-5 and ZUMA-5 patients before and after is shown.
Table S6: Randomly selected index treatment patterns for the final SCHOLAR-5 sample for comparative effectiveness before and after weighting

| Treatment                                                        | Before weighting | After weighting |
|------------------------------------------------------------------|------------------|-----------------|
|                                                                  | Overall | Sub-cohort A | Sub-cohort B | Overall | Sub-cohort A | Sub-cohort B |
| Alkylating chemotherapy                                           | 1       | 0             | 1             | 3.0     | 0             | 3.0           |
| Allogeneic stem cell transplant                                   | 4       | 3             | 1             | 3.2     | 2.6           | 0.6           |
| Autologous stem cell transplant                                   | 12      | 10            | 2             | 10.1    | 7.2           | 2.9           |
| Bruton Tyrosine Kinase Inhibitor (BTKi)                          | 1       | 0             | 1             | 0       | 0             | 0             |
| Anti-CD20 agent + alkylation chemotherapy                        | 5       | 0             | 5             | 7.3     | 0             | 7.3           |
| Anti-CD20 agent + bendamustine                                   | 24      | 24            | 0             | 7.9     | 7.9           | 0             |
| Anti-CD20 agent + CHOP-like chemotherapy                         | 6       | 6             | 0             | 3.1     | 3.1           | 0             |
| Anti-CD20 agent + CVP                                            | 1       | 1             | 0             | 0       | 0             | 0             |
| Anti-CD20 agent + fludarabine-based chemotherapy                 | 3       | 2             | 1             | 1.2     | 0.9           | 0.3           |
| Anti-CD20 agent + immunomodulatory agent                         | 1       | 0             | 1             | 1.1     | 0             | 1.1           |
| Anti-CD20 agent + platinum-based chemotherapy                    | 1       | 0             | 1             | 1.6     | 0             | 1.6           |
| Anti-CD20 agent + other chemotherapy*                            | 12      | 11            | 1             | 6.9     | 6.5           | 0.4           |
| Chemotherapy **                                                   | 12      | 10            | 2             | 5.7     | 4.0           | 1.7           |
| Experimental therapy ***                                         | 23      | 19            | 4             | 13.5    | 12.6          | 0.9           |
| Fludarabine                                                      | 1       | 0             | 1             | 0.5     | 0             | 0.5           |
| Immunomodulatory agent                                           | 18      | 17            | 1             | 9.9     | 7.0           | 3.0           |
| Phosphoinositide 3-kinase inhibitor (PI3Ki)                       | 14      | 14            | 0             | 4.3     | 4.3           | 0             |
| Radioimmunotherapy                                               | 2       | 1             | 1             | 2.1     | 0.2           | 1.9           |
| Stem cell transplant (other)                                     | 2       | 0             | 2             | 3.6     | 0             | 3.6           |
| **Total:**                                                       | 143     | 118           | 25            | 85      | 56.3          | 28.8          |

Note. Numbers may not sum to the overall sample due to rounding.

* Includes: R-IE (rituximab, ifosfamide, etoposide), R-GEMOX (rituximab, Gemcitabine, oxaliplatin), R-IFM VP16+ (rituximab, ifosfamide, etoposide), R-ICE (rituximab ifosfamide, carboplatin, etoposide phosphate), Rituximab + cyclophosphamide + prednisone, R-CVP (rituximab, cyclophosphamide, vincristine sulfate, prednisone), R-DHAP (rituximab, dexamethasone, cytarabine, cisplatin)

** Includes: Bendamustine, Chlorambucil, Fludarabine combination, Mini-DHAOX (dexamethasone, cytarabine and oxaliplatin), ICE (ifosfamide, carboplatin, etoposide phosphate), Ibrutinib, CHOP and mini-CHOP (cyclophosphamide, hydroxydaunorubicin, oncovin, prednisone), Gemcitabine + vinorelbine, cyclophosphamide, Fludarabine endoxa

*** Includes: Anti PD1, Obinutuzumab + atezolizumab + venetoclax, Ibrutinib (trial), anti-CD137 monoclonal antibody, Nivolumab, Obinutuzumab + ONO + entospletinib, Dual-affinity re-targeting antibody, Anti CD20 + anti PD1, Immunotoxin, SYK/JAG inhibitor, BET inhibitor, HDAC inhibitor, BCL2 inhibitor, Cerdulatinib, Bispecific CD20 CD3, Rituximab+investigational drug, Anti-CD79B ADC plus rituximab, Iazomib, R-HOLOXAB VP16
Table S7: Randomly selected index treatment patterns for the final SCHOLAR-5 sample for comparative effectiveness before and after weighting by continent.

| Treatment                                      | US   | US weighted | EU | EU weighted |
|------------------------------------------------|------|-------------|----|-------------|
| Alkylating chemotherapy                        | 0    | 0           | 1  | 3.0         |
| Allogeneic stem cell transplant                | 1    | 0.6         | 3  | 2.6         |
| Autologous stem cell transplant                | 2    | 2.9         | 10 | 7.2         |
| Bruton Tyrosine Kinase Inhibitor (BTKi)        | 1    | 0           | 0  | 0           |
| Anti-CD20 agent + alkylating chemotherapy      | 5    | 7.3         | 0  | 0           |
| Anti-CD20 agent + bendamustine                 | 4    | 1.4         | 20 | 6.4         |
| Anti-CD20 agent + CHOP-like chemotherapy       | 1    | 0           | 5  | 3.1         |
| Anti-CD20 agent + CVP                          | 0    | 0           | 1  | 0           |
| Anti-CD20 agent + fludarabine-based chemotherapy| 1   | 0.3         | 2  | 0.9         |
| Anti-CD20 agent + immunomodulatory agent       | 1    | 1.1         | 0  | 0           |
| Anti-CD20 agent + platinum-based chemotherapy  | 1    | 1.6         | 0  | 0           |
| Anti-CD20 agent + other chemotherapy           | 9    | 5.5         | 3  | 1.4         |
| Other chemotherapy                             | 5    | 2.1         | 7  | 3.6         |
| Experimental therapy                           | 15   | 10.8        | 8  | 2.7         |
| Fludarabine                                     | 0    | 0           | 1  | 0.5         |
| Immunomodulatory agent                         | 8    | 2.3         | 10 | 7.6         |
| Phosphoinositide 3-kinase inhibitor (PI3Ki)     | 7    | 2.0         | 7  | 2.3         |
| Radioimmunotherapy                              | 1    | 1.9         | 1  | 0.2         |
| Stem cell transplant (other)                    | 1    | 1.5         | 1  | 2.2         |
Table S8: Estimated proportions of overall response rate and complete response among the sensitivity analyses

|                     | SCHOLAR-5 | ZUMA-5 |
|---------------------|-----------|--------|
| Overall response rate |           |        |
| Primary analysis     | 42 (49.9%)| 81 (94.2%) |
| Safety cohort        | 68 (54.2%)| 117 (94.4%) |
| Real-world data only | 42 (50.8%)| 81 (94.2%)* |
| Without MI           | 41 (46.8%)| 81 (94.2%)* |
| Modified EAS         | 48 (57%)  | 81 (94.2%)* |
| PS matching          | 39 (58.7%)| 62 (94.2%) |
| Simplified PS        | 40 (46.4%)| 81 (94.2%) |
| Complete response    |           |        |
| Primary analysis     | 25 (29.9%)| 68 (79.1%) |
| Safety cohort        | 44 (35.4%)| 98 (79.0%) |
| Real-world data only | 28 (33.4%)| 68 (79.1%)* |
| Without MI           | 24 (27.5%)| 68 (79.1%)* |
| Modified EAS         | 29 (34.3%)| 68 (79.1%)* |
| PS matching          | 24 (36.2%)| 51.79 (78.2%) |
| Simplified PS        | 21 (24.9%)| 68 (79.1%) |

*Sensitivity analysis only changed SCHOLAR-5 sample, so ZUMA-5 result unchanged from primary analysis. Safety cohort: Including all ZUMA-5 from safety analysis set. Real-world data only: Excluding participants from the DELTA trial. Without MI: Without multiple imputation of missing data. Modified EAS: excluding SCHOLAR-5 patients with missing ECOG score and being within the age range of ZUMA-5 patients. PS matching: Using propensity score matching. See Figure S2 for odds ratios.
Table S9: Estimated proportions of overall response rate and complete response among the subgroup analyses

|                        | SCHOLAR-5 | ZUMA-5 |
|------------------------|-----------|--------|
| **Overall response rate** |           |        |
| Primary analysis       | 42 (49.9%)| 81 (94.2%)|
| POD24                  | 42 (50.8%)| 81 (94.2%)|
| Refractory             | 26 (42.3%)| 59 (93.7%)|
| Prior SCT              | 6 (26.5%) | 20 (95.2%)|
| ≥3 prior LoT           | 24 (40.3%)| 57 (95.0%)|
| **Complete response**  |           |        |
| Primary analysis       | 25 (29.9%)| 68 (79.1%)|
| POD24                  | 28 (33.4%)| 68 (79.1%)|
| Refractory             | 13 (20.7%)| 50 (79.4%)|
| Prior SCT              | 6 (26.5%) | 18 (85.7%)|
| ≥3 prior LoT           | 12 (20.6%)| 48 (80.0%)|

POD24, progression of disease with 24 months of first-line therapy; SCT, stem-cell transplants; LoT, line of therapy. See Figure S3 for odds ratios.
Figure S1: Results from sensitivity analysis

Safety cohort: Including all ZUMA-5 patients from safety analysis set. Real-world data only: Excluding participants from the DELTA trial. Without MI: Without multiple imputation procedures of missing data. Modified EAS: excluding SCHOLAR-5 patients with missing ECOG score and being within the age range of ZUMA-5 patients. PS matching: Using propensity score matching rather than weighting. Simplified PS: including only variables that were pre-specified as high priority.
Figure S2: Results from subgroup analyses

POD24, progression of disease with 24 months of first-line therapy; SCT, stem-cell transplants; LoT, line of therapy.
Figure S3: Duration of response Kaplan-Meier curve in the 3L+ population
Figure S4: Kaplan-Meier plots comparing ZUMA-5 to SCHOLAR-5 among patients that are 4th line of therapy or higher for A. Progression-free survival; B. Overall survival; C. Time-to-next-treatments; and D. Duration of response.
Figure S5: Kaplan-Meier plots comparing ZUMA-5 to SCHOLAR-5 among POD24 patients for A. Progression-free survival; B. Overall survival; C. Time-to-next-treatments; and D. Duration of response
Figure S6: Kaplan-Meier plots comparing ZUMA-5 to SCHOLAR-5 among patients with prior exposure to stem cell transplants for A. Progression-free survival; B. Overall survival; C. Time-to-next-treatments; and D. Duration of response
Figure S7: Flow diagram of study selection from Memorial Sloan Kettering (MSK) Cancer Center

1100 iNHL patients in MSK database

880 patients without 3rd LoT removed*

220 patients with 3rd LoT

166 excluded:
• 131 received ≥ 3rd LoT prior to July 23rd 2014
• 35 transformed prior to 3rd LoT, or prior to July 23rd 2014

54 modern patients

23 removed for not meeting eligibility criteria†

31 eligible patients

*Prior line of therapy with anti-CD20 monotherapy did not count as line of therapy for eligibility.
† Eligibility criteria:
• Patients aged ≥18 years;
• Patients with histologically confirmed diagnosis of iNHL, with histological subtype limited to FL Grade 1, Grade 2, or Grade 3a based on criteria established by the World Health Organization (WHO) 2016 classification;
• Patients with r/r disease (i.e., r/r iNHL)
• Transformed FL excluded;
• FL Histological Grade 3b excluded;
• Excluding prior anti-CD19 CAR T-cell therapy or other genetically modified T-cell therapy;
• Eligible within 12 months before the last updated version of the database (site specific)