GENE THERAPY

Comparative effectiveness of ZUMA-5 (axi-cel) vs SCHOLAR-5 external control in relapsed/refractory follicular lymphoma

Paola Ghione,[1,2,]* M. Lia Palomba,[2,]* Anik R. Patel,[3] Sabela Bobillo,[4] Kevin Deighton,[5] Caron A. Jacobson,[6] Myrna Nahas,[3] Anthony J. Hatswell,[5] A. Scott Jung,[3] Steve Kanters,[7] Julia Thornton Snider,[3] Sattva S. Neelapu,[8] Maria Teresa Ribeiro,[9] M. Alan Brookhart,[10,11] Herve Ghesquieres,[12] John Radford,[13] and John G. Gribben[14]

[1]Roswell Park Comprehensive Cancer Center, Buffalo, NY; [2]Memorial Sloan-Kettering Cancer Center, New York, NY; [3]Kite, A Gilead Company, Santa Monica, CA; [4]Department of Haematology, Vall D'Hebron Institute of Oncology, University Hospital Vall d'Hebron, Barcelona, Spain; [5]Delta Hat, Nottingham, United Kingdom; [6]Dana-Farber Cancer Institute, Boston, MA; [7]RainCity Analytics, Vancouver, BC, Canada; [8]Division of Cancer Medicine, Department of Lymphoma and Myeloma,The University of Texas MD Anderson Cancer Center, Houston, TX; [9]Portuguese Oncology Institute of Porto, Porto, Portugal; [10]Target Real World Evidence (RWE), Durham, NC; [11]Department of Population Health Sciences, Duke University, Durham, NC; [12]Hopital Lyon Sud, Lyon, France; [13]The Christie National Health Service (NHS) Foundation Trust and University of Manchester, Division of Cancer Sciences, Manchester, United Kingdom; and [14]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom

KEY POINTS

- There is a high degree of variability and suboptimal response to available treatments for relapsed or refractory follicular lymphoma.

- ZUMA-5 (axicabtagene ciloleucel) demonstrated substantially improved clinical outcomes relative to these existing therapeutic options.

In the pivotal ZUMA-5 trial, axicabtagene ciloleucel (axi-cel; an autologous anti-CD19 chimeric antigen receptor T-cell therapy) demonstrated high rates of durable response in relapsed/refractory (r/r) follicular lymphoma (FL) patients. Here, outcomes from ZUMA-5 are compared with the international SCHOLAR-5 cohort, which applied key ZUMA-5 trial eligibility criteria simulating randomized controlled trial conditions. SCHOLAR-5 data were extracted from institutions in 5 countries, and from 1 historical clinical trial, for r/r FL patients who initiated a third or higher line of therapy after July 2014. Patient characteristics were balanced through propensity scoring on prespecified prognostic factors using standardized mortality ratio (SMR) weighting. Time-to-event outcomes were evaluated using weighted Kaplan-Meier analysis. Overall response rate (ORR) and complete response (CR) rate were compared using weighted odds ratios. The 143 ScHOLAR-5 patients reduced to an effective sample of 85 patients after SMR weighting vs 86 patients in ZUMA-5. Median follow-up time was 25.4 and 23.3 months for SCHOLAR-5 and ZUMA-5. Median overall survival (OS) and progression-free survival (PFS) in SCHOLAR-5 were 59.8 months and 12.7 months and not reached in ZUMA-5. Hazard ratios for OS and PFS were 0.42 (95% confidence interval [CI], 0.21-0.83) and 0.30 (95% CI, 0.18-0.49). The ORR and CR rate were 49.9% and 29.9% in SCHOLAR-5 and 94.2% and 79.1% in ZUMA-5, for odds ratios of 16.2 (95% CI, 5.6-46.9) and 8.9 (95% CI, 4.3-18.3). Compared with available therapies, axi-cel demonstrated an improvement in meaningful clinical endpoints, suggesting axi-cel addresses an important unmet need for r/r FL patients. This trial was registered at www.clinicaltrials.gov as #NCT03105336.

Introduction

Follicular lymphoma (FL) is the most common indolent non-Hodgkin lymphoma (NHL), representing 17% to 35% of all NHL cases in the United States and Europe.[1,2] Relapsed follicular lymphoma is generally considered incurable[3]; however, CD19-directed chimeric antigen receptor (CAR) T-cell therapy is a potentially curative option for other indications,[4,5] and its effectiveness needs to be explored in FL. In the majority of patients, FL relapses multiple times with a pattern of decreased durability of remission with each subsequent line of therapy (LoT).[6,7] Moreover, a subset of FL patients with high-risk disease defined as those whose disease progresses within 24 months (POD24) of

frontline chemoimmunotherapy have a strikingly worse prognosis, with 5-year overall survival (OS) rates estimated between 26% and 62%,[1,8-11] even after high-dose chemotherapy. Despite advancements in the therapeutic approach to FL, there remains an unmet need for therapies that can generate durable responses, particularly in heavily pretreated and high-risk patients.

There lacks consensus in treatment guidelines for FL, resulting in substantial variability in therapeutic regimens for relapsed/refractory (r/r) FL patients. Moreover, r/r FL patients are heterogeneous in terms of response and duration of response to available therapies, functional status, and prognostic risk factors.

Limited data have been published on the outcomes of patients treated with multiple LoTs, with reported median progression-free survival (PFS) to second-line and third-line of ∼18 and 12 months, respectively.[6,7] A recent systematic literature review confirmed that median OS and PFS decreased with each passing LoT.[12] Both treatment variability and paucity of data in the real-world setting may lead to potential bias and the inability to demonstrate the extent of clinical benefit of therapies reported in r/r FL clinical trials. This variability complicates randomized clinical trial (RCT) design in multiply relapsed FL as there is a lack of uniformity in standard-of-care (SOC) therapies and heterogeneity among patients regarding prior therapies used.

Axicabtagene ciloleucel (axi-cel), a novel CAR T-cell therapy, was approved in March 2021 in the United States for patients with r/r FL following at least 2 prior LoT. Approval was based on ZUMA-5, a pivotal single-arm phase 2 trial reporting a 94% overall response rate (ORR) in patients with r/r FL.[13] ZUMA-5 has demonstrated striking ORR, complete response (CR), OS, and progression-free survival (PFS) within r/r FL patients, but it does not provide comparative measures with other available treatments. To contextualize the clinical benefit associated with axi-cel in r/r FL, we developed an international, retrospective, observational cohort from real-world clinical sites, named SCHOLAR-5. SCHOLAR-5 is one of the few multicountry cohorts of r/r FL patients and was designed to recapitulate the population enrolled in ZUMA-5.[14] In this study, we report a comparative effectiveness analysis between ZUMA-5 and SCHOLAR-5 in patients with r/r FL.

Methods

Design and setting

Data sources for SCHOLAR-5 included patient records from 7 institutions in 5 countries, extracted from 2014 to 2020 (subcohort A), pooled with posttrial data of select patients from the pivotal phase 2 idelalisib trial named DELTA (subcohort B). Posttrial data from subcohort B were more representative of the heterogeneous mix of treatment options available to multiply relapsed FL patients and therefore more appropriate for the comparison. Description of the real-world clinical sites and the DELTA clinical trial can be found in the supplemental Appendix, section 1. Institutional review board approval for the study was obtained separately for each participating site.

To meet eligibility for SCHOLAR-5, patients had to be aged ≥18 years with r/r FL grade 1, 2, or 3a initiating third-line or higher therapy (≥2 prior LoT). Prior LoT with anti-CD20 or radiotherapy monotherapy or surgery alone did not count as LoT for eligibility. Additional eligibility criteria were then applied at LoT level in the analysis stage to ensure comparability to ZUMA-5 (Figure 1). These criteria were: Eastern Cooperative Oncology Group (ECOG) performance status ≤1, no evidence of transformation to diffuse large B-cell lymphoma (DLBCL) or of FL grade 3b histology, no prior anti-CD19 CAR T-cell therapy or genetically modified therapy, index date on July 2014 or later, and index date at least 12 months before the database cutoff date.

Data sources

Subcohort A data were obtained from electronic medical records supplemented by manual extraction from patient charts and imaging reports as needed. Real-world sites were required to identify at least 10 eligible patients along with other feasibility criteria including data availability and completeness. Moreover, to minimize unobserved confounding related to providers and treatment centers, sites were limited to those that were similar to the ZUMA-5 trial sites to attempt to simulate RCT conditions. For subcohort B, the first subsequent LoT was included in comparative analyses to avoid overrepresentation of treatment with PI3K inhibitors in the comparative analysis (Figure 1). These data were collected using a case report form designed for the DELTA trial described in greater detail elsewhere.[15] Details of the ZUMA-5 trial cohort are reported elsewhere.[16] Additional data source details can be found in the supplemental Appendix, section 1.

Variables assessed

Variables used to build the propensity score (PS) model (methods described below) included: POD24, number of prior LoT, relapsed vs refractory to last LoT, prior stem cell transplant, tumor bulk (diameter of largest lesion), time from last treatment, best response to previous line, age, and prior exposure to anti-CD20 alkylator combination therapy. Here, the propensity represents the probability of belonging to the ZUMA-5 cohort of patients (ie, the probability of being treated with axi-cel), and details on the modeling are provided further below. If ECOG score was missing but a Karnofsky's performance score was provided, then an ECOG score was derived from the Karnofsky's performance score. The handling of missing data or partial data included the use of multiple imputation. Multiple imputation was conducted in the assumption that missing values were missing at random. Further details on these variables and handling of missingness are described in the supplemental Appendix. The multiply imputed data were used in all analyses, with the exception of the sensitivity analysis, whereby the analyses were repeated without any multiple imputation to ensure that the imputation did not bias the results.

Patients treated with axi-cel in ZUMA-5 were compared with those treated with other available treatment options in SCHOLAR-5. These included approved and experimental drug therapies and autologous and allogenic transplant. In SCHOLAR-5, the index date was defined as the initiation date of the index LoT. For subcohort A, the index LoT was randomly selected from all eligible LoTs, which was selected as an unbiased approach, which would likely lead to good overlap with the ZUMA-5 LoT distribution.[17,18] Index date must have occurred after July 2014 to reduce time-period bias due to the introduction of PI3K inhibitors and because the Lugano criteria for disease assessment was formalized in 2014. As previously noted, the index LoT in subcohort B was the treatment received immediately after idelalisib and was agnostic to date. Further details on the index date selection method can be found in the supplemental Appendix.

For both subcohorts A and B, single-agent anti-CD20 therapy, surgery, and radiotherapy alone were ineligible index treatments and did not count as prior LoTs. CART-cell therapy or any other cellular therapy were ineligible index treatments, and patients were censored if they received these treatments during follow-up. The index treatment line selection period extended from July 2014 to site-specific dates of abstraction, with the latest date being December 2020 (specific dates presented in supplemental Appendix, section 5).
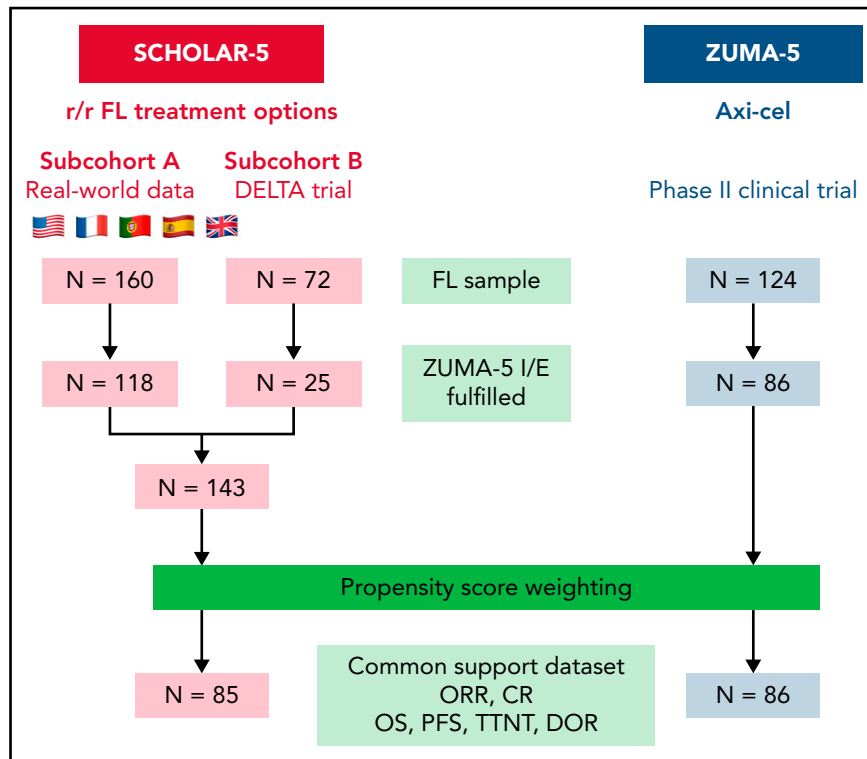
Figure 1. Flow diagram of patient enrolment, selection, and analysis of SCHOLAR-5 and comparative analysis set. Full eligibility criteria are listed in supplemental Appendix, section 2. Key eligibility criteria for SCHOLAR-5 cohort were: (1) diagnosed r/r FL; (2) starting third or higher line of therapy; (3) on or after 23 July 2014. Prior line of therapy with anti-CD20 monotherapy did not count as line of therapy for eligibility. Key exclusion criteria for the SCHOLAR-5 cohort were (1) transformed FL, (2) FL histological grade 3b, and (3) prior anti-CD19 CAR T-cell therapy or other genetically modified T-cell therapy. Patients were eligible for inclusion in the ZUMA-5 sample if they met the criteria for a minimum follow-up of 18 months.

Endpoints

Endpoints included PFS, OS, time to next treatment (TTNT) (included death as an event), ORR, and CR. Of note, method of disease response and progression assessment varied by cohort. In addition to Lugano criteria, response assessments in subcohorts A and B included computed tomography (CT) scans using older criteria. In ZUMA-5, tumor response and progression were evaluated using positron emission tomography (PET) diagnostic CT scans using Lugano criteria. For all effectiveness variables, partial dates (eg, when only month and year were available) in subcohort A were addressed as described in supplemental Appendix, section 5. Notably, the progression dates were not collected for the subsequent LoT in the DELTA trial; therefore, subcohort B was not included in the PFS analysis.

Statistical methods

When comparing ZUMA-5 and SCHOLAR-5, PS methods, specifically standardized mortality ratio (SMR) weighting, were applied to account for the imbalance of confounders. The composition of variables for conducting the PS methods were prespecified by the investigator team and external experts in an effort to include those factors assessed as most clinically relevant. The SMR weighting allowed for the creation of an external comparator arm with a distribution of covariates that resembled those in ZUMA-5. To assess the balance of covariates, the standardized mean differences were computed and required to be <0.1. Variable selection for the PS model was guided by investigators and clinical experts with a goal of minimizing imbalance in prognostically important covariates.[19] Decisions about model

specification were made without knowledge of how those decisions impacted effect estimates. A sensitivity analysis including only the 5 variables rated as high priority was also carried out. PS methods are described in further detail in the supplemental Appendix, with the resulting weights generated used in all analyses.

To enable inclusion of the most prognostic baseline variables in the PS specification, multiple imputation was applied for variables with missing data that were specified as part of the PS model (details in the supplemental Appendix, section 4). Key variables with <40% missing data in either dataset were eligible for multiple imputation.[20] Imputation was chosen over complete case analysis to avoid selection bias.[21] Retrospectively collected data often have missing values and are likely missing at random. Imputation met the dual need of including the largest number of prespecified prognostic factors and retaining a meaningful sample of matched patients. Therefore, in order to ensure the robustness of our primary analytical approach, we included a sensitivity analysis with only complete cases for variables used in the PS model.

A 2-sided 95% confidence interval (CI) was used, and all tests were performed on the 5% α level (2-sided). Differences between ZUMA-5 and SCHOLAR-5 for continuous variables were assessed using weighted linear regression modeling, whereas categorical variables were compared using weighted logistic regression models. For time-to-event variables, the relative difference in hazard of the outcome between groups was estimated using a weighted Cox proportional hazards regression.

Additional subgroup and sensitivity analyses, including removal of the DELTA trial patients from SCHOLAR-5 and analyses without multiple imputation (ie, complete case analysis), were also conducted and described in supplemental Methods. For each subgroup and sensitivity analysis, propensity scoring methods were reapplied to ensure balance in covariates were maintained. Analyses were performed using R Software version 3.6.3 and the MatchIt, tidyverse, lubridate, survival, mice, coxme, sandwich, and boot packages.

Results

Characteristics of each cohort

The 143 r/r FL patients identified in SCHOLAR-5 reduced to an effective sample size of 85 patients after applying SMR weights. Eighty-six FL patients in ZUMA-5 with a minimum potential follow-up of 18 months were included in the primary comparative analyses. Median follow-up time after index treatment was 23.3 months for ZUMA-5 and 25.4 months for SCHOLAR-5 after SMR weighting. Table 1 lists the baseline characteristics of ZUMA-5 and SCHOLAR-5 before and after the SMR weighting. For a complete list of baseline variables, see supplemental Table 5. Notably, ZUMA-5 patients appeared to have a higher proportion of high-risk baseline characteristics than SCHOLAR-5, and these factors appeared to be balanced after the application of SMR weighting. Figure 2 shows the PS distributions before and after SMR weighting.

All PS variables were successfully balanced after SMR weighting (standardized mean difference <0.1) including POD24, number of prior LoT, relapsed vs refractory disease, prior stem cell transplant, size of largest node, response to prior LoT, time since last therapy, and age. We were unable to assess potential imbalance in variables including FLIPI and disease stage before the index due to the extent of missing data (Table 1).

The pattern of index treatments for the final SCHOLAR-5 cohort were heterogeneous and are presented for each subcohort in supplemental Table 6. We also stratified the index treatment patterns from United States and Europe in supplemental Table 7. The most common treatment regimen was an anti-CD20 agent combined with chemotherapy. Experimental treatments were also frequent as expected for patients that had received several previous LoT.

Time-to-event outcomes

PFS was substantially longer in ZUMA-5 (median, NR; 95% CI, 23.5-NE) compared with SCHOLAR-5 (median, 12.7 months; 95% CI, 6.2-14.7), leading to a hazard ratio (HR) of 0.30 (95% CI, 0.18-0.49) (Figure 3; Table 2). At 18 months, 68.8% (95% CI, 57.4-77.8) of patients in ZUMA-5 and 23.8% (95% CI, 11.0-36.5) of patients in SCHOLAR-5 had not progressed or died. Among patients with ≥3 prior LoT, the HR for median PFS was 0.20 (95% CI, 0.11-0.34). In this subgroup analysis, 18-month PFS in ZUMA-5 patients was 67.0 (95% CI, 52.7-77.8) and in SCHOLAR-5 patients was 12.7% (95% CI, 0.7-24.6). As progression dates were not collected for the DELTA trial patients, the effective sample size for SCHOLAR-5 was 56 for PFS. Although statistically significant differences in baseline characteristics did not emerge, the removal of the DELTA trial patients improved the values for several prognostic variables for the SCHOLAR-5

cohort: reduction in POD24, number of prior LoT, and patients who were refractory at index date. See supplemental Table 3 for patient characteristics in this restricted sample. Consequently, the patients in the SCHOLAR-5 sample excluding DELTA trial patients may be considered to have better prognostic factor distribution than the sample with DELTA patients included.

Median OS was not reached in ZUMA-5 (NR; 95% CI, 31.6-NE) compared with a median of 59.8 months in SCHOLAR-5 (95% CI, 21.9-NE), with a HR of 0.42 (95% CI, 0.21-0.83), a 58% reduction in the risk of death. At 18 months, 88.3% (95% CI, 79.4-93.5) of patients in ZUMA-5 and 67.1% (95% CI, 54.1-80.2) of patients in SCHOLAR-5 had survived. Notably among patients with ≥3 prior LoT, OS improvements were more pronounced (HR, 0.31; 95% CI, 0.15-0.66), with a 69% reduction in the risk of death, as was the 18-month survival rate (ZUMA-5: 88.3% [95% CI, 77.0-94.2]; SCHOLAR-5: 55.0% [95% CI, 39.6-70.3]). Findings were maintained across all prespecified sensitivity analyses, including with the DELTA trial cohort removed, (supplemental Figures 1 and 2) highlighting the robustness of the data. The PS–matched sensitivity analysis led to a reduced common support data set and improved PS alignment (Figure 2), in turn leading to an improved OS HR (HR, 0.39; 95% CI, 0.17-0.91). There were similar improvements in other sensitivity analyses with less conservative assumptions including complete case analysis (HR, 0.38; 95% CI, 0.19-0.73) and alternate specification of PS model (HR, 0.36; 95% CI, 0.20-0.68). These improvements were also seen in other endpoints including PFS and response rates (supplemental Figure 1).

Median TTNT was also not reached in ZUMA-5 (NR; 95% CI, NE-NE) compared with 23.4 months in SCHOLAR-5 (95% CI, 9.5-NE) (HR, 0.54; 95% CI, 0.32-0.89). Among patients with ≥3 prior LoT, median TTNT was again not reached in ZUMA-5 (NR; 95% CI, 22.8-NE) compared with 14.2 months in SCHOLAR-5 (95% CI, 5.8-NE) (HR, 0.49; 95% CI, 0.28-0.88). Assessment of TTNT could be considered more comparable between cohorts because date of initiation of next line is captured in routine clinical care, whereas scans are not routine, so follow up in the SCHOLAR5 study, and therefore time to progression, are differently assessed in ZUMA5 and SCHOLAR5. The TTNT results were consistent with the PFS and OS findings.

Response outcomes

ORR and CR were substantially higher in ZUMA-5 (ORR, 94%; CR, 79%) than SCHOLAR-5 (ORR, 50%; CR, 30%), with odds ratios of 16.2 (95% CI, 5.6-46.9) and 8.9 (95% CI, 4.3-18.3), respectively (Table 2). Findings were maintained in subgroup and sensitivity analyses, suggesting robust results. Notably, among patients with ≥3 prior LoT, ORR and CR improvements were greater, with odds ratios of 28.1 (95% CI, 7.4-107.3) and 15.4 (95% CI, 5.8-40.8), respectively. See supplemental Tables 8 and 9 and supplemental Figures 1 and 2 for further details.

Discussion

This study provides a comparative analysis between the ZUMA-5 and SCHOLAR-5 cohorts, where SCHOLAR-5 is an external, observational control dataset from 7 international cancer centers and the DELTA trial. With an unprecedented improvement in OS compared with available therapies for r/r FL, these

Table 1. Demographic and clinical characteristics of patients in the comparative analysis set

	SCHOLAR-5 before weighting (n = 143)	ZUMA-5 (n = 86)	Weighted SCHOLAR-5 (n = 85)	Weighted SMD
Median age* (range)	64 (36-89)	62 (34-79)	61 (36-89)	0.04
Male, n (%)	81 (56.6%)	48 (55.8%)	53 (61.9%)	0.12
Median size of largest nodal mass* (IQR), cm	4.2 (2.8-6.5)	4.4 (3.3-6.4)	4.0 (2.9-6.3)	0.09
FL subtype, n (%)				
Grade 1	56 (42.4)	20 (23.3)	30 (37.3)	0.54
Grade 2	61 (46.2)	43 (50)	42 (52.6)	
Grade 3a	15 (11.4)	23 (26.7)	8 (10.1)	
Missing	11	0	5	
Median number of prior lines of therapy (range)	2 (2-8)	3 (2-9)	3 (2-8)	0.05
Median time since last treatment* (IQR), mo	6.8 (1.2-22.7)	3.5 (1.8-9.0)	2.3 (0.7-8.0)	0.06
Response to prior line of therapy,* n (%)				
CR	41 (28.7)	23.01 (26.8)	19 (22.8)	0.07
PR	49 (34.3)	19.34 (22.5)	19 (22.4)	
SD	22 (15.4)	24.15 (28.1)	26 (31.2)	
PD	31 (21.7)	19.5 (22.7)	20 (23.5)	
Refractory to prior LoT,* n (%)	87 (60.6)	63 (73.3)	65 (76.6)	0.08
POD24*, n (%)	51 (35.7)	49 (57.0)	47 (55.9)	0.02
Prior stem cell transplant,* n (%)	31 (21.7)	21 (24.4)	24 (28.0)	0.08
Median time since diagnosis (IQR), mo	84.8 (53.0-130.5)	59.9 (35.-96.6)	64.6 (41.0-115.8)	0.10
Disease stage, n (%)				
I	4 (6.2)	2 (2.3)	1 (4.6)	NE
II	2 (3.1)	9 (10.5)	0 (1.3)	
III	17 (26.2)	35 (40.7)	8 (27.0)	
IV	42 (64.6)	40 (46.5)	20 (67.1)	
Missing	78	0	55	
Number of nodal sites, n (%)				
1	14 (15.1)	16 (22.5)	8 (14.1)	NE
2	17 (18.3)	12 (16.9)	13 (21.5)	
3	9 (9.7)	7 (9.9)	7 (10.9)	
>4	53 (57)	36 (50.7)	32 (53.6)	
Missing	50	15	25 (29.4)	
FLIPI, n (%)				
0	2 (4)	3 (3.5)	0 (0.4)	NE
1	4 (8)	10 (11.6)	2 (9.5)	
2	11 (22)	33 (38.4)	4 (17.4)	
3	19 (38)	25 (29.1)	7 (32.8)	
4	10 (20)	12 (14.0)	6 (28.1)	
5	4 (8)	3 (3.5)	3 (11.7)	
Missing	93	0	62	

FLIPI, Follicular Lymphoma International Prognostic Index; IQR, interquartile range; NE, not evaluable; PD, progressive disease; PR, partial response; SD, stable disease, SMD; standardized mean difference.

*Variables used in PS weighting. The SMD for disease stage, number of nodal sites, and FLIPI were not evaluable due to missing data. See supplemental Table 5 for the SMD values before weighting.
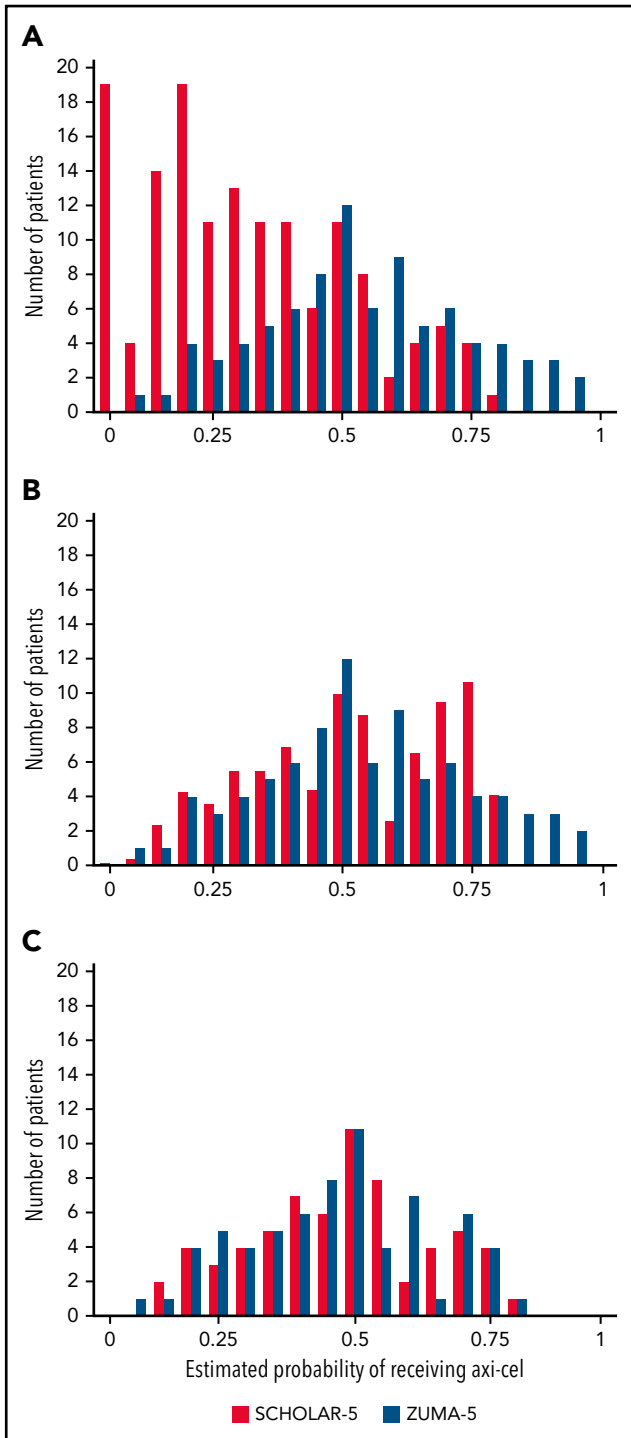
Figure 2. PS distribution before weighting, after weighting, and after matching. Panel A shows a PS distribution before weighting and shows trend for SCHOLAR-5 to have lower propensities and ZUMA-5 to have higher ones. Patients with near-zero PS were all in SCHOLAR-5, and patients with higher propensities were almost all in ZUMA-5. Panel B shows the results after application of SMR weighting, with comparable distributions. Panel C shows that in the matched analysis, patients with scores that were only present in 1 study (eg, scores >0.80 in ZUMA-5) were removed, leading to a better PS overlap but reduced common support data set.

SCHOLAR-5 results highlights the durable treatment effect of axi-cel. Moreover, axi-cel demonstrated superiority in additional clinically meaningful endpoints of PFS, TTNT, ORR, and CR. In
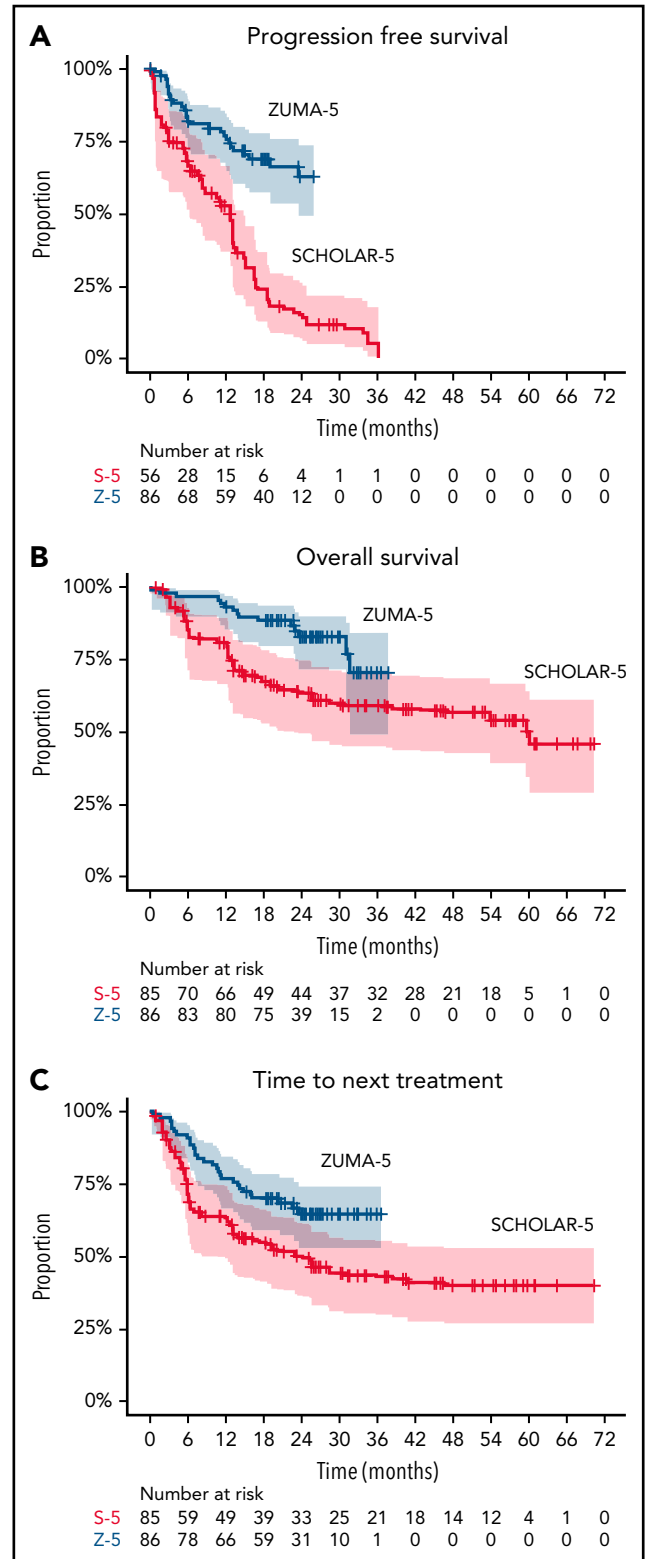
Figure 3. Kaplan-Meier plots comparing ZUMA-5 to SCHOLAR-5 for PFS, OS, and Time-to-next-treatments. Kaplan-Meier curves showing (A) PFS, and (C) time to next treatment in ZUMA-5 (blue) compared with SCHOLAR-5 (red). Shaded area represents 95% CI. Number at risk for the SCHOLAR-5 analysis of PFS was reduced due to the exclusion of DELTA participants from this analysis as PFS was not available in this subgroup. See supplemental Figure 2 for results of all time-to-event outcomes with DELTA participants excluded prior to SMR weight.

GHIONE et al

Table 2. Comparison of clinical outcomes between SCHOLAR-5 and ZUMA-5

	SCHOLAR-5 (n = 85)*	ZUMA-5 (n = 86)	Treatment effect
Time-to-event outcomes	**Median months (95% CI)**	**Median months (95% CI)**	**Hazard ratio (95% CI)**
OS	59.8 (21.9-NE)	NR (31.6, NE)	0.42 (0.21, 0.83)
PFS	12.7 (6.2, 14.7)	NR (23.5, NE)	0.30 (0.18, 0.49)
Time to next treatment	23.4 (9.5, NE)	NR (NE, NE)	0.54 (0.32-0.89)
Response outcomes	**Responders (%)**	**Responders (%)**	**Odds ratio (95% CI)**
ORR	42 (49.9%)	81 (94.2%)	OR: 16.2 (5.6, 46.9)
CR	25 (29.9%)*	68 (79.1%)†	OR: 8.9 (4.3, 18.3)

Note that rounding of patients after classifying as responders or nonresponders in the SCHOLAR-5 weighted sample may lead to a small variability in total sample size.

*SCHOLAR-5 sample size for PFS was 56.

†Response assessments includes CT-based and PET-based scans with limited confirmatory bone marrow biopsies. Thirteen patients with imaging CRs did not receive a confirmatory bone marrow biopsy.

this context, the striking improvement in OS with axi-cel compared with available therapies highlights the suboptimal outcomes with existing therapies and the benefit of therapeutic advances in this population.

For this patient group, large prospective randomized trials using novel immunotherapies in the r/r setting pose feasibility constraints. First, recruitment of large samples is challenging due to the rarity of the disease and its indolent nature, meaning it can take many years for an FL patient to relapse after front-line treatment. This characteristic of r/r FL also led to the relatively modest sample size in the SCHOLAR-5 cohort, which required patients to be initiating a third or higher line of therapy. There are also recent studies demonstrating the challenge of patient identification, Matasar et al identified 100 eligible patients out of 48 685 FL patients in a large US insurance claims database, and the RECORD-FL identified 143 patients from 10 large academic medical centers, going as far back as 2000,[22,23] with only 60 third-line or higher r/r FL patients from 2014 onward. A second feasibility constraint for large RCTs is the lack of consensus on treatment in r/r indolent lymphomas, making the comparator arm difficult to define. Third, because the outcome of standard treatments is poor, clinicians may be reluctant to enroll patients in RCTs where outcomes using novel treatments are uncertain. Clinical study initiation, readout, and subsequent approval is a time-intensive process spanning years. Single-arm studies are typically more feasible, less costly, and can provide a faster timeline for patients to gain access to potentially life-saving therapies.[24] As compared with RCTs, single-arm studies are subject to selection bias and confounding, which may mask an isolated, true therapeutic effect. Nevertheless, between 1999 and 2014, 76 therapeutics were approved by the European Medicines Agency and Food and Drug Administration without investigation in an RCT, with 34 of these indications within hematologic malignancies.[25] In patient populations with substantially high unmet need and/or lack of a SOC, such as this one, the utility of external cohort-matched comparative studies is crucial to answer critical questions faster and ultimately gain novel therapeutic approvals.

The population in the study herein was selected to match key ZUMA-5 clinical trial eligibility criteria to maximize internal validity of the comparative analyses, much like an RCT. The incorporation of external control arms with matched patient-specific data are an increasingly important and relevant tool to demonstrate the comparative effectiveness of novel cancer-directed therapies.[16] SCHOLAR-1, an international, multicohort retrospective study evaluating outcomes in refractory DLBCL, reinforced the striking clinical benefit observed in ZUMA-1,[26] a single-arm registrational study of axi-cel.[27,28] Multiple real-world studies have since demonstrated that axi-cel provides meaningful clinical benefit for patients who otherwise would not have met ZUMA-1 clinical trial eligibility criteria.[29,30] Similarly, data from a single-arm registrational study combining tafasitimab and lenalidomide (L-MIND trial) in concert with an observational, retrospective, 1:1 propensity-matched comparative cohort study (RE-MIND) led to the accelerated approval of tafasitimab and lenalidomide for transplant-ineligible r/r DLBCL.[28]

The analyses performed in this study represent contemporary best practice using statistical methods that account for differences between populations to allow for an optimal, potentially unbiased cross-study comparison. After the application of SMR weights, the baseline characteristics for ZUMA-5 and SCHOLAR-5 were closely balanced from both a statistical and clinical perspective. This was consistent across all sensitivity and subgroup analyses, including when utilizing PS matching (rather than weighting). Although the methods we present here allow a comparison between axi-cel and available r/r FL treatments, they are not a replacement for the rigor and control offered by an RCT. The findings of this study suggest a promising clinical benefit may be associated with axi-cel in r/r FL, which should be studied further in future comparative studies.

The modest sample size of SCHOLAR-5 was a result of strict application of select eligibility criteria from ZUMA-5 to the real-world sample. Our largest participating center, the Memorial Sloan-Kettering Cancer Center, provides a concrete example of the systematic and unbiased patient selection process. As can be seen by the CONSORT diagram (supplemental Appendix), 1100 FL and marginal zone lymphoma patients were reduced to 31 patients eligible for the SCHOLAR-5 cohort. This aligns with the examples, mentioned above, of challenging patient identification in recent studies. Nonetheless, the SCHOLAR-5 sample size was sufficient to achieve the aim of determining the comparative effectiveness of axi-cel vs SOC treatment in a pseudo-

RCT manner. In this regard, a sample size equivalent to the ZUMA-5 population was achieved after PS weighting, with pre-specified prognostic baseline characteristics balanced between groups to the strict threshold of standardized mean difference <0.1. The study results also demonstrate sufficient statistical power in the analysis, with primary and secondary endpoints showing statistically significant differences between ZUMA-5 and SCHOLAR-5 cohorts.

This study has potential limitations. As SCHOLAR-5 is retrospective with data collected from clinical practice databases, missing or incomplete data are expected. Unfortunately, some patient characteristics that could be used to adjust for imbalances were excluded from analyses due to their degree of missingness. These included FLIPI and bone marrow involvement. Although FLIPI and bone marrow biopsies are routinely collected at the time of diagnosis, they are not commonly collected in later lines in real-world clinical practice. Therefore, to maintain an adequate sample size and reduce selection bias, we did not include variables with high rates of missing data in the PS model. To address gaps in electronic medical record or database extractions, trained analysts and clinical teams at participating sites enriched the data by reviewing discrepancies, outliers, and missing values on key data points.

Similarly, differences in unobservable characteristics such as medical history prior to the start of the study period may still exist and lead to residual confounding. Although SMR weighting balances observed covariates between the cohorts, it is not guaranteed to balance unmeasured variables. Again, these analyses cannot match RCTs with respect to internal validity. Moreover, efficacy in ZUMA-5 was assessed by stringent criteria with central review and more frequent disease assessments than would be expected in real-world practice. This potentially biases the ZUMA-5 cohort to have progression events detected earlier than would be the case for the SCHOLAR-5 cohort.

A source of variability and potential bias was the disease progression and response methods. SCHOLAR-5 included some CT-based response assessments and some PET-alone–based response assessments, which may have introduced measurement bias. A CT assessment, absent of a PET assessment, could underestimate response rates in cases where a partial response was later reclassified as an overall disease assessment of CR based on a PET assessment. However, a PET scan alone could overestimate CR as compared with a CT scan alone. Although the effect size may include measurement bias, due to the magnitude of response differences, this bias would be unlikely to impact the response comparison conclusions. Because progressive disease would likely be captured by either imaging modalities, this issue would likely not impact PFS or OS. If anything, the more frequent scanning performed in ZUMA-5 than in real-world settings would bias for earlier detection of disease progression in ZUMA-5.

Finally, safety comparisons were not possible due to lack of safety data in real-world settings. Safety outcomes for axi-cel and individual agents in SCHOLAR-5 have been reported previously, but a comparison was not possible in this study.[13]

Analyses in SCHOLAR-5 add to the growing body of literature describing the critical unmet need for novel therapeutic interventions in r/r FL that result in deep and durable responses. In the single-arm registrational ZUMA-5 phase 2 study, axi-cel demonstrated high OR and CR rates, with evidence of continued clinical benefit in both advanced and high-risk populations.[13] The magnitude of difference in outcomes between the ZUMA-5 and SCHOLAR-5 cohorts demonstrates clinically meaningful benefits for the selected clinical endpoints, with the OS benefits being of particular importance. As response to treatment becomes less durable after each passing line of existing treatments, the high ORR and long durability for axi-cel in the fourth-line or later subgroup translated to even more pronounced treatment effects, including OS. The implication of this finding is that axi-cel continues to have a strong treatment effect despite the number of prior LoTs and that the unmet need for a durable treatment option escalates as patients cycle through many treatments from available options. The differences observed in the primary comparative analysis remained consistent across a variety of sensitivity analyses, indicating the robustness of these results. The comparative analysis between ZUMA-5 and SCHOLAR-5 suggests that axi-cel offers a substantial clinical benefit in patients with r/r FL. Further research is needed to confirm these findings in broader settings and populations.

Authorship

Contribution: The study design and analysis were conducted in a collaboration between Kite, a Gilead Company (study sponsor), and the authors. P.G., M.L.P., H.G., M.T.R., S.B., A.R.P., J.R., and J.G.G. contributed to data collection and verification; S.K., A.J.H., and K.D. contributed to data verification and data analysis; and all authors contributed to results interpretation, writing the manuscript, and approved the final submitted version; the corresponding author had final responsibility for the decision to submit for publication; and the study sponsor, Kite Pharma (a Gilead company), was involved in the study design, conduct, and preparation of the manuscript.

Conflict-of-interest disclosure: M.L.P. reports holding individual stocks and stock options from Seres and Notch; receiving consulting fees from Novartis, Kite, PCYC, and BeiGene; research funding from Seres; and patents and royalty fees from Seres, Juno, Wolters, and Kluwer. A.R.P. reports current employment at and holding stock and stock options from Kite, a Gilead company. S.B. reports receiving fees for serving on a speakers' bureau from Roche and Gilead and consulting fees from Roche. K.D. reports current employment at Delta Hat. C.A.J. reports receiving consulting fees and honoraria from AbbVie, Bristol Myers Squibb/Celgene Corporation, Nkarta, Inc., Novartis Pharmaceuticals Corporation, Precision Biosciences, Kite, a Gilead Company, Lonza, Pfizer,

ORCID profiles: P.G., 0000-0001-6986-7954; A.R.P., 0000-0001-9091-8904; S.B., 0000-0001-7381-680X; K.D., 0000-0001-7994-2137; A.H., 0000-0003-1129-326X; S.K., 0000-0003-0322-5398; J.S., 0000-0003-0017-4863; S.N., 0000-0003-1045-4914; M.T.R., 0000-0002-1176-8184; J.R., 0000-0001-7898-2786; J.G., 0000-0002-8505-7430.

Correspondence: John G. Gribben, Centre for Haemato-Oncology, Charterhouse Square, Barts Cancer Institute, London EC1M 6BQ, United Kingdom; e-mail: j.gribben@qmul.ac.uk.

Footnotes

REFERENCES

1. Al-Hamadani M, Habermann TM, Cerhan JR, Macon WR, Maurer MJ, Go RS. Non-Hodgkin lymphoma subtype distribution, geodemographic patterns, and survival in the US: a longitudinal analysis of the National Cancer Data Base from 1998 to 2011. *Am J Hematol.* 2015;90(9):790-795. https://10.1002/ajh.24086.

2. Teras LR, DeSantis CE, Cerhan JR, Morton LM, Jemal A, Flowers CR. 2016 US lymphoid malignancy statistics by World Health Organization subtypes. *CA Cancer J Clin.* 2016;66(6):443-459.

3. Qualls D, Salles G. Prospects in the management of patients with follicular lymphoma beyond first-line therapy. *Haematologica.* 2022;107(1):19-34.

4. Crump M, Neelapu SS, Farooq U, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study [published correction appears in *Blood.* 2018;131(5):587–588]. *Blood.* 2017;130(16):1800-1808.

5. Pehlivan KC, Duncan BB, Lee DW. CAR-T cell therapy for acute lymphoblastic leukemia: transforming the treatment of relapsed and refractory disease. *Curr Hematol Malig Rep.* 2018;13(5):396-406.

6. Batlevi CL, Sha F, Alperovich A, et al. Follicular lymphoma in the modern era: survival, treatment outcomes, and identification of high-risk subgroups. *Blood Cancer J.* 2020;10(7):74.

7. Link BK, Day BM, Zhou X, et al. Second-line and subsequent therapy and outcomes for follicular lymphoma in the United States: data from the observational National LymphoCare Study. *Br J Haematol.* 2019; 184(4):660-663.

8. Casulo C, Byrtek M, Dawson KL, et al. Early relapse of follicular lymphoma after rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone defines patients at high risk for death: an analysis from the National LymphoCare Study [published correction appears in *J Clin Oncol.* 2016;34(12):1430]. *J Clin Oncol.* 2015;33(23):2516-2522.

9. Casulo C, Dixon JG, Le-Rademacher J, et al. Validation of POD24 as a robust early clinical endpoint of poor survival in FL from 5225 patients on 13 clinical trials. *Blood.* 2022; 139(11):1684-1693.

10. Jurinovic V, Kridel R, Staiger AM, et al. Clinicogenetic risk models predict early progression of follicular lymphoma after first-line immunochemotherapy. *Blood.* 2016; 128(8):1112-1120.

11. Jurinovic V, Metzner B, Pfreundschuh M, et al. Autologous stem cell transplantation for patients with early progression of follicular lymphoma: a follow-up study of 2 randomized trials from the German Low Grade Lymphoma Study Group. *Biol Blood Marrow Transplant.* 2018;24(6):1172-1179.

12. Kanters S, Kahl BS, Wiesinger A, et al. Clinical outcomes in patients relapsed/refractory after ≥ 2 prior lines of therapy for

follicular lymphoma: a systematic literature review and meta-analysis. *J Clin Oncol.* 2021;39(S2):e19548-e19548.

13. 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(suppl 1):40-41.

14. Ghione P, Ghesquieres H, Bobillo S, et al. Outcomes in later-lines of therapy for relapsed/refractory follicular lymphoma: results from the international SCHOLAR-5 study. *Hematol Oncol.* 2021;39(S2): hon.26_2880.

15. Gopal AK, Kahl BS, de Vos S, et al. PI3Kδ inhibition by idelalisib in patients with relapsed indolent lymphoma. *N Engl J Med.* 2014;370(11):1008-1018.

16. Jacobson CA, Chavez JC, Sehgal AR, et al. Axicabtagene ciloleucel in relapsed or refractory indolent non-Hodgkin lymphoma (ZUMA-5): a single-arm, multicentre, phase 2 trial. *Lancet Oncol.* 2022;23(1): 91-103.

17. 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.

18. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC; 2020. This book is only available online.

19. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163(12):1149-1156.

20. Jakobsen JC, Gluud C, Wetterslev 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 Med Res Methodol.* 2017;17(1):162.

21. Ali MS, Prieto-Alhambra D, Lopes LC, et al. Propensity Score methods in health technology assessment: principles, extended applications, and recent advances. *Front Pharmacol.* 2019;10(973):973.

22. Matasar MJ, Shapouri S, Ta JT, To TM, Wu M, Wang R. Healthcare resource utilization and costs of patients with relapsed/ refractory follicular lymphoma receiving 3 or more lines of therapy. *Blood.* 2021; 138(suppl 1):1923.

23. Salles G, Schuster SJ, Dreyling MH, et al. Efficacy comparison of tisagenlecleucel versus standard of care in patients with relapsed or refractory follicular lymphoma. *Blood.* 2021;138(suppl 1):3528.

24. Spreafico A, Hansen AR, Abdul Razak AR, Bedard PL, Siu LL. The future of clinical trial design in oncology. *Cancer Discov.* 2021; 11(4):822-837.

25. Hatswell AJ, Baio G, Berlin JA, Irs A, Freemantle N. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. *BMJ Open.* 2016;6(6): e011666.

26. FDA. Summary Basis for Regulatory Action. https://fda.report/media/114819/Summary-Basis-for-Regulatory-Action—ALBAclone-Anti-D-Fusion-%28Anti-D%29.pdf. Accessed 25 June 2022.

27. Salles G, Duell J, González Barca E, et al. Tafasitamab plus lenalidomide in relapsed or refractory diffuse large B-cell lymphoma (L-MIND): a multicentre, prospective, single-arm, phase 2 study. *Lancet Oncol.* 2020; 21(7):978-988.

28. Zinzani PL, Rodgers T, Marino D, et al. RE-MIND: comparing tafasitamab + lenalidomide (L-MIND) with a real-world lenalidomide monotherapy cohort in relapsed or refractory diffuse large B-cell lymphoma. *Clin Cancer Res.* 2021;27(22): 6124-6134.

29. Jacobson CA, Locke FL, Ma L, et al. Real-world evidence of axicabtagene ciloleucel for the treatment of large B-cell lymphoma in the United States. *Transplant Cell Ther.* 2022. doi: 10.1016/j.jtct.2022.05.026.

30. Nastoupil LJ, Jain MD, Feng L, et al. Standard-of-care axicabtagene ciloleucel for relapsed or refractory large B-cell lymphoma: results from the US lymphoma CAR T consortium. *J Clin Oncol.* 2020;38(27): 3119-3128.