

## ARTICLE OPEN



# Synthetic control arm from mixed clinical trials and real-world data from the LYSA group for untreated diffuse large B-cell lymphoma patients aged over 80 years: a *bona fide* strategy for innovative clinical trials

Valentin Letailleur<sup>1</sup>, Isabelle Chaillol<sup>2</sup>, Fanny Cherblanc<sup>2</sup>, Hervé Ghesquières<sup>3</sup>, Frédéric Peyrade<sup>4</sup>, Stéphanie Guidez<sup>5</sup>, Fontanet Bijou<sup>6</sup>, Pierre Sesques<sup>7</sup>, Cédric Rossi<sup>7</sup>, Luc-Matthieu Fornecker<sup>8</sup>, Ludovic Fouillet<sup>9</sup>, Sylvain Carras<sup>10</sup>, Loïc Chartier<sup>2</sup>, Aurélien Belot<sup>2</sup>, Lucie Oberic<sup>11</sup>, Fabrice Jardin<sup>12</sup> and Benoit Tessoulin<sup>12</sup>✉

© The Author(s) 2025

Patients over 80 years ( $\geq 80$  y.o.) with newly diagnosed diffuse large B-cell lymphomas (DLBCL) are underrepresented in clinical trials (CT). Use of synthetic control arms (SCA) could help to build innovative comparative trials. With this analysis, we aim to demonstrate the clinical performance & use of such SCA. With data from both CT (LNH09-7B) and real-world (REALYSA), we built a mixed SCA composed of  $\geq 80$  y.o. patients with DLBCL in the first line of treatment. In order to display clinically meaningful results, we demonstrate how we can reproduce SENIOR results, a double-arm randomized CT, by switching the internal control arm with our newly built SCA. Patients between arms were balanced using a stabilized inverse probability of treatment weighting approach (sIPTW) based on propensity scores (PS), and the endpoint was overall survival (OS). All covariates included in PS were well balanced after weighting, and OS of Mixed SCA vs. SENIOR experimental arm were not statistically different, with a HR of 0.743 [0.494–1.118] ( $p = 0.1654$ ). Use of SCA built only from real-world data (REALYSA) and sensitivity analyses using different missing data management methods didn't differ from the whole analysis. In newly diagnosed elderly DLBCL patients, the use of SCAs can mimic the control arm of an RCT and could be used to build comparative CT for elderly patients and come up with fast-innovative CTs.

*Blood Cancer Journal* (2025)15:190; <https://doi.org/10.1038/s41408-025-01374-x>

## INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the most common lymphoid malignancy among the elderly, and about 25% [1] of new diagnoses concern patients aged 80 years or older ( $\geq 80$  y.o.). Besides obvious differences related to treatment tolerance and management compared to the younger population, raising knowledge, in molecular cell of origin subclassification, for example, demonstrates that DLBCL of the elderly may have specific aspects [2, 3]. Yet, they are underrepresented in clinical trials, and few trials focus on very elderly patients [1]. Recommended treatment in first-line setting, as established by two phase II clinical trials, LNH03-7B [4] and LNH09-7B [5], is Rituximab and reduced CHOP (Cyclophosphamide, Doxorubicin, Vincristine, Prednisone) combination (R-miniCHOP), with a 2-year overall survival (OS) approaching 60%. Alternatively, regimens replacing or omitting anthracyclines can be proposed for frail patients or those with cardiac function impairment [6–8]. Pre-phase

treatment, combining oral prednisone with or without vincristine and cyclophosphamide, is recommended for the elderly and should be considered after 80 years old, as it has been shown to be beneficial for *performans status* (ECOG) improvement, allowing better tolerance during subsequent R-miniCHOP treatment [5, 9, 10].

More than a decade after LNH09-7B, improving 1<sup>st</sup> line treatment outcomes remains tricky in this population, despite recent overall advances brought by targeted therapies and immunomodulatory agents. Indeed, although the lenalidomide and R-miniCHOP combination (R<sup>2</sup>miniCHOP) had an appealing rationale, as the activated B-cell DLBCL subtype increases with aging [2], SENIOR [11], the first -and only published to date- randomized phase III clinical trial (RCT) focused on  $\geq 80$  y.o. patients, failed to show an improvement in OS, the primary endpoint, which can be partly explained by toxicity issues. Recruitment of the  $\geq 80$  y.o. in clinical trials can be more

<sup>1</sup>Hematology Department, Nantes University Hospital, Nantes, France. <sup>2</sup>Lymphoma Academic Research Organization (LYSARC), Lyon Sud Hospital, Pierre Bénite, France. <sup>3</sup>Hematology Department, Hôpital Lyon Sud, Pierre Bénite, France. <sup>4</sup>Medical Oncology Department, Antoine Lacassagne Center, Nice, France. <sup>5</sup>Hematology Department, Poitiers University Hospital, Poitiers, France. <sup>6</sup>Hematology Department, Bergonié Institute, Bordeaux, France. <sup>7</sup>Hematology Department, Dijon Bourgogne University Hospital, Dijon, France. <sup>8</sup>Hematology Department, Strasbourg Europe Cancer Institute (ICANS), Strasbourg, France. <sup>9</sup>Hematology Department, Saint-Etienne University Hospital, Saint-Etienne, France. <sup>10</sup>Hematology Department, Grenoble-Alpes, Grenoble, France. <sup>11</sup>Hematology Department, Cancer University Institute of Toulouse Oncopole, Toulouse, France. <sup>12</sup>Hematology Department, Henri-Becquerel Center, Rouen, France. ✉email: benoit.tessoulin@chu-nantes.fr

Received: 4 June 2025 Revised: 21 August 2025 Accepted: 17 September 2025

Published online: 03 November 2025

challenging than that of the younger, explaining partly the lack of focused RCT. New combinations are being evaluated in ongoing trials [12], including RCTs [13, 14], that will hopefully bring new treatment options with hypothetical approvals.

At a lower level of proof, phase II clinical trials can bring new opportunities in 1<sup>st</sup> line management. However, comparing results from different phase II trials is not recommended as variations in patient populations, study design, or endpoints can lead to misleading conclusions. Use of synthetic control arms (SCAs) in clinical trials with innovative designs could improve statistical confidence, replacing either entirely the internal control arm (analysis of a phase II CT), or partially (phase III RCT with a mixed control arm composed of randomized patients and historical patients) [15–17]. Even if DLBCL  $\geq 80$  y.o. cannot be considered as a rare disease, the lack of representation and recruitment issues in clinical trials can justify the use of SCA, with rigorous statistical methods to control covariate balance between arms and to reduce biases related to the lack of randomization [18–22] and general rules, as defined by agencies, have been published [23, 24].

In this study, we aimed to build an SCA from mixed real-world and clinical trial data for  $\geq 80$  y.o. DLBCL patients at first line of treatment, and to validate its clinical and statistical relevance by applying it to the SENIOR trial, the only published RCT in this population.

## MATERIAL AND METHODS

### Patients and study design

Patient-level data from REALYSA [25] and LNH09-7B [5] databases were retrieved for SCA patients, and data from the SENIOR trial [11] were used for validation. REALYSA is a French real-world multicentric observational cohort recruiting newly diagnosed lymphoma patients since November 2018. LNH09-7B was a phase II clinical trial assessing the efficacy of Ofatumumab and miniCHOP preceded by a pre-phase treatment (oral vincristine and oral prednisone) as first-line treatment of DLBCL patients  $\geq 80$  y.o., which resulted in a 2-year OS of 64.7% [95% confidence interval (CI): 55.3–72.7]. SENIOR trial was a phase III RCT assessing R-miniCHOP vs. R-Lenalidomide-miniCHOP, with a pre-phase treatment, as first-line treatment of DLBCL patients  $\geq 80$  y.o., with similar inclusion criteria as LNH09-7B. This study failed to show an improvement in OS with the addition of lenalidomide (2-year OS of 66% for R-miniCHOP versus 65.7% for R<sup>2</sup>miniCHOP arm). Patients from LNH09-7B had their treatment started from June 2010 to November 2011, and those from SENIOR from October 2014 to September 2017.

From the REALYSA cohort, we included patients aged  $\geq 80$  y.o., treated with R-miniCHOP combination as first-line treatment, included in the cohort from December 2018 to 31 December 2021. Patients with *performans status* (ECOG) of 3 or 4 and Ann Arbor stage I were not included to match SENIOR inclusion criteria. Data were exported from the registry on the 25th of June 2024.

Clinical trials patients with non-compliant diagnoses (other than DLBCL or high-grade B-cell lymphoma (HGBL)), according to anatomopathological centralized reviews realized in each trial or local diagnosis for REALYSA were excluded.

The study was designed as follows: first step was to build a mixed SCA ("Mixed SCA 1") from real-world data (REALYSA) and clinical trial data (LNH09-7B) adjusted on the SENIOR control arm, and to evaluate if it could mimic the internal control arm from SENIOR. The second step was to build a mixed SCA ("Mixed SCA 2") adjusted on the experimental arm of SENIOR, to evaluate if we could replicate the efficacy results from the SENIOR study by switching the internal control arm to the Mixed SCA 2. For each step, populations were balanced as described below.

Primary and only endpoint for arms comparison after weighting was OS, measured from the date of inclusion (or date of diagnosis in REALYSA) to the date of death. As a high number of REALYSA patients had a limited follow-up period, we chose to censor all patients at 24 months for all cohorts.

### Weighting procedures

Propensity scores (PS) were estimated for each patient included using logistic regression with following covariates: sex, age (spline), Ann Arbor

stage (I–II versus III–IV), *performans status* (ECOG), number of extra-nodal sites involved ( $< 2$  versus  $\geq 2$ ), international prognostic index (IPI) score (0–2 versus 3–5), B symptoms, lactate dehydrogenase (LDH) level (normal versus over the upper limit of normal), bulky mass  $> 10$  cm, albumin level in gram per liter (g/L) (spline).

For comparison between the Mixed SCA 1 and SENIOR internal control arm, PS allows to estimate the probability for a patient to receive "R-miniCHOP via SENIOR." For comparison between the Mixed SCA 2 versus the SENIOR experimental arm, PS allows to estimate the probability for a patient to receive R<sup>2</sup>miniCHOP.

Then, patients' covariates between arms were weighted using the stabilized inverse probability of treatment weighting (sIPTW) statistical approach, the probability of treatment being reflected by PS [26].

To avoid positivity violations, patients with extreme PS (below 0.1 or over 0.9) were excluded [27]. After weighting procedures, standardized mean differences (SMD) were estimated between arms for each covariate included for PS estimation to check balance in covariates' distributions. With the statistical approach used, sIPTW allowed to estimate the average treatment effect.

### Missing data management

Due to the high number of missing data for patients from REALYSA for some covariates, the multiple imputation method "across" was performed (15 imputations by patient) [28, 29]. That generated 15 complete datasets. PS were calculated for each of the datasets. Then, for each patient, the median of the 15 PS was used for weighting. The SMD presented in this manuscript was calculated on real data, i.e., before imputation. With the "across" missing management method, outcome analyses are performed on one dataset.

### Statistics for outcome analyses

As the first objective was to assess if the Mixed SCA 1 could mimic the internal control arm of SENIOR, Hazard Ratios (HR) 95% CIs were expected to include 1. The second objective was to reproduce SENIOR results by replacing the internal control arm with the Mixed SCA 2; HR were expected to overlap with the results obtained in the SENIOR trial (HR [95% CI] of 0.996 [0.66–1.51]), including 1 in the HR 95% CI. Power calculation was not made in our study, but the aim was to build the largest SCA we could from sources we chose (LNH09-7B and REALYSA) with comparable patients at baseline.

OS curves were generated using the Kaplan–Meier method, and OS curves were compared with the log-rank test. HR and 95% CIs were calculated using the Cox proportional hazards model. One-sided *p* value below 0.05 was considered significant. Analyses were performed using SAS software 9.4.

### Sensitivity analyses

Sensitivity analyses were performed using data coming only from the REALYSA cohort to evaluate the possibility to build a well-balanced SCA using only real-world data (REALYSA-SCA), with the same methodology used for the mixed SCA building. Additionally, we performed sensitivity analysis using different missing data management methods on Mixed SCAs analyses or REALYSA-SCA analyses. The "Complete cases" method excludes patients with missing data. For the "within" method, analyses are performed on each imputed dataset (15 analyses), and results were combined using Rubin's rules [28, 29].

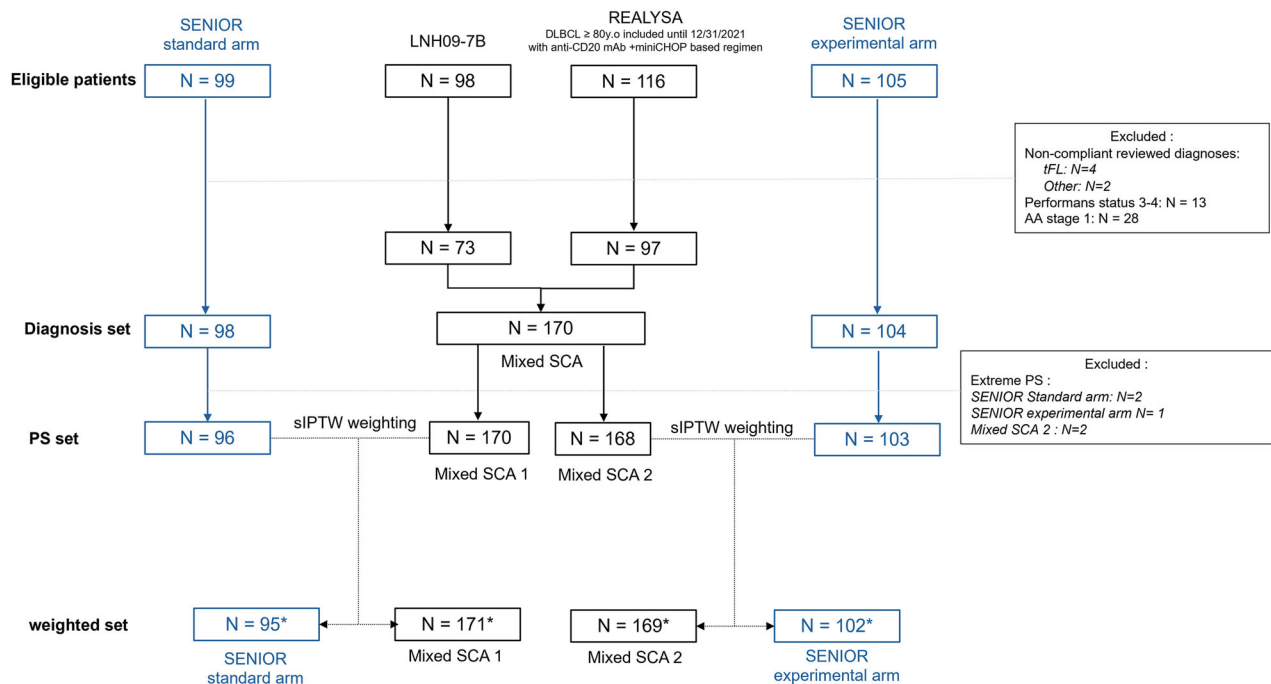
## RESULTS

### Patients' characteristics

Overall, in this study, we included 73 patients from LNH09-7B and 97 patients from REALYSA to source mixed SCAs, and 98 or 104 patients from SENIOR standard or experimental arm, respectively, to assess Mixed SCAs relevance (Fig. 1). This population will be referred to as the confirmed Diagnosis Set. All included patients received anti-CD20 monoclonal antibody (mAb) + miniCHOP-based regimen as first-line treatment.

Table 1 shows characteristics of patients from each cohort included in the confirmed Diagnosis Set. Most patients included in this study had an intermediary-high or high risk DLBCL (77.1% of overall patients with an IPI score at 3 or higher).

PS were calculated for all patients before each sIPTW weighting procedure. Five patients were excluded (Based on Mixed-SCA 1 PS:



**Fig. 1 Study flow-chart.** tFL transformed follicular lymphoma, PS propensity score, mAb monoclonal antibody. \*Effective sample size of the pseudo-population after weighting, which is close to the real samples.

2 from SENIOR standard arm, and based on Mixed-SCA 2: 1 from SENIOR experimental arm and 2 from Mixed-SCA 2) because of extreme values (above 0.9 or below 0.1) (Fig. 1). Final population was included in the “PS set” (Fig. 1).

In the PS set, some covariates were unbalanced before weighting, with an absolute SMD above 0.1 for 4 covariates (Sex, *performans status* (ECOG) 0 and 2, Mass >10 cm and Ann Arbor stage) before sIPTW between SENIOR Standard arm and Mixed SCA 1 and for 5 covariates (Sex, Mass >10 cm, LDH, IPI, B symptoms) between SENIOR experimental arm and Mixed SCA (Fig. 2). SMD were also calculated on a pool of the 15 datasets with imputation and were similar.

### Weighting procedures

Final populations in the PS set were weighted with the sIPTW method: SENIOR Standard arm with Mixed SCA 1 and SENIOR Experimental arm with Mixed SCA 2. Table 2 shows patients' characteristics distribution in the 4 arms before and after weighting.

Using the “across” method to manage missing data, weighting procedures with sIPTW were efficient to balance covariates between arms, as all SMD for covariates included in the PS were below 0.10 (Fig. 3).

To explore if the missing data management method used in this study could impact weighting efficiency, we also analyzed the balance of covariates using the “complete case” or “within” method. SMD for all covariates was <0.1, whatever the method used for missing data management. Moreover, in the “within” method, at different imputations (from 1 to 15), SMD for all covariates were <0.1, and variation of SMD values across the different imputations was very low (Supplementary Fig. 1).

### Outcome analysis

The first step was to assess if Mixed SCA 1 could reproduce the SENIOR control arm's OS. OS was not significantly different between the SENIOR control arm and Mixed SCA 1, with an HR [95% CI] of 0.86 [0.57–1.32] ( $p = 0.493$ ) before weighting, and an HR of 0.79 [0.52–1.20] ( $p = 0.277$ ) after weighting with sIPTW (Fig. 3A, B).

The second step was to assess the reproducibility of SENIOR trial results by switching the SENIOR control arm to Mixed SCA 2. In our study, OS was not significantly different between the SENIOR experimental arm and Mixed SCA 2, with an HR of 0.83 [0.55–1.25] ( $p = 0.3631$ ) before weighting, and an HR of 0.74 [0.49–1.12] ( $p = 0.165$ ) after weighting with sIPTW (Fig. 3C, D).

There was no statistically significant difference in OS between SENIOR arms and Mixed SCAs using different missing data management methods (Supplementary Table 1).

### REALYSA-SCA

We also assessed the feasibility and viability of SCAs (REALYSA-SCAs) built with real-world data from the REALYSA cohort only, with the same methods as the Mixed SCAs.

Table 3 shows patients' characteristics from the SENIOR control arm or experimental arm compared to patients from REALYSA-SCAs before and after sIPTW. More covariates were imbalanced before weighting compared to analyses conducted with Mixed SCAs, but after weighting, SMDs of all covariates were <0.1 (Fig. 4). Weighting procedures were efficient, irrespective of the weighting method or the missing data management method used (Supplementary Fig. 2).

OS was not significantly different between the SENIOR control arm and REALYSA-SCA-1, with an HR [95% CI] of 0.89 [0.56–1.44] ( $p = 0.640$ ) before weighting and an HR of 0.90 [0.56–1.43] ( $p = 0.478$ ) after sIPTW. Similarly, OS was not significantly different between the SENIOR experimental arm and REALYSA-SCA-2 with an HR of 0.84 [0.53–1.35] ( $p = 0.474$ ) before weighting and an HR of 0.88 [0.55–1.41] ( $p = 0.612$ ) after weighting with sIPTW (Fig. 5A–D).

OS comparison results between REALYSA-SCAs and SENIOR standard or experimental arm obtained in sensitivity analyses were close to those obtained with the main methodology, and no statistically significant difference was observed (Supplementary Table 1).

### DISCUSSION

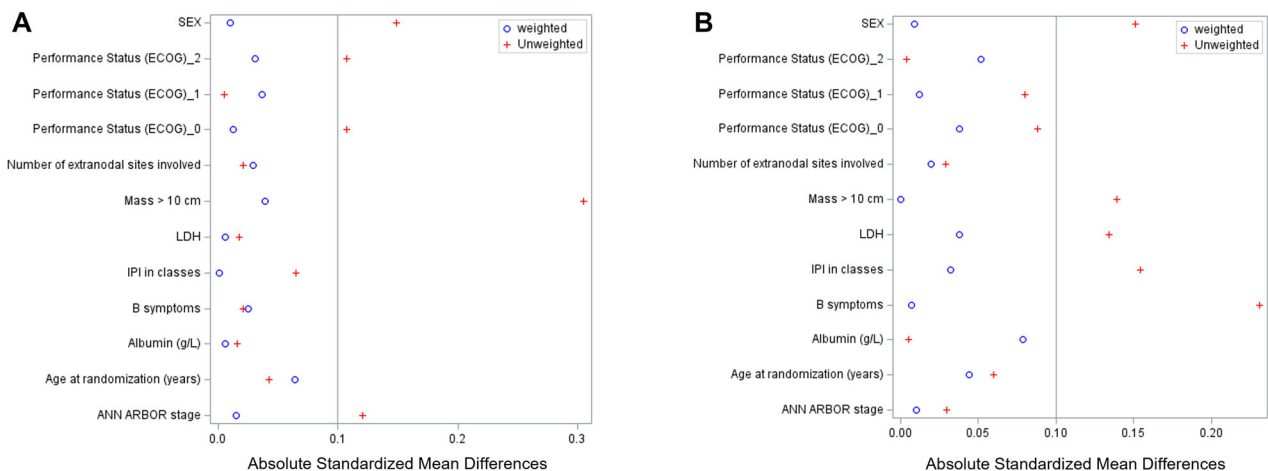
In this study, we showed that a well-balanced SCA designed for ≥80 y.o. DLBCL patients in the first-line setting, with data extracted

**Table 1.** Patients' characteristics from the confirmed diagnosis set.

	Population							
	SENIOR experimental arm N = 104		SENIOR standard arm N = 98		LNHO9-7B N = 73		REALYSA N = 97	
<b>Sex</b>								
Male	45	(43.3%)	44	(44.9%)	34	(46.6%)	53	(54.6%)
Female	59	(56.7%)	54	(55.1%)	39	(53.4%)	44	(45.4%)
<b>Age at randomization (years)</b>								
Mean (SD)	83.7 (3.48)		84.2 (3.72)		83.7 (3.06)		84.1 (3.02)	
≤85	80	(76.9%)	70	(71.4%)	51	(69.9%)	70	(72.2%)
>85	24	(23.1%)	28	(28.6%)	22	(30.1%)	27	(27.8%)
<b>Ann Arbor Stage</b>								
2	12	(11.5%)	15	(15.3%)	9	(12.3%)	9	(9.3%)
3	15	(14.4%)	12	(12.2%)	16	(21.9%)	15	(15.5%)
4	77	(74.0%)	71	(72.4%)	48	(65.8%)	73	(75.3%)
<b>Performans status (ECOG)</b>								
0	28	(26.9%)	18	(18.4%)	23	(31.5%)	16	(16.7%)
1	53	(51.0%)	54	(55.1%)	35	(47.9%)	57	(59.4%)
2	23	(22.1%)	26	(26.5%)	15	(20.5%)	23	(24.0%)
Missing	0		0		0		1	
<b>Extra-nodal involvement ≥2</b>								
Yes	54	(51.9%)	50	(51.0%)	30	(41.1%)	55	(56.7%)
<b>IPI in classes</b>								
1–2	27	(26.7%)	23	(23.5%)	15	(20.5%)	19	(20.0%)
3–5	74	(73.3%)	75	(76.5%)	58	(79.5%)	76	(80.0%)
Missing	3		0		0		2	
<b>B symptoms</b>								
Yes	22	(21.2%)	29	(29.6%)	13	(17.8%)	40	(41.2%)
<b>LDH</b>								
>Upper limit	61	(60.4%)	66	(67.3%)	47	(64.4%)	64	(68.8%)
Missing	3		0		0		4	
<b>Mass &gt; 10 cm</b>								
Yes	17	(16.3%)	22	(22.4%)	5	(6.8%)	13	(15.7%)
Missing	0		0		0		14	
<b>Diagnosis (classification)</b>								
HGBL	3	(2.9%)	1	(1.0%)	0	(0.0%)	16	(16.5%)
DLBCL	101	(97.1%)	97	(99.0%)	73	(100.0%)	81	(83.5%)
<b>Albumin (g/L)</b>								
Mean (SD)	35.8 (7.02)		35.9 (7.46)		35.9 (5.66)		34.7 (6.39)	
Missing	4		1		5		27	

Characteristics of all patients included, from their cohort of origin, are shown before propensity score assessment and weighting procedures. Values are displayed as *n* (%) or mean (SD) when specified.

SD standard deviation, IPI international prognostic index, LDH lactate dehydrogenase, HGBL high-grade B-cell lymphoma, DLBCL diffuse large B-cell lymphoma.

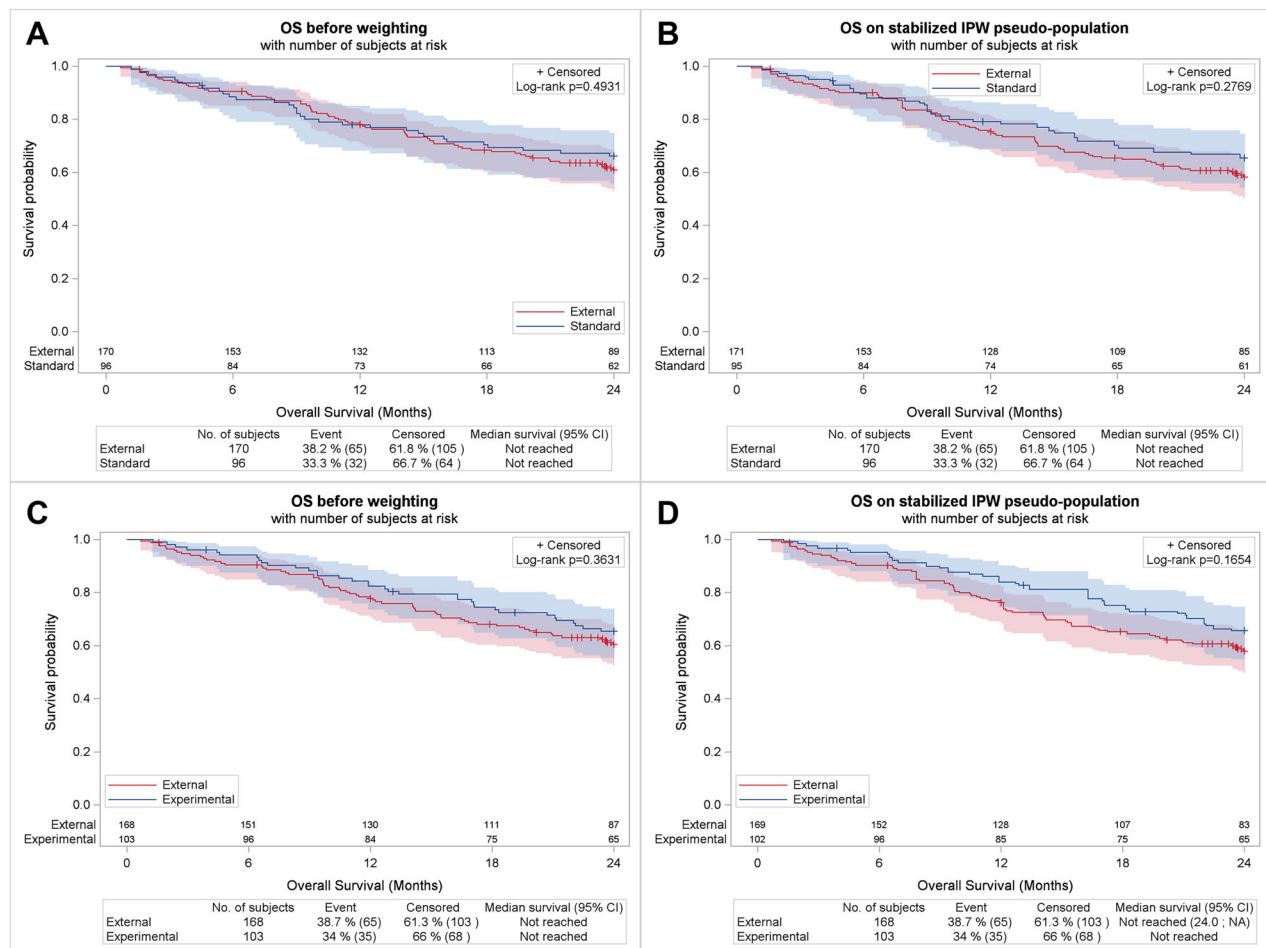


**Fig. 2** Absolute Standardized Mean Differences (SMD) using Mixed SCAs. Balance of covariates included in the propensity score between (A) SENIOR standard arm or (B) SENIOR experimental arm and Mixed SCAs before and after sIPTW weighting.

**Table 2.** Patients' characteristics from the SENIOR standard or SENIOR experimental arm and Mixed SCA 1 or 2 from the PS set before and after sIPTW weighting.

	Before sIPTW weighting			After sIPTW weighting			Before sIPTW weighting			After sIPTW weighting		
	SENIOR standard arm N = 96	Mixed-SCA 1 N = 170	SENIOR standard arm N = 95	Mixed-SCA 1 N = 171	SENIOR experimental arm N = 103	Mixed-SCA 2 N = 168	SENIOR experimental arm N = 102	Mixed-SCA 2 N = 169				
Sex												
Male	42 (43.8%)	87 (51.2%)	45.7 (47.9%)	81.2 (47.4%)	45 (43.7%)	86 (51.2%)	50.0 (49.2%)	82.7 (48.8%)				
Female	54 (56.3%)	83 (48.8%)	49.7 (52.1%)	90.1 (52.6%)	58 (56.3%)	82 (48.8%)	51.6 (50.8%)	86.8 (51.2%)				
Age at randomization (years)												
Mean (SD)	84.0 (3.67)	83.9 (3.04)	84.0 (3.32)	84.2 (3.61)	83.7 (3.48)	83.9 (3.05)	83.9 (3.29)	84.0 (3.57)				
Ann Arbor Stage												
II	14 (14.6%)	18 (10.6%)	11.8 (12.4%)	20.4 (11.9%)	12 (11.7%)	18 (10.7%)	11.2 (11.0%)	18.1 (10.7%)				
III–IV	82 (85.4%)	152 (89.4%)	83.6 (87.6%)	150.9 (88.1%)	91 (88.3%)	150 (89.3%)	90.4 (89.0%)	151.3 (89.3%)				
Performans status (ECOG)												
0	18 (18.8%)	39 (23.1%)	19.7 (20.6%)	36.1 (21.2%)	28 (27.2%)	39 (23.4%)	25.8 (25.4%)	40.1 (23.8%)				
1	52 (54.2%)	92 (54.4%)	53.0 (55.6%)	91.6 (53.7%)	52 (50.5%)	91 (54.5%)	52.7 (51.9%)	86.5 (51.3%)				
2	26 (27.1%)	38 (22.5%)	22.7 (23.8%)	42.8 (25.1%)	23 (22.3%)	37 (22.2%)	23.1 (22.7%)	42.1 (24.9%)				
Missing	0	1	0.0	0.8	0	1	0.0	0.8				
Extra-nodal involvement ≥2												
Yes	49 (51.0%)	85 (50.0%)	47.1 (49.4%)	87.1 (50.9%)	53 (51.5%)	84 (50.0%)	53.3 (52.5%)	87.3 (51.5%)				
IPI in classes												
1–2	22 (22.9%)	34 (20.2%)	19.7 (20.7%)	35.1 (20.7%)	27 (27.0%)	34 (20.5%)						
3–5	74 (77.1%)	134 (79.8%)	75.6 (79.3%)	134.5 (79.3%)	73 (73.0%)	132 (79.5%)	76.2 (77.3%)	131.9 (78.7%)				
Missing	0	2	0.0	1.7	3	2	3.1	1.9				
B symptoms												
Yes	29 (30.2%)	53 (31.2%)	29.4 (30.9%)	50.9 (29.7%)	21 (20.4%)	51 (30.4%)	25.5 (25.1%)	43.0 (25.4%)				
LDH												
>Upper limit	65 (67.7%)	111 (66.9%)	65.6 (68.8%)	114.6 (68.5%)	60 (60.0%)	109 (66.5%)	63.5 (64.5%)	109.2 (66.3%)				
Missing	0	4	0.0	3.9	3	4	3.1	4.7				
Mass > 10 cm												
Yes	22 (22.9%)	18 (11.5%)	15.1 (15.8%)	27.4 (17.3%)	17 (16.5%)	18 (11.7%)	16.1 (15.8%)	24.8 (15.8%)				
Missing	0	14	0.0	12.3	0	14	0.0	12.6				
Albumin (g/L)												
Mean (SD)	35.4 (6.5639)	35.3 (6.0477)	35.3 (6.3026)	35.3 (6.2173)	35.6 (6.6928)	35.5 (5.8327)	35.7 (5.9928)	35.2 (6.0126)				
Missing	1	32	1	32	4	32	4	32				





**Fig. 3 Outcome analysis using Mixed SCAs.** OS comparison between Mixed SCA 1 and SENIOR control arm before (A) and after (B) weighting. OS comparison between Mixed SCA 2 and SENIOR experimental arm (C) before and D after weighting.

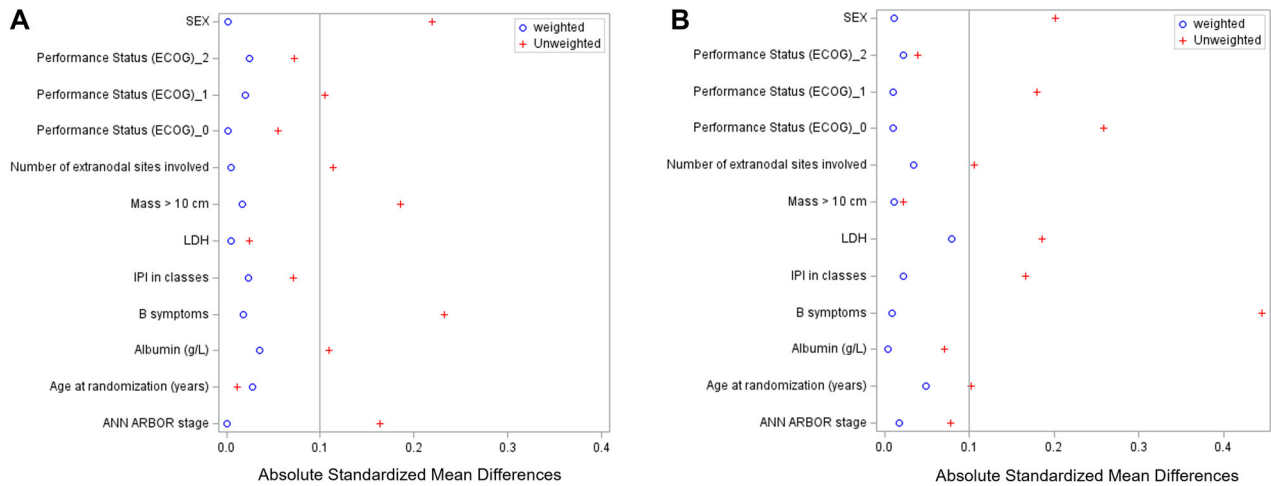
from historical clinical trials and real-world settings, can mimic the internal control arm of an RCT. Switching the internal control arm by a SCA reproduced efficacy results based on OS comparison, with an HR [95% CI] of 0.743 [0.494–1.118] with a mixed SCA from clinical trial and real-world data and an HR of 0.88 [0.55–1.41] ( $p = 0.612$ ) with a SCA only based on real-world data, compared to an HR of 0.996 [0.66–1.51] in SENIOR trial. This research demonstrates that we should open the field of newly designed trials to improve our clinical trial accrual and overall clinical benefit, even in the very elderly setting.

Our selected balancing procedure with the sIPTW method was highly efficient with all SMD  $< 0.1$  for covariates included in the PS calculation, along with a low drop-rate of “extreme patients” (0.7–1.4% (Fig. 1)), ensuring robustness of the results. We used mixed sources of real-world data and clinical trial data to build this mixed SCA, to strengthen the extent of reliability of our results from a perspective of “real-life” incoming phase III clinical trials. Surprisingly, we observed a limited number of real-world-sourced patients who would be excluded after applying inclusion criteria (*performans status* (ECOG)  $< 3$  and Ann Arbor stage  $> I$ ). Regarding patients’ characteristics, those results showed that  $\geq 80$  y.o. patients with DLBCL in the first-line setting are quite similar despite very different inclusion periods between LNH09-7B, SENIOR, and REALYSA. Indeed, most patients’ characteristics were already balanced before weighting, as only a few covariates had a high SMD despite the absence of randomization and the use of real-world data. It suggests that (1) by applying the SENIOR trial population inclusion criteria, we selected a substantially

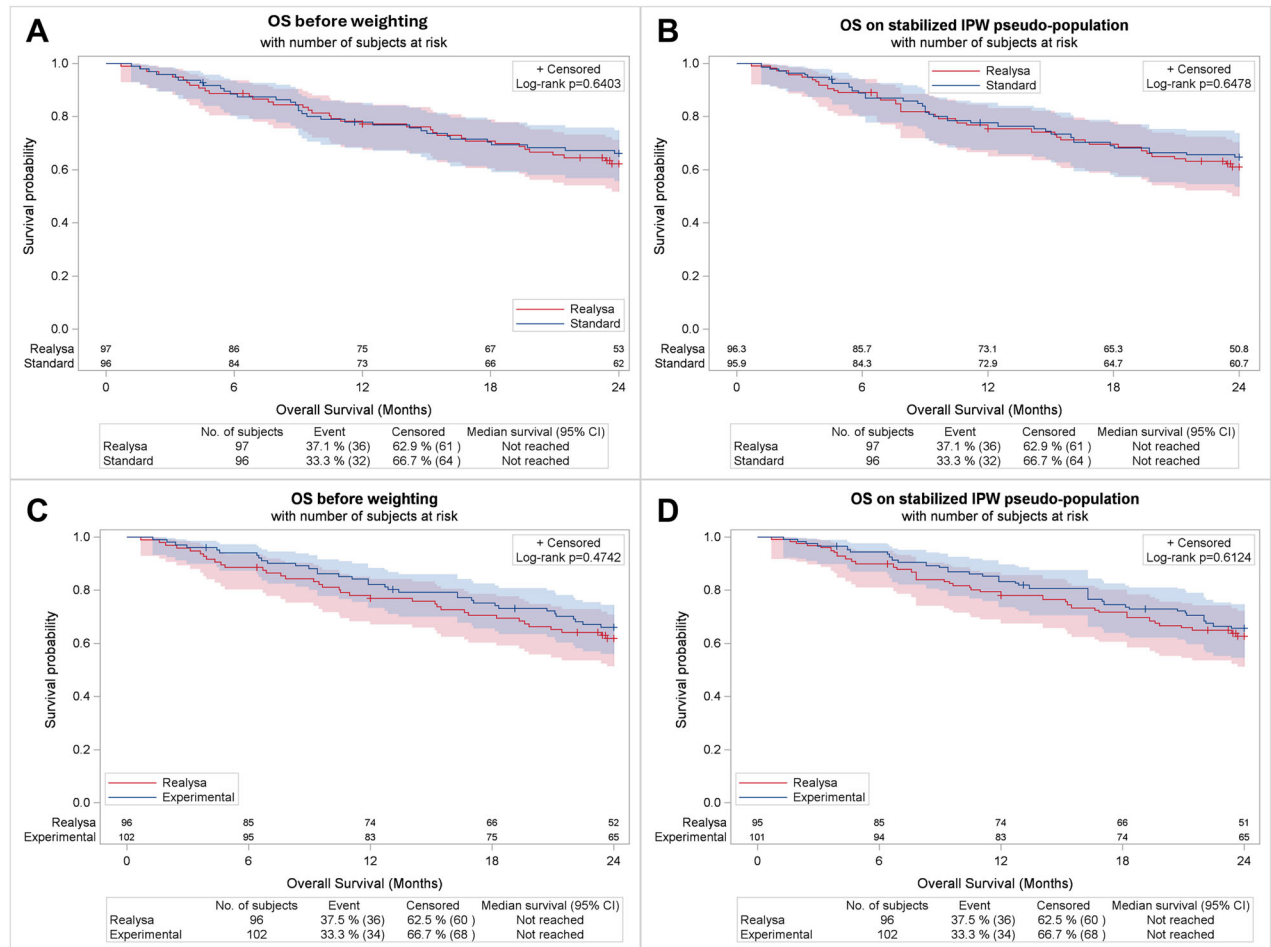
homogeneous population of DLBCL elderly patients to build SCAs, and (2) the SENIOR trial enrolled an unbiased DLBCL elderly population with immunochemotherapy intent of treatment. Of note, analysis based on an SCA only built with these real-world data showed that patients could also be well balanced with SENIOR patients from both internal control arm or experimental arm with the sIPTW method (REALYSA-SCA) with an HR [95% CI] of 0.895 [0.560–1.429] and 0.877 [0.547–1.407], respectively. Some limitations can be taken into account for our study. We did not plan a statistical analysis to compare HRs with 95% between the SENIOR trial and the results of this study because, to our knowledge, there is no such statistical analysis that could properly conclude with a defined threshold if those results can allow us to conclude on a reproduction of efficacy. CIs mainly overlapped with SENIOR results. The fact that the OS of Mixed SCA 1 and SENIOR internal control arm is close can be a result in favor of the study conclusion. Period of data accrual is different between LNH09-7B trial data (2010–2011), and REALYSA (2018–2021), and we cannot rule out that supportive care and improvements in geriatric assessment and management may have improved during this 10-year time frame. It is possible these differences led to improved long-term OS with REALYSA-SCAs compared to Mixed-SCAs. However, we did not plan a comparison between the sources of patients. On the other hand, management of relapse didn’t dramatically change between 2010–2011 and 2018–2021, so differences in OS related to improved post-relapse treatment are unlikely. Furthermore, we couldn’t incorporate prognostic geriatric assessments (e.g., Instrumental Activities of Daily Living

**Table 3.** Patients' characteristics from the SENIOR standard or SENIOR experimental arm and REALYSA-SCA from the PS set, before and after siPTW weighting.

	Before siPTW weighting			After siPTW weighting			Before siPTW weighting			After siPTW weighting		
	SENIOR standard arm N = 96	REALYSA-SCA 1 N = 97	SENIOR standard arm N = 96	REALYSA-SCA 1 N = 96	SENIOR experimental arm N = 102	REALYSA-SCA 2 N = 96	SENIOR experimental arm N = 101	REALYSA-SCA 2 N = 95				
Sex												
Male	42 (43.8%)	53 (54.6%)	47.2 (49.2%)	47.3 (49.1%)	45 (44.1%)	52 (54.2%)	51.2 (50.7%)	48.6 (51.2%)				
Female	54 (56.3%)	44 (45.4%)	48.7 (50.8%)	49.0 (50.9%)	57 (55.9%)	44 (45.8%)	49.8 (49.3%)	46.3 (48.8%)				
Age at randomization (years)												
Mean (SD)	84.0 (3.67)	84.1 (3.02)	84.2 (3.51)	84.1 (3.43)	83.8 (3.48)	84.1 (3.04)	83.9 (3.27)	84.0 (3.40)				
Ann Arbor Stage												
II	14 (14.6%)	9 (9.3%)	11.0 (11.4%)	11.0 (11.4%)	12 (11.8%)	9 (9.4%)	10.5 (10.4%)	10.4 (10.9%)				
III-IV	82 (85.4%)	88 (90.7%)	85.0 (88.6%)	85.3 (88.6%)	90 (88.2%)	87 (90.6%)	90.5 (89.6%)	84.5 (89.1%)				
Performans status (ECOG)												
0	18 (18.8%)	16 (16.7%)	17.2 (17.9%)	17.0 (17.8%)	28 (27.5%)	16 (16.8%)	23.4 (23.2%)	21.4 (22.8%)				
1	52 (54.2%)	57 (59.4%)	54.4 (56.7%)	53.2 (55.7%)	51 (50.0%)	56 (58.9%)	54.5 (54.0%)	50.3 (53.5%)				
2	26 (27.1%)	23 (24.0%)	24.4 (25.4%)	25.3 (26.5%)	23 (22.5%)	23 (24.2%)	23.0 (22.8%)	22.3 (23.7%)				
Missing	0	1	0.0	0.8	0	1	0.0	0.9				
Extra-nodal involvement ≥2												
Yes	49 (51.0%)	55 (56.7%)	52.0 (54.2%)	52.4 (54.5%)	52 (51.0%)	54 (56.3%)	53.2 (52.7%)	51.6 (54.4%)				
IPI in classes												
1-Feb	22 (22.9%)	19 (20.0%)	19.2 (20.0%)	19.8 (20.9%)	27 (27.3%)	19 (20.2%)	23.9 (24.4%)	21.8 (23.4%)				
3-May	74 (77.1%)	76 (80.0%)	76.8 (80.0%)	74.8 (79.1%)	72 (72.7%)	75 (79.8%)	74.3 (75.6%)	71.2 (76.6%)				
Missing	0	2	0.0	1.6	3	2	2.7	1.9				
B symptoms												
Yes	29 (30.2%)	40 (41.2%)	35.5 (37.0%)	34.8 (36.1%)	21 (20.6%)	39 (40.6%)	29.3 (29.0%)	27.9 (29.4%)				
LDH												
>Upper limit	65 (67.7%)	64 (68.8%)	66.7 (69.5%)	64.5 (69.8%)	59 (59.6%)	63 (68.5%)	61.8 (62.9%)	59.9 (66.7%)				
Missing	0	4	0.0	3.8	3	4	2.7	5.0				
Mass > 10 cm												
Yes	22 (22.9%)	13 (15.7%)	19.0 (19.8%)	17.4 (20.5%)	17 (16.7%)	13 (15.9%)	19.2 (19.0%)	16.2 (19.5%)				
Missing	0	14	0.0	11.5	0	14	0.0	11.8				
Albumin (g/L)												
Mean (SD)	35.4 (6.56)	34.7 (6.38)	35.0 (6.50)	35.2 (6.51)	35.4 (6.47)	35.0 (6.17)	35.4 (6.13)	35.4 (6.12)				
Missing	1	27	1	27	4	27	4	27				



**Fig. 4 Absolute Standardized Mean Differences (SMD) using REALYSA SCAs.** Balance of covariates included in the propensity score between (A) SENIOR standard arm or (B) SENIOR experimental arm and REALYSA-SCAs before and after siPTW weighting.



**Fig. 5 Outcome analysis using REALYSA SCAs.** OS comparison between REALYSA-SCA 1 and SENIOR control arm before (A) and after (B) weighting. OS comparison between REALYSA-SCA 2 and SENIOR experimental arm (C) before and (D) after weighting.

(IADL) scale) into the PS calculation, because of the lack of this information in all datasets. In our dataset, data were prospectively collected, with a low rate of missing data (mass > 10 cm and albumin level with >5% of missing data from REALYSA patients) [30]. Moreover, we showed that missing data could be reliably imputed, allowing us to keep a steady sample size with robust

outcome analysis results. Indeed, sensitivity analyses for missing data management with “complete cases,” or “within” multiple imputation methods, led to the same conclusions as for the main (“across”) method.

From a statistical point of view, we chose the siPTW weighting method because it resulted in a pseudo-population with stable



and comparable sample sizes, as opposed to IPTW, which would have resulted in a larger pseudo-population (that may result in an increased rate of type 1 error). As illustrated in sensitivity analyses, CIs were narrower with IPTW compared to sIPTW weighting, but we still obtained consistent results with IPTW.

Other examples of SCA building and validation emerged recently in the hematology field using data at an individual level from historical clinical trials or real-world data with PS-based methods to control for confounding variables [31, 32]. For example, another study also used REALYSA data to build an SCA with newly diagnosed patients with advanced Hodgkin lymphoma and reproduced efficacy results of the phase III RCT AHL2011 [33].

We believe our study demonstrates that in a difficult-to-study population, SCA could be used as a *bona fide* strategy to improve very elderly DLBCL patients' outcomes by allowing trials with "*in silico*" control arms along with more promising experimental arms. Indeed, SENIOR is currently the only published RCT in this setting and recruited 249 patients in 3 years within 100 centers in France and Belgium, illustrating the challenges to enroll these patients.

The mixed SCAs we built could be used to help assessment of new combinations with targeted therapies that could improve OS compared to the R-miniCHOP-based regimen, for which the adjunction of another drug seems tricky due to toxicity issues in this population. In our opinion, OS is the most relevant primary outcome to compare SCA to another arm, especially in the elderly population, as first progression is soon followed by death in this low-reserve population. Use of progression or event-free survival would require the same follow-up for the evaluation of response, which is unlikely with multicentric real-world data. Furthermore, health authorities are now giving weight to these indirect, statistically sound strategies to allow market access to new treatments (if the control arm is the right one), and we could increase this weight by a new design of trials. For instance, comparative trials with control arms composed of 50% of external patients and 50% of randomized patients (in order to keep toxicity comparisons available) could accelerate accrual time by reducing the number of patients to include, or, with a steady number of patients to include, allow more patients to be included in the experimental arm (2:1 randomization). To conclude, we built here an SCA from mixed clinical trial and real-world data for  $\geq 80$  y.o. DLBCL patients in the first-line setting that could be used to build comparative trials with an innovative design, which could help to explore the ability of new treatment combinations to improve survival in this otherwise poorly-represented population in clinical trials.

## DATA AVAILABILITY

Data will be made available upon reasonable request to the corresponding author.

## REFERENCES

- Kanapuru B, Singh H, Kwitkowski V, Blumenthal G, Farrell AT, Pazdur R. Older adults in hematologic malignancy trials: representation, barriers to participation and strategies for addressing underrepresentation. *Blood Rev.* 2020;43:100670.
- Mareschal S, Lanic H, Ruminy P, Bastard C, Tilly H, Jardin F. The proportion of activated B-cell like subtype among de novo diffuse large B-cell lymphoma increases with age. *Haematologica.* 2011;96:1888–90.
- Wilson WH, Wright G, Huang DW, Hodgkinson B, Balasubramanian S, Fan Y, et al. Effect of ibrutinib with R-CHOP chemotherapy in genetic subtypes of DLBCL. *Cancer Cell.* 2021;39:1643–53.e3.
- Peyrade F, Jardin F, Thieblemont C, Thyss A, Emile JF, Castaigne S, et al. Attenuated immunochemotherapy regimen (R-miniCHOP) in elderly patients older than 80 years with diffuse large B-cell lymphoma: a multicentre, single-arm, phase 2 trial. *Lancet Oncol.* 2011;12:460–8.
- Peyrade F, Bologna S, Delwail V, Emile JF, Pascal L, Fermé C, et al. Combination of ofatumumab and reduced-dose CHOP for diffuse large B-cell lymphomas in patients aged 80 years or older: an open-label, multicentre, single-arm, phase 2 trial from the LYSA group. *Lancet Haematol.* 2017;4:e46–55.
- Moccia AA, Schaff K, Freeman C, Hoskins PJ, Klasa RJ, Savage KJ, et al. Long-term outcomes of R-CEOP show curative potential in patients with DLBCL and a contraindication to anthracyclines. *Blood Adv.* 2021;5:1483–9.
- Fields PA, Townsend W, Webb A, Counsell N, Pocock C, Smith P, et al. De novo treatment of diffuse large B-cell lymphoma with rituximab, cyclophosphamide, vincristine, gemcitabine, and prednisolone in patients with cardiac comorbidity: a United Kingdom National Cancer Research Institute Trial. *J Clin Oncol.* 2014;32:282–7.
- Laribi K, Denizon N, Bolle D, Truong C, Besançon A, Sandrini J, et al. R-CVP regimen is active in frail elderly patients aged 80 or over with diffuse large B cell lymphoma. *Ann Hematol.* 2016;95:1705–14.
- Pfreundschuh M, Schubert J, Ziepert M, Schmits R, Mohren M, Lengfelder E, et al. Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomised controlled trial (RICOVER-60). *Lancet Oncol.* 2008;9:105–16.
- Lugtenburg PJ, Mutsaers PGNJ. How I treat elderly patients with DLBCL in the frontline setting. *Blood.* 2023;141:2566–75.
- Oberic L, Peyrade F, Puyade M, Bonnet C, Dartigues-Cuillères P, Fabiani B, et al. Subcutaneous rituximab-MiniCHOP compared with subcutaneous rituximab-MiniCHOP plus lenalidomide in diffuse large B-cell lymphoma for patients age 80 years or older. *J Clin Oncol.* 2021;39:1203–13.
- Tessoulin B, Depaus J, Feugier P, Fouillet L, Abraham J, Amorin S, et al. Verlen, "Very Elderly Rituximab Associated to Lenalidomide - Tafasitamab Combination in Frontline DLBCL Patients", a phase II open-label study evaluating efficacy of lenalidomide and tafasitamab combination associated to rituximab in frontline diffuse large B-cell lymphoma patients of 80 y/o or older from the Lysa Group. *Blood.* 2023;142:3094.
- Jerkeman M, Leppä S, Hamfjord J, Brown P, Ekberg S, José María Ferreri A. S227: initial safety data from the phase 3 POLAR BEAR trial in elderly or frail patients with diffuse large cell lymphoma, comparing R-pola-mini-CHP and R-mini-CHOP. *Hemasphere.* 2023;7:e91359ec.
- Brem EA, Li H, Beaven AW, Caimi PF, Cerchietti L, Alizadeh A, et al. SWOG 1918: A Phase II/III randomized study of R-miniCHOP with or without oral azacitidine (CC-486) in participants age 75 years or older with newly diagnosed aggressive non-Hodgkin lymphomas—aiming to improve therapy, outcomes, and validate a prospective frailty tool. *J Geriatr Oncol.* 2022;13:258–64.
- Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials—a primer for researchers. *Clin Epidemiol.* 2020;12:457–67.
- Lambert J, Lengliné E, Porcher R, Thiébaud R, Zohar S, Chevret S. Enriching single-arm clinical trials with external controls: possibilities and pitfalls. *Blood Adv.* 2022;7:5680–90.
- Cucherat M, Laporte S, Delaitre O, Behier JM, d'Andon A, Binlich F, et al. From single-arm studies to externally controlled studies. Methodological considerations and guidelines. *Therapies.* 2020;75:21–7.
- Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med.* 2009;28:3083–107.
- Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Stat Med.* 2014;33:1242–58.
- Allan V, Ramagopalan SV, Mardekian J, Jenkins A, Li X, Pan X, et al. Propensity score matching and inverse probability of treatment weighting to address confounding by indication in comparative effectiveness research of oral anticoagulants. *J Comp Eff Res.* 2020;9:603–14.
- Schulte PJ, Mascha EJ. Propensity score methods: theory and practice for anesthesia research. *Anesth Analg.* 2018;127:1074.
- Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med.* 2005;24:3089–110.
- Motrinchuk AS, Belousov D YU. Considerations for the design and conduct of externally controlled trials for drug and biological products. *MyRWD.* 2023;3:29–40.
- International Conference on Harmonization. ICH topic E10—Choice of control group in clinical trials. Note for guidance on choice of control group in clinical trials. (CPMP/ICH/364/96). London, UK: European Agency for the Evaluation of Medicinal Products; July 27, 2000.
- Ghesquières H, Rossi C, Cherblanc F, Le Guyader-Peyrou S, Bijou F, Sujobert P, et al. A French multicentric prospective prognostic cohort with epidemiological, clinical, biological and treatment information to improve knowledge on lymphoma patients: study protocol of the "REal world dAta in LYmphoma and survival in adults" (REALYSA) cohort. *BMC Public Health.* 2021;21:432.
- Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health.* 2010;13:273–7.
- Shiba K, Kawahara T. Using propensity scores for causal inference: pitfalls and tips. *J Epidemiol.* 2021;31:457–63.

28. Leyrat C, Seaman SR, White IR, Douglas I, Smeeth L, Kim J, et al. Propensity score analysis with partially observed covariates: How should multiple imputation be used?. *Stat Methods Med Res.* 2019;28:3–19.
29. Nguyen TQ, Stuart EA. Multiple imputation for propensity score analysis with covariates missing at random: some clarity on “within” and “across” methods. *Am J Epidemiol.* 2024;193:1470–6.
30. Ghesquières H, Cherblanc F, Belot A, Micon S, Bouabdallah KK, Esnault C, et al. Challenges for quality and utilization of real-world data for diffuse large B-cell lymphoma in REALYSA, a LYSA cohort. *Blood Adv.* 2024;8:296–308.
31. Yin X, Stuart E, Burcu M, Stewart M, Lamont E, Davi R. The validity of a synthetic control arm derived from historical multiple myeloma clinical trials. *Blood.* 2023;142:6628.
32. Van Le H, Van Naarden Braun K, Nowakowski GS, Sermer D, Radford J, Townsend W, et al. Use of a real-world synthetic control arm for direct comparison of lisocabtagene maraleucel and conventional therapy in relapsed/refractory large B-cell lymphoma. *Leuk Lymphoma.* 2023;64:573–85.
33. Rossi C, Marouf A, Deau Fischer B, Cherblanc F, Cugnod E, Chartier L, et al. First line therapy evaluation using propensity score approach in newly diagnosed advanced classical Hodgkin lymphoma patients from prospective real-world Realysa cohort and phase 3 AHL2011 trial. *Blood.* 2023;142:3056.

## ACKNOWLEDGEMENTS

The authors thank all the patients and their families for their confidence, as well as the LYSA investigators and the CRA teams, LYSARC staff, and the Hospices Civils de Lyon (sponsor of the REALYSA study). The initial studies (SENIOR, LNH09-7B, and REALYSA) were financially supported by Roche, Takeda, Janssen, Amgen, Celgene-Bristol Myers Squibb, Astra-Zeneca, AbbVie, and GSK.

## AUTHOR CONTRIBUTIONS

VL, BT, IC, and FC designed, coordinated the research, analyzed, interpreted the data, and wrote the manuscript. IC performed the statistical analyses. LC and AB advised on statistical analysis methods. HG, FP, SG, FB, PS, CR, L-MF, LF, SC, LC, AB, and LO commented on the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All original studies (SENIOR, LNH09-7B, and REALYSA) were approved by an independent research ethics committee and performed in accordance with the ICH-Good Clinical Practices Guidelines, the Declaration of Helsinki, and local regulatory requirements and laws. All studies were registered at [www.clinicaltrials.gov](http://www.clinicaltrials.gov) (SENIOR: NCT02128061; LNH09-7B: NCT01195714; REALYSA: NCT03869619). Each patient provided written informed consent for their participation. The current study is in compliance with French regulation concerning data reuse and approved by the National Ethics and Scientific Committee for Research (CESREES; Authorization # 8645681) and the National Data Protection Agency (CNIL) (Decision # DR-2023-150).

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41408-025-01374-x>.

**Correspondence** and requests for materials should be addressed to Benoit Tessoulin.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025