

Challenges for quality and utilization of real-world data for diffuse large B-cell lymphoma in REALYSA, a LYSA cohort

Hervé Ghesquière,¹ Fanny Cherblanc,² Aurélien Belot,² Sophie Micon,³ Krimo K. Bouabdallah,⁴ Cyril Esnault,³ Luc-Matthieu Fornecker,⁵ Katia Thokagevistik,³ Maxime Bonjour,² Fontanet Bijou,⁶ Corinne Haioun,⁷ Nadine Morineau,⁸ Loïc Ysebaert,⁹ Gandhi Damaj,¹⁰ Benoit Tessoulin,¹¹ Stéphanie Guidez,¹² Franck Morschhauser,¹³ Catherine Thiéblemont,¹⁴ Adrien Chauchet,¹⁵ Rémy Gressin,¹⁶ Fabrice Jardin,¹⁷ Christophe Fruchart,¹⁸ Gaëlle Labouré,¹⁹ Ludovic Fouillet,²⁰ Pauline Lionne-Huyghe,²¹ Antoine Bonnet,²² Laure Lebras,²³ Sandy Amorim,²⁴ Cécile Leyronnas,²⁵ Gaëlle Olivier,²⁶ Romain Guize,²⁷ Roch Houot,²⁸ Vincent Launay,²⁹ Bernard Drénou,³⁰ Olivier Fitoussi,³¹ Laurence Detournignies,³² Julie Abraham,³³ Carole Soussain,³⁴ Florence Lachenal,³⁵ Gian Matteo Pica,³⁶ Patrick Fogarty,² Pascale Cony-Makhoul,² Adeline Bernier,² Sandra Le Guyader-Peyrou,³⁷ Alain Monnereau,³⁷ Frédéric Boissard,³⁸ Cédric Rossi,³⁹ and Vincent Camus¹⁷

¹Department of Hematology, Hopital Lyon Sud, Claude Bernard Lyon 1 University, Pierre Benite, France; ²Lymphoma Academic Research Organisation, Hopital Lyon Sud, Pierre Benite, France; ³Roche SAS, Boulogne-Billancourt, France; ⁴Hematology and Cell Therapy Department, University Hospital of Bordeaux, Bordeaux, France; ⁵Institut de Cancérologie Strasbourg Europe (ICANS) and University of Strasbourg, Strasbourg, France; ⁶Department of Hematology, Institut Bergonie, Bordeaux, France; ⁷Lymphoid Malignancies Unit, Assistante Publique Hôpitaux de Paris APHP, Hopital Henri Mondor, Creteil, France; ⁸Department of Hematology, Centre Hospitalier Départemental Vendée, La Roche-sur-Yon, France; ⁹Institut universitaire du cancer de Toulouse Oncopole, Toulouse, France; ¹⁰Hematology Institute of Basse Normandie, Centre Hospitalier Universitaire de Caen, Caen, France; ¹¹Department of Hematology, Centre Hospitalier Universitaire de Nantes, Nantes, France; ¹²Department of Hematology, Centre Hospitalier Universitaire de Poitiers, Poitiers, France; ¹³Department of Hematology, Université de Lille, Centre Hospitalier Universitaire de Lille, Groupe de Recherche sur les formes Injectables et les Technologies Associées, Lille, France; ¹⁴Université Paris Cité, Assistante Publique Hôpitaux de Paris, Hôpital Saint-Louis, Service d'Hémo-Oncologie, Paris, France; ¹⁵Department of Hematology, Centre Hospitalier Universitaire de Besançon, Besançon, France; ¹⁶Department of Hematology, Centre Hospitalier Universitaire de Grenoble, Institute for Advanced Biosciences, INSERM U1209/CNRS UMR 5309/Grenoble Alpes University, Grenoble, France; ¹⁷Department of Clinical Hematology, INSERM U1245 Unit, Centre Henri Becquerel, Rouen, France; ¹⁸Department of Hematology, Centre Hospitalier de Dunkerque, Dunkirk, France; ¹⁹Department of Hematology, Centre Hospitalier de Libourne, Libourne, France; ²⁰Department of Hematology, Centre Hospitalier Universitaire de Saint Etienne, Saint Etienne, France; ²¹Department of Hematology, Centre Hospitalier d'Arras, Arras, France; ²²Department of Hematology, Centre Hospitalier de Bretagne Atlantique, Vannes, France; ²³Department of Hematology, Leon Berard Cancer Center, Lyon, France; ²⁴Department of Hematology, Hopital Saint Vincent de Paul, Lille, France; ²⁵Department of Hematology, Groupe Hospitalier Mutualiste de Grenoble, Grenoble, France; ²⁶Department of Hematology, Centre Hospitalier de Niort, Niort, France; ²⁷Department of Hematology, Centre Hospitalier Universitaire de Clermont Ferrand, Clermont Ferrand, France; ²⁸Department of Hematology, Centre Hospitalier Universitaire de Rennes, Rennes, France; ²⁹Department of Hematology, Centre Hospitalier de Saint Brieuc, Saint Brieuc, France; ³⁰Department Hematology, Groupe Hospitalier Mulhouse Sud Alsace, Mulhouse, France; ³¹Department of Hematology, Polyclinique Bordeaux Nord Aquitaine, Bordeaux, France; ³²Department of Hematology, Centre Hospitalier de Roubaix, Roubaix, France; ³³Department of Hematology, Centre Hospitalier Universitaire de Limoges, Limoges, France; ³⁴Department of Hematology, Institut Curie, Saint-Cloud, France; ³⁵Department of Hematology, Centre Hospitalier Pierre Oudot, Bourgoin-Jallieu, France; ³⁶Department of Hematology, Centre Hospitalier Metropole Savoie, Chambéry, France; ³⁷Registre des Hémopathies Malignes de la Gironde, Institut Bergonié, University of Bordeaux, Inserm, Team EPICENE, Bordeaux, France; ³⁸F. Hoffmann-La Roche Ltd, Basel, Switzerland; and ³⁹Department of Hematology, Centre Hospitalier Universitaire de Dijon Bourgogne, Dijon, France

Key Points

- The REALYSA cohort is a source of RWD of high quality for lymphoma thanks to a multistep rigorous data validation process.
- Effectiveness results on patients with first-line DLBCL seem consistent with literature and recent CTs.

Real-world data (RWD) are essential to complement clinical trial (CT) data, but major challenges remain, such as data quality. REal world dAta in LYmphoma and Survival in Adults (REALYSA) is a prospective noninterventional multicentric cohort started in 2018 that included patients newly diagnosed with lymphoma in France. Herein is a proof-of-concept analysis on patients with first-line diffuse large B-cell lymphoma (DLBCL) to (1) evaluate the capacity of the cohort to provide robust data through a multistep validation process; (2) assess the consistency of the results; and (3) conduct an exploratory transportability assessment of 2 recent phase 3 CTs (POLARIX and SENIOR). The analysis population comprised 645 patients with DLBCL included before 31 March 2021 who received immunochemotherapy and for whom 3589 queries were generated, resulting in high data completeness (<4% missing data). Median age was 66 years, with mostly advanced-stage disease and high international prognostic index (IPI) score. Treatments were mostly rituximab, cyclophosphamide, doxorubicin hydrochloride, vincristine, and

Submitted 30 May 2023; accepted 1 October 2023; prepublished online on *Blood Advances* First Edition 24 October 2023; final version published online 10 January 2024. <https://doi.org/10.1182/bloodadvances.2023010798>.

This study was presented in poster form at the 27th annual congress of the European Hematology Association, Vienna, Austria, 9-12 June 2022.

The data that support the findings of this study are available upon reasonable request from the corresponding author, Hervé Ghesquière (hervé.ghesquiere@chu-lyon.fr).

The full-text version of this article contains a data supplement.

© 2024 by The American Society of Hematology. Licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.

prednisone (R-CHOP 75%) and reduced dose R-CHOP (13%). Estimated 1-year event-free survival (EFS) and overall survival rates were 77.9% and 90.0%, respectively (median follow-up, 9.9 months). Regarding transportability, when applying the CT's main inclusion criteria (age, performance status, and IPI), outcomes seemed comparable between patients in REALYSA and standard arms of POLARIX (1-year progression-free survival 79.8% vs 79.8%) and SENIOR (1-year EFS, 64.5% vs 60.0%). With its rigorous data validation process, REALYSA provides high-quality RWD, thus constituting a platform for numerous scientific purposes. The REALYSA study was registered at www.clinicaltrials.gov as #NCT03869619.

Introduction

In France, lymphomas represent most of hematopoietic cancers and are the sixth and seventh most frequent cancers in men and women, respectively.¹ Diffuse large B-cell lymphoma (DLBCL) is the most frequent subtype, with ~5000 new cases every year.¹ Over the past 2 decades, the prognosis of patients with DLBCL has improved as a result of successive interventional trials investigating the intensity of chemotherapy, combinations of chemotherapy with monoclonal antibodies, or targeted therapies, and more recently the evaluation of chimeric antigenic receptor T cells and bispecific monoclonal antibodies in patients with relapsed/refractory disease.² However, although interventional trials remain the gold standard for evaluating new drugs or therapeutic strategies,² recent studies have highlighted limited trial participation.^{3,4} Indeed, <10% of patients are included in prospective trials, because of many factors, including trial availability in care center, organ function–based criteria, comorbidities, performance status (PS), and age.^{5–7} Using recent DLBCL prospective trial criteria, Khurana et al estimated that on the basis of organ function criteria alone, 9% to 24% of real-world patients with newly diagnosed DLBCL would be excluded from trial participation.³ Moreover, Loh et al showed that the number of eligibility criteria in DLBCL trials has increased over the past 30 years and that <50% of real-world patients are actually eligible for the most recent trials.⁴ Lastly, technical requirements (eg, mandatory imaging and tumor biopsy to screen for biomarkers) are time-consuming, can be complex to set up, and might delay protocol treatment initiation. Recently, in order to modernize trial enrollment criteria with the aim to shorten interval between diagnosis and treatment initiation (DTI), 1 US group of experts revised 31 eligibility criteria commonly used in DLBCL randomized clinical trials (RCTs) in rituximab, cyclophosphamide, doxorubicin hydrochloride, vincristine, and prednisone (R-CHOP) era and identified 13 essential and 9 unnecessary criteria, whereas no consensus could be reached for the 9 remaining ones.⁸ The DTI is a strong prognostic factor for patients with DLBCL with a short DTI and having worse prognostic factors and unfavorable outcomes,⁹ thus creating a potential significant selection bias in complex clinical trials (CTs). Overall, these studies highlight the need for a thorough assessment of the generalizability of CT results.

To this end, and in a context of ever greater accessibility to electronic health records, increasing focus is being given to real-world data (RWD).^{10,11} With rigorously managed RWD, noninterventional study designs can provide results that are a notable source

of additional information to that from RCTs.^{12,13} In the field of lymphoma, RWD have, for instance, proven useful in developing clinical scores, defining new outcome end points, assessing the role of imaging in patient follow-up, and addressing long-term toxicities.^{14–16} Interestingly, some prognostic scores and outcome end points were developed using RWD and subsequently validated on data from interventional trials.^{17–21} RWD may come from various sources (eg, institution databases, administrative databases from health insurance, registries, cohorts, or directly from patients via connected electronic devices).^{12,13} The main challenge is to optimize data quality by maximizing information while minimizing missing data, measurement errors, or patients lost to follow-up.

To generate reliable RWD in lymphoma, the Lymphoma Study Association (LYSA) launched in 2018 an observational multicentric prospective cohort in metropolitan France, REALYSA (REal world dAta in LYmphoma and Survival in Adults), to evaluate the real-world prognosis of the 7 most common lymphoma subtypes.²²

In this study, we report the results of a proof-of-concept analysis on patients with DLBCL undergoing firstline treatment (1L) in the REALYSA cohort to (1) evaluate the capacity of a real-world program on lymphoma in France to provide robust data through a specific multistep data validation system; (2) assess whether the characteristics of the population, clinical practices, and estimated effectiveness are consistent with what is expected in real life; and (3) conduct an exploratory assessment of the transportability of recent prospective phase 3 trials evaluating novel 1L agents for patients with DLBCL.^{23,24}

Methods

REALYSA cohort

The REALYSA study was registered at www.clinicaltrials.gov (NCT03869619) and approved by ethics committee, and written informed consent was obtained from all patients. Patients were informed of this specific analysis through a dedicated webpage on the LYSA website. The cohort methodology has been described elsewhere.²² In brief, patients were prospectively recruited in 1 of the 35 hematology centers after signing an informed consent form. Inclusion criteria were the following: aged >18 years, diagnosed with lymphoma in the previous 6 months and before treatment initiation. Patients were managed according to physician's choice, and there were no compulsory visits for the study. Clinical and treatment data at diagnosis for 1L and potential subsequent treatment lines were extracted from medical records.

Data entry checks

Real-time data checks were programmed within the electronic data capture (EDC) system in order to avoid obvious mistakes in data entry: (1) for all numerical variables, a predefined range was programmed, and if a value out of predefined range is entered, a warning message appears on the screen and the value has to be validated and (2) for all variables, an expected format is defined, and if the value is out of the expected format, data entry is blocked and the value has to be modified.

Data validation process and QC

A multistep data validation process inspired by RCT processes but adapted to the RWD constraints was set up to check the internal validity of the data (supplemental Figure 1). The objective was to maximize automation of consistency tests to minimize human resources.

First (level 1), data completeness was assessed with automated tests on preidentified data within the EDC system. Variables were sorted into 3 levels of importance that guide the insistence to obtain the data through queries in case of missing data. Simple consistency checks were also performed automatically within the EDC system to identify different types of inconsistency, such as chronological inconsistency (eg, biopsy date after the start date of treatment), discrepancies between treatment response and the Deauville score, or Ann-Arbor staging inconsistent with reported nodal localizations. Queries are automatically generated and submitted by the data manager. Some 17 000 completeness checks (305 variables throughout) and >35 000 basic consistency checks (645 checks reproduced) were programmed.

Second (level 2), more advanced consistency checks were performed using statistical analysis system software. All inconsistencies were reviewed by the operational study team, and queries generated if deemed necessary. For instance, inconsistencies in the chronology of response evaluation could be identified (eg, a patient reported with stable disease after being in complete response; see supplemental Table 1 for additional examples). When consistency checking could not easily be programmed, a manual review of listings was performed. For example, disease stage for patients with extranodal involvement or treatment patterns were reviewed manually. This second level of validation included 41 advanced checks (baseline characteristics, 19; treatment and follow-up, 22), manual review of 19 free-text entries, and a review of 2 listings.

Last (level 3), to assess the plausibility of patient care pathway, given the baseline characteristics, data were reviewed by LYSA clinicians using patient profiles that were generated using R software (Figure 1). This ad hoc tool was developed to automatically provide a summary of each patient (characteristics, therapeutic management, and evolution over time), allowing for time-efficient data validation. Additional queries could be sent if deemed necessary.

Simultaneously, to check the consistency between data in the electronic case report form (eCRF) and patients' medical records, external quality control (QC) is being implemented for 5% of the patients (included for ≥ 1 year and considered

as validated according to the automatic tests). QC focuses on critical data selected according to 3 criteria: (1) regulatory impact (consent form and inclusion or exclusion criteria), (2) impact on end points (diagnosis date, start date for each treatment line, progression date, and death date), (3) robustness of the data (an imaging examination date will be chosen over a clinical examination date; see supplemental Table 2 for a full list of variables). A concordance rate per patient per center per sample will be used to assess the overall quality of the database and trigger corrective actions if deemed necessary.

As described in the study protocol article,²² an approach based on continuous improvement was implemented with centers, with various tools and regular meetings, to continuously improve data quality and optimize human time.

DLBCL population

The following patients were considered for the study: patients included in REALYSA before 31 March 2021, diagnosed with DLBCL according to World Health Organization classification 2016, and treated with immunochemotherapy (IC-treated) as 1L. To focus the analysis on a real-world population, a non-interventional IC-treated population (hereafter referred to as analysis population) was selected by excluding patients enrolled in a 1L interventional CT.

In addition, to assess the transportability²⁵ of the control arms of 2 recent phase 3 trials (ie, POLARIX²³ and SENIOR²⁴ trials), "POLARIX-like" (P-L) and "SENIOR-like" (S-L) populations were defined by selecting patients who were treated with the standard of care of the control arm (R-CHOP regimen for P-L and reduced dose R-CHOP [R-miniCHOP] regimen for S-L) and met the main inclusion criteria of the trials (POLARIX: age, 18-80 years at treatment initiation; and baseline international prognostic index [IPI] score, 2-5; Eastern Cooperative Oncology Group-Performance Status score, 0-2 and SENIOR: age ≥ 80 years; Ann-Arbor stage, II-IV; and PS, 0-2).

End points of interest

The primary end point was event-free survival (EFS), defined as the time between the start of the first line and either a progression, relapse, new treatment line, or death, whichever occurred first. We also investigated progression-free survival (PFS), overall survival (OS), and end-of-treatment response.

Statistical analysis

Response probabilities were expressed as percentages, with 95% confidence intervals (CIs) calculated according to Exact Pearson-Clopper method. For PFS, EFS, and OS, Kaplan-Meier estimator was used to estimate probabilities of occurrence of a given end point at specific time points (with their 95% CI). Median follow-up was estimated using reverse Kaplan-Meier method. To assess the transportability of standard arms of phase 3 trials, patient characteristics and outcomes for the P-L and S-L populations were described. The analysis was conducted by A. Belot and P.F. in January 2022, using data exported on 15 November 2021 and a data cut-off set on 30 June 2021. Access to primary data was possible for all the authors.

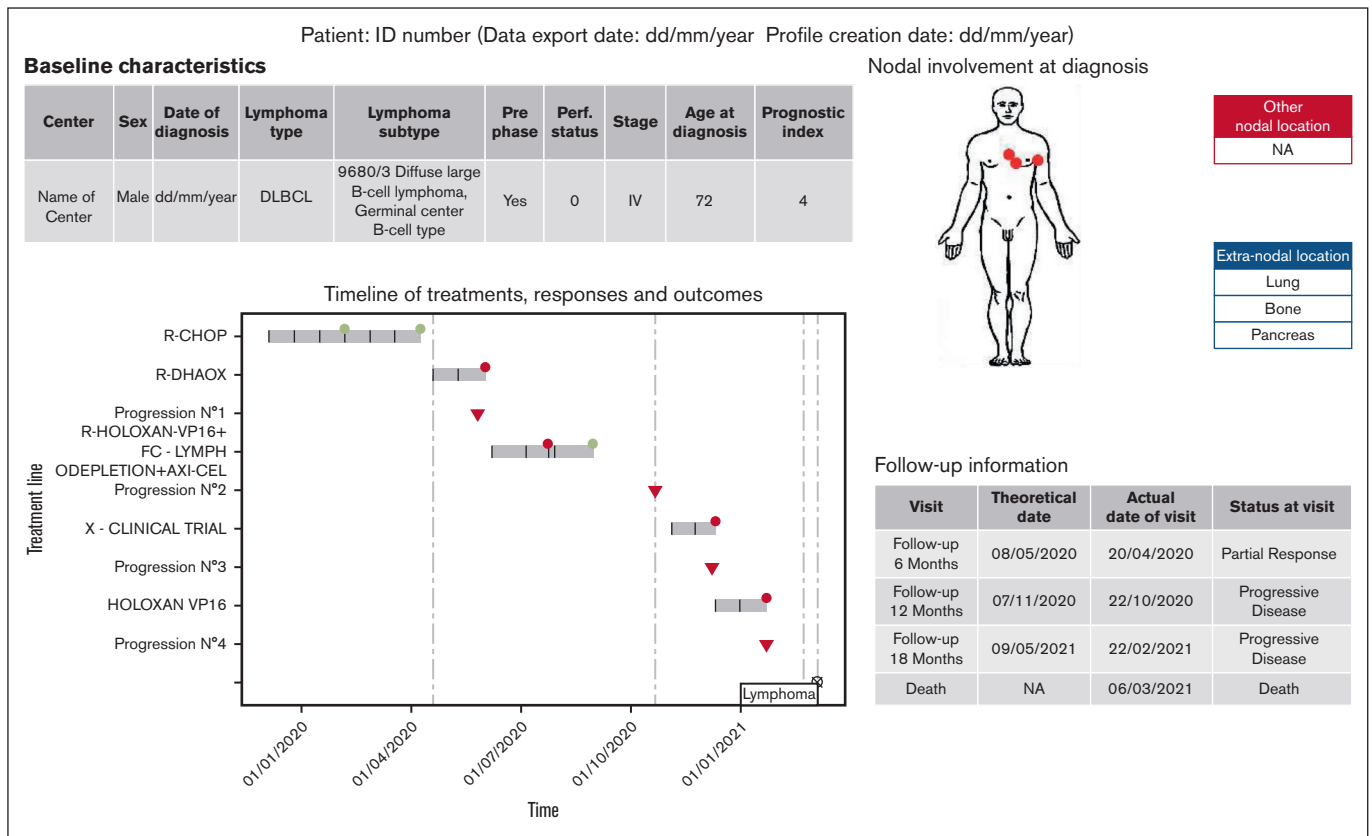


Figure 1. Automated patient profile. The patient profile is divided into 4 main parts. In the upper left part, the patient and tumor characteristics at inclusion are described. In the upper right part, the involvements (nodal and extra-nodal) at diagnosis are reported. Red circles are automatically located on the man/woman (depending on patient's sex) to represent nodal involvements. Nodal involvements reported in the "other" section of the eCRF appear in the red box in the upper right part. Extra-nodal involvements are detailed in the blue box in the lower right part. In the graph on the left, the longitudinal information per line(s) of treatment is indicated, with a gray horizontal bar per line of treatment (the treatment cycles are symbolized by black vertical lines), with the evaluations of the responses (circle above the line with color code according to the response: green for complete response, light green for partial response, orange for stable disease, and red for progressive disease) and the events (progression [inverted red triangle], adverse events [red cross, not present for this patient], and death [crossed out circle]). Note: all colors and symbols are not depicted in this patient profile example. Finally, in the lower right part, follow-up information is reported.

Results

Data validation and quality report

Whole REALYSA population. For the whole REALYSA population, over an 18-month period between 2021 and mid-2022 when all validation tests were running, during which 2700 new patients were recruited and 1300 patients were in follow-up, 7642 queries were sent (Table 1), with the following distribution (supplemental Figure 1): 50% were automated queries generated within the EDC software (level 1) and 50% resulted from advanced consistency testing, using statistical analysis system and manual or medical reviews (levels 2 and 3). Half of the queries (56%) concerned the baseline-data section, 30% treatment data, and 14% follow-up data. After a query, data were modified in ~75% of cases.

DLBCL population of this study. Regarding the analysis population, 3589 queries were transmitted, among which 99% were answered by the centers before data export. The analysis database comprised information on 100% of patients for most variables, including lymphoma subtype, date of diagnosis, and the presence or absence of extranodal involvement (Table 2). Disease stage information was missing for only 1 patient (0.2%). IPI class was

available for 96% of patients, and PS score and lactate dehydrogenase levels for >97% of patients.

DLBCL patient characteristics

The flowchart is presented in Figure 2. Overall, 700 patients with DLBCL were selected from the REALYSA cohort. Three patients (age: 87, 87, and 93 years) who had not received IC were excluded (treated with palliative treatment with rituximab monotherapy and oral cyclophosphamide), leaving an IC-treated population of 697 patients. Of these, 52 (7.5%) enrolled in interventional CTs were excluded, leaving an analysis population of 645 patients. Patients were recruited from 34 centers across 61 French departments. The majority (n = 543; 78% of the IC-treated population) were recruited from university hospitals, 71 (10%) from general hospitals, 70 (10%) from cancer centers, and 13 (2%) from private clinics.

Overall, patient characteristics in the 697 patients of the IC-treated population and the analysis population (N = 645) were comparable (Table 2). Regarding the analysis population, median age was 66 years (range, 19-98 years), and 344 patients were male (53%), with advanced-stage disease (Ann-Arbor stage III/IV, n = 472

Table 1. Number and repartition of queries

	Baseline		Treatment		Follow-up		Whole CRF	
	N	% (levels)	N	% (levels)	N	% (levels)	N	% (levels)
Level 1 queries	2426	57	893	39	480	44	3799	50
Level 2 and 3 queries	1839	43	1385	61	619	56	3843	50
Total	4265	100	2278	100	1099	100	7642	100
Repartition per section	56%		30%		14%			

Bold values are total values, either per query level in the whole eCRF or per eCRF section.

patients; 73%), extranodal locations (n = 499; 77%), elevated lactate dehydrogenase levels (n = 402; 64%), and high IPI score (IPI 2-5, n = 486; 76%). The main histological subtypes were DLBCL (n = 480; 74%), high-grade B-cell lymphoma (n = 81; 13%), and primary mediastinal B-cell lymphoma (PMBL) (n = 45; 7%). Positron emission tomography/computed tomography scan (PET scan) was performed for most patients for initial workup (606 out of 642 patients with available data; 94%).

Treatment patterns

For almost all patients (630 of 645; 98%), 1L was anthracycline-based, with either (1) R-CHOP (n = 482; 75%), (2) R-miniCHOP (n = 86; 13%), or (3) high-dose anthracycline-based regimen (n = 62; 10%, of whom 57 (92%) with R-ACVBP (rituximab, doxorubicin, cyclophosphamide, vindesine, bleomycin, and prednisone). Only 15 patients (2%) received a nonanthracycline-based regimen (see supplemental Table 3 for details). The median age was 65 years (range, 20-90 years) for patients receiving R-CHOP, 83 years (range, 71-95 years) for patients receiving R-miniCHOP, and 40 years (range, 19-67 years) for patients receiving high-dose anthracycline regimens (Table 2). Among patients treated with high-dose anthracycline regimens (n = 62), most were treated in university hospitals (n = 55; 89%) and one-third (n = 21; 34%) had a PMBL subtype. Only 9 patients (1.4%) underwent consolidation therapy with autologous stem cell transplantation (ASCT). Of these, 5 were patients with PMBL, 2 had central nervous system involvement at diagnosis and received a rituximab, high-dose methotrexate, doxorubicin, cyclophosphamide, vincristine, and prednisone regimen before ASCT, and the remaining 2 patients received ASCT after R-CHOP in combination with high-dose methotrexate, according to physician's choice. Radiotherapy was used as consolidation therapy for 19 patients (3%), of whom 8 (42%) had a localized disease (stage I/II) and 7 (37%) a gonadal involvement. Regarding the number of cycles of R-CHOP, 278 patients (58%) received 6 cycles and 102 (21%) 8 cycles (supplemental Table 4). Among the 63 patients with an age-adjusted IPI (aalIPI) of 0, 24 (38%) received 4 cycles of R-CHOP. Overall median DTI was 26.0 days (range, 0-132). Median DTI was similar for patients receiving R-CHOP and R-miniCHOP (26.0 [range, 0-132] and 25.5 [range, 4-84] days, respectively), shorter for patients receiving high-dose anthracycline regimen (19.0 days [range, 0-60]), and longer for patients receiving non-anthracycline-based regimen (35.0 days [range, 6-71]).

Treatment response

Treatment response at the end of 1L therapy was documented for 603 patients (94%). Of these, 483 (80%; 95% CI, 76.7-83.2) had

a complete response, 51 (9%; 95% CI, 6.4-11.0) a partial response (overall response rate [ORR] of 89%; 95% CI, 85.7-91.0), 7 (1.2%, 95% CI, 0.5-2.4) stable disease, and 62 (10%, 95% CI, 8.0-13.0) progressive disease. For most patients (n = 556; 92%), the treatment response was assessed using PET scans.

Outcomes

The median follow-up period from treatment initiation was 9.9 months (range, 0.4-30.5) for the analysis population. Among the 645 patients included in outcome evaluation, 123 EFS events were documented (new treatment initiation for 15 patients [2.3%], progression/relapse for 86 patients [13.3%], and death for 22 patients [3.4%]). Median survival was not reached. The 1-year OS rate was 90.0% (95% CI, 86.5-92.5). The 1-year EFS rate was 77.9% (95% CI, 73.8-81.4). The prognostic value of IPI and aalIPI²⁶ seemed to be reproduced in this RWD set (Figure 3), with the limitation of the short follow-up.

Patient representativeness: comparison with national data registry

As an indicator of representativeness, the analysis population data were compared with national DLBCL incidence data,¹ according to age group and sex (supplemental Figure 2). It suggests a rather good level of comparability in terms of age distribution, though with a slight underrepresentation of older patients (aged >80 years), for women in particular.

Transportability of DLBCL population data: CTs vs REALYSA

A P-L population (n = 320) was isolated from REALYSA, selecting patients with DLBCL who fulfilled the main inclusion criteria of the POLARIX trial and were treated with R-CHOP.²³ This P-L population seemed comparable with the R-CHOP arm population of the POLARIX trial (Table 2): median age, 67 vs 66 years; IPI score, 3-5 in 65.7% vs 62%, respectively. Median DTI was 26.0 days (range, 0-132) for the P-L population and 27 days in the POLARIX R-CHOP arm. The ORR was 89% in real-world P-L population and 84% in the POLARIX R-CHOP arm. The 1-year PFS rates were 79.8% (95% CI, 73.9-84.4) in the R-CHOP control arm of POLARIX²³ and 79.8% (95% CI, 75.9-83.6) in the patients of the P-L population.

Regarding S-L population (n = 59), characteristics also seemed rather comparable with patients from the control arm of the SENIOR trial²⁴ (Table 2): median age, 83.4 vs 83.0 years; PS score, 2 to 4 in 28.8% vs 28.0%; IPI score, 3 to 5 in 79.0% vs 75.0%, respectively. Median DTI was 27.0 days (range,

Table 2. Demographic and clinical characteristics at baseline for IC-treated population, analysis population (total and according to 1L treatment received), P-L population, POLARIX R-CHOP treatment arm, S-L population, and SENIOR trial

	IC-treated population N = 697	Analysis population N = 645					REALYSA P-L population n = 320	POLARIX R-CHOP arm ²³ n = 439	REALYSA S-L population n = 59	SENIOR R-miniCHOP arm ²⁴ n = 127
		Total n = 645	R-high-dose anthracycline n = 62 (9.6%)	R-CHOP n = 482 (74.7%)	R-miniCHOP n = 86 (13.3%)	R-nonanthracycline-based chemotherapy n = 15 (2.3%)				
Male sex	371 (53.2%)	344 (53.3%)	33 (53.2%)	260 (53.9%)	40 (46.5%)	11 (73.3%)	174 (54.4%)	234 (53.3%)	32 (54.2%)	56 (44.0%)
Median age at diagnosis (min-max), y	67.0 (19-98)	66.3 (19-98)	40.1 (19-67)	65.2 (20-90)	82.5 (71-95)	82.3 (60-98)	67.4 (24-80)	66.0 (19-80)	83.4 (80-95)	83.0 (80-96)*
Age at diagnosis, y										
≤60	238 (34.1%)	235 (36.4%)	61 (98.4%)	173 (35.9%)	0 (0.0%)	1 (6.7%)	78 (24.4%)	131 (29.8%)	0 (0.0%)	0 (0.0%)
60 to 80	373 (53.5%)	325 (50.4%)	1 (1.6%)	301 (62.4%)	18 (20.9%)	5 (33.3%)	242 (75.6%)	308 (70.2%)	0 (0.0%)	0 (0.0%)
>80	86 (12.3%)	85 (13.2%)	0 (0.0%)	8 (1.7%)	68 (79.1%)	9 (60.0%)	0 (0.0%)	308 (70.2%)	59 (100.0%)†	127 (100.0%)†
ECOG PS‡										
Data available	680 (97.6%)	628 (97.3%)	61 (98.4%)	470 (97.5%)	84 (97.7%)	13 (86.7%)	320 (100.0%)	439 (100.0%)	59 (100.0%)	127 (100.0%)
2 to 4	128 (18.8%)	125 (19.9%)	7 (11.5%)	79 (16.8%)	33 (39.3%)	6 (46.2%)	56 (17.5%)	75 (17.1%)	17 (28.8%)	36 (28.0%)
Main histological subtypes										
DLBCL	521 (74.7%)	480 (74.4%)	32 (51.6%)	369 (76.6%)	69 (80.2%)	10 (66.7%)	244 (76.3%)	367 (83.6%)	47 (79.7%)	NA
PMBL	46 (6.6%)	45 (7.0%)	21 (33.9%)	24 (5.0%)	0 (0.0%)	0 (0.0%)	9 (2.8%)	0 (0.0%)	0 (0.0%)	
HGBL	85 (12.2%)	81 (12.6%)	5 (8.1%)	61 (12.7%)	11 (12.8%)	4 (26.7%)	45 (14.1%)	50 (11.4%)	9 (15.2%)	
Other large B-cell lymphoma	45 (6.5%)	39 (6.0%)	4 (6.4%)	28 (5.8%)	6 (7.0%)	1 (6.7%)	22 (6.8%)	22 (5.0%)	3 (5.1%)	
Low grade component at diagnosis	42 (6.0%)	41 (6.4%)	2 (3.2%)	31 (6.4%)	7 (8.1%)	1 (6.7%)	20 (6.3%)	NA	5 (8.5%)	NA
PET scan performed at diagnosis										
Data available	694 (99.6%)	642 (99.5%)	62 (100.0%)	480 (99.6%)	85 (98.8%)	15 (100.0%)	318 (99.4%)	NA	58 (98.3%)	NA
Yes	658 (94.8%)	606 (94.4%)	62 (100.0%)	464 (96.7%)	70 (82.4%)	10 (66.7%)	308 (96.9%)		47 (81.0%)	
Ann-Arbor stage‡										
Data available	696 (99.9%)	644 (99.8%)	62 (100.0%)	481 (99.8%)	86 (100.0%)	15 (100.0%)	320 (100.0%)	439 (100.0%)	59 (100.0%)	127 (100.0%)
III to IV	518 (74.4%)	472 (73.3%)	41 (66.1%)	350 (72.8%)	70 (81.4%)	11 (73.3%)	291 (90.9%)	387 (88.2%)	52 (88.1%)	105 (83.0%)
IPI class‡										
Data available	671 (96.3%)	619 (96.0%)	60 (96.8%)	466 (96.7%)	81 (94.2%)	12 (80.0%)	320 (100.0%)	439 (100.0%)	57 (96.6%)	126 (99.2%)
0 to 1	155 (23.1%)	152 (24.6%)	23 (38.3%)	121 (26.0%)	7 (8.6%)	1 (8.3%)	0 (0.0%)	0 (0.0%)	3 (5.3%)	32 (25.0%)
2	153 (22.8%)	141 (22.8%)	15 (25.0%)	110 (23.6%)	12 (14.8%)	4 (33.3%)	110 (34.3%)	167 (38.0%)	9 (15.8%)	32 (25.0%)
3	167 (24.9%)	148 (23.9%)	18 (30.0%)	110 (23.6%)	19 (23.5%)	1 (8.3%)	210 (65.7%)	272 (62.0%)	16 (28.1%)	94 (75.0%)
4-5	196 (29.2%)	178 (28.8%)	4 (6.7%)	125 (26.8%)	43 (53.1%)	6 (50.0%)	210 (65.7%)	272 (62.0%)	29 (50.9%)	94 (75.0%)

ECOG PS, Eastern Cooperative Oncology Group-Performance Status; HGBL, high-grade B-cell lymphoma; NA, not applicable.

*Median age of the whole trial population (ie, both arms, data not available for R-miniCHOP arm only). If "no data available" is not mentioned, it means that data were available for 100% of patients.

†≥80 years of age.

‡When missing data are reported, the percentage of available data is reported, and the distribution of the variable is expressed as a percentage of available data; LDH; R-CHOP; R-mini; R-nonanthracycline-based chemotherapy regimen: R-CEOP (cyclophosphamide, etoposide, vincristine, and prednisone), R-CEP (cyclophosphamide, etoposide, and prednisone), R-COP (cyclophosphamide, vincristine, and prednisone), R-COP + gemcitabine, R-CVP (cyclophosphamide, vincristine, and prednisone), R-CVP + gemcitabine, R-GEMOX (gemcitabine and oxaliplatin), R-ifosfamide, and etoposide.

§Extranodal involvement includes spleen.

Table 2 (continued)

	IC-treated population N = 697	Analysis population N = 645						REALYSA P-L population n = 320	POLARIX R-CHOP arm ²³ n = 439	REALYSA S-L population n = 59	SENIOR R-miniCHOP arm ²⁴ n = 127
		Total n = 645	R-high-dose anthracycline n = 62 (9.6%)	R-CHOP n = 482 (74.7%)	R-miniCHOP n = 86 (13.3%)	R-nonanthracycline-based chemotherapy n = 15 (2.3%)	REALYSA P-L population n = 320				
Data available	662 (95.0%)	610 (94.6%)	59 (95.2%)	458 (95.0%)	81 (94.2%)	12 (80.0%)	315 (98.4%)	NA	57 (96.6%)	NA	
0	77 (11.6%)	76 (12.5%)	2 (3.4%)	66 (14.4%)	8 (9.9%)	0 (0.0%)	0 (0.0%)	4 (7.0%)			
1	213 (32.2%)	197 (32.3%)	22 (37.3%)	153 (33.4%)	17 (21.0%)	5 (41.7%)	102 (32.4%)	14 (24.6%)			
2	277 (41.8%)	244 (40.0%)	30 (50.8%)	181 (39.5%)	30 (37.0%)	3 (25.0%)	173 (54.9%)	25 (43.9%)			
3	95 (14.4%)	93 (15.2%)	5 (8.5%)	58 (12.7%)	26 (32.1%)	4 (33.3%)	40 (12.7%)	14 (24.6%)			
Extranodal involvement ^S	540 (77.5%)	499 (77.4%)	39 (62.9%)	372 (77.2%)	75 (87.2%)	13 (86.7%)	276 (86.3%)	NA	53 (89.8%)	NA	
Elevated LDH level at baseline											
Data available	680 (97.6%)	628 (97.4%)	60 (96.8%)	471 (97.7%)	83 (96.5%)	14 (93.3%)	315 (98.4%)	438 (99.8%)	57 (96.6%)	NA	
Yes	441 (64.9%)	402 (64.0%)	52 (86.7%)	284 (60.3%)	58 (69.9%)	8 (57.1%)	226 (71.7%)	284 (64.8%)	39 (68.4%)		

ECOG PS, Eastern Cooperative Oncology Group-Performance Status; HGBL, high-grade B-cell lymphoma; NA, not applicable.

*Median age of the whole trial population (ie, both arms, data not available for R-miniCHOP arm only). If "no data available" is not mentioned, it means that data were available for 100% of patients.

[†]≥80 years of age.

[#]When missing data are reported, the percentage of available data is reported, and the distribution of the variable is expressed as a percentage of available data; LDH: R-CHOP; R-mini: R-nonanthracycline-based chemotherapy regimen: R-CEOP (cyclophosphamide, etoposide, vincristine, and prednisone), R-CEP (cyclophosphamide, etoposide, and prednisone), R-COP (cyclophosphamide, vincristine, and prednisone), R-COP + gemcitabine, R-GEMOX (gemcitabine and oxaliplatin), R-ifosfamide, and etoposide.

^SExtranodal involvement includes spleen.

5-80 days) for S-L population and 33 days (range, 8-89 days) for R-miniCHOP arm of SENIOR trial. The 1-year EFS was 64.5% (95% CI, 47.8-77.0) and 60.0% (95% CI, 50.8-68.1) in the patients of the S-L population and in the R-miniCHOP arm of SENIOR trial,²⁴ respectively. The 1-year OS rates were 78.3% (95% CI, 61.4-88.5) and 78.5% (95% CI, 70.2-84.7) for the S-L population and R-miniCHOP arm of SENIOR trial,²⁴ respectively.

Discussion

This proof-of-concept analysis of REALYSA patients with DLBCL demonstrate how a nationwide prospective real-world cohort can provide comprehensive robust data on baseline characteristics of patients and treatment effectiveness. Ensuring data quality in a large prospective observational cohort, such as REALYSA is extremely challenging. Tailored processes are essential to ensure effective data management and validation. Furthermore, data collection needs to be regularly updated to keep up with changes in clinical practice. For example, when REALYSA was initiated (2018), chimeric antigenic receptor T-cell therapy was only used as third-line therapy of DLBCL in CT settings, whereas it is now positioned as a second-line treatment for patients with relapsed/refractory DLBCL.²⁷⁻²⁹ The eCRF has been tailored to include information regarding the implementation of these new therapies in daily practice. The data validation process is based on experience from CT and adapted to the constraints of high inclusion rates (~140 patients per month) from multiple recruitment sites, limited human resources, and the necessity to initiate data analysis before the end of the study. Efforts were focused primarily on internal validation of critical data to limit the number of queries but still maintain an appropriate level of quality. For the analysis population (N = 645), 3589 queries were sent to centers, leading to data adjustments in most cases and, consequently, notable improvements in data quality. Despite the inherent limitations of comparing trials with different datasets, it seems that the number of queries generated in our study was similar to the one of an academic phase 3 trial (LNHO9-1B, N = 650 patients and 5180 queries) and 7 to 9 times lower than those of industrial phase 3 trials (REMAR trial,³⁰ N = 650 patients and 31 756 queries; GAINED trial,³¹ N = 671 patients and 26 152 queries). Thus, although the REALYSA validation process may be less stringent than the processes applied to pharmaceutical industry-sponsored CT, it seems rigorous enough to generate meaningful robust data with very few missing data on key variables (<4%). Data validation processes require considerable resources but have a strong impact on data quality. Depending on the situation, dedicating such resources for data validation may not always be feasible, resulting in highly variable levels of RWD data quality. With the recent increase in the use of RWD as a complement to CT data for regulatory decision-making, the challenge is to ensure that data quality is suited to confidently inform on drug use or treatment effectiveness,³²⁻³⁴ in particular if RWD are to be used for marketing authorizations and not only for postmarketing studies. In line with this report, international guidelines stress the importance of clear descriptions of data verification processes.^{35,36} This is, however, rarely done in the literature.³⁷⁻³⁹ However, our process could be improved, notably in terms of verification of source data (ie, medical records). External QC to compare source data with the eCRF data is scheduled for 2023. Future processes will also include reviews of pathology reports by

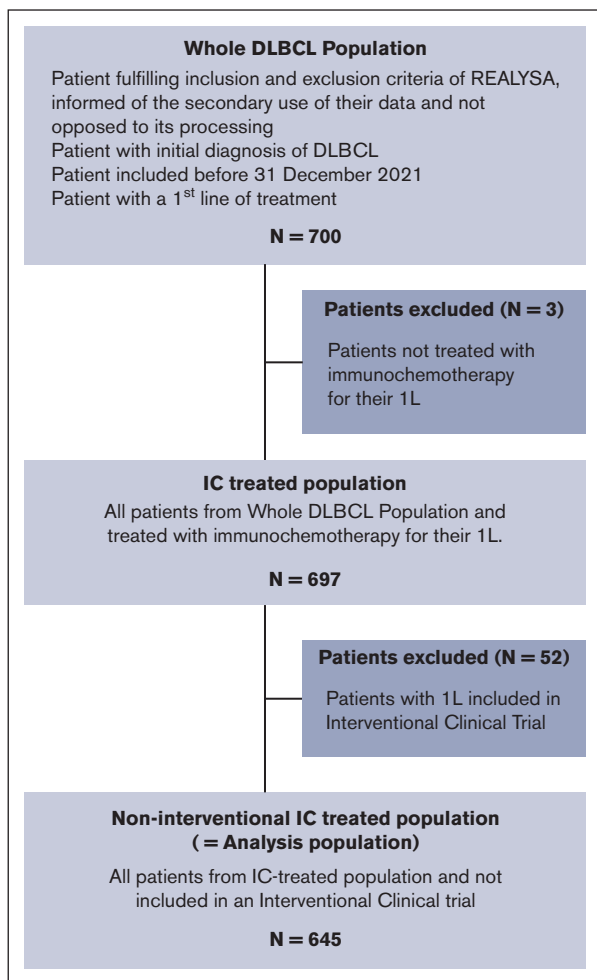


Figure 2. Study flowchart.

hematopathologists from LYSA and LymphoPath networks to reinforce diagnostic accuracy.⁴⁰ Such improvements are key to improve data quality and increase our knowledge of lymphoma biology in real-world settings.⁴¹⁻⁴⁴

The current analysis shows trends in care provided to patients with DLBCL in hematology departments in France. Virtually all patients received curative-intent treatment, with only 3 and 15 patients treated without IC and without anthracycline-based chemotherapy, respectively. Most patients (75%) received R-CHOP, 13% R-miniCHOP, and 10% intensive chemotherapy (mainly rituximab, doxorubicin, cyclophosphamide, vindesine, bleomycin, and prednisone regimen), the latter being mainly younger patients, of whom one-third had the PMBL subtype.^{45,46} Similar observations were made in the US Molecular Epidemiology Resource cohort with 92.6% patients with DLBCL on IC.⁴³ However, population-based registries showed different observations. One Swedish cohort (2007-2014) reported that 14% of patients with DLBCL received noncurative intent therapy (nonanthracycline-based regimen).⁴⁷ Similar results were observed in the British Columbia Cancer registry (16% noncurative intent treatment).⁴⁸ In a study conducted by the Danish National Lymphoma Registry on 1011 older DLBCL patients (age, ≥ 75 years; 2003-2012), Juul et al reported palliative-intent treatment for 21% of patients.⁴⁹ These contrasting results

highlight the importance of considering inclusion bias, especially regarding older/frailer patients less likely to be treated by IC. Indeed, we observed differences in age distribution between the studied cohort and data from the French registry with a bias toward a younger population in our cohort. First, this difference may be explained by considering enrollment modalities. Participants in the REALYSA cohort were required to provide written informed consent, whereas registries were based on an opt-out system (patients were automatically registered). Clinicians may be reluctant to include patients who were frail with limited life-expectancy, a poor clinical condition, or cognitive disorders. Additionally, because participation in REALYSA also includes patient-reported outcomes from epidemiology questionnaires, patients who are old/frail may refuse to participate. This active inclusion process has an impact on patient selection with very unfit patients less likely to be included, thus suggesting that comparisons with registries should be made with caution. To address this selection bias, a specific report is periodically sent to each REALYSA center. It outlines the clinical characteristics of patients included locally and compares them with the global cohort data and national registry data (age, sex, and subtype distribution), thus highlighting the potential selection bias in that center. A nested study conducted in 1 REALYSA center showed that 54% of all patients with lymphoma (151 out of 278 patients) referred to this department of hematology over a 1-year period were included in REALYSA.⁵⁰ Among nonincluded patients ($n = 127$), the following reasons of noninclusion were identified: refusal for 39 (31%) patients, start of a treatment in emergency for 20 (16%) patients, physician's evaluation that inclusion in REALYSA was not feasible for 12 (9%) patients, unknown reason for 56 (44%) patients. The inclusion rate was, therefore, higher than those in CTs, but it confirms the need for a clearer understanding of recruitment bias in prospective real-world cohorts. Second, the bias toward a younger population may also be explained by the recruiting centers. With 35 centers, REALYSA could not ensure the coverage of all patients with DLBCL in France. There was likely an overrepresentation of large teaching hospitals in this report ($n = 543$, 77%), though there is a paucity of relevant French data in recent literature.⁵¹ Thus, REALYSA most likely reflects lymphoma practice in university hospitals in France rather than overall lymphoma practice in France. Further work based on local initiatives,⁵⁰ detailed comparison with registries and potentially with the French National Healthcare Data System may help contextualize these results with recent and more exhaustive data.

Regarding treatment, results are consistent with current guidelines and reflects the implementation of CT results into routine care. Strategies for reducing treatment intensity have been introduced in international guidelines and implemented into routine care,⁵²⁻⁵⁹ with 6 cycles of R-CHOP now the standard of care (58% of R-CHOP-treated patients) and 4 cycles recently implemented in the lowest aalPI group (38% of R-CHOP-treated patients with aalPI = 0). In line with CT results,⁵² only 19 patients (3%) received consolidation treatment with radiotherapy. Although high-dose chemotherapy followed by ASCT used to be a standard of care for patients with high-risk DLBCL in the LYSA group,^{31,60} only 9 patients (1.4%) underwent this procedure in this DLBCL cohort. Interestingly, in line with current guidelines, most patients underwent a PET scan to assess disease stage at diagnosis (94%) and at end of treatment (92%).^{2,53,61} Collecting accurate information

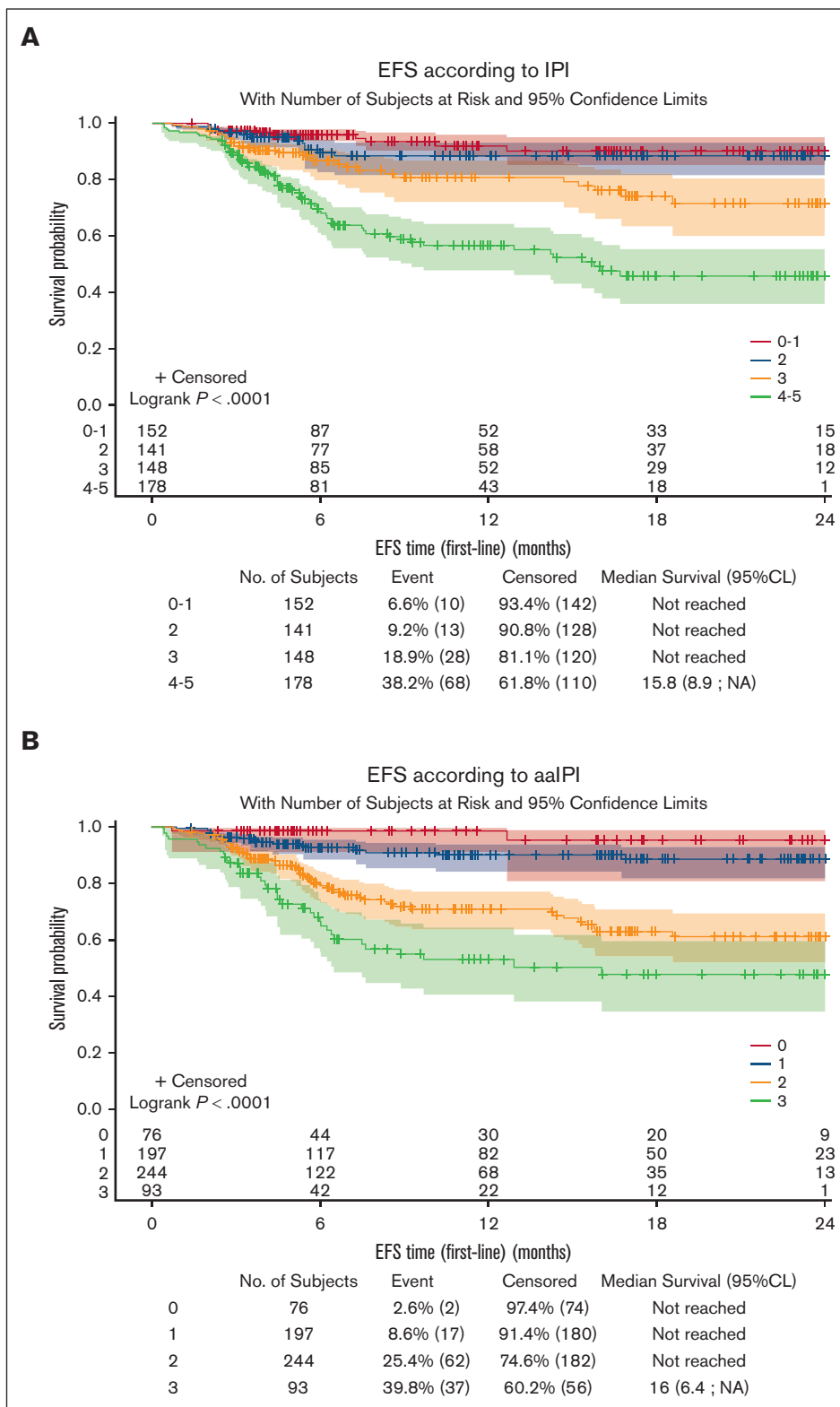


Figure 3. Prognostic value of IPI and aaIPI on event-free survival. EFS according to IPI (A) and aaIPI (B). CL, confidence limit.

on assessment modalities can be particularly challenging in observational cohorts.^{43,62} The REALYSA cohort data include information on the imaging techniques used. Imaging data can,

thus, be uploaded for specific research projects and notably interim PET scans.⁶³ Metabolic imaging and pretreatment circulating tumor DNA levels⁶⁴⁻⁶⁶ have recently been shown to provide

compelling information on patient outcomes. Serial biobanking in the REALYSA cohort ensures the feasibility of such studies.²²

Regarding outcomes, because the REALYSA program was initiated in 2018, follow-up was limited so far to ensure robustness on patient outcomes. Nevertheless, the prognostic value of the IPI and aalPI scores could be reproduced on this DLBCL population, with a plateau on survival curves observed 24 months after diagnosis, as previously reported for patients with DLBCL.¹⁴ Estimated 1-year OS rate in our analysis population was 90.0% (95% CI, 86.5-92.5). On the national scale based on registry data for patients diagnosed between 2010 and 2015, the estimated 1-year OS rate was 71% (95% CI, 70-72).⁶⁷ Comparison between REALYSA outcomes and national registry-based outcomes must be considered with caution. First, as mentioned earlier, REALYSA did not cover all hematologic centers in France and REALYSA centers were mainly university hospitals. Secondly, this study showed that the inclusion rate of patients with palliative care intent was low. Finally, the follow-up of our study was very short for OS estimations, most likely leading to an overestimation of the OS. Of course, the global improvement of OS between these 2 periods (2018-2021 vs 2010-2015) could also reflect the improvement of treatments mainly at relapse, because standard of care in first-line therapy did not change between these 2 periods.

Data from real-world settings are key to contextualizing the results of CTs, in particular regarding the transportability of CT results to the general population. Herein, we conducted an exploratory evaluation of the transportability of 2 CTs using data from REALYSA. First, we could see that when applying inclusion criteria of both CTs, POLARIX, and SENIOR,^{23,24} on REALYSA data, a significant number of patients could be identified from our real-life population, though lower for patients >80 years of age. The ORR and 1-year PFS seemed comparable between the real-world P-L population and the POLARIX control arm (88.8% vs 83.8% and 79.8% vs 79.8%, respectively), with the limitation of different assessment processes (ie, centralized review in the CT vs local review in the real-world cohort). The 1-year EFS and OS rates between S-L population and the SENIOR control arm also seemed comparable (64.5% vs 60% and 78.3% vs 78.5%, respectively), with the limitation of different imaging techniques used to assess treatment response (ie, PET scans in S-L population vs computed tomography scans in SENIOR²⁴ control arm). In both cases, there was also the limitation of different follow-up modalities in real-life vs interventional CTs (with more standardized imaging assessments in the latter). These data suggest that it is feasible with data from REALYSA to assess transportability of CTs. However, these preliminary results are only descriptive, as a proof-of-concept analysis, and must be considered with caution, considering the short median follow-up. Further work including adjustment on variables known to affect outcomes, using more complex statistical techniques (eg, propensity score), and data with longer follow-up should be considered to confirm these preliminary results.

This proof-of-concept study on REALYSA patients with 1L DLBCL suggests that this real-world cohort can generate high-quality data.

The resulting database has minimal missing data on key variables and results are consistent with existing literature in terms of baseline characteristics, treatments, and outcomes. REALYSA is a source of meaningful RWD with significant potential for a multitude of different applications, including a better characterization of the lymphoma population in France as well as innovative study designs, including (but not restricted to) new outcome endpoints or the creation of synthetic control arms.

Acknowledgments

The authors thank all of the patients and their families for their confidence and the Hospices Civils de Lyon (France) who sponsored the study, as well as the LYSA and LYSARC, who coordinate REALYSA, in collaboration with EPICENE (INSERM). The authors also thank all the investigators and the investigating centers.

The REALYSA study is funded by several commercial organizations (Roche, Takeda, Janssen, Amgen, Celgene-Bristol Myers Squibb). This study was specifically supported by F. Hoffmann-La Roche.

Authorship

Contribution: H.G., F.C., A. Belot, S.M., C.E., K.T., A. Bernier, F. Boissard, and V.C. conceptualized and designed the study; H.G., K.-K.B., L.-M.F., F. Bijou, C.H., N.M., L.Y., G.D., B.T., S.G., F.M., C.T., A.C., R. Gressin, F.J., C.F., G.L., L.F., P.L.-H., A. Bonnet, L.L., S.A., C.L., G.O., R. Guieze, R.H., V.L., B.D., O.F., L.D., J.A., C.S., F.L., G.M.P., S.L.G.-P., A.M., C.R., and V.C. managed the provision of study materials or patients; F.C., A. Belot, M.B., P.F., P.C.-M., and A. Bernier oversaw the data collection and assembly; A. Belot and P.F. conducted the statistical analysis; H.G., F.C., A. Belot, and A. Bernier wrote the manuscript; and all authors approved the final version of the manuscript and are accountable for all aspects of the work.

Conflict-of-interest disclosure: H.G. has been involved in consultancy and advisory boards for Roche, Gilead, Bristol Myers Squibb (BMS), and AbbVie. S.M., C.E., K.T., and F. Boissard are employees of Roche. F.M. serves on consultancy and advisory boards for Roche. V.C. has been a consultant and received honoraria from Roche, Janssen, and Celgene-BMS, in addition to receiving research and travel grants from Celgene-BMS. The remaining authors declare no competing financial interests.

ORCID profiles: A. Belot, [0000-0003-1410-5172](https://orcid.org/0000-0003-1410-5172); C.E., [0000-0003-2896-6045](https://orcid.org/0000-0003-2896-6045); F.B., [0000-0001-6768-7175](https://orcid.org/0000-0001-6768-7175); G.D., [0000-0002-4689-3882](https://orcid.org/0000-0002-4689-3882); B.T., [0000-0001-7600-3329](https://orcid.org/0000-0001-7600-3329); F.M., [0000-0002-3714-9824](https://orcid.org/0000-0002-3714-9824); A. Bonnet, [0000-0001-6962-3650](https://orcid.org/0000-0001-6962-3650); B.D., [0000-0002-2175-6785](https://orcid.org/0000-0002-2175-6785); A. Bernier, [0000-0003-3023-429X](https://orcid.org/0000-0003-3023-429X); C.R., [0000-0003-3717-7961](https://orcid.org/0000-0003-3717-7961).

Correspondence: Hervé Ghesquieres, Département d'Hématologie, Hôpital Lyon Sud, Hospices Civils de Lyon, 165 Chem du Grand Revoyet, 69495 Pierre-Bénite, France; email: herve.ghesquieres@chu-lyon.fr.

References

1. Le Guyader-Peyrou S, Defossez G, Dantony E, et al. *Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018. Volume 2 – Hémopathies malignes. Étude à partir des registres des cancers du réseau Francim, Saint-Maurice (Fra) : Santé publique France.* 2019.
2. Sehn LH, Salles G. Diffuse large B-cell lymphoma. *N Engl J Med.* 2021;384(9):842-858.
3. Khurana A, Mwangi R, Nowakowski GS, et al. Impact of organ function–based clinical trial eligibility criteria in patients with diffuse large B-cell lymphoma: who gets left behind? *J Clin Oncol.* 2021;39(15):1641-1649.
4. Loh Z, Salvaris R, Chong G, et al. Evolution of eligibility criteria for diffuse large B-cell lymphoma randomised controlled trials over 30 years. *Br J Haematol.* 2021;193(4):741-749.
5. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-sex-and age-based disparities. *JAMA.* 2004;291(22):2720-2726.
6. Jenei K, Haslam A, Olivier T, Miljkovic M, Prasad V. What drives cancer clinical trial accrual? An empirical analysis of studies leading to FDA authorisation (2015-2020). *BMJ Open.* 2022;12(10):e064458.
7. Unger JM, Vaidya R, Hershman DL, Minasian LM, Fleury ME. Systematic review and meta-analysis of the magnitude of structural, clinical, and physician and patient barriers to cancer clinical trial participation. *J Natl Cancer Inst.* 2019;111(3):245-255.
8. Harkins RA, Patel SP, Lee MJ, et al. Improving eligibility criteria for first-line trials for patients with DLBCL using a US-based Delphi-method survey. *Blood Adv.* 2022;6(9):2745-2756.
9. Maurer MJ, Ghesquières H, Link BK, et al. Diagnosis-to-treatment interval is an important clinical factor in newly diagnosed diffuse large B-cell lymphoma and has implication for bias in clinical trials. *J Clin Oncol.* 2018;36(16):1603-1610.
10. Concato J, Corrigan-Curay J. Real-world evidence – where are we now? *N Engl J Med.* 2022;386(18):1680-1682.
11. Concato J, Stein P, Dal Pan GJ, Ball R, Corrigan-Curay J. Randomized, observational, interventional, and real-world—what's in a name? *Pharmacoepidemiol Drug Saf.* 2020;29(11):1514-1517.
12. El-Galaly TC, Cheah CY, Villa D. Real world data as a key element in precision medicine for lymphoid malignancies: potentials and pitfalls. *Br J Haematol.* 2019;186(3):409-419.
13. Nowakowski G, Maurer MJ, Cerhan JR, Dey D, Sehn LH. Utilization of real-world data in assessing treatment effectiveness for diffuse large B-cell lymphoma. *Am J Hematol.* 2023;98(1):180-192.
14. Maurer MJ, Ghesquières H, Jais J-P, et al. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *J Clin Oncol.* 2014;32(10):1066-1073.
15. Thompson CA, Ghesquieres H, Maurer MJ, et al. Utility of routine post-therapy surveillance imaging in diffuse large B-cell lymphoma. *J Clin Oncol.* 2014;32(31):3506-3512.
16. Alabdjalbar MS, Durani U, Thompson CA, Constine LS, Hashmi SK. The forgotten survivor: a comprehensive review on Non-Hodgkin lymphoma survivorship. *Am J Hematol.* 2022;97(12):1627-1637.
17. Solal-Celigny P, Roy P, Colombat P, et al. Follicular lymphoma international prognostic index. *Blood.* 2004;104(5):1258-1265.
18. Federico M, Bellei M, Marcheselli L, et al. Follicular lymphoma international prognostic index 2: a new prognostic index for follicular lymphoma developed by the International Follicular Lymphoma Prognostic Factor Project. *J Clin Oncol.* 2009;27(27):4555-4562.
19. Bachy E, Maurer MJ, Habermann TM, et al. A simplified scoring system in de novo follicular lymphoma treated initially with immunochemotherapy. *Blood.* 2018;132(1):49-58.
20. 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. *J Clin Oncol.* 2015;33(23):2516-2522.
21. Casulo C, Dixon JG, Le-Rademacher J, et al. Validation of POD24 as a robust early clinical end point of poor survival in FL from 5225 patients on 13 clinical trials. *Blood.* 2022;139(11):1684-1693.
22. Ghesquières H, Rossi C, Cherblanc F, 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(1):432.
23. Tilly H, Morschhauser F, Sehn LH, et al. Polatuzumab vedotin in previously untreated diffuse large B-cell lymphoma. *N Engl J Med.* 2022;386(4):351-363.
24. Oberic L, Peyrade F, Puyade M, 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(11):1203-1213.
25. Degtiar I, Rose S. A review of generalizability and transportability. *Annu Rev Stat Appl.* 2023;10(1):501-524.
26. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med.* 1993;329(14):987-994.
27. Abramson JS, Solomon SR, Arnason JE, et al. Lisocabtagene maraleucel (liso-cel) versus standard of care (SOC) with salvage chemotherapy followed by autologous stem cell transplantation (ASCT) as second-line (2L) treatment in patients with relapsed or refractory large B-cell lymphoma (LBCL): primary analysis of the randomized, phase 3 transform study. *Blood.* 2022;140(suppl 1):1581-1583.

28. Kamdar M, Solomon SR, Arnason J, et al. Lisocabtagene maraleucel versus standard of care with salvage chemotherapy followed by autologous stem cell transplantation as second-line treatment in patients with relapsed or refractory large B-cell lymphoma (TRANSFORM): results from an interim analysis of an open-label, randomised, phase 3 trial. *Lancet*. 2022;399(10343):2294-2308.
29. Locke FL, Miklos DB, Jacobson CA, et al; All ZUMA-7 Investigators and Contributing Kite Members. Axicabtagene ciloleucel as second-line therapy for large B-cell lymphoma. *N Engl J Med*. 2022;386(7):640-654.
30. Thieblemont C, Tilly H, Gomes da Silva M, et al. Lenalidomide maintenance compared with placebo in responding elderly patients with diffuse large B-cell lymphoma treated with first-line rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J Clin Oncol*. 2017;35(22):2473-2481.
31. Le Gouill S, Ghesquières H, Oberic L, et al. Obinutuzumab vs rituximab for advanced DLBCL: a PET-guided and randomized phase 3 study by LYSA. *Blood*. 2021;137(17):2307-2320.
32. Framework for FDA's Real-World Evidence Program. *US Food and Drug Administration*. 2018.
33. HMA-EMA Joint Big Data Taskforce – Summary Report. *Heads of Medicines Agency, European Medicines Agency*. 2019.
34. Vision Paper on the Sustainable Availability of the Proposed Registry Evaluation and Quality Standards Tool (REQueST). *European Network of HTA*. 2019.
35. Public Policy Committee International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP): guidelines for good pharmacoepidemiology practice. *Pharmacoepidemiol Drug Saf*. 2016;25(1):2-10.
36. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457.
37. Arboe B, El-Galaly TC, Clausen MR, et al. The Danish National Lymphoma Registry: coverage and data quality. *PLoS One*. 2016;11(6):e0157999.
38. Lindner L, Weiß A, Reich A, et al. Implementing an automated monitoring process in a digital, longitudinal observational cohort study. *Arthritis Res Ther*. 2021;23(1):181.
39. Blacketer C, Defalco FJ, Ryan PB, Rijnbeek PR. Increasing trust in real-world evidence through evaluation of observational data quality. *J Am Med Inf Assoc*. 2021;28(10):2251-2257.
40. Laurent C, Baron M, Amara N, et al. Impact of expert pathologic review of lymphoma diagnosis: study of patients from the French Lymphopath Network. *J Clin Oncol*. 2017;35(18):2008-2017.
41. Smith A, Howell D, Crouch S, et al. Cohort profile: the Haematological Malignancy Research Network (HMRN): a UK population-based patient cohort. *Int J Epidemiol*. 2018;47(3), 700-700g.
42. Roman E, Kane E, Howell D, et al. Cohort profile update: the Haematological Malignancy Research Network (HMRN) UK population-based cohorts. *Int J Epidemiol*. 2022;51(3):e87-e94.
43. Cerhan JR, Link BK, Habermann TM, et al. Cohort profile: the Lymphoma Specialized Program of Research Excellence (SPORE) Molecular Epidemiology Resource (MER) cohort study. *Int J Epidemiol*. 2017;46(6):1753-1754i.
44. Alduaij W, Collinge B, Ben-Neriah S, et al. Molecular determinants of clinical outcomes in a real-world diffuse large B-cell lymphoma population. *Blood*. 2023;141(20):2493-2507.
45. Camus V, Rossi C, Sesques P, et al. Outcomes after first-line immunochemotherapy for primary mediastinal B-cell lymphoma: a LYSA study. *Blood Adv*. 2021;5(19):3862-3872.
46. Sibon D, Gisselbrecht C, Molina TJ, et al. Outcome of patients with primary mediastinal large B-cell lymphoma after R-CHOP21, R-CHOP14 and R-ACVBP: a pooled analysis of clinical trials from Lysa. *Blood*. 2022;140(suppl 1):1082-1084.
47. Harrysson S, Eloranta S, Ekberg S, et al. Incidence of relapsed/refractory diffuse large B-cell lymphoma (DLBCL) including CNS relapse in a population-based cohort of 4243 patients in Sweden. *Blood Cancer J*. 2021;11(1):9.
48. Lee B, Goktepe O, Hay K, et al. Effect of place of residence and treatment on survival outcomes in patients with diffuse large B-cell lymphoma in British Columbia. *Oncologist*. 2014;19(3):283-290.
49. Juul MB, Jensen PH, Engberg H, et al. Treatment strategies and outcomes in diffuse large B-cell lymphoma among 1011 patients aged 75 years or older: a Danish population-based cohort study. *Eur J Cancer*. 2018;99:86-96.
50. Le Lan C, Belot A, Golfier C, et al. Evaluation of participation and recruitment bias in a prospective real-life multicentric cohort « Real World Data in Lymphoma and Survival in Adults » (REALYSA study) for newly diagnosed lymphoma patients over one year in a hematology department of teaching hospital. *Blood*. 2022;140(suppl 1):5210-5211.
51. Le Guyader-Peyrou S, Orazio S, Dejardin O, Maynadié M, Troussard X, Monnereau A. Factors related to the relative survival of patients with diffuse large B-cell lymphoma in a population-based study in France: does socio-economic status have a role? *Haematologica*. 2017;102(3):584-592.
52. Lamy T, Damaj G, Soubeyran P, et al; LYSA Group. R-CHOP 14 with or without radiotherapy in nonbulky limited-stage diffuse large B-cell lymphoma. *Blood*. 2018;131(2):174-181.
53. Zelenetz AD, Gordon LI, Chang JE, et al. NCCN guidelines® insights: B-cell lymphomas, version 5.2021: featured updates to the NCCN guidelines. *J Natl Compr Cancer Netw*. 2021;19(11):1218-1230.
54. Poeschel V, Held G, Ziepert M, et al. Excellent outcome of young patients (18-60 years) with favourable-prognosis diffuse large B-cell lymphoma (DLBCL) treated with 4 cycles CHOP plus 6 applications of rituximab: results of the 592 patients of the Flyer Trial of the Dshnhl/GLA. *Blood*. 2018; 132(suppl 1):781.

55. Pfreundschuh M, Schubert J, Ziepert M, 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(2):105-116.
56. Sehn LH, Congiu AG, Culligan DJ, et al. No added benefit of eight versus six cycles of CHOP when combined with rituximab in previously untreated diffuse large B-cell lymphoma patients: results from the international phase III GOYA study. *Blood.* 2018;132(suppl 1):783.
57. Poeschel V, Held G, Ziepert M, et al; FLYER Trial Investigators, German Lymphoma Alliance. Four versus six cycles of CHOP chemotherapy in combination with six applications of rituximab in patients with aggressive B-cell lymphoma with favourable prognosis (FLYER): a randomised, phase 3, non-inferiority trial. *The Lancet.* 2019;394(10216):2271-2281.
58. Sehn LH, Scott DW, Villa D, et al. Long-term follow-up of a PET-guided approach to treatment of limited-stage diffuse large B-cell lymphoma (DLBCL) in British Columbia (BC). *Blood.* 2019;134(suppl 1):401.
59. Bologna S, Vander Borgh T, Briere J, et al. Early positron emission tomography response-adapted treatment in localized diffuse large B-cell lymphoma (AAIPI=0): results of the phase 3 LYSA LNH 09-1B Trial. *Hematol Oncol.* 2021;39(S2).
60. Haioun C, Mounier N, Emile JF, et al. Rituximab versus observation after high-dose consolidative first-line chemotherapy with autologous stem-cell transplantation in patients with poor-risk diffuse large B-cell lymphoma. *Ann Oncol.* 2009;20(12):1985-1992.
61. Tilly H, Gomes da Silva M, Vitolo U, et al. Diffuse large B-cell lymphoma (DLBCL): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2015;26(suppl 5):v116-v125.
62. Jakobsen LH, Bøgsted M, Brown PdN, et al. Minimal loss of lifetime for patients with diffuse large B-cell lymphoma in remission and event free 24 months after treatment: a Danish population-based study. *J Clin Oncol.* 2017;35(7):778-784.
63. Kanoun S, Tal I, Berriolo-Riedinger A, et al. Influence of software tool and methodological aspects of total metabolic tumor volume calculation on baseline [18F]FDG PET to predict survival in Hodgkin lymphoma. *PLoS One.* 2015;10(10):e0140830.
64. Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. *J Clin Oncol.* 2018;36(28):2845-2853.
65. Alig S, Macaulay CW, Kurtz DM, et al. Short diagnosis-to-treatment interval is associated with higher circulating tumor DNA levels in diffuse large B-cell lymphoma. *J Clin Oncol.* 2021;39(23):2605-2616.
66. Meriranta L, Alkodsí A, Pasanen A, et al. Molecular features encoded in the ctDNA reveal heterogeneity and predict outcome in high-risk aggressive B-cell lymphoma. *Blood.* 2022;139(12):1863-1877.
67. Monnereau A, Cornet E, Maynadié M, et al. *Survie des personnes atteintes de cancer en France métropolitaine 1989-2018 – Lymphome diffus à grandes cellules B.* Institut national du cancer; 2021.