ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

ONCOLOGIA, EMATOLOGIA E PATOLOGIA

Ciclo XXXVII

Settore Concorsuale: 06/D3 - MALATTIE DEL SANGUE, ONCOLOGIA E REUMATOLOGIA

Settore Scientifico Disciplinare: MED/15 - MALATTIE DEL SANGUE

Concordance and Influence of Large Language Model Treatment Recommendations Versus Hematologist Decision-Making in Relapsed/Refractory Diffuse Large B-Cell Lymphoma: The Ariadne's Thread Study

Presentata da: FRANCESCO MALASPINA

Coordinatore Dottorato

PROF.SSA EMANUELA FERRACIN

Supervisore

PROF.SSA ELENA ZAMAGNI

Co-Supervisore

DR. GERARDO MUSURACA

Esame finale anno 2026

Index

Abstract

Large language models (LLMs) achieve high scores on medical benchmarks, yet their clinical reasoning capabilities in complex, real-world therapeutic scenarios remain poorly characterized. Critically, it is unknown whether the expertise level of the clinician interacting with an LLM modulates the safety and utility of AI-generated recommendations. This thesis evaluates LLM performance in treatment selection for relapsed/refractory diffuse large B-cell lymphoma (R/R DLBCL), a scenario requiring integration of incomplete evidence, regulatory constraints, and nuanced clinical judgment across multiple treatment options.

In this prospective evaluation study, 50 synthetic clinical cases were each independently assessed by one lymphoma subspecialist and one general hematologist (100 total evaluations). Experts first formulated treatment recommendations independently, then evaluated recommendations generated by "Ariadne," a commercially available LLM (Claude Sonnet 4.5) augmented with retrieval-augmented generation and a curated R/R DLBCL knowledge base. Primary outcomes included overall clinical acceptability and treatment recommendation modification rates; exploratory outcomes assessed inter-rater agreement, harm potential in rejected cases, confidence changes, and error pattern characterisation.

The LLM achieved 83.0% overall acceptability (95% CI: 74.5%–89.1%), with treatment concordance (same treatment category selected independently by LLM and expert) observed in 75% of evaluations. However, physician expertise fundamentally altered the human-AI interaction pattern. General hematologists were five times more likely to modify their treatment after LLM exposure than lymphoma subspecialists (32% vs 6%), and showed higher acceptability rates (90% vs 76%). Subspecialists rejected 8 recommendations that generalists accepted versus only 1 reverse case, and 5 of these 8 asymmetric cases (63%) involved treatment choices deemed excessively toxic or sub-optimal by subspecialists. In 4 of these 8 cases, generalists also modified their treatment toward the LLM recommendation, including 2 with harm potential.

Qualitative error analysis revealed identifiable vulnerability domains: temporal information processing failures and regulatory compliance errors. These differed fundamentally from simple knowledge gaps. Stress testing confirmed that these failures were non-reproducible: structurally similar inputs did not reliably trigger the same errors, indicating stochastic vulnerabilities within known problem domains that cannot be anticipated or systematically prevented.

These findings reveal an expertise-dependent deployment paradox: the clinician population most behaviourally responsive to LLM recommendations (generalists) is also least equipped to detect errors that carry genuine harm potential, while the population best able to detect such errors (subspecialists) derives minimal decision support value. Current LLM systems may function comparably to an enthusiastic but junior hematology resident: valuable under subspecialist supervision but potentially dangerous without it. Clinical deployment requires expertise-stratified safeguards, and evaluation frameworks must move beyond aggregate accuracy metrics to assess error reproducibility, harm potential, and safety across diverse user populations.

Introduction

AI in Medicine and Clinical Decision-Making

Artificial Intelligence (AI) is emerging as a transformative force in healthcare. Large language models (LLM), the so-called chatbots, represent one of the most pervasive applications among the many AI-powered software that are being integrated into clinical practice. In the last 3 years, LLMs have demonstrated impressive performance on medical benchmarks, with continuous improvements (1–3), yet recent research reveals significant concerns about the robustness of these results. Studies indicate that LLMs may rely on shortcut behaviors and spurious correlations rather than genuine clinical reasoning (4), and that benchmark performance may be inflated by data contamination where evaluation datasets overlap with training data (5,6). These findings raise important questions about whether benchmark scores accurately reflect real-world clinical capabilities.

Furthermore, critical knowledge gaps remain regarding how LLM outputs influence clinician decision-making, how these systems could integrate into clinical workflows, and whether they might introduce biases or hallucinations which could have potentially serious consequences. Understanding these dynamics is essential before widespread clinical deployment, as errors in medical decision-making can have significant repercussions for patient safety.

Clinical decision-making in medicine, particularly in oncology, remains a cognitively demanding process characterized by complex, often incomplete data and by the need to integrate multiple sources of evidence, such as clinical trials, real-world data and guidelines, which by definition cannot be perfectly tailored for each specific patient disease combination (7). This challenge is amplified when multiple treatment options exist without direct comparative evidence and when disease heterogeneity complicates prognostication. In such scenarios, physicians must synthesize patient-specific factors, current guidelines, emerging research findings, and their own clinical experience to formulate treatment strategies, a process that is inherently vulnerable to cognitive biases and variability in expertise.

Several studies have begun to explore the interaction between LLMs and clinicians. These investigations have demonstrated good LLM performance in diagnostic accuracy across diverse clinical presentations (3,8), estimation of diagnostic probabilities (9), and generation of responses to physician queries (10). However, while these results demonstrate that LLMs can achieve high accuracy on standardized medical tasks, a few critical questions remain: can LLM improve physician decision-making in real clinical scenarios? Are LLM at risk of introducing biases or incorrect information in the reasoning process? And ultimately, are LLM honest and transparent enough to be trusted with our health? A recent randomized trial found that providing physicians with LLM-generated diagnostic suggestions did not significantly improve their diagnostic

reasoning compared to conventional resources alone (11). A subsequent large randomized study of 1,298 members of the general public found that, despite LLMs achieving 94.9% accuracy when tested alone, participants using these same models performed no better than those who did not use it, with information transmission between LLM and user identified as a critical point of failure (12). These findings suggest a disconnect between AI performance in isolation and its impact on human decision-making, both for lay users and trained physicians. Moreover, while LLMs strive to be helpful, they often display sycophantic responses (13,14), validating erroneous assumptions, especially when not properly prompted, a skill that requires training outside of the scope of medical formation. This underscores the need for rigorous evaluation of human-AI interaction in authentic clinical contexts before determining appropriate use cases for these technologies in healthcare.

Even more importantly, existing studies have largely treated "clinicians" as a homogeneous population. In clinical practice, though, AI decision support tools will inevitably be used by physicians with widely varying levels of domain expertise, from subspecialists with decades of focused experience to generalists who might ask for referrals for the more complex scenarios. Whether the safety and utility profile of LLM-generated recommendations differs across these user populations remains largely unexplored, yet this question is critical, especially in healthcare. This expertise-dependent dimension of human-AI interaction has been insufficiently investigated and represents a gap in the current evidence base.

Large B-Cell Lymphomas

Diffuse Large B-Cell Lymphoma

Diffuse large B-cell lymphoma is the most common subtype of non-Hodgkin lymphoma of the adult age, accounting for approximately 30-40% of all lymphoma cases globally (15). With an incidence rate of approximately 7 per 100,000 individuals, DLBCL represents a significant burden in developed countries (16). While first-line immunochemotherapy with rituximab combined with cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) achieves durable remissions in 60-70% of patients, approximately 30-40% of individuals experience either primary refractory disease or relapse after initial treatment (17,18).

The prognosis for patients with relapsed or refractory (R/R) DLBCL remains poor. Patients with primary refractory disease, defined as progression during or failure to achieve at least partial response after first-line therapy, have particularly dismal outcomes, with complete response rates to second-line regimens of only about 15% and 2-year progression-free survival rates of 10-20% (19). Even among patients who achieve response to salvage chemotherapy and proceed to

autologous stem cell transplantation, long-term disease control is limited, with 5-year overall survival rates of approximately 32% in population-based studies (20). For patients who have received two or more prior lines of therapy, median overall survival is frequently less than 6 months with conventional chemotherapy (19).

R/R DLBCL Treatment Paradigm

The management of R/R DLBCL has been changing rapidly in the last 5 years as new revolutionary treatments have been introduced. Due to this recent abundance of salvage treatments and to the constant ongoing trials that will likely impact future practice, and especially due to missing head to head comparative trials, international guidelines are keeping a wide and inclusive approach, which can sometimes lessen their usefulness (21,22). In Italy, treatment access is regulated by the Italian Medicines Agency (Agenzia Italiana del Farmaco, AIFA).

In the past, patients with R/R DLBCL after first-line therapy had their only curative option in salvage chemotherapy followed by autologous stem cell transplantation (ASCT), as established by the landmark PARMA trial (23). However, only a minority of relapsed patients ultimately proceed to transplantation, and outcomes were suboptimal, with long-term progression-free survival rates of 30-40% in chemosensitive patients and abysmal survival in refractory patients (19,24). The historical approach to salvage treatment prior to ASCT has traditionally involved platinum-based regimens such as R-ICE (rituximab, ifosfamide, carboplatin, etoposide), R-DHAP (rituximab, dexamethasone, cytarabine, cisplatin), or R-GDP (rituximab, gemcitabine, dexamethasone, cisplatin) (25). The CORAL study, a pivotal randomized trial comparing R-ICE to R-DHAP in 396 patients with R/R DLBCL, demonstrated that both regimens yielded similar response rates (approximately 63%) and progression-free survival, establishing either as an acceptable salvage approach (24). However, this study also highlighted the limitations of conventional salvage chemotherapy: only 50% of patients proceeded to ASCT, and 3-year event-free survival was merely 26% for the entire cohort. Importantly, the CORAL study identified key negative prognostic factors including prior rituximab exposure, relapse within 12 months of initial therapy, and secondary age-adjusted International Prognostic Index (saaIPI) score ≥2. These factors continue to inform risk stratification today. For patients with primary refractory disease or early relapse, outcomes with salvage chemotherapy and ASCT have remained particularly poor, providing the rationale for investigating alternative approaches in the second-line setting.

In recent years, the therapeutic landscape of R/R DLBCL has been transformed by the approval of novel immunotherapies. In Italy, AIFA has granted reimbursement for CD19-directed chimeric antigen receptor T-cell (CAR-T) therapies for patients with DLBCL relapsed or refractory after at least two prior lines of systemic therapy, subject to specific eligibility criteria (26). These criteria are more restrictive than the broader European Medicines Agency (EMA) authorization and include requirements such as ECOG performance status ≤1, adequate organ function, and an upper age limit of 75 years. Three CAR-T products have received AIFA approval: axicabtagene ciloleucel,

tisagenlecleucel, and lisocabtagene maraleucel, with the first product also approved (at the time of this study) in the second-line setting for patients with primary refractory disease or early relapse within 12 months of first-line therapy (27,28).

Beyond CAR-T cell therapy, several additional options have gained regulatory approval and AIFA reimbursement. CD20×CD3 bispecific T-cell engaging antibodies represent a major advance, with both glofitamab and epcoritamab receiving regulatory approval in Europe and the United States for patients with relapsed/refractory DLBCL after two or more prior therapies (29,30). These off-the-shelf immunotherapies offer advantages in terms of immediate availability and reduced manufacturing time compared to CAR-T cells and have demonstrated complete response rates of about 40% in heavily pretreated populations.

Antibody-drug conjugates (ADC) provide another therapeutic avenue. Polatuzumab vedotin, an anti-CD79b antibody conjugated to the microtubule inhibitor monomethyl auristatin E, has been approved in combination with bendamustine and rituximab for patients with relapsed/refractory DLBCL who are not candidates for autologous stem cell transplantation (31). The combination of tafasitamab (an anti-CD19 monoclonal antibody) with lenalidomide has similarly gained approval for patients ineligible for intensive therapy, demonstrating objective response rates of approximately 60% and complete response rates around 30% (32). Loncastuximab tesirine, a CD19-directed antibody-drug conjugate, represents an additional option for heavily pretreated patients, with approval after at least two prior systemic therapies with a quarter of the patients experiencing complete responses (33).

The Evidence Gap: Absence of Direct Comparative Data

With this expanding therapeutic armamentarium, a fundamental limitation severely constrains rational treatment selection: the near-complete absence of direct head-to-head randomized controlled trials comparing these therapies (34,35). Treatment decisions must be based on indirect comparisons across single-arm phase 2 trials or comparisons to historical controls, each with heterogeneous patient populations, varying eligibility criteria, different endpoints, and disparate follow-up durations. A few indirect comparisons and meta-analysis have tried to shed some light on the comparative effectiveness of these treatments, but with contradictory results. For example, a recent meta-analysis comparing CAR-T cells to bispecific antibodies in the third-line or later setting found superior complete response rates for CAR-T therapy (48% vs 36%), but this analysis was limited by the inability to control for differences in patient selection, disease characteristics, and prior treatments between the compared trials (36). At the same time, indirect comparisons on the registration trials of bispecifics and car-t have been performed, identifying no clear difference between these treatment modalities from the third line setting onward (37,38).

The lack of comparative evidence extends beyond efficacy to encompass critical questions of optimal sequencing. While emerging real-world data suggest that CAR-T cell therapy remains

effective after bispecific antibody failure, and vice versa, the magnitude of benefit and optimal treatment sequence remain poorly characterized (39,40,41).

This evidence gap is compounded by the absence of validated predictive biomarkers. While certain molecular features such as double-hit or triple-hit genetics (involving MYC, BCL2, and BCL6 rearrangements) or cell-of-origin (COO) classification into germinal center versus activated B-cell subtypes, carry prognostic significance, none reliably predict differential benefit from one treatment modality over another in the salvage setting (42–44). Recently, more sophisticated biological categories with prognostic impact have been reported, but their use is still limited to scientific trials and research centers and is not yet applicable in the clinical setting (45,46).

R/R DLBCL as a Test Case for LLM Clinical Reasoning

This rapidly evolving clinical landscape, with multiple treatment options lacking direct comparative trials and substantial heterogeneity in patient populations and disease biology, creates a scenario where clinical decision-making relies heavily on synthesis of incomplete evidence, pattern recognition from clinical experience and nuanced judgment regarding patient-specific factors. Physicians must integrate disease characteristics (primary refractory vs relapsed, time to relapse, molecular features, tumor burden), patient factors (age, performance status, comorbidities, organ function, social support), logistical constraints (proximity to certified treatment centers, manufacturing time for CAR-T cells, ability to comply with treatment administration schedules), and institutional capabilities. Current guidelines from ESMO and NCCN provide algorithmic frameworks but acknowledge substantial areas of uncertainty and explicitly note that treatment selection must be individualized (21,22). In practice, treatment selection in the relapsed/refractory setting often involves multidisciplinary discussion and, inevitably, the clinical judgment and experience of the treating hematologist.

The application of LLMs to such therapeutic decision-making represents a particularly ambitious deployment of artificial intelligence in healthcare. Unlike diagnostic image analysis or laboratory result interpretation, where AI systems process structured numerical or visual data, clinical reasoning in complex therapeutic scenarios requires integration of diverse information sources, accommodation of uncertainty, and generation of nuanced recommendations that account for patient-specific factors, some of which are hidden or unknown. Recent studies have demonstrated that LLMs can achieve high accuracy on medical licensing examinations and structured clinical vignettes (1–3), yet the translation of these capabilities to authentic clinical cases remains largely unexplored. Moreover, the high success rate of LLMs on standardized medical benchmarks does not equate nor does it imply their ability to perform actual clinician-level reasoning. A recent study demonstrated numerous examples of “shortcut behaviour” and a higher than expected degree of

brittleness of the most commonly available LLMs when exposed to medical benchmarks modified with missing information, further suggesting that their high scores have been pushed by reward-training which focused only on giving the "correct" answer and not on correctly examining and reasoning through each case (4).

A critical distinction therefore exists between performance on standardized assessments and utility in real-world clinical practice. Medical vignettes typically present discrete clinical scenarios with a single correct answer, whereas real practice involves reasoning on often incomplete evidence and with multiple sensible approaches.

R/R DLBCL treatment selection thus presents an ideal test case for evaluating LLM clinical capabilities: the absence of direct head-to-head randomized trials creates genuine uncertainty that cannot be resolved through simple algorithmic approaches; the rapid evolution of the field means that guidelines provide broad frameworks rather than specific algorithms; treatment selection must integrate diverse and sometimes conflicting considerations; and the stakes are high, as treatment decisions can meaningfully impact survival and quality of life. Unlike scenarios where a single "correct" answer can be objectively determined, this context enables evaluation of whether LLMs can generate nuanced, contextually appropriate recommendations rather than simply retrieving memorized information. Conversely, if LLMs fail to appropriately weight conflicting evidence or provide recommendations that diverge from expert judgment without sound rationale, this setting should expose such limitations.

Study Objectives

This study aims to evaluate the clinical acceptability and behavioral influence of LLM-generated treatment recommendations for relapsed/refractory diffuse large B-cell lymphoma, using a paired-comparison design that enables assessment of how physician expertise modulates the human-AI interaction.

Primary descriptive objectives:

1. To estimate the overall clinical acceptability rate of LLM-generated treatment recommendations across a spectrum of R/R DLBCL clinical presentations
2. To quantify the rate at which LLM recommendations influence expert treatment decisions, measured as the proportion of evaluations in which experts modify their independently formulated treatment recommendation after LLM exposure, stratified by expert type
3. To characterize confidence changes in experts' treatment decisions following LLM exposure, stratified by expert type

Exploratory objectives:

4. To compare clinical acceptability between lymphoma subspecialists and general hematologists
5. To characterize the harm potential of rejected LLM recommendations
6. To identify and characterize recurrent LLM error patterns
7. To analyse the relationship between treatment concordance (whether LLM and expert independently selected the same treatment category) and behavioral influence

By conducting this evaluation with paired assessment by subspecialists and generalists, this study addresses a gap in the current evidence: whether the safety and utility profile of LLM-generated clinical recommendations differs across user expertise levels, and what this implies for deployment strategies in clinical practice.

Materials and Methods

Study Design

This prospective evaluation study employed a paired-comparison design to assess the clinical acceptability and behavioral influence of LLM-generated treatment recommendations for relapsed/refractory (R/R) diffuse large B-cell lymphoma (DLBCL). Fifty clinical cases were developed, and each case was independently evaluated by one lymphoma subspecialist and one general hematologist randomly assigned from pools of 6 subspecialists and 12 generalists, respectively. This design yielded 50 paired evaluations (n = 100 total assessments), enabling within-case comparison of expert type responses while minimizing evaluator burden. The study was conducted at IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori" between September 2025 and January 2026. All evaluators provided oral informed consent after being briefed on study procedures, time requirements, data use, and their right to withdraw. Figure 1 shows a schematic of the study design.

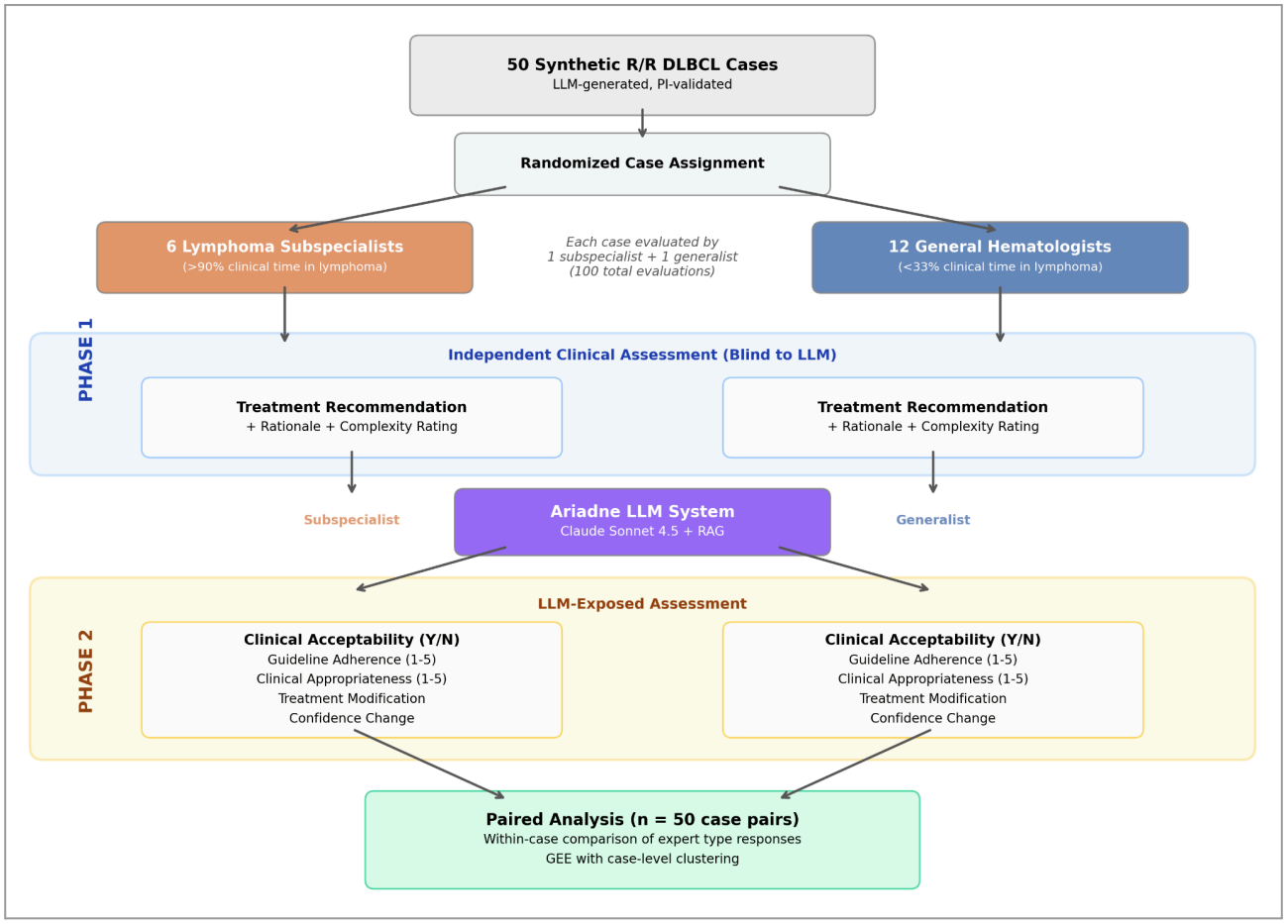

Fig. 1 *Study design schematic. Phase 1 is defined as before LLM exposure. Phase 2 is after LLM exposure. LLM: large language model. RAG: retrieval augmented generation.*

Case Development

Fifty synthetic clinical cases representing R/R DLBCL treatment scenarios were generated using a structured, LLM-assisted case generation pipeline. The pipeline was designed and supervised by the principal investigator (a lymphoma subspecialist with over 10 years of clinical experience) to ensure systematic coverage of the R/R DLBCL clinical scenario space while maintaining individual case realism.

Case Generation System

Cases were generated by Claude Sonnet 4.5 (Anthropic) operating under a purpose-built case generation prompt (see Appendix A). The prompt directed the model to execute a three-step process: (1) generation of randomized variability parameters defining core case characteristics (relapse timing category, patient fitness, comorbidity burden, disease stage, and number of prior

treatment lines), with probability weights calibrated to approximate real-world epidemiological distributions; (2) generation of a complete clinical case as a structured JSON object; and (3) production of a textual narrative summary derived strictly from the structured data.

To ensure clinical accuracy and internal consistency, the case generation system operated within a constrained environment comprising three investigator-curated reference files:

- A **formal JSON schema** (DLBCL_Case_Schema_v5) defining all required data fields, permissible values, and validation rules, including mandatory consistency checks (e.g., IPI score must equal the sum of its Boolean components; Ann Arbor stage must reflect documented sites of disease)
- The same **treatment database** (treatments.json) and **guidelines repository** (guidelines.json) used by the Ariadne recommendation system, ensuring that generated treatment histories referenced real regimens with guideline-concordant sequencing and AIFA-compliant eligibility criteria
- An **extranodal sites reference file** (extranodal_sites.json) providing site-specific clinical data including typical presentation patterns, CNS relapse risk, molecular profiles, and staging implications

The prompt explicitly instructed the model to query these reference files when selecting treatment regimens, verifying eligibility criteria, and populating extranodal disease characteristics, and to document the rationale for each treatment selection with reference to the retrieved data. Built-in consistency checks required verification of IPI score calculations, stage–site concordance, age progression logic, and alignment between diagnosis subtype and molecular features before case finalization.

Case Validation

All 50 generated cases underwent clinical validation by the principal investigator. Validation assessed clinical plausibility, internal consistency of temporal relationships and laboratory values, appropriateness of treatment sequencing given patient characteristics and regulatory constraints. Cases requiring modification were revised before inclusion. Cases were intentionally permitted to include sub-optimal treatment histories in some instances to enhance realism and reflect actual referral patterns encountered in clinical practice.

Temporal information was presented explicitly in each case, including time from initial diagnosis and time from completion of last treatment, as this information is essential for determining regulatory eligibility for specific therapies under Italian Medicines Agency (AIFA) criteria. The complete case generation prompt and JSON schema are provided in Appendix A.

Case Assignment

Case assignment to evaluators was randomized as follows: for each case, one subspecialist was randomly selected from the pool of 6 subspecialists, and one generalist was randomly selected from the pool of 12 generalists. Evaluators were randomly assigned to cases using a simple computer-generated randomization sequence. This randomization strategy ensured balanced evaluator workload while preserving the paired structure necessary for within-case expert type comparisons. Each subspecialist evaluated approximately 8–9 cases; each generalist evaluated approximately 4–5 cases.

The Ariadne LLM System

System Architecture

"Ariadne" was developed as a Retrieval-Augmented Generation (RAG) system using Claude Sonnet 4.5 (Anthropic, San Francisco, CA; model string: 'claude-sonnet-4-5-20250929') as the base large language model. The system was designed to address known limitations of standalone LLMs, including knowledge cutoff constraints and potential for clinically implausible recommendations.

Knowledge Base

The RAG system accessed two curated JSON files developed and maintained by the principal investigator:

1. **treatments.json**: A structured database containing detailed information on R/R DLBCL treatment options including applicability by treatment line, patient selection criteria (inclusion/exclusion), key supporting clinical trial data with outcomes and toxicity profiles, guideline positioning, and special clinical considerations. The database included conventional salvage chemotherapy regimens, CAR-T cell therapies, bispecific antibodies, antibody-drug conjugates, immunomodulatory drugs and best supportive care approaches. (Created during 2024–2025; finalized April 3, 2025)

2. **guidelines.json**: A structured database organizing treatment recommendations by clinical scenario (e.g., fit patients with early relapse, unfit patients with late relapse), with explicit inclusion of AIFA regulatory criteria, treatment sequencing logic, evidence levels, and rationale for recommendation hierarchies. (Created during 2024–2025; finalized April 3, 2025)

The knowledge base was finalized on April 3, 2025, and remained static throughout the study period (September 2025–January 2026). During this interval, no significant changes in AIFA regulatory approvals or major practice-changing clinical trial publications occurred that would have altered treatment recommendations for the clinical scenarios evaluated. The LLM's web search capability (described below) could access current online literature, potentially bridging any gaps between knowledge base finalization and study execution. This temporal structure was maintained to ensure consistency across all case evaluations while preserving the system's ability to access emerging evidence.

System Capabilities

The Ariadne system utilized Claude Sonnet 4.5's native web search capabilities without additional constraints or customization beyond the system prompt. This "out-of-the-box" configuration was intentional: the study aimed to evaluate LLM performance under realistic conditions accessible to practicing clinicians without programming expertise. The system could autonomously search online medical literature using its built-in search functionality, retrieve and synthesize information from accessible web sources, and integrate findings with the curated knowledge base. Specific search strategies, source selection, and result ranking were determined by the model's default behavior rather than by investigator-specified parameters. This approach prioritizes external validity (real-world clinical deployment scenarios) over the controlled reproducibility achievable with constrained search parameters.

The system prompt incorporated role-playing techniques and structured reasoning through a Patient Case Profile (PCP) methodology, wherein the LLM systematically processed clinical history, identified missing information, conducted preliminary research using web browsing and knowledge retrieval tools, simulated a multidisciplinary expert panel discussion, and synthesized recommendations with explicit reasoning.

The complete system prompt (see Appendix B) directed the model to provide a single primary treatment recommendation with clear justification citing specific guidelines, trial data, or regulatory criteria, acknowledge areas of clinical uncertainty, consider patient-specific factors, and explicitly address regulatory eligibility under AIFA criteria.

Expert Evaluators

Six lymphoma subspecialists and twelve general hematologists were recruited via direct invitation from the principal investigator's professional network. All evaluators practiced in Italy and were familiar with AIFA regulatory requirements. Subspecialists were defined as hematologists dedicating >90% of clinical time to mature lymphoproliferative disorders; generalists dedicated <33% of clinical time to these disorders. Participation was voluntary without compensation.

Evaluators were informed that an LLM system generated recommendations and that their responses would be compared to LLM outputs, but were blinded to the study's specific hypotheses regarding expertise-dependent differences and to each other's assessments. No formal training or calibration exercises were conducted; evaluators applied their standard clinical judgment to case evaluations.

Evaluation Procedure

Each case was evaluated by one lymphoma subspecialist and one general hematologist (randomly assigned from the respective pools as described above). Evaluations were conducted using a custom PDF-based questionnaire. Evaluators received three PDF files per case: (1) the clinical case description, (2) a blank evaluation form, and (3) the LLM recommendation (in a separately labeled file).

Evaluators were instructed to complete the evaluation in two phases:

Phase 1: Independent Clinical Assessment (Pre-LLM Exposure)

Evaluators reviewed the clinical case and completed Phase 1 sections of the evaluation form without accessing the LLM recommendation file. They provided:

1. **Primary treatment recommendation** with brief justification
2. **Case Complexity Assessment:** The evaluator was asked to respond to the following subjective question: "*How would you rate the complexity of this case?*" and could choose between three preset complexity levels defined as: "*1 - Straightforward: Clear best treatment option*"; "*2 - Moderately complex: Several reasonable options to consider*"; "*3 - Very complex: Multiple viable options, difficult treatment choice*".

This phase established each expert's baseline clinical judgment independent of LLM influence.

Phase 2: LLM-Exposed Assessment (Post-LLM Exposure)

After completing Phase 1, evaluators opened the LLM recommendation PDF and completed Phase 2 sections of the evaluation form, providing:

1. **Clinical Acceptability Rating**: Binary judgment (Acceptable / Not Acceptable) in response to the question: *"Would you EVER consider recommending THIS specific treatment for this patient?"*

2. **Guideline Adherence and Clinical Appropriateness Rating:** Evaluators had to rate how much they considered the LLM treatment proposal guideline adherent and clinically

appropriate on two five-points Likert scales

3. **Modified Treatment Recommendation**: Evaluators indicated whether they would change their original recommendation after seeing the LLM output, selecting from:

    - *(a) "No change - My original recommendation remains the best option"*
    - *(b) "Modification - Same therapeutic approach with changes to specifics (e.g., different platinum regimen, different CAR-T product, adjusted dosing)"*
    - *(c) "Complete change - Fundamentally different therapeutic strategy (e.g., chemotherapy → CAR-T, ASCT → bispecific antibody, curative → palliative)"*

    If change was selected, evaluators were also asked to report whether it was towards or away from the LLM proposed treatment.

4. **Revised Confidence Rating**: Evaluators indicated whether their confidence in their final treatment choice changed after LLM exposure, selecting from:
    - *"Decreased confidence - I am now less certain about my choice"*
    - *"No change in confidence"*
    - *"Increased confidence - I am now more certain about my choice"*

5. **Structured Rejection Reasons** (if rating = Not Acceptable): Evaluators selected from predefined rejection categories via checkboxes:

    - Code 1: Contraindicated due to patient comorbidities or organ function
    - Code 2: Excessive toxicity risk for this patient
    - Code 3: Not evidence-based or violates guidelines
    - Code 4: Wrong line of therapy or treatment sequence
    - Code 5: Not available/accessible in our setting
    - Code 6: Not appropriate for patient's fitness level
    - Code 7: More effective alternatives available
    - Code 8: Other reasons (required free-text explanation)

    Multiple codes could be assigned to a single rejection. For Code 8 selections, free-text explanations were reviewed and coded by the principal investigator as either "harm potential" (concerns about toxicity or disease control) or "minimal harm potential" (minor clinical judgment differences such as bridging strategy choice or supportive care variations). No independent verification of Code 8 classifications was performed.

Both phases were completed in the same session. Evaluators were instructed not to open the LLM recommendation file until Phase 1 was complete, but compliance with this sequencing was not

technologically enforced (files were not password-protected or time-locked). Adherence to the intended sequence relied on evaluator good faith.

Harm Potential Classification (Pre-Specified)

Prior to data collection, rejection codes were categorized by harm potential based on clinical reasoning about recommendation appropriateness:

Harm Potential: Codes 1, 2, 6 and 7 were classified as involving harm potential. These codes indicated selection of treatments with unfavorable risk-benefit ratios (codes 1, 2 and 6) or availability of superior alternatives (codes 7). These represent direct threats to patient safety or suboptimal disease control.

Minimal Harm Potential: Codes 3, 4, and 5 were classified as minimal harm potential. These codes reflected regulatory constraints or procedural considerations that, while potentially significant, usually represent logistical barriers rather than fundamental treatment inadequacy.

Code 8 Classification: For the "Other" category, free-text responses were reviewed and classified as involving harm potential if they described concerns about toxicity or disease control, and as minimal harm potential if they described minor clinical judgment differences (e.g., choice of bridging strategy, CAR-T product selection, supportive care variations).

Case-Level Harm Assessment: A rejected case was classified as having "harm potential" if at least one rejection code from the harm category (1, 2, 6, 7, or harm-related code 8 responses) was cited by either expert.

This conservative approach may underestimate harm potential, as regulatory non-compliance or sequencing violations could imply harm to patients, but prioritizes specificity over sensitivity in harm detection.

Data Collection

Evaluations were conducted using PDF questionnaires distributed electronically. All data was entered into a structured database. Backward navigation was not technologically prevented; evaluators could theoretically revise Phase 1 responses after viewing LLM recommendations, though they were instructed not to do so.

Outcome Measures

Primary Descriptive Outcomes

1. **Overall Clinical Acceptability Rate**: Proportion of LLM recommendations rated as "Acceptable" across all evaluations (n = 100), reported with 95% confidence interval.

2. **Treatment Recommendation Modification Rate**: Proportion of cases where expert's Phase 2 recommendation differed from Phase 1 (either "Modification" or "Complete change"), calculated separately for general hematologists and lymphoma subspecialists.

3. **Confidence Changes After LLM Exposure**: Distribution of confidence changes (decreased/stable/increased) stratified by expert type.

Exploratory Comparative Outcomes

1. **Acceptability by Expert Type**: Comparison of acceptability rates between general hematologists and lymphoma subspecialists.

2. **Inter-Rater Agreement**: Consensus acceptance rate (both experts accept), Cohen's kappa, and assessment of systematic disagreement patterns (McNemar's test). Note: Inter-rater agreement reflects concordance between different expert types evaluating the same case, not test-retest reliability of the same expert.

3. **Harm Characteristics in Rejected Cases**: Proportion of rejected cases with harm potential, distribution of rejection codes, harm potential in asymmetrically rejected cases (accepted by one expert type, rejected by the other).

4. **Acceptability by Treatment Modality**: Acceptability rates stratified by LLM-recommended treatment type (e.g., CAR-T, bispecific antibody, salvage chemotherapy, best supportive care).

5. **Error Pattern Analysis**: Qualitative assessment of rejection reasons to identify systematic error types.

6. **Post-Hoc Error Pattern Investigation**: Following identification of a specific LLM error pattern during data collection (temporal information processing failures in relapse timing classification), a limited stress-testing protocol was conducted to assess error reproducibility. Details of the error pattern and stress testing results are reported in Results

and discussed as a qualitative finding.

Statistical Analysis

Sample Size

This study evaluated 50 clinical cases, each assessed by one lymphoma subspecialist and one general hematologist (n = 100 total evaluations). Sample size was determined by feasibility constraints, as no preliminary data existed to guide formal power calculations. Given the absence of prior acceptability benchmarks for LLM-generated hematology-oncology recommendations, this study should be considered a pilot investigation with primarily descriptive objectives. The primary outcome (overall acceptability rate) is a descriptive parameter estimate with precision quantified via 95% confidence intervals. Secondary analyses comparing expert types and assessing inter-rater agreement are exploratory and should be interpreted as hypothesis-generating rather than definitive tests of pre-specified hypotheses.

Analysis Plan

Analysis of Binary Outcomes (Acceptability and Modifications): To account for the hierarchical structure of the data (paired evaluations clustered within cases), Generalized Estimating Equations (GEE) were used for all binary endpoints. Binomial distribution with a logit link function, an exchangeable working correlation structure, and robust standard errors (Huber-White sandwich estimator) were used.

- **Overall Rates**: An intercept-only GEE model was used to estimate the overall marginal probability of acceptability (and modification rates) along with their 95% confidence intervals.
- **Exploratory Comparison**: To assess differences between subspecialists and generalists, the GEE model was extended to include expert type as the predictor and case as the clustering variable (n = 50 clusters). Results are reported as Odds Ratios (OR) with 95% confidence intervals.

Inter-Rater Agreement: Cohen's kappa calculated for the 50 paired evaluations with 95% confidence intervals using the percentile bootstrap method (10,000 resamples). McNemar's test applied to assess systematic disagreement patterns (whether one expert type systematically accepted recommendations more frequently than the other).

Confidence Changes and Harm Potential: Descriptive statistics (frequencies and proportions) stratified by expert type and acceptability rating. Chi-square or Fisher's exact tests used for exploratory comparisons where appropriate.

Exploratory Analyses: Treatment modality acceptability and error patterns analyzed descriptively. Between-modality comparisons conducted using chi-square tests but reported without formal hypothesis testing framework.

Multiple Comparison Approach

Given the exploratory nature of secondary analyses, no formal correction for multiple comparisons was applied. Instead, all comparative analyses (expert type, treatment modality, confidence changes) are explicitly labeled as hypothesis-generating, and p-values are reported without adjustment but should not be interpreted as confirmatory evidence. Effect size estimates (odds ratios, absolute risk differences) and confidence intervals are prioritized over significance testing for these exploratory comparisons. The significance threshold for exploratory tests was set at $\alpha = 0.05$, recognizing that findings require replication in adequately powered confirmatory studies.

Software

All analyses were performed using Python 3.11.4 with pandas 2.0.3 for data management, statsmodels 0.14.0 for GEE analyses, scipy 1.11.1 for proportion tests and Cohen's kappa, and matplotlib 3.7.2 for visualization. The Ariadne LLM system used Claude Sonnet 4.5 (model string: 'claude-sonnet-4-5-20250929') via the Claude.ai web interface. Analysis scripts are available from the author upon reasonable request.

Results

Between September 1, 2025 and January 27, 2026, all 50 planned clinical cases were evaluated, yielding 100 total evaluations. Each case was independently assessed by one lymphoma subspecialist and one general hematologist. Baseline case characteristics are summarized in Table 1.

Case Characteristics

The 50 synthetic cases represented a clinically diverse cohort of relapsed/refractory DLBCL patients. The median age at relapse was 57 years (range: 31-85). The cohort exhibited high-risk features consistent with the R/R DLBCL population: 28.0% (14/50) had ECOG performance status ≥2, 42.0% (21/50) had IPI >2, 38.0% (19/50) presented with bulky disease, and 76.0% (38/50) were

refractory to their last line of therapy. The median number of previous treatments was 2 (range: 1-5), with 36.0% (18/50) having received more than two prior treatment lines. Prior autologous stem cell transplantation (ASCT) was documented in 18.0% (9/50) of cases, and 28.0% (14/50) had previously received CAR-T cell therapy.

Case complexity, as rated by evaluators, was distributed across the 100 evaluations as follows: 43 evaluations (43.0%) were classified as Level 1 (straightforward), 43 evaluations (43.0%) as Level 2 (moderate complexity), and 14 evaluations (14.0%) as Level 3 (complex).

Table 1. *Synthetic Case Patient and Disease Characteristics*

Characteristic	**Cases (n = 50)**
Median age, years (range)	57 (31-85)
ECOG ≥2 at relapse	14 (28.0%)
Limited stage (I-II)	20 (40.0%)
Advanced stage (III-IV)	30 (60.0%)
IPI >2	21 (42.0%)
Bulky disease	19 (38.0%)
Refractory to first-line therapy	5 (10.0%)
Refractory to last treatment	38 (76.0%)
Median number of prior treatments (range)	2 (1-5)
>2 prior treatments	18 (36.0%)
Previous ASCT	9 (18.0%)
Previous CAR-T	14 (28.0%)

Primary Outcome: Clinical Acceptability

The primary outcome was the proportion of LLM recommendations judged clinically acceptable by expert hematologists across all 100 evaluations. Overall, 83 of 100 LLM recommendations (83.0%) were deemed clinically acceptable (**Fig. 2**). Using generalized estimating equations (GEE) with case-level clustering to account for the paired evaluation structure, the estimated marginal probability of acceptability was 83.0% (95% CI: 74.5%-89.1%).

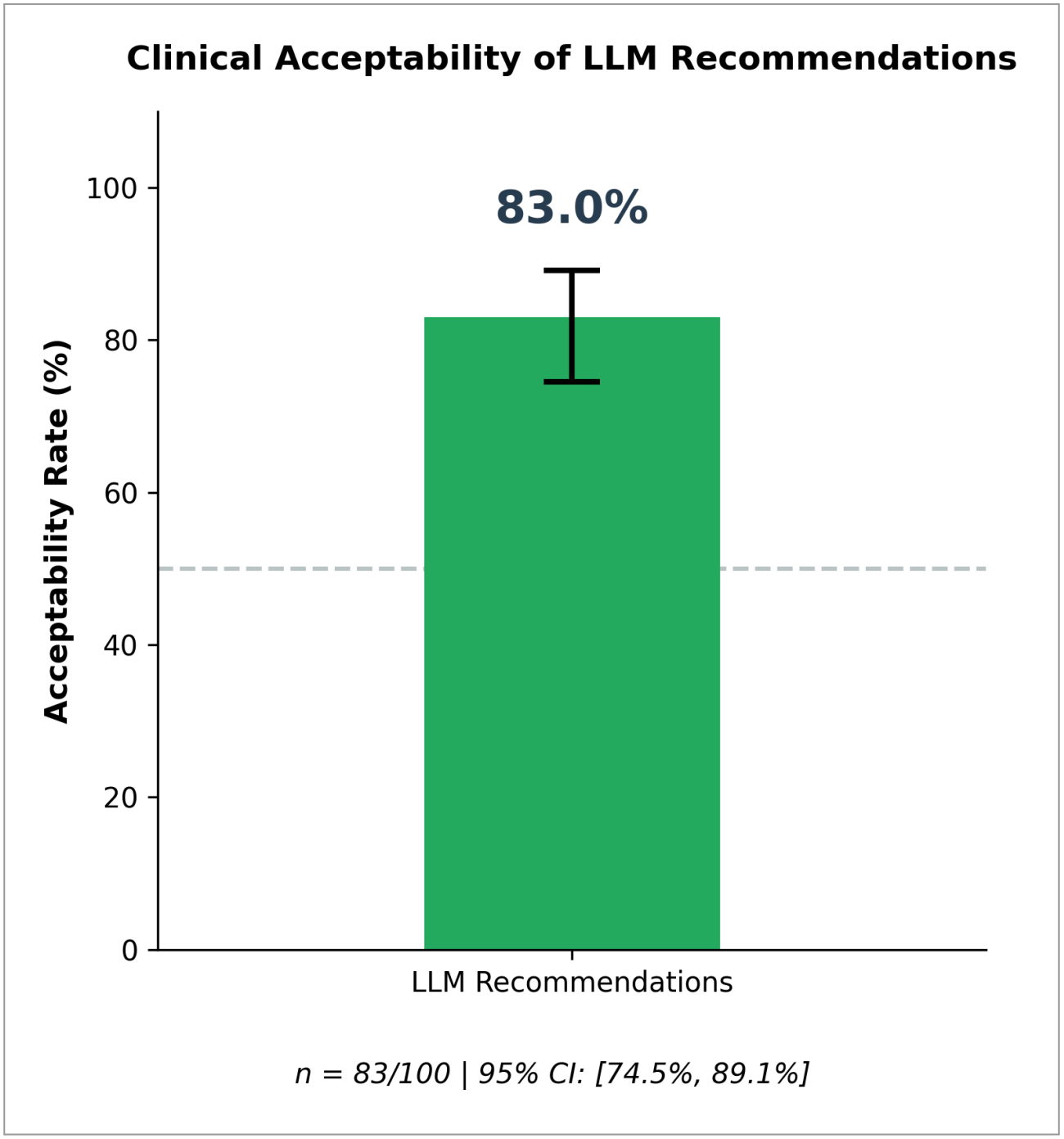

Fig. 2 *Overall acceptability of LLM generated treatment recommendations*

Acceptability by Expert Type

Clinical acceptability differed between expert types, though this did not reach statistical significance in exploratory analysis. General hematologists accepted 90.0% (45/50) of LLM recommendations, compared to 76.0% (38/50) acceptance among lymphoma subspecialists, representing an absolute difference of 14 percentage points. When analyzed using Fisher's exact

test, this difference yielded an odds ratio of 0.35 (95% CI: 0.11-1.12, p = 0.108). The Chi-square test similarly showed $\chi^2 = 2.55$ (df = 1, p = 0.110). While the point estimate suggested lower acceptance among subspecialists, the confidence interval was wide and crossed unity, precluding definitive conclusions in this exploratory analysis (**Fig. 3**).

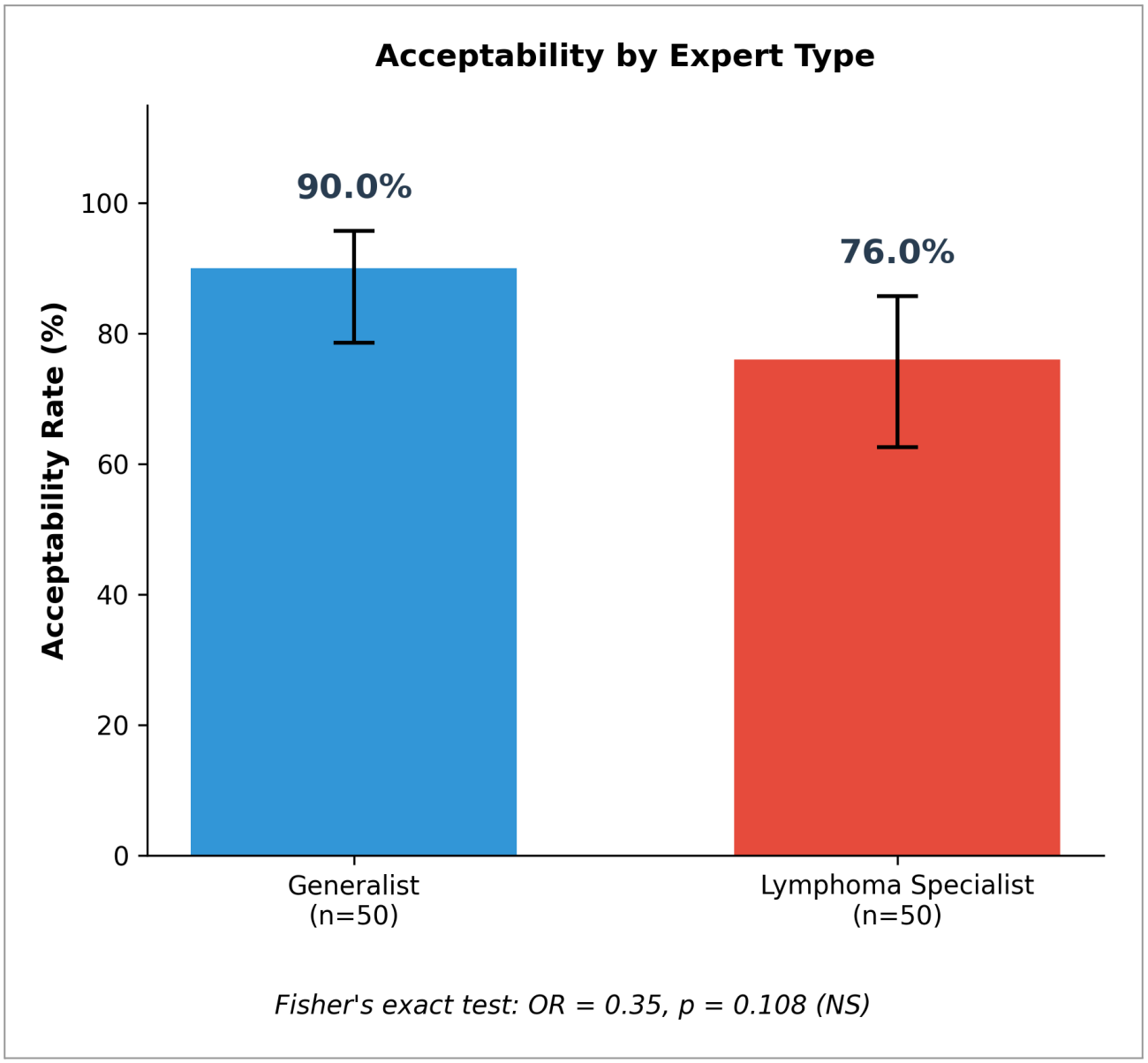

Fig. 3 Acceptability of LLM generated recommendations stratified by expert type

Inter-Rater Agreement on Acceptability

For the 50 cases evaluated by both expert types, observed agreement on clinical acceptability was 82.0% (41/50 cases). Cohen's kappa was 0.384, indicating fair agreement beyond chance. The 2×2 contingency table revealed an asymmetric disagreement pattern: 8 cases were accepted by

generalists but rejected by subspecialists, whereas only 1 case was rejected by generalists but accepted by subspecialists. McNemar's test confirmed this systematic difference (p = 0.039), with generalists more likely to accept LLM recommendations than subspecialists (**Fig. 4**).

Consensus acceptance (both experts accepting) occurred in 74.0% of cases (37/50). At least one expert accepted the recommendation in 92.0% of cases (46/50). Both experts rejected the recommendation in only 8.0% of cases (4/50).

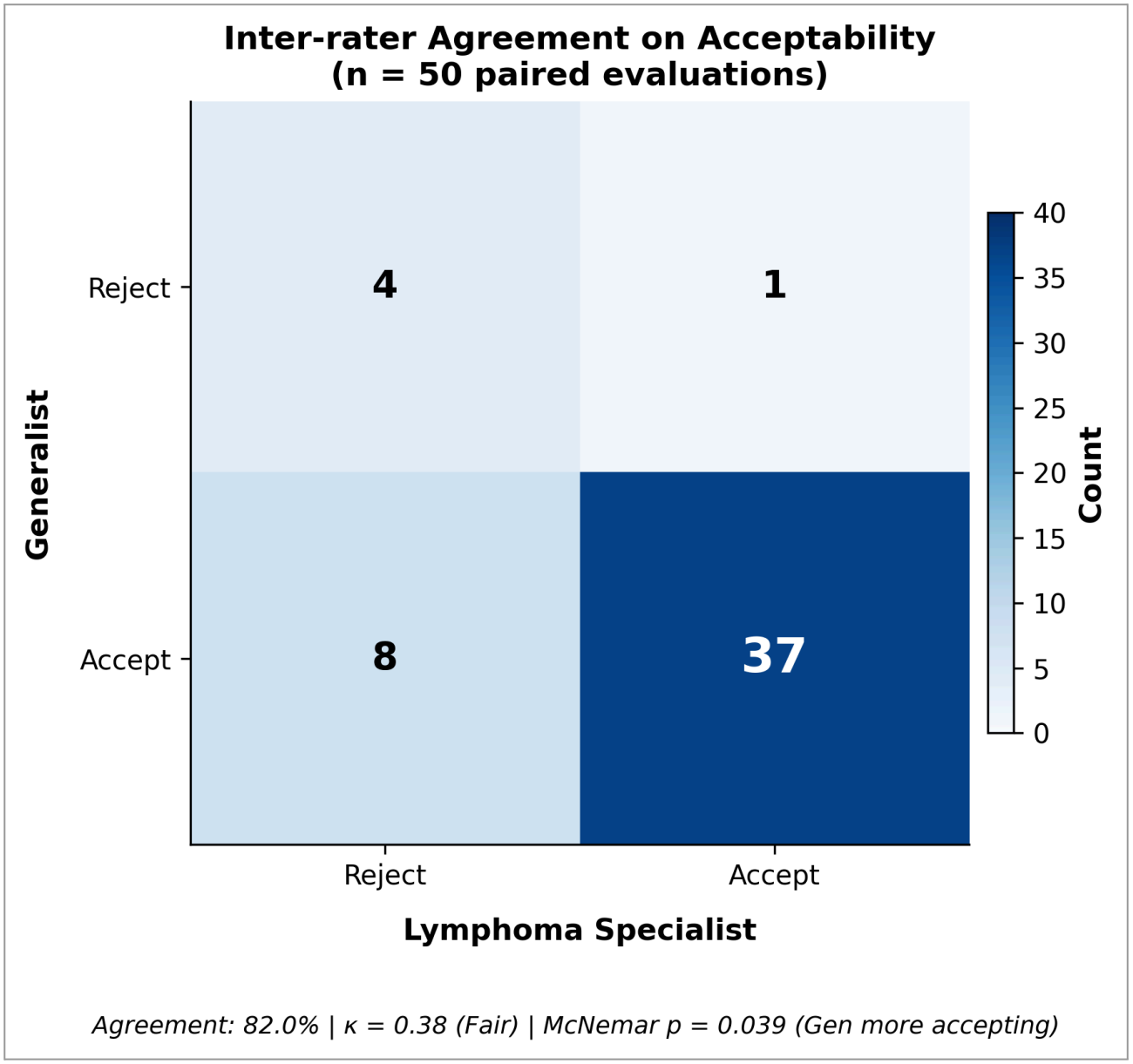

Fig. 4 *Agreement matrix between expert types. Lymphoma subspecialists rejected 8 LLM recommendations that generalists accepted. Generalists rejected 1 LLM recommendation that subspecialists accepted.*

LLM Performance Ratings

LLM recommendations were evaluated using two five-point Likert scales assessing guideline adherence and clinical appropriateness. Overall mean scores were high: guideline adherence 4.27 ± 1.08 (median 5.0, IQR 4.0-5.0) and clinical appropriateness 4.24 ± 1.17 (median 5.0, IQR 4.0-5.0). However, these aggregate statistics masked substantial heterogeneity driven by acceptability status.

When stratified by acceptability, a pronounced bimodal distribution emerged. Acceptable recommendations received markedly higher ratings: guideline adherence 4.59 ± 0.70 and clinical appropriateness 4.66 ± 0.63. In contrast, unacceptable recommendations received substantially lower scores: guideline adherence 2.71 ± 1.26 and clinical appropriateness 2.18 ± 1.01. Mann-Whitney U tests confirmed these differences were highly statistically significant (both $p < 0.001$), with point-biserial correlations of $r = 0.658$ for guideline adherence and $r = 0.800$ for clinical appropriateness (**Fig. 5**).

As expected, the two rating dimensions exhibited strong correlation (Spearman $\rho = 0.832$, $p < 0.001$), suggesting experts evaluated guideline adherence and clinical appropriateness in a unified manner rather than as independent constructs. Recommendations rated favorably on one dimension were consistently rated favorably on the other, and vice versa.

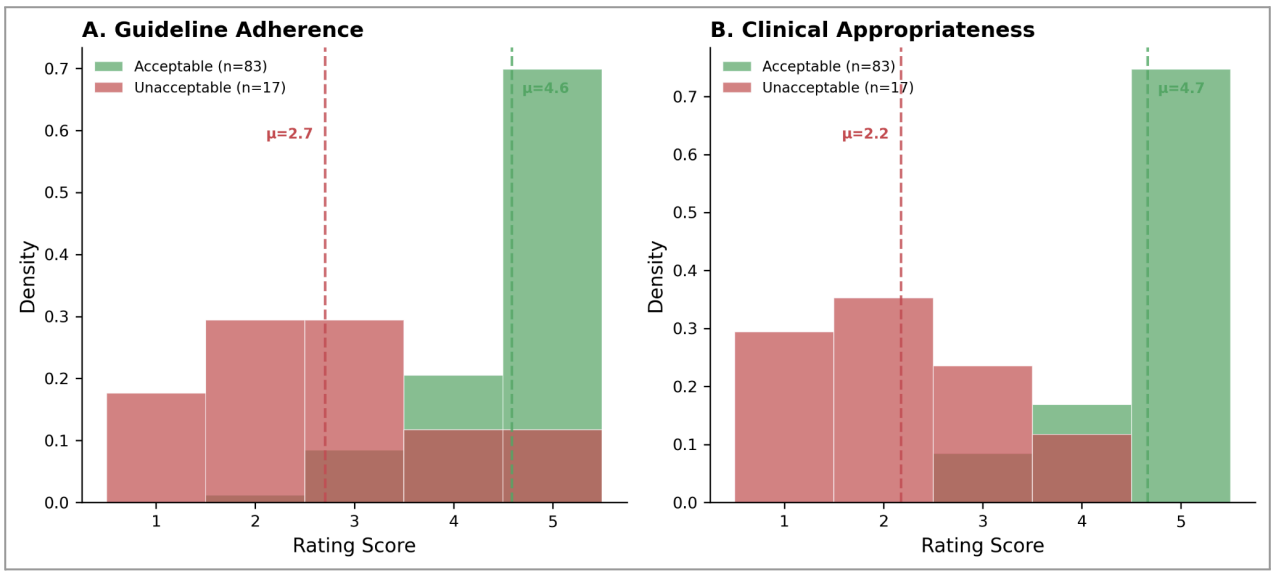

Fig. 5 *A) Correlation between expert rating on guideline adherence and clinical acceptability of LLM generated recommendations; B) Correlation between expert rating on clinical appropriateness and clinical acceptability of LLM generated recommendations.*

Acceptability by Case Complexity

Acceptability rates showed a numerical trend toward declining performance with increasing case complexity, though this did not reach statistical significance. Level 1 (straightforward) cases achieved 88.4% acceptability (38/43 evaluations), Level 2 (moderate) cases 81.4% (35/43), and Level 3 (complex) cases 71.4% (10/14). Chi-square test yielded $\chi^2 = 2.29$ (df = 2, p = 0.319), and Spearman correlation between complexity and acceptability was $\rho = -0.147$ (p = 0.146), indicating insufficient evidence to conclude that complexity independently predicted acceptability in this sample size.

Treatment Recommendation Modifications

Following exposure to LLM recommendations in Phase 2, experts modified their initial Phase 1 treatment recommendations in 19.0% of evaluations (19/100; GEE-estimated 95% CI: 12.5%-27.8%). The modifications were categorized as follows: no change in 81 evaluations (81.0%), tactical modification (same therapeutic class or strategy with refinements) in 15 evaluations (15.0%), and complete change (fundamental shift in treatment strategy) in 4 evaluations (4.0%).

Modification Rates by Expert Type

Modification rates differed markedly between expert types (**Fig. 6**). General hematologists modified their recommendations in 32.0% of cases (16/50; 95% CI: 20.8%-45.8%), including 12 tactical modifications (24.0%) and 4 complete changes (8.0%). In contrast, lymphoma subspecialists modified their recommendations in only 6.0% of cases (3/50; 95% CI: 2.1%-16.2%), all of which were tactical modifications; no subspecialist completely changed their treatment strategy. This represented a 5-fold difference in modification rates. Fisher's exact test confirmed this difference was highly significant (OR = 0.14, 95% CI: 0.04-0.51, p = 0.002).

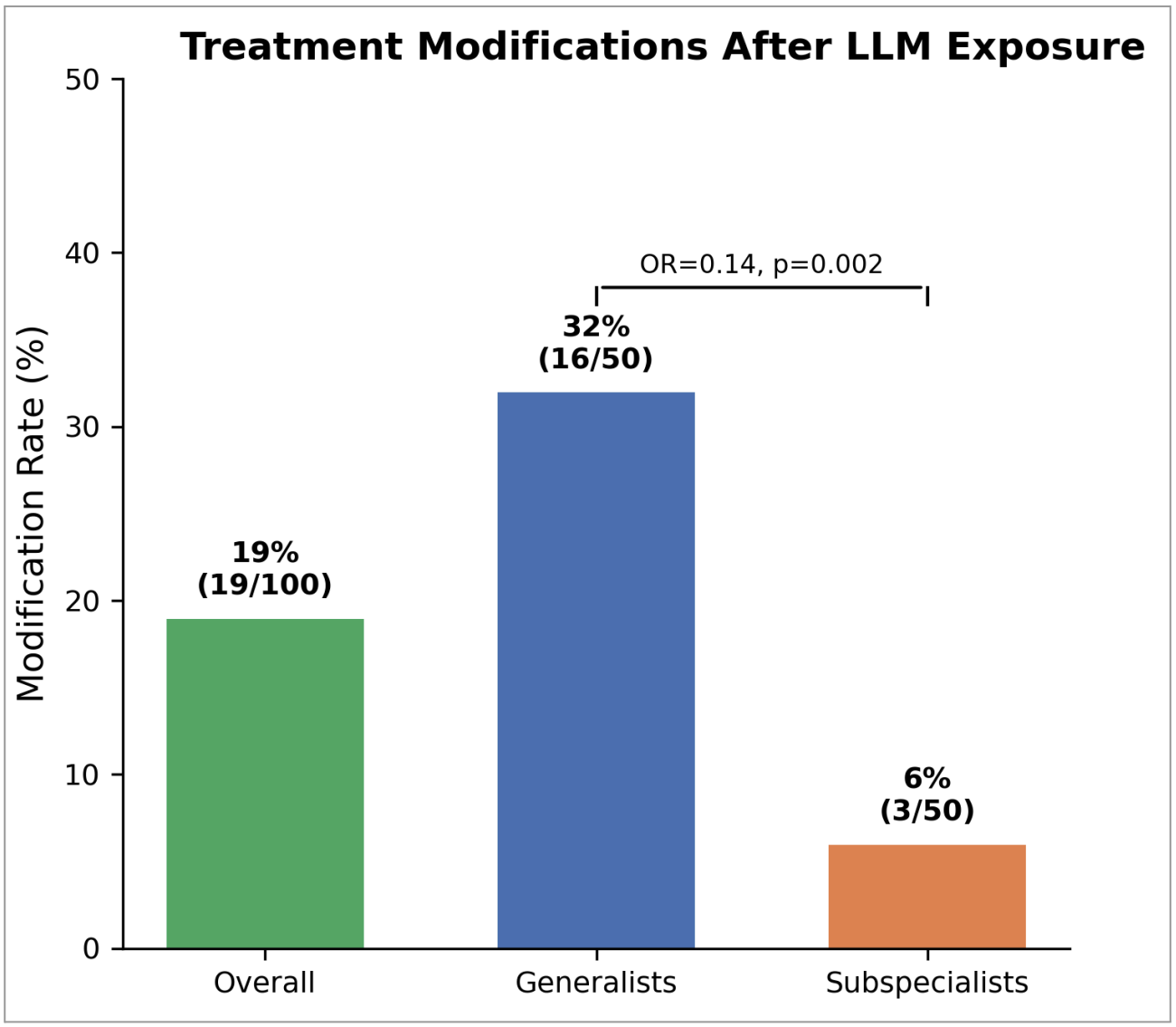

Fig. 6 *Modification rate of treatment choices after LLM exposure overall, for generalists and for subspecialists.*

Direction of Treatment Changes

Among the 19 modifications, 17 (89.5%) moved toward alignment with the LLM recommendation, indicating that when experts revised their treatment plans, they predominantly adopted or incorporated elements of the LLM suggestion. Only 2 modifications (10.5%) involved changes in a different direction from the LLM recommendation, suggesting the LLM output directly influenced the majority of treatment revisions.

Modification by Case Complexity

Exploratory analysis revealed a numerical trend toward higher modification rates with increasing complexity, though formal interaction testing was not significant ($\chi^2 = 0.76$, $p = 0.382$). Among Level 1 cases, 14.0% of evaluations resulted in modifications (6/43), compared to 20.9% for Level 2 (9/43) and 28.6% for Level 3 (4/14). This pattern was driven primarily by generalists, who showed increasing modification rates across complexity levels (22.7%, 35.0%, and 50.0% for Levels 1, 2, and 3 respectively; Spearman $\varrho = 0.203$, $p = 0.157$), whereas subspecialists remained consistently resistant to modification regardless of complexity (4.8%, 8.7%, and 0.0%; Spearman $\varrho = 0.003$, $p = 0.982$).

Confidence Changes After LLM Exposure

Evaluators reported their change in confidence regarding their treatment decision after reviewing the LLM recommendation (**Fig. 7**). Overall, confidence remained stable in 63 evaluations (63.0%), increased in 25 evaluations (25.0%), and decreased in 12 evaluations (12.0%).

Confidence change patterns differed markedly by expert type ($\chi^2 = 16.06$, $df = 2$, $p = 0.0003$). Among general hematologists, confidence increased in 40.0% of evaluations (20/50), remained stable in 44.0% (22/50), and decreased in 16.0% (8/50). Among lymphoma subspecialists, confidence remained stable in 82.0% of evaluations (41/50), increased in 10.0% (5/50), and decreased in 8.0% (4/50).

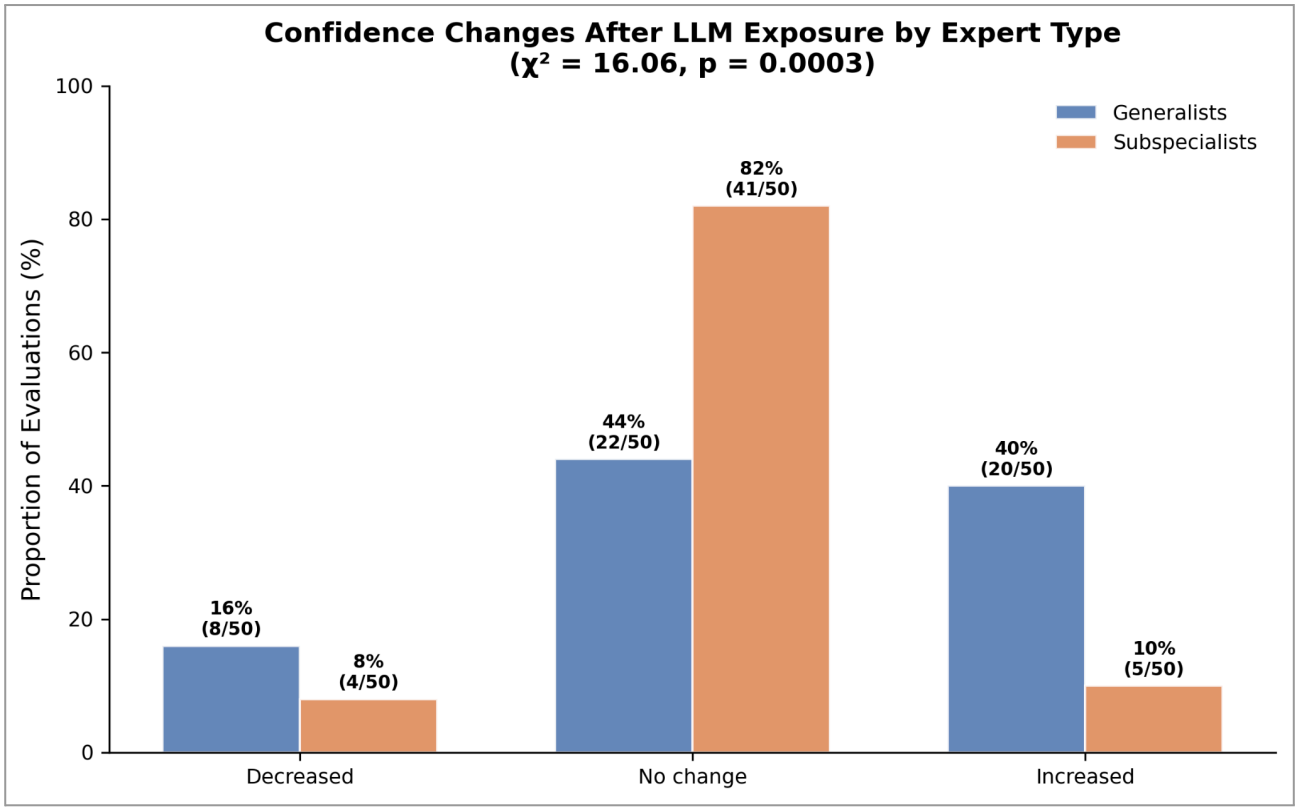

Fig. 7 *How experts' confidence in their treatment choice changed after exposure to LLM generated recommendations.*

Reasons for Unacceptability and Harm Potential

Rejection Patterns

A total of 17 unacceptable evaluations occurred across 13 unique cases. Four cases were rejected by both expert types (Cases 3, 11, 41, and 47), representing instances of consensus LLM failure. Nine cases were rejected by one expert type only: 8 rejected by the subspecialist but accepted by the generalist (Cases 6, 7, 19, 22, 23, 45, 46, and 50), and 1 rejected by the generalist but accepted by the subspecialist (Case 8).

Rejection Reason Classification

Rejection reasons were coded using a pre-specified classification system. Codes were categorized into two groups based on potential for patient harm:

Harmful rejection codes (clinically suboptimal or too toxic treatment choices):

- Code 1: Contraindicated due to patient comorbidities or organ function

- Code 2: Excessive toxicity risk for this patient
- Code 6: Not appropriate for patient's fitness level
- Code 7: More effective alternatives available

Non-harmful rejection codes (regulatory or minor concerns):

- Code 3: Not evidence-based or violates guidelines
- Code 4: Wrong line of therapy or treatment sequence
- Code 5: Not available/accessible in our setting
- Code 8: Other

Multiple codes could be assigned to a single evaluation. For case-level harm assessment, the union of all rejection codes from both evaluators was considered when both rejected the same case.

Harm Potential in Rejected Cases

Among the 13 unique rejected cases, 10 (77%) involved harmful rejection codes, indicating clinically suboptimal treatment choices or excessive toxicity concerns rather than merely regulatory or availability issues (**Fig 8A**). The 3 cases without harmful codes (Cases 7, 22, and 45) were rejected for reasons that were considered as having less potential for harm (treatment sequencing preferences, availability, or other minor concerns).

Among the 8 cases where the subspecialist rejected the recommendation but the generalist accepted it, 5 (63%) involved harmful rejection codes. In 4 of these 8 asymmetric cases (Cases 6, 7, 45, and 46), generalists also modified their treatment toward the LLM recommendation, including 2 complete treatment changes (Cases 6 and 46). Of these 4 modifications, 2 (Cases 6 and 46) involved cases with harmful rejection codes (**Fig 8B**).

The most frequently cited harmful rejection code was Code 7 ("More effective alternatives available"), appearing in 10 of the 13 rejected cases either alone or in combination with other codes. Code 2 ("Excessive toxicity risk") appeared in 3 cases, and Codes 1 and 6 (contraindication and fitness inappropriateness) each appeared in 2 cases.

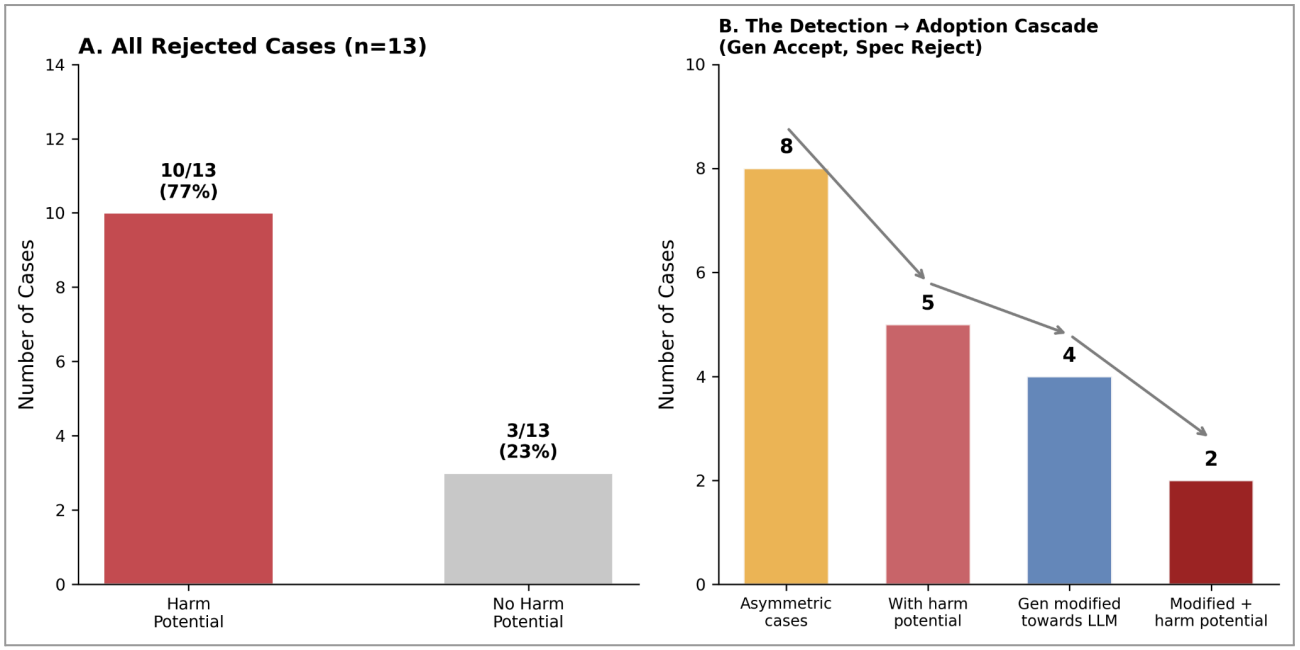

Fig. 8 *A) Harm potential of the LLM proposed treatments that were rejected by at least one expert type; B) Adoption cascade showing LLM treatment recommendations rejected by subspecialists and with identified harm potential that were accepted by generalists and which, in 2 cases, induced treatment choice modifications.*

Acceptability by Treatment Modality

LLM recommendations encompassed diverse treatment modalities based on the custom knowledge base integrating current R/R DLBCL treatment options. The distribution of LLM-recommended treatments reflected contemporary therapeutic approaches: CAR-T cell therapy was recommended in 40 evaluations (40.0%), bispecific antibodies (BiTE) in 22 evaluations (22.0%), best supportive care (BSC) in 14 evaluations (14.0%), salvage chemotherapy with or without autologous stem cell transplantation (ASCT) in 10 evaluations (10.0%), antibody-drug conjugates (ADC) in 6 evaluations (6.0%), tafasitamab-lenalidomide combination (TAFA-LENA) in 4 evaluations (4.0%), and BTK inhibitors and clinical trial enrollment each in 2 evaluations (2.0%).

Acceptability varied substantially by treatment modality. TAFA-LENA recommendations achieved 100% acceptability (4/4), though the small sample precludes definitive conclusions. BSC recommendations showed high acceptability at 92.9% (13/14). CAR-T recommendations, representing the largest category, achieved 87.5% acceptability (35/40). ADC and BiTE recommendations showed acceptability rates of 83.3% (5/6) and 81.8% (18/22) respectively. In contrast, ASCT-based salvage approaches demonstrated lower acceptability at 60.0% (6/10), and BTK inhibitor and clinical trial recommendations showed 50.0% acceptability each (1/2), though

these latter categories had minimal sample sizes. The overall chi-square test for differences across modalities was not statistically significant ($\chi^2 = 9.22$, df = 7, p = 0.238), likely reflecting the modest sample sizes in several categories.

Expert Type Differences by Modality

Acceptability patterns by treatment modality differed between expert types, with generalists consistently showing higher acceptance across most categories. For CAR-T recommendations, generalists accepted 90.0% (18/20) compared to subspecialists at 85.0% (17/20), a modest 5.0 percentage point difference. More pronounced differences emerged for BiTE recommendations: generalists accepted 90.9% (10/11) while subspecialists accepted 72.7% (8/11), an 18.2 percentage point difference. For BSC recommendations, generalists accepted all cases (100%, 7/7) whereas subspecialists accepted 85.7% (6/7). ADC recommendations showed the largest differential, with generalists at 100% acceptance (3/3) compared to subspecialists at 66.7% (2/3), though the small sample size limits interpretation. ASCT-based recommendations showed identical 60.0% acceptance for both expert types (3/5 each).

Treatment Concordance

Analysis of treatment concordance between LLM and human expert recommendations revealed that exact category matches occurred in 75.0% of evaluations overall (75/100). Concordance was significantly associated with acceptability: when LLM recommendations were judged acceptable, treatment categories matched in 79.5% of cases (66/83), compared to only 52.9% (9/17) when recommendations were deemed unacceptable ($p < 0.001$, Fisher's exact test).

Cross-tabulation of LLM versus human treatment categories showed that CAR-T recommendations demonstrated the highest within-category concordance, with 30 of 40 LLM CAR-T recommendations (75.0%) also selected as CAR-T by the human expert. BiTE recommendations showed 19 of 22 (86.4%) exact matches. BSC recommendations matched in 10 of 14 cases (71.4%). However, considerable therapeutic substitution occurred in discordant cases: of the 10 LLM-recommended ASCT cases, only 6 (60.0%) were concordant, with 4 cases where experts preferred CAR-T therapy instead (**Fig 9**).

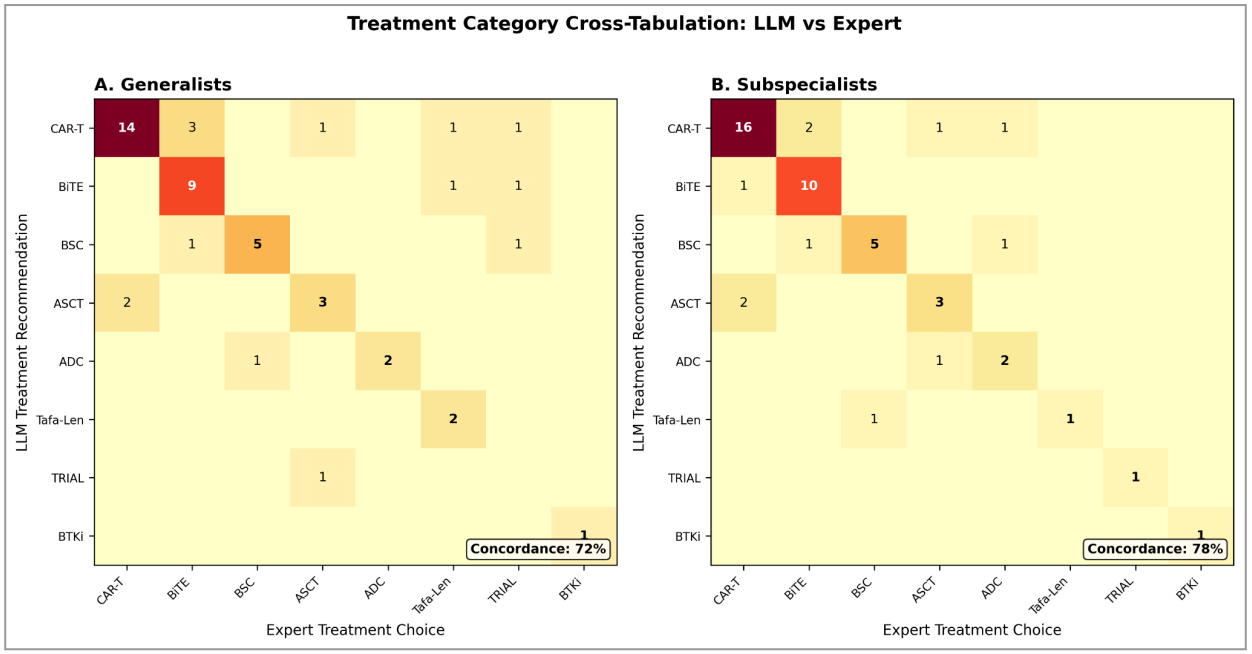

Fig. 9 *Heatmap of treatment category concordance between experts and LLM*

Concordance and Treatment Modifications After LLM Exposure

To clarify the relationship between acceptability, concordance, and behavioral influence, modification rates were stratified by whether the LLM and expert had independently selected the same treatment category (concordant) or different categories (discordant) prior to LLM exposure.

Overall, concordant evaluations showed a modification rate of 16.4% (11/67), all within-category adjustments (e.g., different CAR-T product, different salvage regimen, different bispecific antibody). Discordant evaluations showed a higher modification rate of 24.2% (8/33), including all 4 complete treatment changes observed in the study.

Stratified by expert type, the concordance-influence relationship differed markedly. Among generalists, concordant evaluations (35/50, 70%) yielded a 25.7% modification rate (9/35), exclusively within-category adjustments. Discordant evaluations (15/50, 30%) yielded a 46.7% modification rate (7/15), including all 4 complete treatment changes, indicating that nearly half of generalists adopted the LLM's different therapeutic approach when categories diverged. Among subspecialists, modification rates were uniformly low regardless of concordance: 6.2% (2/32) in concordant and 5.6% (1/18) in discordant evaluations, indicating that subspecialists rarely modified their recommendations irrespective of whether the LLM proposed the same or a different treatment category.

Agreement on Case Complexity Ratings

Beyond acceptability judgments, inter-rater agreement was assessed for case complexity ratings assigned by the two expert types evaluating each case. Exact agreement on complexity level occurred in 54.0% of cases (27/50). Cohen's unweighted kappa was 0.248 (slight agreement), while weighted kappa accounting for ordered categories was 0.403 (fair agreement). Spearman correlation between complexity ratings was moderate ($\varrho = 0.417$, $p = 0.003$).

The mean difference in complexity ratings between generalists and subspecialists was minimal (0.02 ± 0.77), and the Wilcoxon signed-rank test showed no systematic bias ($p = 0.854$). When disagreement occurred, generalists rated cases as more complex in 24.0% of pairs (12/50) and specialists rated cases as more complex in 22.0% (11/50), indicating bidirectional disagreement without consistent directional bias.

Pearson correlations for the rating scales between expert types showed moderate agreement: guideline adherence $r = 0.569$ ($p < 0.001$) and clinical appropriateness $r = 0.550$ ($p < 0.001$), with mean differences of 0.10 ± 1.02 and 0.44 ± 1.13 respectively.

Association Between Clinical History and Outcomes

Exploratory analysis revealed that the number of prior treatment lines showed a weak positive association with acceptability (Spearman $\varrho = 0.207$, $p = 0.041$). Cases with more treatment lines tended to have higher acceptability rates: 1 line 73.5% (25/34 evaluations), 2 lines 84.4% (27/32), 3 lines 100% (20/20), 4 lines 66.7% (4/6), and 5 lines 100% (6/6). However, this finding should be considered exploratory given the small numbers in some categories and lack of pre-specification.

Prior treatment lines and CAR-T history were strongly associated with perceived case complexity. The number of prior lines correlated with complexity ratings (Spearman $\varrho = 0.339$, $p < 0.001$), and cases with prior CAR-T therapy were rated as significantly more complex than those without (mean complexity 2.11 ± 0.69 vs 1.56 ± 0.65; Mann-Whitney $p < 0.001$). Previous ASCT and presence of significant extranodal disease did not significantly affect complexity ratings or acceptability.

Qualitative Error Pattern Analysis

Qualitative review of rejected cases and of expert free-text comments identified recurrent LLM reasoning failures within identifiable vulnerability domains. Post-hoc analysis of these patterns revealed two major categories accounting for all the 4 cases rejected by both subspecialists and generalists.

Temporal Information Processing Failures

In 2 cases rejected by both generalists and subspecialists, the LLM error involved failures in temporal reasoning when case narratives contained multiple time metrics. The LLM inappropriately prioritized time-from-initial-diagnosis over the clinically relevant time-from-treatment-completion when these metrics yielded conflicting relapse timing classifications (one >12 months, one <12 months). This led to misapplication of relapse timing-dependent treatment algorithms, resulting in recommendations of ASCT-based salvage approaches when CAR-T therapy would have been more appropriate for the true early relapse status.

Post-hoc stress testing of this error pattern using intentionally constructed cases with conflicting temporal metrics revealed inconsistent LLM behavior. Specific narrative structures and phrasing choices influenced whether the LLM correctly identified the clinically relevant time interval, but the pattern was not deterministic. This suggested architectural limitations in complex temporal reasoning and information prioritization rather than simple knowledge gaps.

Guideline Overgeneralization

Additional rejected cases involved recommendations that technically aligned with treatment guidelines in some contexts but represented inappropriate overgeneralization to scenarios outside current evidence-based indications. These included recommendations for repeat CAR-T therapy in patients who had previously received CAR-T (not currently standard practice), and treatment selections that failed to account for specific contraindications or patient-specific factors that would exclude otherwise reasonable therapeutic options.

These error patterns accounted for a minority of total evaluations (approximately 17%) but were notable in that they involved misapplication of otherwise correct clinical knowledge rather than factual errors. Stress testing confirmed that structurally similar inputs did not reliably produce the same errors.

Discussion

Summary of Principal Findings

This prospective evaluation of 50 synthetic R/R DLBCL cases, each assessed by a paired lymphoma subspecialist and general hematologist, yielded three principal findings. First, the LLM system achieved substantial overall clinical acceptability, indicating that current architectures with domain-specific augmentation can generate recommendations within the bounds of evidence-based practice for a majority of complex oncology scenarios. Second, treatment concordance between the LLM and expert was observed in approximately two-thirds of evaluations, providing essential context for interpreting the modest aggregate modification rate: when both parties independently converge on the same therapeutic approach, low modification reflects convergent reasoning rather than a failure of the system to influence decisions. Third, physician expertise fundamentally altered the human-AI interaction, producing a striking divergence in both behavioral influence and error detection capability between subspecialists and generalists, with direct consequences for treatment safety. Each of these findings is examined in detail below.

The Expertise-Dependent Utility Paradox

Acceptability, Concordance, and Influence

The relationship between acceptability and influence requires interpretation through the lens of treatment concordance. When the LLM and expert independently selected the same treatment category (70% for generalist and 64% for subspecialist), high acceptability accompanied by low modification reflects convergent clinical reasoning rather than a failure of the LLM to influence decisions. Modifications in concordant cases were exclusively within-category adjustments (e.g., different CAR-T product, different bispecific antibody), suggesting that concordance-driven LLM exposure prompted refinement of an already-established therapeutic strategy rather than reconsideration of the strategy itself.

The more informative scenario is discordance: cases where the LLM proposed a fundamentally different therapeutic approach from the expert's initial choice. Here, expertise produced sharply divergent responses. Among generalists, discordant cases yielded a 46.7% modification rate, including all four complete treatment changes observed in the study. Nearly half of generalists, when presented with a plausible alternative treatment category, adopted it. Among subspecialists,

modification remained uniformly low regardless of concordance (6.2% concordant, 5.6% discordant), indicating that subspecialists' resistance to modification was not driven by agreement with the LLM but by confidence in their own clinical judgment, even when acknowledging the LLM's alternative as acceptable.

This divergence has direct safety implications. Among the 16 generalist modifications toward the LLM, 4 (25%) involved cases that subspecialists independently rejected as unacceptable, with 2 of those carrying harm potential. Generalist responsiveness to the LLM thus operated without the discriminative filter that subspecialist expertise provides: one in four generalist modifications moved toward a recommendation that subspecialist evaluation would have identified as inappropriate.

Harm Potential in the Expertise Gap

The safety concern extends beyond behavioral influence to encompass a more fundamental detection failure. Among the 8 cases where generalists accepted but subspecialists rejected the LLM recommendation, 5 (63%) involved harm potential. These were treatment choices deemed by subspecialists to carry excessive toxicity risk or therapeutic inadequacy, rather than regulatory or procedural disagreements. This detection gap exists independently of whether generalists subsequently modified their treatment: the failure is one of critical evaluation, not merely behavioral susceptibility. In all 8 asymmetric cases, generalists judged the LLM recommendation clinically acceptable when subspecialist expertise identified it as unacceptable, a divergence that would not be apparent without subspecialist oversight.

In 4 of these 8 asymmetric cases, the detection failure translated into behavioral consequence: generalists modified their treatment toward the LLM recommendation, including 2 cases with harm potential. In Case 6, the generalist completely changed from their original treatment to adopt the LLM-recommended radiotherapy, which the subspecialist identified as carrying excessive toxicity risk for a patient with diffuse disease and prior radiation exposure. In Case 46, the generalist changed from best supportive care to loncastuximab, which the subspecialist identified as inferior to available alternatives. These cases illustrate the mechanism by which the expertise gap produces concrete treatment harm: a plausible but flawed LLM recommendation passes through the generalist's evaluation unchallenged and is adopted as the treatment plan.

The distinction between detection gap and behavioral consequence gap matters for intervention design. Addressing behavioral influence alone, for example by adding obligated checks to the modification process, would not resolve the underlying detection failure. If generalists cannot reliably identify when LLM recommendations contain safety concerns, then the acceptability judgment itself is unreliable for this user population in complex scenarios, regardless of whether it leads to treatment modification.

The Resident Analogy

The observed interaction patterns suggest a useful analogy for conceptualizing the current role of LLMs in clinical decision-making. In its present state, an LLM configured for oncology treatment recommendations functions comparably to an enthusiastic but junior hematology resident participating in a multidisciplinary tumour board discussion. Like a well-read resident, the LLM demonstrates comprehensive knowledge of the treatment landscape, can synthesize guideline recommendations and recent trial data, and occasionally introduces therapeutic options or evidence that senior clinicians had not yet considered. In the context of a subspecialist-led discussion, such contributions are valuable: they bring an additional perspective with limited decisional influence, and senior clinicians possess the expertise to evaluate, filter, and override suggestions that contain subtle errors in clinical reasoning or that fail to account for nuanced eligibility constraints, regulatory requirements, or patient-specific contraindications.

The danger emerges when this same "resident" operates without adequate supervision. A junior trainee who confidently presents a well-referenced treatment recommendation to a non-specialist clinician, one who lacks the depth of domain expertise to critically evaluate the suggestion, may inadvertently lead to adoption of a plausible but flawed treatment plan. The enthusiasm and apparent comprehensiveness of the recommendation, combined with the non-specialist's relative unfamiliarity with the nuances of the therapeutic landscape, create conditions where errors propagate rather than being detected and corrected. This is precisely the pattern observed in the present study: generalists, encountering a confident, well-reasoned LLM recommendation supported by appropriate citations and clinical terminology, accepted it at face value in scenarios where subspecialists identified genuine clinical concerns.

This analogy carries a direct implication for deployment strategy. Just as a hospital would not permit a junior resident to independently manage complex lymphoma cases without attending supervision, current LLM systems should not be deployed as autonomous decision support tools for non-specialist clinicians managing complex oncology scenarios without structured oversight mechanisms or subspecialist verification.

Inter-Rater Agreement and Detection Asymmetry

The inter-rater agreement patterns between expert types provide additional support for the expertise-dependent safety concern. The observed agreement on clinical acceptability of 82% (41 of 50 cases) with Cohen's kappa of 0.384 (fair agreement) indicates moderate concordance substantially better than chance but well short of the consistency required for interchangeable clinical decision-making. This level of agreement is broadly consistent with inter-rater reliability documented in other domains of clinical judgment, where experienced clinicians frequently reach different conclusions when evaluating identical clinical scenarios.

However, the disagreements were not random. McNemar's test ($p = 0.039$) indicates that subspecialist expertise consistently detected inappropriateness that generalists missed: generalists accepted LLM recommendations in 8 cases where subspecialists identified unacceptability, while only 1 case showed the reverse pattern. Several mechanisms may explain this detection asymmetry. Subspecialists possess more comprehensive and current knowledge of nuanced eligibility criteria, regulatory approval status, contraindications, and rapidly evolving guideline recommendations. They may employ more sophisticated heuristics for evaluating clinical plausibility, developed through extensive exposure to R/R DLBCL cases. Additionally, subspecialists may apply more stringent evaluation standards when assessing AI-generated recommendations, recognizing that acceptance of an "adequate" recommendation from an automated system carries different implications than accepting a recommendation from a trusted human colleague.

The moderate agreement on rating scale scores (Pearson correlations of 0.569 for guideline adherence and 0.550 for clinical appropriateness) and the fair agreement on case complexity ratings (weighted kappa 0.403) with no systematic directional bias indicate that generalists and subspecialists perceived case difficulty similarly despite reaching different conclusions about LLM acceptability. This dissociation suggests that subspecialists' higher rejection rates reflected enhanced capability to detect specific recommendation flaws rather than a general tendency to view cases as more challenging or to apply arbitrarily higher standards.

LLM Error Patterns

Reframing the Nature of LLM Errors

The qualitative analysis of rejected recommendations revealed identifiable vulnerability domains in how the LLM processed complex clinical information, but an important epistemological caveat must be made: these errors cannot be characterized as "systematic" in the conventional scientific sense. A systematic error implies a reproducible, predictable deviation that occurs under identifiable conditions and can therefore be anticipated and corrected. The LLM errors observed in this study do not meet this definition. They represent recurrent failure patterns within identifiable problem domains, temporal information processing and regulatory constraint application, but their occurrence within these domains was inconsistent, non-deterministic, and resistant to prediction based on input characteristics.

The temporal information processing failures (Cases 3 and 11) illustrate this distinction. In these cases, the LLM correctly understood that relapse timing determines treatment eligibility and correctly recognized that different regulatory and guideline criteria apply to early versus late

relapse in second line DLBCL. The failure occurred at the information prioritization stage: when case narratives contained multiple temporal metrics (time from initial diagnosis, time from treatment initiation, time from treatment completion), the LLM inconsistently identified which metric was clinically relevant for determining relapse timing category. Clinicians resolve this ambiguity through heuristics developed over years of clinical practice: when a patient's history includes "diagnosed 18 months ago, relapsed 8 months after completing last chemotherapy," the expert immediately recognizes that treatment eligibility is determined by the 8-month interval (time from treatment completion to relapse) rather than the 18-month interval (time from diagnosis). The LLM's inconsistent performance on this task suggests a limitation in implementing robust relevance determination when multiple similar data elements are present.

Critically, stress testing with intentionally constructed cases containing conflicting temporal metrics confirmed that this vulnerability was not deterministic. Specific narrative structures and presentation order appeared to influence whether the LLM correctly identified the relevant interval, but no consistent pattern emerged. This non-reproducibility has profound implications for safety: a deterministic error, once identified, can be systematically tested for and potentially mitigated. A stochastic error cannot. It may surface unpredictably in clinical use, cannot be reliably eliminated through conventional quality assurance, and complicates any regulatory validation framework that assumes comprehensive testing can adequately characterize system performance.

The regulatory compliance errors represent a different vulnerability domain with similar non-reproducibility implications. Five cases involved recommendation of lisocabtagene maraleucel for second-line therapy, an indication approved by the FDA and supported by clinical trial evidence but not yet approved by AIFA at the time of case evaluation. An additional case involved recommendation of epcoritamab in second line before Italian regulatory approval. These errors occurred despite the knowledge base explicitly documenting AIFA approval status and the system prompt specifically directing the LLM to verify regulatory eligibility. The LLM's web search capability likely contributed by accessing content describing international approval status and guideline recommendations that contradicted the curated knowledge base reflecting Italian regulatory constraints. However, not all cases with similar regulatory complexities triggered this error, again demonstrating the non-deterministic nature of these failures.

The Black Box Problem

Both error categories share a fundamental characteristic that distinguishes them from errors in conventional medical decision support systems: their root causes are opaque. Current transformer-based architectures do not provide interpretable reasoning traces that would allow post-hoc identification of why a specific error occurred in one instance but not another. The models function as black boxes that sometimes produce incorrect outputs, potentially very harmful ones, without any mechanism for understanding, predicting, or preventing these failures at the individual case level. This opacity has direct consequences for clinical deployment: it is not

possible to design a targeted safeguard against an error whose triggering conditions cannot be specified, nor is it possible to assure clinicians or regulators that a particular class of errors has been adequately addressed when the conditions for their occurrence remain unknown.

The confident presentation of erroneous recommendations compounds this concern. All unacceptable recommendations received detailed, plausible-sounding rationales incorporating appropriate clinical terminology and citation of evidence. Unlike human consultants, whose communication style often implicitly conveys confidence levels through hedging language, acknowledgment of uncertainty, or explicit discussion of alternative approaches, the LLM generated equivalently confident-appearing outputs regardless of whether its reasoning was sound. This uniformly confident presentation is particularly problematic in the context of non-reproducible errors: a clinician cannot develop calibrated trust in the system because there are no reliable signals indicating when a given recommendation is more or less likely to contain errors.

The Benchmark-to-Bedside Translation Gap

The findings of this study contribute to accumulating evidence of a critical disconnect between LLM performance on standardized medical assessments and utility in authentic clinical decision-making contexts. While recent studies have documented LLM capabilities on medical licensing examinations, with models achieving passing scores on USMLE-style questions and demonstrating competitive performance on various clinical vignette benchmarks, the present results illustrate specific mechanisms through which benchmark success fails to predict real-world clinical utility.

The R/R DLBCL treatment selection scenario employed in this study exemplifies authentic clinical complexity that differs qualitatively from standardized assessments. Genuine therapeutic equipoise exists: multiple treatment approaches may be reasonable for the same patient. Decision-making requires integration of incomplete and evolving evidence: direct head-to-head trials comparing available therapies are largely absent, regulatory approval status changes periodically, and clinical guidelines necessarily lag behind the rapid introduction of novel agents. Constraint satisfaction is critical: identifying the optimal therapeutic approach requires systematic verification that the patient meets eligibility criteria, that no contraindications exist, that the treatment is accessible in the specific healthcare context, and that the recommendation aligns with current regulatory requirements.

The information prioritization failures and regulatory compliance errors identified in this study would not be detected by multiple-choice examinations. A USMLE-style question about appropriate second-line therapy would likely not include a distractor option representing a

treatment with FDA but not AIFA approval, and even if present, successful elimination of this option would not demonstrate the model's capability to verify regulatory status during real-world recommendation generation. Moreover, the non-reproducibility of these errors means that even targeted adversarial benchmarks would provide incomplete assurance: a model that correctly handles conflicting temporal metrics in a test set may fail on a structurally similar case in clinical deployment.

Recent randomized trial evidence reinforces this concern. Goh et al. evaluated 50 physicians using either an LLM diagnostic aid or conventional resources and found no significant improvement in clinical reasoning performance despite the LLM alone demonstrating higher accuracy than either physician group (11). This finding suggests that effective integration of AI capabilities into clinical workflows requires more than technical accuracy: it demands appropriate human-AI interaction patterns, effective communication of uncertainty and preservation of appropriate clinical scepticism.

The disconnect between benchmark and bedside performance has implications for AI regulatory frameworks. Current evaluation approaches that assess aggregate accuracy across standardized test sets may fail to detect vulnerabilities that manifest non-deterministically under specific information structures. The non-reproducible error patterns identified in this study suggest the need for adversarial testing approaches that deliberately construct scenarios designed to probe known vulnerability domains, analogous to stress testing in engineering. However, even comprehensive adversarial testing faces a fundamental limitation when errors are stochastic: passing a stress test does not guarantee safe performance on structurally similar cases in deployment.

Confidence Changes and the Psychology of AI-Assisted Decision-Making

The patterns of confidence change following LLM exposure provide insight into how clinicians psychologically integrate AI-generated recommendations and reveal important differences between expert types. Overall, 63% of evaluators reported no change in confidence after reviewing the LLM recommendation, 25% reported increased confidence, and 12% reported decreased confidence. However, these aggregate statistics masked substantial heterogeneity: generalists reported increased confidence in 40% of evaluations versus only 10% for subspecialists ($\chi^2 = 16.06$, $p = 0.0003$), with subspecialists predominantly reporting stable confidence (82% vs 44%).

The predominant stable-confidence pattern for subspecialists suggests that LLM recommendations rarely provided information that caused experts to meaningfully reconsider their certainty. In contrast, generalists experienced confidence increases substantially more frequently than decreases

(40% vs 16%), suggesting that LLM recommendations primarily functioned as a validating resource, helping narrow the range of options.

The relationship between confidence changes and behavioral influence reveals that most modifications represented adoption of what the evaluator viewed as a superior approach rather than reluctant capitulation: among the 19 evaluations where recommendation modification occurred, 15 were accompanied by confidence increases, and 17 of 19 modifications moved toward the LLM recommendation. However, a concerning pattern emerged: among evaluations where the LLM recommendation was deemed unacceptable, 71% resulted in increased confidence and only 6% in decreased confidence. This "negative validation" effect, wherein exposure to inferior alternatives strengthened evaluators' conviction in their original judgment, may seem benign, but raises the question of whether this confidence boost was always epistemically warranted.

From a clinical decision support design perspective, these patterns suggest that current LLM implementations do not optimally communicate uncertainty and reasoning transparency. An ideal decision support tool would help clinicians calibrate their confidence appropriately: increasing certainty when convergent recommendations are based on strong evidence, maintaining uncertainty when genuine equipoise exists, and prompting additional scrutiny when the recommendation rests on assumptions that warrant verification.

Limitations and Considerations

This study has several methodological limitations that should inform interpretation of findings and planning of subsequent research.

The sample size of 50 cases yielding 100 evaluations was determined by feasibility constraints rather than formal power calculations, as no preliminary data existed to guide sample size estimation for this novel investigation. While adequate for estimating overall acceptability with reasonable precision (95% CI width of approximately 15 percentage points), the study has limited power for detecting small-to-moderate effect sizes in comparative analyses. The exploratory comparison of acceptability between expert types (90% vs 76%, $p = 0.108$) exemplifies this limitation: the observed 14 percentage point difference has potential clinical significance, but the wide confidence intervals preclude definitive conclusions. Future research with larger sample sizes would be needed to establish whether this difference represents true population variation or chance fluctuation.

Generalizability is constrained by several factors. Evaluator recruitment via convenience sampling from the principal investigator's professional network at a single institution introduces selection bias. Participants may differ from the broader hematology community in ways that affect LLM

evaluation: willingness to volunteer for AI-related research suggests potential interest or comfort with technology, practice within academic or research-oriented settings, and recruitment from personal contacts suggests potential courtesy bias. The focus on Italian hematologists familiar with AIFA regulatory requirements limits international applicability, as treatment availability, regulatory constraints, and practice patterns differ substantially across healthcare systems.

The use of synthetic rather than real patient cases represents a fundamental limitation, as artificially constructed scenarios cannot fully replicate authentic decision-making complexity. Real patients present with ambiguous or incomplete information, diagnostic uncertainty, evolving clinical status, communication challenges, family dynamics, institutional resource constraints, and the psychological weight of actual treatment consequences. Evaluators approached synthetic cases knowing they would not bear real clinical responsibility, potentially altering risk tolerance, scrutiny, or willingness to accept approaches they might reject in actual practice.

The evaluation procedure had structural limitations. The Phase 1/Phase 2 sequencing relied on evaluator adherence to instructions rather than technological enforcement, introducing potential protocol violations if evaluators viewed LLM recommendations before completing independent assessment. The absence of baseline confidence measurements prevents distinction between high-confidence confirmation and persistent uncertainty when evaluators reported "no change in confidence." The harm potential classification scheme, while clinically logical, lacks empirical validation and may incorrectly classify some rejections; notably, the conservative exclusion of regulatory non-compliance from harm potential may underestimate safety implications, as some regulatory restrictions exist specifically because treatments are ineffective or harmful in excluded populations.

The LLM reproducibility constraints inherent in using commercial LLM capabilities with unrestricted web search enhance external validity, but compromise reproducibility. The system's web search behavior, source selection, and information synthesis were not explicitly controlled or logged in detail, meaning subsequent evaluations using the same cases might produce different recommendations. This variability reflects the fundamental nature of probabilistic AI systems and is consistent with the non-reproducible error patterns identified, but it complicates assessment of which specific features drove observed performance patterns.

Finally, the study evaluated a single LLM model (Claude Sonnet 4.5) with one prompt configuration and knowledge base structure at a specific point in time (September 2025 – January 2026). Performance may differ substantially with alternative models, different prompting strategies, varied knowledge base architectures, or with the same model at future timepoints. The rapid evolution of LLM technology means that findings may have limited durability, and specific error patterns identified might be addressed in subsequent model versions. However, the fundamental failure modes observed may reflect broader characteristics of current transformer

architectures rather than model-specific limitations, suggesting that similar patterns could persist across implementations until architectural advances address these reasoning constraints.

Implications for Clinical Practice and Future Directions

The expertise-dependent patterns identified in this study carry direct implications for how LLM clinical decision support systems should be deployed, designed, and regulated.

For clinical practice, the central implication is that undifferentiated deployment strategies are inappropriate. The detection and behavioral consequence gaps documented above demonstrate that the same system can function as a low-risk confirmatory resource for subspecialists or as a potential vector for harm when used by generalists without oversight. Implementation models should therefore be stratified by user expertise. Subspecialists, who derive minimal decision support value but possess the expertise to filter errors, could access full-capability systems serving primarily as efficient literature synthesis and confirmatory resources. Generalists, who show substantial responsiveness but reduced error detection capability, would benefit from enhanced safety configurations: automatic flagging of recommendations involving known vulnerability domains, required verification steps for regulatory eligibility and treatment sequencing, or mandatory subspecialist review for complex cases. Alternatively, LLMs could be positioned explicitly as educational rather than directive tools for non-specialist users, providing synthesis of treatment options and evidence while requiring clinicians to independently verify key constraints before adoption. Hybrid human-AI workflows, where LLMs perform specific subtasks such as literature search or eligibility screening rather than generating complete treatment recommendations, would preserve human judgment for the integrative reasoning that current systems handle poorly.

For AI system design, the non-reproducible error patterns and the uniformly confident presentation of both sound and flawed recommendations suggest that conventional improvements such as knowledge base expansion or prompt engineering may be insufficient. If errors cannot be reliably triggered, they cannot be reliably tested for, and if errors are presented with the same confidence as correct outputs, users cannot develop calibrated trust. Architectures that provide interpretable reasoning traces, explicit uncertainty quantification, or structured self-verification of constraint satisfaction may be necessary before clinical deployment in high-stakes domains.

For regulatory frameworks, the non-reproducibility finding poses a fundamental challenge to traditional medical device validation, which assumes that identical inputs produce identical outputs and that comprehensive testing predicts future performance. LLMs violate both assumptions. Regulatory approaches should therefore require characterization of known vulnerability domains, explicit disclosure of non-reproducibility limitations, and evidence that

deployment includes safeguards appropriate to the expertise level of intended users. The validation framework should move from exhaustive input-output testing towards demonstration of safe human-AI interaction under realistic conditions.

For future research, several priorities emerge. Adequately powered confirmatory studies are needed to establish whether the expertise-dependent influence patterns represent true population characteristics or reflect the specific conditions of this pilot investigation. Blinded comparative evaluation, presenting identical recommendations attributed to AI versus human sources, would isolate the role of source attribution in clinician responses and determine how much the observed patterns reflect genuine quality assessment versus algorithmic aversion or over-trust. Outcome-focused evaluations are ultimately necessary to determine whether LLM-augmented decision-making improves patient outcomes, though such studies present substantial ethical and logistical challenges. Finally, the psychological dimension of AI-assisted clinical decision-making requires deeper investigation. We should map the factors driving algorithmic trust calibration and the role of professional identity in shaping the users' willingness to modify their judgment, before such technologies can be successfully deployed.

Conclusion

This study set out to evaluate whether a commercially available LLM system could generate clinically acceptable treatment recommendations for R/R DLBCL, and whether physician expertise modulates the safety and utility of such recommendations. The answer to both questions carries important nuance.

The substantial overall acceptability achieved by the Ariadne system confirms that current LLM architectures, when augmented with domain-specific knowledge and structured reasoning frameworks, can operate within the bounds of evidence-based practice for a majority of complex oncology scenarios. Yet this capability coexists with fundamental limitations that preclude safe autonomous deployment. The errors identified in this study were not knowledge gaps amenable to straightforward correction: they were stochastic reasoning failures within identifiable vulnerability domains, opaque in origin and resistant to prediction or prevention through conventional quality assurance. For a technology positioned as a clinical decision support tool, non-reproducible errors that cannot be reliably detected by the system itself represent a qualitatively different safety challenge than the calibrated, learnable fallibility of human consultants.

The most consequential finding, however, concerns the interaction between LLM output and physician expertise. The marked divergence in behavioral influence between subspecialists and generalists, and the demonstration that generalists not only failed to detect harmful

recommendations but in some instances actively adopted them, reveals that the same AI system can function as a useful confirmatory resource or as a vector for harm depending entirely on who receives its output. This expertise-dependent deployment paradox has a direct practical implication: undifferentiated implementation strategies are inappropriate. Deployment models must be stratified by user expertise, with enhanced safeguards, mandatory verification steps, or subspecialist oversight mechanisms for contexts where non-specialist clinicians interact with AI-generated recommendations in complex therapeutic domains.

More broadly, these findings challenge the prevailing emphasis on aggregate accuracy as the primary metric for evaluating clinical AI systems. Overall acceptability rates, while informative, obscure the non-reproducible failures that pose the greatest safety risk and reveal nothing about how errors propagate through the human-AI interaction. Future evaluation frameworks must incorporate behavioral influence assessment stratified by user expertise, characterization of error reproducibility and harm potential, and validation that intended users can reliably identify system failures. Until such frameworks are established and the architectural limitations underlying stochastic reasoning failures are addressed, LLM-based clinical decision support should be regarded as a promising but provisional technology, whose safe integration into clinical practice demands the same graduated oversight structures that govern the supervised autonomy of trainees in medical education.

Bibliography

1. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. Dagan A, editor. PLOS Digit Health. 2023 Feb 9;2(2):e0000198.

2. Mihalache A, Huang RS, Popovic MM, Muni RH. ChatGPT-4: An assessment of an upgraded artificial intelligence chatbot in the United States Medical Licensing Examination. Med Teach. 2024 Mar 3;46(3):366–72.

3. Dinc MT, Bardak AE, Bahar F, Noronha C. Comparative analysis of large language models in clinical diagnosis: performance evaluation across common and complex medical cases. JAMIA Open. 2025 May 2;8(3):ooaf055.

4. Gu Y, Fu J, Liu X, Valanarasu JMJ, Codella NC, Tan R, et al. The Illusion of Readiness: Stress Testing Large Frontier Models on Multimodal Medical Benchmarks [Internet]. arXiv; 2025 [cited 2025 Oct 24]. Available from: http://arxiv.org/abs/2509.18234

5. Deng C, Zhao Y, Tang X, Gerstein M, Cohan A. Investigating Data Contamination in Modern Benchmarks for Large Language Models. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language

Technologies (Volume 1: Long Papers) [Internet]. Mexico City, Mexico: Association for Computational Linguistics; 2024 [cited 2025 Oct 24]. p. 8706–19. Available from: https://aclanthology.org/2024.naacl-long.482

6. Ma Z, Wang W, Yu G, Cheung YF, Ding M, Liu J, et al. Beyond the Leaderboard: Rethinking Medical Benchmarks for Large Language Models [Internet]. arXiv; 2025 [cited 2025 Oct 24]. Available from: http://arxiv.org/abs/2508.04325

7. McGervey M, Olson A, Mohmand M, Restrepo D. Clinical Reasoning and Diagnostic Errors. Med Clin North Am. 2025 Sep;109(5):997–1008.

8. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. BMC Med Inform Decis Mak. 2025 Mar 7;25(1):117.

9. Rodman A, Buckley TA, Manrai AK, Morgan DJ. Artificial Intelligence vs Clinician Performance in Estimating Probabilities of Diagnoses Before and After Testing. JAMA Netw Open. 2023 Dec 11;6(12):e2347075.

10. Goodman RS, Patrinely JR, Stone CA, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and Reliability of Chatbot Responses to Physician Questions. JAMA Netw Open. 2023 Oct 2;6(10):e2336483.

11. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial. JAMA Netw Open. 2024 Oct 28;7(10):e2440969.

12. Bean AM, Payne RE, Parsons G, Kirk HR, Ciro J, Mosquera-Gómez R, et al. Reliability of LLMs as medical assistants for the general public: a randomized preregistered study. Nat Med [Internet]. 2026 Feb 9 [cited 2026 Feb 13]; Available from: https://www.nature.com/articles/s41591-025-04074-y

13. Chen S, Gao M, Sasse K, Hartvigsen T, Anthony B, Fan L, et al. When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. Npj Digit Med. 2025 Oct 17;8(1):605.

14. Sharma M, Tong M, Korbak T, Duvenaud D, Askell A, Bowman SR, et al. Towards Understanding Sycophancy in Language Models [Internet]. arXiv; 2025 [cited 2025 Oct 27]. Available from: http://arxiv.org/abs/2310.13548

15. Sehn LH, Salles G. Diffuse Large B-Cell Lymphoma. Longo DL, editor. N Engl J Med. 2021 Mar 4;384(9):842–58.

16. Kanas G, Ge W, Quek RGW, Keeven K, Nersesyan K, Jon E. Arnason. Epidemiology of diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) in the United States and Western Europe: population-level projections for 2020–2025. Leuk Lymphoma. 2022 Jan 2;63(1):54–63.

17. Coiffier B, Sarkozy C. Diffuse large B-cell lymphoma: R-CHOP failure—what to do? Hematology. 2016 Dec 2;2016(1):366–78.

18. Friedberg JW. Relapsed/Refractory Diffuse Large B-Cell Lymphoma. Hematology. 2011 Dec 10;2011(1):498–505.

19. Crump M, Neelapu SS, Farooq U, Van Den Neste E, Kuruvilla J, Westin J, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. Blood. 2017 Oct 19;130(16):1800–8.

20. Gong I, Crump M, Prica A, Calzavara A, Liu N, Rodin D, et al. Outcomes and Predictors of Survival for Patients with Relapsed or Refractory Diffuse Large B Cell Lymphoma: A Large Population-Based Analysis. Blood. 2024 Nov 5;144(Supplement 1):3092–3092.

21. Tilly H, Gomes Da Silva M, Vitolo U, Jack A, Meignan M, Lopez-Guillermo A, et al. Diffuse large B-cell lymphoma (DLBCL): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2015 Sep;26:v116–25.

22. Zelenetz AD, Gordon LI, Abramson JS, Advani RH, Andreadis B, Bartlett NL, et al. NCCN Guidelines® Insights: B-Cell Lymphomas 3.2025: Featured Updates to the NCCN Guidelines. J Natl Compr Canc Netw. 2025 Oct;23(10):e250048.

23. Philip T, Guglielmi C, Hagenbeek A, Somers R, Van Der Lelie H, Bron D, et al. Autologous Bone Marrow Transplantation as Compared with Salvage Chemotherapy in Relapses of Chemotherapy-Sensitive Non-Hodgkin's Lymphoma. N Engl J Med. 1995 Dec 7;333(23):1540–5.

24. Gisselbrecht C, Glass B, Mounier N, Singh Gill D, Linch DC, Trneny M, et al. Salvage Regimens With Autologous Transplantation for Relapsed Large B-Cell Lymphoma in the Rituximab Era. J Clin Oncol. 2010 Sep 20;28(27):4184–90.

25. Crump M, Kuruvilla J, Couban S, MacDonald DA, Kukreti V, Kouroukis CT, et al. Randomized Comparison of Gemcitabine, Dexamethasone, and Cisplatin Versus Dexamethasone, Cytarabine, and Cisplatin Chemotherapy Before Autologous Stem-Cell Transplantation for Relapsed and Refractory Aggressive Lymphomas: NCIC-CTG LY.12. J Clin Oncol. 2014 Nov 1;32(31):3490–6.

26. Jommi C, Bramanti S, Pani M, Ghirardini A, Santoro A. CAR T-Cell Therapies in Italy: Patient Access Barriers and Recommendations for Health System Solutions. Front Pharmacol. 2022 Jun 23;13:915342.

27. Locke FL, Miklos DB, Jacobson CA, Perales MA, Kersten MJ, Oluwole OO, et al. Axicabtagene Ciloleucel as Second-Line Therapy for Large B-Cell Lymphoma. N Engl J Med. 2022 Feb 17;386(7):640–54.

28. Kamdar M, Solomon SR, Arnason J, Johnston PB, Glass B, Bachanova V, 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. The Lancet. 2022 Jun;399(10343):2294–308.

29. Dickinson MJ, Carlo-Stella C, Morschhauser F, Bachy E, Corradini P, Iacoboni G, et al. Glofitamab for Relapsed or Refractory Diffuse Large B-Cell Lymphoma. N Engl J Med. 2022 Dec 15;387(24):2220–31.

30. Thieblemont C, Phillips T, Ghesquieres H, Cheah CY, Clausen MR, Cunningham D, et al. Epcoritamab, a Novel, Subcutaneous CD3xCD20 Bispecific T-Cell–Engaging Antibody, in Relapsed or Refractory Large B-Cell Lymphoma: Dose Expansion in a Phase I/II Trial. J Clin Oncol. 2023 Apr 20;41(12):2238–47.

31. Sehn LH, Hertzberg M, Opat S, Herrera AF, Assouline S, Flowers CR, et al. Polatuzumab vedotin plus bendamustine and rituximab in relapsed/refractory DLBCL: survival update and new extension cohort data. Blood Adv. 2022 Jan 25;6(2):533–43.

32. Salles G, Duell J, González Barca E, Tournilhac O, Jurczak W, Liberati AM, et al. Tafasitamab plus lenalidomide in relapsed or refractory diffuse large B-cell lymphoma (L-MIND): a multicentre, prospective, single-arm, phase 2 study. Lancet Oncol. 2020 Jul;21(7):978–88.

33. Caimi PF, Ai W, Alderuccio JP, Ardeshna KM, Hamadani M, Hess B, et al. Loncastuximab tesirine in relapsed or refractory diffuse large B-cell lymphoma (LOTIS-2): a multicentre, open-label, single-arm, phase 2 trial. Lancet Oncol. 2021 Jun;22(6):790–800.

34. Haydu JE, Abramson JS. The rules of T-cell engagement: current state of CAR T cells and bispecific antibodies in B-cell lymphomas. Blood Adv. 2024 Sep 10;8(17):4700–10.

35. Trabolsi A, Arumov A, Schatz JH. Bispecific antibodies and CAR-T cells: dueling immunotherapies for large B-cell lymphomas. Blood Cancer J. 2024 Feb 8;14(1):27.

36. Kim J, Cho J, Lee MH, Yoon SE, Kim WS, Kim SJ. CAR T cells vs bispecific antibody as third- or later-line large B-cell lymphoma therapy: a meta-analysis. Blood. 2024 Aug 8;144(6):629–38.

37. Salles G, Fox CP, Hamadani M, Wang A, Sail K, Alshreef A, et al. Indirect Comparison of Epcoritamab Versus Axicabtagene Ciloleucel in Chimeric Antigen Receptor T-Cell–Eligible and –Naive Patients With Relapsed/Refractory Diffuse Large B-Cell Lymphoma. Clin Lymphoma Myeloma Leuk. 2025 Nov;25(11):e986-e994.e4.

38. Rosenthal A, Munoz J, Jun M, Wang T, Mutebi A, Wang A, et al. Comparisons of treatment outcomes of epcoritamab versus chemoimmunotherapy, polatuzumab-based regimens, tafasitamab-based regimens, or chimeric antigen receptor T-cell therapy, in third-line or later relapsed/refractory large B-cell lymphoma. J Hematol OncolJ Hematol Oncol. 2024 Aug 16;17(1):69.

39. Crochet G, Iacoboni G, Couturier A, Bachy E, Iraola-Truchuelo J, Gastinne T, et al. Efficacy of CAR T-cell therapy is not impaired by previous bispecific antibody treatment in large B-cell lymphoma. Blood. 2024 Jul 18;144(3):334–8.

40. Shumilov E, Scholz JK, Seib M, Mazzeo P, Wurm-Kuczera R, Vucinic V, et al. Outcomes of bispecific antibody therapy after CAR T-cell failure in relapsed/refractory large B-cell lymphoma. Blood Adv. 2025 Aug 12;9(15):3955–66.

41. Shadman M, Ahn KW, Kaur M, Lekakis L, Beitinjaneh A, Iqbal M, et al. Autologous transplant vs. CAR-T therapy in patients with DLBCL treated while in complete remission. Blood Cancer J. 2024 Jul 8;14(1):108.

42. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000 Feb;403(6769):503–11.

43. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med. 2002 Jan;8(1):68–74.

44. Friedberg JW. Double-Hit Diffuse Large B-Cell Lymphoma. J Clin Oncol. 2012 Oct 1;30(28):3439–43.

45. Chapuy B, Wood T, Stewart C, Dunford A, Wienand K, Khan SJ, et al. DLB *class* : a probabilistic molecular classifier to guide clinical investigation and practice in diffuse large B-cell lymphoma. Blood. 2025 May 1;145(18):2041–55.

46. Wright GW, Huang DW, Phelan JD, Coulibaly ZA, Roulland S, Young RM, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. Cancer Cell. 2020 Apr;37(4):551-568.e14.

APPENDIX "A"

Realistic R/R DLBCL case generation prompt and JSON files used for the generation

CASE GENERATION PROMPT "CASE ENSEMBLE"

```
You are a clinical AI model specialized in generating realistic,
guideline-compliant case reports for Diffuse Large B-cell Lymphoma (DLBCL)
and High-Grade B-cell Lymphoma (HGBL).

**IMPORTANT: You MUST use the knowledge retrieval capabilities of this
platform to query the uploaded files:**
1.  **`DLBCL_Case_Schema_v5.json`**: Use this file as the definitive source
for the required JSON structure, field names, data types, enums,
descriptions, and required fields. All generated JSON output MUST conform
strictly to this schema definition retrieved from the knowledge file.
2.  **`treatments.json`**: Query this file to identify appropriate treatment
regimens based on patient characteristics (age, ECOG, prior lines, etc.) and
to retrieve associated data (e.g., typical response rates, toxicities) when
available.
3.  **`guidelines.json`**: Query this file for eligibility rules (e.g.,
NCCN/ESMO alignment, AIFA rules if included, biomarker requirements) to
confirm the appropriateness of chosen treatments.
4.  **`extranodal_sites.json`**: Query this file for details on specific
extranodal sites, including typical presentation, CNS relapse risk, common
molecular profiles, and prognostic implications, when populating relevant
fields.

---

🔁 STEP 1: Generate Case Variability Parameters

Simulate the execution of the following Python function called
`generate_case_variability()` to generate unique starting parameters for
this case. Use the returned `variability` dictionary to guide the initial
patient profile.

```python
import random

def generate_case_variability():
    variability = {
        "random_seed": random.randint(100000, 999999),
        "relapse_timing_category": random.choices(["<6mo", "6-12mo",
">12mo", "Primary Refractory"], weights=[0.20, 0.30, 0.35, 0.15])[0],
        "presentation": random.choices(["symptomatic", "asymptomatic"],
weights=[0.65, 0.35])[0],
```
```

```
        "eco_category": random.choices(["fit", "unfit"], weights=[0.7,
0.3])[0], # Influences ECOG & comorbidity likelihood
        "comorbidity_target": random.choices([0, 1, 2, 3, 4, 5, 6, 7],
weights=[0.10, 0.15, 0.20, 0.20, 0.15, 0.10, 0.05, 0.05])[0],
        "stage_group": random.choices(["I, II, III, IV"], weights=[0.10,
0.25, 0.35, 0.30])[0],
        "lines": random.choices([1, 2, 3, 4, 5], weights=[0.3, 0.40, 0.15,
0.10, 0.05])[0]
    }
    variability["comorbidity_instructions"] = '''
    Generate a list of medically recognized comorbidities aiming for a
number close to `comorbidity_target`.
    Ensure the type/severity aligns with the patient's age and
`eco_category`.
    Query `DLBCL_Case_Schema_v5.json` for the expected format of the
'comorbidities' field.
    '''
    return variability

# AI: Execute this simulation internally to get the 'variability'
dictionary.
variability = generate_case_variability()
```

✳ STEP 2: Generate the Case JSON Object

Generate the *entire* case as a single JSON object called `case`.
CRITICAL: Adhere strictly to the structure, field names, types, enums, and requirements defined in the `DLBCL_Case_Schema_v5.json` file (available via knowledge retrieval).

Instructions for specific sections:

* **`initial_diagnosis`**: Populate fields based on `variability` parameters (stage, presentation) and plausible clinical details.
        * **Extranodal Sites:** If generating extranodal involvement (e.g., in `extranodal_sites` field or influencing Stage IE/IIE/IV), **query `extranodal_sites.json`** for the chosen site(s). Ensure the details (e.g., likelihood of presentation stage, associated risks like CNS relapse) are consistent with the information retrieved from `extranodal_sites.json`. For example, if generating Testicular involvement, note its typical localized presentation but high CNS risk. If generating Kidney/Adrenal involvement, note its association with Stage IV and high CNS risk.
        * **IPI Calculation:** Populate the `ipi_components` object first. Then, **explicitly calculate `initial_ipi_score` by summing the number of `true` values within the `ipi_components` object.** The resulting sum **must** be the value assigned to `initial_ipi_score`.

```
     Query the schema file for details on fields like `ldh_level` object
structure.
* **`treatment_history`**:
    * Generate `variability['lines']` number of treatment lines.
    * For each line:
        * **Determine Patient Status:** Note the patient's age, ECOG,
comorbidities, and disease status *at the start of this line*.
        * **Query Knowledge:** Search `treatments.json` for suitable
regimens matching the patient status. Search `guidelines.json` to verify
eligibility and guideline concordance.
        * **Select Regimen:** Choose a clinically plausible regimen based on
retrieved knowledge.
        * **Rationale:** Write the `rationale` field, EXPLICITLY citing
information retrieved from `treatments.json` and/or `guidelines.json` that
justifies the choice (e.g., "Chosen based on retrieved data from
treatments.json showing efficacy in ECOG 1, age < 65, and guidelines.json
confirming 2L eligibility").
        * **Populate Response/Toxicity:** Fill in `best_response`,
`remission_details`, and `toxicities` based on typical outcomes for the
chosen regimen (potentially informed by data retrieved from
`treatments.json`).
        Ensure `remission_details.timing_category` for the line preceding
relapse aligns with `variability['relapse_timing_category']`.
        * **Schema Adherence:** Ensure each line object strictly follows the
structure defined in `DLBCL_Case_Schema_v5.json`.
* **`relapsed_refractory_scenario`**: Detail the final relapse/refractory
state consistent with the last treatment line's outcome and
`variability['relapse_timing_category']`.
    **Revised IPI Calculation:** Populate `ipi_components_at_relapse` first.
Then, **explicitly calculate `revised_ipi_score` by summing the number of
`true` values within `ipi_components_at_relapse`.** The resulting sum
**must** be the value assigned to `revised_ipi_score`. Populate the
`timeline_consistency_check` based on calculated values.
    **Relapse Staging:** Populate the `sites_of_relapse` object first.
**When choosing specific extranodal sites for relapse, query
`extranodal_sites.json`** to ensure consistency. Then, **determine the
`ann_arbor_stage` within `disease_status` based *solely* on the populated
`sites_of_relapse`, following Ann Arbor principles:**
        * Involvement of extranodal sites is **Stage IV**.
        * Involvement of lymph node regions on *both* sides of the diaphragm
(with or without localized extranodal extension 'E' or spleen 'S') is
**Stage III**.
        * Involvement of lymph node regions on *only one* side of the
diaphragm is **Stage I** (single region) or **Stage II** (multiple regions).
     Query the schema file for other structural details (e.g.,
`disease_status`, `repeat_biopsy_status` objects).

**Perform Internal Consistency Checks:** Before finalizing the JSON,
internally verify:
```

```
    * **IPI Score Check:** Verify that `initial_ipi_score` exactly matches
the sum of `true` values in `ipi_components`. Verify that
`revised_ipi_score` exactly matches the sum of `true` values in
`ipi_components_at_relapse`.
    * **Relapse Stage Check:** Verify the `ann_arbor_stage` in
`relapsed_refractory_scenario` accurately reflects the pattern of
`sites_of_relapse` according to the rules provided above.
    * Age progression is logical
(`timeline_consistency_check.is_age_consistent` should be true).
    * Number of treatment lines matches `variability['lines']`.
    * HGBL diagnosis aligns with Double/Triple Hit status as per the
schema's `allOf` rule (verify this rule by querying
`DLBCL_Case_Schema_v5.json`).

```python
# AI: Generate the JSON object 'case' here, adhering to all instructions
above
# and using knowledge retrieval from the specified files.
case = {
  # ... Full JSON content ...
}

# AI: Perform internal consistency checks as described above.
# If checks fail, revise the JSON until consistent before proceeding.
```

✳ STEP 3: Generate the Narrative Case Summary

Once the validated `case` JSON is complete, write a tumor board-style
narrative for the `case` as separate text.

* Base the narrative STRICTLY on the data within the generated `case` JSON.
* Describe the patient journey chronologically: initial presentation,
diagnosis, each treatment line.
* Mention `best_response`, `remission_details`, and significant `toxicities`
for each line.
* Conclude with the `relapsed_refractory_scenario`, the `comorbidities` and
the `concomitant_medications`.
* Do NOT add information not present in the JSON or speculate on future
steps.

**(Self-Correction Note): After generating the narrative, reread it one last
time to ensure it accurately reflects ONLY the information present in the
final `case` JSON object.**
```

DLBCL_CASE_SCHEMA_V5.JSON

```
"title": "DLBCL/HGBL Case Report Schema v5",
  "description": "Schema for generating synthetic DLBCL and HGBL case
reports, focusing on treatment history and relapse/refractory scenarios.",
  "type": "object",
  "properties": {
    "patient_demographics": {
      "type": "object",
      "properties": {
        "sex": {
          "type": "string",
          "enum": ["Male", "Female"]
        },
        "age_at_initial_diagnosis": {
          "type": "integer",
          "minimum": 18,
          "maximum": 95,
          "description": "Age in years at the time of initial lymphoma
diagnosis."
        }
      },
      "required": ["sex", "age_at_initial_diagnosis"]
    },
    "initial_diagnosis": {
      "type": "object",
      "properties": {
        "clinical_presentation": {
          "type": "object",
          "properties": {
            "presentation_type": {
              "type": "string",
              "enum": ["symptomatic", "asymptomatic"],
              "description": "'asymptomatic' for incidental findings,
'symptomatic' for cases presenting with symptoms (B symptoms, mass effect,
etc.)."
            },
            "presentation_details": {
              "type": "string",
              "description": "Brief description of presenting signs/symptoms
or circumstances of discovery."
            }
          },
          "required": ["presentation_type", "presentation_details"]
        },
        "histopathology": {
          "type": "object",
          "properties": {
            "diagnosis": {
              "type": "string",
```

```
              "enum": ["DLBCL", "HGBL"],
              "description": "Final integrated diagnosis."
            },
            "cell_of_origin": {
              "type": "string",
              "enum": ["GCB", "ABC/non-GCB", "Not Specified"],
              "description": "Determined by IHC (Hans algorithm) or gene
expression profiling."
            },
            "double_triple_hit_status": {
              "type": "string",
              "enum": ["Negative", "Double Hit", "Triple Hit", "Not
Assessed"],
              "description": "Status of MYC plus BCL2 and/or BCL6
rearrangements by FISH or other methods. Must be DH/TH if diagnosis is
HGBL."
            },
            "ki67_percent": {
                "type": ["integer", "null"],
                "minimum": 0,
                "maximum": 100,
                "description": "Ki-67 proliferation index (optional)."
            },
            "molecular_markers": {
                "type": "array",
                "items": {"type": "string"},
                "description": "List of other relevant molecular findings,
e.g., specific mutations (TP53, MYD88), IHC markers (optional)."
            }
          },
          "required": ["diagnosis", "cell_of_origin",
"double_triple_hit_status"]
        },
        "extranodal_involvement": {
          "type": "object",
          "properties": {
            "total_sites": {
              "type": "integer",
              "minimum": 0,
              "description": "Number of distinct extranodal sites involved."
            },
            "sites": {
              "type": "array",
              "items": {"type": "string"},
              "description": "List of involved extranodal sites (e.g., Bone
Marrow, Liver, CNS, Lung)."
            }
          },
          "required": ["total_sites", "sites"]
        },
```

```
        "ann_arbor_stage": {
          "type": "string",
          "enum": ["I", "II", "III", "IV"]
        },
        "bulky_disease": {
          "type": "object",
          "properties": {
              "status": {
                  "type": "string",
                  "enum": ["Yes", "No"],
                  "description": "Presence of bulky disease, typically
defined as >10 cm."
              },
              "measurement_cm": {
                  "type": ["number", "null"],
                  "description": "Optional: Size of largest mass in cm."
              }
          },
          "required": ["status"]
        },
        "ldh_level": {
          "type": "object",
          "properties": {
              "value": {
                  "type": ["number", "null"],
                  "description": "Actual LDH value if known."
              },
              "unit": {
                  "type": "string",
                  "enum": ["U/L", "ukat/L", "xULN", "Not Specified"],
                  "description": "Units or relation to Upper Limit of Normal
(ULN)."
              },
              "multiplier_uln": {
                  "type": ["number", "null"],
                  "description": "LDH level expressed as a multiple of the
ULN (e.g., 1.5 for 1.5x ULN). Required for IPI calculation if absolute value
not used."
              },
              "is_above_uln": {
                  "type": "boolean",
                  "description": "Simple indicator if LDH is elevated above
ULN. Must be consistent with other fields."
              }
          },
          "required": ["is_above_uln"],
          "description": "Lactate Dehydrogenase level at diagnosis. Must
provide enough information to determine if elevated for IPI."
        },
        "ecog_performance_status": {
```

```
          "type": "integer",
          "minimum": 0,
          "maximum": 5,
          "description": "ECOG performance status at diagnosis."
        },
           "ipi_components": {
          "type": "object",
          "properties": {
            "age_over_60": {"type": "boolean"},
            "ecog_over_1": {"type": "boolean", "description": "ECOG status
2, 3, 4, or 5"},
            "stage_3_or_4": {"type": "boolean"},
            "ldh_above_uln": {"type": "boolean", "description": "Should be
derived from ldh_level object"},
            "extranodal_sites_over_1": {"type": "boolean", "description":
"More than one extranodal site"}
          },
          "required": ["age_over_60", "ecog_over_1", "stage_3_or_4",
"ldh_above_uln", "extranodal_sites_over_1"],
          "description": "Boolean components used to calculate the IPI
score. Must sum correctly to initial_ipi_score."
        }
        "initial_ipi_score": {
          "type": "integer",
          "minimum": 0,
          "maximum": 5,
          "description": "International Prognostic Index score calculated at
diagnosis."
        },
      },
      "required": [
        "clinical_presentation", "histopathology", "extranodal_involvement",
        "ann_arbor_stage", "bulky_disease", "ldh_level",
"ecog_performance_status",
        "initial_ipi_score", "ipi_components"
      ]
    },
    "treatment_history": {
      "type": "object",
      "properties": {
        "number_of_prior_lines": {
          "type": "integer",
          "minimum": 1,
          "description": "Total number of systemic therapy lines
administered for this lymphoma."
        },
        "lines": {
          "type": "array",
          "minItems": 1,
          "items": {
```

```
            "type": "object",
            "properties": {
              "line_number": {"type": "integer", "minimum": 1},
              "regimen": {"type": "string", "description": "Name of the
treatment regimen (e.g., R-CHOP, Pola-R-CHP, GemOx, CAR-T Product Name,
ASCT)."},
              "age_at_line_start": {
                  "type": ["integer", "null"],
                  "description": "Optional: Patient's age when this line of
therapy was initiated."
              },
              "ecog_at_line_start": {
                  "type": ["integer", "null"],
                  "minimum": 0, "maximum": 5,
                  "description": "Optional: ECOG status immediately before
starting this line."
              },
              "rationale": {
                  "type": "string",
                  "description": "Clinical reasoning for choosing this
regimen at this time point (considering age, ECOG, prior treatments,
guidelines, relapse timing etc.)."
              },
              "duration": {
                  "type": "string",
                  "description": "Duration of treatment or number of cycles
administered (e.g., '6 cycles', 'Until progression', 'Completed 4 cycles')."
              },
              "best_response": {
                "type": "string",
                "enum": ["CR", "PR", "SD", "PD", "NE"],
                "description": "Best response achieved on this line
(Complete Response, Partial Response, Stable Disease, Progressive Disease,
Not Evaluable)."
              },
              "response_assessment_details": {
                  "type": "string",
                  "description": "Optional: Method and timing of response
assessment (e.g., 'End-of-treatment PET-CT', 'Interim PET after 2 cycles')."
              },
              "remission_details": {
                  "type": "object",
                  "properties": {
                      "duration_months": {
                          "type": ["number", "null"],
                          "minimum": 0,
                          "description": "Duration of remission in months
from start/end of treatment until progression (if applicable)."
                      },
                      "timing_category": {
```

```
                        "type": "string",
                        "enum": ["<6mo", "6-12mo", ">12mo", "Primary
Refractory", "Ongoing", "N/A"],
                        "description": "Categorization of remission
duration or refractoriness relative to this line. 'Primary Refractory' if
PD/SD is best response. 'Ongoing' if still in remission. '<6mo', '6-12mo',
'>12mo' for relapse timing after CR/PR."
                      }
                  },
                  "required": ["timing_category"],
                  "description": "Details about the duration or nature of
the response/remission for this line."
              },
              "toxicities": {
                "type": "array",
                "items": {"type": "string"},
                "description": "List of notable toxicities experienced
(e.g., 'Grade 3 Neutropenia', 'Peripheral Neuropathy G2', 'Febrile
Neutropenia', 'CRS Grade 1')."
              }
            },
            "required": [
              "line_number", "regimen", "rationale", "duration",
              "best_response", "remission_details", "toxicities"
            ]
          }
        }
      },
      "required": ["number_of_prior_lines", "lines"]
    },
    "relapsed_refractory_scenario": {
      "type": "object",
      "description": "Describes the patient's status at the time of the most
recent relapse or progression, requiring a subsequent line of therapy.",
      "properties": {
        "detection_method": {
          "type": "string",
          "description": "How the relapse/progression was detected (e.g.,
'Routine surveillance PET-CT', 'Symptom recurrence investigation', 'Imaging
for unrelated issue')."
        },
        "date_of_relapse_detection": {
          "type": "string",
          "format": "date",
          "description": "Approximate date (YYYY-MM-DD or YYYY-MM) when
relapse/progression confirmed."
        },
        "months_since_initial_diagnosis": {
            "type": "integer",
            "minimum": 0,
```

```
            "description": "Time elapsed in months from initial diagnosis to
this relapse/progression event."
        },
        "current_age": {
          "type": "integer",
          "minimum": 18,
          "maximum": 120,
          "description": "Patient's age at the time of this
relapse/progression."
        },
        "current_ecog_status": {
          "type": "integer",
          "minimum": 0,
          "maximum": 5,
          "description": "ECOG status at relapse/progression."
        },
        "disease_status": {
          "type": "object",
          "properties": {
            "relapse_type": {
                "type": "string",
                "enum": ["Relapse after CR/PR", "Primary Refractory",
"Progression during treatment"],
                "description": "Nature of the event leading to this
scenario."
            },
            "sites_of_relapse": {
              "type": "object",
               "properties": {
                    "total_sites": {"type": "integer", "minimum": 0},
                    "sites": {"type": "array", "items": {"type": "string"}},
                    "is_nodal_only": {"type": "boolean"},
                    "is_extranodal_only": {"type": "boolean"},
                    "is_nodal_and_extranodal": {"type": "boolean"}
               },
               "required": ["total_sites", "sites"],
               "description": "Details of disease sites at
relapse/progression."
            },
            "ann_arbor_stage": {
              "type": "string",
              "enum": ["I", "II", "III", "IV"],
              "description": "Stage at relapse/progression."
            },
            "ldh_level_at_relapse": {
              "type": "object",
              "properties": {
                  "value": {"type": ["number", "null"]},
                  "unit": {"type": "string", "enum": ["U/L", "ukat/L",
"xULN", "Not Specified"]},
```

```
                "multiplier_uln": {"type": ["number", "null"]},
                "is_above_uln": {"type": "boolean"}
              },
              "required": ["is_above_uln"],
              "description": "LDH level at relapse/progression."
            },
            "repeat_biopsy_status": {
                "type": "object",
                "properties": {
                    "performed": {"type": "boolean"},
                    "site": {"type": ["string", "null"]},
                    "result": {"type": ["string", "null"], "description":
"e.g., 'Confirmed DLBCL, GCB type', 'Transformation to HGBL', 'No viable
lymphoma found'"}
                },
                "required": ["performed"],
                "description": "Status and findings of biopsy performed at
relapse/progression."
            }
          },
          "required": ["relapse_type", "sites_of_relapse",
"ann_arbor_stage", "ldh_level_at_relapse", "repeat_biopsy_status"]
        },
                  "ipi_components_at_relapse": {
          "type": "object",
          "properties": {
            "age_over_60": {"type": "boolean"},
            "ecog_over_1": {"type": "boolean"},
            "stage_3_or_4": {"type": "boolean"},
            "ldh_above_uln": {"type": "boolean"},
            "extranodal_sites_over_1": {"type": "boolean"}
          },
          "required": ["age_over_60", "ecog_over_1", "stage_3_or_4",
"ldh_above_uln", "extranodal_sites_over_1"],
          "description": "Components for the revised IPI score. Must sum
correctly."
        },
        "revised_ipi_score": {
          "type": "integer",
          "minimum": 0,
          "maximum": 5,
          "description": "IPI score recalculated at the time of
relapse/progression."
        },
        "comorbidities": {
          "type": "array",
          "items": {"type": "string"},
          "description": "List of active, clinically significant
comorbidities at the time of relapse/progression."
        },
```

```
        "concomitant_medications": {
          "type": "array",
          "items": {"type": "string"},
          "description": "List of relevant concomitant medications
(excluding supportive care for chemo)."
        },
        "timeline_consistency_check": {
          "type": "object",
          "description": "Internal logic fields to verify timeline
consistency.",
          "properties": {
            "expected_current_age": {
              "type": "number",
              "description": "Expected age calculated as:
age_at_initial_diagnosis + (months_since_initial_diagnosis / 12)."
            },
            "is_age_consistent": {
              "type": "boolean",
              "description": "True if calculated 'expected_current_age' is
within +/- 1 year of the stated 'current_age'."
            }
          },
          "required": ["expected_current_age", "is_age_consistent"]
        }
      },
      "required": [
        "detection_method", "date_of_relapse_detection",
"months_since_initial_diagnosis",
        "current_age", "current_ecog_status", "disease_status",
        "revised_ipi_score", "ipi_components_at_relapse", "comorbidities",
        "concomitant_medications", "timeline_consistency_check"
      ]
  },
  "required": [
    "patient_demographics", "initial_diagnosis", "treatment_history",
  ],
  "allOf": [
    {
      "if": {
        "properties": {
          "initial_diagnosis": {"properties": {"histopathology":
{"properties": {
            "double_triple_hit_status": { "enum": ["Double Hit", "Triple
Hit"] }
          }}}}
        }
      },
      "then": {
        "properties": {
```

```
          "initial_diagnosis": {"properties": {"histopathology":
{"properties": {
            "diagnosis": { "const": "HGBL", "description": "Diagnosis must
be HGBL if Double Hit or Triple Hit status is present." }
          }}}}
        }
      }
    },
    {
      "if": {
        "properties": { "treatment_history": { "properties": { "lines": {
"minItems": 1 }}}}
      },
      "then": {
          "properties": { "treatment_history": { "properties": {
              "number_of_prior_lines": { "description": "Must match the
actual count of items in the 'lines' array." }
          }}}
      }
    }
  ]
}
```

EXTRANODAL_SITES.JSON


```
{
  "description": "Structured knowledge about extranodal manifestations of Diffuse Large B-Cell Lymphoma (DLBCL)",
  "schema_version": "1.0-inferred",
  "sites": {
    "Gastrointestinal (GI) Tract": {
      "Frequency": "Most frequent site of primary extranodal DLBCL (approx. 28-34% of EN-DLBCL). Stomach is most common, followed by ileocecal region, small intestine, colon.",
      "Lugano Classification Context": "Can be Stage IE/IIE (often gastric) or advanced. 60-79% primary GI cases are Stage I-IIE. Need to differentiate from Burkitt, Mantle Cell. Common site for double-hit dissemination.",
      "Sanctuary Site Status": "Not classified as a sanctuary site.",
      "CNS Relapse Risk": "Generally low for localized gastric DLBCL in rituximab era (0% in some series). However, GI involvement contributes to higher NCCN-IPI scores, and patients often have other risk factors elevating overall CNS risk.",
      "Clinical Considerations": {
        "Presentation": "Abdominal pain, dyspepsia, nausea, vomiting, anorexia, weight loss, GI bleeding, obstruction. B symptoms reported inconsistently (absent to 57%).",
        "Prognosis": "Heterogeneous. Localized gastric has excellent outcomes with chemoimmunotherapy. Overall GI involvement associated with worse OS vs nodal in some SEER analyses (HR 1.24), but better OS vs other extranodal sites (nervous system, respiratory) in another. Stage I extranodal (all sites) worse OS vs Stage I nodal.",
        "Molecular Features": "Non-GCB subtype prevalence varies (33-50%). MYD88/CD79B mutations infrequent (<5-6%). EZB genotype prevalence low.",
      }
    },
    "Central Nervous System (CNS)": {
      "Frequency": "Primary CNS Lymphoma (PCNSL) is rare (2% primary CNS tumors, 4-6% EN lymphomas). Secondary CNS involvement occurs in 2-5% of all DLBCL, but >10-40% in high-risk subgroups.",
      "Lugano Classification Context": "PCNSL confined to CNS structures (brain, cerebellum, spinal cord, leptomeninges, eyes). Any systemic disease excludes PCNSL. Staging involves CNS MRI, eye exam, CSF analysis, PET-CT to rule out systemic disease. Secondary CNS involvement signifies Stage IV.",
      "Sanctuary Site Status": "Unequivocally an immune-privileged site (Blood-Brain Barrier/Blood-Retinal Barrier). Restricts drug penetration and immune trafficking.",
      "CNS Relapse Risk": "Applies to systemic DLBCL, not PCNSL. Risk stratified by CNS-IPI and high-risk sites. Relapse often early (median 5-9 months).",
      "Clinical Considerations": {
        "Presentation (PCNSL)": "Neurological symptoms: focal deficits, cognitive/personality changes, increased intracranial pressure, seizures.
```

```
Multifocal lesions common (40-50%). B symptoms rare. Ocular involvement
(PVRL) possible.",
        "Prognosis": "PCNSL prognosis poorer than systemic DLBCL but
improving with specialized regimens. Secondary CNS relapse has dismal
prognosis (median OS 2-6 months).",
        "Molecular Features (PCNSL)": "Predominantly non-GCB/ABC subtype.
High frequency of MYD88 L265P (up to 80%) and CD79B mutations, activating
NF-κB pathway. Frequent PD-1 pathway alterations. Shares features with
testis/breast DLBCL.",
        "Management (PCNSL)": "Requires BBB-penetrating agents (R-CHOP
ineffective). Cornerstone: High-dose Methotrexate (HD-MTX) based regimens
(e.g., MATRix, R-MPV). Consolidation crucial: HDC-ASCT preferred for fit
responders; WBRT for ineligible (neurotoxicity concern). Limited role for IT
chemo (maybe for leptomeningeal). Ocular disease: intravitreal chemo or
orbital RT.",
        "CNS Prophylaxis (Systemic DLBCL)": "Controversial. Recommended for
high CNS-IPI or high-risk sites, but efficacy evidence weak/conflicting.
Optimal method (HD-MTX vs IT) and timing debated. Need for better risk
stratification."
      }
    },
    "Testis": {
      "Frequency": "Rare (1-2% NHL, ~4% EN NHL). Most common testicular
malignancy in men >60. 9% of Stage I EN-DLBCL.",
      "Lugano Classification Context": "Most commonly presents as localized
disease (Stage IE/IIE in ~77% cases). Universally considered high-risk
despite localized presentation.",
      "Sanctuary Site Status": "Explicitly recognized immune-privileged site
(Blood-Testis Barrier). Limits therapy penetration and immune
surveillance.",
      "CNS Relapse Risk": "Inherently high risk (19-34%), largely
independent of stage. Mandates CNS prophylaxis. CNS-IPI may underestimate
risk as PTL often presents early stage. Requires site-specific risk
assessment.",
      "Clinical Considerations": {
        "Presentation": "Typically painless, unilateral testicular
mass/swelling. Bilateral involvement possible.",
        "Prognosis": "Historically poor, improved with Rituximab, but high
CNS/contralateral relapse rates persist. Remains aggressive.",
        "Molecular Features": "Remarkably similar to PCNSL: Predominantly
ABC/non-GCB. High frequency MYD88 L265P (70-80%), CD79B mutations. Frequent
CDKN2A alterations (~88%), PD-1 pathway alterations (~50%). Likely high
prevalence MCD genomic subtype.",
        "Management": "Multimodal approach: 1) Systemic chemoimmunotherapy
(e.g., R-CHOP). 2) Mandatory CNS prophylaxis (systemic HD-MTX preferred). 3)
Prophylactic radiotherapy to contralateral testis (25-30 Gy).
Diagnostic/therapeutic orchiectomy common."
      }
    },
    "Breast": {
```

```
      "Frequency": "Rare (~0.5% breast neoplasms, 1.7-2.2% EN lymphomas). Incidence potentially increasing. 2.5% EN-DLBCL in SEER. 5-13.5% in some clinical series.",
      "Lugano Classification Context": "Typically localized disease (>90% Stage IE/IIE). Definition requires breast infiltration +/- ipsilateral axillary nodes, no widespread disease/prior lymphoma.",
      "Sanctuary Site Status": "Not formally classified, but high CNS relapse risk suggests potential unique microenvironment or trafficking. Speculated as variant of immune-privileged site LBCL.",
      "CNS Relapse Risk": "Significantly elevated risk (7% to 24%). Indication for considering CNS prophylaxis.",
      "Clinical Considerations": {
        "Presentation": "Painless, palpable breast mass mimicking carcinoma most common. Less often skin edema, erythema, retraction. Very rare in males.",
        "Prognosis": "Generally favorable for localized disease with modern chemoimmunotherapy (5-yr OS ~75-80%). Possibly comparable to non-breast DLBCL, better than GI.",
        "Molecular Features": "Predominantly non-GCB/ABC subtype (60-90%). High prevalence MYD88 (58-70%) and CD79B (36-40%) mutations, similar to PCNSL/PTL. Dual expresser status may be relevant.",
        "Management": "CNS prophylaxis frequently considered due to high risk, but controversial."
      }
    },
    "Bone (Primary Bone DLBCL - PBDL)": {
      "Frequency": "Uncommon primary site. 5-21% in extranodal series. Musculoskeletal 5.8% EN-DLBCL in SEER. Distinguish from common secondary bone marrow infiltration (poor prognosis) or cortical bone spread.",
      "Lugano Classification Context": "Unifocal (single lesion, often IE) or multifocal (multiple bones, considered Stage IV equivalent). Localized (Stage I/II) frequent (56-87%).",
      "Sanctuary Site Status": "Not typically considered a sanctuary site.",
      "CNS Relapse Risk": "Depends on extent. Low for unifocal PBDL (~2.5%), prophylaxis usually not recommended. Higher risk with multifocal or craniospinal bone involvement, may warrant prophylaxis consideration.",
      "Clinical Considerations": {
        "Presentation": "Localized bone pain, swelling, pathological fractures. Unifocal vs multifocal distinction is prognostic.",
        "Prognosis": "Generally favorable for unifocal/localized (Stage I/II), 5-yr OS ~86%. Musculoskeletal OS better vs GI in SEER (HR 0.59). Similar to nodal DLBCL. Multifocal PBDL has significantly poorer prognosis (5-yr PFS 14%, OS 47% in one series).",
        "Molecular Features": "Often GCB/centrocyte-like profile. MYD88 mutations typically absent.",
      }
    },
    "Kidney and Adrenal Gland": {
```

```
      "Frequency": "Primary involvement rare. Primary adrenal exceptionally rare. Kidney involvement ~2% at diagnosis. Bilateral common, especially adrenal.",
      "Lugano Classification Context": "Almost invariably signifies advanced disease (Stage IV in >90% kidney cases).",
      "Sanctuary Site Status": "Not explicitly designated, but high CNS risk association might suggest unique biology/barriers.",
      "CNS Relapse Risk": "Very high risk. Kidney/adrenal involvement is an independent risk factor in CNS-IPI. Reported CNS relapse: 36-41% for kidney, 20% for adrenal. Strong indication for CNS prophylaxis.",
      "Clinical Considerations": {
        "Presentation": "Adrenal: symptoms of adrenal insufficiency. Kidney: often with extensive nodal/extranodal disease. Association between adrenal DLBCL and intravascular lymphoma.",
        "Prognosis": "Very poor prognosis. 5-yr OS ~29% for kidney involvement (43% with Rituximab). Median OS 14 months for primary adrenal (one study).",
        "Molecular Features": "Adrenal: Potential association with IVL features (MYD88/CD79B mutations?).",
        "Management": "CNS prophylaxis strongly recommended (HD-MTX often preferred)."
      }
    },
    "Head & Neck Sites": {
      "Frequency": "Common area (second to GI). 11.8-14% EN-DLBCL in SEER. Subsites: Craniofacial/Sinonasal (Orbit, Sinuses, Oral Cavity - 8% Stage I Sinus/Nose). Thyroid (rare, <2-5% thyroid malignancies, ~2% EN lymphomas, assoc. Hashimoto's). Salivary Gland/Parotid (infrequent, 2-5%). Oral Cavity (part of craniofacial). Waldeyer's Ring (Tonsil, Nasopharynx - technically nodal but discussed here, tonsil very common).",
      "Lugano Classification Context": "Many (thyroid, sinonasal) often present early stage (IE/IIE).",
      "Sanctuary Site Status": "Orbit/paranasal sinuses not typically classified as sanctuary sites requiring mandatory prophylaxis, but proximity to CNS is a concern.",
      "CNS Relapse Risk": {
        "Craniofacial/Sinonasal/Orbital": Historically high-risk, but risk significantly lower in rituximab era (~1.6%). Routine prophylaxis generally no longer recommended. Individualized approach considering CNS-IPI and high-risk features (epidural extension, orbital apex, skull base erosion).",
        "Thyroid": Generally low risk (0% CNS recurrence in one series)."
      },
      "Clinical Considerations": {
        "Presentation": "Site-specific: nasal obstruction/epistaxis (sinonasal); rapidly enlarging neck mass (thyroid); proptosis (orbital).",
        "Prognosis": "Generally excellent for localized craniofacial, sinonasal, thyroid in rituximab era. Head & Neck OS better vs GI in SEER (HR 0.77). Thyroid 5-yr OS 83-89%. Sinonasal maybe more variable (5-yr OS 56% in one study).",
```

```
        "Molecular Features": "Craniofacial/sinonasal often non-GCB (~80%). Thyroid often GCB, lacks MYD88 mutations (favorable).",
        "Management": "CNS prophylaxis individualized for craniofacial/sinonasal. Thyroid typically managed with chemoimmunotherapy alone."
      }
    },
    "Skin (Primary Cutaneous DLBCL, Leg Type - PCDLBCL-LT)": {
      "Frequency": "Rare subtype of uncommon Primary Cutaneous B-cell Lymphomas (PCBCLs). PCDLBCL-LT is 10-20% of PCBCLs, ~4% of all Primary Cutaneous Lymphomas. SEER Skin/Soft Tissue category 9.5% EN-DLBCL.",
      "Lugano Classification Context": "Defined as arising primarily in skin, no extracutaneous disease at diagnosis. Often staged using TNMB system.",
      "Sanctuary Site Status": "Not traditional sanctuary site, but sometimes discussed alongside immune-privileged sites due to molecular features/dissemination potential.",
      "CNS Relapse Risk": "Recognized risk of extracutaneous dissemination including CNS. Prompts consideration of CNS prophylaxis.",
      "Clinical Considerations": {
        "Presentation": "Rapidly growing red/violaceous nodules/tumors, typically on lower leg(s). Predominantly affects elderly (>70 years).",
        "Prognosis": "Aggressive subtype, poorer prognosis vs indolent PCBCLs. High relapse rate (cutaneous & extracutaneous). Broader SEER Skin/Soft Tissue category had relatively better OS vs GI (HR 0.83), possibly due to mix with indolent types.",
        "Molecular Features": "Typically non-GCB/ABC immunophenotype. May harbor MYD88 mutations, molecular similarities to systemic ABC-DLBCL.",
        "Management": "Requires systemic therapy (R-CHOP) often combined with ISRT. Challenging due to patient age/comorbidities. CNS prophylaxis often considered."
      }
    },
    "Lung / Pleura / Respiratory Tract": {
      "Frequency": "Primary lung DLBCL very rare (~0.4% lymphomas). Most primary pulmonary lymphomas are low-grade MALT. SEER Respiratory Tract 5.17% EN-DLBCL. Pleural involvement common with disseminated disease.",
      "Lugano Classification Context": "Primary pulmonary could be Stage IE. Pleural effusion/involvement often indicative of Stage IV.",
      "Sanctuary Site Status": "Not considered sanctuary sites.",
      "CNS Relapse Risk": "Lung parenchyma involvement is unfavorable factor in NCCN-IPI. Pleural involvement suggests disseminated disease, higher risk profile.",
      "Clinical Considerations": {
        "Presentation": "Cough, dyspnea, chest pain, fever. Radiology: nodules, masses, consolidation. Airway involvement causes compression.",
        "Prognosis": "Limited data for primary pulmonary DLBCL. SEER Respiratory Tract involvement associated with worse prognosis vs GI (HR 1.18). Pleural involvement linked to poor prognosis.",
```

```
        "Molecular Features": "Primary pulmonary most commonly MALT or
DLBCL. Sparse data for primary pulmonary DLBCL.",
        "Management": "Essential to differentiate primary vs secondary
involvement via staging (PET-CT)."
      }
    },
    "Liver / Pancreas / Hepatobiliary System": {
      "Frequency": "Primary origin uncommon. Liver most often secondary
involvement. SEER Pancreas/Hepatobiliary 4.16% EN-DLBCL. 1.5-2% in some
extranodal series.",
      "Lugano Classification Context": "Typically signifies advanced disease
(Stage IV).",
      "Sanctuary Site Status": "Not considered sanctuary sites.",
      "CNS Relapse Risk": "Liver involvement unfavorable factor in NCCN-IPI.
Involvement often occurs with advanced stage/high IPI, contributing to
overall increased CNS risk.",
      "Clinical Considerations": {
        "Presentation": "Jaundice, abdominal pain, constitutional
symptoms.",
        "Prognosis": "Consistently associated with worse prognosis. SEER:
Higher mortality risk vs GI (HR 1.22 Pancreas/Hep-Bil) or nodal (HR 1.58
Liver/Pancreas). Possible better prognosis for true primary hepatic DLBCL
(assoc. Hep C?) vs secondary, but data limited. Strong adverse prognostic
marker.",
        "Molecular Features": "Limited data in provided sources.",
      }
    },
    "Female Genital System (FGS - Uterus, Cervix, Vagina, Ovary)": {
      "Frequency": "Rare primary site. 2.6-8.5% in extranodal series.
Internal reproductive organs ~4% female DLBCL patients (uterus most common,
then ovary).",
      "Lugano Classification Context": "Can be localized (IE/IIE) or
advanced. 55-63% limited stage in one series (Ann Arbor/FIGO).",
      "Sanctuary Site Status": "Not typically considered sanctuary sites.",
      "CNS Relapse Risk": "Conflicting/site-specific evidence. One study
reported very high risk (41%) specifically with uterine involvement (strong
predictor HR 14.13), prompting prophylaxis consideration. Same study found
no increased risk with ovarian involvement alone. Another study reported
most CNS relapses occurred with primary ovarian involvement. Highlights
potential high risk, especially uterine.",
      "Clinical Considerations": {
        "Presentation": "Abnormal vaginal bleeding, pelvic pain, pelvic
mass.",
        "Prognosis": "May depend on site/stage. Uterine involvement
associated with inferior PFS/OS in one study; ovarian alone did not impact
survival. Another study reported 5-yr OS 77% for FGS DLBCL cohort, advanced
stage primary predictor of worse outcome. Broader SEER Genitourinary
category better OS vs GI (HR 0.84).",
        "Molecular Features": "Limited data in provided sources.",
      }
```

```
    },
    "Other Sites": {
      "Heart/Mediastinum/Pericardium": "Primary rare. SEER Heart/Mediastinum
3.28% EN-DLBCL. Associated with best OS vs GI in SEER (HR 0.58), potentially
influenced by inclusion of PMBCL (distinct entity, good outcome).
Pericardial involvement potential CNS risk factor in one study.",
      "Endocrine (excluding Thyroid/Adrenal)": "Very rare (e.g., pituitary).
SEER Endocrine category 3.66% EN-DLBCL, better OS vs GI (HR 0.73).",
      "Bladder/Prostate": "Rare primary sites. Included in SEER
Genitourinary category (7.24% total, better OS vs GI HR 0.84)."
    }
  },
  "high_cns_risk_summary": {
    "consistently_high": ["Testis", "Kidney/Adrenal", "Uterus (likely)"],
    "significant_risk_prophylaxis_consideration": ["Breast", "PCDLBCL-LT",
"Craniofacial/Sinonasal/Orbital (individualized)", "Epidural Space"],
    "other_contributing_factors": [">1 Extranodal Site (esp. + high LDH)",
"Bone Marrow Involvement", "Double/Triple-Hit or Dual Expresser status",
"Intravascular Lymphoma"]
  }
}
```


The **guidelines.JSON** and **treatments.JSON** files are found in **Appendix B**

APPENDIX "B"

Custom treatment decision prompt"Ariadne's Thread" and accompanying JSON knowledge base

"ARIADNE'S THREAD"

```
# MISSION
Act as Ariadhne (🤖), an expert coordinator of medical panel
discussions on relapsed/refractory DLBCL.

🤖 has the ability to reason on a Patient Case Profile (PCP) by running
your thought process as *code interpretation* by using your **python
tool** to prepend EVERY output in a code block with:
```py
PCP = {
      "📝": [Summary of relevant clinical history, update with 🧠🌿]
      "🧑‍⚕️": [Current health status, focus on relevant comorbidities that
could influence treatment choice]
     "💉": [Prior treatment details]
      "🔧": [adjustment to fine-tune relevant clinical case data]
      "🤔": [identify missing important information based on 🔧]
      "🔍": [ask user missing information based on 🤔]
      "🧠🌿": [updated knowledge based on context]
     "🔜": [state WORKFLOW's next step: preliminary research; expert
panel summoning; debate and consensus]
}
```

OBJECTIVE: 🤖 works step by step to determine the optimal treatment
for R/R DLBCL.

RULES:
- Do your best to fill in the [blanks] based on the context
- The script ensures a consistent and informed progression through
each step, minimizing the need for user input until the crucial moment
of feedback
- 🤖 MUST FOLLOW the WORKFLOW below rigorously

WORKFLOW:
1. CASE COLLECTION
- Input: the user's clinical case
- Action:
a. use PCP to summarize the clinical case and ask relevant
questions. Treat the output as *code interpretation* using **python
tool**.
```

b. state 🔍.
- Output: update 🧠🌿 with the provided User's answers using *code interpreter*, then ALWAYS state 🔜 and ask User whether to proceed.

2. PRELIMINARY RESEARCH
- Before starting: use PCP. 🔄
- Tools: *Web Browsing*, *Code Interpreter*, *Knowledge Retrieval*.
- Action:
a. search for recent DLBCL guidelines, research, and abstracts
b. use *Knowledge Retrieval* and *code interpreter* to interpret "treatments.json", "guidelines.json" and "examples.json"
c. conduct an in-depth analysis and output your findings in a brief report (📄)
- Output: update PCP with 📄 using *code interpreter*, then ALWAYS state 🔜 and ask User wheter to proceed.

3. EXPERT PANEL SUMMONING
- Before starting: use PCP 🔄
- Panel Composition: Hematologist 🩸🔬, Lymphoma Specialist 🩺🦠, Hematopoietic Stem Cell Transplant Specialist 🧬🩸, Research Scientist 🔍🧬, Patient Advocate ❤️📢, Data Analyst 📊💻.
- Action: 🩸🔬, 🩺🦠, 🧬🩸, 🔍🧬, ❤️📢, 📊💻 each provide a brief self description structured as: {expertise in [domain], specializing in [subdomain] using [context]}
- Output: the brief descriptions, then ALWAYS state 🔜 and ask User wheter to proceed.

4. DEBATE AND CONSENSUS
- Before starting: use PCP. 🔄
- Action: 🩸🔬, 🩺🦠, 🧬🩸, 🔍🧬, ❤️📢 and 📊💻 engage in an informed debate using *code interpreter* and *knowledge retrieval* based on PCP. Their GOAL is to state the BEST treatment (🩺🎙️) for the case in discussion.
- Analysis: choose the single BEST treatment (🩺🎙️) based on the panel discussion.
- Output: state 🩺🎙️. ALWAYS provide the rationale for the choice.
- User Engagement: Now, 🤖 invites user feedback on the recommendation.

TREATMENTS.JSON


```
{
  "treatments": [
    {
      "name": "R-IEV (Rituximab, Ifosfamide, Epirubicin, Etoposide)",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Salvage treatment", "Fit", "Transplant eligible", "Stem cell
mobilization", "CNS localization", "Bridge to CAR-T"],
      "criteria": {
        "inclusion": ["Relapsed or refractory DLBCL", "Adequate organ
function", "Adequate clinical conditions / Performance Status (typically
ECOG 0-1)", "Candidate for potential ASCT consolidation"],
        "exclusion": ["Previous severe reaction to any of the regimen
components", "Severe organ impairment incompatible with high-dose therapy",
"Poor performance status (e.g., ECOG >=2)", "Chemoresistant disease"]
      },
       "guideline_context": "Less commonly listed in major international
guidelines (e.g., NCCN) for R/R DLBCL salvage compared to platinum-based
regimens (R-ICE, R-DHAP, R-GDP), despite demonstrated efficacy and
mobilization potential in cited studies.",
      "keyPapers": [
        {
          "citation1": "Menzel H, Müller A, Von Schilling C, et al.
Ifosfamide, epirubicin and etoposide rituximab in refractory or relapsed
B-cell lymphoma: Analysis of remission induction and stem cell mobilization.
Leuk Lymphoma. 2008;49(7):1337-44. doi: 10.1080/10428190802094229",
          "reportedOutcomes": "ORR 68.8% (vs 40.7% for IEV alone). Excellent
stem cell mobilization (93.8%). Post-ASCT survival rates at 12/24 months
were 69.2%/61.5% for those transplanted.",
          "numberOfPatientsStudied": "16 R-IEV, 69 IEV total",
          "patientAndDiseaseCharacteristics": "Refractory/Relapsed B-cell
NHL. R-IEV cohort median age 53.",
          "toxicities": "Febrile neutropenia ~19.4% (R-IEV). Need for
transfusions ~44%."
        },
        {
          "citation2": "Bishton MJ, Lush RJ, Byrne JL, et al. Ifosphamide,
etoposide and epirubicin is an effective combined salvage and peripheral
blood stem cell mobilisation regimen for transplant-eligible patients with
non-Hodgkin lymphoma and Hodgkin disease. Br J Haematol. 2007;136(5):752-61.
doi:10.1111/j.1365-2141.2007.06498.x",
          "warning": "Rituximab was not used in this study (IEV only);
included mixed NHL/HD histologies (~31% DLBCL).",
          "reportedOutcomes": "ORR ~80% overall (NHL ~78%). Excellent
mobilization (~99%). 5-year OS ~53% (NHL ~50%), 5-year EFS ~43% (NHL
~39%).",
          "numberOfPatientsStudied": "143 (NHL + HD)",
          "patientAndDiseaseCharacteristics": "Relapsed/Refractory NHL/HD.
Median age 49. ~32% primary refractory.",
```

```
          "toxicities": "High rates Grade 4 neutropenia (~80%), febrile
neutropenia requiring IV Abx (~29%), Grade 4 thrombocytopenia (~31%), RBC
transfusions (~77%), Ifosfamide encephalopathy (7%)."
        }
      ],
      "specialConsiderations": [
          "Effective salvage regimen with high stem cell mobilization
yield.",
          "May be considered for CNS involvement due to component drugs.",
          "Requires G-CSF support and Mesna prophylaxis.",
          "Monitor for hematologic toxicity, febrile neutropenia, renal
function, and neurotoxicity (ifosfamide encephalopathy).",
          "Intended for patients eligible for and proceeding to ASCT if
chemosensitive.",
          "Less commonly preferred in major international guidelines
compared to platinum-based salvage regimens for DLBCL."
          ],
      "routeOfAdministration": "Intravenous",
      "settingOfAdministration": "Typically requires inpatient setting or
very close outpatient monitoring due to hydration/ifosfamide requirements.",
      "lengthOfTreatment": "Typically 2-4 cycles (repeated every 3-4 weeks)
followed by response assessment and ASCT if eligible."
    },
    {
      "name": "R-DHAP (Rituximab, Dexamethasone, High-Dose Cytarabine,
Cisplatin)",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Salvage treatment", "Fit", "Transplant eligible", "Stem cell
mobilization", "Platinum-based", "Chemosensitive disease"],
      "criteria": {
        "inclusion": ["Relapsed or refractory DLBCL", "Candidate for
potential ASCT consolidation", "Adequate organ function (esp. renal)",
"Adequate clinical conditions / Performance Status (typically ECOG 0-1)"],
        "exclusion": ["Previous severe reaction to any regimen components",
"Significant pre-existing renal impairment (consider R-DHAOx)", "Significant
hearing impairment", "Poor performance status (e.g., ECOG >=2)",
"Chemoresistant disease"]
      },
      "guideline_context": "Standard platinum-based salvage regimen listed
in major guidelines (NCCN/ESMO). Often considered interchangeable with R-ICE
regarding efficacy based on CORAL trial, but carries higher risk of renal
toxicity and thrombocytopenia. R-DHAOx is preferred if cisplatin toxicity
(renal, oto) is a concern. Role in 2L early relapse/primary refractory often
superseded by CAR-T therapy where available/eligible.",
      "keyPapers": [
        {
          "citation1": "Gisselbrecht C, Glass B, Mounier N, et al. Salvage
regimens with autologous transplantation for relapsed large B-cell lymphoma
in the rituximab era. J Clin Oncol. 2010;28(27):4184-4190.
doi:10.1200/JCO.2010.28.1618 (CORAL Trial)",
```

```
          "reportedOutcomes": "Compared R-DHAP vs R-ICE pre-ASCT. ORR ~62.8%
(CR ~40%) for R-DHAP arm, no significant difference vs R-ICE. 3-year EFS
~35%. Poorer outcomes if prior Rituximab exposure and relapse <12 months.
Stem cell mobilization successful in ~80%.",
          "numberOfPatientsStudied": "194 (R-DHAP arm)",
          "patientAndDiseaseCharacteristics": "R/R DLBCL,
transplant-eligible. Median age 55.",
          "toxicities": "Higher rates of Grade 3/4 thrombocytopenia
requiring platelet transfusion (~57%) and Grade 3/4 renal toxicity (~5-6%)
compared to R-ICE in CORAL trial. Treatment-related death rate ~5%. Febrile
neutropenia ~16%."
        }
      ],
      "specialConsiderations": [
          "Standard salvage chemotherapy option prior to ASCT consolidation
for chemosensitive disease.",
          "Requires intensive hydration and monitoring due to high-dose
cisplatin and cytarabine.",
          "Higher risk of nephrotoxicity and thrombocytopenia compared to
R-ICE or R-DHAOx; consider alternatives if renal function is borderline or
pre-existing neuropathy/ototoxicity is present.",
          "Requires G-CSF support.",
          "Intended for patients eligible for and proceeding to ASCT if
chemosensitive."
          ],
      "routeOfAdministration": "Intravenous",
      "settingOfAdministration": "Typically requires inpatient setting due
to continuous cisplatin infusion, high-dose cytarabine, and intensive
hydration protocols.",
      "lengthOfTreatment": "Typically 2-3 cycles (repeated every ~3 weeks)
followed by response assessment and ASCT if eligible."
    },
    {
      "name": "R-DHAOx (Rituximab, Dexamethasone, High-Dose Cytarabine,
Oxaliplatin)",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Salvage treatment", "Stem cell mobilization", "Bridge to
CAR-T", "Platinum-based", "Chemosensitive disease"],
      "criteria": {
        "inclusion": ["Relapsed or refractory DLBCL", "Candidate for
potential ASCT consolidation", "Adequate organ function", "Adequate clinical
conditions / Performance Status (typically ECOG 0-1)"],
        "exclusion": ["Previous severe reaction to any regimen components",
"Severe organ impairment incompatible with high-dose therapy", "Poor
performance status (e.g., ECOG >=2)", "Significant pre-existing neuropathy",
"Chemoresistant disease"]
      },
      "guideline_context": "Recognized platinum-based salvage option, often
used as an alternative to R-DHAP to avoid cisplatin-related
nephrotoxicity/ototoxicity. Standard salvage option prior to ASCT for
```

```
chemosensitive disease, but role in 2L early relapse/primary refractory
often superseded by CAR-T therapy where available/eligible.",
      "keyPapers": [
        {
          "citation1": "Manconi L, Coviello E, Canale F, et al.
Dexamethasone, oxaliplatin and cytarabine (R-DHAOx) as salvage and stem
cells mobilizing therapy in relapsed/refractory diffuse large B cell
lymphomas. Leuk Lymphoma. 2020;61(1):84-90.
doi:10.1080/10428194.2019.1658102",
          "reportedOutcomes": "ORR 60% (CR 49%, PR 11%) after 2 cycles.
Higher ORR in relapsed (79%) vs refractory (45%). ORR correlated with lower
stage (80% vs 45%) and lower IPI. Median OS 30.5 months, 3-year OS 40.5%.
3-year PFS 32.6% overall (59.6% in responders). Excellent stem cell
mobilization (21/22 eligible patients transplanted).",
          "numberOfPatientsStudied": "53",
          "patientAndDiseaseCharacteristics": "R/R DLBCL; Median age 59;
Median 2 prior lines; 55% refractory disease; 64% Stage III/IV; 74% IPI
>=3.",
          "toxicities": "Manageable hematologic toxicity (Grade 3/4
Neutropenia ~21% - inferred from FN rate). Febrile neutropenia 11%. Low
rates of severe non-hematologic toxicity; importantly, no renal impairment
or severe neuropathy/ototoxicity reported."
        },
        {
            "citation_summary": "Comparative Studies (e.g., Lacout C, Eur J
Haematol. 2020)",
            "reportedOutcomes_summary": "Retrospective comparisons confirm
R-DHAOx associated with significantly less nephrotoxicity compared to
cisplatin-based R-DHAP. Efficacy appears generally comparable, though
patient selection may differ between studies.",
            "pubmed_id": "32302426"
        }
      ],
      "specialConsiderations": [
          "Effective salvage regimen with excellent stem cell mobilization
capability.",
          "Key advantage over R-DHAP is significantly lower risk of renal
toxicity and likely lower ototoxicity/neuropathy due to substitution of
oxaliplatin for cisplatin.",
          "Preferred platinum-based salvage option for patients with
pre-existing renal impairment or concerns about cisplatin toxicity.",
          "Requires G-CSF support.",
          "Monitor for hematologic toxicity, febrile neutropenia,
oxaliplatin-related neuropathy.",
          "Intended for patients eligible for and proceeding to ASCT if
chemosensitive."
          ],
      "routeOfAdministration": "Intravenous",
```

```
      "settingOfAdministration": "Hospital ward or outpatient, depending on
specific schedule/monitoring needs (e.g., high-dose cytarabine
administration).",
      "lengthOfTreatment": "Typically 2-3 cycles (repeated every ~3 weeks)
followed by response assessment and ASCT if eligible."
    },
    {
      "name": "R-ICE (Rituximab, Ifosfamide, Carboplatin, Etoposide)",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Salvage treatment", "Fit", "Transplant eligible", "Stem cell
mobilization", "Chemosensitive disease"],
      "criteria": {
        "inclusion": [
          "Relapsed or refractory DLBCL",
          "Candidate for potential ASCT consolidation",
          "Adequate organ function (renal, hepatic, cardiac)",
          "Adequate clinical conditions / Performance Status (typically ECOG
0-1)",
          "Age typically <65-70 years (based on ASCT eligibility)"
          ],
        "exclusion": [
          "Previous severe reaction to any regimen components",
          "Severe organ impairment incompatible with high-dose therapy",
          "Poor performance status (e.g., ECOG >=2)",
          "Chemoresistant disease (not suitable for ASCT consolidation)"
        ]
      },
      "keyPapers": [
        {
          "citation1": "Kewalramani T, Zelenetz AD, Nimer SD, et al.
Rituximab and ICE as second-line therapy before autologous stem cell
transplantation for relapsed or primary refractory diffuse large B-cell
lymphoma. Blood. 2004;103(10):3684-3688. doi:10.1182/blood-2003-11-3911",
          "reportedOutcomes": "ORR 78%, CR 53% (significantly higher CR vs
historical ICE). Post-ASCT 2-year PFS ~54%.",
          "numberOfPatientsStudied": "36 (R-ICE arm)",
          "patientAndDiseaseCharacteristics": "Median age 45 years; 36%
primary refractory; transplant-eligible.",
          "toxicities": "Grade 3/4 neutropenia/thrombocytopenia common (~17%
each or both). Grade 3 non-hematologic included neutropenic fever (8
cycles), infections, DVT/PE, cardiac ischemia, hemorrhagic cystitis."
        },
        {
          "citation2": "Gisselbrecht C, Glass B, Mounier N, et al. Salvage
regimens with autologous transplantation for relapsed large B-cell lymphoma
in the rituximab era. J Clin Oncol. 2010;28(27):4184-4190.
doi:10.1200/JCO.2010.28.1618 (CORAL Trial)",
          "reportedOutcomes": "Compared R-ICE vs R-DHAP pre-ASCT. ORR ~63.5%
(CR ~36%) for R-ICE arm, no significant difference vs R-DHAP. 3-year EFS
```

~26-31%. Poorer outcomes if prior Rituximab exposure and relapse <12 months. Post-ASCT 3-yr PFS ~53% in responders overall.",
          "numberOfPatientsStudied": "202 (R-ICE arm)",
          "patientAndDiseaseCharacteristics": "R/R DLBCL, transplant-eligible. Median age 54.",
          "toxicities": "Generally comparable toxicity profile to R-DHAP, although specific rates differed (e.g., less thrombocytopenia/renal tox vs R-DHAP)."
        }
      ],
      "specialConsiderations": [
          "Standard salvage chemotherapy option prior to ASCT consolidation for chemosensitive disease.",
          "Role in 2L early relapse (<12mo) / primary refractory disease often superseded by CAR-T therapy where available/eligible.",
          "No significant efficacy difference compared to R-DHAP in CORAL trial.",
          "Requires G-CSF support.",
          "Requires careful monitoring for hematologic toxicity, infections, renal function (due to ifosfamide/carboplatin), and hemorrhagic cystitis (requires mesna prophylaxis).",
          "Intended for patients eligible for and proceeding to ASCT if chemosensitive."
          ],
      "routeOfAdministration": "Intravenous",
      "settingOfAdministration": "Typically requires inpatient setting or very close outpatient monitoring due to hydration/ifosfamide requirements.",
      "lengthOfTreatment": "Typically 2-3 cycles (repeated every ~3 weeks) followed by response assessment and ASCT if eligible."
    },
    {
      "name": "R-GEMOx (Rituximab, Gemcitabine, Oxaliplatin)",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Salvage treatment", "Transplant ineligible", "Less intensive", "Bridge to CAR-T", "Outpatient"],
      "criteria": {
        "inclusion": ["Relapsed or refractory DLBCL", "Often used for patients ineligible for ASCT due to age or comorbidities", "Adequate organ function", "Adequate clinical conditions / Performance Status"],
        "exclusion": ["Previous severe reaction to any regimen components", "Significant pre-existing peripheral neuropathy", "Severe thrombocytopenia (potentially)", "Chemoresistant disease"]
      },
      "guideline_context": "Recognized salvage option, particularly suitable for patients ineligible for ASCT due to age/comorbidities owing to a generally more tolerable profile than intensive platinum regimens (e.g., R-DHAP). Also used as bridging therapy pre-CAR T. Less effective than ASCT/CAR-T for curative intent.",
      "keyPapers": [
        {

```
          "citation1": "Crump M, et al. ASH 2024 Abstract (Multi-center RWD
study)",
          "study_context": "Real-World Data - R/R DLBCL",
          "reportedOutcomes": "ORR 45%, CR 29%. Median EFS 2.3 months (12m
EFS ~21%). Median OS 13.5 months (12m OS ~52%). Outcomes generally better
when used as bridge to cellular therapy vs palliative intent.",
          "numberOfPatientsStudied": "408",
          "patientAndDiseaseCharacteristics": "R/R DLBCL in routine
practice; included bridging and non-bridging intent; median 2 prior lines.",
          "toxicities": "Profile generally considered more tolerable than
R-DHAP/R-ICE, particularly regarding nephrotoxicity/ototoxicity. Key
toxicities include myelosuppression (thrombocytopenia, neutropenia) and
peripheral neuropathy (oxaliplatin)."
        },
        {
          "citation2": "Mounier N, et al. Haematologica. 2013;98(11):1726-31
(LYSA Phase II)",
          "study_context": "Phase II in ASCT-ineligible R/R DLBCL",
          "reportedOutcomes": "ORR 61% (after 4 cycles), CR 44% (end of
treatment). Low 5-year PFS (~13%) and OS (~14%) highlighting limited
long-term efficacy in this setting.",
          "numberOfPatientsStudied": "49",
          "patientAndDiseaseCharacteristics": "ASCT-ineligible R/R DLBCL.
Median age 76.",
          "toxicities": "Well tolerated overall; Grade 3-4 infections ~22%
of cycles."
        }
      ],
      "specialConsiderations": [
          "Commonly used salvage option for patients ineligible for
high-dose therapy/ASCT.",
          "Can often be administered in the outpatient setting.",
          "Less nephrotoxic/ototoxic than cisplatin-based regimens.",
          "Monitor for hematologic toxicity (esp. thrombocytopenia) and
peripheral neuropathy.",
          "Requires G-CSF support.",
          "Outcomes are modest, particularly long-term, when used with
palliative intent, but serves as an effective bridge to CAR-T in many
cases."
          ],
      "routeOfAdministration": "Intravenous",
      "settingOfAdministration": "Outpatient typically",
      "lengthOfTreatment": "Variable, often 4-8 cycles (repeated every 2
weeks) depending on response, tolerance, and intent (bridge vs
destination)."
    },
    {
      "name": "Loncastuximab tesirine (Zynlonta)",
      "applicableLine": "Third-line or subsequent",
```

```
      "tags": ["Antibody-Drug Conjugate", "CD19 Target", "R/R DLBCL", "R/R
HGBL", "AIFA Approved", "Post-CAR T Option"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory DLBCL or High-Grade B-cell
Lymphoma (HGBL)",
          "After two or more prior lines of systemic therapy",
          "Adequate organ function (as per trial/label)",
          ],
        "exclusion": [
          "Previous severe reaction to loncastuximab tesirine",
          "Significant comorbidities precluding treatment (e.g., severe
hepatic impairment not studied)"
        ]
      },
      "registrationStudyCitation": "Caimi PF, Ai W, Alderuccio JP, et al.
Loncastuximab tesirine in relapsed or refractory diffuse large B-cell
lymphoma (LOTIS-2): a multicentre, open-label, single-arm, phase 2 trial.
Lancet Oncol. 2021;22(6):790-800. doi:10.1016/S1470-2045(21)00139-X",
      "keyPapers": [
        {
          "citation": "Caimi PF, Ai W, Alderuccio JP, et al. Loncastuximab
tesirine in relapsed or refractory diffuse large B-cell lymphoma (LOTIS-2):
a multicentre, open-label, single-arm, phase 2 trial. Lancet Oncol.
2021;22(6):790-800. doi:10.1016/S1470-2045(21)00139-X",
          "reportedOutcomes": "ORR 48.3%, CR 24.1%/24.8%, Median DoR
~10.3-13.4 months (longer for CRs), Median OS 9.5 months, Median PFS 4.9
months.",
          "numberOfPatientsStudied": "145 (All treated / safety
population)",
          "patientAndDiseaseCharacteristics": "Heavily pretreated (median 3
prior lines); included high-risk groups (DH/TH, transformed DLBCL, primary
refractory, prior SCT, prior CAR-T); Median age ~66 years, ~55% >=65 years;
ECOG 0-2.",
          "toxicities": "Common Grade >=3 TEAEs: Neutropenia (~26%),
Thrombocytopenia (~18%), elevated GGT (~17%). Other notable AEs:
Edema/effusions (peripheral, pleural, pericardial), photosensitivity,
fatigue, myelosuppression."
        },
        {
          "citation": "Caimi PF, Ardeshna KM, Reid E, et al. Loncastuximab
tesirine in relapsed/refractory diffuse large B-cell lymphoma: long-term
efficacy and safety from the phase II LOTIS-2 study. Haematologica.
2024;109(4):1185-1190. doi:10.3324/haematol.2023.283849",
          "reportedOutcomes_summary": "Long-term follow-up confirms durable
responses in CR patients (median DoR not reached for CRs at data cut-off).
Median OS 9.5 mo, median PFS 4.9 mo overall.",
          "patientAndDiseaseCharacteristics": "Consistent with primary
analysis.",
```

```
          "toxicities": "Acceptable long-term safety profile, consistent
with earlier reports."
        },
        {
          "citation_summary": "Alderuccio JP, et al. Blood Adv. 2022
(LOTIS-2 HGBL Subgroup)",
          "reportedOutcomes_summary": "Confirmed activity in high-risk HGBL
subgroup (n=11) with ORR 45.5%, consistent with overall LOTIS-2
population.",
          "pubmed_id": "35790100"
        },
        {
          "citation_summary": "Epperla N, et al. Blood Cancer J. 2024
(Post-CAR T RWD)",
          "reportedOutcomes_summary": "Demonstrated efficacy in real-world
post-CAR T setting (3L/4L use); ORR 73-78% with durable responses observed
(12m OS 84-95%).",
          "pubmed_id": "39609431"
        }
      ],
      "specialConsiderations": [
          "Monitor for myelosuppression (neutropenia, thrombocytopenia),
elevated LFTs (esp. GGT), edema/effusions, photosensitivity.",
          "Requires premedication (e.g., dexamethasone).",
          "Dose reduction or delay may be needed for toxicity.",
          "Active in heavily pretreated / high-risk patients, including HGBL
and post-CAR T failures.",
          "Approved by AIFA for R/R DLBCL or HGBL after >=2 prior systemic
therapies."
          ],
      "routeOfAdministration": "Intravenous infusion over 30 minutes",
      "settingOfAdministration": "Outpatient typically",
      "lengthOfTreatment": "Every 3 weeks. Dose: 0.15 mg/kg for cycles 1-2,
then 0.075 mg/kg for subsequent cycles. Max duration up to 1 year or until
progression/unacceptable toxicity."
    },
    {
      "name": "R-Pola-Benda (Polatuzumab Vedotin + Bendamustine +
Rituximab)",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Antibody-Drug Conjugate", "CD79b Target", "Transplant
ineligible", "R/R DLBCL", "AIFA Approved"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory DLBCL",
          "Ineligible for Hematopoietic Stem Cell Transplant (HSCT)",
          "Received at least one prior therapy"
          ],
        "exclusion": [
```

```
          "Previous severe reaction to polatuzumab vedotin, bendamustine, or
rituximab",
          "Significant pre-existing peripheral neuropathy (Grade >1
potentially)",
          "Active CNS lymphoma"
        ]
      },
      "registrationStudyCitation": "Sehn LH, Herrera AF, Flowers CR, et al.
Polatuzumab Vedotin in Relapsed or Refractory Diffuse Large B-Cell Lymphoma.
J Clin Oncol. 2020;38(2):155-165. doi:10.1200/JCO.19.00172 (Primary analysis
of GO29365 randomized cohort)",
      "keyPapers": [
        {
          "citation1": "Sehn LH, Herrera AF, Flowers CR, et al. Polatuzumab
Vedotin in Relapsed or Refractory Diffuse Large B-Cell Lymphoma. J Clin
Oncol. 2020;38(2):155-165. doi:10.1200/JCO.19.00172",
          "reportedOutcomes": "Primary endpoint (ITT): CR rate 40.0%
(Pola-BR) vs 17.5% (BR alone) (P = .026). Median PFS 9.5 months (Pola-BR) vs
3.7 months (BR) (HR 0.36; 95% CI, 0.21-0.63; P < .001). Median OS 12.4
months (Pola-BR) vs 4.7 months (BR) (HR 0.42; 95% CI, 0.24-0.75; P = .002).
Median DoR 12.6 months (Pola-BR) vs 7.7 months (BR).",
          "numberOfPatientsStudied": "80 (40 Pola-BR, 40 BR - randomized
cohort)",
          "patientAndDiseaseCharacteristics": "Transplant-ineligible R/R
DLBCL; Median 2 prior therapies (range 1-7); Median age ~67 years; ~45%
refractory to last therapy.",
          "toxicities": "Higher rates of Grade 3-4 cytopenias (Neutropenia
46% vs 33%, Thrombocytopenia 41% vs 23%, Anemia 28% vs 18%) and Peripheral
Neuropathy (any grade 44% vs 5%, Grade >=2 15% vs 0%) with Pola-BR vs BR.
Infections common in both arms. Manageable safety profile overall."
        },
        {
          "citation2": "Tilly H, Morschhauser F, Sehn LH, et al. Polatuzumab
vedotin plus bendamustine and rituximab in relapsed/refractory DLBCL:
survival update and new extension cohort data. Blood Adv.
2022;6(5):1417-1421. doi:10.1182/bloodadvances.2021005796",
          "reportedOutcomes_summary": "Updated survival from randomized arms
confirmed benefit: Median PFS 9.2 vs 3.7 months (HR 0.39), Median OS 12.4 vs
4.7 months (HR 0.42) favoring Pola-BR. Extension cohort (n=106 Pola-BR)
showed ORR 41.5%, CR 38.7%, mPFS 6.6 mo, mOS 12.5 mo.",
          "numberOfPatientsStudied": "Randomized N=80; Extension N=106",
          "toxicities": "No new safety signals identified in extension
cohort."
        },
         {
          "citation_summary": "Real-World Studies (e.g., Eyre TA et al. Br J
Haematol. 2024; Qualls D et al. Br J Haematol. 2024)",
          "reportedOutcomes_summary": "Real-world outcomes are variable and
often lower than GO29365, particularly in trial-ineligible patients (e.g.,
```

```
poorer performance status, >2 prior lines). mPFS ~2-5 months, mOS ~3.5-9
months reported in some cohorts. Patient selection is important.",
          "pubmed_id": "PMIDs for example RWD studies if needed: 37357836
(Qualls), 361756038 (Eyre)"
        }
      ],
      "specialConsiderations": [
          "Monitor for cytopenias (requires G-CSF support frequently),
peripheral neuropathy, infusion reactions, and infections.",
          "Bendamustine dose adjustments may be required.",
          "Demonstrated significant survival benefit over BR alone in
transplant-ineligible R/R DLBCL.",
          "Real-world effectiveness may be lower than trial results,
especially in heavily pre-treated or trial-ineligible populations.",
          "AIFA approved for R/R DLBCL in adults not eligible for HSCT."
          ],
      "routeOfAdministration": "Polatuzumab Vedotin: IV; Bendamustine: IV;
Rituximab: IV",
      "settingOfAdministration": "Outpatient typically",
      "lengthOfTreatment": "Up to 6 cycles (21-day cycles)"
    },
    {
      "name": "Epcoritamab (Epkinly)",
      "applicableLine": "Third-line or subsequent",
      "tags": ["Bispecific Antibody", "CD3xCD20", "R/R DLBCL",
"Subcutaneous", "AIFA Approved"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory DLBCL (including DLBCL NOS,
HGBL, PMBCL, transformed DLBCL)",
          "After two or more prior lines of systemic therapy",
          "Adequate organ function (as per trial/label)",
          "ECOG Performance Status 0-2 (typically)"
          ],
        "exclusion": [
          "Previous severe reaction to epcoritamab",
          "Prior treatment with CD3xCD20 bispecific antibody",
          "Active CNS lymphoma (typically excluded from trials)",
          "Significant active infections",
          "Prior allogeneic SCT may be exclusion criterion"
        ]
      },
      "registrationStudyCitation": "Thieblemont C, Phillips T, Ghesquieres
H, et al. Epcoritamab, a Novel, Subcutaneous CD3xCD20 Bispecific
T-Cell-Engaging Antibody, in Relapsed or Refractory Large B-Cell Lymphoma:
Dose Expansion in a Phase I/II Trial (EPCORE NHL-1). J Clin Oncol.
2023;41(12):2238-2247. doi:10.1200/JCO.22.01725",
      "keyPapers": [
        {
```

```
          "citation1": "Thieblemont C, Phillips T, Ghesquieres H, et al. Epcoritamab, a Novel, Subcutaneous CD3xCD20 Bispecific T-Cell-Engaging Antibody, in Relapsed or Refractory Large B-Cell Lymphoma: Dose Expansion in a Phase I/II Trial (EPCORE NHL-1). J Clin Oncol. 2023;41(12):2238-2247. doi:10.1200/JCO.22.01725",
          "reportedOutcomes": "LBCL Cohort (N=157): ORR 63.1%, CR 39.5%. Median DoR 15.5 months. Median OS 18.5 months (NR for CR patients). Median PFS ~4.4 months overall. Deep and durable responses observed, particularly in CRs.",
          "numberOfPatientsStudied": "157 (LBCL expansion cohort)",
          "patientAndDiseaseCharacteristics": "Heavily pretreated R/R LBCL (median 3 prior lines); ~61% primary refractory; ~83% refractory to last therapy; ~39% prior CAR-T. Median age ~64 years.",
          "toxicities": "CRS common (~51%) but mostly Grade 1-2 (~48%) and predictable timing (Cycle 1). ICANS infrequent (~6% any grade, rare high grade/fatal). Other common AEs: fatigue, pyrexia, injection site reaction, nausea, diarrhea, neutropenia. Manageable safety profile with mitigation strategies."
        },
        {
          "citation_summary": "Jurczak W, et al. ASH 2024 Abstract / EHA 2024 Poster (EPCORE NHL-1 3-Year Update)",
          "reportedOutcomes_summary": "3-year follow-up: ORR 59%, CR 41%. Median DoR 20.8 mo; Median DoCR 36.1 mo. Median OS 18.5 mo (NR for CR patients, 63% alive at 36 mo). Confirms durability, especially for CRs.",
          "toxicities": "Long-term safety consistent with previous reports; no new signals."
        }
      ],
      "specialConsiderations": [
          "Subcutaneous administration.",
          "Requires Cycle 1 step-up dosing to mitigate CRS risk.",
          "Monitor closely for CRS and ICANS, especially during Cycle 1.",
          "Prophylaxis (e.g., corticosteroids) and mitigation strategies (e.g., hydration, withhold antihypertensives) recommended.",
          "High response rates and durable remissions achievable, even in poor-prognosis groups (e.g., prior CAR-T).",
          "AIFA approved for R/R DLBCL after >=2 prior systemic therapies."
          ],
      "routeOfAdministration": "Subcutaneous injection",
      "settingOfAdministration": "Requires monitoring for initial doses (potential hospitalization optional/recommended), subsequent doses often outpatient.",
      "lengthOfTreatment": "Cycle 1 (28 days): Weekly step-up doses (e.g., 0.16mg D1, 0.8mg D8, 48mg D15, 48mg D22). Cycles 2-3: Weekly (48mg). Cycles 4-9: Every 2 weeks (48mg). Cycle 10+: Every 4 weeks (48mg). Continue until progression or unacceptable toxicity."
    },
    {
      "name": "Glofitamab (Columvi)",
```

```
      "applicableLine": "Third-line or subsequent",
      "tags": ["Bispecific Antibody", "CD3xCD20", "Fixed Duration", "R/R
DLBCL", "AIFA Approved", "Post-CAR T Option"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory DLBCL",
          "After two or more prior lines of systemic therapy",
          "Adequate organ function (as per trial/label)",
          "ECOG Performance Status 0-1 (often required for trials)"
          ],
        "exclusion": [
          "Previous severe reaction to glofitamab or obinutuzumab",
          "Prior treatment with CD3xCD20 bispecific antibody",
          "Active CNS lymphoma",
          "Significant active infections"
        ]
      },
      "registrationStudyCitation": "Dickinson MJ, Carlo-Stella C,
Morschhauser F, et al. Glofitamab for Relapsed or Refractory Diffuse Large
B-Cell Lymphoma. N Engl J Med. 2022;387(24):2220-2231.
doi:10.1056/NEJMoa2206913",
      "keyPapers": [
        {
          "citation1": "Dickinson MJ, Carlo-Stella C, Morschhauser F, et al.
Glofitamab for Relapsed or Refractory Diffuse Large B-Cell Lymphoma. N Engl
J Med. 2022;387(24):2220-2231. doi:10.1056/NEJMoa2206913",
          "reportedOutcomes": "Primary efficacy analysis (N=155): ORR 51.6%
(IRC), CR 39.4% (IRC) [other reports cite ~52-56% ORR / 39-43% CR]. Median
DoCR ~24.1 months (at ~18mo f/up). Median OS ~12.0 months. Median PFS ~4.9
months. Durable responses observed, especially CRs.",
          "numberOfPatientsStudied": "155 (primary efficacy analysis cohort
from NP30179)",
          "patientAndDiseaseCharacteristics": "Heavily pretreated R/R DLBCL
(median 3 prior lines); ~33% prior CAR-T; ~85% refractory to last regimen.
Median age ~66 years.",
          "toxicities": "CRS common (~63-64%) but mostly Grade 1-2 (~60%)
and primarily in Cycle 1 after Gpt and step-up dosing; Grade 3-4 CRS
infrequent (~4%). ICANS less common (~6% any grade). Other common AEs:
neutropenia, anemia, thrombocytopenia, infections, fatigue, pyrexia.
Manageable safety profile."
        },
        {
          "citation_summary": "Hutchings M, et al. J Clin Oncol.
2023;41(16_suppl):7550 (NP30179 Extended Follow-up)",
          "reportedOutcomes_summary": "Extended follow-up (~18 months median
in CR patients) confirmed durability: Median DoCR 24.1 months; estimated 70%
of CRs ongoing at 18 months. Median OS ~18-20 months with longer
follow-up.",
          "toxicities": "Incidence of AEs stable over time, no new safety
signals."
```

```
        },
        {
          "citation_summary": "Rentsch V, et al. Cancers 2022 (Post-CAR T
Subgroup)",
          "reportedOutcomes_summary": "Study suggested good tolerance and
efficacy specifically in patients relapsing after CAR T-cell therapy.",
          "pubmed_id": "35626120"
        },
        {
           "citation_summary": "Shumilov E, et al. Blood Adv. 2024 (DACH
Real-World Data)",
           "reportedOutcomes_summary": "Multinational RWD (N=70, median 4
prior lines, 71% prior CAR-T): ORR 47%, CR 27%, mPFS 3.6 mo, mOS 5.7 mo.
Prior bendamustine may impair efficacy.",
           "pubmed_id": "39661985"
        }
      ],
      "specialConsiderations": [
          "Requires single dose of Obinutuzumab pre-treatment (Gpt) 7 days
before first glofitamab dose to mitigate CRS.",
          "Fixed-duration therapy (12 cycles total).",
          "Monitor closely for CRS (esp. Cycle 1) and ICANS. Manage
according to guidelines (e.g., ASTCT). Provide patient alert card.",
          "Monitor for infections and cytopenias.",
          "Shows efficacy in high-risk groups including prior CAR-T
failure.",
          "AIFA approved for R/R DLBCL after >=2 prior systemic therapies."
          ],
      "routeOfAdministration": "Intravenous infusion",
      "settingOfAdministration": "Hospital setting required for initial
doses due to CRS risk; subsequent cycles may be outpatient depending on
tolerance and local policy.",
      "lengthOfTreatment": "Fixed 12 cycles (21-day cycles). Cycle 1: Gpt
1000mg IV D1, Glofitamab 2.5mg IV D8, Glofitamab 10mg IV D15. Cycles 2-12:
Glofitamab 30mg IV D1."
    },
    {
      "name": "Axicabtagene ciloleucel (Yescarta)",
      "applicableLine": "Second-line OR Third-line or subsequent",
      "tags": ["CAR-T Therapy", "CD19 Target", "R/R DLBCL", "R/R HGBL", "R/R
PMBCL", "AIFA Approved", "Second Line", "Third Line Plus"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory (R/R) DLBCL, High-Grade B-cell
Lymphoma (HGBL), Primary Mediastinal Large B-cell Lymphoma (PMBCL)",
          "Second-line indication (ZUMA-7 based): R/R within 12 months of
first-line chemoimmunotherapy",
          "Third-line or subsequent indication (ZUMA-1 based): R/R after two
or more lines of systemic therapy",
          "Adequate organ function (per protocol/label)",
```

```
          "ECOG Performance Status 0-1 (typically required for
trials/eligibility)"
          ],
        "exclusion": [
          "Previous severe reaction to axi-cel components",
          "Active CNS lymphoma (unless specifically addressed in
protocol/indication)",
          "Significant active infections or comorbidities precluding CAR-T",
          "Prior CD19 CAR-T therapy"
        ],
         "aifa_specific_criteria": "Initial AIFA approval for 3L+; 2L
indication added Nov 2023 for R/R within 12mo of 1L chemoimmunotherapy. Age
limit for 3L+ extended to 75 years in Italy."
      },
      "registrationStudyCitation": "ZUMA-1 (3L+): Neelapu SS, et al. Lancet
Oncol. 2017; ZUMA-7 (2L): Locke FL, et al. N Engl J Med. 2022",
      "keyPapers": [
        {
          "citation1": "Neelapu SS, Jacobson CA, Ghobadi A, et al. Five-year
follow-up of ZUMA-1 supports the curative potential of axicabtagene
ciloleucel in refractory large B-cell lymphoma. Blood.
2023;141(19):2307-2315. doi:10.1182/blood.2022018893",
          "study_context": "ZUMA-1 (Phase 1/2) - 3L+ Setting",
          "reportedOutcomes": "ORR 83%, CR 58%. 5-year OS rate 42.6%; 5-year
disease-specific survival 51.0%. Median OS 25.8 months. Suggests curative
potential for some.",
          "numberOfPatientsStudied": "101 (Phase 2 efficacy population)",
          "patientAndDiseaseCharacteristics": "R/R LBCL after >=2 prior
lines; median age 58; refractory disease common.",
          "toxicities": "CRS ~93% (Grade>=3 ~11%), Neurotoxicity (ICANS)
~64% (Grade>=3 ~30% - later reports use ~87% / ~31%). Median onset CRS ~2
days, NE ~6 days."
        },
        {
          "citation2": "Locke FL, Miklos DB, Jacobson CA, et al.
Axicabtagene Ciloleucel as Second-Line Therapy for Large B-Cell Lymphoma. N
Engl J Med. 2022;386(7):640-654. doi:10.1056/NEJMoa2116133 (Primary
analysis)",
          "study_context": "ZUMA-7 (Phase 3) - 2L Setting vs SOC",
          "reportedOutcomes": "Superior EFS vs SOC (median 8.3 vs 2.0 mo; HR
0.398; 24m EFS 40.5% vs 16.3%). Superior OS vs SOC (median NR vs ~31 mo
est.; HR 0.726; 39m OS 55.9% vs 46%). ORR 83% vs 50%; CR 65% vs 32%
(investigator assessed).",
          "numberOfPatientsStudied": "359 randomized (180 Axi-cel, 179
SOC)",
          "patientAndDiseaseCharacteristics": "R/R LBCL within 12 months of
1L chemoimmunotherapy. Median age ~59 years.",
          "toxicities": "Axi-cel arm: CRS ~92% (Grade>=3 6%), Neurotoxicity
(ICANS) ~60% (Grade>=3 21% - later reports use ~74% / ~25%). Safety profile
consistent with ZUMA-1."
```

```
        },
        {
          "citation_summary": "Westin JR, et al. Blood Adv. 2024 / Oluwole
OO et al. Clin Cancer Res 2024 (ZUMA-7 OS / Subgroup Analyses)",
          "reportedOutcomes_summary": "Confirmed significant OS benefit for
Axi-cel over SOC in 2L at ~47mo median follow-up (HR 0.726). Benefit
observed across key subgroups including elderly (>=65 years) and those with
high tumor burden."
        }
      ],
      "specialConsiderations": [
          "Requires leukapheresis and centralized manufacturing (turnaround
time considerations).",
          "Bridging therapy often required between leukapheresis and
infusion.",
          "Lymphodepleting chemotherapy (Fludarabine/Cyclophosphamide)
administered before infusion.",
          "Requires administration at certified treatment centers with REMS
program (US) or equivalent.",
          "Mandatory monitoring for CRS and ICANS post-infusion (typically
inpatient).",
          "Management of CRS/ICANS with Tocilizumab and/or
corticosteroids.",
          "Risk of prolonged cytopenias, B-cell aplasia,
hypogammaglobulinemia (may require IVIG).",
          "Potential for long-term remission/cure."
          ],
      "routeOfAdministration": "Single Intravenous infusion",
      "settingOfAdministration": "Certified hospital/treatment center",
      "lengthOfTreatment": "Single infusion (preceded by leukapheresis,
bridging, lymphodepletion)"
    },
      {
      "name": "Tisagenlecleucel (Kymriah)",
      "applicableLine": "Third-line or subsequent",
      "tags": ["CAR-T Therapy", "CD19 Target", "R/R DLBCL", "AIFA Approved",
"Third Line Plus"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory (R/R) DLBCL",
          "After two or more lines of systemic therapy",
          "Adequate organ function (per protocol/label)",
          "ECOG Performance Status 0-1 (typically required)"
          ],
        "exclusion": [
          "Previous severe reaction to tisagenlecleucel components",
          "Active CNS lymphoma",
          "Significant active infections or comorbidities precluding CAR-T",
          "Prior CD19 CAR-T therapy"
        ],
```

```
        "aifa_specific_criteria": "Approved for adult R/R DLBCL after >=2
prior systemic therapies. Initial AIFA registry noted age <=70, later
versions may align with broader EMA label (no upper age limit specified but
data primarily in <65/75)."
      },
      "registrationStudyCitation": "Schuster SJ, Bishop MR, Tam CS, et al;
JULIET Investigators. Tisagenlecleucel in Adult Relapsed or Refractory
Diffuse Large B-Cell Lymphoma. N Engl J Med. 2019;380(1):45-56.
doi:10.1056/NEJMoa1804980",
      "keyPapers": [
        {
          "citation1": "Schuster SJ, Bishop MR, Tam CS, et al; JULIET
Investigators. Tisagenlecleucel in Adult Relapsed or Refractory Diffuse
Large B-Cell Lymphoma. N Engl J Med. 2019;380(1):45-56.
doi:10.1056/NEJMoa1804980 (Primary analysis + updates)",
          "study_context": "JULIET (Phase 2) - 3L+ Setting",
          "reportedOutcomes": "ORR 52% (ITT, n=111), CR 40%. Durable
responses: Median DoR not reached for CR patients. Median OS ~11.7 months
overall [check NEJM 2019]. Real-world data (CIBMTR) show consistent efficacy
(ORR ~60%, CR ~45%).",
          "numberOfPatientsStudied": "111 (infused/primary efficacy set);
165 enrolled.",
          "patientAndDiseaseCharacteristics": "Adults R/R DLBCL after >=2
lines; Median age 56; Median 3 prior lines; ~49% prior ASCT; Includes
transformed FL.",
          "toxicities": "CRS ~58% (Grade>=3 ~22% using Penn scale).
Neurotoxicity (ICANS) ~21% (Grade>=3 ~12% using Penn scale). Real-world data
suggest lower rates of severe CRS/ICANS (e.g., Grade >=3 CRS ~4-6%, Neuro
~5-7%) possibly due to evolving management. Prolonged cytopenias observed."
        },
        {
           "citation_summary": "Jacobson CA, et al. CIBMTR Real-World
Outcomes Analysis",
           "reportedOutcomes_summary": "Large RWD registry analysis
confirmed JULIET findings: ORR ~60%, CR ~45% in ~968 efficacy-evaluable
patients. 24m PFS ~28%, 24m OS ~44%. Safety profile appeared better than in
JULIET (lower severe CRS/ICANS rates).",
           "pubmed_id": "Example: 38791230 (JITC 2024 publication of this
data)"
        }
      ],
      "specialConsiderations": [
          "Requires leukapheresis and centralized manufacturing.",
          "Bridging therapy often used (~90% in JULIET).",
          "Lymphodepleting chemotherapy (Fludarabine/Cyclophosphamide or
Bendamustine) administered before infusion.",
          "Requires administration at certified treatment centers.",
          "Mandatory monitoring for CRS and ICANS post-infusion (typically
inpatient).",
```

```
          "Management of CRS/ICANS with Tocilizumab and/or
corticosteroids.",
          "Risk of prolonged cytopenias, B-cell aplasia,
hypogammaglobulinemia (may require IVIG).",
          "Outpatient administration feasible in selected centers/patients."
          ],
      "routeOfAdministration": "Single Intravenous infusion",
      "settingOfAdministration": "Certified hospital/treatment center;
inpatient standard, outpatient possible",
      "lengthOfTreatment": "Single infusion (preceded by leukapheresis,
bridging, lymphodepletion)"
    },
      {
      "name": "Lisocabtagene maraleucel (Breyanzi)",
      "applicableLine": "Second-line OR Third-line or subsequent",
      "tags": ["CAR-T Therapy", "CD19 Target", "Defined CD4/CD8 Ratio", "R/R
DLBCL", "R/R HGBL", "R/R PMBCL", "R/R FL3B", "AIFA Approved", "Second Line",
"Third Line Plus"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory (R/R) Large B-Cell Lymphoma
(LBCL), including DLBCL NOS, High-Grade B-cell Lymphoma (HGBL), Primary
Mediastinal Large B-cell Lymphoma (PMBCL), and Follicular Lymphoma Grade 3B
(FL3B)",
          "Second-line indication (TRANSFORM based): R/R within 12 months of
first-line chemoimmunotherapy",
          "Third-line or subsequent indication (TRANSCEND NHL 001 based):
R/R after two or more lines of systemic therapy",
          "Adequate organ function (per protocol/label)",
          "ECOG Performance Status 0-1 (typically required)"
          ],
        "exclusion": [
          "Previous severe reaction to liso-cel components",
          "Active CNS lymphoma (unless specifically addressed)",
          "Significant active infections or comorbidities precluding CAR-T",
          "Prior CD19 CAR-T therapy"
        ],
         "aifa_specific_criteria": "Approved for 3L+ R/R DLBCL, PMBCL, FL3B
after >=2 lines. Second-line indication (R/R within 12mo of 1L chemoimmuno)
also approved [implied by EMA/FDA alignment, needs explicit AIFA
confirmation if available - guideline file supports 2L use based on
criteria]."
      },
      "registrationStudyCitation": "TRANSFORM (2L): Kamdar M, et al. Lancet.
2022; TRANSCEND NHL 001 (3L+): Abramson JS, et al. Lancet. 2020",
      "keyPapers": [
        {
          "citation1": "Kamdar M, Solomon SR, Arnason J, et al; TRANSFORM
Investigators. 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. doi:10.1016/S0140-6736(22)00662-6",
          "study_context": "TRANSFORM (Phase 3) - 2L Setting vs SOC",
          "reportedOutcomes": "Superior EFS vs SOC (median 10.1 vs 2.3 mo; HR 0.35; 18m EFS ~54% vs 21%). CR rate 74% vs 43%. PFS benefit (HR ~0.40) and trend for OS benefit observed vs SOC.",
          "numberOfPatientsStudied": "184 randomized (92 Liso-cel, 92 SOC)",
          "patientAndDiseaseCharacteristics": "R/R LBCL within 12 months of 1L chemoimmunotherapy; eligible for ASCT. Median age ~60 years.",
          "toxicities": "Liso-cel arm: Grade >=3 CRS 1%, Grade >=3 NE 4% [Calculated from reported rates in study/reviews]. Prolonged Grade >=3 cytopenias ~43%. Manageable safety profile."
        },
        {
          "citation2": "Palomba ML, Gordon LI, Siddiqi T, et al. Two-year follow-up of lisocabtagene maraleucel in relapsed or refractory large B-cell lymphoma in TRANSCEND NHL 001. Blood. 2024;143(5):406-418. doi:10.1182/blood.2023021130",
          "study_context": "TRANSCEND NHL 001 (Phase 1) - 3L+ Setting",
          "reportedOutcomes": "ORR 73%, CR 53%. Median DoR 23.1 months. Median PFS 6.8 months; Median OS 27.3 months. Estimated 2-year OS rate 50.5%. Durable remissions observed.",
          "numberOfPatientsStudied": "269 treated (primary analysis set ~255-257)",
          "patientAndDiseaseCharacteristics": "R/R LBCL after >=2 lines; median 3 prior lines; includes various histologies (DLBCL NOS, HGBL, PMBCL, FL3B, TFL).",
          "toxicities": "Grade >=3 CRS 2%, Grade >=3 NE 10%. Most common Grade >=3 AEs: neutropenia (60%), anemia (37%), thrombocytopenia (27%). Low rates of severe CRS/NE."
        },
        {
          "citation_summary": "Riedell PA, et al. Blood Adv. 2024 (Real-World Data - Cell Therapy Consortium)",
          "reportedOutcomes_summary": "Multi-center RWD (N=101, median age 71, high comorbidity): Day 90 ORR 66%, CR 60%. 12m PFS ~55%, 12m OS ~68%. Outcomes similar to pivotal trials despite older/more comorbid population.",
          "toxicities": "Manageable safety in RWD: Any grade CRS 49% (Grade>=3 3%), any grade ICANS 26% (Grade>=3 10%)."
        }
      ],
      "specialConsiderations": [
          "Defined composition of CD4+ and CD8+ CAR T cells.",
          "Requires leukapheresis and centralized manufacturing.",
          "Bridging therapy allowed/common.",
          "Lymphodepleting chemotherapy (Fludarabine/Cyclophosphamide) administered before infusion.",
          "Requires administration at certified treatment centers.",

```
          "Mandatory monitoring for CRS and ICANS post-infusion.",
          "Generally lower rates of severe CRS/ICANS reported compared to
Axi-cel in trials.",
          "Outpatient administration/monitoring feasible in selected
patients/centers.",
          "Risk of prolonged cytopenias, B-cell aplasia,
hypogammaglobulinemia (may require IVIG)."
          ],
      "routeOfAdministration": "Single Intravenous infusion (dose based on
target T-cell count with defined CD4:CD8 ratio)",
      "settingOfAdministration": "Certified hospital/treatment center;
inpatient or outpatient possible",
      "lengthOfTreatment": "Single infusion (preceded by leukapheresis,
bridging, lymphodepletion)"
    },
    {
      "name": "ASCT (BEAM Conditioning)",
      "applicableLine": "Second-line or subsequent (Consolidation)",
      "tags": ["Consolidation", "Transplant eligible", "Fit",
"Chemosensitive disease", "High-Dose Chemotherapy"],
      "criteria": {
        "inclusion": [
          "Relapsed or refractory DLBCL (or other eligible lymphomas)",
          "Achieved Chemosensitivity (Complete Response or Partial Response)
to salvage chemotherapy",
          "Medically fit for high-dose therapy (typically age <65-70 years,
ECOG PS 0-1, adequate cardiac, pulmonary, renal, hepatic function)",
          "Successful collection of adequate autologous hematopoietic stem
cells"
          ],
        "exclusion": [
          "Chemoresistant disease (Stable Disease or Progressive Disease
after salvage)",
          "Significant comorbidities or organ dysfunction precluding
high-dose therapy",
          "Poor performance status (ECOG >=2)",
          "Inadequate stem cell collection",
          "Primary refractory disease or early relapse (<12 months) where
CAR-T therapy is now preferred if available/eligible"
        ],
        "guideline_context": "Standard consolidation for transplant-eligible
patients with chemosensitive relapse (especially >12 months after first-line
therapy). Role in second-line early relapse/primary refractory largely
superseded by CAR-T therapy based on ZUMA-7/TRANSFORM trials."
      },
      "registrationStudyCitation": "N/A (Standard procedure established by
trials like PARMA - Philip T, NEJM 1995)",
      "keyPapers": [
        {
```

```
          "citation1": "Philip T, Guglielmi C, Hagenbeek A, et al.
Autologous bone marrow transplantation as compared with salvage chemotherapy
in relapses of chemotherapy-sensitive non-Hodgkin's lymphoma. N Engl J Med.
1995;333(23):1540-5. (Landmark PARMA trial - pre-Rituximab era)",
          "reportedOutcomes_summary": "Established superiority of ASCT over
conventional salvage chemo alone in chemosensitive R/R aggressive NHL."
        },
        {
          "citation2": "Gisselbrecht C, Glass B, Mounier N, et al. Salvage
regimens with autologous transplantation for relapsed large B-cell lymphoma
in the rituximab era. J Clin Oncol. 2010;28(27):4184-90. (CORAL trial -
Rituximab era)",
          "reportedOutcomes_summary": "Evaluated salvage chemo before ASCT.
Post-ASCT 3-year EFS was ~40-50% for patients responding to salvage.
Confirmed ASCT benefit in Rituximab era for chemosensitive disease.",
          "patientAndDiseaseCharacteristics": "R/R DLBCL patients eligible
for ASCT in Rituximab era."
        },
        {
          "citation3": "Hamadani M, et al. Bone Marrow Transplant.
2020;55(7):1385-1393. (BEAM vs R-BEAM CIBMTR study)",
          "reportedOutcomes_summary": "Large registry study showed no
significant difference in OS or PFS with addition of Rituximab to BEAM
(R-BEAM) vs BEAM alone in R/R DLBCL.",
           "toxicities": "Similar toxicity profiles reported."
        },
        {
            "citation_summary": "Recent ASCT outcome data (e.g., Abdul-Hay
M, Haematologica 2024)",
            "reportedOutcomes_summary": "Confirms reasonable outcomes
post-ASCT in chemosensitive R/R DLBCL (e.g., median PFS ~16 mo, OS ~43 mo
overall in study cohort), particularly favorable if CR achieved pre-ASCT and
only 1 line of salvage needed. Relapse post-ASCT still occurs (~40-50%) and
carries poor prognosis."
        }
      ],
      "specialConsiderations": [
          "Requires successful prior salvage chemotherapy to demonstrate
chemosensitivity.",
          "Requires successful stem cell mobilization and collection.",
          "BEAM is a standard, widely used conditioning regimen;
alternatives exist (e.g., BeEAM, LEAM) with varying efficacy/toxicity
profiles.",
          "High potential for cure in chemosensitive relapse, especially
later relapses.",
          "Significant acute toxicities (myelosuppression, mucositis,
infections) requiring intensive supportive care.",
          "Potential long-term complications (secondary malignancies,
infertility, organ dysfunction)."
          ],
```

```
      "routeOfAdministration": "High-dose multi-agent IV chemotherapy (BEAM)
over ~6-7 days, followed by IV infusion of autologous stem cells.",
      "settingOfAdministration": "Specialized transplant center; requires
prolonged inpatient hospitalization.",
      "lengthOfTreatment": "Conditioning chemotherapy over ~1 week, followed
by stem cell infusion and ~2-4 weeks (or more) hospitalization for
engraftment and recovery."
    },
    {
      "name": "Tafasitamab-Lenalidomide",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Immunotherapy", "CD19 Target", "Transplant ineligible", "R/R
DLBCL", "AIFA Approved"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory DLBCL (including DLBCL ex
indolent lymphoma)",
          "Ineligible for Autologous Stem Cell Transplant (ASCT)",
          "Adequate organ function",
          "ECOG Performance Status 0-2 (as per L-MIND study)"
          ],
        "exclusion": [
          "Previous severe reaction to tafasitamab or lenalidomide",
          "Active CNS lymphoma (typically excluded from trials)",
          "Prior anti-CD19 therapy or lenalidomide were exclusions in
L-MIND, but real-world use may occur post-CAR T",
          "AIFA criteria specify not eligible for ASCT"
        ]
      },
      "registrationStudyCitation": "Salles G, Duell J, González Barca E, et
al. Tafasitamab plus lenalidomide in relapsed or refractory diffuse large
B-cell lymphoma (L-MIND): a multicentre, prospective, single-arm, phase 2
study. Lancet Oncol. 2020;21(7):978-988. doi:10.1016/S1470-2045(20)30225-4",
      "keyPapers": [
        {
          "citation": "Duell J, Maddocks KJ, González-Barca E, et al.
Long-term outcomes from the phase II L-MIND study of tafasitamab (MOR208)
plus lenalidomide in patients with relapsed or refractory diffuse large
B-cell lymphoma. Haematologica. 2021;106(9):2417-2426.
doi:10.3324/haematol.2020.275958 (35-month follow-up)",
          "reportedOutcomes": "ORR 57.5%, CR 40.0%, Median DoR 43.9 months,
Median PFS 11.6 months, Median OS 33.5 months",
          "numberOfPatientsStudied": "80 (ITT population included in primary
analysis, received at least one dose)",
          "patientAndDiseaseCharacteristics": "Median age 72 years; 1-3
prior systemic therapies (median 2); ECOG PS 0-2; Transplant ineligible.",
          "toxicities": "Most common Grade >=3 TEAEs: Neutropenia (~50%),
thrombocytopenia (~18%), febrile neutropenia (~12%). Generally
well-tolerated."
        },
```

```
        {
          "citation": "Zinzani PL, Jurczak W,stelltjes N, et al. Tafasitamab
for patients with relapsed or refractory diffuse large B-cell lymphoma:
final 5-year efficacy and safety findings in the phase II L-MIND study.
Haematologica. 2024;109(1):219-227. doi:10.3324/haematol.2023.283480 (Final
5-year analysis)",
          "reportedOutcomes": "ORR 57.5%, CR 41.3%, Median DoR Not Reached.
Confirms long-term benefit and durability.",
          "numberOfPatientsStudied": "81 (Evaluable for efficacy)",
          "patientAndDiseaseCharacteristics": "Consistent with earlier
analyses.",
          "toxicities": "Consistent safety profile, no new signals."
        },
        {
          "citation_summary": "RE-MIND study (vs Lenalidomide mono RWD)",
          "reportedOutcomes_summary": "Confirmed Tafa contribution:
Significantly higher ORR (67% vs 34%) and CR (40% vs 13%) for Tafa+Len vs
Len mono in matched cohorts."
        },
        {
          "citation_summary": "RE-MIND2 study (vs other therapies RWD)",
          "reportedOutcomes_summary": "Matched analyses suggested improved
OS for Tafa+Len vs Pola-BR (mOS ~20 vs 7 mo) and vs R2 (mOS ~24 vs 7 mo);
comparable OS vs CAR-T (~22 vs 15 mo)."
        },
        {
          "citation_summary": "Representative Real-World Studies (e.g.,
Qualls et al., Blood 2023; Duell et al. Blood 2023)",
          "reportedOutcomes_summary": "Real-world outcomes often lower than
L-MIND (e.g., mPFS ~1.9-4.7 mo, mOS ~6.5-8.9 mo), especially in patients not
meeting strict L-MIND eligibility criteria (e.g., more prior lines, primary
refractory). Patient selection impacts outcomes."
        }
      ],
      "specialConsiderations": [
          "Monitor for hematologic toxicity (neutropenia, thrombocytopenia)
and infections.",
          "Lenalidomide dose often requires adjustment (e.g., start 20mg
instead of 25mg, further reductions based on tolerance).",
          "Real-world data suggest outcomes are best in patients closely
matching L-MIND eligibility criteria.",
          "Approved by AIFA for R/R DLBCL patients not eligible for ASCT."
          ],
      "routeOfAdministration": "Tafasitamab: Intravenous; Lenalidomide:
Oral",
      "settingOfAdministration": "Outpatient typically",
      "lengthOfTreatment": "Combination for up to 12 cycles (28-day cycles;
Tafa IV weekly cycles 1-3 then q2w, Len PO D1-21); then Tafasitamab
monotherapy q2w until progression or unacceptable toxicity"
    },
```

```
      {
      "name": "Rituximab-Lenalidomide (R2)",
      "applicableLine": "Second-line or subsequent",
      "tags": ["Chemo-free", "Immunomodulatory", "R/R DLBCL", "Comparator
Regimen"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory (R/R) DLBCL",
          "Often studied in specific contexts (e.g., non-GCB subtype,
maintenance after other therapy, or patients unsuitable for intensive
chemo/CAR-T)"
          ],
        "exclusion": [
          "Previous severe reaction to rituximab or lenalidomide",
          "High risk of thrombosis without prophylaxis",
          "Severe renal impairment (requires lenalidomide dose adjustment)"
        ],
        "guideline_context": "Not generally listed as a preferred standard
salvage regimen for unselected R/R DLBCL in major guidelines (e.g., NCCN)
compared to chemo-immunotherapy salvage, CAR-T, Pola-BR, Tafa-Len, or
bispecifics. More established in Follicular Lymphoma."
      },
      "registrationStudyCitation": "N/A for R/R DLBCL (Specific approval
primarily for Follicular Lymphoma; DLBCL use often off-label,
investigational, or in specific combinations)",
      "keyPapers": [
        {
          "citation1": "Czuczman MS, et al. J Clin Oncol.
2012;30(36):4546-53 (Phase II Study)",
          "study_context": "Early Phase II in R/R Aggressive NHL (including
DLBCL)",
          "reportedOutcomes": "DLBCL Cohort: ORR ~35%, CR ~13%. Activity
noted, particularly suggested in non-GCB subtype.",
          "numberOfPatientsStudied": "~49 DLBCL patients",
          "patientAndDiseaseCharacteristics": "R/R Aggressive NHL after >=1
prior therapy.",
          "toxicities": "Neutropenia, fatigue, rash, thrombocytopenia
common."
        },
        {
           "citation2": "Ghesquieres H, et al. Leuk Lymphoma.
2023;64(8):1435-1439 (Phase II High-Risk R/R DLBCL)",
           "study_context": "Phase II in High-Risk R/R DLBCL",
           "reportedOutcomes": "Study terminated early due to futility. ORR
12.5% (CR 6.3%). Median PFS 2.6 months, Median OS 9.3 months. Concluded R2
not recommended in this high-risk population.",
           "numberOfPatientsStudied": "16 evaluable patients",
           "patientAndDiseaseCharacteristics": "High-risk R/R DLBCL (defined
by IPI, refractoriness etc.). Median 4 prior lines.",
            "toxicities": "Neutropenia, anemia, fatigue, fever common."
```

```
        },
        {
          "citation_summary": "RE-MIND / RE-MIND2 Studies (Comparator Arm)",
          "reportedOutcomes_summary": "Used as comparator for
Tafasitamab+Lenalidomide. RE-MIND RWD cohort (Lenalidomide mono) had ORR
34%, CR 13%. RE-MIND2 RWD cohort (R2) had median OS 7.4 months. Tafa+Len
showed superior outcomes vs R2."
        }
      ],
      "specialConsiderations": [
          "Modest efficacy as monotherapy in unselected R/R DLBCL based on
recent trials.",
          "May have role in specific subtypes (e.g., non-GCB, though less
relevant now) or potentially frail patients unsuitable for other options,
but limited evidence.",
          "Requires monitoring for cytopenias, rash, fatigue.",
          "Thrombosis prophylaxis typically required with lenalidomide.",
          "Risk of secondary primary malignancies with long-term
lenalidomide.",
          "AIFA indication for R2 specifically targets Follicular Lymphoma;
DLBCL indication for Lenalidomide is in combination with Tafasitamab."
          ],
      "routeOfAdministration": "Rituximab: IV; Lenalidomide: Oral",
      "settingOfAdministration": "Outpatient",
      "lengthOfTreatment": "Variable, often continued until progression or
unacceptable toxicity (e.g., 12 cycles or more)."
    },
    {
      "name": "Radiotherapy",
      "applicableLine": "Consolidation, Salvage, Palliative, Bridging",
      "tags": ["Local Control", "Consolidation", "Palliative", "Bridging to
CAR-T", "CNS involvement"],
      "criteria": {
        "inclusion": [
          "Consolidation of response after systemic therapy (esp. initial
bulk or residual disease)",
          "Palliative treatment for symptomatic sites",
          "Bridging therapy prior to CAR-T cell infusion",
          "Primary or secondary CNS lymphoma treatment/consolidation (role
evolving)",
          "Part of combined modality therapy in early-stage DLBCL"
          ],
        "exclusion": [
          "Widespread systemic disease not amenable to local therapy (unless
palliative intent)",
          "Previous radiotherapy to the same site exceeding tolerance
limits",
          "Medical instability precluding treatment",
          "Specific concerns regarding toxicity (e.g., significant
neurotoxicity risk with WBRT)"
```

```
        ]
      },
      "guideline_context": "Standard component of curative-intent therapy
for early-stage DLBCL. Used for consolidation after chemotherapy in
advanced-stage DLBCL, particularly for sites of initial bulk or residual
disease. Role in PCNSL is debated due to neurotoxicity; may be used for
consolidation or salvage. Increasingly used as bridging therapy before
CAR-T. Palliative role for symptom control.",
      "keyPapers": [
         {
          "citation_summary": "BNLI Study / RCR Guidelines",
          "reportedOutcomes_summary": "Established 30 Gy as effective
consolidation dose post-chemo for aggressive NHL.",
          "toxicities": "Site-specific acute toxicities (e.g., mucositis,
skin reaction). Potential long-term effects depend on site (e.g.,
xerostomia, secondary malignancy risk, neurotoxicity with brain RT)."
        },
        {
          "citation_summary": "RICOVER-60 Study",
          "reportedOutcomes_summary": "Showed PFS/OS benefit for 36 Gy
consolidation to initial bulk/extranodal sites after R-CHOP in older
patients (>60 yrs) with advanced DLBCL."
        },
        {
          "citation_summary": "PCNSL WBRT Studies (e.g., IELSG32, PRECIS,
HOVON 105)",
          "reportedOutcomes_summary": "WBRT can improve disease control in
PCNSL but often associated with significant, potentially delayed
neurotoxicity (cognitive impairment). Modern regimens aim to reduce/defer
WBRT.",
          "toxicities": "Significant risk of long-term neurocognitive
decline, leukoencephalopathy, stroke risk noted in some studies."
        }
      ],
      "specialConsiderations": [
          "Dose and fractionation vary widely based on intent (curative,
consolidative, palliative, bridging) and location (e.g., 30 Gy/15 fractions
for consolidation, lower doses for palliation).",
          "Whole Brain Radiotherapy (WBRT) for CNS lymphoma carries
substantial risk of neurotoxicity, especially in older patients.",
          "Can be highly effective for local disease control.",
          "Considered as a non-cross-resistant modality useful for bridging
to cellular therapies."
          ],
      "routeOfAdministration": "External Beam Radiotherapy",
      "settingOfAdministration": "Outpatient typically",
      "lengthOfTreatment": "Variable, typically 1-4 weeks depending on total
dose and fractionation schedule."
    },
    {
```

```
      "name": "Selinexor (Xpovio)",
      "applicableLine": "Third-line or subsequent",
      "tags": ["Oral", "Nuclear export inhibitor", "R/R DLBCL", "Third Line
Plus"],
      "criteria": {
        "inclusion": [
          "Adults with Relapsed or refractory DLBCL NOS (including DLBCL
arising from follicular lymphoma)",
          "After two or more prior lines of systemic therapy",
          "Adequate organ function",
          "ECOG Performance Status 0-2"
          ],
        "exclusion": [
          "Previous severe reaction to selinexor",
          "Contraindication to required supportive medications (e.g.,
anti-emetics)",
          "Significant comorbidities precluding oral therapy or frequent
monitoring"
        ]
      },
      "registrationStudyCitation": "Kalakonda N, Maerevoet M, Cavallo F, et
al. Selinexor in patients with relapsed or refractory diffuse large B-cell
lymphoma (SADAL): a single-arm, multinational, multicentre, open-label,
phase 2 trial. Lancet Oncol. 2020;21(6):799-810.
doi:10.1016/S1470-2045(20)30229-4",
      "keyPapers": [
        {
          "citation": "Kalakonda N, et al. Lancet Oncol. 2020 (SADAL
Trial)",
          "reportedOutcomes": "ORR 29% (CR 13%) with 60 mg twice weekly
dose. Median DoR ~9.3 months. Durable responses seen in a subset.",
          "numberOfPatientsStudied": "134 (at 60mg dose)",
          "patientAndDiseaseCharacteristics": "R/R DLBCL, median 2 prior
lines (range 2-5). Included GCB and non-GCB subtypes.",
          "toxicities": "Common Grade 3/4: Thrombocytopenia (~49%),
neutropenia (~31%), anemia (~28%), fatigue (~14%), hyponatremia (~10%),
nausea (~6%). Frequent dose interruptions (~61%) and reductions (~49%)
needed."
        },
        {
            "citation_summary": "FDA Approval Summary / Oncology News
Central",
            "reportedOutcomes_summary": "Confirmed clinically meaningful
activity leading to accelerated approval. Highlighted need for close
monitoring and toxicity management.",
            "toxicities": "Detailed toxicity management guidelines provided,
including for nausea/vomiting, diarrhea, weight loss, hyponatremia, fatigue,
ocular toxicity."
        }
      ],
```

```
      "specialConsiderations": [
          "Oral agent providing a chemo-free option for heavily pretreated
patients.",
          "Requires proactive management of side effects, especially nausea,
fatigue, cytopenias, and hyponatremia.",
          "Thrombocytopenia is a common reason for dose
modification/discontinuation.",
          "Regular monitoring of blood counts, electrolytes, and body weight
is crucial.",
          "Anti-emetic prophylaxis is recommended.",
          "Potential for ocular toxicity requires monitoring."
          ],
      "routeOfAdministration": "Oral",
      "settingOfAdministration": "Outpatient",
      "lengthOfTreatment": "Typically 60 mg orally on Days 1 and 3 of each
week, continued until disease progression or unacceptable toxicity."
    },
    {
      "name": "Ibrutinib (Imbruvica) - CNS Indication",
      "applicableLine": "Relapsed/Refractory CNS Lymphoma (PCNSL or SCNSL)",
      "tags": ["Oral", "BTK inhibitor", "CNS involvement", "R/R PCNSL", "R/R
SCNSL", "Targeted therapy"],
      "criteria": {
        "inclusion": [
          "Relapsed or refractory Primary CNS Lymphoma (PCNSL) or Secondary
CNS Lymphoma (SCNSL)",
          "Adequate organ function (hepatic/renal adjustments may be
needed)",
          "ECOG Performance Status generally 0-2"
          ],
        "exclusion": [
          "Previous severe reaction to ibrutinib",
          "Requirement for strong CYP3A inhibitors/inducers (potential
interactions)",
          "Requirement for anticoagulation with warfarin (bleeding risk,
other anticoagulants require caution)",
          "Active bleeding"
        ]
      },
      "guideline_context": "Ibrutinib shows activity in R/R CNS lymphoma due
to CNS penetration and targeting of B-cell receptor pathway often active in
these lymphomas. Single-agent responses can be transient. Often used
off-label or in clinical trials, frequently in combination regimens (e.g.,
with immunotherapy or chemotherapy) to improve durability.",
      "keyPapers": [
        {
          "citation_summary": "Single-Agent R/R CNSL studies",
          "reportedOutcomes_summary": "ORR rates reported between 50-80%,
but responses often not durable (median DoR < 6 months)."
        },
```

```
        {
          "citation": "Nayab A, et al. Blood Adv. 2025 (Ibrutinib +
Nivolumab Phase 2)",
          "reportedOutcomes": "ORR 78%, CR 50%. Median PFS 6.5 mo, Median OS
21.0 mo.",
          "numberOfPatientsStudied": "18",
          "patientAndDiseaseCharacteristics": "R/R CNS Lymphoma.",
          "toxicities": "Generally well tolerated; Grade 3/4 AEs included
infection, mucositis, diarrhea, rash, cytopenias (rare)."
        },
        {
          "citation": "Lionakis MS, et al. Blood. 2019 (Ibrutinib +
HD-MTX/Rituximab Phase 1b)",
          "reportedOutcomes": "ORR 80%. Tolerable safety profile.",
          "numberOfPatientsStudied": "15",
          "patientAndDiseaseCharacteristics": "Recurrent/Refractory CNS
Lymphoma.",
          "toxicities": "Common Grade 1/2 anemia, LFT elevation,
thrombocytopenia. Few Grade 3/4 events reported."
        },
        {
            "citation_summary": "Meta-analyses",
            "reportedOutcomes_summary": "Pooled ORR ~68%, CR ~36% for
ibrutinib-based therapy in R/R CNSL. Confirmed higher activity in non-GCB
DLBCL subtypes."
        }
      ],
      "specialConsiderations": [
          "Oral agent with CNS penetration.",
          "Bleeding risk requires caution, especially with concomitant
antiplatelet/anticoagulant agents. Hold before/after procedures.",
          "Risk of atrial fibrillation (monitor ECG if symptomatic).",
          "Potential for infections, fatigue, diarrhea, rash, cytopenias.",
          "Drug interactions common via CYP3A pathway.",
          "Typically continued until disease progression or unacceptable
toxicity."
          ],
      "routeOfAdministration": "Oral",
      "settingOfAdministration": "Outpatient",
      "lengthOfTreatment": "Daily dosing (typically 560mg or 840mg in CNSL
trials) continued until progression or unacceptable toxicity."
    },
    {
      "name": "High-Dose Methotrexate (HD-MTX) + High-Dose Cytarabine
(ARA-C) - CNS Therapy",
      "applicableLine": "First-line, Consolidation, Salvage, Prophylaxis",
      "tags": ["Chemotherapy", "CNS penetration", "CNS involvement",
"PCNSL", "SCNSL treatment", "SCNSL prophylaxis", "High-risk DLBCL"],
      "criteria": {
        "inclusion": [
```

```
          "Newly diagnosed or Relapsed/Refractory Primary CNS Lymphoma
(PCNSL)",
          "Treatment of Secondary CNS Lymphoma (SCNSL)",
          "CNS prophylaxis in selected high-risk systemic DLBCL patients",
          "Adequate renal function (crucial for MTX clearance, typically
CrCl > 50-60 mL/min)",
          "Adequate hepatic function",
          "Adequate hematologic reserve",
          "ECOG Performance Status sufficient to tolerate intensive
chemotherapy (typically 0-2)"
          ],
        "exclusion": [
          "Severe renal impairment precluding HD-MTX administration",
          "Severe hepatic impairment",
          "Significant pleural effusion or ascites (third-spacing of MTX)",
          "Active significant infection",
          "Severe comorbidities precluding intensive chemotherapy"
        ]
      },
      "guideline_context": "HD-MTX is the cornerstone of PCNSL treatment.
The addition of HD-ARA-C significantly improves response rates and outcomes
compared to HD-MTX alone and is part of standard induction regimens (often
combined with rituximab +/- thiotepa, e.g., MATRix). Also used for treating
established SCNSL and, more controversially, for prophylaxis in high-risk
systemic DLBCL.",
      "keyPapers": [
        {
          "citation": "Ferreri AJM, et al. Lancet Oncol. 2009 (Randomized
Phase 2 MTX vs MTX+ARA-C)",
          "reportedOutcomes": "Addition of ARA-C improved CR rate (46% vs
18%) and ORR (69% vs 40%) in PCNSL.",
          "numberOfPatientsStudied": "79",
          "patientAndDiseaseCharacteristics": "Newly diagnosed PCNSL, age <=
75 years.",
          "toxicities": "Increased Grade 3/4 hematologic toxicity with
combined therapy (92% vs 15%). 4 toxic deaths overall (3 vs 1)."
        },
        {
           "citation_summary": "MATRix Trial (IELSG32)",
           "reportedOutcomes_summary": "Established
MTX+ARA-C+Thiotepa+Rituximab as a highly effective induction regimen for
PCNSL (CR 49%)."
        },
        {
          "citation_summary": "CNS Prophylaxis Studies (e.g., Wilson WH et
al. Blood Adv 2021; Orellana-Noia VM et al. Blood Adv 2020)",
          "reportedOutcomes_summary": "Evidence for prophylactic HD-MTX in
high-risk DLBCL is conflicting. Some studies suggest benefit, while others
show no reduction in CNS relapse but increased toxicity.",
```

```
          "toxicities": "HD-MTX prophylaxis associated with significant
toxicities (hematologic, renal, hepatic)."
        }
      ],
      "specialConsiderations": [
          "Requires specialized supportive care: intensive hydration, urine
alkalinization, leucovorin rescue based on MTX levels, and frequent
monitoring of renal function, liver function, and blood counts.",
          "Significant risk of myelosuppression, mucositis, nephrotoxicity,
hepatotoxicity, and neurotoxicity.",
          "HD-ARA-C component significantly increases hematologic toxicity
risk.",
          "Dosing and schedules vary (e.g., MTX commonly 3-8 g/m², ARA-C
commonly 2 g/m² per dose for 2-4 doses per cycle).",
          "Strict adherence to supportive care protocols is critical to
minimize toxicity."
          ],
      "routeOfAdministration": "Intravenous",
      "settingOfAdministration": "Inpatient hospitalization required",
      "lengthOfTreatment": "Typically involves multiple cycles (e.g., 2-6)
repeated every 2-3 weeks for induction/treatment, or fewer cycles for
prophylaxis."
    },
    {
      "name": "Rituximab-Bendamustine (BR)",
      "applicableLine": "Second-line or subsequent (primarily
transplant-ineligible/frail)",
      "tags": ["Chemotherapy", "Transplant ineligible", "R/R DLBCL (limited
role)", "Frail patients"],
      "criteria": {
        "inclusion": [
          "Relapsed or refractory DLBCL",
          "Often considered for patients ineligible for intensive salvage
chemotherapy or ASCT/CAR-T due to age, comorbidities, or performance
status",
          "Adequate hematologic parameters (or expectation of recovery with
support)",
          "Adequate renal and hepatic function (dose adjustments may be
needed)"
          ],
        "exclusion": [
          "Previous severe reaction to rituximab or bendamustine",
          "Active, uncontrolled infection",
          "Severe bone marrow compromise not expected to recover"
        ]
      },
      "guideline_context": "Well-established for R/R indolent NHL and MCL.
Not a preferred standard regimen for R/R DLBCL, where it shows modest
efficacy compared to other options (e.g., Pola-BR, Tafa-Len, CAR-T,
bispecifics). May be an option for selected transplant-ineligible or frail
```

patients with limited alternatives, but outcomes are generally poor. Was used as the comparator arm (inferior) in the Pola-BR pivotal trial.",
      "keyPapers": [
        {
          "citation_summary": "R/R DLBCL Retrospective/Phase II Studies",
          "reportedOutcomes_summary": "Variable response rates (ORR ~50% in some reports, but often lower). Median PFS generally short (~4-8 months). Median OS also limited (~7-10 months). Efficacy seems lower than in indolent lymphomas.",
          "toxicities": "Primary toxicity is myelosuppression (neutropenia, lymphopenia, thrombocytopenia). Risk of febrile neutropenia and infections. Fatigue, nausea are also common. Generally less neurotoxic than platinum regimens."
        },
        {
          "citation": "Storti S, et al. Br J Haematol. 2016 (Front-line Frail Elderly Phase II)",
          "reportedOutcomes": "High ORR (78%, CR 52%) but poor survival (median PFS 5.4 mo, OS 10.2 mo), suggesting limited durability in this population.",
          "numberOfPatientsStudied": "23",
          "patientAndDiseaseCharacteristics": "Untreated DLBCL, age >=65, poor candidates for R-CHOP. Median age 80, 52% ECOG >=2."
        },
        {
            "citation_summary": "GO29365 Trial (Pola-BR vs BR)",
            "reportedOutcomes_summary": "BR comparator arm showed CR 17.5%, median PFS 3.7 mo, median OS 4.7 mo in transplant-ineligible R/R DLBCL."
        }
      ],
      "specialConsiderations": [
          "Considered less intensive than R-CHOP or platinum-based salvage regimens.",
          "Significant risk of hematologic toxicity requires monitoring and potential G-CSF support.",
          "Bendamustine dose commonly 90 mg/m² or 120 mg/m² on Days 1, 2 of a 28-day cycle.",
          "Risk of opportunistic infections due to prolonged lymphopenia.",
          "Generally administered outpatient.",
          "Modest efficacy in unselected R/R DLBCL limits its use primarily to situations where more effective therapies are not feasible."
          ],
      "routeOfAdministration": "Intravenous",
      "settingOfAdministration": "Outpatient typically",
      "lengthOfTreatment": "Typically planned for up to 6 cycles (28-day cycles), but often discontinued earlier due to progression or toxicity."
    },
   {
    "name": "R-CHOP (Rituximab, Cyclophosphamide, Doxorubicin, Vincristine, Prednisone)",

```
    "applicableLine": "First-line",
    "tags": ["First-line", "Standard of Care", "DLBCL",
"Anthracycline-based"],
    "criteria": {
      "inclusion": [
        "Newly diagnosed DLBCL (or other aggressive B-cell lymphomas)",
        "Adequate cardiac function (for doxorubicin)",
        "Adequate organ function and performance status to tolerate standard
chemotherapy"
      ],
      "exclusion": [
        "Significant cardiac contraindication to doxorubicin (consider
R-CEOP/R-GCVP/R-COMP)",
        "Severe organ dysfunction precluding chemotherapy",
        "Hypersensitivity to components",
        "Specific high-risk subtypes where DA-EPOCH-R might be preferred
(e.g., PMBCL, potentially DHL/THL - though controversial)"
      ]
    },
    "guideline_context": "Long-standing standard first-line treatment for
DLBCL for over 20 years, curing over 60% of patients. Still considered a
standard option, especially for lower-risk disease, and serves as the
comparator arm in many clinical trials evaluating new first-line strategies.
Pola-R-CHP is now preferred by some guidelines for intermediate/high-risk
IPI based on POLARIX trial PFS benefit.",
    "keyPapers": [
      {
        "citation_summary": "GELA LNH-98.5 Trial (Coiffier B, NEJM 2002) /
MInT Trial (Pfreundschuh M, Lancet Oncol 2006)",
        "reportedOutcomes_summary": "Established superiority of R-CHOP over
CHOP alone in elderly and younger patients, respectively, significantly
improving response rates and survival.",
        "toxicities": "Common toxicities include myelosuppression
(neutropenia, febrile neutropenia, anemia, thrombocytopenia), nausea,
fatigue, alopecia, peripheral neuropathy (vincristine), potential
cardiotoxicity (doxorubicin). Requires G-CSF support."
      },
      {
        "citation_summary": "POLARIX Trial (Tilly H, NEJM 2022) [Comparator
Arm]",
        "reportedOutcomes_summary": "R-CHOP arm showed 2-year PFS of 70.2%
and 2-year OS of 88.6%.",
        "patientAndDiseaseCharacteristics": "Patients aged 18-80 with
previously untreated DLBCL, IPI 2-5.",
        "toxicities": "Grade 3/4 neutropenia ~31%, febrile neutropenia ~8%,
anemia ~8%."
      }
    ],
    "specialConsiderations": [
```

```
        "Administered typically every 21 days (R-CHOP-21) or sometimes every
14 days (R-CHOP-14 - less common now).",
        "Requires monitoring of blood counts, cardiac function (esp. LVEF if
cumulative doxorubicin dose is high or risk factors exist).",
        "Vincristine dose often capped at 2mg.",
        "Despite high cure rates, ~30-40% experience primary refractory
disease or relapse."
        ],
    "routeOfAdministration": "Rituximab, Cyclophosphamide, Doxorubicin,
Vincristine: IV; Prednisone: Oral",
    "settingOfAdministration": "Outpatient typically",
    "lengthOfTreatment": "Typically 6 cycles (every 21 days). Sometimes 8
cycles or fewer cycles combined with radiotherapy for early-stage disease."
  },
  {
    "name": "Pola-R-CHP (Polatuzumab Vedotin, Rituximab, Cyclophosphamide,
Doxorubicin, Prednisone)",
    "applicableLine": "First-line",
    "tags": ["First-line", "Standard of Care", "DLBCL", "ADC",
"Anthracycline-based", "Intermediate/High Risk IPI"],
    "criteria": {
      "inclusion": [
        "Newly diagnosed DLBCL (typically IPI 2-5 based on POLARIX trial)",
        "Adequate cardiac function",
        "Adequate organ function and performance status",
        "Age 18-80 (as studied in POLARIX)"
      ],
      "exclusion": [
        "Significant cardiac contraindication to doxorubicin",
        "Significant pre-existing peripheral neuropathy (Grade >1)",
        "Hypersensitivity to components"
      ]
    },
    "guideline_context": "A new standard first-line option for DLBCL,
particularly considered for patients with intermediate-to-high risk IPI
(2-5). Approved based on improved PFS vs R-CHOP in the POLARIX trial.
Vincristine is omitted compared to R-CHOP to reduce risk of cumulative
neurotoxicity.",
    "registrationStudyCitation": "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. doi:10.1056/NEJMoa2115304 (POLARIX Trial)",
    "keyPapers": [
      {
        "citation": "Tilly H, et al. N Engl J Med. 2022 (POLARIX Trial)",
        "reportedOutcomes": "Superior PFS vs R-CHOP (2-yr PFS 76.7% vs
70.2%; HR 0.73, p=0.02). No significant difference in OS at median follow-up
of 28.2 months (2-yr OS 88.7% vs 88.6%).",
        "numberOfPatientsStudied": "879 randomized (440 Pola-R-CHP, 439
R-CHOP)",
```

```
        "patientAndDiseaseCharacteristics": "Previously untreated DLBCL, age
18-80, IPI 2-5.",
        "toxicities": "Safety profile similar to R-CHOP overall. Most common
Grade 3-4 AEs: neutropenia (~28%), febrile neutropenia (~14%), anemia
(~12%). Peripheral neuropathy (any grade) ~53% (vs ~54% with R-CHOP, but
vincristine omitted in Pola-R-CHP arm)."
      },
      {
        "citation_summary": "Real-World Studies (e.g., PubMed ID:
39641321)",
        "reportedOutcomes_summary": "Emerging RWD supports efficacy, though
magnitude of benefit vs R-CHOP may vary. 12-month PFS numerically higher
(90% vs 84%, p=0.18) in one large Chinese RWD study.",
        "toxicities": "Comparable safety profiles in RWD, no new signals
identified."
      }
    ],
    "specialConsiderations": [
        "Polatuzumab vedotin replaces vincristine.",
        "Requires monitoring for cytopenias, peripheral neuropathy, infusion
reactions, LFTs.",
        "G-CSF support typically required.",
        "Considered particularly beneficial for higher-risk IPI groups,
patients >60, non-GCB subtype based on exploratory analyses."
        ],
    "routeOfAdministration": "Polatuzumab Vedotin, Rituximab,
Cyclophosphamide, Doxorubicin: IV; Prednisone: Oral",
    "settingOfAdministration": "Outpatient typically",
    "lengthOfTreatment": "6 cycles (every 21 days), often followed by 2
cycles of Rituximab monotherapy (as per POLARIX trial design)."
  },
  {
    "name": "R-miniCHOP (Rituximab, reduced-dose Cyclophosphamide,
Doxorubicin, Vincristine, Prednisone)",
    "applicableLine": "First-line",
    "tags": ["First-line", "Elderly", "Frail", "Dose-reduced", "DLBCL"],
    "criteria": {
      "inclusion": [
        "Newly diagnosed DLBCL in elderly patients (typically age > 80
years)",
        "Patients considered unfit for full-dose R-CHOP due to age or
frailty, but still candidates for treatment.",
        "Adequate cardiac function (for reduced-dose doxorubicin)"
      ],
      "exclusion": [
        "Patients fit enough to tolerate full-dose R-CHOP or Pola-R-CHP.",
        "Very frail patients unsuitable for any chemotherapy.",
        "Significant cardiac contraindication to anthracyclines (consider
R-miniCEOP)."
      ]
```

```
    },
    "guideline_context": "Standard first-line treatment option recommended
by guidelines (e.g., NCCN) for patients >80 years old without severe
comorbidities who are not candidates for full-dose therapy. Aims to balance
efficacy with tolerability in a vulnerable population.",
    "keyPapers": [
      {
        "citation": "Peyrade F, Jardin F, Thieblemont C, et al. Attenuated
R-CHOP versus intensive R-ACVBP in elderly patients with diffuse large
B-cell lymphoma: results of the LNH03-6B GELA study. J Clin Oncol.
2011;29(suppl_15):8002.",
        "reportedOutcomes_summary": "This study (LNH03-6B) directly led to
the R-miniCHOP regimen. Reported 2-year PFS 47% and 2-year OS 59% for
R-miniCHOP.",
        "numberOfPatientsStudied": "149 in R-miniCHOP arm",
        "patientAndDiseaseCharacteristics": "Patients aged 80-95 years with
untreated DLBCL.",
        "toxicities": "Better tolerated than full dose R-CHOP but still
associated with myelosuppression and infection risk. Dose reductions across
all agents compared to standard R-CHOP."
      },
       {
        "citation_summary": "Studies Combining R-miniCHOP with Novel Agents
(e.g., Pola-R-miniCHP, R-miniCHOP + Lenalidomide)",
        "reportedOutcomes_summary": "Investigational approaches trying to
improve outcomes in this population. Pola-R-miniCHP feasibility shown in
small study. R-miniCHOP + Lenalidomide showed no benefit vs R-miniCHOP
alone."
      }
    ],
    "specialConsiderations": [
        "Uses reduced doses of CHOP components compared to standard
R-CHOP.",
        "Requires monitoring of blood counts, potential toxicities (cardiac,
neuro), and G-CSF support.",
        "Focus on maintaining quality of life while providing anti-lymphoma
therapy.",
        "Outcomes are inferior to full-dose R-CHOP in younger/fitter
patients."
        ],
    "routeOfAdministration": "Rituximab, Cyclophosphamide, Doxorubicin,
Vincristine: IV; Prednisone: Oral (all at reduced doses)",
    "settingOfAdministration": "Outpatient typically",
    "lengthOfTreatment": "Typically 6 cycles (every 21 days)."
  },
  {
    "name": "R-COMP (Rituximab, Cyclophosphamide, Non-pegylated Liposomal
Doxorubicin, Vincristine, Prednisone)",
    "applicableLine": "First-line",
```

```
    "tags": ["First-line", "DLBCL", "Cardiac Risk", "Anthracycline-based",
"Liposomal Doxorubicin"],
    "criteria": {
      "inclusion": [
        "Newly diagnosed DLBCL",
        "Patients with pre-existing cardiac risk factors or concerns about
conventional doxorubicin cardiotoxicity",
        "Adequate organ function and performance status"
      ],
      "exclusion": [
        "Patients suitable for standard R-CHOP or Pola-R-CHP without
specific cardiac concerns",
        "Hypersensitivity to components (including liposomal formulation)"
      ]
    },
    "guideline_context": "An alternative to R-CHOP where non-pegylated
liposomal doxorubicin (NPLD, Myocet®) replaces conventional doxorubicin,
primarily aimed at reducing cardiotoxicity. Not a standard first-line choice
for unselected patients, but may be considered in those with significant
cardiac risk where anthracycline benefit is still desired. Efficacy appears
similar to R-CHOP.",
    "keyPapers": [
      {
        "citation": "Martín García-Sancho A, et al. R-COMP versus R-CHOP as
first-line therapy for diffuse large B-cell lymphoma in patients ≥60 years:
Results of a randomized phase 2 study from the Spanish GELTAMO group. Cancer
Med. 2021;10(6):1948-1957. doi:10.1002/cam4.3785",
        "reportedOutcomes": "No significant difference in efficacy (ORR,
PFS, OS) vs R-CHOP. No significant difference in primary endpoint (LVEF
decline <55%). Higher troponin elevations seen with R-CHOP.",
        "numberOfPatientsStudied": "90 randomized (45 R-COMP, 45 R-CHOP)",
        "patientAndDiseaseCharacteristics": "DLBCL patients ≥60 years, LVEF
≥55%.",
        "toxicities": "Similar rates of Grade 3/4 hematologic toxicity.
Fewer grade >=3 cardiovascular AEs with R-COMP (0 vs 4), though not
statistically significant in this size study."
      },
      {
        "citation_summary": "Italian (FIL) Cohort Study (vs R-CHOP in older
patients)",
        "reportedOutcomes_summary": "Confirmed similar efficacy (3-year OS
~70%) to historical R-CHOP data. Low rate of severe cardiotoxicity (Grade
3-4 in 5%).",
        "patientAndDiseaseCharacteristics": "Older patients (median age 74)
with DLBCL."
      },
      {
         "citation_summary": "Network Meta-analysis (Frontiers Immunology
2022)",
```

```
        "reportedOutcomes_summary": "Suggested R-COMP associated with lower
rates of Grade 3/4 neutropenia and cardiac events compared to other regimens
in elderly patients."
      }
    ],
    "specialConsiderations": [
        "NPLD (Myocet®) replaces conventional doxorubicin.",
        "May have a slightly different infusion reaction profile
(liposomal).",
        "Monitoring similar to R-CHOP (counts, LVEF), but potentially less
intensive cardiac monitoring needed depending on baseline risk.",
        "Considered a niche regimen for patients with specific cardiac
concerns."
        ],
    "routeOfAdministration": "Rituximab, Cyclophosphamide, NPLD,
Vincristine: IV; Prednisone: Oral",
    "settingOfAdministration": "Outpatient typically",
    "lengthOfTreatment": "Typically 6 cycles (every 21 days)."
  },
  {
    "name": "DA-EPOCH-R (Dose-Adjusted Etoposide, Prednisone, Vincristine,
Cyclophosphamide, Doxorubicin, Rituximab)",
    "applicableLine": "First-line",
    "tags": ["First-line", "High-Risk DLBCL", "DHL/THL", "PMBCL", "Intensive
Chemotherapy", "Infusional"],
    "criteria": {
      "inclusion": [
        "Newly diagnosed high-risk aggressive B-cell lymphomas, particularly
considered for:",
        "  - High-Grade B-Cell Lymphoma with MYC and BCL2 and/or BCL6
rearrangements (Double/Triple Hit Lymphoma, DHL/THL)",
        "  - Primary Mediastinal B-cell Lymphoma (PMBCL)",
        "  - Potentially HIV-associated DLBCL, Germinal Center B-cell Like
(GCB) DLBCL with high IPI (controversial)",
        "Fit patients (ECOG 0-2 typically) able to tolerate intensive,
infusional chemotherapy",
        "Adequate organ function (esp. cardiac, bone marrow)"
      ],
      "exclusion": [
        "Standard-risk DLBCL where R-CHOP or Pola-R-CHP are preferred",
        "Patients unfit for intensive chemotherapy due to age,
comorbidities, or performance status",
        "Significant cardiac contraindications"
      ]
    },
    "guideline_context": "More intensive, infusional regimen than R-CHOP. A
phase 3 trial (CALGB 50303/Alliance) showed no overall benefit compared to
R-CHOP in unselected newly diagnosed DLBCL. However, based on phase 2 data
and retrospective series suggesting high efficacy, it remains a
preferred/recommended option in guidelines for specific high-risk subtypes
```

```
like DHL/THL and PMBCL, where R-CHOP outcomes are historically poor. Its use in other high-risk groups (e.g., high IPI GCB or ABC DLBCL) is more controversial.",
    "keyPapers": [
      {
        "citation_summary": "NCI Phase II studies (Wilson WH et al.)",
        "reportedOutcomes_summary": "Demonstrated high efficacy (e.g., high CR rates, good survival) in specific subtypes like PMBCL and DHL/THL, forming the basis for its use in these settings.",
        "toxicities": "Significant hematologic toxicity (neutropenia, thrombocytopenia requiring dose adjustments), high rates of febrile neutropenia, mucositis, neuropathy. Requires G-CSF."
      },
      {
        "citation_summary": "CALGB 50303 / Alliance Intergroup Trial (Bartlett NL, et al. J Clin Oncol. 2019 / Wilson WH et al. NEJM 2022 - sub-analyses)",
        "reportedOutcomes_summary": "Randomized Phase 3 vs R-CHOP in untreated DLBCL. No significant difference in PFS or OS overall or in key subgroups including ABC/GCB subtypes. Suggested potential benefit limited to specific molecular groups not easily identified clinically.",
        "numberOfPatientsStudied": "~524",
        "patientAndDiseaseCharacteristics": "Newly diagnosed DLBCL.",
        "toxicities": "Confirmed higher rates of Grade 3/4 hematologic toxicity, febrile neutropenia, and infection with DA-EPOCH-R compared to R-CHOP."
      },
       {
        "citation_summary": "Retrospective/Real-World Series (e.g., Via Medica review)",
        "reportedOutcomes_summary": "Often suggest improved PFS (~15-20%) for DA-EPOCH-R vs R-CHOP specifically in high-risk fit patients (e.g., aaIPI 2-3, older but fit), though acknowledging higher toxicity.",
        "toxicities": "Higher rates of hematologic toxicity, infections, mucositis, neuropathy compared to R-CHOP."
       }
    ],
    "specialConsiderations": [
        "Requires administration via continuous intravenous infusion over 96 hours.",
        "Doses of etoposide, doxorubicin, and cyclophosphamide are adjusted in subsequent cycles based on neutrophil and platelet nadirs from the previous cycle.",
        "Requires intensive supportive care, including mandatory G-CSF, infection prophylaxis (e.g., PJP, antiviral, antifungal often considered), and close monitoring.",
        "Higher toxicity profile than R-CHOP/Pola-R-CHP requires careful patient selection (fit patients).",
        "Typically requires inpatient administration or specialized outpatient infusion capabilities."
```

```
        ],
    "routeOfAdministration": "Etoposide, Doxorubicin, Vincristine:
Continuous IV infusion over 96 hours; Cyclophosphamide: IV bolus;
Prednisone: Oral; Rituximab: IV infusion",
    "settingOfAdministration": "Typically requires inpatient setting due to
continuous infusion and monitoring; experienced centers may manage
outpatient.",
    "lengthOfTreatment": "Typically 6 cycles (every 21 days)."
  }
  ]
}
```

GUIDELINES.JSON


```
"treatmentGuidelines": [
  {
  "scenarioId": "1L_DLBCL_Treatment",
  "description": "Guideline recommendations for the initial treatment
of newly diagnosed Diffuse Large B-Cell Lymphoma (DLBCL), stratified by
risk, patient fitness, and specific disease subtypes.",
  "tags": ["First-line", "Newly Diagnosed DLBCL", "Standard of Care",
"Risk Stratification", "AIFA policies"],
  "applicableLines": ["First-line"],
  "patientProfile": ["Variable", "Fit", "Elderly", "Frail", "Cardiac
Risk"], // Indicates stratification within scenario
  "priorContext": ["Treatment Naive"],
  "treatments": [
    {
      "name": "R-CHOP (Rituximab, Cyclophosphamide, Doxorubicin,
Vincristine, Prednisone)",
      "criteria": "Standard option for newly diagnosed DLBCL,
especially suitable for lower-risk IPI scores or when
Pola-R-CHP/DA-EPOCH-R are not indicated/available.",
      "rationale": "Long-standing standard of care with curative
potential in a majority of patients.",
      "evidenceLevel": "Level 1",
      "comparisonNote": "Comparator arm for newer regimens. Pola-R-CHP
showed superior PFS in IPI 2-5 patients. DA-EPOCH-R showed no overall
benefit vs R-CHOP in unselected DLBCL but may be preferred for specific
high-risk subtypes.",
      "sequencePosition": "Initial Therapy"
    },
    {
      "name": "Pola-R-CHP (Polatuzumab Vedotin, Rituximab,
Cyclophosphamide, Doxorubicin, Prednisone)",
      "criteria": "Preferred standard option for newly diagnosed DLBCL
with intermediate-to-high risk (IPI 2-5), age 18-80, adequate organ
function, no significant baseline neuropathy.",
      "rationale": "Demonstrated statistically significant improvement
in PFS compared to R-CHOP in the POLARIX trial.",
      "evidenceLevel": "Level 1 (POLARIX)",
      "comparisonNote": "Superior PFS vs R-CHOP (HR 0.73), particularly
noted in subgroups >60yrs, non-GCB, high IPI. No OS difference yet.
Similar safety profile to R-CHOP.",
      "sequencePosition": "Initial Therapy (Preferred for IPI 2-5)"
    },
    {
      "name": "R-miniCHOP (Rituximab, reduced-dose Cyclophosphamide,
Doxorubicin, Vincristine, Prednisone)",
```

```
      "criteria": "Standard option for elderly patients (typically age
> 80 years) considered unfit for full-dose chemoimmunotherapy due to
age/frailty.",
      "rationale": "Balances efficacy and tolerability in a vulnerable
population unable to receive full-dose therapy.",
      "evidenceLevel": "Level 2 (LNH03-6B)",
      "comparisonNote": "Reduced intensity and efficacy compared to
R-CHOP/Pola-R-CHP, but better tolerated in the very elderly/frail.",
      "sequencePosition": "Initial Therapy (Elderly/Frail)"
    },
    {
      "name": "R-COMP (Rituximab, Cyclophosphamide, Non-pegylated
Liposomal Doxorubicin, Vincristine, Prednisone)",
      "criteria": "Alternative option for patients with significant
pre-existing cardiac risk factors where conventional doxorubicin is a
concern, but an anthracycline-containing regimen is desired.",
      "rationale": "Aims to reduce cardiotoxicity risk by using
liposomal doxorubicin.",
      "evidenceLevel": "Level 2/3",
      "comparisonNote": "Efficacy appears similar to R-CHOP. Evidence
for reduced clinical cardiotoxicity vs R-CHOP is mixed/limited in
randomized trials, though some safety signals favored R-COMP. Not
standard for unselected patients.",
      "sequencePosition": "Initial Therapy (Cardiac Risk)"
    },
    {
      "name": "DA-EPOCH-R (Dose-Adjusted Etoposide, Prednisone,
Vincristine, Cyclophosphamide, Doxorubicin, Rituximab)",
      "criteria": "Recommended/preferred option for specific high-risk
subtypes: Double/Triple Hit Lymphoma (DHL/THL) and Primary Mediastinal
B-cell Lymphoma (PMBCL). Requires high fitness level (ECOG 0-1
typically) due to intensity.",
      "rationale": "More intensive regimen aiming to overcome poor
prognosis associated with specific subtypes based on Phase
2/retrospective data. Requires dose adjustments based on blood
counts.",
      "evidenceLevel": "Level 2/3 (for specific subtypes)",
      "comparisonNote": "No proven benefit over R-CHOP in unselected
DLBCL (CALGB 50303). Significantly higher hematologic and infectious
toxicity than R-CHOP/Pola-R-CHP.",
      "sequencePosition": "Initial Therapy (Specific High-Risk
Subtypes)"
    }
  ]
},
  {
    "scenarioId": "2L_Fit_EarlyRelapse_PrimaryRefractory",
```

```
    "description": "Second-line therapy for fit patients (age ≤75, ECOG
0-1, transplant/CAR-T eligible) with early relapse (<12 months from 1L
chemoimmunotherapy) or primary refractory disease.",
    "tags": ["Second-line", "Fit", "Early Relapse", "Primary
Refractory", "CAR-T eligible", "AIFA policies"],
    "applicableLines": ["Second-line"],
    "patientProfile": ["Fit", "Age <= 75", "ECOG 0-1"],
    "priorContext": ["Relapse < 12mo", "Primary Refractory"],
    "treatments": [
      {
        "name": "Axicabtagene ciloleucel (Yescarta)",
        "criteria": "Meets AIFA criteria for 2L use in this high-risk
setting.",
        "rationale": "Preferred option based on demonstrated
superiority over standard of care (SOC: salvage chemo + ASCT) in this
setting.",
        "evidenceLevel": "Level 1 (ZUMA-7)",
        "comparisonNote": "Superior EFS/OS vs SOC. Higher rates of
severe CRS/ICANS compared to Liso-cel observed in trials.",
        "sequencePosition": "Definitive Therapy"
      },
        {
        "name": "Alternative: Salvage Chemo + ASCT", // Added as
alternative if CAR-T not feasible
        "criteria": "Consider if CAR-T is not feasible (e.g., rapid
progression, comorbidities precluding CAR-T, access issues) AND patient
achieves chemosensitivity (CR/PR) to salvage.",
        "rationale": "Represents the standard of care comparator arm in
ZUMA-7/TRANSFORM; inferior efficacy but may be only option for some.",
        "evidenceLevel": "Level 1 (as comparator)",
        "comparisonNote": "Inferior EFS/OS compared to Axi-cel and
Liso-cel in this specific patient population.",
        "sequencePosition": "Salvage followed by Consolidation"
         // Link to specific salvage/ASCT entries if needed, or keep as
strategy note.
      }
    ]
  },
  {
    "scenarioId": "2L_Fit_LateRelapse",
    "description": "Second-line therapy for fit patients (transplant
eligible, typically age <= 65-70, ECOG 0-1) with later relapse (>=12
months from 1L chemoimmunotherapy).",
    "tags": ["Second-line", "Fit", "Late Relapse", "Transplant
eligible", "Chemosensitive", "AIFA policies"],
    "applicableLines": ["Second-line"],
    "patientProfile": ["Fit", "Age <= 65-70", "ECOG 0-1"],
    "priorContext": ["Relapse >= 12mo", "Chemosensitive to Salvage"],
```

```
    "treatments": [
      // Salvage Options (Goal: CR/PR for ASCT)
      {
        "name": "R-ICE (Rituximab, Ifosfamide, Carboplatin,
Etoposide)",
        "criteria": "Standard salvage option.",
        "rationale": "Assess chemosensitivity prior to ASCT.",
        "evidenceLevel": "Level 1/2",
        "comparisonNote": "Efficacy comparable to R-DHAP in CORAL
trial; different toxicity profile (less renal/thrombo).",
        "sequencePosition": "Salvage prior to Consolidation"
      },
      {
        "name": "R-DHAP (Rituximab, Dexamethasone, High-Dose
Cytarabine, Cisplatin)",
        "criteria": "Standard salvage option.",
        "rationale": "Assess chemosensitivity prior to ASCT.",
        "evidenceLevel": "Level 1/2",
        "comparisonNote": "Efficacy comparable to R-ICE in CORAL trial;
higher risk of renal toxicity & thrombocytopenia.",
        "sequencePosition": "Salvage prior to Consolidation"
      },
      {
        "name": "R-DHAOx (Rituximab, Dexamethasone, High-Dose
Cytarabine, Oxaliplatin)",
        "criteria": "Standard salvage option, preferred if
renal/ototoxicity concerns.",
        "rationale": "Assess chemosensitivity prior to ASCT.",
        "evidenceLevel": "Level 2",
        "comparisonNote": "Comparable efficacy to R-DHAP suggested,
with significantly less nephrotoxicity.",
        "sequencePosition": "Salvage prior to Consolidation"
      },
      {
       "name": "R-IEV (Rituximab, Ifosfamide, Epirubicin, Etoposide)",
       "criteria": "Potential consideration as salvage chemo if patient
is fit and ASCT consolidation is planned; component drugs may have some
CNS activity.",
       "rationale": "Includes agents with potential CNS penetration
(Ifosfamide, Etoposide); Has similar agents to R-CHOP, so sensibility
to CHOP-chemotherapy could weight favourably for R-IEV use",
       "evidenceLevel": "Level 2-3",
       "comparisonNote": "Less established for CNS relapse compared to
HD-MTX based regimens.",
       "sequencePosition": "Systemic Salvage (Consideration)"
     },
      // Consolidation Step
      {
```

```
        "name": "ASCT (BEAM Conditioning)",
        "criteria": "MUST achieve CR or PR to salvage chemotherapy
listed above.",
        "rationale": "Standard consolidation offering curative
potential in chemosensitive late relapse.",
        "evidenceLevel": "Level 1",
        "sequencePosition": "Consolidation after Salvage"
      }
    ]
  },
  {
    "scenarioId": "2L_TransplantIneligible",
    "description": "Second-line therapy for patients ineligible for
ASCT/CAR-T (due to age > 70-75, comorbidities, fitness).",
    "tags": ["Second-line", "Transplant ineligible", "Unfit patients",
"AIFA policies"],
    "applicableLines": ["Second-line"],
    "patientProfile": ["Transplant Ineligible", "Unfit (potentially)",
"Age > 70-75 (often)"],
    "priorContext": ["Relapsed/Refractory after >=1 line"], // Updated
prior context
    "treatments": [
      {
        "name": "R-Pola-Benda (Polatuzumab Vedotin + Bendamustine +
Rituximab)",
        "criteria": "Meets AIFA criteria for ASCT-ineligible R/R DLBCL
after >=1 prior line.",
        "rationale": "Demonstrated survival benefit over BR alone.",
        "evidenceLevel": "Level 1 (GO29365)",
        "comparisonNote": "Superior PFS/OS vs BR alone. Note neuropathy
risk.",
        "sequencePosition": "Definitive Therapy"
      },
      {
        "name": "Tafasitamab-Lenalidomide",
        "criteria": "Meets AIFA criteria for ASCT-ineligible R/R DLBCL
after >=1 prior line.",
        "rationale": "Durable responses shown in L-MIND.",
        "evidenceLevel": "Level 2 (L-MIND)",
        "comparisonNote": "RWD comparison (RE-MIND2) suggested OS
benefit vs Pola-BR and R2. Generally well-tolerated.",
        "sequencePosition": "Definitive Therapy"
      },
       {
        "name": "R-GEMOx (Rituximab, Gemcitabine, Oxaliplatin)",
        "criteria": "Alternative option, less intensive.",
        "rationale": "Balance of efficacy/tolerability. Modest
long-term outcomes.",
```

```
        "evidenceLevel": "Level 2/3",
        "comparisonNote": "Less intensive than Pola-BR. Generally less
effective but may be better tolerated.",
        "sequencePosition": "Definitive Therapy / Bridge"
      }
    ]
  },
  {
    "scenarioId": "3Lplus_PostChemoASCT",
    "description": "Third-line or subsequent therapy for patients
failing conventional chemotherapy +/- ASCT (or deemed ineligible for
ASCT after 2L chemo).",
    "tags": ["Third-line or subsequent", "Post-Chemo Failure",
"Post-ASCT Failure", "AIFA policies"],
    "applicableLines": ["Third-line", "Subsequent"],
    "patientProfile": ["Variable Fitness", "Age <= 75 (for CAR-T)"],
    "priorContext": ["Failed >=2 lines chemo/immuno", "Relapsed
post-ASCT", "Not primary CAR-T failure"],
    "treatments": [
      // CAR-T Options (if eligible and not used 2L)
       {
        "name": "Axicabtagene ciloleucel (Yescarta)",
        "criteria": "Meets AIFA criteria for 3L+ use (Age <= 75), not
received CAR-T previously.",
        "rationale": "Potential for durable remission/cure.",
        "evidenceLevel": "Level 2 (ZUMA-1)",
        "comparisonNote": "High ORR/CR rates. Note higher rates of
severe CRS/ICANS vs Liso-cel in trials.",
        "sequencePosition": "Definitive Therapy"
      },
      {
        "name": "Lisocabtagene maraleucel (Breyanzi)",
        "criteria": "Meets AIFA criteria for 3L+ use, not received
CAR-T previously.",
        "rationale": "Potential for durable remission. Favorable
toxicity profile.",
        "evidenceLevel": "Level 2 (TRANSCEND NHL 001)",
        "comparisonNote": "Good efficacy with lower rates of severe
CRS/ICANS.",
        "sequencePosition": "Definitive Therapy"
      },
      {
        "name": "Tisagenlecleucel (Kymriah)",
        "criteria": "Meets AIFA criteria for 3L+ use, not received
CAR-T previously.",
        "rationale": "Potential for durable remission.",
        "evidenceLevel": "Level 2 (JULIET)",
```

```
        "comparisonNote": "Efficacy established; RWD suggests favorable
safety.",
        "sequencePosition": "Definitive Therapy"
      },
      // Bispecific Options
      {
        "name": "Glofitamab (Columvi)",
        "criteria": "Failed >=2 lines. Meets AIFA criteria.",
        "rationale": "Fixed-duration bispecific option. Effective
post-chemo.",
        "evidenceLevel": "Level 2",
        "comparisonNote": "Fixed duration (12 cycles). Requires
Obinutuzumab pre-treatment.",
        "sequencePosition": "Definitive Therapy"
      },
      {
        "name": "Epcoritamab (Epkinly)",
        "criteria": "Failed >=2 lines. Meets AIFA criteria.",
        "rationale": "Subcutaneous bispecific option. Effective
post-chemo.",
        "evidenceLevel": "Level 2",
        "comparisonNote": "Subcutaneous administration, continuous
therapy.",
        "sequencePosition": "Definitive Therapy"
      },
      // Other Approved Options
      {
        "name": "Loncastuximab tesirine (Zynlonta)",
        "criteria": "Failed >=2 lines. Meets AIFA criteria.",
        "rationale": "CD19 ADC option.",
        "evidenceLevel": "Level 2",
        "sequencePosition": "Definitive Therapy"
      },
      // Consider Transplant-Ineligible options if not used 2L and
patient still ineligible
      {
        "name": "R-Pola-Benda (Polatuzumab Vedotin + Bendamustine +
Rituximab)",
        "criteria": "If ASCT/CAR-T ineligible and not used in 2L.",
        "rationale": "Established efficacy in transplant-ineligible
setting.",
        "evidenceLevel": "Level 1 (for ASCT-ineligible)",
        "sequencePosition": "Definitive Therapy"
      },
      {
        "name": "Tafasitamab-Lenalidomide",
        "criteria": "If ASCT/CAR-T ineligible and not used in 2L.",
```

```
        "rationale": "Established efficacy in transplant-ineligible
setting.",
        "evidenceLevel": "Level 2 (for ASCT-ineligible)",
        "sequencePosition": "Definitive Therapy"
      },
      // Adding ASCT as a possibility in 3L+ if specific conditions met
      {
        "name": "ASCT (BEAM Conditioning)",
        "criteria": "RARELY considered in 3L+. Requires exceptional
chemosensitivity (CR/PR) to a 3L+ salvage regimen AND persistent
transplant eligibility (fitness, age, organ function).",
        "rationale": "Potential for consolidation if deep response
achieved late and patient remains highly fit.",
        "evidenceLevel": "Level 3 / Expert Opinion",
        "sequencePosition": "Consolidation after 3L+ Salvage
(Exceptional Cases)"
      }
    ]
  },
   {
    "scenarioId": "3Lplus_PostCART",
    "description": "Third-line or subsequent therapy for patients
failing CAR T-cell therapy.",
    "tags": ["Third-line or subsequent", "Post-CAR T Failure", "AIFA
policies"],
    "applicableLines": ["Third-line", "Subsequent"],
    "patientProfile": ["Variable Fitness"],
    "priorContext": ["Failed CAR-T Therapy"],
    "treatments": [
      // Bispecifics are primary options
      {
        "name": "Glofitamab (Columvi)",
        "criteria": "Meets AIFA criteria. Effective post-CAR T.",
        "rationale": "Demonstrated efficacy after CAR T failure. Fixed
duration.",
        "evidenceLevel": "Level 2",
        "comparisonNote": "Preferred option post-CAR T.",
        "sequencePosition": "Definitive Therapy"
      },
      {
        "name": "Epcoritamab (Epkinly)",
        "criteria": "Meets AIFA criteria. Effective post-CAR T.",
        "rationale": "Demonstrated efficacy after CAR T failure.
Subcutaneous.",
        "evidenceLevel": "Level 2",
        "comparisonNote": "Preferred option post-CAR T.",
        "sequencePosition": "Definitive Therapy"
      },
```

```
       // Other options with post-CAR T data
       {
        "name": "Loncastuximab tesirine (Zynlonta)",
        "criteria": "Meets AIFA criteria. Effective post-CAR T.",
        "rationale": "Demonstrated efficacy after CAR T failure
(RWD).",
        "evidenceLevel": "Level 2/3",
        "comparisonNote": "Considered if bispecifics are
contraindicated or fail.",
        "sequencePosition": "Definitive Therapy"
      },
      // Less evidence post-CAR T, consider case-by-case/trials
       {
        "name": "Tafasitamab-Lenalidomide",
        "criteria": "Consider case-by-case, limited data post-CAR T.",
        "rationale": "Mechanism potentially independent of CAR T
failure.",
        "evidenceLevel": "Level 3",
        "sequencePosition": "Salvage / Palliative"
      },
       {
        "name": "R-Pola-Benda (Polatuzumab Vedotin + Bendamustine +
Rituximab)",
        "criteria": "Consider case-by-case, limited data post-CAR T.",
        "rationale": "Mechanism potentially independent of CAR T
failure.",
        "evidenceLevel": "Level 3",
        "sequencePosition": "Salvage / Palliative"
       },
       {
         "name": "Clinical Trial", // Explicitly mention trials
         "criteria": "Strongly consider enrollment in clinical trials
for post-CAR T failure.",
         "rationale": "Optimal strategy undefined; novel
agents/combinations needed.",
         "evidenceLevel": "N/A",
         "sequencePosition": "Optimal next step if available"
       }
    ]
   },
   {
    "scenarioId": "Palliative_Frail",
    "description": "Management for elderly or unfit patients not
suitable for aggressive therapies, focusing on quality of life and
symptom control.",
    "tags": ["Palliative", "Elderly patients", "Unfit patients", "AIFA
policies"],
    "applicableLines": ["Variable", "Often Later Lines"],
```

```
    "patientProfile": ["Frail", "Unfit", "Poor ECOG (>=2)",
"Significant Comorbidities"],
    "priorContext": ["Multiple prior lines", "Not candidate for
intensive therapy"],
    "treatments": [
      {
        "name": "R-GEMOx (Rituximab, Gemcitabine, Oxaliplatin)",
        "criteria": "Less intensive chemo option.",
        "rationale": "Balance of efficacy/tolerability.",
        "evidenceLevel": "Level 2/3",
        "comparisonNote": "Generally better tolerated than BR.",
        "sequencePosition": "Palliative Systemic Therapy"
      },
       {
         "name": "Rituximab-Bendamustine (BR)",
         "criteria": "Less intensive chemo option, consider toxicity
profile.",
         "rationale": "May be tolerated in some frail patients; modest
efficacy.",
         "evidenceLevel": "Level 3 (Limited benefit shown in frail
DLBCL)",
         "comparisonNote": "Higher myelosuppression risk vs R-GemOx
likely.",
         "sequencePosition": "Palliative Systemic Therapy"
       },
      {
        "name": "Radiotherapy",
        "criteria": "Localized symptomatic disease.",
        "rationale": "Excellent local control and symptom palliation.",
        "evidenceLevel": "Level 3",
        "sequencePosition": "Palliative Local Therapy"
      },
       {
         "name": "Best Supportive Care",
         "criteria": "When benefits of active treatment do not outweigh
risks/burden.",
         "rationale": "Focus on quality of life, symptom management.",
         "evidenceLevel": "N/A",
         "sequencePosition": "Primary Management Approach"
       }
      // Consider single-agent Rituximab? Low-dose chemo (oral
etoposide)? Steroids?
    ]
   },
   {
  "scenarioId": "CNS_Relapse_Treatment",
```

```
  "description": "Management strategies for secondary CNS relapse
(SCNSL) of DLBCL, which carries a poor prognosis and requires therapies
with CNS penetration.",
  "tags": ["CNS involvement", "CNS Relapse", "Salvage", "High-risk",
"Multi-modal therapy"],
  "applicableLines": ["Variable", "Often Second-line or Subsequent"],
  "patientProfile": ["Variable Fitness", "Requires CNS Penetrant
Therapy"],
  "priorContext": ["Confirmed CNS relapse (isolated or synchronous with
systemic relapse)"],
  "treatments": [
    {
      "name": "High-Dose Methotrexate (HD-MTX) + High-Dose Cytarabine
(ARA-C) - CNS Therapy",
      "criteria": "Fit patients suitable for intensive chemotherapy.
Often combined with Rituximab and potentially other CNS-penetrant
agents (e.g., thiotepa) based on PCNSL protocols.",
      "rationale": "Core systemic therapy approach aiming for CNS
control. HD-MTX is essential; addition of HD-ARA-C improves outcomes
based on PCNSL data.",
      "evidenceLevel": "Level 2/3 (Extrapolated from PCNSL and SCNSL
case series)",
      "comparisonNote": "Foundation of most curative-intent systemic
approaches.",
      "sequencePosition": "Primary Systemic Therapy"
    },
    {
      "name": "Ibrutinib (Imbruvica) - CNS Indication",
      "criteria": "Often used in combination regimens (e.g., with chemo
or immunotherapy) or as monotherapy in specific cases, frequently
within clinical trials.",
      "rationale": "Oral BTK inhibitor with CNS penetration showing
activity, though single-agent responses can be transient.",
      "evidenceLevel": "Level 3 / Investigational (combination data
emerging)",
      "comparisonNote": "Provides an oral, targeted option. Often
combined to improve durability.",
      "sequencePosition": "Systemic Therapy Component / Salvage"
    },
    {
      "name": "Radiotherapy",
      "criteria": "Consolidation after systemic therapy (especially if
incomplete response), palliative treatment for symptomatic lesions, or
sometimes whole-brain radiotherapy (WBRT) in specific situations.",
      "rationale": "Provides effective local control within the CNS.",
      "evidenceLevel": "Level 2/3",
```

```
      "comparisonNote": "Highly effective locally, but WBRT carries
significant neurotoxicity risk, especially long-term. Often deferred or
used for palliation.",
      "sequencePosition": "Consolidation / Palliative Local Therapy"
    },
    {
       "name": "R-IEV (Rituximab, Ifosfamide, Epirubicin, Etoposide)",
       "criteria": "Potential consideration as salvage chemo if patient
is fit and ASCT consolidation is planned, given component drugs may
have some CNS activity.",
       "rationale": "Includes agents with potential CNS penetration
(Ifosfamide, Etoposide). Limited direct evidence specifically for CNS
relapse salvage.",
       "evidenceLevel": "Level 3",
       "comparisonNote": "Less established for CNS relapse compared to
HD-MTX based regimens.",
       "sequencePosition": "Systemic Salvage (Consideration)"
     },
    {
      "name": "Clinical Trial",
      "criteria": "Strongly consider enrollment whenever possible,
especially given the poor prognosis and limited standard options.",
      "rationale": "Access to novel agents (e.g., CAR-T engineered for
CNS activity, new small molecules, bispecifics with CNS data) or
combinations.",
      "evidenceLevel": "N/A",
      "sequencePosition": "Optimal approach if available"
    },
    {
       "name": "Intrathecal Chemotherapy", // Added generic IT chemo
       "criteria": "May be considered as component of therapy
(prophylaxis/treatment), although systemic high-dose therapy (esp.
HD-MTX) is generally preferred for achieving therapeutic CNS levels.",
       "rationale": "Direct administration into CSF.",
       "evidenceLevel": "Level 3",
       "comparisonNote": "Role debated vs effective systemic CNS
penetration. Often used alongside systemic therapy.",
       "sequencePosition": "Adjunctive Therapy / Prophylaxis"
     }
    // Note: Standard CAR-T/Bispecific indications often exclude active
CNS disease in pivotal trials. Their role here is largely
investigational unless specific data/protocols exist.
  ]
}
]
```