Title Page

Patient-Reported Challenges in Lymphoma Diagnosis: Analysis of Online Forum Narratives Using Artificial Intelligence

Running title

Patient-reported diagnostic challenges in lymphoma

Author(s)

Feng He [a], Jose M Valderas [a]

Affiliation

[a] Centre for Research in Health Systems Performance (CRiHSP), National University of Singapore, Singapore

Correspondence

Feng He

Centre for Research in Health Systems Performance (CRiHSP), National University of Singapore, Singapore

Email: feng.he@u.nus.edu

Abstract word count: 250

Main text word count: 3623

Figure/Tables: 5

Version: 2025-11-19

ABSTRACT

Background: Lymphoma diagnosis remains challenging due to diverse subtypes and nonspecific presentations. While prior research focused primarily on clinical accuracy, how patients experience and describe these challenges remain understudied. This study systematically analyzed online patient narratives to investigate their perspectives on diagnostic difficulties.

Methods: We developed an artificial intelligence (AI) pipeline integrating DeepSeek large language model, optical character recognition, and transformer-embedded keyword clustering to analyse online narratives reporting diagnostic discrepancies or misdiagnosis from China's largest lymphoma forum for patients and caregivers (house086.com). The pipeline extracted patient demographics, timelines, diagnostic barriers, facilitators, outcomes, and AI-graded severity. External validation (n=400) against manually-derived labels assessed pipeline reliability. Multivariable logistic regression examined associations between barriers, facilitators, and the binarized severity scores.

Results: Over the study period (2011-2025), patients reporting diagnostic difficulties doubled, while the rate of severe outcomes declined. From 2016 narratives (median patient age 47; 59% family-authored), AI-assisted keyword taxonomy identified 7 diagnostic facilitators, 11 barriers, and 5 consequences. Psychological distress was the most common consequence of diagnostic challenges (90%). Clinician-related issues (91%) and case complexity (77%) were the most prevalent barriers, but inappropriate initial treatment conferred the greatest risk (OR 19.06, 11.30-32.17). Among facilitators, specialist input reduced severe outcomes by 40% (OR 0.60, 0.44-0.81), while peer networks (OR 0.62) and clinician expertise (OR 0.65) provided additional protection.

Conclusion: This large-scale analysis of patient narratives identified factors underlying patient-perceived diagnostic difficulties in lymphoma. AI enables scalable analysis of patient-generated data, offering insights into targeted quality improvement and digital health interventions in patient safety.

Keywords: Lymphoma, Patient-Reported Experience, Artificial Intelligence, Patient Safety, Social Media

INTRODUCTION

Lymphoma accounts for 3-4% of all cancers worldwide, with more than half a million new cases reported annually [1]. Accurate diagnosis of lymphoma is a major challenge in clinical practice, mainly because of its diverse subtypes and nonspecific symptoms in early stages [2]. Expert reviews in the United States, United Kingdom, and Europe have reported diagnostic change in 5.8-27.3% of the referred cases [2-4], while studies in China documented rates as high as 39.8% [5]. These diagnostic uncertainties can delay appropriate care, expose patients to harmful treatment, and may contribute to long-term health disparities [3, 6].

Prior research has focused on accuracy improvement and error prevention through clinician training, advanced imaging, or pathology support [3-8], but little is known about how patients themselves cope with diagnostic difficulties and uncertainties. Treatment navigation often requires patients and families to advocate for second opinions, pursue tertiary care, and endure financial and psychological strain [6, 9], as overturning an initial diagnosis can be difficult for clinicians and institutional support is often lacking. Global Patient Survey (GPS, 2024) of 11,170 lymphoma patients and caregivers by Lymphoma Coalition further highlighted ongoing gaps in communication, shared decision-making, and psychosocial support across countries, underscoring the need to better understand patient experiences during the diagnostic process [citation]. Patient narratives provide critical insights into these processes, but conventional surveys and registries capture them only at small scale and limited depth [10-12].

Recent advances in artificial intelligence (AI), particularly large language models (LLMs), enable scalable and systematic analyses of patient-generated data from social media [13-15]. While these self-reported narratives inherently carry limitations regarding clinical verification [10], their vast scale and richness provide a complementary, patient-cantered perspective to existing clinical literature [10, 16]. Applying LLMs to these narratives may therefore help identify factors associated with patient-perceived diagnostic difficulties, providing actionable insights into service delivery gaps [13-16].

Building on this, the present study aimed to (1) systematically identify patient-reported barriers and facilitators in lymphoma diagnosis, and (2) quantify the influence of specific barriers and facilitators on the severity of patient-perceived consequences, using an AI-driven analysis of online forum narratives.

METHODS

Study Design and Data Collection

We conducted a retrospective cross-sectional study of patient narratives from House086.com, China's largest online forum dedicated to lymphoma patients and caregivers [13]. To capture diverse patient terminology related to diagnostic difficulties (e.g., misdiagnosis, diagnostic error, pathological discrepancy), a keyword list was iteratively developed through exploratory manual review (**Supplementary Table S1**). All relevant posts published between April 2011 and October 2025 were then identified through keyword searches and manually reviewed to generate a curated list of candidate webpage links (URLs) at the time of data collection (June-October, 2025).

To retrieve and process raw webpage content from URLs, we developed an AI-based extraction pipeline (**Figure 1**), with optical character recognition (easyOCR [17]) to extract texts from attached medical report images. Text and OCR-derived content were analysed by DeepSeek, an LLM with strong performance in Chinese text comprehension [18]. From multi-user discussion threads, DeepSeek identified users with potential diagnostic difficulties, then standardized their narratives into structured case profiles for quantitative analysis.

Keyword Extraction and Taxonomy

Given the heterogeneity in content and style of patient narratives, the LLM was initially allowed to freely propose salient keywords for patient-reported diagnostic facilitators, barriers, and consequences during pilot analysis. The resulted keywords were mostly Chinese colloquial expressions and sometimes, English medical abbreviations. Hence we generated word embeddings using a multilingual sentence-transformer model (paraphrase-multilingual-MiniLM-L12-v2) to capture contextual and cross-lingual semantic similarity. Agglomerative hierarchical clustering with ward distance was then applied to group semantically related terms. Resulting clusters were manually reviewed and consolidated further into 7 facilitators (e.g., specialist input, family support, peer network), 11 barriers (e.g., case complexity, specimen issues, clinician related issues), and 5 consequences (disease progression, treatment-related harm, psychological distress, loss of time, and financial strain) based on clinical literature and pilot analysis. In formal analyses, the LLM with our predefined keyword taxonomy as prompt was allowed to choose up to three keywords per category (barriers, facilitators, and consequences) for each patient narrative.

Variable Definitions and Outcome

Patient-reported diagnostic challenges were defined as the description of misdiagnosis or diagnostic errors directly, or providing sufficient evidence of a diagnosis that was later contradicted or revised, as determined by the LLM from context. Username, user type (patient, family member, or others), and posting year were directly parsed from the HTML content, whereas patient demographics (age, sex), diagnostic encounters (date, hospital, diagnosis), and patient experience (barriers, facilitators, and consequences) were inferred directly from post contents by the LLM. For text quality control, a confidence score (1-5) was assigned by LLM to each narrative.

For the primary outcome, an overall severity of patient-reported consequences was graded by LLM on a scale of 1 to 5, with 5 indicating the worst consequences. Due to score concentration, we binarized the outcome into severe consequences (score ≥4) versus moderate-or-lower severity (≤3). A secondary outcome was defined as reporting clinically relevant consequences (disease progression or treatment-related harm) in those having a severity score ≥4.

Data Preprocessing

The inclusion and exclusion criteria workflow was summarized in **Figure 2.** From over three million forum posts, 15776 were returned by keyword search. After manual review, 13220 posts irrelevant to the study scope or duplicates of already selected posts were excluded. The remaining 2613 URLs for AI processing contributed 2865 narratives related to diagnostic difficulties. Of these, 849 narratives were excluded as they lacked meaningful description of the diagnostic history (n=1), flagged by DeepSeek as borderline case profiles (n=355), or had low content quality (confidence score ≤2, n=16). For multiple narratives from the same user, only the most recent and highest-quality version was retained (excluding n=423). Cases reporting single diagnostic encounter were manually reviewed and dropped if no evidence of diagnostic discrepancy was retrieved from the post (n=50). The final analytic dataset comprised 2016 unique profiles for analyses. For the secondary outcome, a refined subset of 917 high-quality profiles (confidence score ≥4, with known patient sex, known delay duration, and multiple encounters) was used.

Due to the anonymous nature of forum discussions, missing data was found in patient age, sex, delay duration, and the number of encounters. To preserve the sample size and possible

information, all missing values were as a separate category ("Not Disclosed"), with continuous variables converted to categorical ones for quantitative analysis.

Human Validation

To validate the reliability of AI-assisted data extraction, an external human validation was performed by three healthcare professionals fluent in Chinese. A stratified random subsample of 400 cases (approximately 20% of study population) was independently reviewed by coders with access to the original websites for discussion texts and uploaded medical records (including pathology reports, clinical notes, and treatment records where available). For case inclusion validation, each reviewer independently assessed whether each case genuinely reported diagnostic difficulties, defined as patient-reported misdiagnosis, or discrepancies between initial diagnosis and subsequent opinions. Final human labels were determined by majority vote. Positive predictive value (PPV) was computed by comparing human labels against AI classifications. For severity validation, human coders, blind to AI outputs, independently graded consequence severity on the same 1-5 scale. A median human score was computed for each case. Agreement between median human ratings and AI-generated scores was assessed using Cohen's κ for binarized outcome (severe vs moderate-or-lower severity).

Statistical Analyses

For population characteristics, continuous variables were summarized as mean with standard deviation (SD) or medians with interquartile ranges (IQRs) as appropriate for the variable, while categorical variables were summarized as counts (row %). For keyword frequencies, counts with column % were used. Inter-group differences were tested with Student's t-Test, Mann-Whitney U test, or chi-square as appropriate.

Associations between AI-identified barriers, facilitators, and binarized consequence severity were estimated using multivariable logistic regression, adjusting for patient age, sex, user type, posting year, delay duration, and the number of diagnostic encounters. All numeric variables were categorized to account for potential non-linear effects (see **Table 1** for details). Odds ratios (ORs) with 95% confidence intervals (CIs), and corresponding p-values were reported. Multicollinearity was assessed using the Cramér's V [19].

All analyses were done using R version 4.2.0 and Python 3.11.13. All codes used for data collection, preprocessing, and statistical analysis will be made publicly available upon publication.

Ethics

This study analysed patient narratives from an open-access online forum in China (House086.com). All data were publicly accessible without registration or login requirements, and the platform terms of use permit analysis of forum contents without further consent. No identifiable personal information was collected or analysed. Only de-identified and aggregate results are reported. Use of publicly available, legally obtained, and de-identified data that do not harm individuals or involve sensitive personal information is consistent with relevant data-use regulations in China. The protocol has also been submitted to the National University of Singapore Institutional Review Board and is currently under review. The study adhered to the principles of the Declaration of Helsinki.

RESULTS

Population Characteristics

Of the 2016 analyzed cases, 929 (46.1%) were classified as severe consequences (score ≥4), whereas 1087 (53.9%) were classified as moderate-or-lower (≤3) (**Table 1)**. A high proportion did not disclose patient age (65.7%) or sex (31.1%). The majority of posts were authored by family members (59.2%), followed by patient themselves (38.6%). Self-authored posts were significantly more likely to report moderate-or-lower consequences compared to family-authored posts (60.3% vs 50.3%).

From 2011-2025, the number of users reporting diagnostic challenges increased by over 100%. Concurrently, the proportion of moderate-or-lower severity outcomes significantly improved, rising from 41.9% in 2011-2015, to 53.6% in 2016-2020, and 59.1% in 2021-2025.

Patients who reported three or more diagnostic encounters were more likely to have severe consequences (58.1% vs 43.5% in those reporting only two encounters). Furthermore, patients with moderate-or-lower severity reported a significantly shorter median diagnostic delay (1.0 month, IQR 1.0-3.0) than those with severe consequences (4.0 months, 2.0-11.0).

Keyword Taxonomy and Frequencies

AI-assisted taxonomy development yielded 11 patient-reported barriers, 7 facilitators, and 5 consequences. These categories, spanning clinical, systemic, and personal domains, are detailed with their corresponding definitions in **Table 2**.

The most frequently described consequence of diagnostic difficulties was psychological distress, reported in 90.1% of the narratives (**Table S2**). Analysis by consequence severity revealed significant concentration of clinically relevant consequences in the severe group, with disease progression four times more prevalent (66.8% vs 15.5%), and treatment-related harm twice as prevalent (52.2% vs 24.5%). Conversely, loss of time was significantly more prevalent in the moderate-or-lower severity group (69.8% vs 61.6%).

Among barriers, clinician-related issues (90.6%) and case complexity (77.3%) were overwhelmingly reported (**Table S3**). While patient-level barriers (e.g., patient psychological resistance, limited health literacy, and physical limitations) and family/financial constraints were mentioned less frequently (all <7%). The most common facilitators were specialist input (66.7%), patient self-advocacy (61.1%), and tertiary-hospital care (60.3%).

Multivariable Regression Analysis

In the primary model (outcome: severe consequences, score ≥4), all 11 AI-identified barriers were significantly associated with an increased likelihood of severe consequences (**Figure 4**, **Table S3**). The strongest predictor was misguided prior treatment (OR 19.06, 95% CI 11.30-32.17). Other factors with exceptionally high risk were system level inefficiency (OR 8.93, 6.00-13.29), family and financial constraints (OR 8.47, 4.05-17.73), and patient physical limitations (OR 7.40, 3.37-16.26). Notably, the most frequently reported barrier, clinician-related issues, was also a major risk factor (OR 6.00, 3.74-9.65).

Among facilitators, specialist input (OR 0.60, 0.44-0.81), peer network engagement (OR 0.62, 0.45-0.87) and clinician expertise (OR 0.65, 0.46-0.90) were significantly associated with a decreased likelihood of severe consequences.

A notable finding was the association of family support with increased odds of severe consequences in the primary model (OR 1.49, 1.02-2.17), though this effect disappeared in the secondary model (outcome: clinically critical consequences). The secondary model focusing on disease progression or treatment harm, generally showed consistent directionality

but larger effect sizes for all barriers, and highlighted specialized interventions as a significant protective factor (OR 0.40, 0.20-0.81), which was non-significant in the primary model.

Collinearity testing showed moderate correlation between two pairs of facilitators: tertiary hospitals and peer networks (phi=0.51), and specialized interventions and specialist input (0.41), but both were retained due to their distinct clinical implications. Human validation of 400 randomly sampled cases yielded PPV of [%] for case identification and inter-rater agreement of κ=[to be added] for severity grading.

DISCUSSION

Key Findings

This large-scale AI-driven study of online patient narratives yielded several key insights for diagnostic challenges in lymphoma. First, while the number of narratives describing diagnostic difficulties more than doubled from 2011 to 2025, the overall trend suggested a significant improvement in patient-reported outcomes over the same period. Our analyses revealed clinician-related issues (90.6%) and case complexity (77.3%) as the most frequently reported barriers, yet the strongest predictors of severe consequences included misguided prior treatment (OR 19.06), system-level inefficiency (OR 8.93), and family or financial constraints (OR 8.47). Specialist input (66.7%) and patient self-advocacy emerged as the leading facilitators in treatment navigation, with specialist input, peer network, and clinician expertise serving as protective factors against severe outcomes. Together, these results provide a patient-centered taxonomy and a quantitative risk map that complements existing literature on diagnostic safety.

Comparison with Existing Evidence

Our barrier-facilitator framework echoes prior work on cancer care quality, where delays, fragmentation, and inadequate psychosocial support were key barriers, and multidisciplinary coordination acted as facilitators [11]. The recent global survey by Lymphoma Coalition similarly revealed significant information gaps and treatment regret, with Chinese respondents reporting particularly severe challenges: only 15-18% understood their diagnostic process (vs. 59% globally), and 53% expressed treatment regret (vs. 11% globally) (citation). However, that survey, while valuable, could not identify the specific risk factors or mechanisms underlying these disparities. Our study extends this literature by providing a

quantitative risk map derived directly from the patient experience in a Chinese context. For example, while the GPS found that 91% of Chinese patients considered treatment cost a major concern, our regression analysis demonstrates that financial constraints were associated with an 8.5-fold increase in odds of severe consequences (OR 8.47). Similarly, the "information gaps" highlighted in the GPS correspond in our data to clinician-related issues (OR 6.00) and inadequate specialist access.

The high prevalence of case complexity (77.3%) and specimen issues (23.8%) as AI-identified barriers is consistent with long-standing recognition that pathology and histopathological interpretation in lymphoma are inherently challenging [2-4]. From the patient perspective, clinician-related issues were the most common barrier (90.6%). While this high frequency does not necessarily imply malpractice, it highlights prevalent gaps in training and communication that compromise quality care [6, 20]. More importantly, our analysis distinguished between these high-frequency issues and high-impact risk factors. Inappropriate initial treatment, although reported in a minority of cases (7.8%), manifested as the strongest barrier in logistic regression (OR 19.06), exemplifying the risks of initiating therapy without diagnostic certainty. This is aligned with evidence by Proctor et al. that approximately 11% of lymphoma treatment plans were revised following discordant diagnoses [3].

Similarly, the high OR for system level inefficiency (OR 8.93), together with specialist access barriers and medical resources shortages, points to the structural and systemic deficiencies as critical drivers of severe consequences, often surpassing individual clinician errors. Beyond the hospital domain, patient and family-level barriers, including psychological resistance, health illiteracy, financial pressure, and physical limitations, also pose substantial, though less frequent risks.

Among the facilitators, specialist input and clinician expertise represented positive forces from within the healthcare system, while peer networks, formed by patients and caregivers, also emerged as an important source of support. According to Engler et al., cancer patients actively seek out stories from peers to better understand treatment choices, manage emotional impact, and reduce feelings of isolation [16]. Such experience-sharing may complement clinical information and foster emotional coping, particularly in the face of diagnostic uncertainty or difficulty.

A key finding was the counterintuitive association of family support with increased odds of severe consequences in the primary model. The predominance of family-authored narratives (59.2%) highlights the central role of family advocacy in treatment navigation. Rather than suggesting family support is inherently harmful, this association likely reflects reverse causality or confounding by indication, where family support may only become intensely necessary and reported in narratives when a case is already complicated, delayed, or progressing poorly, emerging as a "last resort" strategy after prolonged uncertainty [6, 21]. These findings suggest that individual-level efforts are insufficient without systemic support, in contrast to specialist input that represents structured intervention.

Finally, psychological distress was the most frequently mentioned consequence (90.1%) regardless of consequence severity. While not directly predictive of clinical harm, such distress may evolve into resistance to further diagnostic testing and follow-up compliance, potentially undermining long-term health status. Clinically relevant harms (disease progression and treatment-related harm) and financial strain were markedly more common in those reporting severe consequences. Similar to our taxonomy, prior work also described psychological, family, and financial consequences of problematic care events, revealing the multifaceted nature of patient-perceived harm [6, 9].

Implication for Practice and Digital Health Applications

Our findings suggest that targeted interventions are needed to systematically address barriers to patient-centred lymphoma care. International initiatives, such as the Nebraska Lymphoma Study Group, the North Central London Lymphoma network, and the French Lymphopath Network show how centralized expert review can reduce diagnostic errors [2-4]. However, such frameworks remain difficult to implement in countries with limited resources and large patient populations, where infrastructural gaps often leave patients responsible for complex tasks like arranging the transfer of pathology slides for second opinions.

Digital health tools may offer complementary and scalable solutions to address diagnostic challenges at multiple levels. First, to combat the frequent barriers of case complexity and specimen issues, AI-assisted pathology systems have achieved high accuracy in lymphoma detection and hold promise for reducing time and man-power costs [7, 8]. Second, patient-facing LLM applications, such as chatbots, may enhance health literacy, provide immediate

psychological support, and facilitate shared decision-making. By providing clear, accessible information, these tools could potentially mitigate friction associated with clinician-related issues and fostering trust in diagnostic process [20, 22, 23]. Third, system-level tools such as decision-support platforms, triage algorithms, telemedicine, and outpatient robotics could help address financial, logistical, and physical barriers, particularly in under-resourced regions.

Critically, we observed that patients are already adopting AI tools independently to interpret pathology reports and disease information. While recent work has raised concerns that patients may over-trust AI-generated advice even when inaccurate [24]. Our findings complement this perspective by examining a population already exposed to potential diagnostic inaccuracies, often in settings where specialist access is limited. In such contexts, patients described turning to AI for its perceived neutrality and convenience, underscoring the urgent need for health systems to integrate safe, evidence-based digital tools with appropriate medical oversight.

Strengths and Limitations

This study has several strengths. To our knowledge, it is the first large-scale, AI-driven analysis focusing on patients who experienced diagnostic difficulties in oncology, an important but often overlooked population. By systematically analysing over 2000 online patient narratives, we were able to capture authentic, patient-cantered perspectives that are less constrained by traditional clinical or institutional framing, highlighting the value of patient voices in chronic disease management [14]. The AI pipeline provided a consistent, scalable, and reproducible method for structuring patient-generated text, thereby offering a transferable framework to digital health research across different health systems and disease areas. The external human validation sub-study, using independent clinical records, enhances the overall credibility of the AI-extracted data.

Several limitations must be acknowledged, primarily related to the nature of patient-reported data. First, the data reflects patient-perceived diagnostic discordance rather than clinically verified error rates. Given the retrospective account, potential for recall bias or limited medical terminology, patient narratives are inherently framed from their perspectives rather than representing objective clinical truth. Nevertheless, these narratives carry unique value by capturing the patient's lived experience of uncertainty and providing direct insights into

specific healthcare delivery gaps. Second, the anonymous nature of the online forum resulted in substantial missing demographic data (e.g., 65.7% missing age), which limited the scope of subgroup analyses. While we mitigated this by coding "Not Disclosed" as a separate category, this approach cannot fully eliminate potential confounding effects. Third, methodological limitations in data extraction included occasional OCR failure on low-resolution images and the necessity of limiting AI analysis to the initial text and up to 10 images per post, which may have introduced some information loss. This was partly mitigated by the LLM's contextual inference capability and the design of a case confidence score to exclude low-quality narratives from the final analysis. Moreover, the LLM-graded consequence severity score clustered at 3 (49.4%) and 4 (42.3%), requiring us to binarize the outcome (score ≥4 vs ≤3) for a balanced and stable regression model. Finally, as all narratives were extracted from a single, Chinese-language platform, the results may not be generalizable to other populations.

Future Research

Future work will focus on longitudinal analyses of patient narratives to examine the dynamic interplay of barriers and facilitators over time, distinguishing short- from long-term consequences perceived by patients and families. Item response theory [25] may be incorporated to improve the precision and reliability of patient-reported outcome measurements. Health economics methods could help provide a more comprehensive assessment of the societal and financial burden imposed by diagnostic challenges. Beyond lymphoma, the AI-driven analytic framework developed in this study could be extended to patient-authored texts across other diseases and healthcare systems, enabling systematic incorporation of patient perspectives into clinical research and practice.

CONCLUSION

This large-scale, AI-based analysis of more than 2000 online forum narratives identified patient-perceived barriers and facilitators in lymphoma diagnosis. Our findings reveal that diagnostic safety requires not only hospital-based interventions but also active engagement of patients, families, peer networks, and health systems. More broadly, this study demonstrates the feasibility of leveraging patient-generated social media data to inform clinical practice and digital health innovation, underscoring the potential of AI tools to systematically evaluate health services and improve patient-perceived quality of care.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

All authors declare no competing interests.

AUTHOR CONTRIBUTIONS

(Provisional, may be updated in subsequent revisions)

Feng He: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Writing, Visualization, Project Administration.

José M. Valderas: Supervision, Methodology, Writing.

ACKNOWLEDGMENTS

We thank participants of online patient forums for sharing their experiences, which made this research possible. ChatGPT was used for language polishing, and Claude was used for code development. All scientific content and analyses were conducted and verified by the author.

DATA & CODE AVAILABILITY

All codes used for data collection, preprocessing, and statistical analysis will be made publicly available upon publication. The de-identified dataset may be shared upon reasonable request.

REFERENCES

REFERENCES

1. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
2. Laurent, C., et al., *Impact of expert pathologic review of lymphoma diagnosis: study of patients from the French Lymphopath Network.* Journal of clinical Oncology, 2017. **35**(18): p. 2008-2017.
3. Proctor, I.E., et al., *Importance of expert central review in the diagnosis of lymphoid malignancies in a regional cancer network.* Journal of Clinical Oncology, 2011. **29**(11): p. 1431-1435.
4. Bowen, J.M., et al., *Lymphoma diagnosis at an academic centre: rate of revision and impact on patient care.* Br J Haematol, 2014. **166**(2): p. 202-8.
5. Deng, J., et al., *Misdiagnosis analysis of 2291 cases of haematolymphoid neoplasms.* Frontiers in Oncology, 2023. **13**: p. 1128636.
6. Singh, H., et al., *The global burden of diagnostic errors in primary care.* BMJ Qual Saf, 2017. **26**(6): p. 484-494.
7. Syrykh, C., et al., *Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning.* NPJ digital medicine, 2020. **3**(1): p. 63.
8. Li, D., et al., *A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals.* Nature communications, 2020. **11**(1): p. 6004.
9. Mazor, K.M., et al., *Toward patient-centered cancer care: patient perceptions of problematic events, impact, and response.* Journal of Clinical Oncology, 2012. **30**(15): p. 1784-1790.
10. Tsianakas, V., et al., *Using patients' experiences to identify priorities for quality improvement in breast cancer care: patient narratives, surveys or both?* BMC Health Services Research, 2012. **12**(1): p. 271.
11. Wagner, E.H., et al., *The quality of cancer patient experience: perspectives of patients, family members, providers and experts.* Quality and Safety in Health Care, 2010. **19**(6): p. 484.
12. Nartey, Y., et al., *Is the English Cancer Patient Experience Survey representative? A comparative analysis with the National Lung Cancer Audit.* Lung Cancer, 2020. **140**: p. 27-34.
13. Ning, K.G., H; Franklin, M; Yang, X; Wei, R; Song, Z; Xu, H; Leng, L; Liu, M; Dai, J; Zhang, J; Zeng, R; Hou, Y; Wang, R; Liu, Z; Huang, C; Cai, R; Liu, H; Xia, LC, *Analysis of China's Largest Blood Cancer Patient Forum: Regional Inequities in Health-Seeking Behavior and Need for Enhanced Online Patient Support.* J Med Internet Res, 2025.
14. Dreisbach, C., et al., *A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data.* Int J Med Inform, 2019. **125**: p. 37-46.
15. Siddiqui, Z.A., et al., *Leveraging social media data to study disease and treatment characteristics of Hodgkin's lymphoma Using Natural Language Processing methods.* PLOS Digit Health, 2025. **4**(3): p. e0000765.

16. Engler, J., et al., *Using others' experiences. Cancer patients' expectations and navigation of a website providing narratives on prostate, breast and colorectal cancer.* Patient Education and Counseling, 2016. **99**(8): p. 1325-1332.
17. JaidedAI. *EasyOCR: Ready-to-use Optical Character Recognition with 80+ Supported Languages*. 2020 2025-10-27 [cited 2025 2025-09-02]; Available from: https://github.com/JaidedAI/EasyOCR.
18. Gibney, E., *China's cheap, open AI model DeepSeek thrills scientists.* Nature, 2025. **638**(8049): p. 13-14.
19. Rea, L.M. and R.A. Parker, *Designing and conducting survey research: A comprehensive guide*. 2014: John Wiley & Sons.
20. Salwei, M.E., J.S. Ancker, and M.B. Weinger, *The decision aid is the easy part: workflow challenges of shared decision making in cancer care.* JNCI: Journal of the National Cancer Institute, 2023. **115**(11): p. 1271-1277.
21. Rezaei, M., et al., *Caregiving consequences in cancer family caregivers: a narrative review of qualitative studies.* Frontiers in Public Health, 2024. **12**: p. 1334842.
22. Thirunavukarasu, A.J., et al., *Large language models in medicine.* Nature Medicine, 2023. **29**(8): p. 1930-1940.
23. Arora, A. and A. Arora, *The promise of large language models in health care.* The Lancet, 2023. **401**(10377): p. 641.
24. Shekar, S., et al., *People Overtrust AI-Generated Medical Advice despite Low Accuracy.* NEJM AI, 2025. **2**(6): p. AIoa2300015.
25. Chang, C.-H. and B.B. Reeve, *Item response theory and its applications to patient-reported outcomes measurement.* Evaluation & the health professions, 2005. **28**(3): p. 264-282.

TABLES AND FIGURES

Figure 1. AI Pipeline for Extracting and Structuring Patient Forum Narratives.

Forum posts from House086.com were curated through keyword searches and manual review into a list of webpage links. Text contents were parsed from the HTML source, and text appearing within attached medical images was transcribed using optical character recognition. All textual information was analysed by a large language model (DeepSeek), which identified individual users reporting potential diagnostic challenges and converted their narratives into structured case profiles capturing demographics, diagnostic history, and patient-reported barriers, facilitators, and consequences. A predefined keyword taxonomy, refined through embedding-based semantic clustering, guided the classification of keywords. Quality scores (1–5) were assigned to each case; a stratified 20% subsample underwent independent human validation for both case inclusion and outcome severity.

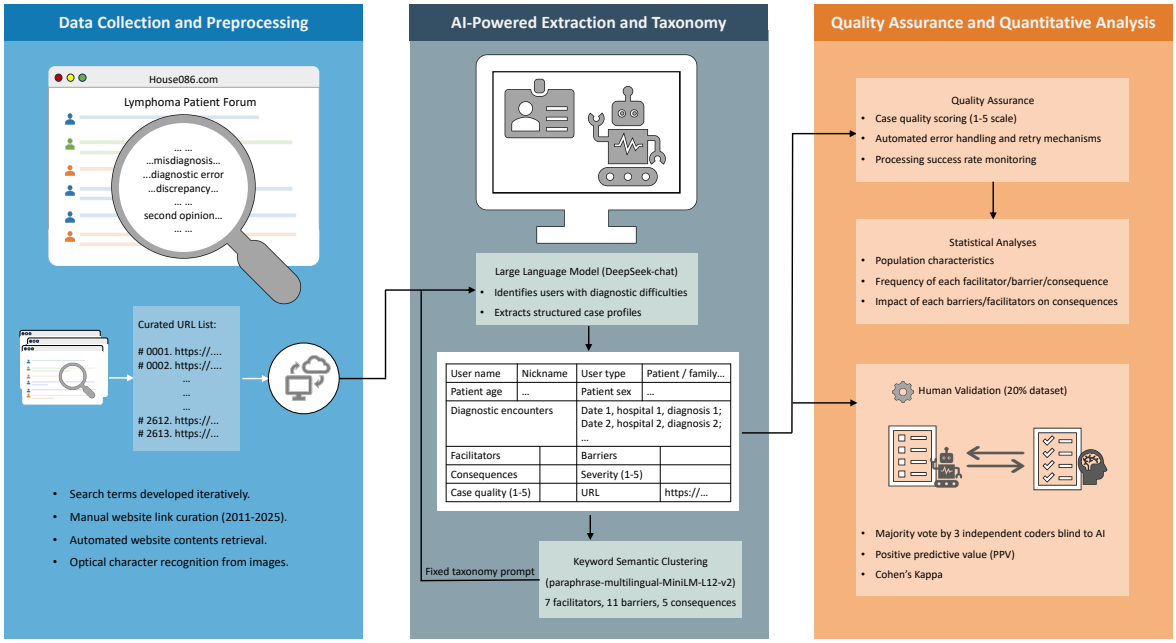


Figure 2. Data Cleaning Workflow for Patient Forum Posts.

From more than three million posts, irrelevant entries were excluded using keyword screening and manual review. Candidate posts were processed through the AI pipeline to extract diagnosis-related details. After quality control and de-duplication, 2016 unique case profiles were retained for analysis.

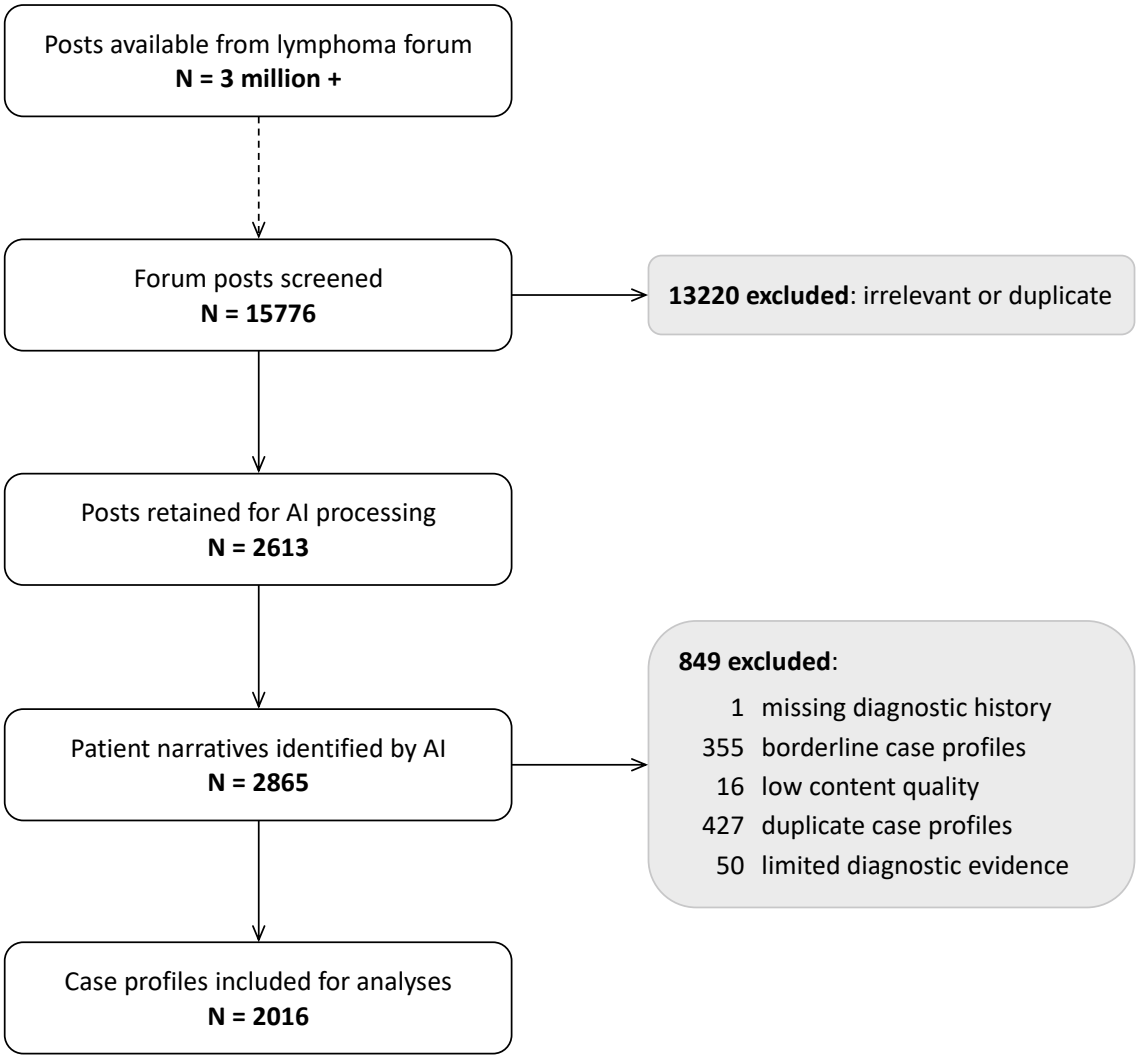


Table 1. Population characteristics by the patient-reported consequence severity

	Overall (N=2016)	Severity score ≤3 (n=1087)	Severity score ≥4 (n=929)	p
Patient age, years	47.0 [29.0, 59.0]	49.0 [31.0, 60.0]	45.0 [28.0, 58.0]	0.051
Patient age category				0.116
<18	56	22 (39.3)	34 (60.7)	
18-40	243	126 (51.9)	117 (48.1)	
40-60	227	115 (50.7)	112 (49.3)	
>=60	166	93 (56.0)	73 (44.0)	
Not disclosed	1324	731 (55.2)	593 (44.8)	
Patient sex				<0.001
Female	609	314 (51.6)	295 (48.4)	
Male	780	370 (47.4)	410 (52.6)	
Not disclosed	627	403 (64.3)	224 (35.7)	
Post user type				<0.001
Patient-self	779	470 (60.3)	309 (39.7)	
Family member	1194	601 (50.3)	593 (49.7)	
Others (admin, doctors)	43	16 (37.2)	27 (62.8)	
Posting year				<0.001
2011 – 2015	346	145 (41.9)	201 (58.1)	
2016 – 2020	822	441 (53.6)	381 (46.4)	
2021 – 2025	848	501 (59.1)	347 (40.9)	
No. of diagnostic encounters*	2.5 (1.0)	2.3 (0.8)	2.7 (1.1)	<0.001
No. of encounters				<0.001
2	1141	645 (56.5)	496 (43.5)	
>= 3	637	267 (41.9)	370 (58.1)	
Not disclosed	238	175 (73.5)	63 (26.5)	
No. of months delayed	2.0 [1.0, 6.0]	1.0 [1.0, 3.0]	4.0 [2.0, 11.0]	<0.001

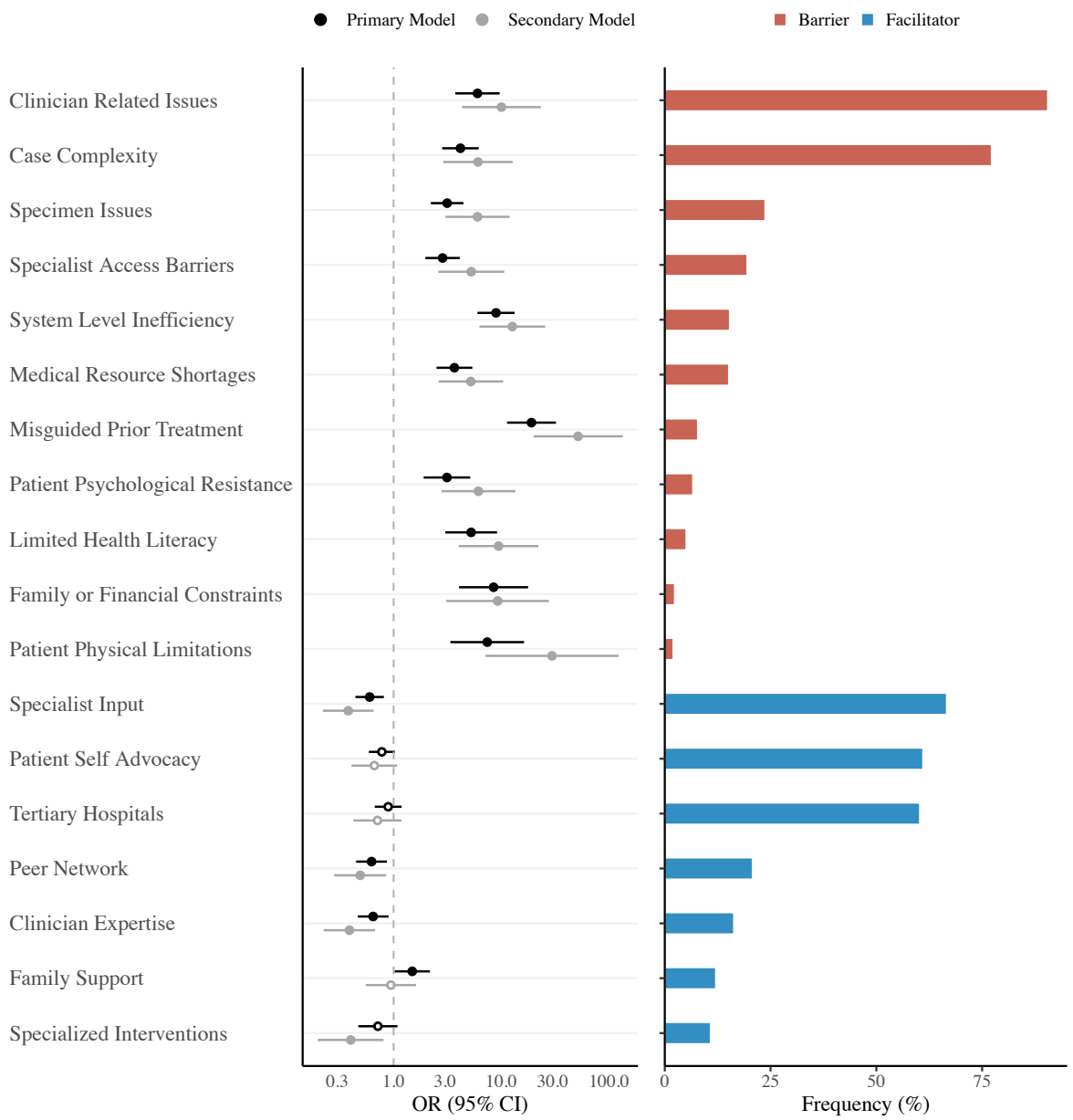
* Reported as mean (SD) due to limited variability in median values.

Table 2. Embedding-based mapping of barriers, facilitators, and consequences.

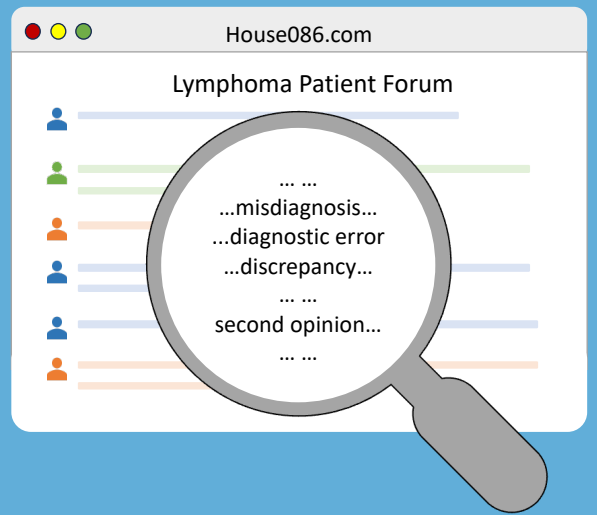
Barriers (p=11)	
Clinician Related Issues	Inadequate communication, low diagnostic accuracy, errors, or unprofessional attitudes among treating clinicians.
Case Complexity	Diagnostic challenges due to rare subtypes, atypical presentations, ambiguous pathology, or conflicting expert opinions.
Specimen Issues	Problems with sample acquisition, adequacy, handling, or quality that limit diagnostic reliability.
Specialist Access Barriers	Difficulty securing timely appointments, referrals, or consultations with relevant subspecialists.
System Level Inefficiency	Healthcare process barriers such as pandemic restrictions, complex administrative procedures, slow pathology turnaround, and poor interdepartmental coordination.
Medical Resource Shortages	Lack of essential facilities, equipment, medications, or specialist staff needed for timely and accurate diagnosis or treatment.
Misguided Prior Treatment	Inappropriate or harmful interventions before an accurate diagnosis, leading to complications or disease progression.
Patient Psychological Resistance	Reluctance or refusal to undergo recommended tests or treatments due to fear, anxiety, denial, or mistrust.
Limited Health Literacy	Insufficient understanding of medical information, disease severity, treatment options, or referral processes, including reliance on unproven remedies.
Family or Financial Constraints	Household-level barriers such as caregiver disagreement, lack of family support, or inability to afford care despite available services.
Patient Physical Limitations	Health or functional conditions that hinder diagnostic procedures or treatment, such as frailty, comorbidities, or pregnancy.
Facilitators (p=7)	
Specialist Input	Additional insights from subspecialists or senior consultants through formal consultations or second opinions.
Patient Self-Advocacy	Active involvement in seeking information, contacting specialists, evaluating treatment options, and adhering to care.
Tertiary Hospitals	High-level urban medical centres with comprehensive subspecialty services and advanced infrastructure.
Peer Network	Patient and caregiver communities that exchange experiences, treatment information, and healthcare navigation advice.
Clinician Expertise	Advanced knowledge, skills, and judgement of treating clinicians that improve diagnostic and treatment outcomes.
Family Support	Emotional, practical, or financial assistance from relatives that facilitates diagnosis and treatment.
Specialized Interventions	Advanced diagnostic tests or treatments such as PET-CT, genomic testing, novel drugs, clinical trials, or CAR-T therapy.
Consequences (p=5)	
Psychological Distress	Emotional harm such as anxiety, depression, or trauma caused by misdiagnosis or diagnostic discrepancies.
Loss of Time	Unnecessary time spent on ineffective diagnostic or treatment pathways.
Disease Progression	Worsening of the underlying disease due to diagnostic discordance or treatment delays.
Treatment-Related Harm	Injury or complications directly resulting from inappropriate or unnecessary treatment.
Financial Strain	Increased economic burden from repeated consultations, tests, or ineffective treatments.

Figure 3. Barriers and Facilitators Associated With Patient-Reported Outcomes

Left: Forest plot from multivariable logistic regression showing odds ratios (95% CI) for each barrier and facilitator in relation to unfavorable outcome (defined as severity ≥4 in the primary model, additionally with the presence of clinically relevant consequences in the secondary model). Filled dots indicate $p \leq 0.05$; open dots indicate $p > 0.05$. Right: Frequency of barriers and facilitators across the analytic cohort (n = 2016). Bars show the proportion of patient narratives mentioning each factor, with barriers in red and facilitators in blue, ordered by descending frequency within each category.



Data Collection and Preprocessing

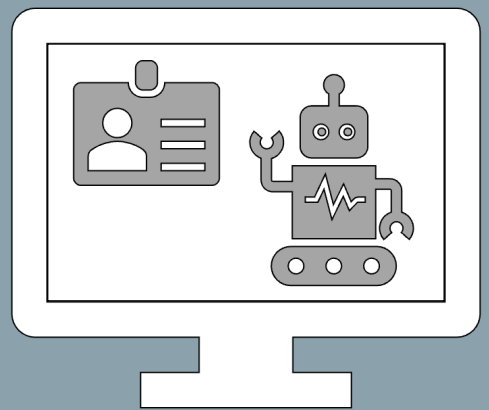


Curated URL List:

0001. https://....
0002. https://....
…
…
…
2612. https://...
2613. https://...

- Search terms developed iteratively.
- Manual website link curation (2011-2025).
- Automated website contents retrieval.
- Optical character recognition from images.

AI-Powered Extraction and Taxonomy



Large Language Model (DeepSeek-chat)

- Identifies users with diagnostic difficulties
- Extracts structured case profiles

User name	Nickname	User type	Patient / family…
Patient age	…	Patient sex	…
Diagnostic encounters		Date 1, hospital 1, diagnosis 1; Date 2, hospital 2, diagnosis 2; …	
Facilitators		Barriers	
Consequences		Severity (1-5)	
Case quality (1-5)		URL	https://…

Fixed taxonomy prompt

Keyword Semantic Clustering

(paraphrase-multilingual-MiniLM-L12-v2)

7 facilitators, 11 barriers, 5 consequences

Quality Assurance and Quantitative Analysis

Quality Assurance

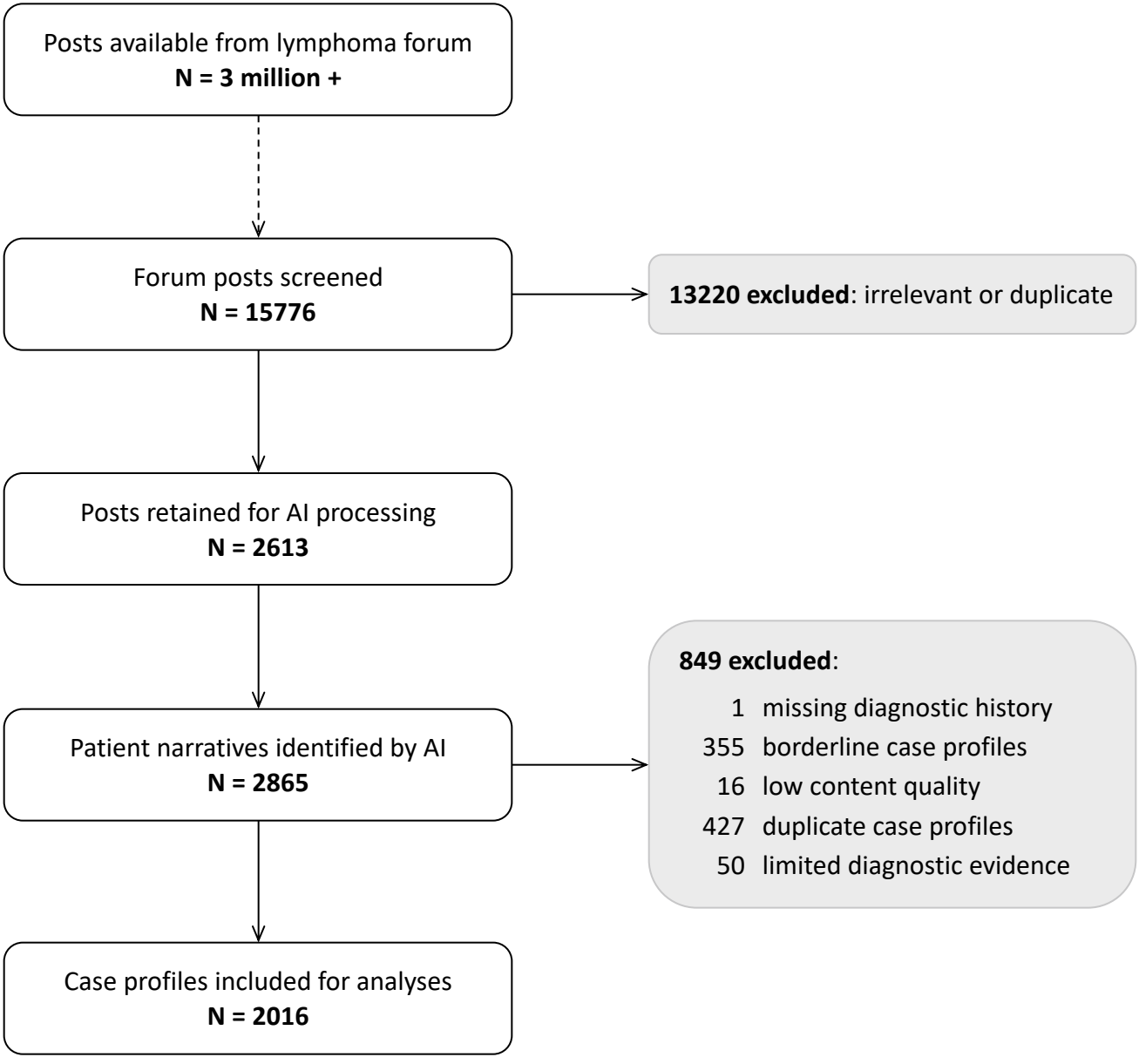
- Case quality scoring (1-5 scale)
- Automated error handling and retry mechanisms
- Processing success rate monitoring

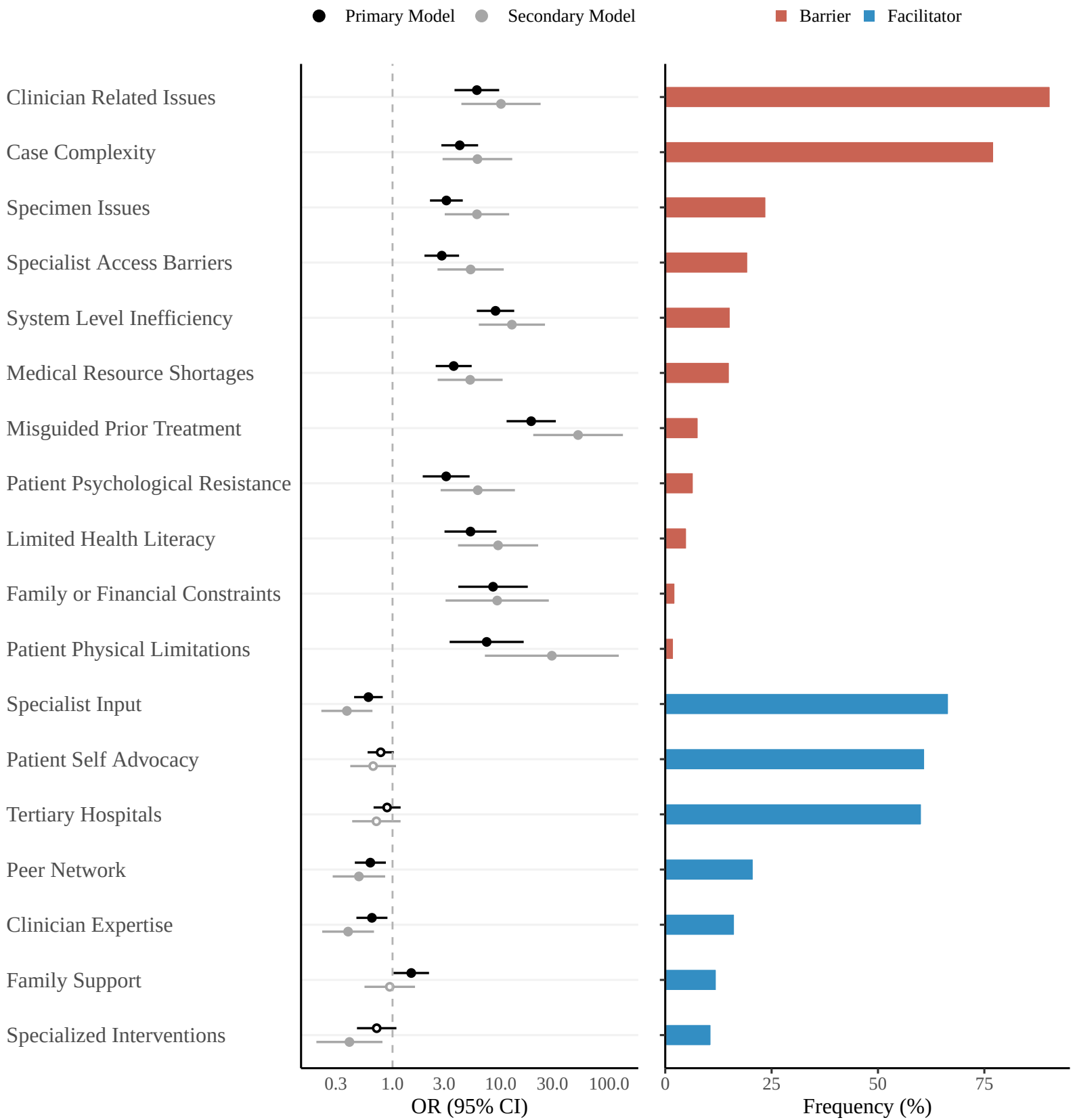
Statistical Analyses

- Population characteristics
- Frequency of each facilitator/barrier/consequence
- Impact of each barriers/facilitators on consequences

Human Validation (20% dataset)

- Majority vote by 3 independent coders blind to AI
- Positive predictive value (PPV) for inclusion
- Cohen's Kappa for severity rating


Posts available from lymphoma forum
N = 3 million +
Forum posts screened
N = 15776
13220 excluded: irrelevant or duplicate
Posts retained for AI processing
N = 2613
Patient narratives identified by AI
N = 2865
849 excluded:
1 missing diagnostic history
355 borderline case profiles
16 low content quality
427 duplicate case profiles
50 limited diagnostic evidence
Case profiles included for analyses
N = 2016


Primary Model
Secondary Model
Barrier
Facilitator
Clinician Related Issues
Case Complexity
Specimen Issues
Specialist Access Barriers
System Level Inefficiency
Medical Resource Shortages
Misguided Prior Treatment
Patient Psychological Resistance
Limited Health Literacy
Family or Financial Constraints
Patient Physical Limitations
Specialist Input
Patient Self Advocacy
Tertiary Hospitals
Peer Network
Clinician Expertise
Family Support
Specialized Interventions
0.3
1.0
3.0
10.0
30.0
100.0
OR (95% CI)
0
25
50
75
Frequency (%)