Title: An individualized prediction model for early-stage classic Hodgkin Lymphoma

Authors: Angie Mae Rodday* PhD[1], Andrew M. Evens* DO[2], Matthew J. Maurer ScD[3], Jenica N. Upshaw MD[4], Nicholas Counsell MSc[5], Sara Rossetti MD[6], Cheryl Chang BA[7], Zhu Cui MD[8], Qingyan Xiang PhD[4], Raphael Mwangi MS[3], Ranjana Advani MD[7], Marc Andre MD[9], Andrea Gallamini MD[10], Annette E. Hay MB ChB[11], David C. Hodgson MD[12], Richard T. Hoppe MD[13], Martin Hutchings MD[6], Peter Johnson MD[14], Eric Mou MD[15], Stephen Opat MBBS[16], John Raemaekers MD[17], Kerry J. Savage MD[18], Susan K. Parsons** MD[1], John Radford** MD[19]

*Contributing equally as co-first authors; **Contributing equally as co-senior authors

Affiliations:

[1]Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

[2]Division of Blood Disorders, Rutgers Cancer Institute, New Brunswick, NJ, USA;

[3]Division of Clinical Trials and Biostatistics and Division of Hematology, Mayo Clinic, Rochester, MN, USA

[4]Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

[5]Cancer Research UK and University College London Cancer Trials Centre, Cancer Institute, University College London, London, UK

[6]Department of Hematology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

[7]Stanford Cancer Institute, Stanford University, Stanford, CA

[8]Division of Hematology/Oncology, Tufts Medical Center, Boston, MA, USA

[9]Department of Haematology, CHU UCL Namur, Yvoir, Belgium

[10]Research and Innovation Department, A. Lacassagne Cancer Center, Nice, France

[11]Department of Medicine, Queen's University, Kingston, ON, Canada

[12]Department of Radiation Oncology, Princess Margaret Cancer Centre and University of Toronto, Toronto, Canada

[13]Department of Radiation Oncology, Stanford University, Stanford, CA, USA.

[14]School of Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

[15]College of Medicine, Division of Hematology, Oncology, and Blood & Marrow Transplant, University of Iowa, Iowa City, IA, USA

[16]Lymphoma Research Group, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia

[17]EORTC Lymphoma Group, Brussels, Belgium

[18]University of British Columbia and the Department of Medical Oncology, BCCancer, Centre for Lymphoid Cancer, Vancouver, BC, Canada

[19]University of Manchester, Christie NHS Foundation Trust, and NIHR Manchester Biomedical Research Centre

Corresponding Author: Andrew M. Evens, Rutgers Cancer Institute, 195 Little Albany St, New Brunswick, NJ, 08901; e-mail: andrew.evens@rutgers.edu

Acknowledgements: We are indebted to several people in the conduct of this research. We specifically would like to acknowledge the expertise, effort, and generosity of the group statisticians from the clinical trials groups and registries in assembling the de-identified data used in this analysis. We are grateful to Rachel Murphy-Banks, MA, our project manager, who tirelessly oversees work of the HoLISTIC Consortium. Further, we thank Brian Felter, BS, for his assistance in manuscript preparation and Jonathan Huang, MS, for his assistance with internal-external validation. Finally, we thank Ralph M. Meyer, MD, for his contributions to the consortium from its inception and wish him a glorious retirement.

Keywords: Hodgkin lymphoma; prognosis; prediction modeling; cancer; outcomes

Conflict of Interest Statement:
Andrew M. Evens: Advisory board/honoraria: Pfizer, Incyte, Genentech,and CRISPR Therapeutics; Consultant/honoraria (Data and Safety Monitoring Board): Novartis and Pharmacyclics

Martin Hutchings: Advisory boards/honoraria: AbbVie, AstraZeneca, Genmab, Johnson&Johnson, Merck, Roche, Takeda. Research support (institution): AbbVie, AstraZeneca, Bristol Myers-Squibb, Celgene, Genentech, Genmab, Incyte, Johnson&Johnson, Merck, Novartis, Roche, Takeda

Kerry J. Savage: KJS Honoraria/consulting BMS, Roche, Seagen, Abbvie; Research support BMS; Steering committee Corvus; DSMC Regeneron

Authorship:
First draft of manuscript was written by Angie Mae Rodday, Andrew M. Evens, Susan K. Parsons, and John Radford

Abstract: 300/300
Word count for original article: 2882/3500
Tables/figures: 5/5
References: 34/40

Abstract

Background: A predictive model for early-stage classic Hodgkin lymphoma (cHL) does not exist. Leveraging patient-level data from large clinical trials and registries, we developed and validated the Early-stage cHL International Prognostication Index (E-HIPI) to predict 2-year progression-free survival (PFS).

Methods: The model was developed using Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines in 3000 adults with newly-diagnosed early-stage cHL from four international phase III clinical trials conducted from 1994-2011. External validation was performed in two cohorts, totaling 2360 treated patients from five international cHL registries (1996-2019). Two-year PFS was estimated with a Cox model with pre-treatment variables selected using backwards elimination. Internal validation corrected for overfitting. External validation assessed discrimination and calibration. The final model was also compared against EORTC favorable/unfavorable status.

Results: Median age in the development cohort was 31.2 years; 77.4% had stage II disease. The estimated 2-year PFS was 93.7%. Final variables retained in the model were sex and *continuous values* of maximum tumor diameter (MTD), hemoglobin, and albumin. The optimism-corrected C-statistic in the development cohort was 0.63 (95%CI, 0.60-0.69). Two-year PFS was lower in validation cohorts 1 (90.3%) and 2 (91.6%). In validation cohort 1, the C-statistic was 0.63, the calibration slope was near 1, but overall calibration indicated underprediction, which improved by updating the intercept. Performance was similar in validation cohort 2. In addition, higher-risk E-HIPI scores were associated with worse outcomes within both the EORTC unfavorable and favorable subgroups. When included altogether in one Cox model, the E-HIPI was associated with PFS, while EORTC favorable/unfavorable status was not. Online risk calculators were developed (https://rtools.mayo.edu/holistic_ehipi/).

Conclusions: Utilizing objective, continuous, and readily available variables, we developed and validated a new prediction model for early-stage cHL. Male sex, lower hemoglobin or albumin, and higher MTD were associated with worse PFS.

INTRODUCTION

Classic Hodgkin lymphoma (cHL) predominantly occurs in younger adults, often presenting at early-stage, and generally associated with excellent outcomes.[1] However, there is no global consensus regarding the optimal treatment approach beyond use of multi-agent chemotherapy with or without radiotherapy. Unlike in advanced-stage cHL,[2,3] a unified prognostic index or clinical prediction model for patient outcomes does not exist in early-stage cHL. Given the excellent survival outcomes, an early-stage cHL prediction model could identify patients at different risks of progression or relapse and potentially guide treatment selection.

Early-stage cHL has been classified into favorable and unfavorable subgroups based on the absence or presence of prognostic factors (e.g., large mediastinal mass, number of nodal groups, B symptoms, etc) for the past 40-50 years. The European Organization for Research and Treatment of Cancer (EORTC) prognostication data stem from historical clinical trials of patients treated from 1964 to 1982 when staging laparotomy and treatment with mantle field radiation alone were frequently used.[4] Also, the factors used in this classification are dichotomized, subjective, or difficult to measure; and they have not been calibrated for accuracy of absolute risk estimates. Furthermore, classification schemes vary across the world and include different factors or criteria (i.e., German Hodgkin Lymphoma Study Group (GHSG), the EORTC, and the Canadian Cancer Trials Group (CCTG, formerly National Cancer Institute of Canada).[5]

The *HoLISTIC* (Hodgkin Lymphoma International Study for Individual Care) consortium consists of detailed individual patient data obtained from seminal, phase III, international cHL clinical trials and international cHL registries that have been harmonized using a common data model with detailed data dictionary.[6] Harnessing this rich multi-source data and employing modern methods, we developed and validated a clinical prediction model, known as the Early-stage cHL International Prognostication Index (E-HIPI), including

continuous forms of objective variables to predict 2-year progression-free survival (PFS).

METHODS

Data sources and study population

We obtained individual patient data through formal data-sharing agreements with international cHL clinical trial groups and cHL registries. Once the data were harmonized, the model was developed on 3000 patients with newly diagnosed cHL treated on four international early-stage phase III clinical trials conducted from 1994-2011 (***Table S1***).[7-11] External validation was initially performed on 1488 patients with early-stage cHL treated with curative intent using multi-agent chemotherapy (e.g., doxorubicin/Adriamycin, bleomycin, vinblastine, and dacarbazine [ABVD]; bleomycin, etoposide, doxorubicin/Adriamycin, cyclophosphamide, vincristine/Oncovin, procarbazine, prednisone [BEACOPP]; , Stanford V, etc) from four major cancer registries (BC Cancer, Princess Margaret Cancer Centre, Iowa/Mayo SPORE, Stanford Registry) from 1996-2019 ("validation cohort 1") (***Table S2***).[12-15] As a sensitivity analysis, a second external validation was performed on 872 early-stage cHL patients from the Danish National Lymphoma Registry from 1996-2019 ("validation cohort 2") (***Table S2***).[16]

Development and validation cohorts were restricted to patients aged 18-65 years with stages I or II supra-diaphragmatic disease as this constituted the majority of patients enrolled in the early-stage cHL clinical trials. To reflect contemporary treatment approaches, patients who received only radiotherapy without multi-agent chemotherapy were excluded. The validation cohorts excluded registry patients who were enrolled on trials utilized in the model development.

Outcome

The primary outcome was 2-year PFS. Two years was chosen because most relapses occur within this timeframe in early-stage cHL [7,9,10,17] Time was defined as days from registration (clinical trials) or pathologic diagnosis (registries) to the event; censoring occurred if patients

were lost to follow-up or at two years. PFS events were defined as progression, relapse, or death from any cause. We used the adjudication of primary outcomes from the original data source.

Candidate Predictors

We considered the following baseline predictors based on clinical relevance and availability from the clinical trials data: age, sex, stage (I, II), histology (lymphocyte depleted, lymphocyte rich, mixed cellularity, nodular sclerosis, not otherwise specified (biopsy too limited to define cHL subtype), maximum tumor diameter (MTD), nodal groups, white blood cell (WBC) count, absolute lymphocyte count (ALC), hemoglobin, albumin, and erythrocyte sedimentation rate (ESR). Nodal groups were defined by EORTC criteria;[17] the primary analysis treated count of nodal groups as a continuous variable, but we also explored it as a categorical or binary variable.

Laboratory values were standardized using their mean and standard deviation. B symptoms were not included as a candidate predictor because of concerns about reliability and reproducibility; however, as a sensitivity analysis, we added B symptoms and re-evaluated model performance. We excluded patients with missing data on >50% of the candidate predictors. Multiple imputation was performed to address missing data on candidate predictors[18] (***Table S3***). Linearity of continuous variables was assessed based on partial residual plots and fitting-penalized smoothing splines.

Statistical analysis

We followed the TRIPOD reporting guidelines[19] (***Table S5***). The purpose of our model was to predict patient prognosis based on 2-year PFS using a Cox proportional hazards (PH) model. Discrimination and calibration were used to assess model performance in the development and validation cohorts. Discrimination was assessed using Harrell's C-statistic.[20,21]

Calibration was assessed by comparing observed and predicted probabilities of 2-year PFS within quintiles of predicted probabilities and estimating overall calibration (predicted minus observed) and calibration slopes.

First, variables were selected from the development cohort using backward elimination ($p<0.1$). Second, we performed internal validation to obtain a shrinkage factor for the final model coefficients to decrease the risk of overfitting and correct the development C-statistic for optimism. Third, as part of internal-external validation, we performed cross-validation using each trial in the development cohort to refit the model and assess the C-statistic within the included and omitted trials. Finally, the baseline hazard for the "average patient" and the final model coefficients (after applying the shrinkage factor) were applied to patients in the external validation cohorts. Discrimination and calibration were evaluated in validation cohort 1. Due to lower 2-year PFS in validation cohort 1 and the resulting underestimation of PFS events, we performed a sensitivity analysis updating the model intercept using validation cohort 1 and assessed performance in both validation cohorts 1 and 2. Additional methods for model development and validation, comparison of the E-HIPI with EORTC favorable/unfavorable status, and sensitivity analyses (e.g., stratified baseline hazards by key subgroups, model building restricted to those without bulky disease) are in the ***Supplemental Methods (pages 4-6)***.

RESULTS

Patient characteristics

See ***Table 1*** for detailed patient and disease characteristics for the development and validation cohorts, including comment on the representativeness of the patient sample and the generalizability findings. Consort flow diagrams for the development and validation cohort eligibility are detailed in ***Figure S1***. The median follow-up in the development cohort was 60

months (Q1=45, Q3=75), and by 2 years, 38 patients had died from any cause and 185 had experienced progression, relapse, or death.

E-HIPI development

The Kaplan-Meier (KM) estimator for 2-year PFS was 93.7% (95% CI: 92.9, 94.6). Univariable associations between candidate predictors and 2-year PFS are reported in ***Table S4***, ***Figure 1***, and ***Figures S2-S5***. No continuous variables violated the linearity assumption (***Figure 1***, ***Figures S6-S10***). The following predictors were eliminated from the model: stage, histology, count of nodal groups (continuous, categorical, or binary), and continuous values of age, WBC count, lymphocyte count, and ESR. The inclusion of B symptoms did not improve model performance in sensitivity analysis *(**Supplemental Results***, page 7).

Sex, MTD, albumin, and hemoglobin were retained in the model (***Table 2***). Female sex and both higher hemoglobin and albumin were predictive of better PFS, while higher MTD was predictive of worse PFS. Continuous, categorical and binary versions of count of nodal groups were not retained in the model (***Table S4, Figure S7***). See ***Figure S11*** for the distribution of predicted risk in the development cohort. The C-statistic in the development cohort was 0.650 (95% CI: 0.607, 0.690), with an optimism-corrected value of 0.629 (95% CI: 0.602, 0.689). Based on internal-external validation, C-statistics in the included trials ranged from 0.618 to 0.657 and C-statistics in the omitted trial ranged from 0.484 to 0.703 (***Table S6***).

Stratified baseline hazards for the trials in the development cohort were similar (0.932 to 0.953); as were those stratified by presence of bulky disease (no bulk=0.942, bulk=0.955) (***Table S7)***. When model building was repeated in the development cohort restricted to those without bulky disease, sex, MTD and hemoglobin were retained (***Table S8***).

External validation

Baseline characteristics and outcomes of validation cohort 1 were similar to the development cohort except for more stage II disease, more NOS histology, longer follow-up (median 108 months, Q1=64, Q3=165), and lower 2-year PFS (90.3%, 95% CI: 88.8%, 91.8%) (***Table 1***). The C-statistic for the E-HIPI model was 0.626 (95% CI: 0.583, 0.669) in validation cohort 1. The predicted PFS event distribution and the observed outcomes stratified by tertile of predicted risk in validation cohort 1 are shown in ***Figure 2***. The calibration plot in the validation cohort 1 is shown in ***Figure 3A***. Overall calibration was -3.6% and the calibration slope was 0.96 (95% CI: 0.59, 1.33). After updating the E-HIPI intercept in validation cohort 1 to account for underprediction of events, the overall calibration improved to -0.1% and the plots showed better calibration (***Figure 3B***).

Baseline characteristics and outcomes of validation cohort 2 were similar to the development cohort except for fewer females, more mixed cellularity histology and less NOS pathology, smaller MTD, longer follow-up (median 150 months, Q1=95, Q3=215), and slightly lower 2-year PFS (91.6%, 95% CI: 89.9%, 93.5%) (***Table 1***). When applying the E-HIPI with the updated intercept to validation cohort 2, the C-statistic was 0.593 (95% CI: 0.526, 0.659). Overall calibration was 0.6% and the calibration slope was 0.75 (95% CI: 0.19, 1.30) in validation cohort 2 (***Figure S12***).

Comparison of E-HIPI with EORTC favorable/unfavorable status in development cohort

Higher risk E-HIPI scores were associated with worse outcomes when considered within either the EORTC favorable (HR=1.34, 95% CI: 1.14, 1.57) or unfavorable (HR=1.13, 95% CI: 1.08, 1.19) subgroups. Additionally, the inclusion of favorable/unfavorable status did not improve the performance of the E-HIPI in the development cohort; when the E-HIPI and EORTC status were included together in a Cox PH model, the E-HIPI was associated with 2-year PFS (HR=1.15, 95% CI: 1.10-1.20), while EORTC favorable/unfavorable status was not (HR=1.18, 95% CI: 0.83, 1.96) (***Table S8***).

Online Risk Calculator

An interactive online tool for the E-HIPI that generates 2-year PFS in an individualized "risk calculator" as well as dynamic applications for "risk comparison" and "risk stratification" across user-defined cut-points can be found at https://rtools.mayo.edu/holistic_ehipi/.

DISCUSSION

The E-HIPI, as developed and validated on nearly 5,400 patients, provides a novel clinical prediction model for early-stage cHL. Following rigorous TRIPOD recommendations,[19] we identified four objective prognostic factors determined at the time of diagnosis, i.e., sex (male/female), with MTD, albumin, and hemoglobin considered as continuous variables. This resulted in a clinically meaningful prediction model across a diverse cohort of patients with early-stage cHL treated with modern multiagent chemotherapy regimens with or without radiation. Notably, the E-HIPI included patients from multiple countries and continents, further supporting its worldwide utility for clinical prognosis and the design of prospective clinical trials.

The original EORTC early-stage prognostication analyses were done in the early 1970s when staging laparotomy was still commonly performed for patient staging,[22,23] and the majority of patients from the original multivariable prognostic classification were treated with mantle field radiation alone without chemotherapy.[4] Additionally, there are varying early-stage classification schemes (e.g., GHSG, EORTC, CCTG) that characterize "favorable" or "unfavorable" disease that use different clinical factors and criteria that can cause confusion amongst clinicians and patients.[5] In 2013, the GHSG analyzed the different early-stage cHL classification schemes and found that age was not a significant factor, extranodal disease and ESR had borderline significance, while bulk mediastinal mass and number of node sites were prominent factors.[24] Furthermore, there was low specificity with high rates of false-positive results in all classification schemes (1-specificity of 53-55%).

In addition, the clinical factors used in current classification schemes are either subjective, e.g., B symptoms, or inconsistently measured, e.g., count of nodal groups, and all factors are dichotomized in these models. Dichotomization of variables is discouraged in prognostic models as it diminishes statistical power and may conceal clinically meaningful non-linear associations.[25] Indeed, we found that EORTC favorable/unfavorable status was not prognostic once the E-HIPI was known, and that the E-HIPI was prognostic within both favorable and unfavorable subgroups. Given concerns about reliability and reproducibility of B symptoms, we chose not to include it as a candidate predictor in the E-HIPI; sensitivity analysis found its inclusion did not improve model performance. Although the count of nodal groups was considered as an E-HIPI candidate predictor, it was not associated with PFS whether analyzed as a continuous, categorical, or binary variable.

Sex, MTD, albumin, and hemoglobin emerged as important pre-treatment factors for predicting 2-year PFS in the E-HIPI. While sex has not emerged on previous analyses for early-stage cHL, male sex is a prominent adverse factor in advanced-stage cHL, as reported from the International Prognostic Score and the more recent and robust Advanced-Stage Hodgkin Lymphoma International Prognostic Index (A-HIPI).[2,3] The causality of this association is unclear. In a prior GHSG analysis, female patients with cHL had more prominent leukopenia *during* treatment versus males, which was associated with improved freedom from treatment failure.[26] Other sex-dependent genomic or single nucleotide polymorphisms may be contributing factors.

Presence of bulky disease has been considered a poor prognostic factor in early-stage cHL for decades.[27-29] The importance of bulky disease has been maintained when including either computerized tomography or positron emission tomography (PET) imaging, using varying cut-points (e.g., 7-10cm). Unlike prior dichotomization of bulky disease as a risk factor, the E-HIPI includes continuous MTD, which revealed a striking linear association of worsening

prognosis with each 1-centimeter increase. The prognostic impact of MTD in early-stage cHL was described in the UK RAPID trial[30] and subsequently validated in data from the RAPID and the EORTC/LYSA H10 trials.[31] These analyses, as well as the E-HIPI development, demonstrate that each centimeter of bulk is critical. In a sensitivity analysis, we found that continuous MTD was still highly predictive even among those without bulky disease, i.e., patients with a maximum lesion diameter of less than 10cm. There is considerable interest in the prognostic power of PET-derived metabolic tumor volume (MTV).[31] Studies comparing the predictive performance of MTV and MTD are needed. The latter has the advantage of simplicity and applicability to geographic regions where PET facilities are less accessible.

The other two factors in the E-HIPI were albumin and hemoglobin. These also were associated with outcomes in advanced-stage cHL models.[2,3] However, they have not been studied in patients with early-stage disease. Despite having radiographic evidence of early-stage cHL, the presence of anemia or hypoalbuminemia may reflect underlying cytokine-mediated inflammation, which also induces changes in the tumor microenvironment.[32,33] Additionally, these laboratory changes may be a marker of coexistent patient comorbidities affecting tolerance to therapy.

Limitations of these analyses are the E-HIPI's modest performance represented by discrimination and calibration. However, the E-HIPI outperformed existing favorable/unfavorable classification. Given the low event rate in early-stage cHL, it may be difficult to identify patients at risk for worse outcomes. In addition, as evidenced by low discrimination in internal-external validation in the HD6 trial (despite similar baseline hazards across trials), small studies with few events are more susceptible to random variation. Our goal was to develop the model in an idealized setting (i.e., clinical trial patients) and then to subsequently assess model performance in an external "real-world" validation dataset as part of one analysis (i.e., group/country registries), which could be efficiently accomplished as part of the global HoLISTIC Consortium.

The higher event rates among patients from the real-world registries compared to those in clinical trials used in model development required updating the model intercept (not the coefficients) to improve calibration. The purpose of updating the intercept was to show that the model could be updated for a given setting, not to provide a new intercept for use in all other settings. For applications in settings with different event rates than our development cohort, the model may require updating to improve calibration. We excluded patients aged >65 years because of the paucity of older adults treated in the clinical trials utilized for development. For the previously published A-HIPI model, it was compared with advanced machine learning-based prognostic models.[34] However, both models were shown to have similar statistical strength. Nonetheless, we speculate that future analyses that incorporate machine learning methodology with the E-HIPI model could enhance its predictive ability.

As should be standard practice in predictive modeling, we *did not* consider post-baseline factors such as treatments received or imaging results. However, the HoLISTIC consortium dataset and associated models have been built to allow for the incorporation of new studies and baseline data as they emerge, which will allow for re-calibration of models over time.

In conclusion, harnessing a large, multi-source, international database of recent clinical trials and cancer registries, we developed and validated a clinical prediction model using data from 1994-2019, the E-HIPI, for adult patients under age ages 65 years with early-stage cHL. We identified novel linear relationships of lower hemoglobin and albumin and higher MTD with worse PFS and established male sex as an adverse risk factor. Crucially, each of these variables is objective and easily measurable, meaning the E-HIPI is reproducible and applicable in most healthcare settings. To enhance the use of the E-HIPI, we also developed an interactive online calculator to assist clinicians and patients in estimating individualized prognosis and risk comparisons.

Funding: NCI R01 CA 262265-04 (**Modeling Multi-Source Data in Hodgkin Lymphoma**, Parsons/Evens, MPI)

Data availability statement: The harmonized data that support the findings of this study are not available for third-party distribution according to existing data use agreements. Data from individual trials and registries may be available directly from the source entities.

References

1. Dores GM, Curtis RE, Dalal NH, Linet MS, Morton LM. Cause-Specific Mortality Following Initial Chemotherapy in a Population-Based Cohort of Patients With Classical Hodgkin Lymphoma, 2000-2016. J Clin Oncol 2020;38(35):4149-4162. DOI: 10.1200/JCO.20.00264.
2. Hasenclever D, Diehl V. A prognostic score for advanced Hodgkin's disease. International Prognostic Factors Project on Advanced Hodgkin's Disease. N Engl J Med 1998;339(21):1506-14. DOI: 10.1056/NEJM199811193392104.
3. Rodday AM, Parsons SK, Upshaw JN, et al. The Advanced-Stage Hodgkin Lymphoma International Prognostic Index: Development and Validation of a Clinical Prediction Model From the HoLISTIC Consortium. J Clin Oncol 2023;41(11):2076-2086. DOI: 10.1200/JCO.22.02473.
4. Tubiana M, Henry-Amar M, van der Werf-Messing B, et al. A multivariate analysis of prognostic factors in early stage Hodgkin's disease. Int J Radiat Oncol Biol Phys 1985;11(1):23-30. DOI: 10.1016/0360-3016(85)90358-x.
5. Hernandez-Ilizaliturri FJ.. Hodgkin lymphoma staging. Medscape. (https://emedicine.medscape.com/article/2007081-overview?form=fpf).
6. (https://www.hodgkinconsortium.com/).Accessed May 22, 2025.
7. Meyer RM, Gospodarowicz MK, Connors JM, et al. Randomized comparison of ABVD chemotherapy with a strategy that includes radiation therapy in patients with limited-stage Hodgkin's lymphoma: National Cancer Institute of Canada Clinical Trials Group and the Eastern Cooperative Oncology Group. J Clin Oncol 2005;23(21):4634-42. DOI: 10.1200/JCO.2005.09.085.
8. Meyer RM, Gospodarowicz MK, Connors JM, et al. ABVD alone versus radiation-based therapy in limited-stage Hodgkin's lymphoma. N Engl J Med 2012;366(5):399-408. DOI: 10.1056/NEJMoa1111961 [doi].
9. Ferme C, Thomas J, Brice P, et al. ABVD or BEACOPP(baseline) along with involved-field radiotherapy in early-stage Hodgkin Lymphoma with risk factors: Results of the European Organisation for Research and Treatment of Cancer (EORTC)-Groupe d'Etude des Lymphomes de l'Adulte (GELA) H9-U intergroup randomised trial. Eur J Cancer 2017;81:45-55. DOI: 10.1016/j.ejca.2017.05.005.
10. Radford J, Illidge T, Counsell N, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. N Engl J Med 2015;372(17):1598-1607. DOI: 10.1056/NEJMoa1408648 [doi].
11. Andre MPE, Girinsky T, Federico M, et al. Early Positron Emission Tomography Response-Adapted Treatment in Stage I and II Hodgkin Lymphoma: Final Results of the Randomized EORTC/LYSA/FIL H10 Trial. J Clin Oncol 2017;35(16):1786-1794. DOI: 10.1200/JCO.2016.68.6394.
12. Hapgood G, Zheng Y, Sehn LH, et al. Evaluation of the Risk of Relapse in Classical Hodgkin Lymphoma at Event-Free Survival Time Points and Survival Comparison With the General Population in British Columbia. J Clin Oncol 2016;34(21):2493-500. DOI: 10.1200/JCO.2015.65.4194.
13. Hahn E, Jiang H, Ng A, et al. Late Cardiac Toxicity After Mediastinal Radiation Therapy for Hodgkin Lymphoma: Contributions of Coronary Artery and Whole Heart Dose-Volume Variables to Risk Prediction. Int J Radiat Oncol Biol Phys 2017;98(5):1116-1123. DOI: 10.1016/j.ijrobp.2017.03.026.

14. Cerhan JR, Link BK, Habermann TM, et al. Cohort Profile: The Lymphoma Specialized Program of Research Excellence (SPORE) Molecular Epidemiology Resource (MER) Cohort Study. Int J Epidemiol 2017;46(6):1753-1754i. DOI: 10.1093/ije/dyx119.
15. Weil CR, Qian Y, Von Eyben R, et al. Long-term outcomes of patients with unfavorable stage I-II classic Hodgkin lymphoma treated with Stanford V chemotherapy and limited field irradiation. Leuk Lymphoma 2020;61(10):2428-2434. DOI: 10.1080/10428194.2020.1768385.
16. Arboe B, Josefsson P, Jorgensen J, et al. Danish National Lymphoma Registry. Clin Epidemiol 2016;8:577-581. DOI: 10.2147/CLEP.S99470.
17. Raemaekers JM, Andre MP, Federico M, et al. Omitting radiotherapy in early positron emission tomography-negative stage I/II Hodgkin lymphoma is associated with an increased risk of early relapse: Clinical results of the preplanned interim analysis of the randomized EORTC/LYSA/FIL H10 trial. J Clin Oncol 2014;32(12):1188-1194. DOI: JCO.2013.51.9298 [pii];10.1200/JCO.2013.51.9298 [doi].
18. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011;30(4):377-99. DOI: 10.1002/sim.4067.
19. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med 2015;162(10):735-6. DOI: 10.7326/L15-5093-2.
20. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247(18):2543-6. (https://www.ncbi.nlm.nih.gov/pubmed/7069920).
21. Harrell J, F.E. Regression modeling strategies: With applications to linear models, logistic regression and survival analysis. New York: Springer, 2001.
22. Carbone PP, Kaplan HS, Musshoff K, Smithers DW, Tubiana M. Report of the Committee on Hodgkin's Disease Staging Classification. Cancer Res 1971;31(11):1860-1. (https://www.ncbi.nlm.nih.gov/pubmed/5121694).
23. Tubiana M, Attie E, Flamant R, Gerard-Marchant R, Hayat M. Prognostic factors in 454 cases of Hodgkin's disease. Cancer Res 1971;31(11):1801-10. (https://www.ncbi.nlm.nih.gov/pubmed/5121686).
24. Klimm B, Goergen H, Fuchs M, et al. Impact of risk factors on outcomes in early-stage Hodgkin's lymphoma: an analysis of international staging definitions. Ann Oncol 2013;24(12):3070-6. DOI: 10.1093/annonc/mdt413.
25. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332(7549):1080. DOI: 10.1136/bmj.332.7549.1080.
26. Klimm B, Reineke T, Haverkamp H, et al. Role of hematotoxicity and sex in patients with Hodgkin's lymphoma: an analysis from the German Hodgkin Study Group. J Clin Oncol 2005;23(31):8003-11. DOI: 10.1200/JCO.2005.205.60.
27. Mauch P, Goodman R, Hellman S. The significance of mediastinal involvement in early stage Hodgkin's disease. Cancer 1978;42(3):1039-45. DOI: 10.1002/1097-0142(197809)42:3<1039::aid-cncr2820420302>3.0.co;2-r.
28. Schomberg PJ, Evans RG, O'Connell MJ, et al. Prognostic significance of mediastinal mass in adult Hodgkin's disease. Cancer 1984;53(2):324-8. DOI: 10.1002/1097-0142(19840115)53:2<324::aid-cncr2820530225>3.0.co;2-e.
29. Mendenhall NP, Cantor AB, Barre DM, Lynch JW, Jr., Million RR. The role of prognostic factors in treatment selection for early-stage Hodgkin's disease. Am J Clin Oncol 1994;17(3):189-95. DOI: 10.1097/00000421-199406000-00002.

30. Illidge TM, Phillips EH, Counsell N, et al. Maximum tumor diameter is associated with event-free survival in PET-negative patients with stage I/IIA Hodgkin lymphoma. Blood Adv 2020;4(1):203-206. DOI: 10.1182/bloodadvances.2019001023.
31. Phillips E, Counsell, N, Illidge TM, Andre M, Aurer I, Fiaccadori V, Fortpied C, Neven A, Federico M, Barrington SF, Raemaekers JM, Radford J. Maximum Tumor Diameter is Associated with Relapse Risk in Limited-Stage Hodgkin Lymphoma: An International Study. Blood Adv 2025. DOI: https://doi.org/10.1182/bloodadvances.2024015140.
32. Hohaus S, Massini G, Giachelia M, et al. Anemia in Hodgkin's lymphoma: the role of interleukin-6 and hepcidin. J Clin Oncol 2010;28(15):2538-43. DOI: 10.1200/JCO.2009.27.6873.
33. da Silva PB, Perini GF, Pereira Lde A, et al. Imbalance of Pro- and Anti-Inflammatory Cytokines in Patients With cHL Persists Despite Treatment Compared With Control Subjects. Clin Lymphoma Myeloma Leuk 2015;15 Suppl:S151-7. DOI: 10.1016/j.clml.2015.02.002.
34. Rask Kragh Jorgensen R, Bergstrom F, Eloranta S, et al. Machine Learning-Based Survival Prediction Models for Progression-Free and Overall Survival in Advanced-Stage Hodgkin Lymphoma. JCO Clin Cancer Inform 2024;8:e2300255. DOI: 10.1200/CCI.23.00255.

Table 1. Baseline characteristics for development and validation cohorts.[a, b]

	Development Cohort, n=3000	Validation Cohort 1, n=1488	Validation Cohort 2, n=872
Study/Registry [c]			
HD6	266 (8.9%)	--	--
H9U	676 (22.5%)	--	--
RAPID	341 (11.4%)	--	--
H10	1717 (57.2%)	--	--
BC Cancer	--	596 (40.1%)	--
Iowa/Mayo	--	227 (15.3%)	--
PMR	--	325 (21.8%)	--
Stanford Registry	--	340 (22.8%)	--
LYFO	--	--	872 (100.0%)
Age (years), mean (SD)	34.4 (12.2)	34.0 (11.8)	37.5 (13.6)
Age (years), median (Q1, Q3)	31.2 (24.2, 43.0)	31.0 (25.0, 42.0)	35.0 (26.0, 49.0)
Female sex, n (%)	1526 (50.9%)	795 (53.4%)	401 (46.0%)
Stage, n (%)			
I	678 (22.6%)	215 (14.4%)	214 (24.5%)
II	2322 (77.4%)	1273 (85.6%)	658 (75.5%)
B symptoms, n (%)	795 (26.5%)	404 (27.2%)	302 (34.7%)
Stage, n (%)			
I	678 (22.6%)	215 (14.4%)	214 (24.5%)
IIA	1651 (55.0%)	894 (60.1%)	397 (45.6%)
IIB	671 (22.4%)	379 (25.5%)	261 (29.9%)
Histology, n (%)			
Lymph Depleted	18 (0.6%)	1 (0.1%)	5 (0.6%)
Lymph Rich	90 (3.0%)	50 (3.4%)	50 (5.7%)
Mixed Cellularity	398 (13.3%)	103 (6.9%)	170 (19.5%)
Nodular Sclerosis	2435 (81.2%)	1166 (78.4%)	549 (63.0%)
NOS	59 (2.0%)	168 (11.3%)	98 (11.2%)
MTD (cm), mean (SD)	6.5 (3.5)	6.5 (3.3)	5.3 (2.8)
MTD (cm), median (Q1, Q3)	5.7 (3.5, 9.0)	6.0 (4.0, 9.0)	4.8 (3.3, 6.0)

Bulky disease (≥10 cm)	584 (19.5%)	325 (21.9%)	79 (9.1%)
Count of nodal groups, median (Q1, Q3)	2.0 (1.0, 3.0)	2.0 (2.0, 3.0)	Not collected
WBC count (10^3/uL), mean (SD)	10.0 (4.0)	9.2 (3.8)	9.5 (3.6)
Lymphocyte count (10^3/uL), mean (SD)	1.6 (0.7)	1.5 (0.7)	1.7 (0.7)
Hemoglobin (g/dL), mean (SD)	13.0 (1.6)	13.1 (1.6)	13.4 (1.8)
Albumin (g/dL), mean (SD)	4.2 (0.5)	4.0 (0.5)	4.1 (0.5)
ESR (mm/hr), mean (SD)	37.5 (29.8)	32.9 (28.8)	33.3 (31.2)

Abbreviations: BC, British Columbia; ESR, erythrocyte sedimentation rate; LYFO, Danish National Lymphoma Registry; MTD, maximum tumor diameter; NOS, not otherwise specified; PMR, Princess Margaret cancer center Registry; Q1, quartile 1; Q3, quartile 3; SD, standard deviation; WBC, white blood cell.

[a] Summary statistics after multiple imputation. See ***Table S3*** for missingness prior to multiple imputation.

[b] This is a heterogeneous global database and there are not full comparable population-level data available internationally; however, the age and sex-based data herein are representative of the broader population affected in the United States (https://seer.cancer.gov/statfacts/html/hodg.html); there were not data based on race collected on the majority of data herein.

[c] Number of patients by trial and registry refer to the subset of patients who met the eligibility criteria (***Figure S1***).

Table 2. Model parameters for 2-year PFS prediction model.[a]

	Beta Coefficient [b]	**HR [b]**	**Optimism-Corrected Beta Coefficient [b]**
Female sex	-0.529	0.59	-0.437
MTD per 1 centimeter	0.090	1.09	0.074
Hemoglobin per 1 SD	-0.191	0.83	-0.158
Albumin per 1 SD	-0.181	0.83	-0.149

Abbreviations: HR, hazard ratio; MTD, maximum tumor diameter; PFS, progression-free survival; SD, standard deviation.

[a] Variables dropped from the model are not included (i.e., age, stage (I vs. II), histology (lymphocyte depleted, lymphocyte rich, mixed cellularity, not otherwise specified, and nodular sclerosis), extranodal site, number of nodal sites/groups, white blood cell count, absolute lymphocyte count, and erythrocyte sedimentation rate); every variable was analyzed as a continuous, categorical or binary variable. For hemoglobin and albumin, the interpretation is standardized for a 1-SD increase.

[b] Beta coefficients and HRs are from the multivariable model after backwards elimination that was fit in the development cohort. Optimism-corrected beta coefficients were derived using internal validation to obtain shrinkage factors for the beta coefficients to decrease the risk of overfitting. We do not provide 95% confidence intervals (CIs) for hazard ratios (HRs) because we used variable selection, and our goal was prediction rather than statistical inference.