The Advanced-Stage Hodgkin Lymphoma International Prognostic Index: Development and Validation of a Clinical Prediction Model From the HoLISTIC Consortium

The Advanced-Stage Hodgkin Lymphoma International Prognostic Index (A-HIPI): A Report from HoLISTIC Consortium

Authors: Angie Mae Rodday, PhD, MS,[1*] Susan K. Parsons, MD, MRP,[1*] Jenica N. Upshaw, MD, MSc[1,2], Jonathan W Friedberg, MD[3] Andrea Gallamini, MD[4] Eliza Hawkes, MBBS, DMedSc[5] David Hodgson, MD[6] Peter Johnson, MD[7] Brian K. Link, MD[8] Eric Mou, MD[8] Kerry J. Savage, MD MSc[9] Pier Luigi Zinzani, MD, PhD[10] Matthew Maurer, DMSc, MS[11] Andrew M. Evens, DO, MBA, MSc[12]

Affiliations: [1]Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA; [2]The CardioVascular Center and Advanced Heart Failure Program, Tufts Medical Center, Boston, MA; [3]James P. Wilmot Cancer Center, University of Rochester Medical Center, Rochester, NY; [4]Research and Clinical Innovation Department, Antoine Lacassagne Cancer Center, Nice, France; [5]Australasian Lymphoma and Related Diseases Registry, Monash University, Melbourne, Australia; [6]Department of Radiation Oncology, Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada; [7] School of Cancer Sciences, Faculty of Medicine, University of Southampton, United Kingdom; [8]Division of Hematology, Oncology, and Blood & Marrow Transplantation, University of Iowa, Iowa City, IA; [9]Centre for Lymphoid Cancer, BC Cancer, Vancouver, BC, Canada; [10]IRCCS Azienda Ospedaliero-Universitaria di Bologna, Istituto di Ematologia “Seragnoli” Dipartimento di Medicina Specialistica, Diagnostica Sperimentale Università di Bologna, Bologna, Italy; [11]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN; [12]Division of Blood Disorders, Rutgers Cancer Institute New Jersey, New Brunswick, NJ

**Co-first authorship.*

Running head: Hodgkin International Prognostic Index

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

Corresponding Author: Andrew M. Evens, DO, MBA, MSc
Professor of Medicine, Rutgers Robert Wood Johnson Medical School
Associate Vice Chancellor, Clinical Innovation and Data Analytics, Rutgers Biomedical and Health Sciences
Associate Director, Clinical Services, Rutgers Cancer Institute of New Jersey
195 Little Albany Street, New Brunswick, New Jersey 08903-2681
Phone: 732-235-9289
Email: ae378@cinj.rutgers.edu

Manuscript word count: 3000

Abstract word count: 275

References: 40

Number of Tables: 3

Number of Figures: 2

Supplementary Appendix: 12 figures, 13 tables

Disclaimer: to be presented at the 64th ASH Annual Meeting and Exposition, Oral Presentation, December 10-13, 2022.

Abstract

Background: The International Prognostic Score (IPS) has been used in classic Hodgkin lymphoma (cHL) for 25 years. However, analyses have documented poor calibration of the IPS among contemporarily treated patients. Harnessing multi-source individual patient data from the Hodgkin Lymphoma International STudy for Individual Care *(*HoLISTIC) consortium, we developed and validated a modern prediction model.

Methods: Model development was performed on 4,022 newly-diagnosed advanced-stage adult cHL patients from eight international phase 3 clinical trials, conducted from 1996-2014. External validation was performed on 1,431 contemporaneously treated patients from four real-world cHL registries. To consider prognostic association over a full range of continuous variables, we evaluated piecewise linear splines for potential *non-linear* relationships. Five-year progression-free survival (PFS) and overall survival (OS) were estimated using Cox proportional-hazards models.

Results: Median age in the development cohort was 33 years (18-65); 55% were male; nodular sclerosis was the most common histology. Kaplan-Meier estimators were 0.77 for 5-year PFS and 0.92 for 5-year OS. Significant variables predictors included age, sex, stage, bulk, absolute lymphocyte count (ALC), hemoglobin, and albumin, with slight variation for PFS versus OS. Moreover, age and ALC yielded *non-linear* relationships with outcomes. Optimism-corrected c-statistics in the development model for 5-year PFS and OS were 0.590 and 0.720, respectively. There was good discrimination and calibration in external validation and consistent performance in internal-external validation. Compared with the IPS, there was improved discrimination for OS and improved calibration for PFS and OS.

Conclusion: We rigorously developed and externally validated a modern-day prediction model in >5,000 advanced-stage cHL patients. Furthermore, we identified several novel non-linear relationships and improved the prediction of patient outcomes. An online calculator was created for individualized point-of-care use.

Introduction

Classic Hodgkin lymphoma (cHL) is a B-cell malignancy that occurs predominantly in younger adults and is generally associated with favorable disease outcomes.[1] However, there is no single consensus-based or individualized treatment approach globally beyond use of multi-agent chemotherapy with curative intent.

The most widely used prognostication tool in cHL has been the International Prognostic Score (IPS), published in 1998 by the German Hodgkin Study Group.[2] The IPS identified seven clinical factors prognostic for survival at five years in newly-diagnosed advanced-stage disease. While used over the past 25 years, this score requires updating for several reasons. First, management strategies and outcomes for cHL have improved over the past 10-20 years.[1,3] Indeed, analyses documented poor calibration of the IPS7 among contemporarily treated patients.[4,5] Second, predictive modeling techniques have grown in sophistication since the IPS7 was developed. For example, the IPS7 relied on the dichotomous categorization of patient and disease factors with limited information about model performance (discrimination and calibration). Further, the IPS7 was developed based on complete case analyses, but missing data were frequent, requiring the use of imputation strategies that were not externally validated.

In 2018, we formed a trans-discipline international consortium, *HoLISTIC* (Hodgkin Lymphoma International STudy for Individual Care), consisting of worldwide clinical cHL experts, decision scientists, statisticians, epidemiologists, and patient advocates.[6] We subsequently obtained and harmonized individual patient data (IPD) from multiple seminal phase III international clinical trials in newly-diagnosed cHL, conducted from 1996-2014.[7] In addition, we incorporated detailed IPD from contemporaneously treated patients recorded in "real world" cHL registries across North America. Applying established data science methods,[8] we created a common data model with a detailed data dictionary. Data were normalized, standardized, and harmonized, resulting in the creation of a comprehensive, annotated master database of more

than 5,000 adult patients with newly-diagnosed, advanced-stage cHL.

Harnessing this rich multi-source data and employing modern methods, we sought to build upon the original IPS to develop and validate a robust, modern prediction model known as the Advanced-stage cHL International Prognostication Index (A-HIPI) to predict five-year progression-free survival (PFS) and overall survival (OS). An associated objective was to create a model at the individual patient level, based on continuous forms of data, and to harness a newly created online calculator.

METHODS

Data sources and study population

We obtained IPD obtained through formal data-sharing agreements with global cHL clinical trial groups and cancer registries. Once the data were harmonized, model development was performed on a cohort of 4,022 cHL newly-diagnosed patients treated on eight international advanced-stage phase III clinical trials from 1996-2014: HD9601, HD2000, UK Stanford V, ECOG2496, SWOG0816, RATHL, HD0801, HD0607 (***Table S1***).[9-16] External validation was performed in a separate cohort of 1,431 advanced-stage cHL patients treated with curative intent from four major cancer registries (BC Cancer, Princess Margaret Cancer Centre, Iowa/Mayo SPORE, Australia) from 1996-2019.[17-20] Registry patients enrolled on trials utilized in the model development were excluded from the validation cohort. Model development and validation were restricted to patients ages 18-65 years with stages IIB, III, or IV disease as this constituted the vast majority of patients enrolled in the cHL clinical trials.

Outcomes

The two primary outcomes were 5-year PFS and 5-year OS in part as most of cHL relapses occur within the first 5 years from diagnosis.[21] 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 five years. PFS events were defined as progression, relapse, or death. Adjudication of primary outcomes was performed by the data source and was not revised in the data harmonization process.

Candidate Predictors

We considered the following baseline predictors based on clinical relevance and standard data capture on the clinical trials: sex, stage (IIB, III, IV), B symptoms, histology (lymphocyte depleted, lymphocyte rich, mixed cellularity, nodular sclerosis, NOS), any bulk (by trial/registry definition: ***Table S2***), age at diagnosis, white blood cell count (WBC), absolute lymphocyte count (ALC), hemoglobin, albumin, and erythrocyte sedimentation rate (ESR). ESR had not been routinely collected across all trials, particularly among studies that did not include early-stage unfavorable disease. We excluded patients who were missing data on more than 50% of the candidate predictor variables. See ***Table S3*** for full details of missing data by variable. Multiple imputation was performed to address missing data on the remaining patients.[22] Stage was considered as a three-level categorical variable (IIB, III, IV) and a binary variable (IIB/III, IV). Linearity of continuous variables was assessed based on partial residual plots and fitting-penalized smoothing splines. Based on this, and to reduce the risk of overfitting when using restricted cubic splines, we created piecewise linear splines with one inflection point (i.e., knot) for relevant variables.

Statistical analysis

Separate models for PFS and OS were fit using Cox proportional hazards (PH) models. Backward elimination with $p<0.05$ was used to select variables to include in the prediction models. Because variable selection was used, and because the purpose of our modeling was parameter estimation rather than statistical inference, we do not provide 95% confidence intervals. Discrimination was assessed using Harrell's c-statistic.[23,24] Calibration was assessed

by comparing observed and predicted probabilities of 5-year outcomes within deciles of predicted probabilities and estimating calibration intercepts and slopes. We also performed internal validation to obtain shrinkage factors for the final model coefficients to decrease the risk of overfitting.

As part of internal-external validation, we performed cross-validation on the model development cohort, where each clinical trial was left out 'one at a time' to account for between-trial heterogeneity (e.g., use of baseline imaging, definitions of bulk, treatment regimen), to refit the model and assess its performance (using the c-statistic) within the excluded trial. Final model coefficients, including the baseline hazard for 5-year PFS and OS for the "average patient", were applied to the external validation cohort of cancer registries. Discrimination and calibration were evaluated in the validation cohort as above. We also examined performance of the IPS7 and IPS3 in our validation cohort. Sensitivity analyses included imputation of out-of-range lab values and stratification by clinical trial in model development. Additional methods for model development and validation are in the ***Supplement***.

Results

Patient characteristics

The median age in the development cohort was 33 years. See ***Table 1*** for all other patient and disease characteristics. The median follow-up was 62 months (IQR 36-90), and by 5 years, 858 patients experienced progression, relapse, or death (21%), and 278 died (7%). Kaplan Meier estimators at 5 years were 77% (95% CI: 76%-78%) for PFS and 92% (95% CI: 91%-93%) for OS. Consort flow diagrams for the development and validation cohort eligibility are detailed in ***Figure S1***. Seven patients (<1%) were excluded because of >50% missing values, while 135 (3%) patients were excluded because of out-of-reference range laboratory values.

Development of the A-HIPI

Continuous values of age and ALC had non-linear relationships with PFS and OS when modeled with piecewise linear splines (***Figure 1***). These findings were consistent across all datasets. See ***Figures S2-S5*** for relationships of total WBC count, hemoglobin, albumin, and ESR with survival. Univariate associations between baseline clinical variables and 5-year PFS and OS are reported in ***Tables S4 and S5.*** Additionally, Kaplan-Meier plots depicting PFS and OS for sex, stage, bulk, histology, and B-symptoms are in ***Figures S6-S10***. The following clinical predictors were eliminated from the PFS model: sex, B symptoms, histology, bulk, WBC count, hemoglobin, and ESR; the following predictors were eliminated from the OS model: B symptoms, histology, WBC count, and ESR.

Model variables with parameter estimates, hazard ratios, and optimism-corrected parameter estimates on multivariable analyses for the A-HIPI model are delineated in ***Table 2***. Please see the ***Supplemental Results*** for full model equations. Age, stage, ALC, and albumin were significant for both 5-year PFS and OS; sex, bulk disease, and hemoglobin were significant only for 5-year OS. C-statistics in the model development dataset were 0.605 (95% CI: 0.585, 0.624) for 5-year PFS and 0.732 (95% CI: 0.700, 0.761) for 5-year OS, with optimism-corrected values of 0.590 (95% CI: 0.584, 0.623) and 0.720 (95% CI: 0.689, 0.756), respectively. Based on internal/external validation, c-statistics in the omitted trial ranged from 0.54 to 0.65 for PFS and 0.61 to 0.77 for OS (***Tables S6-S7***).

External validation

Baseline characteristics and outcomes of the external validation cohort were similar besides a difference in stage (38% IIB, 30% III, 32% IV) and longer median follow-up (74 months, IQR 31-132) ***(Table 1***). C-statistics for the A-HIPI model were 0.590 (95% CI: 0.557, 0.622) for 5-year PFS and 0.730 (95% CI: 0.681, 0.774) for 5-year OS in external validation. By comparison, c-statistics for 5-year PFS and OS with the IPS7 were 0.597 and 0.692, respectively, and 0.579 and 0.657, respectively, for the IPS3[4,25], as derived from the validation

cohort (see ***Tables S8-S9***). Sensitivity analyses demonstrated no change in model parameters or c-statistic (***Tables S10-S12)***. In addition, the predicted PFS and OS distribution and the observed outcomes stratified by quartile of predicted risk in the external validation cohort are shown in ***Figure 2***. See ***Figure S11*** for the distribution of predicted risk in the development cohort.

Calibration plots for the A-HIPI, showing deciles of predicted PFS and OS plotted against observed values in the external validation cohort are shown in ***Figure 3***. The A-HIPI model showed good calibration in the external validation cohort except for a slight overestimation of risk in the highest risk decile. For the A-HIPI PFS model, the calibration intercept was -0.40 (95% CI: -0.82, 0.00) and the slope was 0.83 (95% CI: 0.50, 1.16); for the A-HIPI OS model, the calibration intercept was -0.43 (95% CI: -0.96, 0.09) and the slope was 0.89 (95% CI: 0.66, 1.11). The calibration plots for the IPS7 in the external validation cohort are shown in ***Figure S12***. IPS7 showed an overestimation of risk among all risk groups.

To simplify the application of the A-HIPI model, a calculator was created for point-of-care use (https://holistic-calculator.web.app/).

Discussion

Via unique worldwide partnerships fostered through the HoLISTIC Consortium and the application of rigorous data science principles, we obtained and harmonized detailed IPD for more than 5,000 newly-diagnosed, advanced-stage cHL patients. This work leveraged a carefully constructed common data model and included many seminal clinical trials conducted in the field over the past 15-20 years, together with IPD from prominent cancer registries. In this inaugural output from HoLISTIC, we developed and validated a modern-day prognostic model for adults with advanced-stage cHL, the A-HIPI. The model contains several novel non-linear relationships of key variables. Furthermore, while discrimination

did not differ significantly for PFS from the IPS7, there was improved discrimination for OS, and improved calibration for both PFS and OS with the A-HIPI.

The IPS7 was a seminal publication in the field.[2] It was largely developed on data from the 1980s in an era that included patients who received heavier alkylator therapy +/- total nodal irradiation (one-quarter of patients in IPS7 received non-anthracycline alkylator-based therapy). Clinical trials attempting to use the IPS7 as a tool to predict treatment effect have been negative.[26-28] In addition, subsequent analyses showed a diminished utility of the original IPS among contemporarily treated cHL patients.[4,5] Moccia et al. examined 740 patients treated with advanced-stage cHL in British Columbia.[5] The IPS7 remained prognostic, but the range of PFS and OS rates narrowed considerably across IPS scores of 0-7 in patients treated with ABVD or ABVD-like regimens, indicating poor calibration. Furthermore, Diefenbach et al. evaluated the IPS7 among 854 patients on E2496, one of the eight clinical studies included herein.[4,25] Among the original IPS7 variables, only age, stage, and hemoglobin were prognostic; it was labeled the IPS3. However, both of these were smaller datasets and neither had a formal assessment of model calibration.

Outcomes have continued to improve for cHL patients over the past two decades.[29,30] This is partly a result of improved pathology capabilities leading to more accurate diagnosis, use of positron-emission tomography (PET) imaging for more precise delineation of stage,[31] and preservation of treatment dose intensity, especially with ABVD-based therapy.[32] Other noteworthy changes in practice have included response-adapted therapy,[33,34] and improved post-relapse salvage treatment modalities, including stem cell transplantation and integration of novel targeted therapeutics, that have helped maintain OS despite treatment failure.[1]

Just as therapy has changed for advanced-stage cHL over the ensuing 25 years since the original IPS, so have predictive modeling methods. This includes assessment of internal validation to obtain shrinkage factors to decrease the risk of overfitting, internal-external

validation to assess heterogeneity of model performance across studies, and calibration for predicted versus observed outcomes in the external validation cohort. Additionally, candidate predictive variables in the original IPS were dichotomized to yield a score up to seven points, with total scores reflecting different prognostic significance for survival. In contrast, we utilized continuous values for age and all laboratory values, characterizing them based on the shape of their observed relationship with the outcomes. We found a striking non-linear effect with PFS outcomes *improving* for patients from ages 18 through 30 years, before a more typical association (e.g., worsening outcomes associated with advancing age) after age 30.[35] This finding is intriguing and quite distinct from a simple dichotomization of age at 45 years. It may represent unique biology in older adolescents and young adults that warrants further study. Alternatively, it may represent barriers or challenges to cancer care delivery in this age group, including non-adherence or dose delays.

In addition, we identified a similar non-linear relationship between ALC and survival. Lymphopenia has been identified as an adverse prognostic factor across multiple lymphoma subtypes and it was a variable in the IPS7 as a dichotomized value (i.e., $<600^3/\mu L$ or <8% of WBC); however, laboratory values were not examined as continuous variables in that analysis. Patients with values towards the higher range of normal and above the normal range for ALC (i.e., $>2.0^3/\mu L$) here also had inferior outcomes. Increasing WBC count is an adverse factor in the IPS7 and other lymphoid malignancies (e.g., mantle cell lymphoma);[36] increased ALC may represent a hyper-inflammatory state in cHL. Additional translational and biologic studies are warranted to delineate this relationship.

In the original IPS, the sample size of the discovery cohort was 1,618 cHL patients with complete data.[2] This was validated among 2,643 patients with missing serum albumin levels estimated by linear regression from hemoglobin levels and missing ALC that relied on inguinal node involvement as a surrogate indicator. It is important to highlight in our database that we

had little missing data across disease and patient characteristics, and we used multiple imputation to address missing data. In total, implausible or missing lab values were seen in a minority of patients. In sensitivity analyses, we compared two approaches to address missing lab data, including eliminating those cases versus multiple imputation and comparing the resultant model performance (***Tables S10-S11***); the results were similar.

We demonstrated superior discrimination for 5-year OS with the inclusion of continuous variables (and non-linear relationships) compared with the IPS7 and IPS3, although there was an overestimate in the highest deciles of predicted risk in the A-HIPI. Interestingly, despite use of dichotomous variables and missing data as highlighted above, the IPS7 had similar discrimination to the A-HIPI for prediction of PFS, suggesting that there are still unaccounted for predictors of 5-year treatment failure. Response-adapted trials with midstream therapy adjustment (e.g., de-escalation or escalation of treatment based on interim PET-based imaging) and other potential factors may make baseline data less relevant in PFS outcomes. We assessed potential differences in model performance by trials in the development cohort using leave-one-out internal-external validation; performance was relatively stable across trials, which varied in their use of baseline imaging, definitions of bulk, and treatment regimens.

However, the A-HIPI had improved calibration, comparing observed versus expected events, for both PFS and OS compared with IPS7 (***Figure 3 and Figure S12***). Unlike discrimination, which refers to a model's ability to separate high versus low-risk patients, calibration refers to the ability of a model to accurately predict outcomes in other cohorts and future patients.[37] To this end, the A-HIPI may be used to compare outcomes across cHL clinical trials and to reliably identify high-risk future patient groups for clinical trial design.

Limitations of these data include a paucity of older adults ages >65 years as treated in the clinical trials utilized for model development. There remains a need to build collaborative datasets for the entire age spectrum that include older adults who have been generally

understudied and have inferior outcomes.[38,39] We intend to enrich the current large annotated dataset with older patients over ages 60-65 years, as well as with younger adolescent and young cHL patients who are also often sparsely represented in clinical trials. Additionally, these data will need to be validated in cohorts who received more immediate intensive chemotherapy (e.g., BEACOPP) and novel targeted agents as part of frontline therapy. As is standard practice in predictive modeling, we did not consider the impact of specific treatments on outcomes since treatment is downstream of the baseline prediction. However, this dataset will serve as a platform to incorporate important post-baseline variables such as receipt of different treatment regimens, response-adapted imaging results, and key biologic findings into the model utilizing multistate modeling, and also the use of simulation modeling that can estimate the individualized risk of post-acute and late effects that occur decades after the completion of therapy.[40] Moreover, we have intentionally constructed this dataset and the associated model to allow for the *dynamic incorporation* of new studies and data and to allow for model re-calibration as outcomes evolve.

In conclusion, harnessing a large annotated database of recent seminal clinical trials together with cHL registries, we developed and validated a modern prognostic model, A-HIPI, to predict five-year PFS and OS. We identified novel non-linear relationships between age and ALC with outcomes and unique prognostic variables significant for PFS versus OS. Finally, to enhance the use of A-HIPI, we developed an online calculator to assist clinicians and future patients in estimating individualized prognosis (https://holistic-calculator.web.app/). Collectively, the multi-source predictive modeling approach herein may also inform the development of similar methods, treatment models, and analytic algorithms for other cancers and diseases.

Acknowledgments: We gratefully acknowledge the statistical and administrative staff at each of the clinical trials groups and registries for their assistance in data procurement and harmonization. We acknowledge Jason Nelson, MPH, for his assistance with the prediction model and coding; the patients and hospitals who participate in the Australian Lymphoma and Related Diseases Registry, steering committee and registry management team of LaRDR; Joseph M. Connors and Dr. Randy D. Gascoyne for the creation of the Centre for Lymphoid Cancer, BC Cancer Lymphoma Database and Dr. Laurie H. Sehn for ongoing management; Amy Kirkwood for data management and statistical advice for the Stanford V and RATHL trials coordinated in the University College London Clinical Trials Unit, and Peter Hoskin for leadership of the UK Stanford V study; we acknowledge the Princess Margaret Cancer Centre Foundation, and Dr. Mary Gospodarowicz, Dr. Richard Tsang, Dr. Michael Crump, and Krystyna Tybinkowski (in memoriam) for their dedicated management of the Princess Margaret Cancer Centre lymphoma registry, and the patients who contribute to this registry.

Funding: NCI R01 262265-01A1 (**Modeling Multi-Source Data in Hodgkin Lymphoma**, Parsons/Evens, MPI); NCI P50 CA97274 (Mayo/Iowa SPORE); U10 CA180820 (ECOG Support); U10 CA180888 and UG1 CA233230 (SWOG support); K08 HL146959 (Upshaw); Stanford V and RATHL studies were supported by Cancer Research United Kingdom grants CRUK/02/002 and CRUK/07/033, respectively; the Italian GITIL/FIL HD 0607 clinical trial was supported by the Associazione Italiana per la Ricerca sul Cancro IG n.2013 (A.G.), by Associazione Italiana Lotta alla Leucemia Bergamo Section, and by a research grant of Cassa di Risparmio di Cuneo.

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. Ansell SM: Hodgkin lymphoma: 2023 update on diagnosis, risk-stratification, and management. Am J Hematol 97:1478-1488, 2022
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 339:1506-14, 1998
3. Evens AM, Hutchings M, Diehl V: Treatment of Hodgkin lymphoma: the past, present, and future. Nat Clin Pract Oncol 5:543-56, 2008
4. Diefenbach CS, Li H, Hong F, et al: Evaluation of the International Prognostic Score (IPS-7) and a Simpler Prognostic Score (IPS-3) for advanced Hodgkin lymphoma in the modern era. Br J Haematol 171:530-8, 2015
5. Moccia AA, Donaldson J, Chhanabhai M, et al: International Prognostic Score in advanced-stage Hodgkin's lymphoma: altered utility in the modern era. J Clin Oncol 30:3383-8, 2012
6. www.hodgkinconsortium.com
7. Evens A, Advani R, Aleman B: The Hodgkin Lymphoma International Study for Individual Care (HOLISTIC): A Multi-National Collaborative to Enhance Decision Making for Pediatric and Adult Hodgkin Lymphoma (HL), European Hematology Association (EHA), 2020, pp EP1149
8. https://datacommons.cancer.gov/
9. Chisesi T, Bellei M, Luminari S, et al: Long-term follow-up analysis of HD9601 trial comparing ABVD versus Stanford V versus MOPP/EBV/CAD in patients with newly diagnosed advanced-stage Hodgkin's lymphoma: a study from the Intergruppo Italiano Linfomi. J Clin Oncol 29:4227-33, 2011
10. Merli F, Luminari S, Gobbi PG, et al: Long-Term Results of the HD2000 Trial Comparing ABVD Versus BEACOPP Versus COPP-EBV-CAD in Untreated Patients With Advanced Hodgkin Lymphoma: A Study by Fondazione Italiana Linfomi. J Clin Oncol 34:1175-81, 2016
11. Hoskin PJ, Lowry L, Horwich A, et al: Randomized comparison of the stanford V regimen and ABVD in the treatment of advanced Hodgkin's Lymphoma: United Kingdom National Cancer Research Institute Lymphoma Group Study ISRCTN 64141244. J Clin Oncol 27:5390-6, 2009
12. Gordon LI, Hong F, Fisher RI, et al: Randomized phase III trial of ABVD versus Stanford V with or without radiation therapy in locally extensive and advanced-stage Hodgkin lymphoma: an intergroup study coordinated by the Eastern Cooperative Oncology Group (E2496). J Clin Oncol 31:684-91, 2013
13. Press OW, Li H, Schoder H, et al: US Intergroup Trial of Response-Adapted Therapy for Stage III to IV Hodgkin Lymphoma Using Early Interim Fluorodeoxyglucose-Positron Emission Tomography Imaging: Southwest Oncology Group S0816. J Clin Oncol 34:2020-7, 2016
14. Johnson P, Federico M, Kirkwood A, et al: Adapted Treatment Guided by Interim PET-CT Scan in Advanced Hodgkin's Lymphoma. N Engl J Med 374:2419-29, 2016
15. Zinzani PL, Broccoli A, Gioia DM, et al: Interim Positron Emission Tomography Response-Adapted Therapy in Advanced-Stage Hodgkin Lymphoma: Final Results of the Phase II Part of the HD0801 Study. J Clin Oncol 34:1376-85, 2016
16. Gallamini A, Tarella C, Viviani S, et al: Early Chemotherapy Intensification With Escalated BEACOPP in Patients With Advanced-Stage Hodgkin Lymphoma With a Positive

Interim Positron Emission Tomography/Computed Tomography Scan After Two ABVD Cycles: Long-Term Results of the GITIL/FIL HD 0607 Trial. J Clin Oncol 36:454-462, 2018

17. 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 46:1753-1754i, 2017

18. 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 98:1116-1123, 2017

19. 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 34:2493-500, 2016

20. Lymphoma, Related Diseases Registry I: Improving outcomes for patients with lymphoma: design and development of the Australian and New Zealand Lymphoma and Related Diseases Registry. BMC Med Res Methodol 22:266, 2022

21. Biccler JL, Glimelius I, Eloranta S, et al: Relapse Risk and Loss of Lifetime After Modern Combined Modality Treatment of Young Patients With Hodgkin Lymphoma: A Nordic Lymphoma Epidemiology Group Study. J Clin Oncol 37:703-713, 2019

22. White IR, Royston P, Wood AM: Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 30:377-99, 2011

23. Harrell FE, Jr., Califf RM, Pryor DB, et al: Evaluating the yield of medical tests. JAMA 247:2543-6, 1982

24. Harrell FE, Jr., Lee KL, Califf RM, et al: Regression modelling strategies for improved prognostic prediction. Stat Med 3:143-52, 1984

25. Hayden AR, Lee DG, Villa D, et al: Validation of a simplified international prognostic score (IPS-3) in patients with advanced-stage classic Hodgkin lymphoma. Br J Haematol 189:122-127, 2020

26. Carde P, Karrasch M, Fortpied C, et al: Eight Cycles of ABVD Versus Four Cycles of BEACOPPescalated Plus Four Cycles of BEACOPPbaseline in Stage III to IV, International Prognostic Score >/= 3, High-Risk Hodgkin Lymphoma: First Results of the Phase III EORTC 20012 Intergroup Trial. J Clin Oncol 34:2028-36, 2016

27. Mounier N, Brice P, Bologna S, et al: ABVD (8 cycles) versus BEACOPP (4 escalated cycles >/= 4 baseline): final results in stage III-IV low-risk Hodgkin lymphoma (IPS 0-2) of the LYSA H34 randomized trial. Ann Oncol 25:1622-8, 2014

28. Viviani S, Zinzani PL, Rambaldi A, et al: ABVD versus BEACOPP for Hodgkin's lymphoma when high-dose salvage is planned. N Engl J Med 365:203-12, 2011

29. Brenner H, Gondos A, Pulte D: Ongoing improvement in long-term survival of patients with Hodgkin disease at all ages and recent catch-up of older patients. Blood 111:2977-83, 2008

30. Pulte D, Jansen L, Brenner H: Changes in long term survival after diagnosis with common hematologic malignancies in the early 21st century. Blood Cancer J 10:56, 2020

31. Barrington SF, Kirkwood AA, Franceschetto A, et al: PET-CT for staging and early response: results from the Response-Adapted Therapy in Advanced Hodgkin Lymphoma study. Blood 127:1531-8, 2016

32. Evens AM, Cilley J, Ortiz T, et al: G-CSF is not necessary to maintain over 99% dose-intensity with ABVD in the treatment of Hodgkin lymphoma: low toxicity and excellent outcomes in a 10-year analysis. Br J Haematol 137:545-52, 2007

33. Diefenbach CS, Connors JM, Friedberg JW, et al: Hodgkin Lymphoma: Current Status and Clinical Trial Recommendations. J Natl Cancer Inst 109, 2017

34. Armitage JO: The importance of testing PET-directed therapy in patients with Hodgkin lymphoma. Curr Hematol Malig Rep 6:155-6, 2011

35. Parsons S, Rodday A, Friedberg J, et al: Inferior Outcomes for Young Adults Treated on Advanced Stage Clinical Trials: Report from the HoLISTIC Consortium, International Symposium on Hodgkin Lymphoma. Cologne, Germany, October 23, 2022.

36. Hoster E, Dreyling M, Klapper W, et al: A new prognostic index (MIPI) for patients with advanced-stage mantle cell lymphoma. Blood 111:558-65, 2008

37. Alba AC, Agoritsas T, Walsh M, et al: Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. JAMA 318:1377-1384, 2017

38. Kumar AJ, Nelson J, Rodday AM, et al: Development and validation of a prediction model for 1-year mortality among older adults with Hodgkin Lymphoma who receive dose-intense chemotherapy. J Geriatr Oncol 12:1233-1239, 2021

39. Evens AM, Carter J, Loh KP, et al: Management of older Hodgkin lymphoma patients. Hematology Am Soc Hematol Educ Program 2019:233-242, 2019

40. Evens AM, Parsons SK: Continuum of Care for Hodgkin Lymphoma: Impact of Modern Therapy on Postacute Morbidity and Mortality. J Clin Oncol 38:4131-4134, 2020

Table 1. Baseline characteristics for development and validation cohorts.

	Development (N=4022)	Validation (N=1431)
Study*, n (%)		
ECOG2496	632 (15.7%)	
SWOG0816	322 (8.0%)	
HD2000	276 (6.9%)	
HD9601	310 (7.7%)	
HD0607	769 (19.1%)	
HD0801	466 (11.6%)	
Stanford V	314 (7.8%)	
RATHL	933 (23.2%)	
PMH		318 (22.2%)
Iowa/Mayo SPORE		251 (17.5%)
Australia		260 (18.2%)
BC Cancer		602 (42.1%)
Age (years), mean (SD)	35.1 (12.0)	35.6 (13.1)
Age (years), median (Q1, Q3)	33.0 (25.0, 43.0)	32.0 (25.0, 44.5)
Categorical age (years), n (%)		
18 to 30	1618 (40.2%)	613 (42.8%)
>30	2404 (59.8%)	818 (57.2%)
Female sex, n (%)	1828 (45.5%)	622 (43.5%)
Stage, n (%)		
Stage IIB	1106 (27.5%)	536 (37.5%)
Stage III	1568 (39.0%)	427 (29.8%)
Stage IV	1348 (33.5%)	468 (32.7%)
Histology, n (%)		

Lymphocyte depleted	46 (1.1%)	7 (0.5%)
Lymphocyte rich	102 (2.5%)	22 (1.5%)
Mixed cellularity	521 (13.0%)	85 (5.9%)
Nodular sclerosis	2986 (74.2%)	1023 (71.5%)
NOS	367 (9.1%)	294 (20.6%)
B symptoms, n (%)	2938 (73.1%)	1104 (77.1%)
Any bulk †, n (%)	1408 (35.0%)	433 (30.3%)
WBC count (10^3/µL), mean (SD)	10.7 (5.3)	10.8 (5.2)
Lymphocyte count (10^3/µL), mean (SD)	1.5 (0.7)	1.4 (0.7)
Categorical lymphocyte count (10^3/µL), n (%)		
0.1 to 2	3183 (79.1%)	1160 (81.0%)
2 to 5	839 (20.9%)	271 (19.0%)
Hemoglobin (g/dL), mean (SD)	12.0 (1.9)	12.0 (1.9)
Albumin (g/dL), mean (SD)	3.7 (0.6)	3.7 (0.6)
ESR (mm/hour), mean (SD)	59.0 (35.7)	52.8 (35.6)

Abbreviations: PMH refers to Princess Margaret Hospital; BC to British Columbia; Australia to Australasian Lymphoma and Related Diseases Registry; NOS to not otherwise specified; WBC to white blood cell; n, number; ESR to erythrocyte sedimentation rate.

* Number of patients by trial and registry refer to the subset of patients who met our eligibility criteria (see ***Table S1*** and ***Figure S1***).

† Bulk was defined per clinical trial or registry as either mediastinal and/or non-mediastinal bulk; clinical trial and registry definitions were used (see ***Table S2***).

Table 2. Model parameters for PFS and OS.

	5-year PFS*			5-year OS*		
	Beta coefficient	**HR**	**Optimism-corrected beta coefficient**	**Beta coefficient**	**HR**	**Optimism-corrected beta coefficient**
Age (years)						
Linear effect in 18 to 30†	-0.026	0.97	-0.024	-0.022	0.98	-0.020
Linear effect in >30†	0.016	1.02	0.014	0.049	1.05	0.046
Female				-0.251	0.78	-0.234
Stage^						
Stage IIB						
Stage III	0.207	1.23	0.184			
Stage IV	0.423	1.53	0.377	0.285	1.33	0.266
Any bulk				0.312	1.37	
Lymphocyte count (10^3/μL)						
Linear effect in .1 to 2†	-0.287	0.75	-0.255	-0.497	0.61	-0.463
Linear effect in 2 to 5†	0.188	1.21	0.167	0.396	1.49	0.369
Hemoglobin (g/dL)				-0.124	0.88	-0.116
Albumin (g/dL)	-0.307	0.74	-0.274	-0.406	0.67	-0.379

PFS refers to progression-free survival; OS to overall survival; and HR to hazard ratio.

* Separate models were built for OS and PFS with some variables included in one model but not the other. Blank cells indicate that a variable is not included in the model. Variables dropped from both models are not included. For continuous variables, the effect estimates represent a 1-unit increase in the predictor variable. 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. Because variable selection was used, and because the purpose of our modeling was parameter estimation rather than statistical inference, we do not provide 95% confidence intervals.

† Effect estimate shown is the combination of the two piecewise linear splines for the group with values above the inflection point. Piecewise linear splines are interpreted as follows. *Age:* for those aged 18 to 30 years, a 1-

year increase in age indicates improving outcomes, while for those >30 years, a 1-year increase in age indicates worse outcomes. *Lymphocyte count:* for an ALC of 0.1 to 2.0 (10^3/μL), a 1-unit increase in lymphocyte count indicates improving outcomes, while for an ALC 2.0 to 5.0 (10^3/μL), a 1-unit increase in lymphocyte count indicates worse outcomes. See ***Table S13*** for the optimism-corrected beta coefficients that correspond to each individual piece of the spline.

^ For the PFS model, the reference is stage IIB; for the OS model, the reference is stage IIB and III.

Figure Legends

Figure 1. Relationship between continuous variables and survival. Plots of penalized smoothing splines based on partial residuals from multivariable Cox PH model for continuous variables. Plots demonstrate non-linear associations between age with ***(A)*** progression-free survival (PFS) and ***(B)*** overall survival (OS), and for absolute lymphocyte count (ALC) with ***(C)*** PFS and ***(D)*** OS.

Figure 2. Distribution of predicted risk and observed outcomes stratified by predicted risk quartiles in the validation cohort. Distribution of predicted ***(A)*** progression-free survival (PFS) and ***(B)*** overall survival (OS) in the external validation cohort. Kaplan-Meier observed ***(C)*** PFS and ***(D)*** OS rates stratified by quartile of predicted risk in the external validation cohort. The term “event” on upper left x axis and lower left y axis of PFS plots indicates progression, relapse, or death.

Figure 3. Calibration plots in the validation cohort. Deciles of predicted ***(A)*** progression-free survival (PFS) ***(B)*** or overall survival (OS) plotted against observed values in the external validation cohort. The diagonal line with a slope of 1 denotes perfect calibration. The red line shows the calibration slope using local regression (loess). Calibration curves show moderate calibration but with overestimate of risk in the highest decile of predicted risk. The term “event” on the y axis of the PFS plot on the left indicates progression, relapse, or death.

Figure 1.

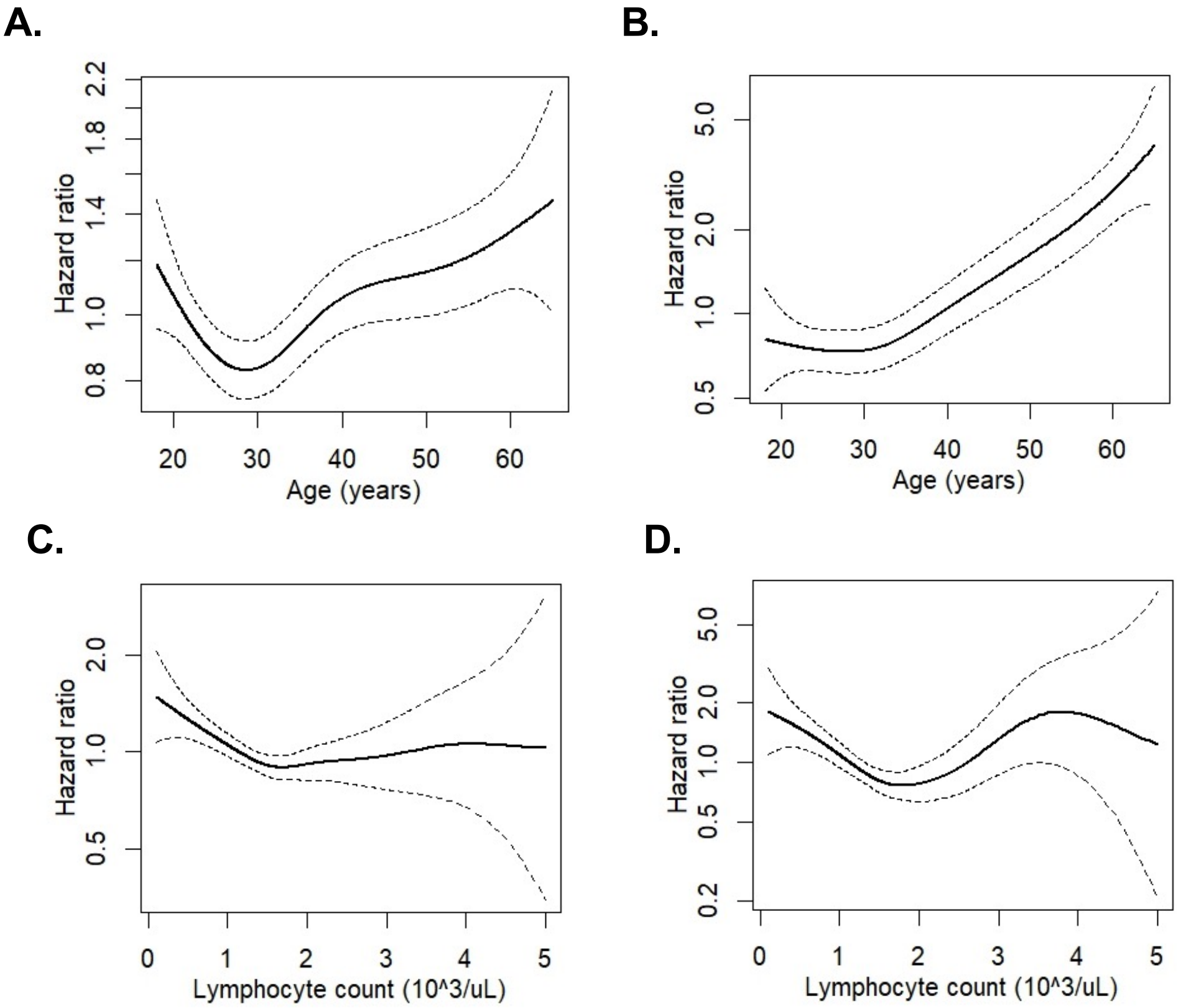

A.
Hazard ratio
0.8
1.0
1.4
1.8
2.2
20
30
40
50
60
Age (years)
B.
Hazard ratio
0.5
1.0
2.0
5.0
20
30
40
50
60
Age (years)
C.
Hazard ratio
0.5
1.0
2.0
0
1
2
3
4
5
Lymphocyte count (10^3/uL)
D.
Hazard ratio
0.2
0.5
1.0
2.0
5.0
0
1
2
3
4
5
Lymphocyte count (10^3/uL)

Figure 2.

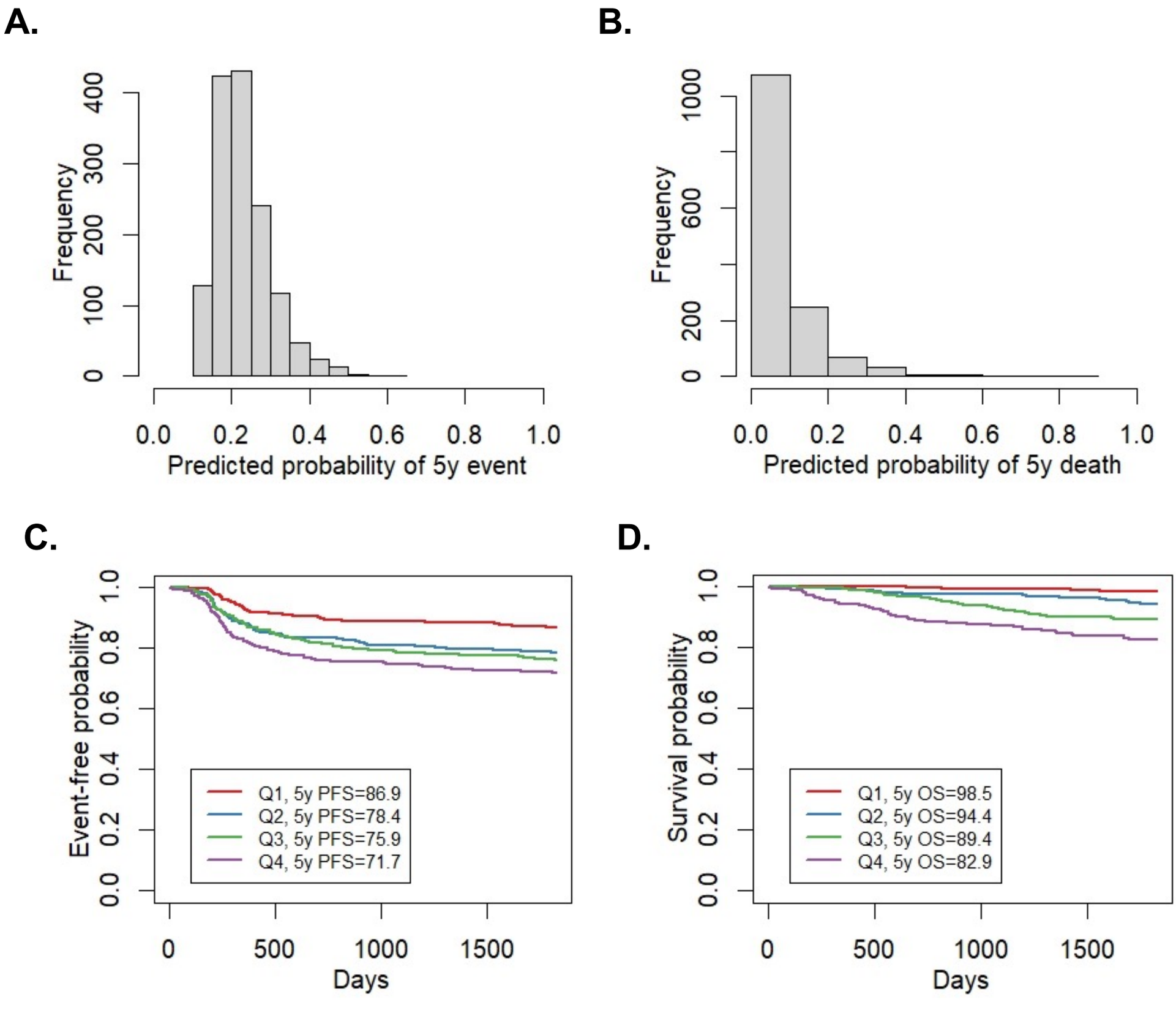
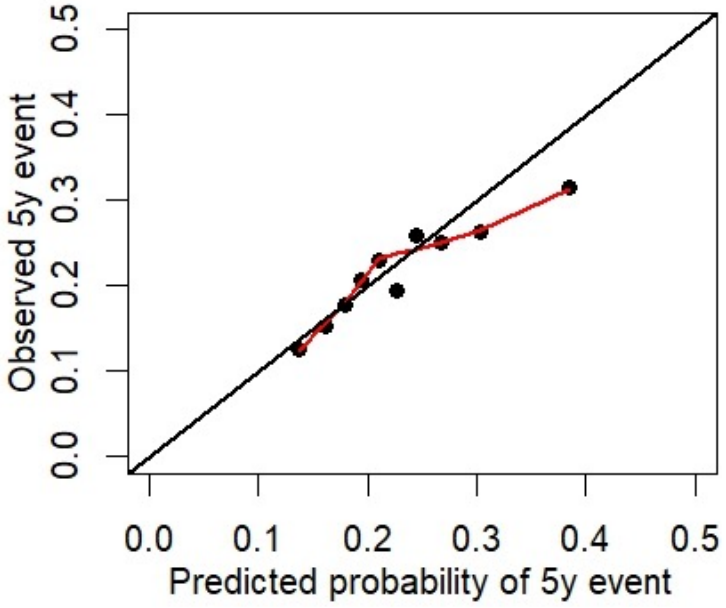
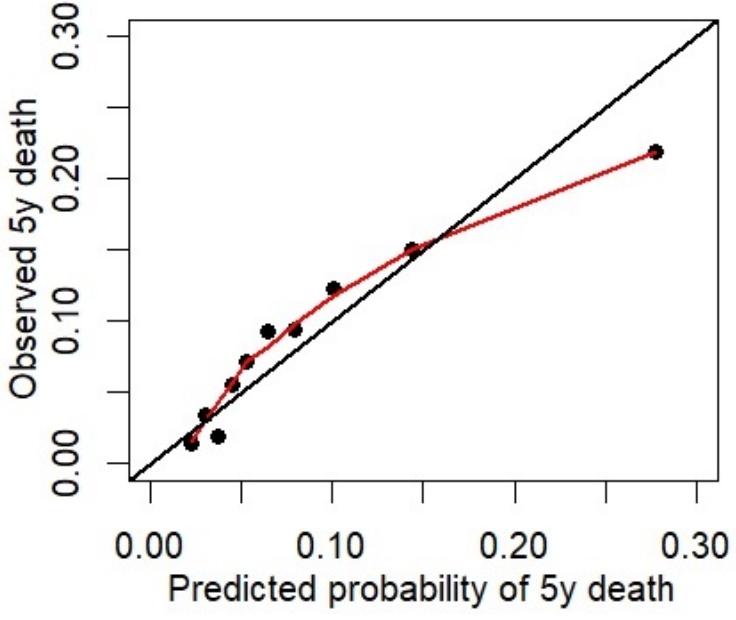

A.
Frequency
0
100
200
300
400
0.0
0.2
0.4
0.6
0.8
1.0
Predicted probability of 5y event
B.
Frequency
0
200
600
1000
0.0
0.2
0.4
0.6
0.8
1.0
Predicted probability of 5y death
C.
Event-free probability
0.0
0.2
0.4
0.6
0.8
1.0
Q1, 5y PFS=86.9
Q2, 5y PFS=78.4
Q3, 5y PFS=75.9
Q4, 5y PFS=71.7
0
500
1000
1500
Days
D.
Survival probability
0.0
0.2
0.4
0.6
0.8
1.0
Q1, 5y OS=98.5
Q2, 5y OS=94.4
Q3, 5y OS=89.4
Q4, 5y OS=82.9
0
500
1000
1500
Days

Figure 3.

A.


Observed 5y event
0.0
0.1
0.2
0.3
0.4
0.5
Predicted probability of 5y event

B.


Observed 5y death
0.00
0.10
0.20
0.30
Predicted probability of 5y death