ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ANNALS OF ONCOLOGY DRIVING INNOVATION IN ONCOLOGY

ORIGINAL ARTICLE

Correlation between progression-free and overall survival in patients with Hodgkin lymphoma: a comprehensive analysis of individual patient data from randomized German Hodgkin Study Group (GHSG) trials☆

P. J. Bröckelmann[1], H. Müller[1], M. Fuchs[1], S. Gillessen[1], D. A. Eichenauer[1], S. Borchmann[1], A. S. Jacob[1], K. Behringer[1], J. Momotow[1], J. Ferdinandus[1], B. Böll[1], X. Yang[2], C. Kobe[3], H.-T. Eich[4], C. Baues[5], W. Klapper[6], A. Engert[1], P. Borchmann[1] & B. von Tresckow[1,7]*

[1]University of Cologne, Faculty of Medicine and University Hospital of Cologne, Department I of Internal Medicine, Center for Integrated Oncology Aachen Bonn Cologne Düsseldorf (CIO ABCD) and German Hodgkin Study Group (GHSG), Cologne, Germany; [2]Merck & Co., Inc., Rahway, USA; [3]Department of Nuclear Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne; [4]Department of Radiation Oncology, University Hospital of Muenster, Muenster; [5]Department of Radiation Oncology, Ruhr-University Bochum, Medical Faculty, Marien Hospital, Herne; [6]Hematopathology Section and Lymph Node Registry, Department of Pathology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel; [7]Department of Hematology and Stem Cell Transplantation, West German Cancer Center and German Cancer Consortium (DKTK Partner site Essen), University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Available online 18 December 2024

Background: This study aimed to evaluate the correlation between progression-free (PFS) and overall survival (OS) after first-line treatment of classical Hodgkin lymphoma (HL) and to assess the potential of PFS as a surrogate parameter for OS.

Patients and methods: We analyzed individual patient data collected during and after treatment with polychemotherapy in nine randomized phase III trials [German Hodgkin Study Group (GHSG) HD7-HD15] between January 1993 and August 2018. The effects of 16 experimental treatments on PFS and OS at the trial level were evaluated using Cox proportional hazards (PH) regression and linear weighted least squares regression. At the patient level, marginal Cox PH models for multiple endpoints were applied using the Wei–Lin–Weissfeld method.

Results: At least one PFS and OS event was recorded in 1682 and 1064 of 10 605 patients, respectively. At the trial level, there was a strong correlation between treatment effects on PFS and OS (weighted Pearson $r = 0.72$, $R^2 = 0.54$, $P < 0.001$). At the patient level, a moderate to strong correlation between treatment effects on PFS and OS was observed, with Pearson r values ranging between 0.61 and 0.85 (each $P < 0.001$) and an overall $r = 0.74$. A regression model that accounted for different types of experimental treatments and historical progress across trial generations achieved a very strong correlation ($R^2 = 0.93$). When applied to data from the contemporary first-line ECHELON-1 trial, this model successfully predicted OS from PFS {prognosticated ln[HR(OS)] $= -0.68$ as compared with observed ln [HR(0.59)] $= -0.53$}.

Conclusion: In first-line trials of HL, PFS and OS, as well as treatment effects and prognostic effects on these endpoints, are strongly correlated. PFS serves as a strong predictor of treatment effects on OS, providing valuable insights many years before OS can be reliably assessed.

Key words: Hodgkin lymphoma, progression-free survival, overall survival, correlation

INTRODUCTION

Classical Hodgkin lymphoma (HL) accounts for ∼10%-15% of all malignant lymphomas and occurs with two age peaks (around the age of 25 and >60 years) at an incidence of 2-3/100 000 per year. Approximately 80%-90% of patients with HL are cured with adequate risk-adapted first-line treatment.[1-3] However, in case of refractory or relapsed disease, ∼50% of patients receiving second-line therapy relapse again and most of these patients eventually die from their disease.[4] Novel therapeutic approaches investigating new compounds for the treatment of HL should

*Correspondence to: Prof. Dr med. Bastian von Tresckow, Department of Hematology and Stem Cell Transplantation, West German Cancer Center and German Cancer Consortium (DKTK Partner Site Essen), University Hospital Essen, University of Duisburg-Essen, Hufelandstraße 55, 45147 Essen, Germany. Tel: +49-201-723-4341

E-mail: bastian.vontresckow@uk-essen.de (B. von Tresckow).

@B_von_Tresckow

☆Note: This work has been presented in part as oral abstract talks at the Annual Conference of the European Hematology Association (EHA) in June 2023 in Frankfurt, Germany, and the 17th International Conference on Malignant Lymphoma (ICML) in June 2023 in Lugano, Switzerland.

0923-7534/© 2024 The Author(s). Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

hence at least maintain current cure and survival rates in the first-line setting, while ideally reducing treatment-related morbidity and mortality.[5] For high-risk patients and in the relapsed setting, however, improved efficacy with prolonged progression-free (PFS) and overall survival (OS) remains a major goal.[6]

In clinical HL research, PFS and OS are the two outcomes predominating as measures for treatment efficacy. PFS is one of the most relevant endpoints from the perspective of HL patients[7,8] and most randomized HL trials apply PFS as a primary endpoint aiming at significant PFS differences (or exclusion of those in noninferiority trials). Although OS is considered an important outcome measure for regulatory authorities[9,10] and health technology assessment bodies, evaluating OS differences in HL is often unfeasible: In addition, in conditions with high cure rates such as HL, it is not necessarily a disease-specific endpoint. Evaluation of OS, especially at the time of initial trial analysis, is usually constrained by limited statistical power and might be influenced by subsequent therapies. Because of the overall favorable prognosis, evaluation of OS after first-line treatment for HL requires trials with extended duration spanning decades and the recruitment of unrealistically high patient numbers to achieve sufficient power, given the low OS event rates. Conducting trials aimed at demonstrating OS differences to measure the relative and cost-effectiveness of new drugs in this setting would result in significant delays or restrictions in access to effective and potentially life-saving medicines. Harnessing strongly correlated endpoints of similar relevance to patients, such as PFS, as surrogate parameters for OS, hence appears relevant. The promises and challenges associated with such approaches were recently also discussed by an author group from the Food and Drug Administration (FDA), highlighting the urgent need for comprehensive studies in this area.[11]

The relationship between PFS with OS in HL is hence of utmost importance, but has not been studied to date. Here, we report the correlation of treatment effects on PFS and OS, and the direct correlation between PFS and OS, in 10 605 patients with HL treated in randomized academic clinical trials in the first-line setting.

METHODS

Patients

The aim of this study was to evaluate the correlation between PFS and OS in patients with HL undergoing first-line treatment. For this purpose, we re-analyzed nine randomized controlled multicenter phase III trials involving patients with histologically confirmed first-occurrence HL. These trials, namely, the academic HD7-HD15 trials conducted by the German Hodgkin Study Group (GHSG), were each designed to address one of three well-established, clinically defined risk groups: early-stage favorable, early-stage unfavorable, or advanced-stage HL. All multicenter trials analyzed in this study were initiated, coordinated, and conducted by the GHSG using comparable methods and a consistent research framework. The first patient was enrolled in January 1993, and the last validated follow-up data were collected in August 2018. Recruitment for three consecutive trial generations, G3 (HD7-HD9), G4 (HD10-HD12), and G5 (HD13-HD15), was conducted between 1993 and 1998, 1998 and 2003, and 2003 and 2009, respectively. Each generation included trials for early-stage favorable (HD7, HD10, and HD13), early-stage unfavorable (HD8, HD11, and HD14) and advanced-stage HL (HD9, HD12, and HD15) patient groups.

Treatment arms of the randomized GHSG trials that were prematurely closed, patients without histologically confirmed classical HL (e.g. patients with nodular lymphocyte predominant HL), and patients who terminated trial participation before starting study treatment were excluded from this analysis. The primary analysis included individual patient data from 10 605 participants in the GHSG HD7-HD15 trials. Supplementary Table S1, available at https://doi.org/10.1016/j.annonc.2024.12.009, provides an overview of the randomized trials and patient numbers. All trials were approved by the review boards of the participating sites and conducted in accordance with the Declaration of Helsinki. Details on inclusion and exclusion criteria, patient and study characteristics, and results and conclusions have been published previously elsewhere.[1,12-15]

Definitions

PFS was calculated for each patient as the time between the date of completion of initial staging examinations and the date of first progression/relapse or death from any cause. For patients with a continuing response, PFS was defined as ending on the date of the last documented follow-up examination of tumor status and was right-censored. OS was calculated for each patient as the time from the date of completion of the initial staging examinations to either the date of death from any cause or the date the patient was last documented as alive; in the latter case, OS was right-censored. Unless otherwise specified, correlation measures (r) are weighted Pearson product—moment correlations, and the reported P values are based on two-sided tests. The strength of correlations was interpreted according to widely accepted ranges for these correlation measures.[16-18]

Statistical analyses

Different methodological approaches to estimate PFS—OS correlations were applied.

Treatment effects on PFS and OS and their correlation at the trial level. First, we examined the correlation between the effects of 16 experimental treatments on PFS and OS at the trial level using a two-step approach. We then estimated the experimental treatment effects—relative to the respective standard treatments—using Cox proportional hazards (PH) regression, with the treatment effects expressed as the natural logarithm of the estimated hazard ratios [ln(HR)]. These effects were then visualized and analyzed using meta-analytic techniques to compute correlation measures, including point estimates and their 95%

confidence intervals. We estimated the weighted Pearson correlation of the treatment effects on PFS and OS [ln(HR)] using linear weighted least squares regression. The corresponding r was adjusted for age and sex, based on their prognostic relevance in the international prognostic score,[19] when calculating the ln[HR]. This represents a weighted population estimate derived through Fisher z transformation with bias correction. In addition, a multiple regression model incorporating three additional covariates was developed to account for additional factors and enhance the prediction of treatment effects from PFS to OS. As detailed in Supplementary Appendix S1, available at https://doi.org/10.1016/j.annonc.2024.12.009, this model was applied to an external contemporary trial.

Marginal Cox PH model of treatment effects on PFS and OS and their correlation at the patient level. Moreover, we analyzed the treatment effects on PFS and OS at the patient level by applying a marginal Cox PH model for multiple failure time data (Wei—Lin—Weissfeld method, WLW[20,21]) within each trial. This analysis was conducted separately for early (early-stage favorable and unfavorable) and advanced HL stages and for the total sample.

Correlation of PFS and OS themselves with copula models at the patient level. Finally, we applied copula models to estimate Kendall's tau and other correlation measures of PFS and OS at the patient level.[22,23] The marginal empirical cumulative distribution function was applied to transform PFS and OS, followed by fitting Archimedean copulas within each trial. The three Archimedean copulas (Clayton, Frank, and Gumbel) differ in their weighting of earlier versus later survival times. Model fit, parameter estimates, and Kendall's tau with 95% confidence intervals are reported. The model with the lowest Akaike information criterion and the lowest Schwartz—Bayes criterion was selected as the best fit and used to compute correlation measures, including Kendall's, Pearson's, and Spearman correlations. Results from individual trials were pooled across the same subgroups of trials as already noted (e.g. early versus advanced stages).

RESULTS

Patient characteristics and events

Individual patient data from 10 605 participants in the randomized GHSG HD7-HD15 trials were analyzed (Table 1). Although the risk profile differed markedly between early-stage favorable, early-stage unfavorable, and advanced-stage HL, the variations within each GHSG risk group over time and across trial generations were minimal (e.g. the percentage of patients with Ann Arbor stage IV disease in advanced stages ranged between 33% and 35% for the HD9, HD12, and HD15 trials). The median observation time (OT) for PFS and OS was 87.5 and 100.1 months, respectively. In addition, OT for PFS ranged from 53.9 to 150.7 months across trials, and OT for OS ranged from 60.1 to 180.9 months (Supplementary Table S2, available at https://doi.org/10.1016/j.annonc.2024.12.009). We recorded at least one PFS event and one OS event in 1682 and 1064 of the 10 605 patients, respectively (Supplementary Table S3, available at https://doi.org/10.1016/j.annonc.2024.12.009).

Treatment effects on PFS and OS and their correlation at the trial level

To evaluate the correlation of treatment effects at the trial level, we estimated ordinary Cox PH regression coefficients in separate statistical analyses of PFS and OS (Table 2). These estimated Cox regression coefficients were subsequently analyzed using an inverse variance-weighted linear regression model adjusted for sex and age $\geq$45 years to account for key patient variables influencing PFS and OS.[19] At the trial level, the analysis revealed a high and significant correlation between treatment effects on PFS and OS ($r = 0.72$, $R^2 = 0.54$, $P < 0.001$; Figure 1).[18]

Interestingly, in cases involving significantly superior experimental treatment regimens with different backbones [e.g. HD14B, bleomycin, etoposide, doxorubicine, cyclophosphamide, vincristin, procarbazine and prednisone (BEACOPP)-based versus doxorubicine, bleomycin, vinblastin, dacarbazine (ABVD)-based], the effects on OS were less predictable from treatment effects on PFS compared with cases with similar yet nonsuperior experimental treatments (e.g. HD12D, comparisons of BEACOPP-based approaches; Figure 1).[24] To account for these factors, we developed a multiple regression model that included the type of compared treatments (equivalent or not equivalent, e.g. ABVD-based versus ABVD-based considered equivalent and BEACOPP-based versus ABVD-based considered not equivalent) and historic effects across the GHSG trial generations (G3-G5) as additional predictors. This model achieved a very strong correlation ($R_2 = 0.93$, $P < 0.001$; Table 3).

As an example of how our model (Table 3) can be used in an external dataset, we applied it to the more recent ECHELON-1 phase III trial results that compared two ABVD variants (6 $\times$ ABVD versus 6 $\times$ AVD + brentuximab vedotin) in patients with advanced-stage HL.[3] ECHELON-1 reported an HR of 0.68 for the treatment effect on PFS and 0.59 for the treatment effect on OS. Our model prognosticated an $\ln(HR_{OS})$ of -0.68 which is very close to the actually observed $\ln[HR_{OS}(0.59)]$ of -0.53 (Figure 2 and Supplementary Table S4, available at https://doi.org/10.1016/j.annonc.2024.12.009 and Appendix 1). Therefore our prognostic model is not only applicable to the GHSG dataset it was developed from, but is also able to predict OS from PFS in a contemporary randomized first-line trial (ECHELON-1) with reasonable accuracy.

Marginal Cox PH model of treatment effects on PFS and OS and their correlation at the patient level

The statistical analysis at the patient level confirmed a strong correlation between treatment effects on PFS and OS (Table 4 and Supplementary Table S5, available at https://doi.org/10.1016/j.annonc.2024.12.009 summarizes the

https://doi.org/10.1016/j.annonc.2024.12.009

Table 1. Patient characteristics

	HD7	HD8	HD9	HD10	HD11	HD12	HD13	HD14	HD15	Total (*N* = 10 872)
	Early-stage favorable (*n* = 267)	Early-stage unfavorable (*n* = 1104)	Advanced stage HL (*n* = 770)	Early-stage favorable (*n* = 1098)	Early-stage unfavorable (*n* = 1345)	Advanced stage HL (*n* = 1513)	Early-stage favorable (*n* = 1142)	Early-stage unfavorable (*n* = 1607)	Advanced-stage HL (*n* = 2026)	
Sex										
Female	105 (39)	559 (51)	312 (41)	435 (40)	691 (51)	599 (40)	484 (42)	854 (53)	810 (40)	4849 (45)
Male	162 (61)	545 (49)	458 (59)	663 (60)	654 (49)	914 (60)	658 (58)	753 (47)	1216 (60)	6023 (55)
Age ≥45 years										
No	199 (75)	861 (78)	598 (78)	748 (68)	1049 (78)	1159 (77)	715 (63)	1319 (82)	1597 (79)	8245 (76)
Yes	68 (25)	243 (22)	172 (22)	350 (32)	296 (22)	354 (23)	427 (37)	288 (18)	429 (21)	2627 (24)
Ann Arbor stage										
Missing	0	0	0	3	0	1	0	0	0	4
I	115 (43)	93 (8)	5 (1)	368 (34)	81 (6)	1 (0)	378 (33)	77 (5)	0 (0)	1118 (10)
II	152 (57)	986 (89)	113 (15)	727 (66)	1264 (94)	255 (17)	761 (67)	1525 (95)	329 (16)	6112 (56)
III	—	25 (2)	401 (52)	—	—	726 (48)	3 (0)	1 (0)	998 (49)	2154 (20)
IV	—	—	251 (33)	—	—	530 (35)	—	4 (0)	699 (35)	1484 (14)
Treatment group										
A (standard)	—	552 (50)	248 (32)	276 (25)	342 (25)	370 (24)	564 (49)	811 (50)	669 (33)	3832 (35)
B	267 (100)	552 (50)	264 (34)	276 (25)	341 (25)	386 (26)	—	796 (50)	684 (34)	3566 (33)
C	—	—	258 (34)	269 (24)	330 (25)	379 (25)	578 (51)	—	673 (33)	2487 (23)
D	—	—	—	277 (25)	332 (25)	378 (25)	—	—	—	987 (9)
Large mediastinal mass										
Missing	1	0	0	2	0	14	0	2	16	35
Mediastinal mass limited	266 (100)	893 (81)	554 (72)	1096 (100)	1074 (80)	1058 (71)	1136 (99)	1300 (81)	1390 (69)	8767 (81)
Large mediastinal mass	—	211 (19)	216 (28)	—	271 (20)	441 (29)	6 (1)	305 (19)	620 (31)	2070 (19)
Extranodal disease										
Missing	3	0	0	1	0	6	0	1	0	11
No extranodal disease	264 (100)	1020 (92)	496 (64)	1097 (100)	1209 (90)	1185 (79)	1137 (100)	1478 (92)	1629 (80)	9515 (88)
Extranodal disease	—	84 (8)	274 (36)	—	136 (10)	322 (21)	5 (0)	128 (8)	397 (20)	1346 (12)
≥3 lymph nodes affected										
Missing	2	0	0	1	0	6	0	1	0	10
0-2 lymph node areas	265 (100)	397 (36)	114 (15)	1097 (100)	439 (33)	249 (17)	1087 (95)	532 (33)	293 (14)	4473 (41)
≥3 lymph node areas	—	707 (64)	656 (85)	—	906 (67)	1258 (83)	55 (5)	1074 (67)	1733 (86)	6389 (59)
High erythrocyte sedimentation rate										
Missing	3	3	2	2	2	39	2	14	67	134
Normal erythrocyte sedimentation rate	264 (100)	551 (50)	240 (31)	1096 (100)	639 (48)	428 (29)	1134 (99)	755 (47)	609 (31)	5716 (53)
Increased erythrocyte sedimentation rate	—	550 (50)	528 (69)	—	704 (52)	1046 (71)	6 (1)	838 (53)	1350 (69)	5022 (47)

Continued

Table 1. Continued

	HD7	HD8	HD9	HD10	HD11	HD12	HD13	HD14	HD15	Total
	Early-stage favorable (n = 267)	Early-stage unfavorable (n = 1104)	Advanced stage HL (n = 770)	Early-stage favorable (n = 1098)	Early-stage unfavorable (n = 1345)	Advanced stage HL (n = 1513)	Early-stage favorable (n = 1142)	Early-stage unfavorable (n = 1607)	Advanced-stage HL (n = 2026)	**(N = 10 872)**
Albumin <4 g/dl										
Missing	55	180	140	192	196	307	61	244	80	2818
No	169 (80)	605 (65)	315 (50)	777 (86)	695 (60)	528 (44)	891 (82)	825 (61)	862 (44)	4842 (60)
Yes	43 (20)	319 (35)	315 (50)	129 (14)	454 (40)	678 (56)	190 (18)	538 (39)	1084 (56)	3212 (40)
Hb <10.5 g/dl										
Missing	2	12	10	13	4	8	6	3	10	1672
No	258 (97)	1019 (93)	632 (83)	1076 (99)	1260 (94)	1225 (81)	1130 (99)	1499 (93)	1648 (82)	8248 (90)
Yes	7 (3)	73 (7)	128 (17)	9 (1)	81 (6)	280 (19)	6 (1)	105 (7)	368 (18)	952 (10)
Leukocytes <15 000/mm^3										
Missing	2	10	7	10	4	7	6	4	3	1656
No	259 (98)	973 (89)	608 (80)	1070 (98)	1182 (88)	1208 (80)	1108 (98)	1426 (89)	1591 (79)	7999 (87)
Yes	6 (2)	121 (11)	155 (20)	18 (2)	159 (12)	298 (20)	28 (2)	177 (11)	432 (21)	1217 (13)
Stage IV										
Missing	0	0	0	1	0	1	0	0	0	1609
No	267 (100)	1104 (100)	516 (67)	1097 (100)	1345 (100)	979 (65)	1142 (100)	1607 (100)	1327 (65)	7777 (84)
Yes	—	—	254 (33)	—	—	533 (35)	—	—	699 (35)	1486 (16)
Lymphocytes low										
Missing	7	50	29	53	41	56	62	60	86	1991
No	255 (98)	998 (95)	688 (93)	1036 (99)	1246 (96)	1331 (91)	1072 (99)	1507 (97)	1790 (92)	8416 (95)
Yes	5 (2)	56 (5)	53 (7)	9 (1%)	58 (4)	126 (9)	8 (1)	40 (3)	150 (8)	465 (5)

Data are presented as n or n (%).
HL, Hodgkin lymphoma.

https://doi.org/10.1016/j.annonc.2024.12.009

Table 2. Treatment effects on PFS and OS

Trial	N	PFS			OS	
		Arm	Estimate, ln(HR)	SE	Estimate, ln(HR)	SE
HD8	1104	B	0.03404	0.11750	−0.12337	0.13609
HD9	770	B	−0.32484	0.14510	−0.19896	0.17158
		C	−0.54328[a]	0.15387	−0.31415	0.17714
HD10	1098	B	0.02068	0.25443	0.09532	0.32116
		C	0.14816	0.24658	0.11697	0.31650
		D	−0.03641	0.25917	−0.26256	0.34881
HD11	1345	B	0.40784	0.17861	0.19469	0.24878
		C	0.05329	0.19080	−0.00117	0.25414
		D	0.07225	0.19085	0.08136	0.25038
HD12	1513	B	0.17742	0.18842	0.13631	0.21679
		C	0.15597	0.19053	0.10266	0.22130
		D	0.37012	0.18255	0.17175	0.21670
HD13	1142	C	0.47130	0.20496	0.23998	0.31764
HD14	1607	B	−0.71568	0.18401	−0.10627	0.24687
HD15	2026	B	−0.40353	0.15063	−0.52930	0.19439
		C	−0.08574	0.14069	−0.26999	0.18269

Regression estimates with SE, separate Cox PH regression of PFS and OS adjusted for sex and age ≥45 years.
HR, hazard ratio; OS, overall survival; PH, proportional hazards; PFS, progression-free survival; SE, standard error.
[a]Assumption of PH violated.

treatment effects). Within the trials, the weighted Pearson *r* ranged between 0.61 and 0.85 (each $P < 0.001$), and with two exceptions, all correlations were $r > 0.70$. Overall, the weighted Pearson *r* was 0.74, with correlations being higher in advanced-stage HL ($r = 0.78$) than in early-stage disease ($r = 0.72$; Supplementary Table S6, available at https://doi.org/10.1016/j.annonc.2024.12.009).

Correlation of PFS and OS themselves with copula models at the patient level

To assess the direct correlation between PFS and OS at the patient level, we applied copula models. Across all trials, the Gumbel copula provided the best fit, as indicated by the lowest Akaike information criterion and Bayesian information criterion, compared with the Clayton and Frank copulas (Supplementary Table S7, available at https://doi.org/10.1016/j.annonc.2024.12.009). Observed correlations derived from Gumbel copula models of PFS and OS at the patient level showed a weighted Pearson *r* ranging between 0.72 and 0.83 (each $P < 0.001$, Supplementary Table S8, available at https://doi.org/10.1016/j.annonc.2024.12.009). The strength and the homogeneity of these correlations are remarkably similar to other results observed at the patient level.

DISCUSSION

To the best of our knowledge, this is the first analysis examining correlations between PFS and OS in HL. Our study analyzed nine large randomized trials evaluating 16 experimental treatments in >10 000 patients with newly diagnosed HL. The findings revealed strong correlations

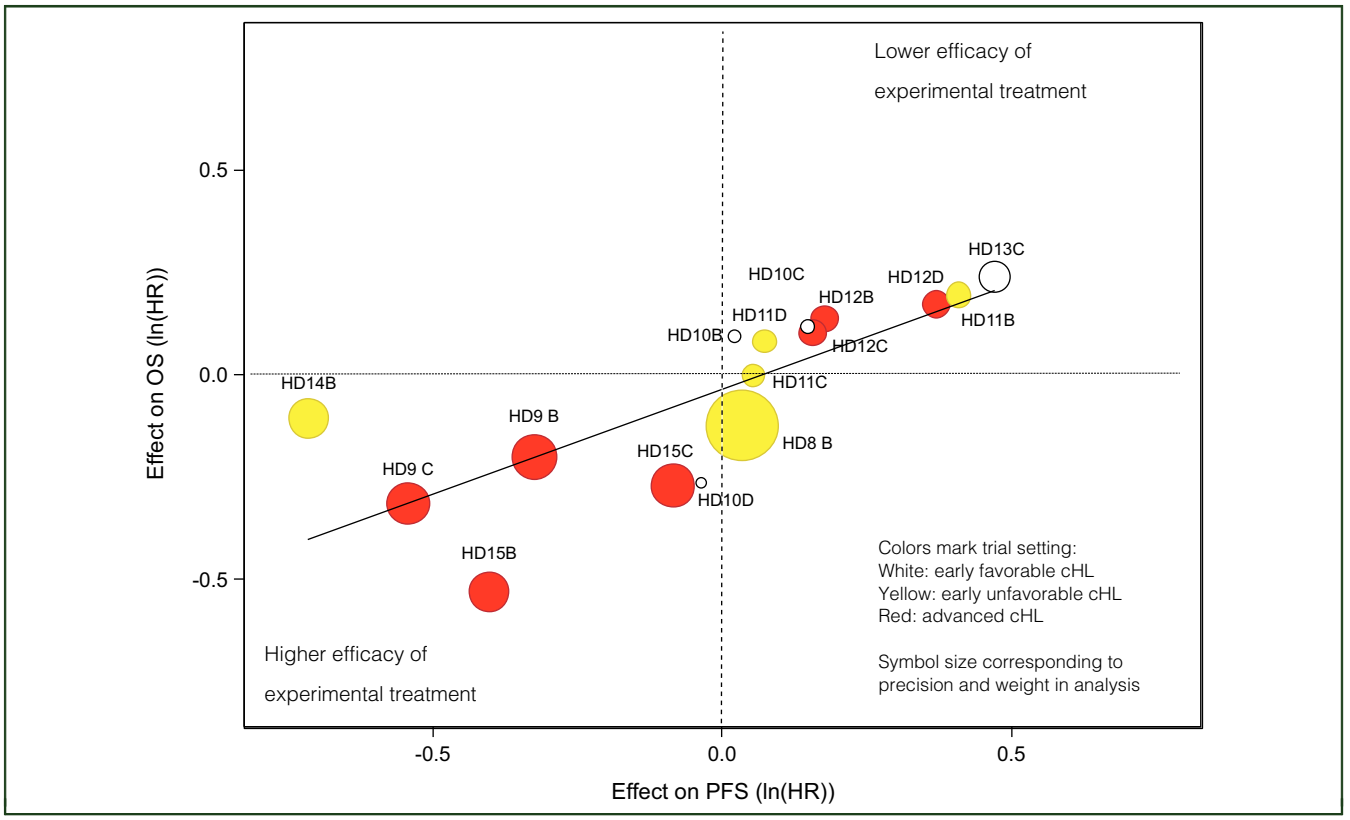

Figure 1. Sixteen separately estimated treatment effects on PFS and OS at the trial level and the resulting linear regression line with $r = 0.721$ (95% CI 0.350-0.896).
cHL, classical Hodgkin lymphoma; CI, confidence interval; HR, hazard ratio; OS, overall survival; PFS, progression-free survival.

Table 3. Multiple linear regression model of the estimated treatment effects of PFS on OS at the trial level ($R^2 = 0.93$, $P < 0.001$)

Variable	*b*	beta	*P*
Intercept	−0.122	0	0.0013
1. Treatment effects on PFS	0.346	0.550	0.0001
2. Type of standard chemotherapy	−0.425	−0.770	<0.0001
3. Trial generation G4 (versus G3)	0.152	0.337	0.0067
4. Trial generation G5 (versus G3)	0.237	0.523	0.0010

Type of standard chemotherapy = 1 with significant treatment effects on PFS and the same type of treatment (i.e. in the HD15 trial), = 0 otherwise.
Trial generation G4 = 1 in the HD10-12 trials, = 0 otherwise.
Trial generation G5 = 1 in the HD13-15 trials, = 0 otherwise.
b, regression coefficient; beta, standardized regression coefficient; *P*, significance; OS, overall survival; PFS, progression-free survival.

between treatment effects on PFS and OS and between the survival measures of PFS and OS themselves.

The principal findings from this analysis are as follows: (1) treatment effects on PFS and OS were highly correlated at the trial level, with a regression model incorporating known moderating factors achieving a very strong correlation; (2) a strong correlation between treatment effects on PFS and OS was also confirmed at the patient level; (3) Gumbel copula models applied at the patient level demonstrated strong correlations between PFS and OS themselves.

At the trial level, we observed a strong overall correlation ($r = 0.721$, 95% confidence interval 0.350-0.896; $R^2 = 0.54$, $P < 0.001$) between randomized treatment effects on PFS and OS. In meta-analyses, correlations with $R^2 > 0.50$ are typically considered large,[18] indicating strong to very strong correlations in this context. With respect to treatment effects at the patient level, our study revealed moderate to strong correlations with *r* between 0.61 and 0.85 across all HL stages, risk groups, and a wide range of polychemotherapy-based treatment approaches. Consistently, correlations between PFS and OS themselves ($r = 0.72$-0.83) at the patient level were also strong. These findings underscore the remarkably strong and consistent nature of the observed correlations, despite the inclusion of all risk groups and a wide range of treatments. Nevertheless, heterogeneity in treatment effects was observed at the trial level when the experimental treatment was superior and based on a different backbone compared with the standard treatment.

In this context, we identified additional variables predictive of treatment effects on OS: the type of standard chemotherapy in the trials (equivalent to experimental treatments or not) and trial generation (more recent G4 and G5 versus G3). This aligns with published factors influencing the relationship between PFS and OS, including the type of treatment and the trial recruitment period.[24] Incorporating these factors into the regression model at the trial level resulted in a very high (strong) correlation ($R^2 = 0.93$), surpassing even ambitious surrogacy criteria (e.g. $R^2 \geq 0.6$ or $R^2 \geq 0.8$).[25-27] However, as the model and the surrogacy margin were not prespecified in our study, further prospective validation is required to conclusively establish PFS

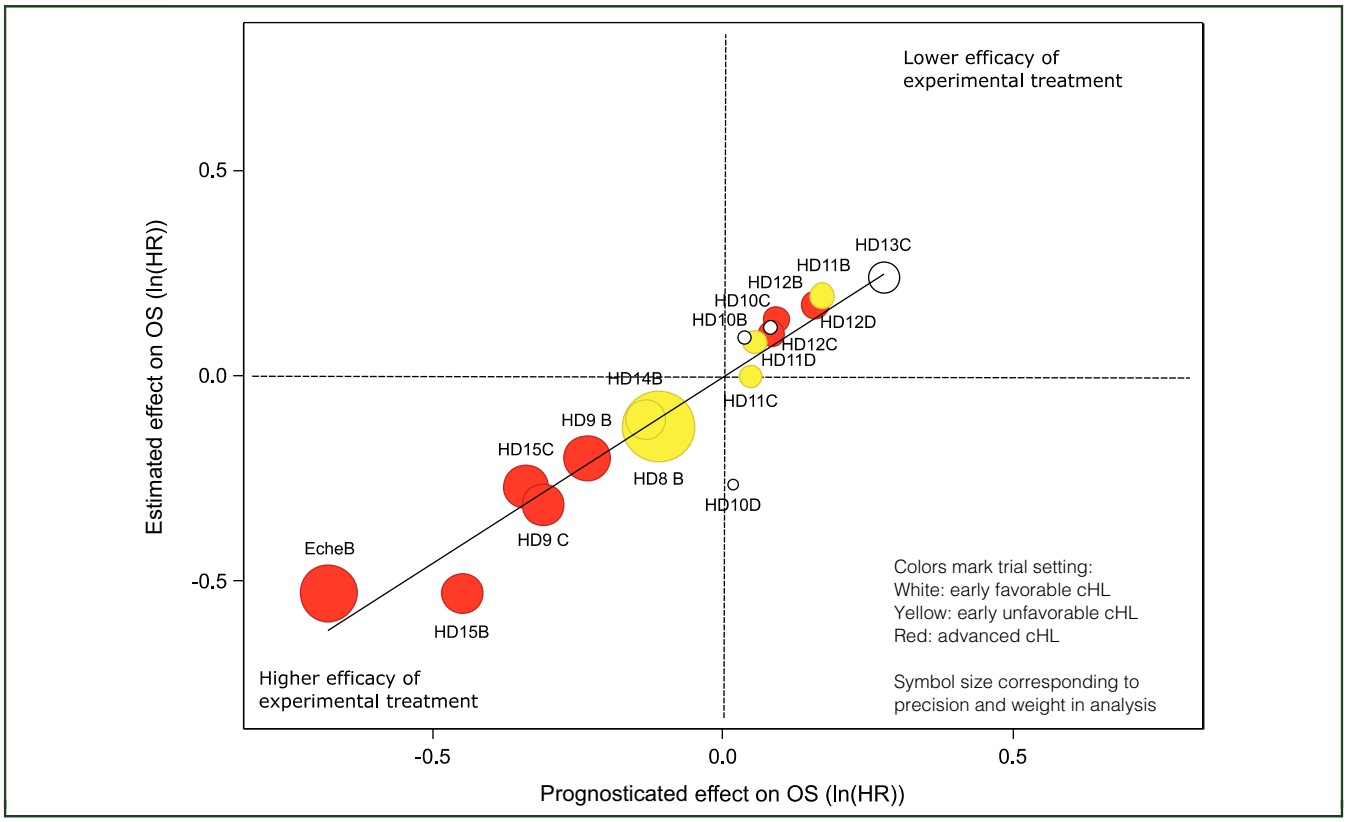

Figure 2. Prognosticated and observed treatment effects on OS using the regression model of Table 2. The treatment effects in the ECHELON-1 trial[3] are marked with 'Eche B' in the lower left corner.
cHL, classical Hodgkin lymphoma; HR, hazard ratio; OS, overall survival.

Table 4. Pearson product–moment correlations (*r*) of treatment effects on PFS and OS at the patient level (marginal Cox model, WLW)

Trial	*N*	Treatment group	Pearson *r*	LL 95% CI	UL 95% CI
HD8	1104	B	0.795	0.773	0.816
HD9	770	B	0.772	0.729	0.810
		C	0.790	0.749	0.824
HD10	1098	B	0.762	0.715	0.802
		C	0.735	0.684	0.779
		D	0.714	0.660	0.761
HD11	1345	B	0.688	0.636	0.734
		C	0.723	0.675	0.764
		D	0.730	0.684	0.771
HD12	1513	B	0.849	0.822	0.871
		C	0.841	0.814	0.865
		D	0.813	0.782	0.841
HD13	1142	C	0.609	0.571	0.644
HD14	1607	B	0.727	0.703	0.749
HD15	2026	B	0.753	0.725	0.778
		C	0.739	0.709	0.765

CI, confidence interval; LL 95% CI, lower limit of 95% confidence interval; OS, overall survival; PFS, progression-free survival; UL, upper limit of 95% CI; WLW, Wei—Lin—Weissfeld method.

as a surrogate for OS in first-line HL trials. In line with our findings, such validation must account for the influence of historic advancements and a core design factor: the types of systemic treatments being compared. This suggests that PFS, particularly when derived from data spanning multiple studies over longer periods, is not a simple surrogate for OS. Both historical progress and the choice of standard treatment can influence OS independently of PFS. This distinction arises because PFS is a composite endpoint marking the first event, whereas OS represents the final event. Thus in the era of our study, the increasing availability of new treatment options for relapsed HL may have extended OS after relapses without influencing PFS. As OS is influenced by all relapses/progressions and their treatment, its evaluation at a later stage, after potentially numerous salvage treatments, can obscure the shortcomings of a new experimental first-line treatment. In addition, the availability of effective salvage options in HL reduces the number of OS events recorded within reasonable time frames. This, in turn, results in reduced statistical power and necessitates a large, unfeasible number of patients and decades-long follow-up periods for first-line trials conducted in this rare disease when OS is used as the primary endpoint. Importantly, PFS reflects both primary cure and survival, making it regularly reported as the most important endpoint from a patient's perspective.[7,8,28]

Our analysis has several potential limitations. The most important limitation of the WLW method in the context of our study is its inability to disentangle the different causal pathways affecting OS. As a result, the correlation of two one-parametric effect estimates may oversimplify the true, more complex relationships between treatments, progressions/relapses, and OS. Although this represents an important and fundamental restriction of the method, it aligns with common statistical practices in oncology, where the effects of new experimental treatments on PFS and OS are frequently summarized using single estimates, such as hazard ratios. Another potential limitation is the assumption of PH, which is the basic requirement of Cox PH models. In our study, tests for the assumption of PH were conducted in all Cox PH regression analyses and revealed nonrandom deviations only in the PFS data for arm C of the HD9 trial. Additionally, we used the robust variance estimates proposed by Wei et al.[20] to account for potential violations of the PH assumption. Moreover, the Archimedean copulas applied in our analysis are a frequently used and simple one-parametric family of copulas, which may, once again, oversimplify the relationship between PFS and OS. Another limitation of our study is the wide range of treatments and OTs. Although these variations do not affect the treatment effects within the trials, they do influence the overall results, especially the prognostic model, making it sensitive to these confounding factors. Although we used elaborated statistical methods to account for the aforesaid factors, achieving perfect statistical control over all these confounding factors is not entirely feasible.

An additional potential limitation of this study is that it focused solely on traditional polychemotherapy-based first-line treatments. New therapeutic agents, such as brentuximab vedotin, nivolumab, or pembrolizumab, were not investigated in the clinical trials analyzed. However, as shown in Figure 2 and Supplementary Table S4, available at https://doi.org/10.1016/j.annonc.2024.12.009 and Appendix 1, our findings are applicable to a trial investigating the only approved targeted agent for first-line treatment of HL, brentuximab vedotin. Specifically, our regression model demonstrated the ability to predict the treatment effect on OS from PFS with reasonable accuracy in the ECHELON-1 trial dataset, despite the typically wide 95% confidence interval in HL PFS and OS estimates due to the low number of events, even in large trials.

Altogether, high cure rates are achieved with state-of-the-art therapies at the first diagnosis of HL. Therefore recent, current, and future first-line trials in HL are often designed as noninferiority trials, where OS events are usually very rare and occur very late. In these trials, PFS is an earlier, more sensitive, and more specific endpoint than OS.[29] The effects of reduced treatment efficacy on OS may take years or even decades to become apparent due to confounding factors, such as the availability of increasingly effective salvage therapies for progressive and relapsed disease. Therefore, in the context of present and future non-inferiority trials for newly diagnosed HL, OS appears to be a less practical primary endpoint compared with PFS.

In summary, PFS and OS, along with their respective treatment and prognostic effects, are highly correlated in first-line trials of HL. Given the high cure rates achieved with first-line treatments, PFS serves as a more sensitive and practical endpoint compared with OS, providing a reliable prediction of treatment effects on OS many years before OS can be accurately evaluated. Our findings strongly support the use of PFS as the primary endpoint in clinical trials for newly diagnosed HL.

FUNDING

PJB is supported by an Excellence Stipend of the Else Kröner-Fresenius-Foundation (EKFS). This work was financially supported through a research grant by Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc., Rahway, NJ, USA. No grant number is applicable.

DISCLOSURE

PJB is an advisor or consultant for Hexal, Merck Sharp & Dohme, Need Inc., Stemline, and Takeda; holds stock options in Need Inc.; has received honoraria from AstraZeneca, BeiGene, BMS/Celgene, Lilly, Merck Sharp & Dohme, Need Inc., Stemline, and Takeda; and reports research funding from BeiGene (Inst), BMS (Inst), Merck Sharp & Dohme (Inst), and Takeda (Inst). MF reports honoraria from BMS/Celgene, Janssen, and Takeda and is the CEO of KARO GmbH. DAE reports honoraria from Takeda and Sanofi-Genzyme. SB reports a consultant role for Galapagos; has received honoraria from Takeda; reports travel support from Takeda; and reports being a founder and shareholder of Liqomics. JF reports an advisory role for Takeda Oncology and speaker fees from Roche Pharma and Takeda Oncology. BB has received honoraria and research funding from Amgen, Gilead, MSD, Miltenyi, Noscendo, Novartis, Pfizer, and Takeda. WK reports research grants from Roche, Amgen, Takeda, Janssen, and Incyte, all to the institution. BvT is an advisor or consultant for Allogene, Amgen, BMS/Celgene, Cerus, Gilead Kite, Incyte, IQVIA, Janssen-Cilag, Lilly, Merck Sharp & Dohme, Miltenyi, Novartis, Noscendo, Pentixapharm, Pfizer, Pierre Fabre, QualWorld, Regeneron, Roche, Sobi, and Takeda; has received honoraria from AbbVie, AstraZeneca, BMS/Celgene, Gilead Kite, Incyte, Lilly, Janssen-Cilag, Merck Sharp & Dohme, Novartis, Roche, and Takeda; reports research funding from Esteve (Inst), Merck Sharp & Dohme (Inst), Novartis (Inst), and Takeda (Inst); and travel support from AbbVie, AstraZeneca, Gilead Kite, Janssen-Cilag, Lilly, Merck Sharp & Dohme, Pierre Fabre, Roche, Takeda, and Novartis; is a member of steering committees for Regeneron and Takeda. The remaining authors have declared no conflicts of interest.

REFERENCES

1. Sasse S, Brockelmann PJ, Goergen H, et al. Long-term follow-up of contemporary treatment in early-stage Hodgkin lymphoma: updated analyses of the German Hodgkin Study Group HD7, HD8, HD10, and HD11 trials. *J Clin Oncol*. 2017;35:1999-2007.
2. Kreissl S, Goergen H, Buehnen I, et al. PET-guided eBEACOPP treatment of advanced-stage Hodgkin lymphoma (HD18): follow-up analysis of an international, open-label, randomised, phase 3 trial. *Lancet Haematol*. 2021;8:e398-e409.
3. Ansell SM, Radford J, Connors JM, et al. Overall survival with brentuximab vedotin in stage III or IV Hodgkin's lymphoma. *N Engl J Med*. 2022;387:310-320.
4. Brockelmann PJ, Muller H, Gillessen S, et al. Clinical outcomes of relapsed and refractory Hodgkin lymphoma patients after contemporary first-line treatment: a German Hodgkin Study Group analysis. *Leukemia*. 2022;36:772-780.
5. Brockelmann PJ. Unfolding the potential of anti-programmed cell death protein 1 blockade in Hodgkin lymphoma - combination and personalisation? *Br J Haematol*. 2022;196:10-12.
6. Brockelmann PJ, Muller H, Kucuksarioglan E, et al. Outcomes of patients with the third or higher relapsed classical Hodgkin lymphoma: results from the German Hodgkin Study Group. *Ann Oncol*. 2019;30:490-491.
7. Kreissl S, Goergen H, Muller H, et al. Survivors' perspectives on risks and benefits of Hodgkin lymphoma treatment: results of a survey by the German Hodgkin Study Group. *Leuk Lymphoma*. 2019;60:1389-1398.
8. Brockelmann PJ, McMullen S, Wilson JB, et al. Patient and physician preferences for first-line treatment of classical Hodgkin lymphoma in Germany, France and the United Kingdom. *Br J Haematol*. 2019;184:202-214.
9. Guideline on the clinical evaluation of anticancer medicinal products - Revision 6. European Medicines Agency; 2024. Available at https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-evaluation-anticancer-medicinal-products-revision-6_en.pdf. Accessed 13 January 2025.
10. U.S. Food & Drug Administration. *Development and approval process—Drugs*. Beltsville: Center for Drug Evaluation and Research; 2022.
11. Merino M, Kasamon Y, Theoret M, Pazdur R, Kluetz P, Gormley N. Irreconcilable differences: the divorce between response rates, progression-free survival, and overall survival. *J Clin Oncol*. 2023;41:2706-2712.
12. von Tresckow B, Kreissl S, Goergen H, et al. Intensive treatment strategies in advanced-stage Hodgkin's lymphoma (HD9 and HD12): analysis of long-term survival in two randomised trials. *Lancet Haematol*. 2018;5:e462-e473.
13. Engert A, Haverkamp H, Kobe C, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *Lancet*. 2012;379:1791-1799.
14. Behringer K, Goergen H, Hitz F, et al. Omission of dacarbazine or bleomycin, or both, from the ABVD regimen in treatment of early-stage favourable Hodgkin's lymphoma (GHSG HD13): an open-label, randomised, non-inferiority trial. *Lancet*. 2015;385:1418-1427.
15. Gillessen S, Plutschow A, Fuchs M, et al. Intensified treatment of patients with early stage, unfavourable Hodgkin lymphoma: long-term follow-up of a randomised, international phase 3 trial of the German Hodgkin Study Group (GHSG HD14). *Lancet Haematol*. 2021;8:e278-e288.
16. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg*. 2018;126:1763-1768.
17. Swinscow TDV. Chapter 11: Correlation and regression. In: *Statistics at Square One*. 9th ed. London, UK: BMJ Publishing Group; 1997.
18. Cohen J. A power primer. *Psychol Bull*. 1992;112:155-159.
19. 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;339:1506-1514.
20. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc*. 1989;84:1065-1073.
21. Kelly PJ. A review of software packages for analyzing correlated survival data. *Am Stat*. 2004;58:337-342.
22. Weber EM, Titman AC. Quantifying the association between progression-free survival and overall survival in oncology trials using Kendall's. *Stat Med*. 2019;38:703-719.
23. Genest C, Rivest LP. Statistical-inference procedures for bivariate Archimedean copulas. *J Am Stat Assoc*. 1993;88:1034-1043.
24. Hernandez-Villafuerte K, Fischer A, Latimer N. Challenges and methodologies in using progression free survival as a surrogate for overall survival in oncology. *Int J Technol Assess Health Care*. 2018;34:300-316.
25. Belin L, Tan A, De Rycke Y, Dechartres A. Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. *Br J Cancer*. 2020;122:1707-1714.
26. Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? an analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the biomarker-surrogacy (BioSurrogate) evaluation schema (BSES). *BMC Med Res Methodol*. 2012;12:27.

27. Validity of surrogate endpoints in oncology. IQWiG Reports. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG). 2011.
28. van Leeuwen M, Husson O, Alberti P, et al. Understanding the quality of life (QOL) issues in survivors of cancer: towards the development of an EORTC QOL cancer survivorship questionnaire. *Health Qual Life Outcomes*. 2018;16:114.
29. Borchmann P, Moccia A, Greil R, et al. Comprehensive analysis of treatment related morbidity and progression-free survival in the GHSG phase III HD21 trial. *Blood*. 2023;142:3057.