Published in final edited form as:

Pediatr Blood Cancer. 2017 April ; 64(4): . doi:10.1002/pbc.26278.

Childhood Hodgkin International Prognostic Score (CHIPS) Predicts event-free survival in Hodgkin Lymphoma: A Report from the Children's Oncology Group

Cindy L. Schwartz¹, Lu Chen², Kathleen McCarten³, Suzanne Wolden⁴, Louis S. Constine⁵, Robert E. Hutchison⁶, Pedro A. de Alarcon⁷, Frank G. Keller⁸, Kara M. Kelly⁹, Tanya A. Trippet⁴, Stephan D. Voss¹⁰, and Debra L. Friedman¹¹

¹Division of Pediatrics, UT MD Anderson Cancer Center, Houston, Texas ²Children's Oncology Group, Monrovia, California ³Department of Radiology R.I. Hospital, Warren Alpert Medical School, Providence, Rhode Island ⁴Department of Pediatrics, Memorial Sloan Kettering Cancer Center, NY, New York ⁵Department of Radiation Oncology, University of Rochester, Rochester, New York ⁶Department of Pathology, S.U.N.Y. Upstate Medical University, Syracuse, New York ⁷Department of Pediatrics, University of Illinois College of Medicine, Peoria, Illinois ⁸Department of Pediatrics, Emory University, Atlanta, Georgia ⁹Department of Pediatrics, Roswell Park Cancer Institute, Buffalo, New York ¹⁰Department of Radiology, Boston Children's Hospital, Boston, Massachusetts ¹¹Vanderbilt University School of Medicine, Ingram Cancer Center, Nashville, Tennessee

Abstract

Background—Early response to initial chemotherapy in Hodgkin lymphoma (HL) measured by computed tomography (CT) and/or positron emission tomography (PET) after two to three cycles of chemotherapy may inform therapeutic decisions. Risk stratification at diagnosis could, however, allow earlier and potentially more efficacious treatment modifications.

Patients and Methods—We developed a predictive model for event-free survival (EFS) in pediatric/adolescent HL using clinical data known at diagnosis from 1103 intermediate-risk HL patients treated on Children's Oncology Group protocol AHOD0031 with doxorubicin, bleomycin, vincristine, etoposide, prednisone, cyclophosphamide (ABVE-PC) chemotherapy and radiation. Independent predictors of EFS were identified and used to develop and validate a prognostic score (Childhood Hodgkin International Prognostic Score [CHIPS]). A training cohort

Correspondence: Cindy L. Schwartz, Division of Pediatrics, UT MD Anderson Cancer Center, 1515 Holcombe Blvd Box 87, Houston, TX 77030. clschwartz@mdanderson.org.

AUTHORS' CONTRIBUTIONS

C. L. S. wrote the manuscript, designed and performed research, analyzed and interpreted data, and provided study material and patients; L. C. wrote the manuscript, performed research, and analyzed and interpreted data; S. D. V. and R. E. H. performed research and analyzed and interpreted data; T. A. T. analyzed and interpreted data; K. M., S. W., F. G. K., K. M. K., L. S. C., and P. A. analyzed and interpreted data and provided study material and patients; D. L. F. wrote the manuscript, performed research, analyzed and interpreted data, and provided study material and patients.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

was randomly selected to include approximately half of the overall cohort, with the remainder forming the validation cohort.

Results—Stage 4 disease, large mediastinal mass, albumin (<3.5), and fever were independent predictors of EFS that were each assigned one point in the CHIPS. Four-year EFS was 93.1% for patients with CHIPS = 0, 88.5% for patients with CHIPS = 1, 77.6% for patients with CHIPS = 2, and 69.2% for patients with CHIPS = 3.

Conclusions—CHIPS was highly predictive of EFS, identifying a subset (with CHIPS 2 or 3) that comprises 27% of intermediate-risk patients who have a 4-year EFS of <80% and who may benefit from early therapeutic augmentation. Furthermore, CHIPS identified higher risk patients who were not identified by early PET or CT response. CHIPS is a robust and inexpensive approach to predicting risk in patients with intermediate-risk HL that may improve ability to tailor therapy to risk factors known at diagnosis.

Keywords

adolescent; Hodgkin lymphoma; outcomes; pediatric; prognostic score

1 | INTRODUCTION

Hodgkin lymphoma (HL) is a highly curable malignancy, but the extent of treatment necessary for the individual patient differs across the population.^{1–4} Significant risks (late toxicities^{5–11} vs. recurrence^{12,13}) are associated with over- and undertreatment. Therefore, it is critical to examine methods of allocation to treatment strata that are optimally tailored to disease extent and biology. Although early response to initial chemotherapy as measured by computed tomography (CT) and/or positron emission tomography (PET) has proven useful in allocating patients to subsequent therapy^{1,3} (e.g., elimination of radiotherapy and augmentation of chemotherapy), risk stratification at diagnosis may allow earlier modification of the treatment approach. Until recently, selection of HL treatment regimens was based primarily on stage and B symptoms, occasionally with the addition of other clinical factors recognized as univariate predictors of outcome. The International Prognostic Score,¹⁴ developed based on multivariable analysis using a large cohort of adults with advanced-stage disease, has been used to identify risk but includes predictors that are not applicable to the pediatric/adolescent population. Pediatric studies that explored clinical features as multivariate factors were limited by cohort size and heterogeneity of treatments.^{15–17} Because a prognostic score for event-free survival (EFS) in pediatric/adolescent HL represents an important gap in knowledge that can impact patient outcome, we evaluated a large cohort of children and adolescents receiving uniform treatment for intermediate-risk HL. This has provided opportunity to identify predictors of EFS in HL based on clinical factors known at diagnosis and to develop a prognostic score using these factors.

2 | METHODS

2.1 | Patients and treatment

AHOD0031 was an intergroup study including patients from the Children's Oncology Group (COG), Stichting Kinderoncologie Nederland, Dutch Childhood Oncology Group (SKION), and Israeli Society of Pediatric Hematology and Oncology (ISPHO).¹ The COG includes patients from United States, Canada, Switzerland, Australia, and New Zealand in their HL trials. The study was open to enrollment from September 2002 to July 2009. Enrolled patients were <22 years of age with intermediate-risk HL (excluding IA/IIA without bulk and IIIB/IVB). Written informed consent and assent was obtained according to institutional review board guidelines from the patient, his/her parent, or legal guardian. The chemotherapy backbone was ABVE-PC: doxorubicin (25 mg/m² days 1 and 2), bleomycin (5 Units/m² day 1 and 10 Units/m² day 8), vincristine (1.4 mg/m² days 1 and 8, maximum dose 2.8 mg), etoposide (125 mg/m² days 1, 2, and 3), prednisone (40 mg/m²/day days 1–7), and cyclophosphamide (800 mg/m² day 1) with filgrastin 5 µg/kg daily days 6, 7, and 9 and until count recovery (modified from Schwartz).³ After two cycles, CT scan defined rapid early responders (RERs) with a 60% reduction in tumor size versus the slow early responders (SERs). Detailed information can be found in the primary study report.¹

The “standard” therapy was four cycles of ABVE-PC plus 21 Gy involved field radiation therapy (IFRT). Central review of early response was performed for RER who achieved complete response (CR) and for all patients beginning in 2006. With evolution of imaging techniques, PET scans became the usual choice of physicians. The study was completed before the Deauville criteria¹⁸ were defined in 2009; institutions therefore considered PET to be negative if there was definitive resolution of PET activity using qualitative measures. At the time of data analysis, PET scans read institutionally as equivocal prior to Deauville were re-evaluated at the Quality Assurance Review Center (Providence, RI) and assigned as negative if the Deauville score (DS) was ≤ 2.

2.2 | Patient cohorts for analyses

The study population for predictive modeling included patients randomized or assigned to receive standard therapy (four cycles of ABVE-PC and 21 Gy IFRT). Patients randomized to the experimental arms (no IFRT or chemotherapy augmentation) were excluded. Predictors of EFS were therefore evaluated in homogeneously treated patients (four cycles of ABVE-PC plus 21 Gy IFRT). Mutually exclusive training and validation cohorts were selected by random assignment of approximately half of the population to each cohort using the RANUNI function of SAS. Predictors of EFS identified in the training cohort were used to develop the Childhood Hodgkin International Prognostic Score (CHIPS), and then evaluated in the validation cohort. Log rank testing was used to compare EFS for each CHIPS (0–3). Additional analyses evaluated patients by histology and stage.

A secondary analysis was performed to evaluate whether use of early response by PET or CT might provide information that would mitigate the utility of the CHIPS. This analysis included all patients in the training and validation cohorts who had both PET and CT response reported after course 2 and data available to determine the CHIPS prognostic score.

2.3 | Statistical analysis

Clinical variables available within the COG database were included in the univariate analysis. Continuous variables were dichotomized based on clinical significance and/or quartiles of the cohort. Class variables were evaluated to identify subsets of the population with differential risks. Failures for EFS included disease relapse/progression, second malignancy, and death due to any cause. Time-to-event was computed from study enrollment; patients without failures were censored at date of last contact.

Cox proportional hazard model¹⁹ was used for univariate and multivariable analyses of the clinical predictors of EFS. Multivariable predictive modeling and stepwise selection were performed with predictors having $P \leq 0.25$ in univariate analysis. In instances of collinearity of predictors (e.g., bulk disease vs. large mediastinal mass), the most robust univariate variable was evaluated in the multivariable model. Analyses were performed using SAS 9.3.

CHIPS was devised using each factor associated with reduced EFS (from stepwise selection), with one point assigned to each relevant predictor (based on similar hazard ratios [HRs]). Only patients with data available for all predictive factors comprising CHIPS were used in the analysis of EFS. P -values from the log rank test comparing the EFS curves defined by CHIPS were used to determine whether the survival curves between CHIPS levels were different.

3 | RESULTS

3.1 | Patients

A total of 1,712 of 1,734 patients enrolled on AHOD0031 were eligible for study (Fig. 1). Twenty-two were ineligible: 18 did not meet diagnosis criteria, one received pretreatment steroid, and three did not have required pretreatment evaluations. Early response was not reported in 74 patients; they were not subsequently randomized or assigned to protocol therapy. Patients assigned to experimental regimens ($N = 535$) were not included in this analysis, leaving 1,103 who received four cycles of ABVE-PC/IFRT and serve as the focus of this report. These 1,103 patients were randomly divided into a training cohort ($N = 562$) and validation cohort ($N = 541$). Of these patients, 760 patients had both PET (optional) and CT (mandatory) after two cycles of chemotherapy. Patient characteristics of the 760 patients were similar to the 343 patients who did not have PET scans after two cycles, although PET scans were obtained more frequently in patients with private or military insurance versus Medicaid/Medicare or no insurance (74.3% vs. 60.2%, $P < 0.0001$) and in those without a large mediastinal mass (73.5% vs. 63.6%, $P = 0.0005$). Data were also available for all four CHIPS predictors in 704 of 760 patients. These 704 patients were used to examine CHIPS versus early CT or PET in allocation of therapy.

3.2 | Development of CHIPS: Predictors identified in the cohort receiving 4 ABVE-PC + radiation therapy

All variables in the AHOD0031 database were considered for inclusion in the model. Table 1 shows factors available for the majority of patients, the distribution of factors, and predictors identified by univariate and multivariable analyses and by multivariable analysis

with stepwise selection on the training cohort. The distribution of factors is similar within the training and validation cohorts.

Continuous variables dichotomized to differentiate clinical significance or “normal” from “abnormal” (based on developmental issues, well-defined normal values, or quartiles) included age (<13), number of involved sites (<3), hemoglobin (<10.5), albumin (<3.5), and erythrocyte sedimentation rate (ESR) (<20). ESR <50 was evaluated, but was less robust than ESR <20 in predicting an event. Univariate collinear variables related to disease burden included a large mediastinal mass (mediastinal tumor/thoracic diameter ratio >0.33), nodal aggregates (continuous aggregate of nodal tissue ≥ 6 cm in the transverse diameter), number of nodal sites (based on protocol-defined nodal regions), and bulk disease (defined as a large mediastinal mass or nodal aggregates). The presence of a large mediastinal mass was most robust and used in further analyses. B symptoms (fever, weight loss, night sweats individually, and the B symptom complex) were also collinear variables. Fever was used in further analyses because it was measurable and the most robust of the collinear variables.

A multivariable model was fit (Table 1, middle section) on the training cohort using the following variables: nodular sclerosis (NS) histology, stage 4, large mediastinal mass, fever, hemoglobin <10.5, albumin < 3.5, and ESR <20. A subsequent step-wise selection (Table 1, right section) resulted in a final model with four predictors of adverse EFS: stage 4, large mediastinal mass >0.33, fever, and albumin <3.5. The HRs for these factors were modest and similar (1.56–2.70); each were therefore assigned a point value of 1 toward the CHIPS. The CHIPS ranged from 0 to 3 in this intermediate-risk cohort (only IVB patients can achieve CHIPS 4).

3.3 | CHIPS and EFS

Figure 2 shows EFS curves for the training (2a), validation (2b), and combined (2c) cohorts. CHIPS could be assigned to 1,018 of the 1,103 patients in the combined (training and validation) cohorts. CHIPS functioned similarly in both the training and validation cohorts. *P*-values from log rank testing comparing the EFS curves defined by CHIPS confirmed that survival curves between CHIPS levels were different ($P < 0.0001$ and $P < 0.013$ for training and validation cohorts, respectively). Patients having CHIPS 0 and 1 in the combined cohort (1,018 patients) had excellent 4-year EFS (93.1 ± 1.4 and $88.5 \pm 1.6\%$, respectively) versus $77.6 \pm 2.9\%$ and $69.2 \pm 6.5\%$ for CHIPS 2 and 3 patients, respectively (Fig. 2). Overall, 73% were patients with CHIPS 0 or 1 having a 4-year EFS of 90.6%. CHIPS 2 or 3 patients comprised 27% of the combined cohort and had a 4-year EFS of 75.9%.

There were more patients with CHIPS 0 in the RER/CR than in the RER/no CR. SER had the fewest patients with CHIPS 0. Conversely, the SER had a larger percentage of CHIPS 2–3 than either RER cohort. Overall, there was a trend for CHIPS to be lower for RER/CR versus RER/no CR ($P < 0.001$) and for RER versus SER ($P < 0.0001$).

Table 2 shows the results of applying CHIPS to patients based on stage. Within each stage, CHIPS identified cohorts with different EFS. EFS within each CHIPS did not differ by stage. CHIPS was also evaluated in specific histologic cohorts. Eighty percent (817 of 1,018 patients) had NS histology; they show a virtually identical pattern to that seen in the entire

cohort. Outcomes could be predicted using CHIPS for patients (8%) with mixed cellularity (MC) histology: CHIPS 0–1 had >95% 4-year EFS versus 75% 4-year EFS for CHIPS 2–3.

3.4 | CHIPS and CT/PET

Data from the secondary analysis performed to evaluate whether the use of early response by PET or CT provided information that would mitigate the utility of the CHIPS are shown in Table 3.

This analysis included all patients in the population (training and validation cohorts) who had both PET and CT response reported after course 2 and had data available to determine the CHIPS prognostic score. There were 704 patients who had data to calculate CHIPS and both CT and PET response after two cycles. A high-risk cohort was identified with poor response by both CT and PET after two ABVE-PC; these individuals have 60% EFS. In comparison, those with evidence of response by CT and/or PET had 4-year EFS of 85–88%. Among this better risk cohort, CHIPS can further identify subcohorts with different EFS outcomes; EFS is 91.5 % for the 76% of patients with CHIPS 0 or 1 versus 74.9% for the 24% with CHIPS 2 or 3. Thus, patients with CHIPS 2 or 3 have an adverse outcome, even in the setting of response by PET or CT scan, or both.

The benefit of using CHIPS versus early response based on CT and PET scan is clear in this cohort. While both approaches detected very small populations of patients ($n = 34$ with 59–60% EFS for either CHIPS 3 or PET/CT nonresponders), the former was better able to separate the larger population with reasonably good outcomes into CHIPS 0–1 with an outcome of >90% and CHIPS 2 with an outcome of 75%. Early response could not separate cohorts with EFS of 88% for PET and CT responders or 85–88% for CT or PET responders. But even more important is the feasibility of defining a prognostic group at diagnosis rather than after two cycles, thus enabling earlier therapeutic modifications.

4 | DISCUSSION

Prognostic factors are clinical or biologic factors that predict likelihood of successful therapy. They are of relevance only when treatments available are sufficient for some but not all: if all or none of the patients are cured, there are no prognostic factors. Prognostic factors are not inherent biologic features of disease; instead, they are determined in direct relation to specific therapeutic modalities. Clinical trials are often tailored to previously identified prognostic factors, which then lose prognostic significance. Hasenclever noted that in the German Hodgkin Lymphoma Study Group trials, introduction of escalated bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, prednisone (BEACOPP) resulted in the disappearance of predictive power.²⁰ Patients may be overtreated when prognostic factors are the basis to augment therapy; once this occurs, the lack of observable events means that there is no information to discriminate among cured patients. Once a truly curative or nearly curative therapy exists, prognostic factors vanish.

Spoto considered a quantitative assessment of risks and benefits of treatment strategies,²¹ noting that optimal “cut-points” in criteria for high- versus low-risk disease must be based on both the morbidity of augmentation therapy and incremental efficacy. In HL, potential

gain in population outcome is achievable only by giving (i) minimally toxic/minimally beneficial treatment to a larger cohort or (ii) more intensive/toxic therapy to a small high-risk cohort. The history of prognostic factors for HL reflects the process of therapeutic tailoring leading toward more uniform prognosis and subsequent restriction of the utility of these factors.

We used early-response-based therapy in our legacy³ and COG¹ trials to tailor therapy for the individual. In the current trial (AHOD0031), we found that early-response could guide (i) reduction of therapy for RER by identifying patients without the need for IFRT and (ii) therapeutic augmentation with additional chemotherapy in SER.¹ Since tailoring of therapy occurred only after two or four cycles of standard chemotherapy, we sought to determine whether a cohort could be identified at diagnosis that might be targeted for earlier intensification or reduction of therapy.

AHOD0031 enrolled sufficient numbers of homogeneously treated patients to fully explore, with training and validation cohorts, the association of specific clinical parameters and outcome. Since individual univariate factors are highly associated in HL, we identified the most robust of collinear variables and then evaluated them in multivariable regression analysis. This led to the development of CHIPS based on four factors: fever, stage 4 disease, large mediastinal mass, and hypoalbuminemia (<3.5). Although CHIPS can theoretically range from 0 to 4, only scores of 0–3 were possible in this intermediate-risk cohort. Patients with CHIPS 0–1 had 90.5% 4-year EFS versus 76.9% for those with CHIPS 2–3. This differential in expected outcome can be defined at diagnosis, allowing for early implementation of tailored therapy.

Stage has been a highly relevant predictor of outcome in HL. In a small study of 69 children with newly diagnosed HL, stage IV disease was the only predictor of outcome determined by multivariable analysis.²² Our larger study found additional factors. Table 2 shows that CHIPS provides additional information beyond that provided by stage alone. The majority of patients on this trial had NS HL, which dominated the analysis of predictors but the CHIPS was also valid in MC HL. We did not evaluate CHIPS for other histology (lymphocyte predominant or depleted) as they were rare in this clinical trial. We also have not yet evaluated CHIPS in low- or high-risk HL.

Of the 670 patients with good response by CT and/or PET after two chemotherapy cycles, 162 (24%) had CHIPS 2–3 with increased risk of adverse outcome. This represents a significant cohort of the intermediate-risk patients who would have been treated without augmentation, or possibly with reduction in therapy. Instead, this cohort may benefit from more intensive therapy upfront. Optimizing initial therapy may improve overall EFS of this cohort. If the outcome of a one-fourth of patients can be improved, with as much as a 50% reduction in risk (from 76 to 88% EFS) by additional chemotherapy or introduction of newer targeted agents (e.g., brentuximab vedotin²³), then overall outcome of both this higher risk cohort and the overall cohort maybe enhanced to a greater extent than considered likely by Sposto et al.

We could not identify any CHIPS 2 or 3 cohorts defined by PET or CT response that had an EFS > 80%. This cohort therefore represents a higher risk group who may benefit from enhanced therapy. While it is unknown whether we can mitigate the adverse outcomes in those with high CHIPS, this would be a fruitful avenue for research. With new targeted therapy, this might be accomplished without excessive additional toxicity. We would suggest that the CHIPS might be used upfront for risk assignment, with CHIPS 2 and 3 receiving high-risk HL treatment such as the higher dose cyclophosphamide (current COG trials) or brentuximab vedotin (if ongoing studies show improved efficacy). Reduction in therapy could be considered for patients with lower CHIPS. If we were to modify initial therapy based on CHIPS, we would want to determine whether either the early response or EFS is (i) improved for the higher risk patients who might have augmented therapy and (ii) maintained for lower risk patients who might have reduced treatment.

With the use of CHIPS, response measured by PET or CT after two chemotherapy cycles did not contribute much additional information (Table 3). Only 19 patients (2.7% of this cohort) had poor response as measured by both PET and CT and had CHIPS 0 or 1. Although their therapies might not be optimally tailored if only CHIPS (not PET or CT) was used for allocation of therapy, only six of 19 recurred, too few to significantly impact the overall treatment failure rate even if all six would have benefitted from augmentation.

Early PET response to ABVD was evaluated by standardized uptake value in an early study by Hutchings²⁴ and more recently²⁵ by a DS of 3. PET scans on our study were investigational and not used for clinical decisions; a DS 2 identified early PET responders. Since those with DS = 3 would have been in the PET nonresponders/CT responders cohort that had virtually identical outcomes to the PET responder/CT responder cohort, it is unlikely using a different DS criteria for evaluating PET response would have impacted our results.

The criteria for response assessment must be defined by the study goals. In trials designed to reduce therapy for responders, more stringent criteria for response may reduce risk of relapse while more lenient criteria will reduce therapeutic exposures. Similar decisions may ultimately be made in determining therapeutic allocation by CHIPS.

Our data suggest that in this cohort, CHIPS eliminated any benefit of early response assessment. Most important, however, is that it is known at diagnosis and therefore allows earlier intervention. The robustness of CHIPS is of relevance not only because it facilitates earlier identification of risk, but also because it suggests feasibility of eliminating the costly early PET/CT scans that are increasingly used for therapeutic determination in HL.

Even more critical than cost reduction in the developed nations is that CHIPS could potentially allow developing nations to incorporate an equivalently predictive risk-based therapeutic algorithms using accessible parameters (stage 4 disease, large mediastinal mass, albumin level, and fever), thus improving access to state-of-the-art care throughout the world. This would, however, require careful analysis of CHIPS in the context of different cohorts and with different treatment regimens. Our results are based on a specific treatment regimen and did not include either the lowest or highest stage patients.

While it likely is possible to augment therapy to the extent that this prognostic score (CHIPS) will no longer be predictive, it is our goal to preserve ability to identify cohorts with excellent outcomes that may be considered for therapeutic reduction. Further reduction of therapy in the three-fourth of the population with low CHIPS may be feasible if higher risk patients receive more optimal treatment.

5 | CONCLUSION

CHIPS is a robust and inexpensive approach to predicting risk in patients with intermediate-risk HL that may improve ability to tailor therapy to risk factors known at diagnosis. While we advanced the concept of early response in HL as the method of titrating therapy³ because we did not have either biological or other markers as predictors, it is possible that this reliance on measuring early response will be replaced in the next era by an emphasis on biological and prognostic factors at diagnosis. Further assessments will be needed to determine the utility of the score in the context of additional cohorts and other therapeutic regimens and to determine biological correlates of the clinical factors.

Acknowledgments

Grant sponsor: Chair's Grant U10 CA98543 of the Children's Oncology Group from the National Cancer Institute, National Institutes of Health.

Abbreviations

ABVE-PC	Doxorubicin, bleomycin, vincristine, etoposide, prednisone, cyclophosphamide
CHIPS	Childhood Hodgkin International Prognostic Score
COG	Children's Oncology Group
CR	complete response
CT	computed tomography
DS	Deauville score
EFS	event-free survival
ESR	erythrocyte sedimentation rate
HL	Hodgkin lymphoma
HR	hazard ratio
IFRT	Involved field radiation therapy
MC	mixed cellularity
NS	nodular sclerosis
PET	positron emission tomography

RER	rapid early responders
SER	slow early responders

References

1. Friedman DL, Wolden S, Constone L, et al. AHOD0031: A phase III study of dose-intensive therapy and radiotherapy for intermediate risk Hodgkin lymphoma: A report from the Children's Oncology group. *J Clinical Oncol*. 2014; 32(32):3651–3658. [PubMed: 25311218]
2. Hasenclever, 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):1506–1514. [PubMed: 9819449]
3. Schwartz CL, Constone LS, Villaluna D, et al. A risk-adapted, response-based approach using ABVE-PC for children and adolescents with intermediate- and high-risk Hodgkin lymphoma: The results of P9425. *Blood*. 2009; 114(10):2051–2059. [PubMed: 19584400]
4. Wolden SL, Chen L, Kelly KM, et al. Long-term results of CCG 5942: A randomized comparison of chemotherapy with and without radiotherapy for children with Hodgkin's lymphoma—A report from the Children's Oncology Group. *J Clin Oncol*. 2012; 30(26):3174–3180. [PubMed: 22649136]
5. Armstrong GT, Liu Q, Yasui Y, et al. Late mortality among 5-year survivors of childhood cancer: A summary from the Childhood Cancer Survivor Study. *J Clin Oncol*. 2009; 27(14):2328–2338. [PubMed: 19332714]
6. Castellino SM, Geiger AM, Mertens AC, et al. Morbidity and mortality in long-term survivors of Hodgkin lymphoma: A report from the Childhood Cancer Survivor Study. *Blood*. 2011; 117(6):1806–1816. [PubMed: 21037086]
7. Friedman DL, Constone LS. Late effects of treatment for Hodgkin lymphoma. *J Natl Comp Canc Network*. 2006; 4(3):249–257.
8. Mertens AC, Yasui Y, Liu Y, et al. Pulmonary complications in survivors of childhood and adolescent cancer. A report from the Childhood Cancer Survivor Study. *Cancer*. 2002; 95(11):2431–2441. [PubMed: 12436452]
9. Schellong G, Riepenhausen M, Bruch C, et al. Late valvular and other cardiac diseases after different doses of mediastinal radiotherapy for Hodgkin disease in children and adolescents: report from the longitudinal GPOH follow-up project of the German-Austrian DAL-HD studies. *Pediatr Blood Cancer*. 2010; 55(6):1145–1152. [PubMed: 20734400]
10. Sklar C, Whitton J, Mertens A, et al. Abnormalities of the thyroid in survivors of Hodgkin's disease: Data from the Childhood Cancer Survivor Study. *J Clin Endocrinol Metab*. 2000; 85(9):3227–3232. [PubMed: 10999813]
11. Sklar CA, Mertens AC, Mitby P, et al. Premature menopause in survivors of childhood cancer: A report from the childhood cancer survivor study. *J Natl Canc Inst*. 2006; 98(13):890–896.
12. Nachman JB, Sposto R, Herzog P, et al. Children's Cancer Group. Randomized comparison of low-dose involved-field radiotherapy and no radiotherapy for children with Hodgkin's disease who achieve a complete response to chemotherapy. *J Clin Oncol*. 2002; 20(18):3765–3771. [PubMed: 12228196]
13. Friedmann AM, Hudson MM, Weinstein HJ, et al. Treatment of unfavorable childhood Hodgkin's disease with VEPA and low-dose, involved-field radiation. *J Clin Oncol*. 2002; 20(14):3088–3094. [PubMed: 12118022]
14. Mauz-Körholz C, Hasenclever D, Dörffel W, et al. Procarbazine-free OEPA-COPDAC chemotherapy in boys and standard OPPA-COPP in girls have comparable effectiveness in pediatric Hodgkin's lymphoma: The GPOH-HD-2002 study. *J Clin Oncol*. 2010; 28(23):3680–3686. [PubMed: 20625128]
15. Metzger ML, Weinstein HJ, Hudson MM, et al. Association between radiotherapy vs. no radiotherapy based on early response to VAMP chemotherapy and survival among children with favorable-risk Hodgkin lymphoma. *JAMA*. 2012; 307(24):2609–2616. [PubMed: 22735430]

16. Smith RS, Chen Q, Hudson MM, et al. Prognostic factors for children with Hodgkin's disease treated with combined-modality therapy. *J Clin Oncol.* 2003; 21(10):2026–2033. [PubMed: 12743158]
17. Bazzeh F, Rihani R, Howard S, Sultan I. Comparing adult and pediatric Hodgkin lymphoma in the Surveillance, Epidemiology and End Results Program, 1988–2005: An analysis of 21 734 cases. *Leuk Lymphoma.* 2010; 51(12):2198–2207. [PubMed: 21054151]
18. Meignan M, Gallamini A, Meignan, et al. Report on the first international workshop on interim-PET-scan in lymphoma. *Leuk Lymphoma.* 2009; 50:1257–1260. [PubMed: 19544140]
19. Cox DR. Regression models and life-tables. *J R Stat Soc Series B.* 1972; 34(2):187–220.
20. Hasenclever D. The disappearance of prognostic factors in Hodgkin's disease. *Ann Oncol.* 2002; 13(S1):75–78. [PubMed: 12078907]
21. Spoto R, London WB, Alonzo TA. Criteria for optimizing the prognostic risk groups in pediatric cancer: Analysis of data from the Children's Oncology Group. *J Clin Oncol.* 2007; 25(15):2070–2077. [PubMed: 17513812]
22. Oguz A, Karadeniz C, Okur FV, et al. Prognostic factors and treatment outcome in childhood Hodgkin disease. *Pediatr Blood Cancer.* 2005; 45(5):670–675.
23. Younes A, Connors JM, Park SI, et al. Brentuximab vedotin combined with ABVD or AVD for patients with newly diagnosed Hodgkin's lymphoma: A phase 1, open-label, dose-escalation study. *Lancet Oncol.* 2013; 14(13):1348–1356. [PubMed: 24239220]
24. Hutchings M, Loft A, Hansen M, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood.* 2006; 107:52–59. [PubMed: 16150944]
25. Biggi A, Gallamini A, Chauvie S, et al. International validation study for interim PET in ABVD-treated, advanced-stage Hodgkin lymphoma: Interpretation criteria and concordance rate among reviewers. *J Nuc Med.* 2013; 54(5):683–690.

FIGURE 1.

The cohort diagram shows the allocation of patients on AHOD0031 to standard (N = 1,103) versus the two experimental regimens (N = 382 and 153), to the training and validation cohort (N = 562, 541), and those on the standard arm who had PET and CT performed (N = 760 of whom 704 also had CHIPS known)

FIGURE 2.
The EFS by CHIPS for four ABVE-PC + IFRT patients on the training cohort (2A), the validation cohort (2B), and the combined cohort (2C)

TABLE 1

Predictors of EFS

	Training cohort, N = 562						Validation cohort, N = 541		
	Univariate			Multivariable			Multivariable stepwise selection		
	#	%	HR	p	HR	p	HR	p	%
Gender: Male	562	53.9	0.79	0.32					47.9
Age < 13	562	23.1	0.83	0.55					26.1
Race:	542			0.93					532
White (reference)		77.3	1	n/a					82.3
Black		13.3	0.66	0.29					10.2
Asian		3.5	1.05	0.94					2.4
Native American		0.2	0.00	0.99					0.38
Hawaii/Pacific Isl.		0.4	0.00	0.99					0
Other		5.3	0.71	0.57					4.7
Race: black vs. other	542	13.3	0.64	0.28					532
Hispanic	542	12.6	0.65	0.34					529
Histology	538			0.57					516
NS (reference)		86.6	1	n/a					86.6
MC		7.8	0.62	0.35					8.5
LP		5.4	0.44	0.25					4.5
LD		0.2	0.00	0.99					0.4
Histology: NS	538	86.6	1.82	0.18	0.94	0.90			516
									86.6

	Training cohort, N = 562						Validation cohort, N = 541		
	Univariate			Multivariable			Multivariable stepwise selection		
	#	%	HR	P	HR	P	HR	P	#
Histology: MC	538	7.8	0.66	0.44					516
									%
Stage: versus stage 4	562			0.13					541
1		5.5	0.44	0.18					6.1
2		59.3	0.63	0.08					58.6
3		18.0	0.45	0.04					20.2
4 (reference)		17.3	1	n/a					15.2
Stage: 4	562	17.3	1.83	0.035	2.36	0.0054	2.31	0.004	541
Large mediastinal mass	541	44.0	2.81	<0.001	2.60	0.0006	2.70	<0.001	522
Nodal aggregates 6 cm	549	55.9	1.36	0.22					57.7
Bulk disease	558	75.8	2.73	0.0049					76.4
Involved sites < 3	492	67.7	1.05	0.85					69.0
Extralymphatic	538	26.0	1.10	0.73					23.8
B symptoms	562	23.7	2.02	0.0062					21.2
Weight loss	556	10.3	1.54	0.23					10.7
Night sweats	560	15.0	2.11	0.011					13.7
Fever	561	10.7	2.32	0.0086	2.33	0.017	2.26	0.011	8.8
Hemoglobin < 10.5	548	18.6	1.43	0.22	1.02	0.96			23.2
Albumin < 3.5	541	30.7	1.89	0.012	1.45	0.18	1.56	0.073	30.3
ESR < 20	531	18.8	0.49	0.064	0.55	0.26			19.1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Training cohort, N = 562										Validation cohort, N = 541			
Univariate			Multivariable				Multivariable stepwise selection			Univariate			
#	%	HR	p	HR	p	HR	p	HR	p	#	%	HR	p
ESR < 50	55.2	0.67	0.11							513			
											53.8		

TABLE 2

Four-year EFS for stage by CHIPS

CHIPS	N	Stage 1	Stage 2	Stage 3	Stage 4	Entire cohort
0	341	94.7 ± 3.7 (N = 41)	93.0 ± 2.0 (N = 185)	91.1 ± 2.9 (N = 115)	(N = 0)	93.1 ± 1.4 NSS
1	405	91.7 ± 8.0 (N = 14)	88.7 ± 2.0 (N = 251)	88.1 ± 4.2 (N = 63)	87.8 ± 3.8 (N = 77)	88.5 ± 1.6 NSS
2	217	80.0 ± 17.9 (N = 5)	78.6 ± 3.7 (N = 129)	80 ± 8.9 (N = 20)	74.5 ± 5.7 (N = 63)	77.6 ± 2.9 NSS
3	55	100 (N = 2)	64.5 ± 9.6 (N = 27)	(N = 0)	72.7 ± 8.8 (N = 26)	69.2 ± 6.5 NSS
Total	1018					$P < 0.001$

TABLE 3

Four-year EFS for patients with CT and PET evaluation after two cycles

	PET and CT responders	PET nonresponder/CT responder	CT nonresponder/PET responder	PET and CT nonresponders	All
	% 4-year EFS/total number				
CHIPS 0	92% (182)	93% (45)	100% (10)	53% (9)	91% (246)
CHIPS 1	91% (179)	89% (73)	88% (19)	80% (10)	90% (281)
CHIPS 2	78% (78)	74% (34)	79% (19)	63% (12)	76% (143)
CHIPS 3	68% (20)	50% (6)	75% (5)	0% (3)	59% (34)
All	88% (459)	85% (158)	86% (53)	60% (34)	86% (704)
COHORTS DEFINED BY CHIPS					
	Either PET and/or CT response			PET and CT Nonresponders	
CHIPS 0	93 % (237)	CHIPS 0–1	91.5% (508)	60% (34)	
CHIPS 1	90 % (271)				
CHIPS 2	77 % (131)	CHIPS 2–3	74.9% (162)		
CHIPS 3	65 % (31)				