



Published in final edited form as:

*Clin Cancer Res.* 2013 December 1; 19(23): . doi:10.1158/1078-0432.CCR-13-1120.

## A Comparative Analysis of Prognostic Factor Models for Follicular Lymphoma based on a Phase III Trial of CHOP-Rituximab vs CHOP + <sup>131</sup>Iodine-Tositumomab

Oliver W. Press<sup>1,2,\*</sup>, Joseph M. Unger<sup>3,\*</sup>, Lisa M. Rimsza<sup>4</sup>, Jonathan W. Friedberg<sup>5</sup>, Michael LeBlanc<sup>3</sup>, Myron S. Czuczman<sup>6,7</sup>, Mark Kaminski<sup>8</sup>, Rita M. Braziel<sup>9</sup>, Catherine Spier<sup>4</sup>, Ajay K. Gopal<sup>1,2</sup>, David G. Maloney<sup>1,2</sup>, Bruce D. Cheson<sup>7,10</sup>, Shaker R. Dakhil<sup>11</sup>, Thomas P. Miller<sup>12</sup>, and Richard I. Fisher<sup>5</sup>

<sup>1</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

<sup>2</sup>Medical Oncology, University of Washington Medical Center, Seattle, WA

<sup>3</sup>SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA

<sup>4</sup>Department of Pathology, University of Arizona, Tucson, AZ

<sup>5</sup>James P. Wilmot Cancer Center, University of Rochester Medical Center, Rochester, NY

<sup>6</sup>Roswell Park Cancer Institute, Buffalo, NY

<sup>7</sup>Cancer and Leukemia Group B

<sup>8</sup>University of Michigan, Ann Arbor, MI

<sup>9</sup>Department of Pathology, Oregon Health & Science University, Portland, OR

<sup>10</sup>Georgetown University Hospital, Washington DC

<sup>11</sup>Wichita Community Clinical Oncology Program, Wichita, KA

<sup>12</sup>Arizona Cancer Center, University of Arizona, Tucson, AZ

### Abstract

**Purpose**—There is currently no consensus on optimal front-line therapy for patients with follicular lymphomas (FL). We analyzed a Phase III randomized intergroup trial comparing 6 cycles of CHOP-R with six cycles of CHOP followed by iodine I-131 tositumomab radioimmunotherapy (RIT) to assess whether any subsets benefitted more from one treatment or the other, and to compare three prognostic models.

---

Corresponding Author: Oliver W. Press, M.D., Ph.D., Fred Hutchinson Cancer Research Center, D3-190, 1100 Fairview Avenue, Seattle, WA 98109. Phone: (206) 667-1872, Fax: (206) 667-1874, press@u.washington.edu.

\*Oliver Press and Joseph Unger contributed equally to this manuscript and are co-first authors.

### Conflict of Interest Disclosure

Oliver W. Press (Principal Investigator) – consultancy with Roche/Genentech and clinical trial research funding from Roche/Genentech to the Fred Hutchinson Cancer Research Center. Joseph M. Unger – no conflict of interest. Lisa M. Rimsza – no conflict of interest. Jonathan W. Friedberg – consultancy with Genentech. Michael LeBlanc – no conflict of interest. Myron S. Czuczman – consultancy with Genentech Pharmaceuticals and Spectrum Pharmaceuticals. Mark S. Kaminski – research funding from GlaxoSmithKline; royalties from patents on CD20 Radioimmunotherapy. Rita M. Braziel – no conflict of interest. Catherine M. Spier – no conflict of interest. Ajay K. Gopal – consultancy with Seattle Genetics; honoraria from Seattle Genetics and Millennium/Takeda; research funding from Seattle Genetics, GlaxoSmithKline, Spectrum, Lilly, SBio, Piramal, Abbott, and Emergent Biosolutions. David G. Maloney – honoraria from Genentech, Roche, and GlaxoSmithKline; research funding from Genentech. Bruce D. Cheson – consultancy with Roche-Genentech. Shaker Dakhil – no conflict of interest. Thomas P. Miller – no conflict of interest. Richard I. Fisher – consultancy with Roche.

**Experimental Design**—We conducted univariate and multivariate Cox regression analyses of 532 patients enrolled on this trial and compared the prognostic value of the FLIPI, FLIPI2, and LDH +  $\beta$ 2M models.

**Results**—Outcomes were excellent, but not statistically different between the two study arms (5 year PFS of 60% with CHOP-R and 66% with CHOP-RIT [ $p=0.11$ ]; 5-yr OS of 92% with CHOP-R and 86% with CHOP-RIT [ $p=0.08$ ]; overall response rate of 84% for both arms). The only factor found to potentially predict the impact of treatment was serum  $\beta$ 2 microglobulin ( $\beta$ 2M); among patients with normal  $\beta$ 2M, CHOP-RIT patients had better PFS compared to CHOP-R patients, whereas among patients with high serum  $\beta$ 2M, PFS by arm was similar (interaction  $p$ -value=.02).

**Conclusions**—All three prognostic models (FLIPI, FLIPI2, LDH +  $\beta$ 2M) predicted both PFS and OS well, though the LDH +  $\beta$ 2M model is easiest to apply and identified an especially poor risk subset. In an exploratory analysis using the latter model, there was a statistically significant trend suggesting that low risk patients had superior observed PFS if treated with CHOP-RIT, whereas high risk patients had a better PFS with CHOP-R.

## Keywords

Follicular Lymphoma; Prognostic Factors; Subset Analysis;  $\beta$ 2 microglobulin; Front-Line Therapy

## Introduction

Follicular lymphoma (FL) is a common, indolent Non-Hodgkin's lymphoma (NHL) associated with long-term survival with a variety of initial treatment approaches.(1, 2) Recent longitudinal and epidemiologic studies suggest that survival of FL patients has markedly improved in the last 15 years concurrent with the implementation of immunochemotherapy regimens incorporating both chemotherapy and anti-CD20 monoclonal antibodies,(3–8) but there is no consensus on which of these regimens is optimal. In an attempt to address this question, SWOG and CALGB designed a Phase III study in 1999–2000 comparing two of the most promising chemotherapy regimens for FL at the time, namely 6 cycles of CHOP chemotherapy administered with 6 doses of rituximab vs six cycles of CHOP chemotherapy, followed by dosimetric and therapeutic doses of tositumomab and  $^{131}\text{I}$ -tositumomab as consolidative radioimmunotherapy, based on previous promising pilot studies of these regimens.(9–11) The results of this Phase III trial (S0016) have recently been reported(12) and demonstrated that the PFS and OS were excellent on both arms of the study, but not statistically different with 4.9 years of median follow-up. It remains possible, however, that some subsets of patients might benefit more from one regimen or the other. To address this hypothesis, we conducted an exploratory analysis using univariate and multivariate Cox regression to identify subgroups of FL patients with differential outcomes using CHOP-R or CHOP-RIT. In addition, we used this data set to compare and contrast the relative values of three prognostic models for FL, namely, the original follicular lymphoma international prognostic index (FLIPI) model(13), an updated FLIPI2 model(14), or a lab-based model consisting of only the baseline LDH and  $\beta$ 2M values. This manuscript presents the results of these exploratory analyses.

## Materials and Methods

### Eligibility

Details of the protocol eligibility and exclusion criteria have been published elsewhere.(12) In brief, patients over the age of 18 with untreated, measurable bulky stage II or stage III–IV FL (grade 1, 2, or 3) expressing CD20 were eligible if they had a SWOG performance status of 0–2, granulocytes  $\geq 1,500$  cells/ $\mu\text{l}$ , and platelets  $\geq 100,000$ / $\mu\text{l}$ . Bulky adenopathy was

defined as > 10 cm in diameter or greater than one-third the thoracic diameter. Excisional biopsies or large core needle biopsies showing follicular architecture were required; fine needle aspirates and marrow biopsies alone were not sufficient. Diagnostic biopsies were all reviewed centrally by expert SWOG pathologists to confirm the diagnosis of FL according to published consensus morphologic, immunophenotypic, and genetic criteria.(15) Cases with >25% diffuse architecture and >15 centroblasts per high power field were considered diffuse large B cell lymphoma and excluded. Investigators were asked to enroll only patients with FL requiring therapy and not asymptomatic, low tumor burden patients for whom watchful waiting would be appropriate. All patients signed a written informed consent in accordance with institutional and federal guidelines.

### Study Design and Protocol Treatment

Baseline and serial follow-up patient evaluations, laboratory testing, and imaging studies were performed as previously described.(12) Patients received CHOP chemotherapy every 21 days for six cycles using standard doses, supportive care and dose reductions as published.(10–12) Patients on the CHOP-R arm were treated as described by Czuczman et al. with 375 mg/m<sup>2</sup> of rituximab on days 1, 6, 48, 90, 134, and 141 and CHOP chemotherapy on days 8, 29, 50, 71, 92, and 113.(16) Patients on the CHOP-RIT arm received consolidative RIT with tositumomab/<sup>131</sup>I-tositumomab administered 4–8 weeks after the sixth cycle of CHOP.(10–12)

### Statistical Considerations

The primary objective of the research trial was to determine which of the two regimens tested (CHOP-R or CHOP-RIT) was superior in terms of PFS. Patients were randomized to CHOP-R or CHOP-RIT using a dynamic allocation scheme at the time of registration. Data were centrally reviewed and clinical responses assigned by the SWOG statistical center and the principal investigator according to the criteria established in two international workshops.(17, 18) Remission and survival status was assessed 200 days and 365 days after initiation of therapy and then every six months until death. Re-staging was also performed whenever patients developed symptoms or signs of relapse. Patient randomization was balanced according to  $\beta_2$  microglobulin ( $\beta_2$ M) level (>IULN [institutional upper limit of normal] vs. IULN). PFS was defined as the time from registration to the first observation of progressive disease or death due to any cause. Analyses of PFS and OS were performed based on modified “intent-to-treat”, where only patients known to be ineligible were excluded. Multivariate PFS and OS analyses were performed by Cox regression(19) and survival was estimated according to the method of Kaplan and Meier.(20) This study was continuously monitored by a Data and Safety Monitoring Committee (DSMC) and formal interim analyses were performed after 50% and after 75% of eligible patients were randomized.

### Prognostic Factor Analyses

Univariate analyses of the association of baseline clinical and selected laboratory factors with both PFS and OS were performed, with adjustment for treatment arm. In addition, we assessed whether the association of any factor and PFS or OS differed according to treatment using interaction terms in multivariate Cox regression models. Since the analyses were exploratory, the results were not corrected for multiple comparisons.

### Prognostic Model Comparison

FLIPI and FLIPI2 risk scores were calculated as described in the original publications.(13, 14) Each patient’s chart was individually reviewed by both the principal investigator and by a data coordinator in the SWOG statistical center to confirm accurate assignment of patients

to risk groups according to these models. In particular, careful attention was paid to lymph node group assignments with comparison of baseline tumor assessment forms and radiographic reports to published FLIPI lymph node maps.(13) In addition, maximal lymph node diameters were double checked for each patient to assure correct designation of FLIPI2 scores. We compared the performance of the FLIPI and FLIPI2 models to a model based entirely on baseline laboratory tests. Candidate factors for this lab-based risk model must have been significant predictors of both PFS and OS in both the univariate and multivariate regression settings. The factors that remained statistically significant independent predictors of both PFS and OS were standardized according to institutional normal limits. A panel of risk models was then generated based on varying the split points of the continuous factors together in increments of 10%. Best lab-based risk models were determined by comparing Wald Chi-square statistics. Models were fitted using both ordinal categorical variables and dummy variables. To adjust for multiple comparisons for the cut-point models, permutation sampling was used to control the family-wise type 1 error for each cutpoint, based on 1000 samples.(21)

## Results

### Patient Characteristics

Key baseline characteristics of the 532 eligible patients enrolled on this study were as follows: median age, 54 years (range: 23–87 years); male gender, 54%; white race, 90%; elevated  $\beta$ 2M, 54%; median  $\beta$ 2M, 2.2 (range: 0.1–41); B symptoms, 27%; bulk >10 cm or > one third the thoracic diameter, 25%; stage IV, 61%; and intermediate or high risk FLIPI scores, 70%. These (and other) baseline characteristics were well balanced between the two treatment arms without any statistically significant differences.(12) Only 3% of patients had missing baseline LDH, with no difference by arm; all patients had baseline  $\beta$ 2M data.

### Clinical Outcomes

After a median follow-up of 4.9 years, the estimated 5-year PFS was 66% with CHOP-RIT and 60% with CHOP-R (two-sided  $p=0.11$ ) and estimates of 5-year OS were 86% with CHOP-RIT and 92% with CHOP-R ( $p=0.08$ ). The overall response rate was 84% in both the CHOP-R and CHOP-RIT arms, while the complete remission rate was (40%) in the CHOP-R arm and (45%) in the CHOP-RIT arm ( $p=0.30$ ). We further analyzed survival outcomes according to type of remission, though to avoid potential selection and lead-time biases, we only considered outcomes in responders who had achieved at least one year survival without progression using landmark survival analysis. The 5-year PFS for patients with partial remissions was 61%, compared to 70% for patients with complete remissions ( $p=.01$ ). There was no difference in 5-year overall survival between patients with partial (91%) and complete (93%) remissions ( $p=.28$ ). Finally, there was no evidence that the pattern of better PFS for CR patients and similar OS between CR and PR patients were different within treatment arms (data not shown).

### Prognostic Factor Analysis

Univariate and multivariate Cox proportional hazards regression analyses were performed to identify clinical features significantly associated with PFS and OS.(19) Table 1 shows univariate results by baseline factors. Seven factors were significant predictors of PFS at the two-sided,  $\alpha=.05$  level, including maximum lymph node diameter, hemoglobin, performance status,  $\beta$ 2M, LDH, number of lymph nodes and stage (listed in order of decreasing statistical significance). A statistically significant interaction with treatment was evident in only one case, specifically the association between treatment and PFS differed by  $\beta$ 2M level (interaction  $p$ -value=.02; Figure 1). In patients with a normal baseline  $\beta$ 2M level, those treated with CHOP-RIT had a better PFS than those treated with CHOP-R (76% vs

61% after 5 years), while patients with an elevated  $\beta$ 2M had similar PFS with both regimens.

Six factors were statistically significantly associated with OS in a univariate Cox regression analysis, including maximum lymph node diameter, performance status, LDH,  $\beta$ 2M, age, and hemoglobin (listed in order of decreasing statistical significance [Table 1]). None of these variables, however, interacted significantly with treatment arm at the  $p=0.05$  level.

We conducted a multivariable analysis of the factors that were statistically significantly associated with both PFS and OS (that is, hemoglobin, LDH, maximal lymph node size, performance status, and  $\beta$ 2M). The results showed that hemoglobin (HR=1.96, 95% CI: 1.29–2.97,  $p=.002$ ) and maximal lymph node size (HR=1.81, 95% CI: 1.31–2.49,  $p=.0003$ ) retained statistical significance in the multivariable model for PFS, and that LDH (HR=1.90, 95% CI: 1.09–3.31,  $p=.02$ ) and maximal lymph node size (HR=1.96, 95% CI: 1.10–3.49,  $p=.02$ ) retained statistical significance in a multivariable model for OS (Supplemental Table 1).

### Comparison of 3 Prognostic Factor Models for Follicular Lymphoma

We next conducted a multivariable Cox regression analysis to test the ability of the FLIPI and FLIPI2 prognostic models to predict outcomes of patients treated on this protocol (Table 2). For the FLIPI model, each increase in risk level was associated with a HR of 1.55 for PFS ( $p<.0001$ ) and 1.95 for OS ( $p=.0001$ ). The FLIPI2 model performed even better, with each increase in risk level associated with a HR of 1.90 for PFS ( $p<.0001$ ) and 2.65 for OS ( $p<.0001$ ). Figure 2 shows PFS and OS Kaplan-Meier curves for each model. There was marginal evidence that the association of FLIPI and PFS differed by treatment arm (interaction  $p$ -value=.08). A comparison of the risk group distributions in our trial with those of selected other major follicular lymphoma studies demonstrates that our study had fewer low risk patients than either the original FLIPI(13) or FLIPI2 studies(14), but somewhat more than the recent PRIMA trial(6) (Supplemental Table 2).

We also derived a prognostic model based exclusively on routine laboratory tests. Univariate models demonstrated that  $\beta$ 2M, LDH, and hemoglobin were all statistically significantly associated with PFS and OS (Table 1), but only  $\beta$ 2M and LDH remained independent for both PFS and OS in multivariate models that included all 3 factors ( $p >.05$ ). Therefore, among the three laboratory variables analyzed, the candidate lab-based factors for our model were  $\beta$ 2M and LDH. Both factors were standardized by dividing the observed value by the institutional upper limit of normal (IULN), creating a ratio. Each ratio, in parallel, was then allowed to vary by a percentage of the IULN. For instance, patients were categorized according to whether both factors were less than 150% of IULN (coded 0), one or the other factor was greater than 150% of IULN (coded 1), or both factors were greater than 150% of IULN (coded 2). A range of models with ratios from 50% of IULN to 200% (twice) the IULN was explored in 10% increments on each variable; models were then compared using model Wald chi-square statistics. Since each model had zero, one, or two risk factors, we explored models using both ordinal categorical variables, which assumes a linear association – that is, the same increase in risk moving from low to intermediate risk as from intermediate to high risk – as well as a dummy variable model, which does not make a linearity assumption. Supplemental figure 1 shows how the Wald chi-square values, for both PFS (top panel) and OS (bottom panel), achieved maxima at 150% of the IULN. The close tracking of the respective curves for the dummy variable models and the ordinal categorical models suggests an approximately linear association between low, intermediate, and high categories, so the ordinal categorical approach was chosen for analysis. Importantly, multiple comparisons testing showed that each of the cutpoints at 70% of IULN or above was statistically significant at the permutation-adjusted  $p=.01$  level.



Two models were constructed, one splitting each factor at the IULN (for clinical simplicity) and the other model at the optimal cutpoint of 150% of the IULN (for maximum separation of prognostic groups). Results are shown in the bottom of Table 2 and in Figure 3. For the model splitting at the IULN, each increase in risk level was associated with a HR of 1.61 for PFS ( $p < .0001$ ; Figure 3A) and 2.05 for OS ( $p < .0001$ ; Figure 3B). For the model splitting at 150% of IULN, each increase in risk level was associated with a HR of 2.00 for PFS ( $p < .0001$ ; Figure 3C) and 2.61 for OS ( $p < .0001$ ; Figure 3D). Figures 3C and 3D compare the PFS and OS, respectively, of patients in whom both LDH and  $\beta 2M$  were above 150% of the IULN (high risk) compared to those with only one factor  $>150\%$  of the IULN (intermediate risk), and to those with neither factor  $>150\%$  of the IULN (low risk), and shows the outstanding separation of the curves with indicated 5 yr PFS of 67%, 56%, and 20%. As illustrated in this figure, the model splitting at 150% of the IULN identifies an extremely high risk, albeit small, group of patients with very poor PFS and OS.

For both  $\beta 2M + LDH$  models, significant interactions ( $p = .03$  for each) were evident for the association of risk level and PFS, likely reflecting the influence of  $\beta 2M$ . Low risk patients appear to have a better PFS if treated with CHOP-RIT and high risk patients may have a better PFS if treated with CHOP-R (Figure 4A). High risk patients treated with CHOP-RIT had very poor OS (Figure 4B).

Table 2 demonstrates that in this dataset, the simple laboratory-based model split at the IULN performed as well as the FLIPI model, and the 150% model performed as well as the FLIPI 2 model and better than the original FLIPI model. The statistically significant interaction of the lab based model with treatment for PFS suggests the model has predictive as well as prognostic utility. However, it should be noted that since the event rate for overall survival in this single phase III study was low, a split sample validation into training and test datasets was not possible. Because the 150% model represents an optimized model, confirmation that this model outperforms a simpler model based on splitting at IULN would require an independent validation dataset.

## Discussion

Management strategies for FL remain highly controversial. There is no consensus concerning the optimal induction strategy for front-line therapy of indolent NHL and widely diverging views persist on the roles of rituximab “maintenance” therapy or consolidation therapy with RIT, though several well-designed phase III clinical trials have recently been reported addressing these issues. These recent studies have shown that R-CHOP and R-fludarabine-based induction regimens provide superior PFS compared to R-CVP, but that overall survival does not differ significantly with any of the regimens.(6, 22) These trials also demonstrate that fludarabine based induction regimens are more myelosuppressive and increase the risk of secondary malignancies compared to the alkylator-based regimens, which most investigators now prefer.(6, 22) A single Phase III study suggests that Bendamustine plus rituximab induction may afford a superior PFS compared to R-CHOP, and be less toxic, though overall survival is similar.(23) Administration of an extended course of “maintenance” rituximab for two years following completion of induction immunochemotherapy, or a single dose of  $^{90}\text{Y}$ trium-ibritumomab tiuxetan as consolidation have each been shown to markedly prolong PFS in FL patients in separate phase III studies(6, 24, 25), but neither approach has yet been shown to improve overall survival compared to patients treated with induction regimens alone.

The results reported in this manuscript are based on the largest Phase III clinical trial of advanced FL conducted in North America in the last 20 years, which sought to assess whether FL patients consolidated with  $^{131}\text{I}$ odine-tositumomab following six cycles of CHOP

chemotherapy (without any rituximab exposure) would have superior outcomes compared with patients treated with 6 cycles of CHOP plus 6 doses of rituximab. The results of the trial did not demonstrate a statistically significant improvement in PFS or OS for the experimental arm (CHOP + RIT) compared to patients treated with CHOP-R, though outcomes were excellent with either regimen. Though the study demonstrated no differences between the treatment arms in the primary analysis of the entire group of patients,(12) many investigators have queried whether specific subsets of patients might have benefitted more from one regimen or another. We have conducted an exploratory, “hypothesis-generating” subset analysis using Cox analysis to address this question, as reported in this manuscript. In addition, we have used this large patient cohort to compare three prognostic models of FL to ascertain their relative merits.

Univariate regression analysis demonstrated that elevations of baseline LDH or  $\beta$ 2M levels, anemia, large lymph node size, and poor performance status all adversely affected both PFS and OS for patients enrolled on the trial. In addition, the number of enlarged lymph nodes and the stage of disease had univariate prognostic significance for PFS but not OS. Conversely, older age was significantly associated with worse OS, but not PFS. Of all these individual factors, only an elevated  $\beta$ 2M was shown to exhibit a statistically significant interaction with the treatment arm in terms of outcome (Table I, Figure 2). Patients with a normal baseline  $\beta$ 2M level treated with CHOP-RIT showed improved PFS compared to those treated with CHOP-R (76% vs 61% after 5 years, whereas PFS by arm in patients with high  $\beta$ 2M was similar (interaction p-value=.02). This finding is consistent with several prior studies demonstrating the powerful prognostic value of  $\beta$ 2M for patients with indolent NHL(14, 26–29), myeloma(30, 31), and other hematopoietic malignancies(32–35) Although the reason that low risk patients might have superior PFS if treated with frontline RIT consolidation is unclear, this finding is consistent with an independent report that FL patients treated with  $^{90}\text{Y}$ trium-Ibritumomab-Tiuxetan as frontline therapy had a much longer PFS if the baseline LDH was normal than if it was elevated.(36) We are, of course, acutely aware of the limitations of such retrospective subset analyses, particularly when many variables are interrogated, and therefore believe that this finding should be considered hypothesis-generating rather than treatment-defining, until independent prospective studies supporting this analysis are reported.

Since the outcome of patients with FL is highly variable, several prognostic factor models have been developed based on clinical and laboratory features to estimate the survival of newly presenting patients. These prognostic factor models have proven very useful as guides to assist patient counseling, treatment planning and clinical trial interpretation. The most widely employed model, the FLIPI index, was based on a large retrospective analysis of patients diagnosed between 1985 and 1992 who were treated in the “pre-rituximab” era. This index incorporates 5 adverse prognostic factors (age >60, advanced stage, hemoglobin level <120 g/L, >4 nodal areas of involvement, and elevated LDH) and stratifies patients into low, intermediate, and high risk groups based on the presence of 0–1, 2, or >3 adverse risk features, respectively. Importantly, the original FLIPI study excluded  $\beta$ 2M from the multivariable analysis because of the “very high proportion of patients with missing data” in the retrospective analysis.(13) The FLIPI index was found to be more discriminant in terms of OS than the IPI (devised based on features of patients with aggressive lymphomas) and identified three groups of patients with projected 10 year overall survivals of 71% (low risk), 51% (intermediate risk), and 37% (high risk). The FLIPI index has been widely adopted in clinical practice for prognostication and by investigators for patient stratification in clinical trials. Nevertheless, concerns have emerged about the applicability of the index in the modern era with the routine incorporation of rituximab into both front-line and salvage therapies. In addition, reservations have been expressed about the difficulties experienced by clinicians in properly assigning lymph node “groups” defined by the FLIPI index model.

The FLIPI lymph node “groups” do not correspond to the nodal groups defined by the Ann Arbor staging system. Furthermore, the inconsistent designation of bilateral adenopathy as constituting either one or two nodal groups has led to many errors in assigning FLIPI scores, and in the experience of SWOG, has generally led to an over-estimation of the number of lymph node groups and, consequently, higher FLIPI scores. We have therefore found it essential to perform central review of all FLIPI score assignments on SWOG studies of FL to assure they are accurate.

To address these and other concerns, the FLIPI investigators conducted a prospective study to confirm the value of the original FLIPI index in the “rituximab era”, and also to develop a “more accurate” prognostic index. This effort culminated in the development of the FLIPI2 index based on the  $\beta$ 2M level, the longest diameter of the largest involved lymph node ( $>6$  cm), presence of bone marrow involvement, hemoglobin level  $<12$  g/dL, and age over 60 years. The 3 year PFS rates in the FLIPI2 study were 91%, 69% and 51% respectively for patients in the low (0–1 factors), intermediate (2 factors), and high risk (3 factors) groups. Despite its obvious strengths compared to the original FLIPI index, the FLIPI2 index has not been widely adopted, at least in North America, where the original FLIPI index is still more widely employed in both clinical practice and research trials.

We exploited the large size and prospective nature of the S0016 study to compare the strengths of the FLIPI and FLIPI2 indices to each other and to a third, simpler prognostic index based strictly on two baseline laboratory tests (LDH and  $\beta$ 2M). Our data strongly confirm the powerful prognostic value of the FLIPI and FLIPI2 indices for stratifying patients into risk groups for both PFS and OS, and also support the contention that the FLIPI2 index is superior to the original FLIPI score in separating risk groups (Figure 1, Table 2). Furthermore, our data suggest that the simpler LDH +  $\beta$ 2M index is at least comparable to the FLIPI index in prognostic power and may even perform as well as the FLIPI2 index in defining risk groups (Table 2, Figures 1 & 3). Importantly, this simpler model is much easier to compute and utilize. These findings strongly support a previous report from a smaller study of FL patients treated with fludarabine and mitoxantrone conducted by SWOG(29) and a study of intermediate grade lymphomas at MD Anderson(37), validating the prognostic power of the LDH +  $\beta$ 2M risk stratification for NHL. One of the largest unmet clinical needs in management of FL is the upfront identification of the small subgroup of patients destined for early relapse within 2 years of R-CHOP induction. The LDH +  $\beta$ 2M index, particularly with the 150% cutpoint, performs exceptionally well at identifying this poor risk group, which merits consideration of front-line novel therapeutics. The results for the LDH +  $\beta$ 2M models at cutpoints of 70% of IULN or above are statistically significant under multiple comparisons testing, however, there is less assurance that the 150% model would be the optimal model in different settings. We believe that widespread adoption of this simple, powerful prognostic index for FL will prove to be as useful as the International Staging System (ISS) for myeloma (composed of  $\beta$ 2M + albumin(38)) and encourage utilization of the LDH +  $\beta$ 2M index by other investigators and cooperative groups studying FL. Although Wald Chi-Square analysis has demonstrated the maximal power of this model using cutpoints of 150% of the IULN for both variables in this dataset, the model has powerful prognostic significance using any cutpoint at or above 70% of the IULN, and for simplicity in clinical practice we advocate merely using the IULNs for both variables (Figure 3).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



## Acknowledgments

The authors are grateful to Jeri Jardine, Scott Kurrak, Nancy Press and the SWOG Operations and Statistical Offices for data management and administrative assistance with the conduct of this trial and the preparation of this manuscript. In addition, we acknowledge the generous support of Corixa and GlaxoSmithKline, who provided tositumomab and iodine/<sup>131</sup>I iodine tositumomab for this trial, as well as a research grant covering nuclear medicine costs and HAMA testing.

**Financial Support:** This investigation was supported in part by the following PHS Cooperative Agreement grant numbers awarded by the National Cancer Institute, DHHS: CA32102, CA38926, CA11083, CA27057, CA13612, CA35431, CA20319, CA35090, CA35261, CA46282, CA45560, CA58861, CA35281, CA67575, CA35176, CA45377, CA35128, CA76429, CA22433, CA63844, CA12644, CA04919, CA45450, CA35119, CA35178, CA58416, CA67663, CA46441, CA35192, CA37981, CA76447, CA52654, CA16385, CA073590, CA45808, CA58882, CA14028, CA58723, CA74811, CA31946 to SWOG; NCI R01 CA076287 to Dr. Press, and in part by a grant from Corixa and GlaxoSmithKline to SWOG.

## References

1. Press, OW. Follicular Lymphoma. In: Kaushansky, K.; Lichtman, M.; Beutler, E.; Kipps, T.J.; Seligsohn; Prchal, editors. *Williams' Hematology*. 8. New York: McGraw-Hill, Medical Pub. Division; 2010. p. 1565-74.
2. Zelenetz AD, Abramson JS, Advani RH, Andreadis CB, Bartlett N, Bellam N, et al. Non-Hodgkin's lymphomas. *J Natl Compr Canc Netw*. 2011; 9:484–560. [PubMed: 21550968]
3. Fisher RI, LeBlanc M, Press OW, Maloney DG, Unger JM, Miller TP. New treatment options have changed the survival of patients with follicular lymphoma. *J Clin Oncol*. 2005; 23:8447–52. [PubMed: 16230674]
4. Swenson WT, Wooldridge JE, Lynch CF, Forman-Hoffman VL, Chrischilles E, Link BK. Improved survival of follicular lymphoma patients in the United States. *J Clin Oncol*. 2005; 23:5019–26. [PubMed: 15983392]
5. Liu Q, Fayad L, Cabanillas F, Hagemester FB, Ayers GD, Hess M, et al. Improvement of overall and failure-free survival in stage IV follicular lymphoma: 25 years of treatment experience at The University of Texas M.D. Anderson Cancer Center. *J Clin Oncol*. 2006; 24:1582–9. [PubMed: 16575009]
6. Salles G, Seymour JF, Offner F, Lopez-Guillermo A, Belada D, Xerri L, et al. Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomised controlled trial. *Lancet*. 2011; 377:42–51. [PubMed: 21176949]
7. Marcus R, Imrie K, Solal-Celigny P, Catalano JV, Dmoszynska A, Raposo JC, et al. Phase III study of R-CVP compared with cyclophosphamide, vincristine, and prednisone alone in patients with previously untreated advanced follicular lymphoma. *J Clin Oncol*. 2008; 26:4579–86. [PubMed: 18662969]
8. Hiddemann W, Kneba M, Dreyling M, Schmitz N, Lengfelder E, Schmits R, et al. Frontline therapy with rituximab added to the combination of cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) significantly improves the outcome for patients with advanced-stage follicular lymphoma compared with therapy with CHOP alone: results of a prospective randomized study of the German Low-Grade Lymphoma Study Group. *Blood*. 2005; 106:3725–32. [PubMed: 16123223]
9. Czuczman MS, Weaver R, Alkuzweny B, Berlfein J, Grillo-Lopez AJ. Prolonged clinical and molecular remission in patients with low-grade or follicular non-Hodgkin's lymphoma treated with rituximab plus CHOP chemotherapy: 9-year follow-up. *J Clin Oncol*. 2004; 22:4711–6. [PubMed: 15483015]
10. Press OW, Unger JM, Brazier RM, Maloney DG, Miller TP, LeBlanc M, et al. A phase 2 trial of CHOP chemotherapy followed by tositumomab/iodine I 131 tositumomab for previously untreated follicular non-Hodgkin lymphoma: Southwest Oncology Group Protocol S9911. *Blood*. 2003; 102:1606–12. [PubMed: 12738671]
11. Press OW, Unger JM, Brazier RM, Maloney DG, Miller TP, Leblanc M, et al. Phase II trial of CHOP chemotherapy followed by tositumomab/iodine I-131 tositumomab for previously untreated

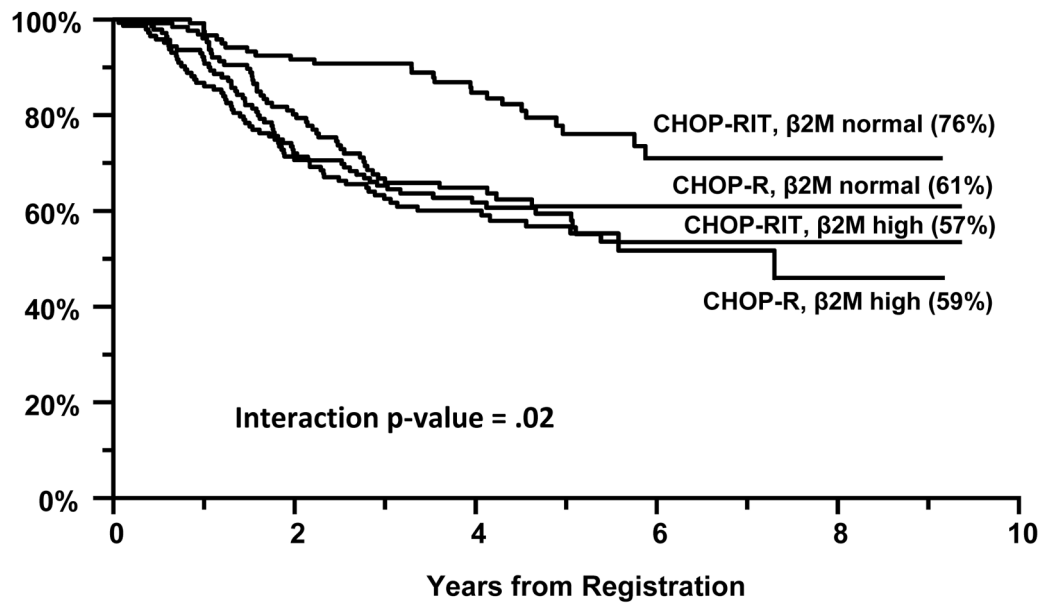
- follicular non-Hodgkin's lymphoma: five-year follow-up of Southwest Oncology Group Protocol S9911. *J Clin Oncol.* 2006; 24:4143–9. [PubMed: 16896003]
12. Press OW, Unger JM, Rimsza LM, Friedberg JW, LeBlanc M, Czuczman MS, et al. Phase III randomized intergroup trial of CHOP plus rituximab compared with CHOP chemotherapy plus (131)iodine-tositumomab for previously untreated follicular non-Hodgkin lymphoma: SWOG S0016. *J Clin Oncol.* 2013; 31:314–20. [PubMed: 23233710]
  13. Solal-Celigny P, Roy P, Colombat P, White J, Armitage JO, Arranz-Saez R, et al. Follicular lymphoma international prognostic index. *Blood.* 2004; 104:1258–65. [PubMed: 15126323]
  14. Federico M, Bellei M, Marcheselli L, Luminari S, Lopez-Guillermo A, Vitolo U, et al. Follicular lymphoma international prognostic index 2: a new prognostic index for follicular lymphoma developed by the international follicular lymphoma prognostic factor project. *J Clin Oncol.* 2009; 27:4555–62. [PubMed: 19652063]
  15. Harris, NL.; Swerdlow, SH.; Jaffe, ES.; Ott, BN.; Nathwani, B.; de Jong, D., et al. Follicular Lymphoma. In: Swerdlow, SH.; Campo, E.; Harris, NL., et al., editors. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues.* Lyon: International Agency for Research on Cancer; 2008. p. 220-6.
  16. Czuczman MS, Grillo-Lopez AJ, White CA, Saleh M, Gordon L, LoBuglio AF, et al. Treatment of patients with low-grade B-cell lymphoma with the combination of chimeric anti-CD20 monoclonal antibody and CHOP chemotherapy. *J Clin Oncol.* 1999; 17:268–76. [PubMed: 10458242]
  17. Cheson BD, Horning SJ, Coiffier B, Shipp MA, Fisher RI, Connors JM, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol.* 1999; 17:2454–60. [PubMed: 10561309]
  18. Cheson BD, Pfistner B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol.* 2007; 25:579–86. [PubMed: 17242396]
  19. Cox DR. Regression models and life tables (with discussion). *J Roy Statist Soc Serv B.* 1972; 34:187–220.
  20. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958; 53:457–81.
  21. Rimsza LM, Unger JM, Tome ME, Leblanc ML. A strategy for full interrogation of prognostic gene expression patterns: exploring the biology of diffuse large B cell lymphoma. *PloS one.* 2011; 6:e22267. [PubMed: 21829609]
  22. Federico M, Luminari S, Dondi A, Sacchi S, Franco V, Pileri S, et al. R-CVP versus R-CHOP versus R-FM as first-line therapy for advanced-stage follicular lymphoma: Final results of FOLL05 trial from the Fondazione Italiana Linfomi (FIL). *J Clin Oncol.* 2012; 30 (suppl; abstr 8006).
  23. Rummel M, Niederle N, Maschmeyer G, Banat AG, von Gruenhagen U, Losem C, et al. Bendamustine plus rituximab (B-R) versus CHOP plus rituximab (CHOP-R) as first-line treatment in patients with indolent and mantle cell lymphomas (MCL): Updated results from the StiL NHL1 study. *J Clin Oncol.* 2012; 30 (suppl; abstr 3).
  24. Morschhauser F, Radford J, Van Hoof A, Vitolo U, Soubeyran P, Tilly H, et al. Phase III trial of consolidation therapy with yttrium-90-ibritumomab tiuxetan compared with no additional therapy after first remission in advanced follicular lymphoma. *J Clin Oncol.* 2008; 26:5156–64. [PubMed: 18854568]
  25. Morschhauser F, Radford J, Van Hoof A, Botto B, Rohatiner AZ, Salles G, et al. 90Yttrium-Ibritumomab Tiuxetan Consolidation of First Remission in Advanced-Stage Follicular Non-Hodgkin Lymphoma: Updated Results After a Median Follow-Up of 7.3 Years From the International, Randomized, Phase III First-Line Indolent Trial. *J Clin Oncol.* 2013; 31:1977–83. [PubMed: 23547079]
  26. Amlot PL, Adinolfi M. Serum beta 2 microglobulin and its prognostic value in lymphomas. *Eur J Cancer.* 1979; 15:791–6. [PubMed: 92408]
  27. Canovas A, Alonso JJ, Barreiro G, Aguirre C. Prognostic factors in follicular lymphoma: the importance of beta-2 microglobulin. *Tumori.* 2010; 96:117–21. [PubMed: 20437868]

28. Litam P, Swan F, Cabanillas F, Tucker SL, McLaughlin P, Hagemester FB, et al. Prognostic value of serum beta-2 microglobulin in low-grade lymphoma. *Annals of internal medicine*. 1991; 114:855–60. [PubMed: 2014946]
29. Velasquez WS, Lew D, Grogan TM, Spiridonidis CH, Balcerzak SP, Dakhil SR, et al. Combination of fludarabine and mitoxantrone in untreated stages III and IV low-grade lymphoma: S9501. *J Clin Oncol*. 2003; 21:1996–2003. [PubMed: 12743154]
30. Durie BG, Stock-Novack D, Salmon SE, Finley P, Beckord J, Crowley J, et al. Prognostic value of pretreatment serum beta 2 microglobulin in myeloma: a Southwest Oncology Group Study. *Blood*. 1990; 75:823–30. [PubMed: 2405920]
31. Bataille R, Klein B. Serum beta-2-microglobulin (beta 2m) in myeloma: toward a simple prognostic stratification using beta 2M and acute-phase proteins? *Blood*. 1991; 77:1616–7. [PubMed: 1706956]
32. Neumann F, Gattermann N, Barthelmes HU, Haas R, Germing U. Levels of beta 2 microglobulin have a prognostic relevance for patients with myelodysplastic syndrome with regard to survival and the risk of transformation into acute myelogenous leukemia. *Leuk Res*. 2009; 33:232–6. [PubMed: 18639338]
33. Kantarjian HM, Smith T, Estey E, Polyzos A, O'Brien S, Pierce S, et al. Prognostic significance of elevated serum beta 2-microglobulin levels in adult acute lymphocytic leukemia. *The American journal of medicine*. 1992; 93:599–604. [PubMed: 1466355]
34. Chronowski GM, Wilder RB, Tucker SL, Ha CS, Sarris AH, Hagemester FB, et al. An elevated serum beta-2-microglobulin level is an adverse prognostic factor for overall survival in patients with early-stage Hodgkin disease. *Cancer*. 2002; 95:2534–8. [PubMed: 12467067]
35. Johnson PW, Whelan J, Longhurst S, Stepniewska K, Matthews J, Amess J, et al. Beta-2 microglobulin: a prognostic factor in diffuse aggressive non-Hodgkin's lymphomas. *British journal of cancer*. 1993; 67:792–7. [PubMed: 8471438]
36. Scholz CW, Pinto A, Linkesch W, Linden O, Viardot A, Keller U, et al. Yttrium-90-Ibritumomab-Tiuxetan as First Line Treatment for Follicular Lymphoma. 30 Months Follow up Data from an International Multicenter Phase II Clinical Trial. *J Clin Oncol*. 2012 (in press).
37. Swan F Jr, Velasquez WS, Tucker S, Redman JR, Rodriguez MA, McLaughlin P, et al. A new serologic staging system for large-cell lymphomas based on initial beta 2-microglobulin and lactate dehydrogenase levels. *J Clin Oncol*. 1989; 7:1518–27. [PubMed: 2674337]
38. Greipp PR, San Miguel J, Durie BG, Crowley JJ, Barlogie B, Blade J, et al. International staging system for multiple myeloma. *J Clin Oncol*. 2005; 23:3412–20. [PubMed: 15809451]

### Statement of Translational Relevance

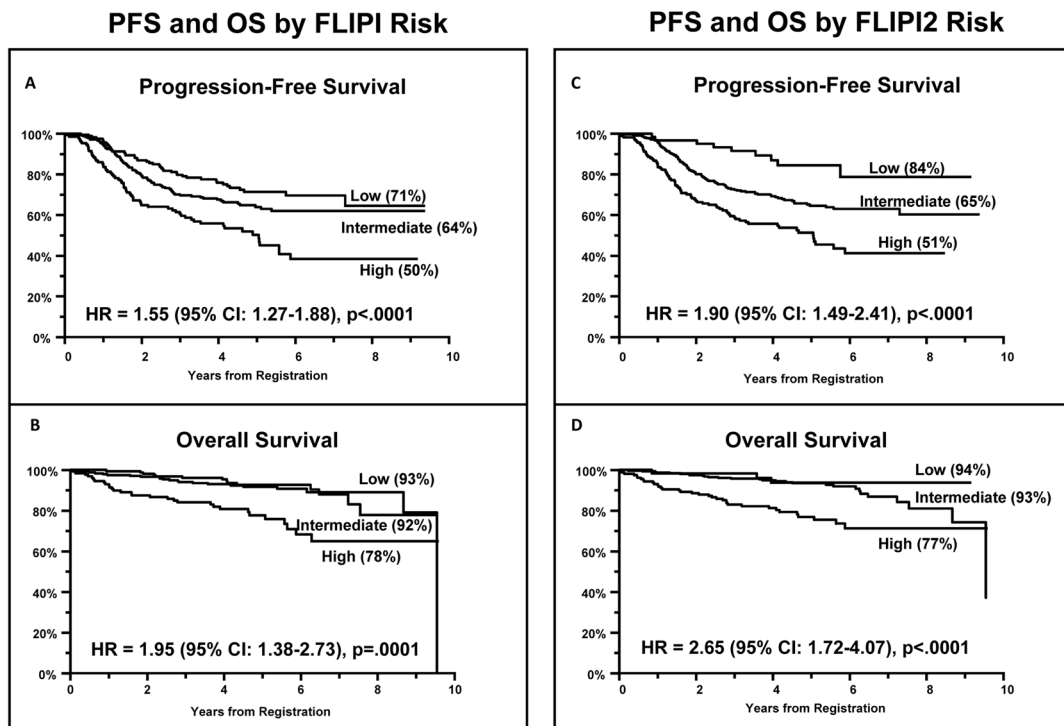
This manuscript reports results from the largest Phase III clinical trial of follicular lymphoma conducted in North America in the last 20 years, comparing two immunochemotherapy options, namely, CHOP-Rituximab and CHOP + <sup>131</sup>Iodine-Tositumomab. It presents a Cox analysis suggesting that patients with a normal  $\beta 2$  microglobulin level treated with CHOP-RIT show an improved progression-free survival compared to those with CHOP-R (76% vs 61% after 5 years, respectively [interaction p-value=.02]). It also compares 3 prognostic models for follicular lymphoma (FLIPI1, FLIPI2, and  $\beta 2$  + LDH) and assesses their relative merits. The findings of this large study will be helpful to hematologists and oncologists as they evaluate their patients and choose immunochemotherapy regimens.

### Progression-Free Survival By Treatment and $\beta$ 2M

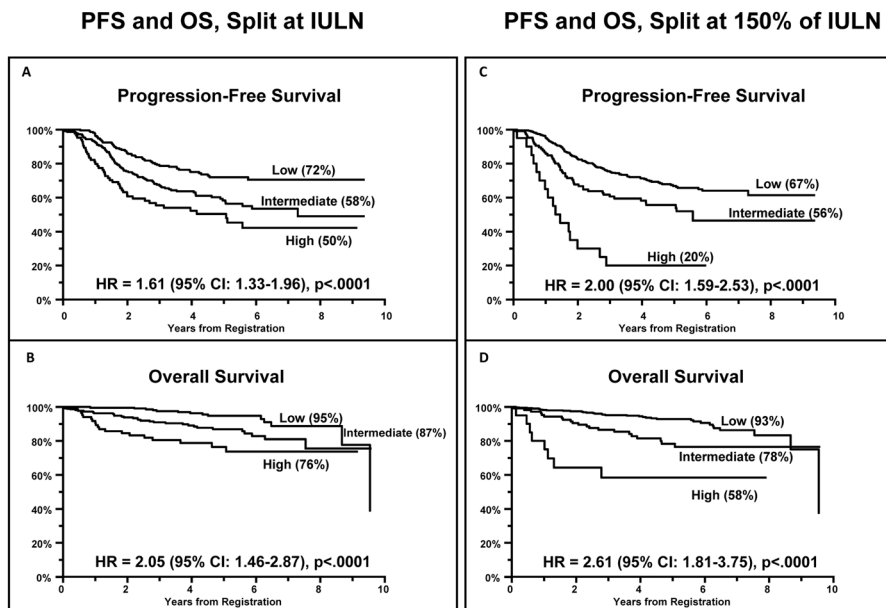


**Figure 1.** Progression Free Survival of Advanced Follicular Lymphoma Patients Enrolled on Intergroup Protocol S0016, stratified according to treatment arm and by serum beta 2 microglobulin Level.

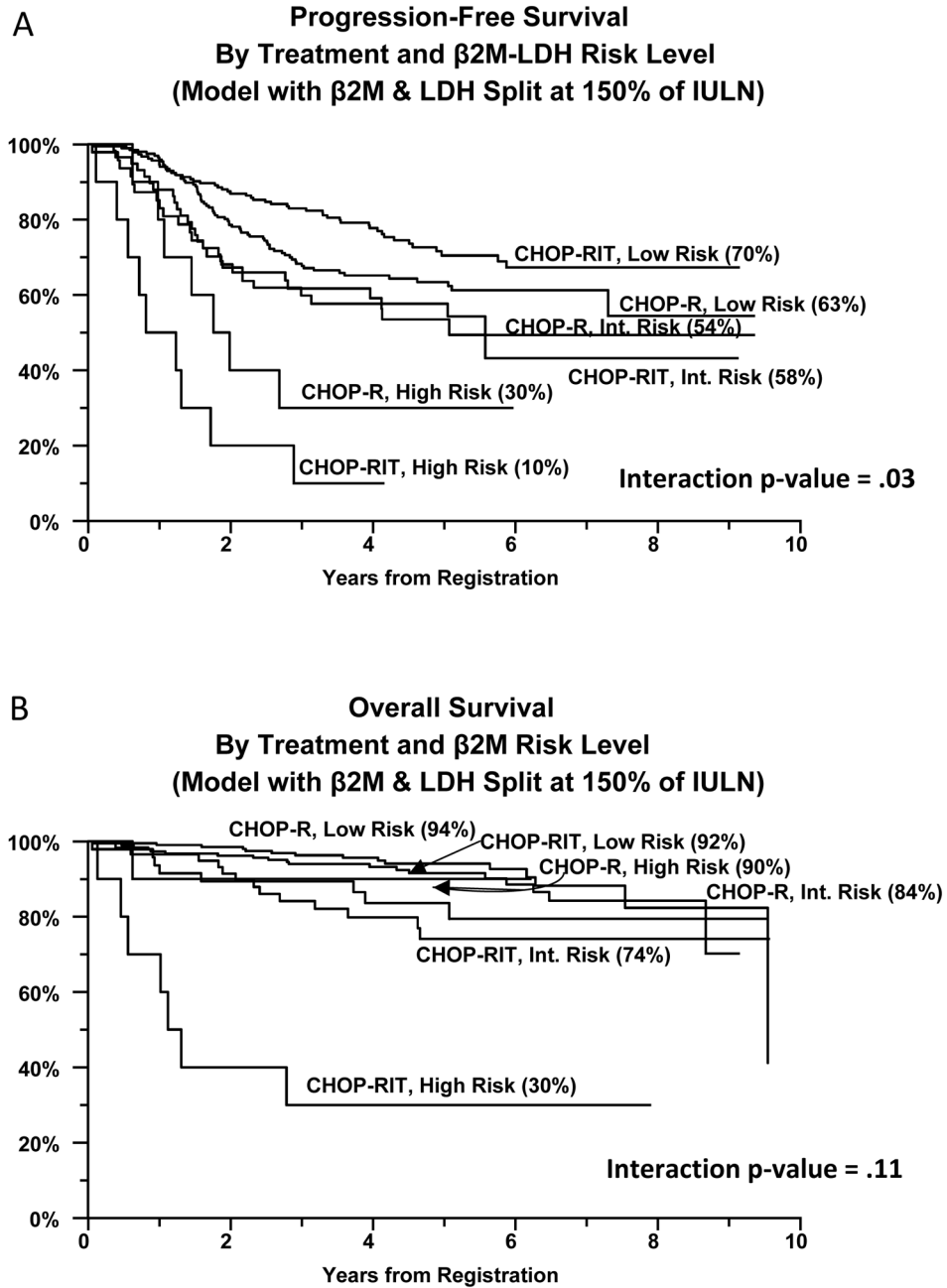




**Figure 2.** Progression Free Survival (2a, 2c) and Overall Survival (2b, 2d) of Advanced Follicular Lymphoma Patients Enrolled on Intergroup Protocol S0016, stratified according to FLIPI (2a, 2b) or FLIPI2 Risk Groups (2c, 2d).



**Figure 3.** Progression Free Survival (3A, 3C) and Overall Survival (3B, 3D) of Advanced Follicular Lymphoma Patients Enrolled on Intergroup Protocol S0016, stratified according to the  $\beta$ 2M + LDH Prognostic Model using either the Institutional Upper Limit of Normal (3A, 3B) or 150% of the IULN (3C, 3D) as cutpoints for both variables.



**Figure 4.** Progression Free Survival (4A) and Overall Survival (4B) of Advanced Follicular Lymphoma Patients Enrolled on Intergroup Protocol S0016 stratified according to risk group defined by the LDH + beta 2 microglobulin prognostic model (split at 150% of the IULN) and treatment arm.

**Table 1**  
Univariate Analyses of Baseline Factors of 554 Patients Enrolled on Intergroup Protocol S0016

Factor	Factor Levels (Codes)	Progression-Free Survival			Overall Survival		
		CHOP-RIT	CHOP-R	Univariate HR (95% CI) P=	Interaction With Tx P=	Univariate HR (95% CI) P=	Interaction With Tx P=
Age	60 (0)	72%	69%	1.21 (0.83-1.52) P=.47	P=.10	1.94 (1.18-3.18) P=.009	P=.06
	>60 (1)	28%	31%				
Bone Marrow Involvement	No (0)	44%	44%	1.29 (0.96-1.72) P=.09	P=.73	1.00 (0.61-1.62) P=.99	P=.35
	Yes (1)	56%	56%				
Hemoglobin	12 (0)	89%	89%	2.14 (1.46-3.16) P=.0001	P=.39	2.14 (1.13-4.05) P=.02	P=.39
	<12 (1)	11%	11%				
LDH	IULN (0)	74%	78%	1.60 (1.17-2.18) P=.003	P=.59	2.45 (1.48-4.05) P=.0005	P=.79
	>IULN (1)	26%	22%				
Maximal LN Size	6 cm (0)	61%	60%	1.95 (1.47-2.59) P<.0001	P=.20	2.62 (1.59-4.30) P=.0002	P=.34
	>6 cm (1)	39%	40%				
Number of Nodes	4 (0)	61%	61%	1.52 (1.14-2.02) P=.004	P=.75	1.22 (0.75-1.99) P=.43	P=.92
	>4 (1)	39%	39%				
Performance Status	0 (0)	74%	74%	1.72 (1.27-2.34) P=.0004	P=.93	2.12 (1.27-3.51) P=.004	P=.69
	1 or 2 (1)	26%	26%				
Sex	Female (0)	45%	47%	1.12 (0.84-1.49) P=.43	P=.68	1.50 (0.90-2.47) P=.12	P=.30
	Male (1)	55%	53%				
Serum-β2M	IULN (0)	46%	47%	1.69 (1.26-2.27) P=.0004	P=.02	2.22 (1.31-3.75) P=.003	P=.10
	>IULN (1)	54%	53%				
Spleen Involvement	No (0)	77%	80%	1.14 (0.81-1.59) P=.45	P=.06	0.85 (0.46-1.56) P=.59	P=.10
	Yes (1)	23%	20%				
Stage	<Stage IV (0)	36%	41%	1.42 (1.05-1.93) P=.02	P=.19	1.11 (0.67-1.85) P=.68	P=.49
	IV (1)	64%	59%				

Factor	Factor Levels (Codes)	CHOP-RIT		CHOP-R		Progression-Free Survival			Overall Survival		
		74%	26%	71%	29%	Univariate HR (95% CI)	Interaction With Tx	Univariate HR (95% CI)	Interaction With Tx	Univariate HR (95% CI)	Interaction With Tx
Symptoms	A (0) B (1)					0.85 (0.61-1.17) P=.32	P=.19	1.16 (0.68-1.96) P=.58	P=.65		



**Table 2**  
Comparison of Multivariable Models for Outcome Parameters of Patients Enrolled on Intergroup Protocol S0016

Factor	Factor Levels (Codes)	Percent in Group			Progression-Free Survival			Overall Survival		
		Overall	CHOP-RIT	CHOP-R	Univariate HR (95% CI)	Interaction With Tx	Univariate HR (95% CI)	Interaction With Tx	Univariate HR (95% CI)	Interaction With Tx
<b>FLIPI</b>	Low (0)	30%	31%	30%	1.55 (1.27-1.88) P<.0001	.08	1.95 (1.38-2.73) P=.0001	.27		
	Intermediate (1)	45%	43%							
	High (2)	24%	26%							
<b>FLIPI2</b>	Low (0)	11%	13%	10%	1.90 (1.49-2.41) P<.0001	.11	2.65 (1.72-4.07) P<.0001	.11		
	Intermediate (1)	59%	56%	62%						
	High (2)	30%	31%	28%						
<b>B2M-LDH</b> Both split at the IULN	Low (0)	41%	40%	43%	1.61 (1.33-1.96) P<.0001	.03	2.05 (1.46-2.87) P<.0001	.23		
	Intermediate (1)	42%	41%	43%						
	High (2)	17%	19%	15%						
<b>B2M-LDH</b> Both split at 150% of IULN	Low (0)	75%	73%	78%	2.00 (1.59-2.53) P<.0001	.03	2.61 (1.81-3.75) P<.0001	.11		
	Intermediate (1)	21%	23%	19%						
	High (2)	4%	4%	4%						