Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Zhuowei Wang Date

Effect of Surveillance Imaging in Relapse Detection on Survival of Patients with Follicular Lymphoma

By

Zhuowei Wang
Master of Public Health

Biostatistics and Bioinformatics

Zhengjia (Nelson) Chen
Thesis Advisor

Paul Weiss
Reader

Effect of Surveillance Imaging in Relapse Detection on Survival of Patients with Follicular Lymphoma

By

Zhuowei Wang

B.S.
Tianjin University
2018

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics and Bioinformatics
2020

Abstract

Effect of Surveillance Imaging in Relapse Detection on Survival of Patients with Follicular Lymphoma

By Zhuowei Wang

Background: Serving as an approach for early detection of tumor relapse, the routine implementation of surveillance imaging (SI) is still highly controversial due to its disadvantages like low specificity, huge costs and so on. Most importantly, previous studies did not show improvement in overall survival (OS) by SI. This study focused on patients with recrudescent follicular lymphoma (FL) and divided them into clinical detection group and SI detection group to examine whether SI results in better survival outcomes compared with traditional clinical detection methods.

Methods: We first conducted descriptive analysis on patients' demographic and disease characteristics. Then survival analyses were performed starting by comparing survival probabilities and survival curves between two groups of patients using Kaplan-Meier method. After that, simple and multiple Cox proportional hazard models were built to investigate the association between detection methods and three survival outcomes. Supremum test and standardized score process plots were finally applied to check the proportional hazard assumption.

Results: In LEAD cohort, the hazard ratios of SI detection versus clinical detection were 1.71 (95% CI: 0.77, 3.80, p=0.187), 0.14 (95% CI: 0.10-3.36, p=0.222) and 0.28 (95% CI: 0.03-2.51, p=0.255) for PFS, OS from diagnosis and OS from relapse, respectively, in each multiple Cox model. In MER cohort, the values were 1.05 (95% CI: 0.71-1.55, p=0.806), 1.17 (95% CI: 0.53-2.59, p=0.691) and 1.10 (95% CI: 0.50-2.47, p=0.799), respectively. Although Cox models and hazard ratios are different in these two cohorts, similar results of no statistically significant association between detection method and survival outcomes were generated.

Conclusion: The early detection of relapse by SI did not bring in much improvement in survival outcome for patients with FL in our study so the necessity of routine utilization of SI should be reconsidered. Further studies in prospective cohorts with more balanced patient and disease characteristics would be needed to validate these findings in FL and in other indolent lymphomas.

Effect of Surveillance Imaging in Relapse Detection on Survival of Patients with Follicular Lymphoma

By

Zhuowei Wang

B.S.
Tianjin University
2018

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD/

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics and Bioinformatics
2020

Acknowledgements

Upon the time of completing my graduation project and this thesis, I want to express my sincere gratitude to people who accompanied me in the two years.

I am grateful to Professor Nelson Chen for his help and instruction on this thesis project. He was always available when I faced problems. His patience, enthusiasm on his work and spirit of exploration encouraged me and are also valuable qualities that I should learn from in my future career.

I also wish to show my gratitude to all the faculty and staff of Department of Biostatistics and Bioinformatics in Rollins School of Public Health, especially my thesis reader - Professor Paul Weiss, thank you all for your time, instruction, assistance, and support.

Finally, I am indebted to my parents. I cannot finish my master's degree without their financial and emotional support.

Table of Contents

1. Introduction

Follicular lymphoma (FL) is a type of cancer that is characterized by the clonal proliferation of neoplastic lymphoid cells which have similar morphological, immunophenotypic and molecular genetic properties to germinal center B cells.[1] As the second most common B-cell non-Hodgkin lymphomas (NHL), it makes up approximately 35% of NHLs and 70% of indolent lymphomas.[2] Each year, there are 13,000 to 15,000 newly diagnosed cases of FL in USA.[3] Previous study showed that the incidence of FL was positively associated with age, which is also explained by a median age at diagnosis of 65 years.[2, 3] To determine the invasiveness of FL, the number of centroblasts can be quantified as an indication. Based on this, the World Health Organization (WHO) classifies FL into four grades, among which grades 1, 2 and 3A have typical morphological characteristics of FL, while grade 3B is closer to de novo aggressive lymphoma.[4]

Being described as indolent, FL progresses slowly and responds very well to treatments.[5, 6] Usually, different treatments are assigned to patients based on their tumor stage.[2] According to the Ann Arbor staging system, tumors are divided into four principal stages by their location. At present, radiation therapy is best choice for patients with stage I/II tumor, which leads to 10-year overall survival (OS) rates of 60% to 80%. However, patients with early stage tumors make up less than 10% of the population diagnosed with FL.[2] Unfortunately, most patients have stage III or IV at diagnosis, although generally without symptoms.[7, 8] For these patients, anti-CD20 antibody-based therapy (e.g. rituximab) and chemotherapy are the recommended standard of care and lead to significantly improved outcomes in FL.[2]

Despite the development of treatments and the high response rate, FL is still incurable due to the occurrence of relapse.[9] Relapse is the phenomenon that the lymphoma comes back after the

achievement of a complete remission (CR). Approximately 20% of patients with stage II, III and IV FL will relapse within 2 years of first-line therapy (as defined as: the initial treatment recommended for a disease) and most patients will experience several relapses over their lifetime.[3, 10] Moreover, these relapses can gradually become aggressive and hard to manage, with some cases transforming into aggressive lymphoma, such as diffuse large B cell lymphoma (DLBCL).[3]

In order to detect relapse and to provide timely salvage therapies, a long follow-up period is necessary for patients with FL. National Cancer Institute (NCI)-sponsored international working group suggests that, after finishing treatments, patients should have follow-up visits quarterly for 2 years, then semiannually for 3 years and finally once a year for at least 5 years. [11] During these visits, patients are asked about symptoms like weight loss, night sweats and fevers and clinical and laboratory studies are also performed. [12] Apart from these, surveillance imaging (SI), also called surveillance scanning, plays the role of a secondary screening assessment for relapse detection.[10] With the emergence of advanced imaging technologies like computed tomography (CT) and positron emission tomography (PET) scanning, SI has become a common and standard practice in most of the US for screening assessments.[13] These improved SI methods are more efficient for diagnosis than X-rays and are less invasive than surgery and lymphangiograms.[14] According to guidelines from National Comprehensive Cancer Network (NCCN), routine CT scanning is suggested to be conducted semiannually for 2 years for patients who have achieved CR. [13]

Although SI may lead to a higher rate of early relapse detection in asymptomatic patients, the implementation of SI is still highly debated. Firstly, partially due to its low specificity, SI has

caused anxiety among patients by the potential for false-positive outcomes.[10] In addition, in practice SI screening is not cost-effective with the detection of one recurrence requiring an estimated 8.3 or more scans. This results in a cost of several thousand dollars on average.[14] Moreover, the radiation-induced risk of secondary malignancies is another limitation of SI. In a recent study conducted by Brenner and Hall, up to 2% of all cancers are estimated to be caused by radiation from CT scans.[15] Finally, previous studies focusing on Hodgkin lymphomas or aggressive NHL (DLBCL) have reported that SI for earlier relapse detection did not result in significant improvement in overall survival (OS) and progression-free survival (PFS).[14, 15] However, this conclusion is still unproved for indolent NHL like FL.

In this study, we retrospectively evaluated the role of routine SI following first-line therapy in relapse detection and the subsequent effect on overall survival (OS) and progression-free survival (PFS) for patients with FL. This was conducted using survival analyses on an institutional cohort of FL patients at Emory University. Same analyses were then carried out in another patient cohort from Molecular Epidemiology Resource (MER) of the University of Iowa/Mayo Clinic to compare with the results in LEAD cohort. For the following part of this thesis, we will describe in detail the two patient cohorts as well as Kaplan-Meier method and cox-proportional hazards model that are used for survival analyses in Section 2. Then we will summarize data analysis results and our findings in Section 3. Finally, we will discuss the conclusion and limitations of our study in Section 4.

2. Method

2.1 Study Population

The data of LEAD cohort came from Lymphoid Malignancies Enterprise Architecture Database (LEAD) which was approved by Institutional Review Board (IRB) and contained medical information for patients with lymphoma diagnosed or treated at Emory University. Patients who did not experience disease relapse were not included in the study. Among patients with relapse, those who were diagnosed with FL between July 1991 and July 2016 and older than 18 years old at diagnosis were included. Another criterion for inclusion was that patients should achieve partial or complete response or a stable disease status after receiving first-line therapy. Also, those with grade 3B FL were excluded from the study since FL in grade 3B is more similar to an aggressive lymphoma but not an indolent lymphoma that our study focused on.

The data of MER cohort which was used for result validation came from the University of Iowa/Mayo Clinic (UI/MC) Lymphoma Specialized Program of Research Excellence (SPORE). Patients in this program were enrolled from 2002 to 2015 and must be diagnosed with FL within 9 months of enrollment. The inclusion or exclusion criteria were the same as those for LEAD cohort.

These two datasets contained patients' demographic, disease, and relapse detection information. The variable of interest for our study was the detection methods for relapse, which was classified into two categories – clinical detection and radiographic detection. Clinical detection meant that the relapse was confirmed by patient-reported symptoms and abnormal physical exam or

laboratory study results, without indication from radiological surveillance. Radiographic detection was defined as the relapse that was first indicated by routine surveillance imaging (SI) and then confirmed by other diagnostic methods. Other covariates included gender, race, age at diagnosis, grade and Ann Arbor stage of tumor, presence of B-symptoms, Eastern Cooperative Oncology Group (ECOG) score and FLIPI score. Time information used for generating outcome variables OS and PFS included date of diagnosis, date of death or date of last contact.

2.2 Statistical Analysis Method

2.2.1 Descriptive analysis

The descriptive statistics were generated for all the covariates stratified by relapse detection methods in LEAD and MER cohorts, separately. For categorical variables, the frequencies and corresponding percentages of each level were presented. For continuous variables, the means and 95% confidence intervals were presented. The difference in distribution of covariates within each relapse detection group was compared using Chi-square test or Fisher ‘s exact test for categorical variables and ANOVA test or Kruskal-Wallis test for continuous variables.

2.2.2 Survival analysis

Three survival outcomes – PFS, OS from diagnosis and OS from relapse – were of interest in the study. PFS was defined as the time in years from the date of diagnosis to the date of relapse. OS from diagnosis was defined as the time in years from the date of diagnosis to the date of death or last contact. OS from relapse was defined as the time in years from the date of relapse to the date of death or last contact.

Kaplan-Meier method was used for estimating survival probabilities in 1 year, 3 years and 5 years for all the three outcomes. Also, median survival time were estimated for PFS as no one was censored in terms of this outcome. The estimator for survival probability was given by:

$$\hat{S}_{KM(t)} = \prod_{k:t_{(k)} \leq t} \left(1 - \frac{d_k}{n_k}\right)$$

where $\hat{S}_{KM(t)}$ was the estimated survival probability at time t; $t_{(k)}$ was the sorted observed event time; n_k was the size of the risk set at time $t_{(k)}$ and d_k was the number of events at time $t_{(k)}$. The Greenwood's formula for standard error of $\hat{S}_{KM(t)}$ was given by:

$$\widehat{SE}\{\hat{S}_{KM(t)}\} = \sqrt{\hat{V}\{\hat{S}_{KM(t)}\}} = \{\hat{S}_{KM(t)}\}^2 \sum_{k:t_{(k)} \leq t} \frac{d_k}{n_k(n_k - d_k)}$$

The 95% confidence interval was calculated using log(-log) transformation in SAS. Survival curves of three survival outcomes for clinical detection group and SI detection group were also plotted.

Simple Cox proportional hazard (Cox PH) models were used for analyzing the association between survival outcomes and each risk factor. Multivariable Cox models were also built to compare survival outcomes between clinical detection group and radiographic detection group while controlling for the effect of other covariates. To get the best model for explaining each survival outcome, variables included in the model were selected using backward elimination method with a criterion of removal at the alpha level of 0.2. The formulation of a Cox proportional hazard model was given by:

$$h(t|Z_i) = h_0(t) \cdot \exp\left(\beta_i{}^T Z_i\right)$$

where $h(t|Z_i)$ was the hazard rate at time t given values of all the covariates Z_i within a subject; $h_0(t)$ is the baseline hazard when $Z_i = 0$.

To evaluate the proportional hazard assumption in Cox models, Supremum test was conducted, and the standardized score process plots were generated.

All the analyses were performed using SAS 9.4 with two-sided tests and a significant level of 0.05.

3. Result

3.1 Descriptive analysis

Descriptive statistics of two patient cohorts were combined and shown in Table 1.

There were totally 53 eligible patients in the LEAD cohort. In this cohort, 35 (66.0%) patients had their relapses detected clinically while the other 18 (34.0%) were detected by SI. Only the proportion of patients who had B-symptoms was significantly different (p-value=0.018) between clinical detection group (n=13, 40.6%) and radiographic detection group (n=1, 5.9%). In MER cohort, there were totally 113 eligible patients, among which 63 (55.8%) had their relapse detected by clinical approach and 50 (44.2%) were detected radiographically. There was no significant difference in all the variables between two groups.

3.2 Survival analysis

3.2.1 LEAD cohort

The Kaplan-Meier survival probabilities in 1 year, 3 years and 5 years in terms of three outcomes were summarized in Table 2 and the survival curves comparing two relapse detection groups were shown in Figure 1. Hazard ratios (HRs) and p-values given by all the Cox PH models in LEAD cohort were summarized in supplementary Table S1. The estimated median PFS time in this cohort is 4.2 years (95% CI: 2.5-6) in clinical detection group and 3.6 years (95% CI: 1.8-4.4) in SI detection group. Although it seemed that SI led to an earlier detection of relapse based on median PFS time, there was no statistically significant difference in survival probabilities between two groups with a p-value of 0.2895 given by the two-sample log-rank test. The simple Cox PH model gave the same result with a p-value of 0.291 in Wald test (HR of SI group=1.37, 95% CI: 0.76-2.45). After controlling for the confounders – ethnicity, ECOG, FLIPI, B-symptoms and age, no apparent difference in patient survival was detected from the multivariable model (HR of SI group=1.71, 95% CI: 0.77-3.80, p-value=0.187). In terms of the other two survival outcomes, survival probabilities and risk of death were also not significantly different between patients in two groups. For OS from diagnosis, the p-value was 0.5324 in log-rank test and 0.537 in simple Cox model (HR of SI group=0.61, 95% CI: 0.13-2.94). In the multivariable Cox model with covariates gender, stage, grade, ECOG, B-symptoms and age, the HR was 0.14 (95% CI: 0.01-3.36) with a p-value of 0.222. For OS from relapse, the p-value was 0.3346 in log-rank test and 0.346 in simple Cox model (HR of SI group=0.47, 95% CI: 0.10-2.28). In the multivariable Cox model with covariates grade and ECOG, the HR was 0.28 (95% CI: 0.03-2.51) with a p-value of 0.255.

Results of the Supremum tests for checking proportional hazard assumption in the three multivariate Cox models in LEAD cohort were summarized in Table S3. None of the variables included in the three models violated the PH assumption as all the test results were non-significant at the alpha level of 0.05.

3.2.2 MER cohort

For this study cohort, the Kaplan-Meier median survival time and survival probabilities were summarized in Table 3 and the survival curves were shown in Figure 2. Hazard ratios (HRs) and p-values given by all the Cox PH models were summarized in supplementary Table S2. The estimated median PFS time in this cohort is 2 years (95% CI: 1.6-2.6) in clinical detection group and 2.4 years (95% CI: 1.9-2.8) in SI detection group. Similar to LEAD cohort, there was no statistically significant difference in PFS between two groups with a p-value of 0.9260 in log-rank test. Moreover, SI did not even show its advantage of earlier relapse detection in this cohort for a longer median PFS time in SI group rather than in clinical group. The simple Cox PH model gave the same result with a p-value of 0.926 in Wald test (HR of SI group=0.98, 95% CI: 0.67-1.44). After controlling for the effects of covariates – stage and grade, the HR of SI group was 1.05 (95% CI: 0.71-1.55) and this difference was still not significant (p-value=0.806). There was also no significant difference in the risk of death between the two groups. For OS from diagnosis, the p-value was 0.9624 in log-rank test and 0.962 in one-variable Cox model (HR of SI group=1.02, 95% CI: 0.47-2.21). In the multivariable Cox model with stage, ECOG and age as covariates, the HR was 1.17 (95% CI: 0.53-2.59) with a p-value of 0.691. For OS from relapse, the p-value was 0.9603 in log-rank test and 0.960 in one-variable Cox model (HR of SI group=1.02, 95% CI: 0.47-2.21). In the multivariable Cox model with covariates ECOG, FLIPI and age, the HR was 1.10 (95% CI: 0.50-2.47) with a p-value of 0.799.

Results of the Supremum tests for checking proportional hazard assumption in the three multivariate Cox models in MER cohort were summarized in Table S4. The PH assumption held well for nearly all the variables except for ECOG in the model where OS from relapse was the outcome.

4. Conclusion and Discussion

This study evaluated whether earlier detection of relapse by Surveillance Imaging had a positive effect on survival outcomes in patients with FL achieving remission after receiving first-line therapy in two cohorts. By first comparing the estimated survival probabilities in clinical detection group versus radiographic detection group, we found that patients who had their disease relapse detected by SI did not show significant improvements in PFS, OS from diagnosis and OS from relapse in both of the two cohorts. This result was reasonable for PFS since detection methods definitely will not have effect on disease relapse. However, the longer median PFS time in SI detection group in MER cohort raised our doubt about the ability of SI for earlier relapse detection. In addition, although in the plot of OS from diagnosis in Figure 1, the radiographic detection group showed an improved survival probability compared to clinical detection group, this improvement did not appear until about 10 years of survival. However, this time is relatively long for patients with FL relapse and a 10-year survival probability of about 0.8 is considerable as mentioned in introduction part, which means SI still did not bring obvious benefit in OS from diagnosis. In single-variable and multivariable analyses based on Cox models built for each survival outcome, although variables included in the models and corresponding values of hazard ratio were different in two cohorts, the similar thing was that no statistically significant

association was observed between detection methods and survival outcomes, which also weakened the usefulness of SI.

In a similar study conducted by Liedtke et al, they found that for patients with aggressive NHL (DLBCL), there was no significant difference in median overall 5-year survival between those whose relapses were detected by routine imaging and those by abnormal exam results or reported symptoms.[16] Our results were consistent with their findings of no significant associations between detection method and survival outcomes although they also mentioned that routine surveillance scanning helped to identify patients who might have a better outcome based on the age-adjusted international prognostic index determined at the time of relapse (sAAIPI). Moreover, our findings refined the results in Truong's study, which also indicated no difference in survival outcomes between the two groups for both aggressive and indolent NHL, by focusing on a specific type of indolent NHL – FL.[13]

Obviously, there are some limitations in our study. Firstly, since this was a retrospective study using past medical records, we were unable to make sure that patients were randomized to receive SI. This may lead to bias in our results because potential differences may exist between patients whose relapse was detected by clinical method and those by SI. For example, patients with more serious disease are more likely to be recommended by their physicians to receive routine SI. Moreover, as SI is a costly method, financial status of patients is another key factor to be considered, especially for that this factor will also influence the quality of treatment and follow-up care that patients received. Secondly, data quality may be another problem in this study. Wrong data like irrational date was found when I calculated survival times and got negative ones. Also, data regarding whether the relapse was detected clinically or radiographically were easy to

be misclassified. Finally, since PH assumption was violated for covariate ECOG in MER cohort, stratified Cox model is expected to be considered in future analysis.

In conclusion, our study suggested that surveillance imaging may not be necessary for patients with FL and achieving remission after first-line treatment as it did not bring notable benefit in survival while highly increase health care costs. Future studies are recommended to be conducted prospectively with more rigorous study design like ensuring all the patients included have access to SI to balance patient characteristics in clinical detection group and SI detection group to a large extent thus providing more robust results in FL and in other indolent lymphomas. In addition, it is meaningful to investigate whether there is difference in tumor aggressiveness between relapses detected by clinical approaches and radiographic methods and give further information on the necessity of SI.

Tables and Figures

Table 1. Descriptive statistics of baseline covariates stratified by detection method in two cohorts

			LEAD Cohort			MER Cohort		
			Clinical vs. SI			Clinical vs. SI		
Covariate	**Statistics**	**Level**	**clinical N=35**	**radio N=18**	**P-value***	**clinical N=63**	**radio N=50**	**P-value***
Gender	N (Col %)	female	21 (60.00)	8 (44.44)	0.281	20 (31.75)	24 (48.00)	0.078
	N (Col %)	male	14 (40.00)	10 (55.56)		43 (68.25)	26 (52.00)	
Ethnicity	N (Col %)	white	24 (68.57)	14 (77.78)	0.539	61 (98.39)	47 (94.00)	0.323
	N (Col %)	other	11 (31.43)	4 (22.22)		1 (1.61)	3 (6.00)	
Stage at diagnosis	N (Col %)	1,2	7 (21.88)	2 (11.11)	0.459	8 (12.70)	9 (18.00)	0.434
	N (Col %)	3,4	25 (78.13)	16 (88.89)		55 (87.30)	41 (82.00)	
Grade at diagnosis	N (Col %)	1,2	29 (82.86)	14 (82.35)	1.000	54 (85.71)	43 (86.00)	0.965
	N (Col %)	3	6 (17.14)	3 (17.65)		9 (14.29)	7 (14.00)	
ECOG	N (Col %)	0	8 (25.00)	7 (43.75)	0.186	36 (58.06)	36 (72.00)	0.126
	N (Col %)	≥1	24 (75.00)	9 (56.25)		26 (41.94)	14 (28.00)	
FLIPI	N (Col %)	low	6 (25.00)	1 (6.67)	0.308	15 (23.81)	12 (24.00)	0.767
	N (Col %)	intermediate	6 (25.00)	6 (40.00)		24 (38.10)	16 (32.00)	
	N (Col %)	high	12 (50.00)	8 (53.33)		24 (38.10)	22 (44.00)	
B-symptoms	N (Col %)	yes	13 (40.63)	1 (5.88)	0.018	10 (16.39)	7 (14.00)	0.728
	N (Col %)	no	19 (59.38)	16 (94.12)		51 (83.61)	43 (86.00)	
Age at diagnosis	Mean (95% CI)		54.20 (27.47, 80.93)	57.48 (35.35, 79.61)	0.453	58.63 (35.17, 82.09)	59.69 (36.39, 82.99)	0.619

* The p-values were calculated by ANOVA or Kruskal-Wallis test for numerical covariates and chi-square test or Fisher's exact test for categorical covariates.

Table 2. K-M survival probability estimates for three outcomes in LEAD cohort

	Group	N	Event	Censored	Median Survival Time (95% CI)	1-Year Survival	3-Year Survival	5-Year Survival	P-value*
PFS	clinical	35	35 (100%)	0 (0%)	4.2 (2.5, 6)	94.3% (79.0%, 98.5%)	57.1% (39.3%, 71.5%)	40.0% (24.0%, 55.5%)	0.2895
	SI	18	18 (100%)	0 (0%)	3.6 (1.8, 4.4)	94.4% (66.6%, 99.2%)	55.6% (30.5%, 74.8%)	22.2% (6.9%, 42.9%)	
OS from diagnosis	clinical	34	7 (21%)	27 (79%)	-	100.0% (NA, NA)	97.1% (80.9%, 99.6%)	97.1% (80.9%, 99.6%)	0.5324
	SI	18	2 (11%)	16 (89%)	-	100.0% (NA, NA)	93.3% (61.3%, 99.0%)	86.2% (55.0%, 96.4%)	
OS from relapse	clinical	34	7 (21%)	27 (79%)	-	93.4% (76.2%, 98.3%)	89.7% (71.3%, 96.6%)	80.7% (51.7%, 93.3%)	0.3346
	SI	17	2 (12%)	15 (88%)	-	92.9% (59.1%, 99.0%)	85.7% (53.9%, 96.2%)	85.7% (53.9%, 96.2%)	

* The p-values were calculated by log-rank test.

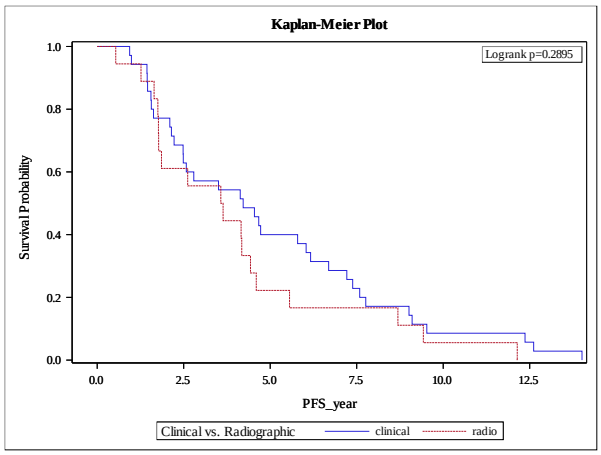
Table 3. K-M survival probability estimates for three outcomes in MER cohort

	Group	N	Event	Censored	Median Survival Time (95% CI)	1-Year Survival	3-Year Survival	5-Year Survival	P-value*
PFS	clinical	63	63 (100%)	0 (0%)	2 (1.6, 2.6)	88.9% (78.1%, 94.5%)	33.3% (22.1%, 45.0%)	9.5% (3.9%, 18.2%)	0.9260
	SI	50	50 (100%)	0 (0%)	2.4 (1.9, 2.8)	96.0% (84.9%, 99.0%)	32.0% (19.7%, 45.0%)	18.0% (8.9%, 29.7%)	
OS from diagnosis	clinical	63	14 (22%)	49 (78%)	-	100.0% (NA, NA)	95.2% (85.7%, 98.4%)	90.0% (79.0%, 95.4%)	0.9624
	SI	50	12 (24%)	38 (76%)	-	100.0% (NA, NA)	98.0% (86.6%, 99.7%)	96.0% (84.8%, 99.0%)	
OS from relapse	clinical	63	14 (22%)	49 (78%)	-	96.7% (87.6%, 99.2%)	88.4% (77.2%, 94.3%)	84.1% (71.5%, 91.5%)	0.9603
	SI	50	12 (24%)	38 (76%)	-	98.0% (86.6%, 99.7%)	93.7% (81.8%, 97.9%)	86.4% (72.1%, 93.7%)	

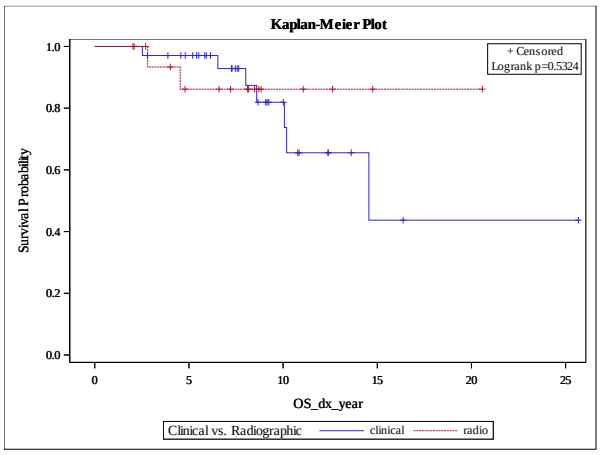
* The p-values were calculated by log-rank test.

Figure 1. K-M survival curves stratified by method of relapse detection in LEAD cohort

A. PFS



B. OS from diagnosis



C. OS from relapse

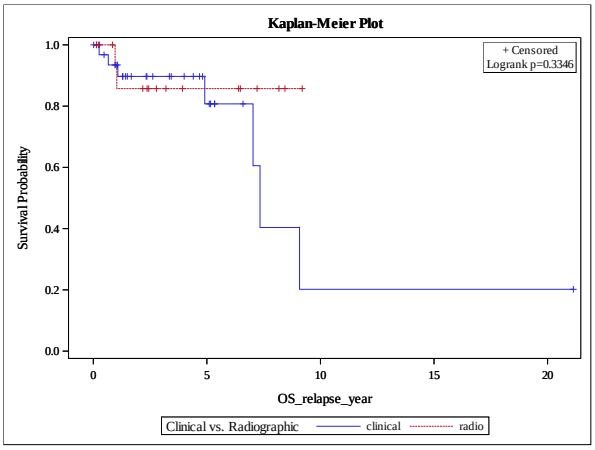
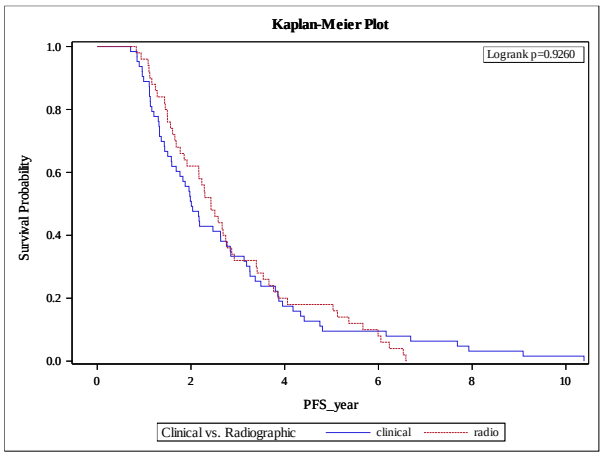
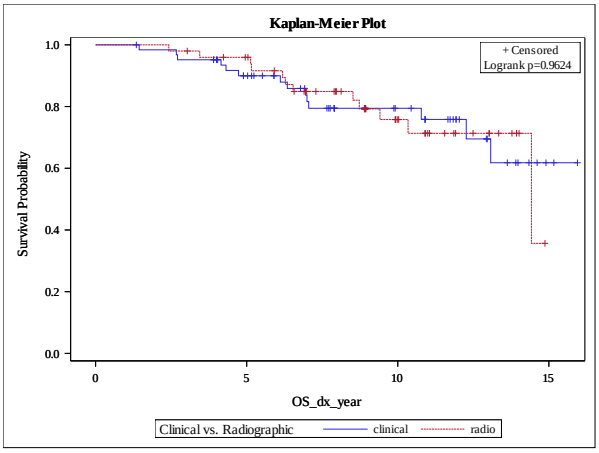


Figure 2. K-M survival curves stratified by method of relapse detection in MER cohort

A. PFS



B. OS from diagnosis



C. OS from relapse

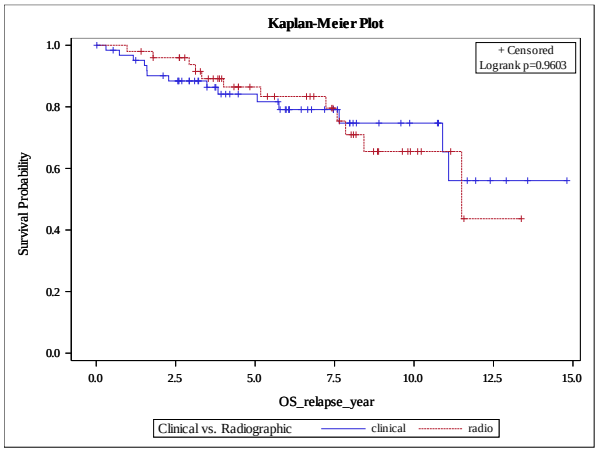


Table S1. Univariate and Multivariate associations between each variable and three outcomes in LEAD cohort

		PFS				OS from diagnosis				OS from relapse			
		Univariate		Multivariate		Univariate		Multivariate		Univariate		Multivariate	
		HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*
Type	SI	1.37 (0.76-2.45)	0.291	1.71 (0.77-3.80)	0.187	0.61 (0.13-2.94)	0.537	0.14 (0.01-3.36)	0.222	0.47 (0.10-2.28)	0.346	0.28 (0.03-2.51)	0.255
	Clinical	Ref		Ref		Ref		Ref		Ref		Ref	
Gender	Male	0.95 (0.54-1.66)	0.854	-	-	0.85 (0.23-3.19)	0.811	0.00 (0.00, 0.93)	**0.046**	1.06 (0.28-4.04)	0.930	-	-
	Female	Ref				Ref		Ref		Ref			
Ethnicity	White	0.88 (0.48-1.60)	0.665	0.47 (0.18, 1.21)	0.116	0.35 (0.09-1.43)	0.145	-	-	0.39 (0.10-1.46)	0.161	-	-
	Others	Ref		Ref		Ref				Ref			
Stage	1, 2	0.68 (0.33-1.42)	0.306	-	-	0.84 (0.17-4.20)	0.834	0.00 (0.00, 2.15)	0.083	1.02 (0.18-5.77)	0.983	-	-
	3, 4	Ref				Ref		Ref		Ref			
Grade	1,2	1.03 (0.50-2.15)	0.933	-	-	0.55 (0.11-2.78)	0.470	0.00 (0.00, 1.80)	0.075	0.29 (0.05-1.75)	0.178	0.21 (0.03, 1.55)	0.127
	3	Ref				Ref		Ref		Ref		Ref	
ECOG	0	1.01 (0.54-1.87)	0.985	0.44 (0.18, 1.09)	0.078	0.78 (0.20-3.00)	0.722	0.00 (0.00, 1.00)	0.050	0.93 (0.23-3.78)	0.921	0.24 (0.04, 1.37)	0.107
	>=1	Ref		Ref		Ref		Ref		Ref		Ref	
FLIPI	Low	0.60 (0.25-1.43)	0.248	4.00 (1.05, 15.20)	**0.042**	0.75 (0.14-4.19)	0.745	-	-	1.39 (0.24-8.12)	0.713	-	-
	Inter	0.65 (0.31-1.36)	0.251	5.24 (0.88, 31.41)	0.070	0.00 (0.00-.)	0.995	-	-	0.00 (0.00-.)	0.995	-	-
	High	Ref		Ref		Ref				Ref			
B-Symp.	Yes	1.48 (0.78-2.84)	0.233	9.94 (2.32, 42.57)	**0.002**	1.36 (0.33-5.60)	0.673	0.00 (0.00, 0.89)	**0.001**	0.99 (0.24-4.05)	0.993	-	-
	No	Ref		Ref		Ref		Ref		Ref			
Age		1.02 (1.00-1.04)	0.070	1.06 (1.02-1.10)	**0.008**	1.06 (1.00-1.12)	**0.041**	1.26 (1.01-1.59)	**0.043**	1.04 (0.98-1.09)	0.184	-	-

* The p-values that are smaller than 0.05 are marked in bold.

Table S2. Univariate and Multivariate associations between each variable and three outcomes in MER cohort

		PFS				OS from diagnosis				OS from relapse			
		Univariate		Multivariate		Univariate		Multivariate		Univariate		Multivariate	
		HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*
Type	SI	0.98 (0.67-1.44)	0.926	1.05 (0.71-1.55)	0.806	1.02 (0.47-2.21)	0.962	1.17 (0.53-2.59)	0.691	1.02 (0.47-2.21)	0.960	1.10 (0.50-2.47)	0.799
	Clinical	Ref		Ref		Ref		Ref		Ref		Ref	
Gender	Male	0.99 (0.67-1.44)	0.937	-	-	1.05 (0.47-2.31)	0.913	-	-	1.05 (0.47-2.32)	0.906	-	-
	Female	Ref				Ref				Ref			
Ethnicity	White	1.08 (0.40-2.95)	0.874	-	-	1.15 (0.16-8.56)	0.890	-	-	1.19 (0.16-8.89)	0.864	-	-
	Others	Ref				Ref				Ref			
Stage	1, 2	0.65 (0.38-1.12)	0.125	0.63 (0.36-1.08)	0.093	0.42 (0.10-1.80)	0.244	0.24 (0.05, 1.10)	0.066	0.46 (0.11-1.95)	0.293	-	-
	3, 4	Ref		Ref		Ref		Ref		Ref			
Grade	1,2	1.25 (0.74-2.13)	0.404	1.45 (0.83, 2.54)	0.193	1.68 (0.50-5.62)	0.401	-	-	1.61 (0.48-5.40)	0.441	-	-
	3	Ref		Ref		Ref				Ref			
ECOG	0	1.06 (0.71-1.57)	0.788	-	-	0.39 (0.18-0.86)	**0.020**	0.44 (0.20-0.99)	**0.048**	0.39 (0.18-0.86)	**0.019**	0.49 (0.22-1.11)	0.086
	>=1	Ref				Ref		Ref		Ref		Ref	
FLIPI	Low	0.72 (0.45-1.17)	0.182	-	-	0.19 (0.04-0.83)	**0.028**	-	-	0.20 (0.05-0.88)	**0.033**	4.43 (0.98-19.96)	0.053
	Inter	0.86 (0.56-1.32)	0.483	-	-	0.65 (0.28-1.50)	0.314	-	-	0.69 (0.30-1.56)	0.368	4.94 (1.02-23.83)	**0.045**
	High	Ref				Ref				Ref		Ref	
B-Symp.	Yes	1.31 (0.78-2.21)	0.303	-	-	2.40 (1.00-5.76)	0.050	-	-	2.34 (0.98-5.60)	0.056	-	-
	No	Ref											
Age		1.00 (0.98-1.01)	0.819	-	-	1.04 (1.00-1.07)	**0.028**	1.04 (1.01-1.08)	**0.026**	1.03 (1.00-1.07)	**0.035**	1.03 (1.00-1.07)	0.075
* The p-values that are smaller than 0.05 are marked in bold.													

Table S3. Supremum test results in LEAD cohort

		PFS		OS from diagnosis		OS from relapse	
Variable	**Level**	**Maximum Absolute Value (MAV)**	**Pr>MAV**	**Maximum Absolute Value (MAV)**	**Pr>MAV**	**Maximum Absolute Value (MAV)**	**Pr>MAV**
Type	SI	1.0120	0.1690	1.1428	0.1440	0.6076	0.3040
	Clinical	Ref		Ref		Ref	
Gender	Male	-	-	1.5703	0.3120	-	-
	Female			Ref			
Ethnicity	White	0.5072	0.9310	-	-	-	-
	Others	Ref					
Stage	1, 2	-	-	1.7883	0.1360	-	-
	3, 4			Ref			
Grade	1,2	-	-	1.6322	0.1820	0.9381	0.0770
	3			Ref		Ref	
ECOG	0	0.9035	0.3950	1.7483	0.2590	0.6880	0.3470
	>=1	Ref		Ref		Ref	
FLIPI	Low	1.9985	0.1190	-	-	-	-
	Inter	1.0267	0.9130	-	-	-	-
	High	Ref					
B-Symp.	Yes	1.1288	0.4330	2.1622	0.1370	-	-
	No	Ref		Ref			
Age		0.7104	0.8260	0.3000	0.6070	-	-

Table S4. Supremum test results in MER cohort

		PFS		OS from diagnosis		OS from relapse	
Variable	**Level**	**Maximum Absolute Value (MAV)**	**Pr>MAV**	**Maximum Absolute Value (MAV)**	**Pr>MAV**	**Maximum Absolute Value (MAV)**	**Pr>MAV**
Type	SI	1.1308	0.1310	0.5174	0.7790	0.7430	0.4090
	Clinical	Ref		Ref		Ref	
Gender	Male	-	-	-	-	-	-
	Female						
Ethnicity	White	-	-	-	-	-	-
	Others						
Stage	1, 2	1.2197	0.0910	0.6913	0.3830	-	-
	3, 4	Ref		Ref			
Grade	1,2	1.1072	0.1320	-	-	-	-
	3	Ref					
ECOG	0	-	-	1.1442	0.0960	1.3696	**0.0390**
	>=1			Ref		Ref	
FLIPI	Low	-	-	-	-	1.0828	0.6990
	Inter	-	-	-	-	1.3570	0.5090
	High					Ref	
B-Symp.	Yes	-	-	-	-	-	-
	No						
Age		-	-	0.5953	0.6820	0.7351	0.4420

Reference

1. Lackraj T, Goswami R, Kridel R. Pathogenesis of follicular lymphoma. *Best Pract Res Clin Haematol.* Mar 2018;31(1):2-14.

2. Freedman A, Jacobsen E. Follicular lymphoma: 2020 update on diagnosis and management. *Am J Hematol.* Dec 8 2019.

3. Becnel MR, Nastoupil LJ. Follicular Lymphoma: Past, Present, and Future. *Curr Treat Options Oncol.* May 24 2018;19(7):32.

4. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* May 19 2016;127(20):2375-2390.

5. Horning SJ, Rosenberg SA. The natural history of initially untreated low-grade non-Hodgkin's lymphomas. *N Engl J Med.* Dec 6 1984;311(23):1471-1475.

6. Kahl BS, Yang DT. Follicular lymphoma: evolving therapeutic strategies. *Blood.* Apr 28 2016;127(17):2055-2063.

7. Hiddemann W, Cheson BD. How we manage follicular lymphoma. *Leukemia.* Jul 2014;28(7):1388-1395.

8. Nastoupil LJ, Sinha R, Byrtek M, et al. Comparison of the effectiveness of frontline chemoimmunotherapy regimens for follicular lymphoma used in the United States. *Leuk Lymphoma.* May 2015;56(5):1295-1302.

9. Fischer T, Zing NPC, Chiattone CS, Federico M, Luminari S. Transformed follicular lymphoma. *Ann Hematol.* Jan 2018;97(1):17-29.

10. Cohen JB, Flowers CR. Optimal disease surveillance strategies in non-Hodgkin lymphoma. *Hematology Am Soc Hematol Educ Program.* Dec 5 2014;2014(1):481-487.

11. Nakamura K, Sasaki M, Kunitake N, et al. Relapse patterns of localized non-Hodgkin's lymphoma of the head and neck after clinical remission: results of a strict follow-up procedure. *Int J Clin Oncol.* Dec 2001;6(6):302-305.

12. Elis A, Blickstein D, Klein O, Eliav-Ronen R, Manor Y, Lishner M. Detection of relapse in non-Hodgkin's lymphoma: role of routine follow-up studies. *Am J Hematol.* Jan 2002;69(1):41-44.

13. Truong Q, Shah N, Knestrick M, et al. Limited utility of surveillance imaging for detecting disease relapse in patients with non-Hodgkin lymphoma in first complete remission. *Clin Lymphoma Myeloma Leuk.* Feb 2014;14(1):50-55.

14. Phillips T, Mercer J. Surveillance Scans in Lymphoma: Friend or Foe? *Curr Treat Options Oncol.* Feb 2017;18(2):10.

15. Cohen JB, Kurtz DM, Staton AD, Flowers CR. Next-generation surveillance strategies for patients with lymphoma. *Future Oncol.* 2015;11(13):1977-1991.

16. Liedtke M, Hamlin PA, Moskowitz CH, Zelenetz AD. Surveillance imaging during remission identifies a group of patients with more favorable aggressive NHL at time of relapse: a retrospective analysis of a uniformly-treated patient population. *Ann Oncol.* Jun 2006;17(6):909-913.