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:

_____ _____

Chaejin Kim Date

Partitioning Around Medoids Clustering in Follicular Lymphoma Patients: Comparison with FLIPI and FLIPI2 in PFS

By

Chaejin Kim

Master of Public Health

Biostatistics and Bioinformatics

_____

Zhengjia (Nelson) Chen, PhD

(Thesis Advisor)

_____

Michael Kutner, PhD

(Reader)

Partitioning Around Medoids Clustering in Follicular Lymphoma Patients: Comparison with FLIPI and FLIPI2 in PFS

By

Chaejin Kim

Pharm.D.

Ewha Womans University

2015

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD

Reader: Michael Kutner, 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

2018

Abstract

Partitioning Around Medoids Clustering in Follicular Lymphoma Patients: Comparison with FLIPI and FLIPI2 in PFS

By Chaejin Kim

Background: The Follicular Lymphoma International Prognostic Indices (FLIPI and FLIPI2) are widely used prognostic indices in follicular lymphoma (FL), with some limitations because they dichotomize continuous variables and classify patients based on the number of adverse factors. In this thesis, Gower's distance and Partitioning Around Medoids (PAM) clustering were employed to overcome these issues, and progress free survival (PFS) analysis was performed based on PAM clustering results to compare with PFS using the conventional methods.

Methods: The demographic and baseline characteristics are summarized descriptively using frequencies (%) and means (SD). Gower's distance was calculated to measure dissimilarity between observations, considering that the prognostic factors of FLIPI and FLIPI2 are mixed type data. Using the distance matrix, the silhouette width was calculated to find the optimal number of clusters. Based on the results, PAM clustering was conducted. The clustering results were compared to the classification of FLIPI or FLIPI2 in terms of PFS.

Results: When using FLIPI prognostic factors, PAM in 3 clusters showed the smaller P-value (P- value=0.094) from the log-rank test than that in 3 risk groups of FLIPI (P-value=0.27). When using FLIPI2 prognostic factors, PAM in 3 cluster also showed the smaller P-value (P=0.03) from the log-rank test than that in 3 risk groups of FLIPI2 (P-value=0.50). The Kaplan-Meier (KM) curves and comparison tables between PAM and FLIPI or FLIPI2 indicated that, although PAM reflected the scale of FLIPI or FLIPI2 in some sense, it showed somewhat counterintuitive survival results considering the composition of patients in FLIPI or FLIPI2 scales. PFS stratified by PAM showed better differentiation in survival.

Conclusion: Classification based on FLIPI or FLIPI2 versus PAM clustering provided us with different results. For both FLIPI and FLIPI2, PAM clustering showed better classification in terms of PFS. This may indicate that the issues observed in process of establishing FLIPI and FLIPI2 may indeed contribute to the loss of power to classify patients with FL. A large scale study may be warranted to validate these results.

Partitioning Around Medoids Clustering in Follicular Lymphoma Patients: Comparison with FLIPI and FLIPI2 in PFS

By

Chaejin Kim

Pharm.D.

Ewha Womans University

2015

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD

Reader: Michael Kutner, 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
2018

Acknowledgements

I would like to show my best appreciation to my thesis advisor Dr. Nelson Chen. Not only for thesis advising but also he gave me various opportunity of practice as a statistician. Also, I would like to thank my thesis reader Dr. Michael Kutner. He gave an advice and comments for my thesis. I deeply appreciate Dr. Chen and Dr. Kutner's help and advice. And I appreciate to Dr. Goldman Max who kindly allowed me to use the dataset. Lastly but not least, I would like to show gratitude to my parents and my two brothers for supporting and encouraging me throughout my years of study. Thank you.

Table of Contents

1. Introduction

In medical practice, prognostic models or indices are widely used to determine the direction of therapeutic treatments and predict prognoses. This allows us to predict future outcomes for a given disease utilizing baseline characteristics [1, 2].Here, there have been several prognostic indies for follicular lymphoma (FL), the most common subtype of indolent non-Hodgkin's lymphoma (NHL) developed [3]. The Follicular Lymphoma International Prognostic Index (FLIPI) is an index which has been widely used to predict the prognosis of FL patients [4]. Five prognostic risk factor parameters for overall survival (OS) were selected in the FLIPI to classify patients with FL: age, Ann Arbor stage[5, 6], hemoglobin (Hb) level, serum LDH level and the number of nodal sites. Adverse risk factor levels for the five parameters were defined as age $\geq$ 60 years old, Ann Arbor stage III-IV, Hb level < 120 *g/L*, LDH level > upper limit of normal (ULN) and number of nodal sites > 4. The FLIPI index score for the patients is calculated as the number of adverse risk factor levels ranging from 0 to 5. Finally, the FLIPI index classifies patients with FL into three risk groups for OS: low (0, 1 risk factor), intermediate (2 risk factors) and high ($\geq$ 3 risk factors)[4]. In 2009, a revised improved FLIPI index, namely, FLIPI2 overcame the weakness of the FLIPI index. The five prognostic risk factor parameters and their adverse risk factor levels were defined as follows: β2-microglobulin (B2M; >ULN), longest diameter of the largest involved node (LoDLIN; > 6 cm), BMI (presence), Hb level (<120 *g/L*) and age (> 60 years old). For the FLIPI2 index, patients were stratified into three risk groups for progress free survival (PFS): low (0 risk factors), intermediate (1 or 2 risk factors), and high (3 to 5 risk factors)[7]. Both FLIPI and FLIPI2 are widely used in clinical practice and give healthcare providers a convenient tool to help determine therapeutic medical decisions[3].

With the popularity of the prognosis research, however, some issues were raised regarding current prognostic models or indices development. According to a Prognosis Research Strategy (PROGRESS) study conducted by the PROGRESS group [1], main concerns include modelling continuous prognostic factors[8], the complexity of the model[9], dealing with missing data[10] and checking the model assumptions.

In FLIPI and FLIPI2, some of these same issues have also been observed, which may lead to bias of the prognosis prediction. First of all, continuous variables were dichotomized by using usual clinical thresholds[4, 7]. It is known that dichotomization of continuous variable results in loss of information and power[8]. Also, there is no evidence that clinical thresholds are also optimal cut-off point for prognostic models. Furthermore, FLIPI and FLIPI2 classify FL patients into risk groups based on the number of adverse factors that patients have. This approach is very simple and intuitive, so can be widely accepted in clinical practice. However, only considering the number of adverse factors for dividing risk groups cannot take into account the absolute and relative correlations between prognostic factors and outcomes as well as within relationships of prognostic factors.

In this thesis, clustering was performed to classify FL patients. Cluster analysis is the most common unsupervised learning method. It aims to group the objects based on the definition of similarity (or dissimilarity). Given that FLIPI and FLIPI2 selected good prognostic factor levels, clustering was performed using all the information of the factors rather than dichotomized factors. To measure the dissimilarity of the mixed type data, not only contain continuous variables but also binary, ordinal or categorical variables, Gower's distances were calculated [11, 12]. Then, partitioning around medoids (PAM) was employed as a clustering algorithm. The results of

clustering were compared with the risk groups by using FLIPI and FLIPI2. PFS based on clustering results by PAM was also conducted. All the analyses were conducted by R[13] and R packages, ‘survminer[14]’ and ‘cluster[15]’.

2. Methods

2.1 Patient Sample

The data for these analyses were collected from a retrospective chart review on all patients with previously untreated FL and complete records at Emory University in Atlanta, GA diagnosed between July 1991 and July 2016. Patients were identified using the Institutional Review Board (IRB)-approved institutional Lymphoid Malignancies Enterprise Architecture Database (LEAD), which includes patient data for each previously consented lymphoma patient evaluated or treated at Emory. Inclusion criteria of patients were those above the age of 18 who had achieved a documented partial response, complete response, or stable disease following first-line induction therapy as assessed and documented by clinicians and radiologists after the initial period of treatment and who did not proceed immediately to a second-line regimen[16]. Patients in grade 3b of FL were excluded as well as patients after their first treatment if they had incomplete or missing outcome data, or if they experienced a failure in treatment defined as a time to progression of less than 26 weeks, or if the patient pursued watchful waiting without any therapy during the observation period.

The following medical records were collected (not necessary for all patients) from the patients: demographic data (age at diagnosis, gender, and ethnicity), disease characteristics at the time of diagnosis (including Ann Arbor stage, grade, presence of B-symptoms, extranodal involvement,

Eastern Cooperative Oncology Group [ECOG] performance score, Groupe d'Etude des Lymphomes Folliculaires [GELF] score, FLIPI, and FLIPI2 scores), and care information (date of diagnosis, date of progress and date of last contact)[4, 7, 16]. PFS time was calculated from date of diagnosis to date of progress or date of last contact, as appropriate.

2.2 Statistical Analyses

2.2.1 Descriptive Summarization

The descriptive summary for patients' demographics and baseline characteristics were calculated. Medians and interquartile ranges were summarized for continuous variables, while the frequencies and percentages were calculated for binary or categorical variables. Patients having at least one medical record were included in the descriptive statistical summarization.

2.2.2 Gower's Distance (Gower's Dissimilarity Coefficient)

Gower's dissimilarity coefficient was first proposed by J.C Gower in 1971 [11], and extended by Kaufman and Rousseuw in 1990 [12]. In this analysis, Kaufman and Rousseuw's version of Gower's dissimilarity coefficient was used to deal with mixed-type data.

The dissimilarity between i^{th} and j^{th} observations in n dimensional data is defined as

$$S_{ij} = \frac{\sum_k^n w_{ijk} S_{ijk}}{\sum_k^n w_{ijk}}$$

where S_{ijk} indicates the dissimilarity for the k^{th} variable between the two observations, and w_{ijk} is weight for the k^{th} variable between the two observations. In this analysis, w_{ijk} is equal to 0 when S_{ijk} is missing, else w_{ijk} is 0 (for binary variables, see below).

S_{ijk} is defined as follows depending on data type of variables, where x_{ik} is the i^{th} observation of k^{th} variable:

- For ordinal and continuous variables

$$S_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{r_k}$$

  where r_k is the range of values for the k^{th} variable, and.

- For nominal variables, S_{ijk} is equal to 1 if $x_{ik} = x_{jk}$, or $S_{ijk} = 0$ if $x_{ik} \neq x_{jk}$.
- For binary variables, S_{ijk} and w_{ijk} are defined as,

Value of obs. for k^{th} variable: 1 for present, 0 for absent of the feature				
i^{th} obs.	1	1	0	0
j^{th} obs.	1	0	1	0
Value of S_{ijk} and w_{ijk} for the corresponding obs.				
S_{ijk}	1	0	0	0
w_{ijk}	1	1	1	0

2.2.3 Partitioning Around Medoids (PAM)

1. Partitioning Around Medoids is a clustering method that classifies observations into *k* medoids, where *k* is chosen in advance [12]. PAM is very similar to *k*-means clustering in that observations are partitioned into *k* clusters. However, the two methods differ in that the PAM set medoids, represent observations chosen from a cluster as center of each

cluster, while *k*-means clustering set the center of clusters by calculating the mean of the observation in each cluster [12, 17].

Let *S* be a set of selected observations, i.e, medoids, *O* is the set of the selected observations. Then the set of unselected objects, *U*, is defined as $U = O - S$. The goal of PAM is minimization of the average dissimilarity of objects (belongs to *U*) to their closest selected object (belongs to *S*). In other words, the goal is to minimize the sum of the dissimilarities between objects in the set *U* and their closest medoid in the set *S*. PAM consists of the following two phases [18]:

(i) In the first phase, BUILD, *k* objects are selected for an initial set *S*.

(ii) In the second phase, SWAP, the quality of the clustering is improved by exchanging selected objects with unselected objects.

For each object *p* we maintain two numbers [18]:

• D_p, the dissimilarity between *p* and the closest medoid (object in *S*) and

• E_p, the dissimilarity between *p* and the second closest medoid (object in *S*)

D_p, E_p are updated every iterate when the sets S and U change.

Note that $D_p = 0$ only when $p \in S$ and $D_j \leq E_j$.

The BUILD phase has the following steps [18]:

1. Initialize the set *S* by adding an object to *S* so that the sum of the distances to all other objects is minimal.

2. Consider an object $i \in U$ to include into the set *S*.

3. For an object j that $j \in U - \{i\}$, calculate D_j .

4. Compute $g_i = \sum_{j \in U-\{i\}} C_{ji}$, where $C_{ji} = \max\{D_j - d(j,i), 0\}$

5. Repeat 2-4 steps until we find the object *i* such that maximizes g_i.

6. Update *S* and *U* as $S = S \cup \{i\}$ and $U = U - \{i\}$

The steps above are now implemented until *k* objects have been selected.

The second phase, SWAP, tries to improve the quality of the clustering by improving the set of selected objects [18].

Let T_{ih}=$\sum\{K_{jih} \mid j \in U\}$ where K_{jih} is defined as follow.

For an object $j \in U - \{i\}$ where *i* is moved from *S* to *U* and h is transferred from *U* to *S*,

1. If $d(j,i) > D_j$, then $K_{jih} = \min\{d(j,h) - D_j, 0\}$
2. If $d(j,i) = D_j$, then $K_{jih} = \min\{d(j,h),\ E_j\} - D_j$

Then select a pair $(i,h) \in S \times U$ which minimizes T_{ih}. If $T_{ih} < 0$, then the transfer between *i* and *h* is implemented, while D_p and E_p are also updated for every object *p*. And go back to the first step of SWAP to test other objects in *U* [18].

Finally, when min(T_{ih})>0, the algorithm stops [18].

2.2.4 Kaplan-Meier Estimator

The Kaplan-Meier(KM) estimator is widely used estimation method to estimate a survival function[19].

For the survivor function $S(t) = \Pr(T > t)$, the KM estimator is defined as in [19]:

$$\hat{S}(t) = \prod_{i:t_i \le t}(1 - \frac{d_i}{n_i})$$

where $d_i$is the number of events and n_i is the total number of individuals at risk at time *i*.

2.2.5 Log-Rank Test

The Log-Rank Test is the commonly used to compare the survival between groups. It tests the following set of hypotheses where $h_j(t)$ is hazard rate of group *j* at time *t* and τ is the largest time at which all the groups have at least one subject at risk[20].

$H_0: h_1(t) = h_2(t) = \cdots = h_K(t)$, for all $t \le \tau$

H_1: at least one of the $h_j(t)$'s is different for some $t \le \tau$

For the distinct times of observe failures as $t_1 < t_2 < \cdots < t_D$, the following are defined[20]:

d_{ij} is the number of events in group *j* at time t_i

Y_{ij} is the number of individuals at risk in group j at time t_i where $i = 1,2,\cdots D$ and $j = 1,2,\cdots K$

Then the null hypothesis test is based on the statistics $Z_j(\tau)$ [20]:

$$Z_j(\tau) = \sum_{i=1}^{D} W_j(t_i)\left\{\frac{d_{ij}}{Y_{ij}} - \frac{d_i}{Y_i}\right\},$$

where $W_j(t_i)$ is a positive weight function with the property that $W_j(t_i)$ is zero whenever Y_{ij} is zero.

In general, we usually choose $W_j(t_i) = Y_{ij}W(t_i)$ where $W(t_i)$ is a common weight shared by each group. Then $Z_j(\tau)$ becomes

$$Z_j(\tau) = \sum_{i=1}^{D} W(t_i)\left\{d_{ij} - Y_{ij}(\frac{d_i}{Y_i})\right\}, j = 1,2,\cdots K$$

where the variance of $Z_j(\tau)$ is given by

$$\widehat{\sigma_{JJ}} = \sum_{i=1}^{D} W(t_i)^2 \frac{Y_{ij}}{Y_i}\left(1 - \frac{Y_{ij}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i, j = 1,2, \cdots K$$

and the covariance of $Z_j(\tau)$, $Z_g(\tau)$ is given by

$$\widehat{\sigma_{Jg}} = -\sum_{i=1}^{D} W(t_i)^2 \frac{Y_{ij}}{Y_i}\frac{Y_{ig}}{Y_i}\left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i, g \neq j$$

For any $K - 1$ of Z_j's and estimated variance-covariance matrix Σ,

the test statistics is expressed by the quadratic form [20]

$$\chi^2 = (Z_1(\tau), \cdots, Z_{K-1}(\tau))\, \Sigma^{-1} (Z_1(\tau), \cdots, Z_{K-1}(\tau))^t$$

3. Results

3.1 Descriptive Summarization

The summarization results of the demographic and baseline characteristics are shown in Table 1. There were 148 patients in the data set. From Table 1, the median age and IQR of the patients was 57.6 and [47.5, 66.0], respectively. The proportion of female patients was not statistically larger (55.4%) than that of males (44.6%), P = 0.19. Most of patients were classified into 1, 2 or 3 of FLIPI score (81.9%), while according to FLIPI2, 83.8% of patients were in risk group of 0, 1 or 2. Except gender and bone marrow involvement, distributions of all the other categorical variables were not statistically equal at $P < 0.05$.

3.2 Classification using Prognostic Factors in FLIPI

By using age, Ann Arbor stage, hemoglobin (Hb) level, serum LDH level and the number of nodal sites, the PAM dissimilarity matrix was calculated. Figure 1 shows the average silhouette width based on the PAM dissimilarity matrix. Considering the result of the silhouette width and for comparing with FLIPI risk group, the number of clusters, 2, 3 and 4, were tested.

3.2.1 Number of Clusters: 2

Based on the PAM dissimilarity matrix, PAM clustering was performed. Table 2 represents the results of clustering compared to the original FLIPI risk group. PAM1, cluster 1 with PAM clustering was more likely to include patients with higher FLIPI risk score compared to PAM2. The log-rank test of FLIPI risk group yielded a P= 0.22 (Figure 2) whereas that of PAM with 2 clusters yielded a P= 0.33 (Figure 3).

NOTE: Make similar changes to sections below.

3.2.2 Number of Clusters: 3

Based on the PAM dissimilarity matrix, PAM clustering was performed. Table 3 represents the results of clustering compared to the original FLIPI risk group. Note that all individuals of PAM2 in PAM with 2 clusters (Table 2) were transferred to PAM3 in PAM with 3 clusters (Table 3). PAM1 in the 2 clusters (Table 2) was stratified into PAM1 and PAM2 in the 3 clusters (Table 3). The log-rank test of PAM with 3 clusters yielded a P=0.09 (Figure 4). The newly separated groups, PAM 1 and PAM 2 in the 3 clusters (Table 3) show quite different survival estimate. Figure 5 represents the KM curve with the log-rank P-value of 0.27 stratified by FLIPI 3 risk group (0,1-low, 2-intermediate, 3,4,5-high).

3.2.3 Number of Clusters: 4

Based on the PAM dissimilarity matrix, PAM clustering was performed. Table 4 represents the results of clustering compared to original FLIPI risk group. The log-rank test of PAM with 4 clusters yielded a P=0.05 (Figure 6), showing the lowest P-value among Figure 2-6.

3.3 Classification using Prognostic Factors in FLIPI2

By using β2-microglobulin, longest diameter of the largest involved node, BMI, Hb level and age, the PAM dissimilarity matrix was calculated. Figure 7 shows the average silhouette width based on the PAM dissimilarity matrix. Considering the result of the silhouette width and for comparing with FLIPI2 risk group, the number of clusters, 2, 3 and 4, were tested.

3.3.1 Number of Clusters: 2

Based on the PAM dissimilarity matrix, PAM clustering was performed. Table 5 represents the results of clustering compared to the original FLIPI2 risk group. PAM1 was more likely to include patients with higher FLIPI2 risk compared to PAM2. The log-rank test of FLIPI2 risk group yielded a P=0.67 (Figure 8) whereas that of PAM with 2 clusters yielded a P=0.21 (Figure 9).

3.3.2 Number of Clusters: 3

Based on the PAM dissimilarity matrix, PAM clustering was performed. Table 6 represents the results of clustering compared to the original FLIPI2 risk group. Note that all individuals of PAM2 in PAM with 2 clusters (Table 5) were transferred to PAM2 in PAM with 3 clusters (Table 6). PAM1 in the 2 clusters (Table 5) was stratified into PAM1 and PAM3 in PAM with 3 clusters (Table 6). The log-rank test of PAM with 3 clusters yielded P=0.03 (Figure 10), showing the lowest P-value among Figure 7-12. The newly separated groups, PAM 1 and PAM 3 in the 3 clusters (Table 6) show quite different survival estimate.. Figure 11 represents the KM curve with the log-rank P-value of 0.5 stratified by FLIPI2 3 risk group (0-low, 1,2-intermediate, 3,4,5-high).

3.3.3 Number of Clusters: 4

Based on the PAM dissimilarity matrix, PAM clustering was performed. Table 7 represents the results of clustering compared to the original FLIPI2 risk group. Note that all individuals of PAM1 and PAM3 in PAM with 3 clusters (Table 6) were transferred to PAM1 and PAM4, respectively, in PAM with 4 clusters (Table 7). PAM2 in the 3 clusters (Table 6) was stratified into PAM2 and PAM3 in the 4 clusters (Table 7). The log-rank test of PAM with 4 clusters yielded P=0.06 (Figure 12).

4. Discussion and **Conclusion**

For PAM clustering using FLIPI prognostic factors, PAM with 2 clusters showed the highest average silhouette width followed by PAM with 4 clusters and 5 clusters. Although average silhouette width of the 3 clusters is smaller than that of the 5 clusters, considering clinical meaning and to compare the results with FLIPI risk groups, 2, 3 and 4 clusters were examined. Overall, the results of PAM clustering were roughly consistent with FLIPI in that some clusters

only include those who are in FLIPI0, FLIPI1 or FLIPI2, which means those clusters did not include high-risk group of FLIPI. However, as the number of cluster increases, each cluster was, again, stratified into sub-clusters. One interesting thing is that this stratification showed a difference in the KM curve. PAM2 in the 2 cluster is directly transferred into PAM3 in the 3 clusters. When comparing Figure 3 and Figure 4, we can visually confirm that PAM1 and PAM2 in Figure 4, which were originally PAM1 in Figure 3 showed substantial difference in survival. This difference may contribute to the lower P-value of log-rank test. When compared to log-rank test for survival difference stratified by FLIPI 3 risk groups (Figure 5), that stratified by PAM with 3 clusters showed the lower P-value. Figure 6 showed the lowest P-value among Figure 2-6. PAM1 and PAM3 in the 4 clusters showed the biggest differences in survival. Considering composition of each cluster, PAM3 and PAM4 in the 4 clusters were supposed to show similar results, while PAM4 was expected to have slightly better survival. Although the two clusters showed similar pattern until t=2,800. However, PAM3 showed much better survival, after that time point, which was opposite of the expectation. This suggests that PAM clustering may capture more details, but clusters classified by PAM roughly reflect FLIPI scale. Also, it was remarkable that as the number of clusters increased, new medoids were added, while medoids observed in the smaller number of clusters were kept.

For the clustering using FLIPI2 prognostic factors, PAM with 2 clusters showed the highest average silhouette width followed by the 3 clusters and the 7 clusters. Although average silhouette width of the 4 clusters is smaller than that of the 7 clusters, considering clinical meaning while comparing the results with FLIPI risk groups, 2, 3 and 4 clusters were examined. Overall, it was observed that PAM clustering was roughly similar to FLIPI2 in that a certain cluster only included patients who were in FLIP2-0, FLIPI2-1 or FLIPI2-2, which means those clusters did not include high-risk group of FLIPI2. However, as number of cluster increases, it

was observed that each cluster was stratified into sub-clusters. As observed in the result of PAM using FLIPI scale, the sub-stratification showed indeed difference in KM curve. PAM2 in the 2 cluster of FLIPI2 (Table 5) is directly transferred into PAM2 in the 3 clusters of FLIPI2 (Table 6). When comparing the 2 clusters (Figure 9) and the 3 clusters (Figure 10), it was visually confirmed that PAM1 and PAM3 in the 3 clusters which were originally PAM1 in the 2 clusters showed a significant difference in survival (P=0.027). This difference may contribute to the lower P- value of log-rank test. This was remarkable because, considering the composition of PAM3 and PAM1 in 3 clusters, PAM3 was expected to have the worst survival, but results indicated the opposite. When compared to log-rank test for a survival difference stratified by FLIPI2 3 risk groups (Figure 11), that stratified by PAM with 3 clusters showed the much lower P-value. Figure 12 and Table 7 also supported the counterintuitive results inferred from FLIPI2 scale. It was obvious that PAM1 in the 4 cluster had the worst survival. However, when considering the composition of patients with respect to the FLIPI2 scale, PAM4 in the 4 clusters was expected to have the worst survival, or at least PAM1 and PAM4 were supposed to have the similar survival. Also, PAM2 in the 4 clusters was expected to have the best survival. Unfortunately, Figure 12 showed a somewhat different result from expectations. This suggests that PAM clustering and FLIPI2 showed quite different results both in terms of classification of FL patients and in terms of PFS. Also, similar with FLIPI scale, as the number of cluster increases, new medoids were added while medoids observed in the smaller number of clusters were kept.

This study has some limitations. First of all, the data set only included patients with FL who achieved first remission. Given that this analysis includes both patients who achieved first remission and those who didn’t achieve it, the results may be changed. Second, the sample size may not large enough. FLIPI[4] and FLIPI2[7] were established using information from about 1,000 patients or more. To develop solid results, further study may be needed using a larger data-

set. Also, only patients who have complete information on prognostic factors for FLIPI or FLIPI2, respectively, were included in this analysis to compare the result with FLIPI or FLIPI2, as appropriate. For this reason, only 127 patients and 80 patients were included when analyzing prognostic factors of FLIPI and FLIPI2, respectively. There may be pattern for the missing data. Lastly, this study was carried out by using retrospective chart review. It is believed that, to find prognostic factors, investigating prospective association between covariates and the outcome is more reasonable than that of retrospective relationship[1, 2]. Hence, when the next study is designed, a prospective study design should be considered to get more definitive results.

Overall, this study showed that classification based on FLIPI or FLIPI2 versus PAM clustering may provide us different results. And for both FLIPI and FLIPI2, PAM clustering showed better classification in terms of PFS. This may indicate that there are limitations in the way of classification in FLIPI and FLIPI2. FLIPI and FLIPI2 are very easy to understand and be used in clinical practice. However, FLIPI and FLIPI2 may sacrifice accuracy and precision because continuous variables were dichotomized and risk groups were classified based on simply how many adverse factors patients have. This may result in providing suboptimal treatment to FL patients. In advanced from the population approach, precise treatment for each patient has gained significant attention in healthcare field and people believe, it could be attained through remarkable development in technology. Ease of use in clinical practice is important. However, we should not trade off the precise treatment with convenient use in the field. And now technology can provide the healthcare provider and easy solution even when using complicated prognostic models. Although this study was more like a pilot study, it suggested that a conventional approach may not provide us the optimal results. Hence, a larger scale study may be needed to validate our results and provide a better therapeutic treatment.

5. References

1. Ewout W. Steyerberg, K.G.M.M., Danielle A. van der Windt, Jill A. Hayden, Pablo Perel, Sara Schroter, Richard D. Riley, Harry Hemingway, Douglas G. Altman, for the PROGRESS Group, *Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research.* PLoS Med, 2013. **10**(2): p. e1001381.
2. Harry Hemingway, P.C., Pablo, Jill A Hayden, Keith Abrams, Adam Timmis, Andrew Briggs Lindsay , Ruzan Udumyan research assistant 1, Karel G M Moons professor, et al., *Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes.* The BMJ, 2013. **346**: p. e5595.
3. Andrew D. Zelenetz, L.I.G., William G. Wierda, Jeremy S. Abramson, *NCCN Clinical Practice Guidelines in Oncology: B-cell Lymphomas*. 2017, National Comprehensive Cancer Network.
4. Philippe Solal-Cé ligny, P.R., Philippe Colombat, Josephine White, Jim O. Armitage, Reyes Arranz-Saez, Wing Y. Au, Monica Bellei, Pauline Brice, Dolores Caballero, Bertrand Coiffier, Eulogio Conde-Garcia, Chantal Doyen, Massimo Federico, Richard I. Fisher,, et al., *Follicular Lymphoma International Prognostic Index.* Blood, 2004. **First Edition**.
5. Paul P. Carbone (Chairman), H.S.K., Karl Musshoff, David W. Smithers, and Maurice Tubiana, *Report of the Committee on Hodgkin's Disease Staging Classification.* Cancer Research, 1971. **31**: p. 1860-1861.
6. Armitage, J.O., *Staging Non-Hodgkin Lymphoma.* A Cancer Journal for Clinicians, 2005. **55**: p. 368-376.
7. Massimo Federico, M.B., Luigi Marcheselli, Stefano Luminari, Armando Lopez-Guillermo, Umberto Vitolo, Barbara Pro, Stefano Pileri, Alessandro Pulsoni, Pierre Soubeyran, Sergio Cortelazzo, Giovanni Martinelli, Maurizio Martelli, Luigi Rigacci, Luca Arcaini, Francesco Di Raimondo, Francesco Merli, Elena Sabattini, Peter McLaughlin, and Philippe Solal-Ć eligny, *Follicular Lymphoma International Prognostic Index 2: A New Prognostic Index for Follicular Lymphoma Developed by the International Follicular Lymphoma Prognostic Factor Project.* Journal of Clinical Oncology, 2009. **27**.
8. Patrick Royston, D.G.A., Willi Sauerbrei, *Dichotomizing continuous predictors in multiple regression: a bad idea.* Statistics in Medicine, 2006. **25**: p. 127-141.
9. W, S., *The use of resampling methods to simplify regression models in medical statistics.* Appl Stat, 1999. **48**: p. 313-329.
10. Sterne JAC, W.I., Carlin JB, Spratt M, Royston P, *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* The BMJ, 2009. **338**.
11. Gower, J.C., *A General Coefficient of Similarity and Some of Its Properties.* Biometrics, 1971. **27**(4): p. 857-874.
12. Leonard Kaufman, P.J.R., *Finding Groups in Data- An Introduction to Cluster Analysis*. 1990: Wiley-Interscience.
13. R Core Team, *A language and environment for statistical computing*. 2017, R Foundation for Statistical Computing: Vienna, Austria.
14. Kosinski, A.K.a.M., *Drawing Survival Curves using 'ggplot2*. 2018.
15. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., *Cluster Analysis Basics and Extensions*. 2017.
16. Cheson BD, F.R., Barrington SF, et al., *Recommendations for initial evaluation, staging, and response assessment of hodgkin and non-hodgkin lymphoma: The lugano classification.* J Clin Oncol, 2014. **32**(27): p. 3059-3067.

17. A. P. REYNOLDS, G.R., B. DE LA IGLESIA and V. J. RAYWARD-SMITH, *Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms.* Journal of Mathematical Modelling and Algorithms, 2006. **5**: p. 475–504.
18. K. Ramya, R.S., *Integration of PAM Clustering Scheme with Dynamic Ensemble Membership Selection for Medical Data Analysis.* International Journal on Research in Exploring Engineering and Science, 2016. **I**(IV).
19. Kaplan, E.L., Meier, P, *Nonparametric estimation from incomplete observations.* J. Amer. Statist. Assn., 1958. **53**(282): p. 457-481.
20. John Klein, M.M., *Survival Analysis Techniques for Censored and Truncated Data*. 2nd edition ed. 2003, New York: Springer-Verlag.

6. Tables and Figures

Table 1. Demographic and Baseline Characteristics

Variable	N	Category	Summary Statistics	P-value
Gender	**148**			
		Male	66 (44.6%)	0.19
		Female	82 (55.4%)	
Ethnicity	**148**			
		White	104 (70.3%)	< 0.0001
		Black	12 (8.1%)	
		Others	32 (21.6%)	
Stage at Diagnosis	**144**			
		1, 2	36 (25.0%)	0.01
		3	42 (29.2%)	
		4	66 (45.8%)	
Grade at Diagnosis	**145**			
		1, 2	120 (82.8%)	< 0.0001
		3	25 (17.2%)	
ECOG	**140**			
		0	58 (41.4%)	0.043
		1, 2	82 (58.6%)	
Age at Diagnosis			57.6 [47.5, 66.0]	
Hgb			13.7 [12.3, 14.6]	
LDH			158 [143.5, 188]	
4 nodal sites	**141**			
		0	50 (35.5%)	0.0006
		1	91 (64.5%)	

Variable	N	Category	Summary Statistics	P-value
Node Diameter (cm)			3.7 [2.6, 5.9]	
B2 Micro			1.9 [1.5, 2.4]	
FLIPI	**127**			
		0	10 (7.9%)	< 0.0001
		1	21 (16.5%)	
		2	40 (31.5%)	
		3	43 (33.9%)	
		4	10 (7.9%)	
		5	3 (2.4%)	
FLIPI2	**80**			
		0	12 (15%)	< 0.0001
		1	39 (48.8%)	
		2	16 (20.0%)	
		3	10 (12.5%)	
		4	3 (3.8%)	
		5	0 (0.0%)	
B Symptoms	**135**			
		Yes	27 (20.0%)	< 0.0001
		No	108 (80.0%)	
Extranodal	**140**			
		Yes	93 (66.4%)	0.0001
		No	47 (33.6%)	
Bone Marrow Involvement	134			
		Yes	63 (47.0%)	0.49
		No	71 (53.0%)	
GELF*	**133**			
		Yes	79 (59.4%)	0.03
		No	54 (40.6%)	

- Data are presented as number of patients (%), or median (IQR, interquartile range). P-value is calculated by chi-square test or Fisher's exact test for categorical variables
* Groupe d'Etude des Lymphomes Folliculaires

Table 2. PAM with 2 Clusters vs FLIPI

	FLIPI0	FLIPI1	FLIPI2	FLIPI3	FLIPI4	FLIPI5	P- value
PAM1	0	4	30	43	10	3	<0.0001*
PAM2	10	17	10	0	0	0	
				Medoids			
	ID	Stage	Age	Hb	LDH	> 4 node site?	
PAM1	82	4	55.3	13.4	174	YES	
PAM2	43	1	62.7	13.4	152	NO	

* Fisher's exact test

Table 3. PAM with 3 Clusters vs FLIPI

	FLIPI0	FLIPI1	FLIPI2	FLIPI3	FLIPI4	FLIPI5	P-value
PAM1	0	3	22	25	6	2	<0.0001*
PAM2	0	1	8	18	4	1	
PAM3	10	17	10	0	0	0	
				Medoids			
	ID	Stage	Age	Hb	LDH	> 4 node site?	
PAM1	82	4	55.3	13.4	174	YES	
PAM2	122	3	64.0	13.0	152	YES	
PAM3	43	1	62.7	13.4	152	NO	

* Fisher's exact test

Table 4. PAM with 4 Clusters vs FLIPI

	FLIPI0	FLIPI1	FLIPI2	FLIPI3	FLIPI4	FLIPI5	P-value
PAM1	0	0	18	25	6	2	<0.0001*
PAM2	0	1	8	18	4	1	
PAM3	4	5	10	0	0	0	
PAM4	6	15	4	0	0	0	
				Medoids			
	ID	Stage	Age	Hb	LDH	> 4 node site?	
PAM1	82	4	55.3	13.4	174	YES	
PAM2	122	3	64.0	13.0	152	YES	
PAM3	73	3	49.4	13.7	123	NO	
PAM4	43	1	62.7	13.4	152	NO	

* Fisher's exact test

Table 5. PAM with 2 Clusters vs FLIPI2

	FLIPI2-0	FLIPI2-1	FLIPI2-2	FLIPI2-3	FLIPI2-4	P- value
PAM1	0	16	15	9	4	<0.0001*
PAM2	12	23	1	1	1	
			Medoids			
	ID	Age	Hb	Node Diam.	$\beta 2$-microglobulin	Bone Marrow Involvement?
PAM1	94	65.2	13.8	3.80	1.91	YES
PAM2	73	49.4	13.7	3.20	1.52	NO

* Fisher's exact test

Table 6.PAM with 3 Clusters vs FLIPI2

	FLIPI2-0	FLIPI2-1	FLIPI2-2	FLIPI2-3	FLIPI2-4	P- value
PAM1	0	13	14	2	0	<0.0001*
PAM2	12	23	1	1	1	
PAM3	0	3	11	7	2	
			Medoids			
	ID	Age	Hb	Node Diam.	$\beta 2$-microglobulin	Bone Marrow Involvement?
PAM1	56	44.8	13.6	3.50	1.80	YES
PAM2	73	49.4	13.7	3.20	1.52	NO
PAM3	94	65.2	13.8	3.80	1.91	YES

* Fisher's exact test

Table 7. PAM with 3 Clusters vs FLIPI2

	FLIPI2-0	FLIPI2-1	FLIPI2-2	FLIPI2-3	FLIPI2-4	P- value
PAM1	0	13	14	2	0	<0.0001*
PAM2	12	11	1	0	0	
PAM3	0	12	0	1	1	
PAM4	0	3	11	7	2	
			Medoids			
	ID	Age	Hb	Node Diam.	$\beta 2$-microglobulin	Bone Marrow Involvement?
PAM1	56	44.8	13.6	3.50	1.80	YES
PAM2	73	49.4	13.7	3.20	1.52	NO
PAM3	219	59.0	14.7	7.8	1.6	NO
PAM4	94	65.2	13.8	3.80	1.91	YES

* Fisher's exact test

Figure 1. Average Silouette Width using Prognostic Factor in FLIPI

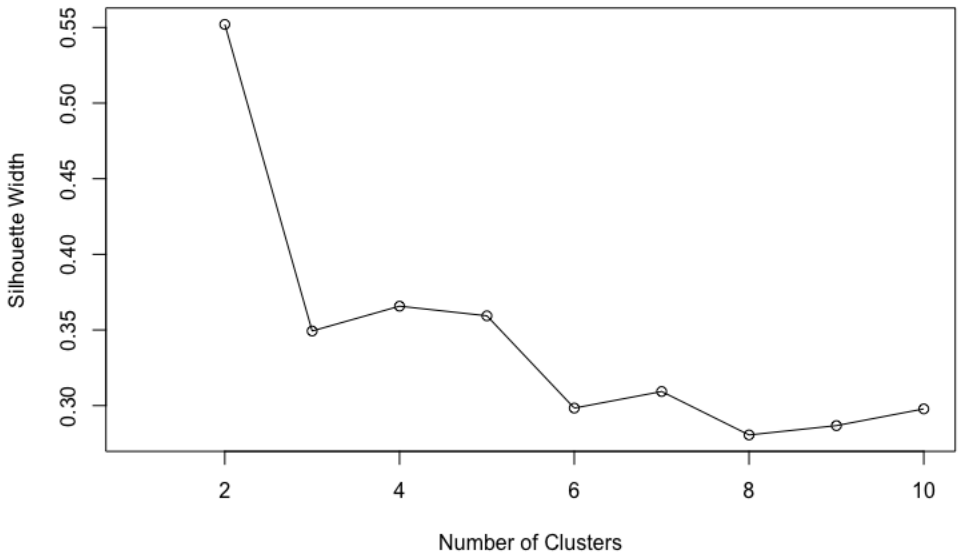

Figure 2. Kaplan-Meier Curves Stratified by FLIPI

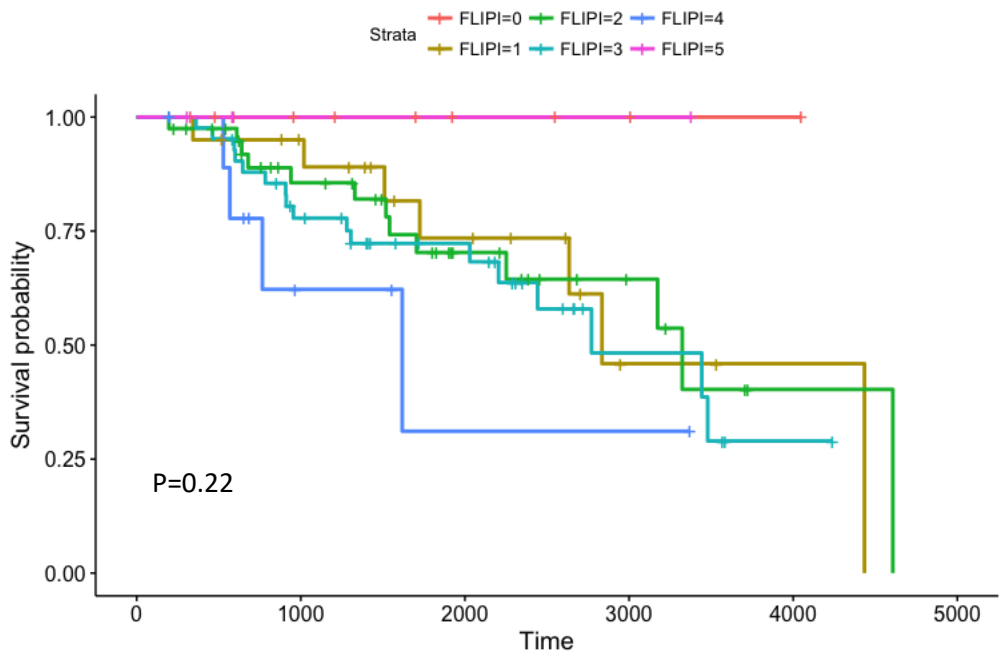

Figure 3. Kaplan-Meier Curves Stratified by PAM with 2 Clusters -FLIPI Prognostic Factors

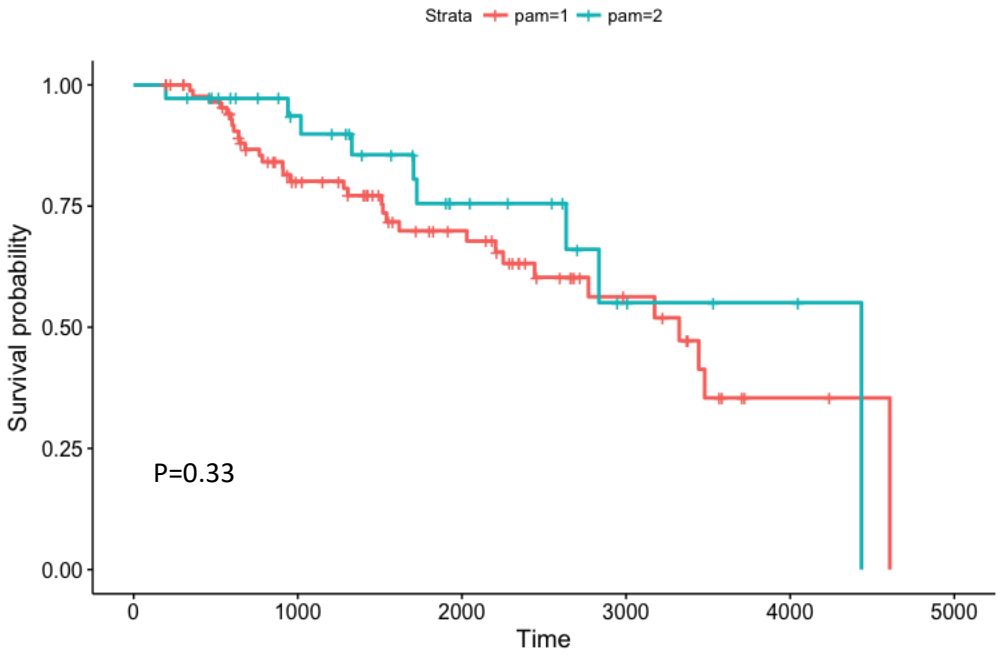

Figure 4. Kaplan-Meier Curves Stratified by PAM with 3 Clusters -FLIPI Prognostic Factors

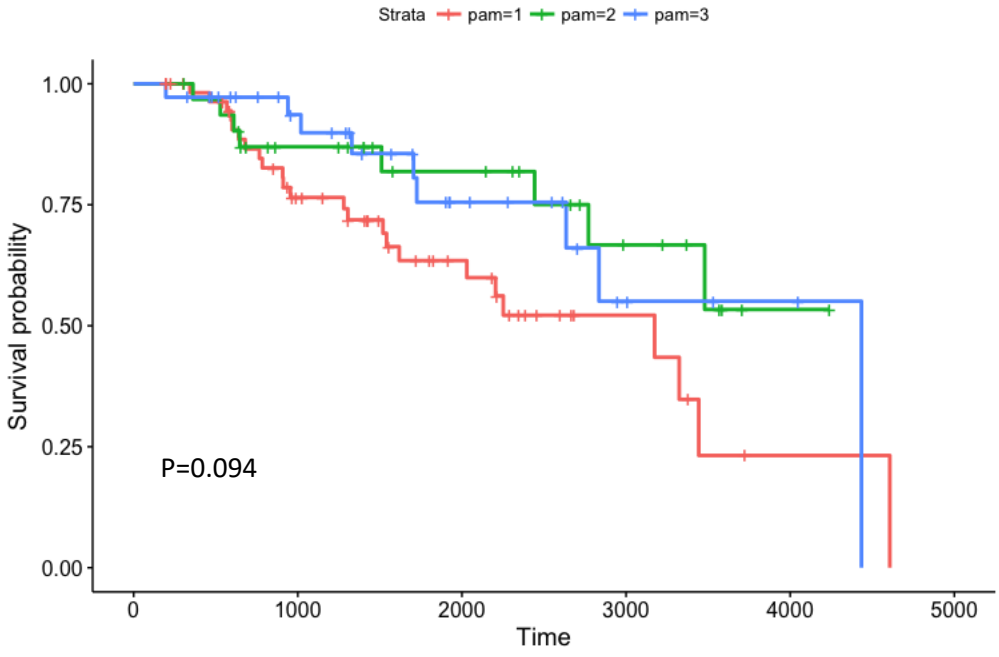

Figure 5. Kaplan-Meier Curves Stratified by FLIPI 3 Risk Groups

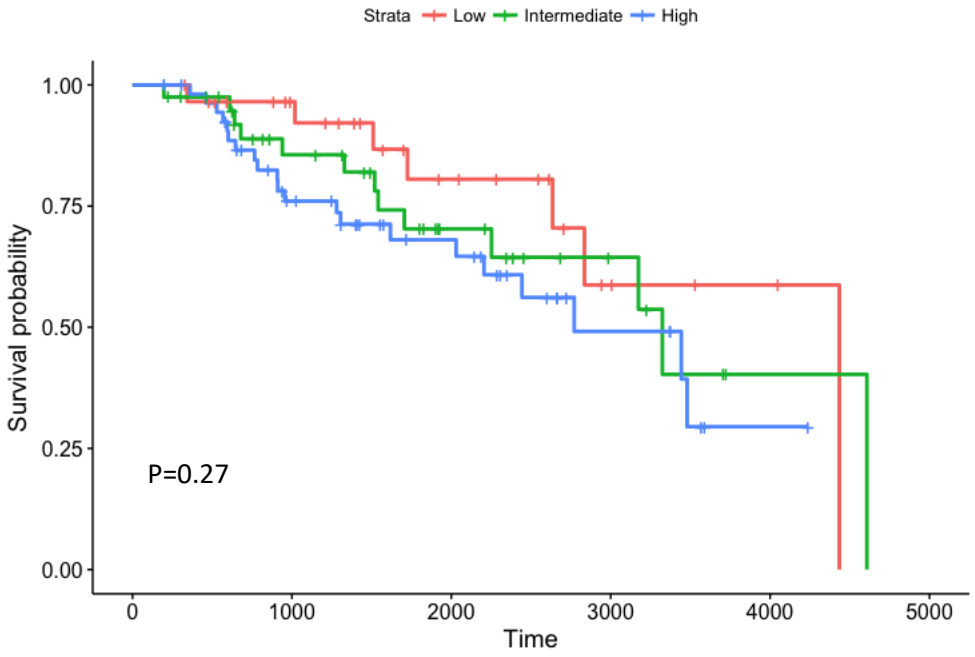

Figure 6. Kaplan-Meier Curves Stratified by PAM with 4 Clusters -FLIPI Prognostic Factors

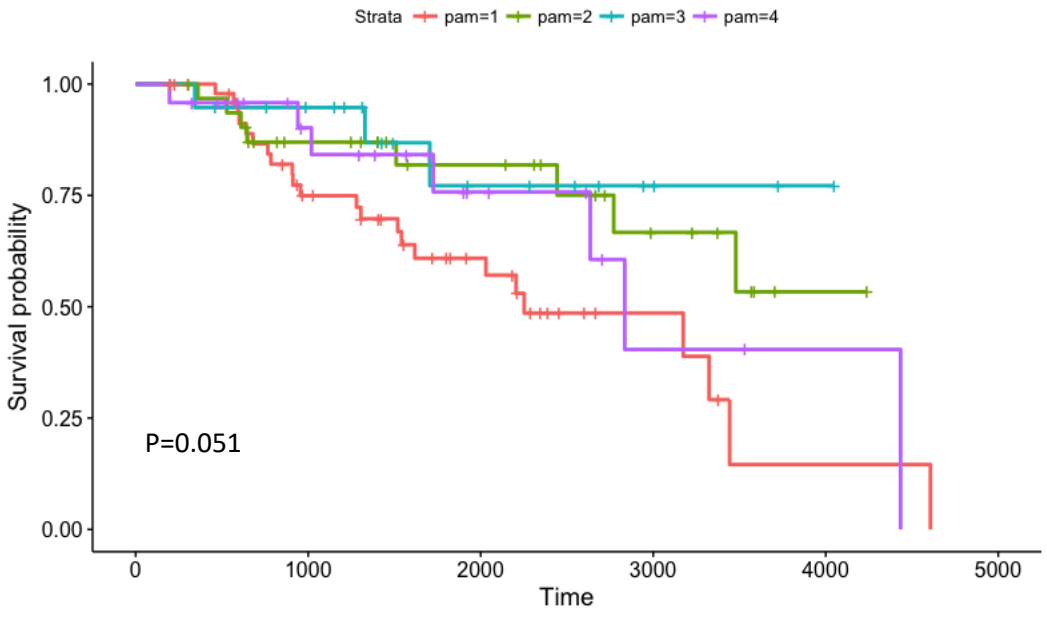

Figure 7. Average Silouette Width using Prognostic Factor in FLIPI2

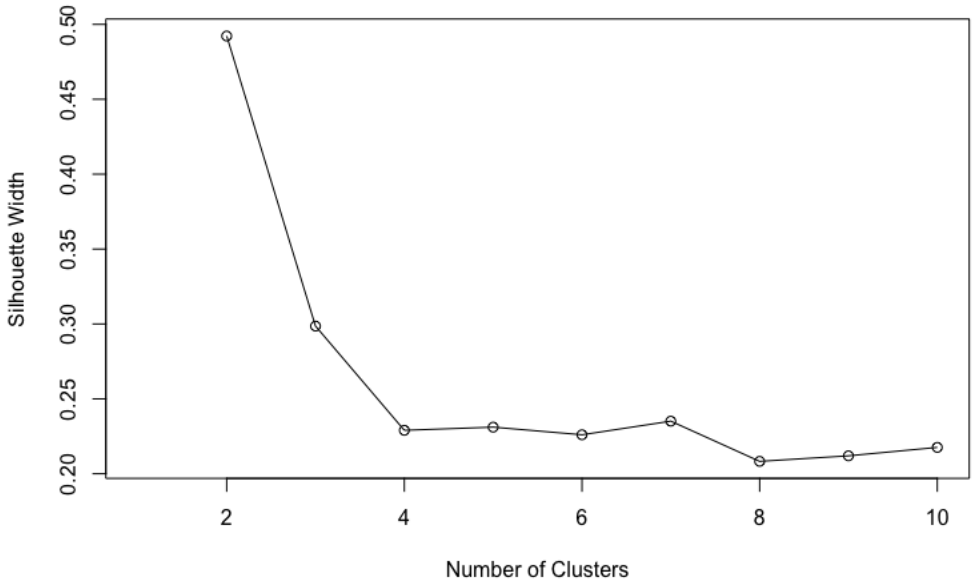

Figure 8. Kaplan-Meier Curves Stratified by FLIPI2

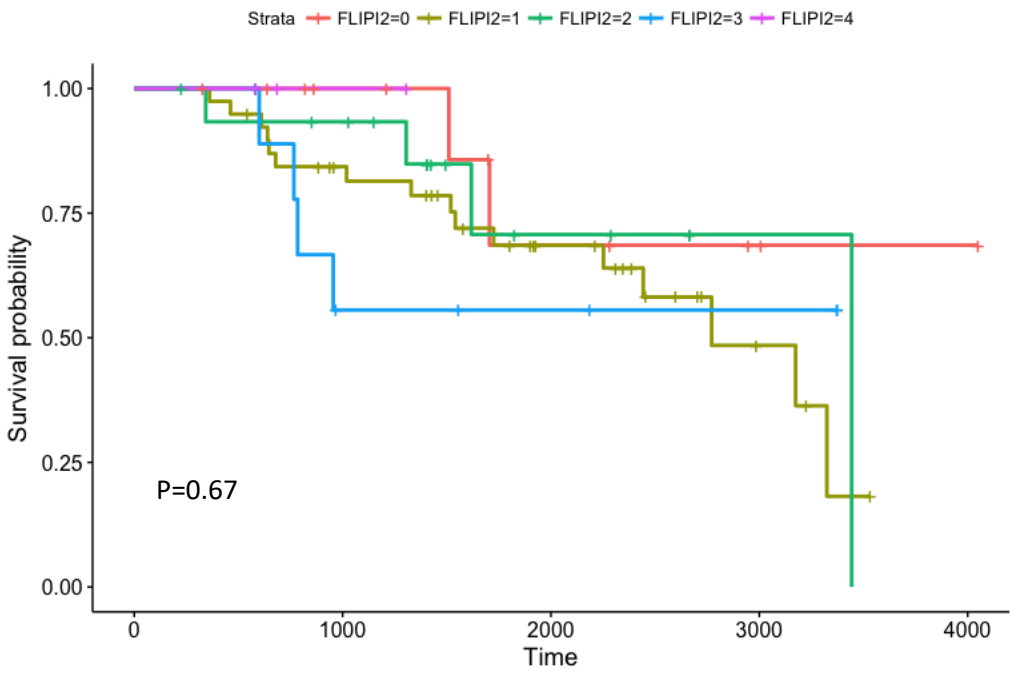

Figure 9. Kaplan-Meier Curves Stratified by PAM with 2 Clusters -FLIPI2 Prognostic Factors

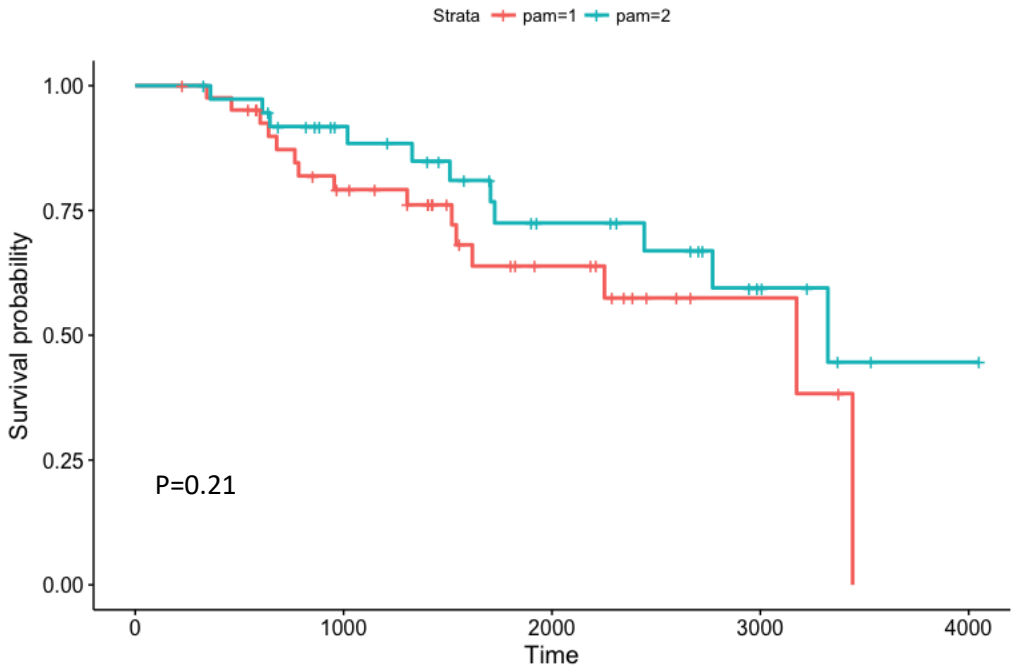

Figure 10. Kaplan-Meier Curves Stratified by PAM with 3 Clusters -FLIPI2 Prognostic Factors

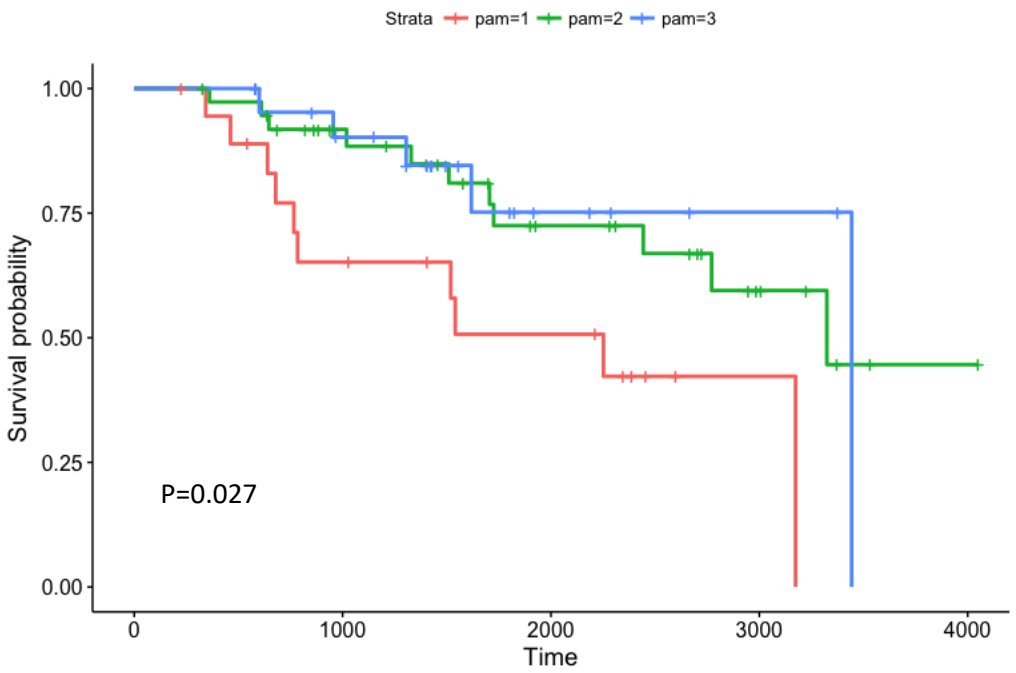

Figure 11. Kaplan-Meier Curves Stratified by FLIPI2 3 Risk Groups

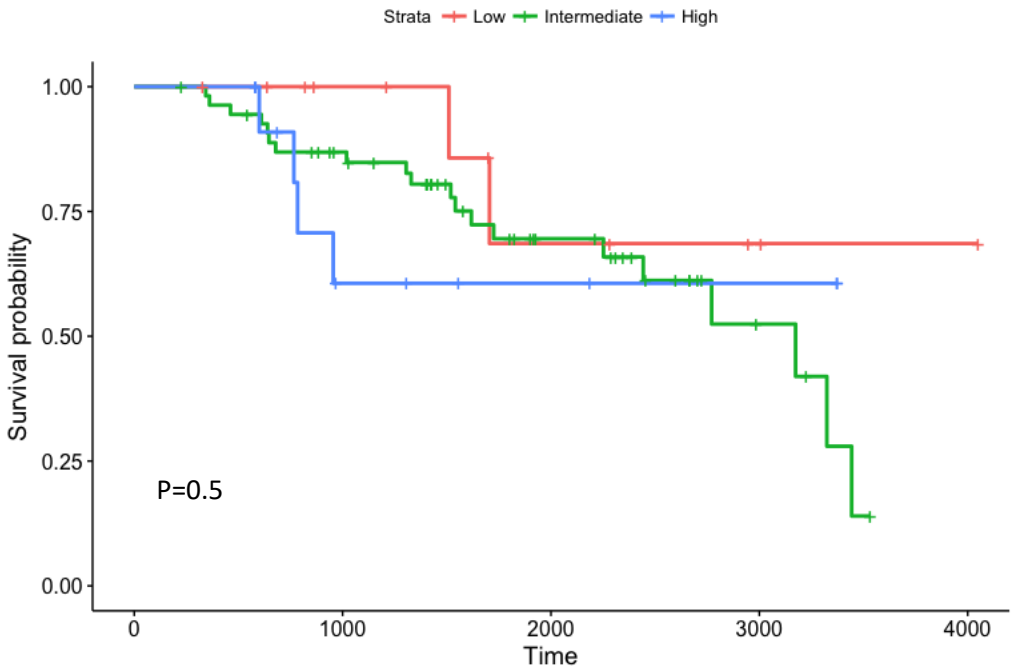

Figure 12. Kaplan-Meier Curves Stratified by PAM with 4 Clusters -FLIPI2 Prognostic Factors