

# 找出影響食道癌患者的重要變數

存活分析期末報告 B082040005 高念慈

## 一、動機與目的

食道雖然不像肝臟一樣是沉默器官，但也是個大家平常不會特別注意到的器官，很容易發現時也是中後期，所以要是能找出影響食道癌患者的特徵，有助於更好的處理這個疾病，早點發現，早點治療。

## 二、資料介紹

資料來源: [https://www.cbioportal.org/study/clinicalData?id=stes\\_tcga\\_pub](https://www.cbioportal.org/study/clinicalData?id=stes_tcga_pub)

資料來自 cBioPortal，是一份關於 559 位食道癌患者在一段期間內的檢查報告，有 108 個欄位，資料含缺失值。

變數介紹:

手動移除缺失值超過八成的變數，再去掉重複、相關性高和變異數小的變數，對剩下變數做介紹。

事件:是否死亡

“Overall.Survival.Status”：總體.生存.狀態，(chr.1:DECEASED，0:LIVING)

存活時間:從首次診斷出疾病到死亡或是研究結束，單位:月

“Overall.Survival..Months.”：總體.生存..月，(num.)

1. "Absolute.Extract.Ploidy"：絕對.提取物.倍性，(num.)，倍性水平是指雙倍體體細胞 (2n) 或單倍體配子體細胞 (1n) 中染色體組的數量。
2. "Absolute.Extract.Purity"：絕對提取物純度，(num.)
3. "Diagnosis.Age"：診斷年齡，(num.)，首次診斷出病症或疾病的年齡。
4. "ARID1A.Protein.Coding"：ARID1A 蛋白質編碼，(int.0:NO,1:YES)
5. "CDKN2A.Methylation"：CDKN2A 甲基化，(logi.FALSE,TRUE)，CDKN2A 甲基化可能會導致基因的表達受到抑制或關閉，從而影響細胞週期的調控。
6. "Country"：國家，(chr.)
7. "EBV.Positive"：伊普斯坦-巴爾病毒陽性，(int.)，EBV 是一種 DNA 病毒，它可以感染人類的 B 淋巴細胞，並在體內引起多種疾病。(喉嚨痛、疲勞、淋巴結腫大相關)
8. "Percent.Tumor.Cells"：癌細胞百分比，(num.)
9. "Percent.Tumor.Nuclei"：癌細胞核百分比，(num.)
10. "Person.Neoplasm.Cancer.Status"：人體腫瘤癌症狀態，

(chr.TUMOR\_FREE/WITH\_TUMOR)，描述腫瘤的有無、腫瘤的種類、腫瘤的大小和擴散程度，以及其他與腫瘤狀態相關的信息。

11. "EC"，(int.0:NO,1:YES)，EC 可能代表食道攝影檢查、食道黏膜刷取檢查、食道鏡檢查。
12. "EC.SMG"，(int.0:NO,1:YES)，EC.SMG 可能代表食道超聲內視鏡檢查，也可能指食道超聲引導下淋巴結活檢。
13. "Estimated.Leukocyte.Percentage"：估計的白血球百分比，(num.)，提供關於白血球組成的信息，幫助醫生判斷患者的免疫狀態和健康狀況。
14. "Fraction.Genome.Altered"：基因組改變比例，(num.)，腫瘤細胞中發生基因組改變的比例，如果腫瘤細胞中有多個區域發生了基因組改變，值就會較高。
15. "Gastric.Classification"：胃.分類，(chr.)
16. "GEA.CIN.SMG"，(int.0:NO,1:YES)，可能表示在食道癌檢查中同時進行食道細胞異常檢查和腹部超聲檢查。
17. "GEA.GEJ"，(int.0:NO,1:YES)，可能表示在食道癌檢查中同時進行上消化道鏡檢查以及專注於胃食道連接處的評估。
18. "GEA.SMG"，(int.0:NO,1:YES)，可能表示在食道癌檢查中同時進行腹部超聲和上消化道鏡檢查。
19. "Genome.Doublings"：基因體倍增，(int.)，細胞的染色體數目增加為正常倍數的兩倍，它會影響細胞的功能和癌症的發展。
20. "PIK3CA.Protein.Coding"：PIK3CA 蛋白質編碼，(int.0:NO,1:YES)，PIK3CA 基因突變或過度表達與多種腫瘤的發生和進展有關，包括乳腺癌、卵巢癌、大腸癌、胃癌等。
21. "Neoplasm.Histologic.Grade"：腫瘤.組織學.分級，(chr.)，低分級通常表示腫瘤細胞較接近正常組織，分化程度較好，高分級則表示腫瘤細胞較不正常，分化程度較差。
22. "HER2.Amplification.Status"：人類表皮生長因子受體 2 基因放大的狀態，(int.0:正常,1:放大)，HER2 放大可能促進癌症細胞的生長和擴散，在乳腺癌中相對常見。
23. "KRAS.Protein.Coding"：KRAS 蛋白編碼，(int.0:NO,1:YES)，KRAS 基因是一個重要的腫瘤抑制基因，它在調節細胞生長和分化中起著關鍵的作用，特別是大腸癌和肺癌等。
24. "MLH1.Methylation"：MLH1 甲基化，(logi.FALSE,TRUE)，MLH1 甲基化導致基因組中的 MLH1 基因失去表達，這意味著蛋白質無法正常產生。缺乏 MLH1 蛋白質會干擾 DNA 配對修復機制，使得 DNA 錯誤無法被及時修復，增加細胞的突變風險。
25. "MSI.Status"：微衛星不穩定性的狀態，(chr.)，微衛星穩定 (MSS)：腫瘤組織中微衛星序列保持穩定，沒有明顯的不穩定性，微衛星不穩定 (MSI)：

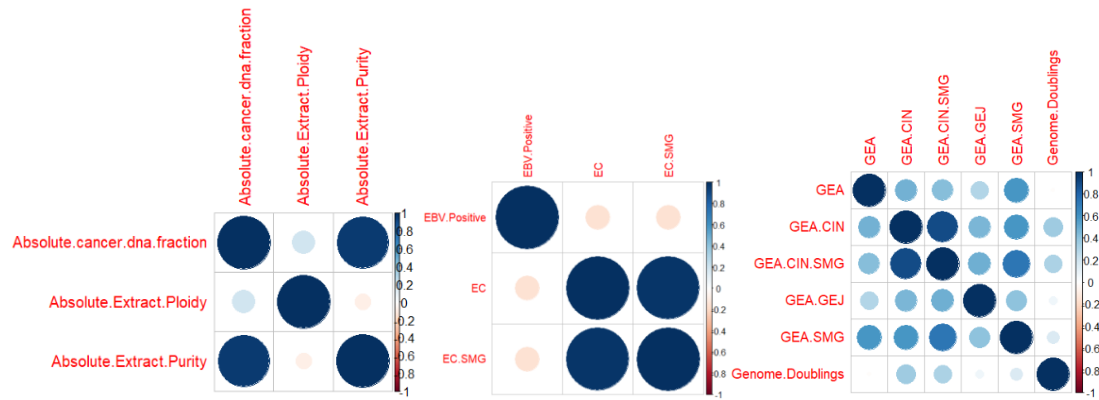
腫瘤組織中出現微衛星不穩定性，表明微衛星序列的錯誤或缺陷。

26. "Mutation.Rate": 突變率，(num.)，在一個特定基因或基因群中突變事件發生的頻率。
27. "New.Neoplasm.Event.Post.Initial.Therapy.Indicator": 新的腫瘤。事件。後。初始。治療。指標，(chr,YES/NO)，在初次治療後是否出現新的腫瘤事件。  
Positive/Yes：可能意味著癌症復發、轉移或出現了新的原發腫瘤，  
Negative/No：意味著患者在初次治療後沒有出現新的腫瘤或腫瘤復發的跡象。
28. "RHOA.Protein.Coding": RHOA 蛋白質編碼，(int.0:NO,1:YES)，RHOA 基因的突變或異常活化可能導致細胞的遷移、侵襲和轉移能力增強，從而促進腫瘤的進展和轉移。
29. "Surgical.Margin.Resection.Status": 手術邊緣切除狀態，(chr.)，指在手術過程中腫瘤或病變組織周圍的切緣是否被完全切除的狀態。正常組織（Negative Margin，切緣未檢測到腫瘤或病變組織）不確定（Close Margin，切緣附近可能存在腫瘤或病變組織）腫瘤或病變組織殘留（Positive Margin，切緣檢測到腫瘤或病變組織）。
30. "SCC": 鱗狀細胞癌，(int.,0:NO,1:YES)
31. "SCC.SMG", (int.,0,1)，(int.,0:NO,1:YES)，"SMG"可以代表 SLNM、Mediastinoscopy、Groove Node Dissection 等檢查項目。
32. "SCNA.Cluster.GEA": 基因組拷貝數異常在胃食道腺癌中的聚類分析，(chr.)，描述基因組拷貝數異常的模式，使用聚類分析方法將相似的拷貝數異常模式的腫瘤分到同一個聚類中。
33. "Sex": 性別。
34. "TMB..nonsynonymous.": 腫瘤組織中非同義突變的數量，(num.)，高 TMB.nonsynonymous 暗示著腫瘤組織中存在大量的基因突變，可能對免疫系統的識別和攻擊產生影響。
35. "TP53.Protein.Coding": TP53 蛋白質編碼，(int.0:NO,1:YES)，TP53 是人類體內最為重要的腫瘤抑制基因之一，TP53 的突變可能導致 p53 蛋白質的功能喪失或減弱，從而降低細胞對 DNA 損傷的應答和防止腫瘤發展的能力。
36. "Tumor.Stage": 腫瘤階段，(chr.)

### 三、資料前處理

(圖一)至(圖六)為高度相關的變數。

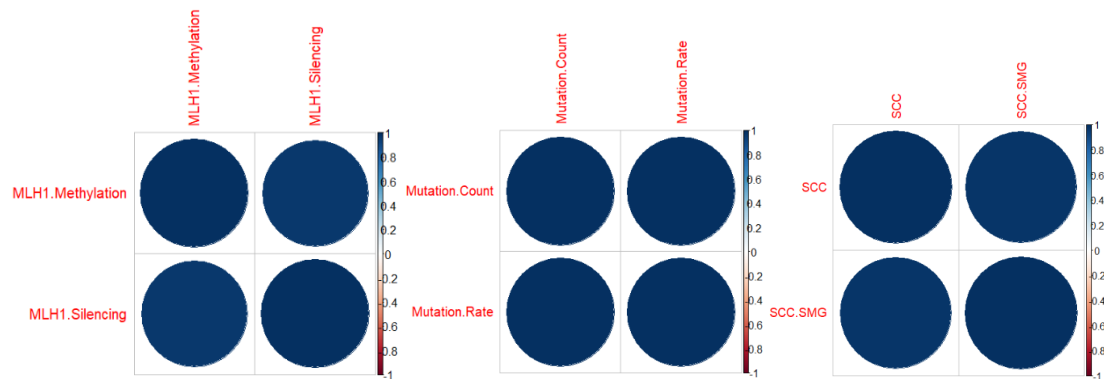
對剩下資料，(559 obs. of 35 variables)，補缺失值，考慮到資料取得不易，且不想隨意創造數據，於是把資料分出存活和死亡，分別對數值變數填補各列中位數，類別變數填補眾數，(圖七)、(圖八)。



(圖一)

(圖二)

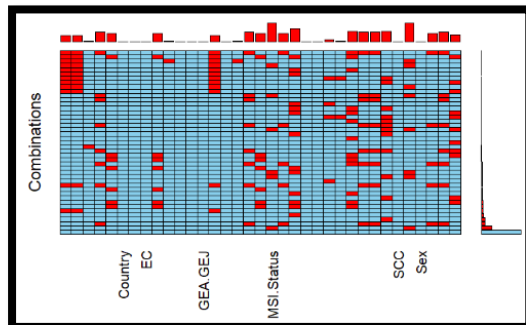
(圖三)



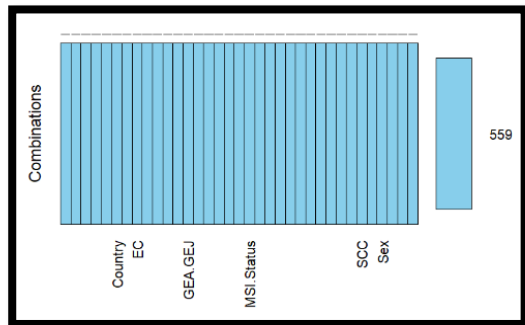
(圖四)

(圖五)

(圖六)



(圖七)，補缺失值前



(圖八)，補缺失值後

#### 四、挑選變數

從剩下 33 個變數中找出會顯著影響食道癌風險的變數，significant level 皆為 0.05，對 24 個類別變數採取 Log rank test；9 個數值變數採取 Wald test、Likelihood ratio test 或 Score test。(圖九)、(圖十)

##### (一)Log rank test

Test 整個風險函數有無差異， $H_0: h_1(t) = h_2(t)$ ，不考慮與其他變數一起作用，變數中沒有像藥效那種前期強，後期弱的變數，故把權重設為:  $p = q = 0$ 。

顯著變數: 9 個, (圖十一)至(圖十九)

1. Country : 國家, (chr.),  $p < 0.0001 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
2. Genome.Doublings : 基因體倍增, (int.),  $p = 0.002 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
3. Neoplasm.Histologic.Grade : 腫瘤.組織學.分級, (chr.),  $p = 0.013 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
4. MLH1.Methylation : MLH1 甲基化, (logi.FALSE,TRUE),  $p = 0.023 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
5. New.Neoplasm.Event.Post.Initial.Therapy.Indicator : 新的腫瘤.事件.後.初始.治療.指標, (chr,YES/NO),  $p < 0.0001 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
6. Person.Neoplasm.Cancer.Status : 人體腫瘤癌症狀態, (chr.TUMOR\_FREE/WITH\_TUMOR),  $p < 0.0001 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
7. Surgical.Margin.Resection.Status : 手術邊緣切除狀態, (chr.),  $p < 0.0001 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
8. Sex : 性別,  $p = 0.017 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數
9. Tumor.Stage : 腫瘤階段, (chr.),  $p < 0.0001 < 0.05$ , 拒絕  $H_0$ , 顯著影響風險函數

(更新資料變數: DATA5, 559 obs. of 20 variables)

## (二) Local test

看連續變數是否對食道癌生存有顯著影響, 9 個連續變數中, 只有 3 個對食道癌的生存有顯著影響,  $H_0: \beta_1 = \beta_{10} = (0)$ 。

Wald test :

1. pvalue:0.4105631  $> 0.05$ , 不拒絕  $H_0$ , 係數顯著為 0, Absolute.Extract.Ploidy 對食道癌生存沒有顯著影響
2. pvalue:0.1194808  $> 0.05$ , 不拒絕  $H_0$ , 係數顯著為 0, Absolute.Extract.Purity 對食道癌生存沒有顯著影響
3. pvalue:0.01779964  $< 0.05$ , 拒絕  $H_0$ , 係數不顯著為 0, 年齡對食道癌生存有顯著影響

Likelihood ratio test :

4. p-value:0.0496991  $< 0.05$ , 拒絕  $H_0$ , 係數不顯著為 0, 白血球百分比對食道癌生存有顯著影響
5. p-value:0.0483102  $< 0.05$ , 拒絕  $H_0$ , 係數不顯著為 0, 基因組改變比例對食道癌生存有顯著影響

6.  $p\text{-value}: 0.4504393 > 0.05$ ，不拒絕  $H_0$ ，係數顯著為 0，突變率對食道癌生存沒有顯著影響

Score test :

7.  $p\text{-value}: 0.7248046 > 0.05$ ，不拒絕  $H_0$ ，係數顯著為 0，癌細胞百分比對食道癌生存沒有顯著影響

8.  $p\text{-value}: 0.7152439 > 0.05$ ，不拒絕  $H_0$ ，係數顯著為 0，癌細胞核百分比對食道癌生存沒有顯著影響

9.  $p\text{-value}: 0.7928437 > 0.05$ ，不拒絕  $H_0$ ，係數顯著為 0，腫瘤組織中非同義突變的數量對食道癌生存沒有顯著影響

顯著變數: 3 個，年齡、白血球百分比、基因組改變比例

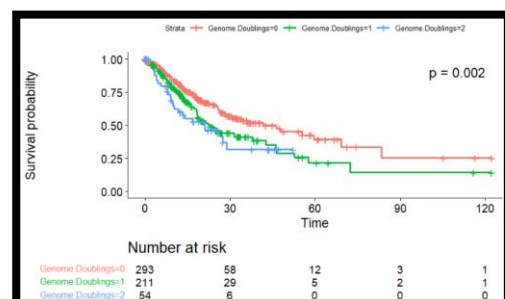
(更新最後資料變數: DATA6，559 obs. of 14 variables)

```
Group(類別變數):
1. ARID1A.Protein.Coding : ARID1A 蛋白質編碼, (int.0:NO,1:YES)
2. CDKN2A.Methylation : CDKN2A 甲基化, (logi.FALSE,TRUE)
3. Country : 國家, (chr.)
4. EBV.Positive : 伊普斯坦-巴爾病毒陽性, (int.)
5. EC, (int.0:NO,1:YES)
6. Gastric.Classification : 胃.分類, (chr.)
7. GEA.CIN.SMG, (int.0:NO,1:YES)
8. GEA.GEJ, (int.0:NO,1:YES)
9. Genome.Doublings : 基因體倍增, (int.)
10. Neoplasm.Histologic.Grade : 腫瘤.組織學.分級, (chr.)
11. HER2.Amplification.Status" : 人類表皮生長因子受體2基因放大的狀態, (int.0:正常,1:放大)
12. KRAS.Protein.Coding : KRAS蛋白質編碼, (int.0:NO,1:YES)
13. MLH1.Methylation : MLH1甲基化, (logi.FALSE,TRUE)
14. MSI.Status : 微衛星不穩定性的狀態, (chr.)
15. New.Neoplasm.Event.Post.Initial.Therapy.Indicator" : 新的腫瘤.事件.後.初始.治療.指標, (chr,YES/NO)
16. Person.Neoplasm.Cancer.Status : 人體腫瘤癌症狀態, (chr.TUMOR_FREE/WITH_TUMOR)
17. PIK3CA.Protein.Coding : PIK3CA 蛋白質編碼, (int.0:NO,1:YES)
18. RHOA.Protein.Coding : RHOA 蛋白質編碼, (int.0:NO,1:YES)
19. Surgical.Margin.Resection.Status : 手術邊緣切除狀態, (chr.)
20. SCC : 鱗狀細胞癌, (int.,0:NO,1:YES)
21. SCNA.Cluster.GEA : 基因組拷貝數異常在胃食道腺癌中的聚類分析, (chr.)
22. Sex : 性別
23. TP53.Protein.Coding : TP53 蛋白質編碼, (int.0:NO,1:YES)
24. Tumor.Stage : 腫瘤階段, (chr.)
```

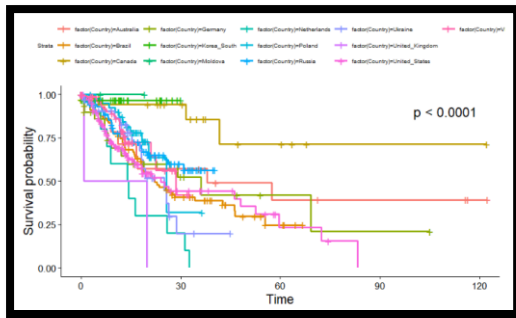
(圖九)，類別變數

```
1. Absolute.Extract.Ploidy
2. Absolute.Extract.Purity
3. Diagnosis.Age
4. Estimated.Leukocyte.Percentage
5. Fraction.Genome.Altered
6. Mutation.Rate
7. Percent.Tumor.Cells
8. Percent.Tumor.Nuclei
9. TMB..nonsynonymous.
```

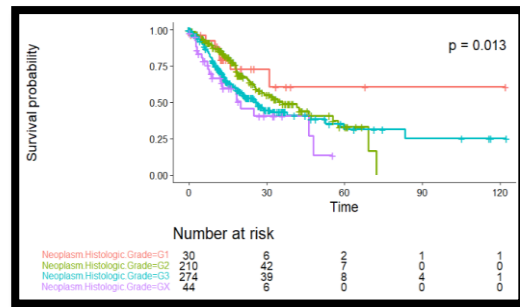
(圖十)，數值變數



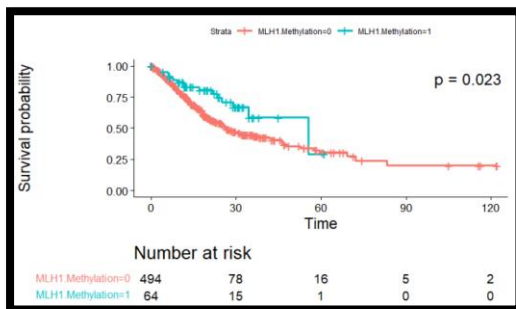
(圖十一)



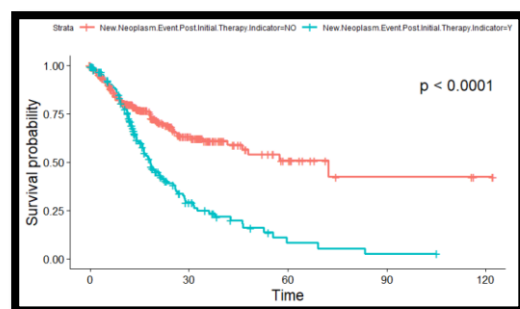
(圖十二)



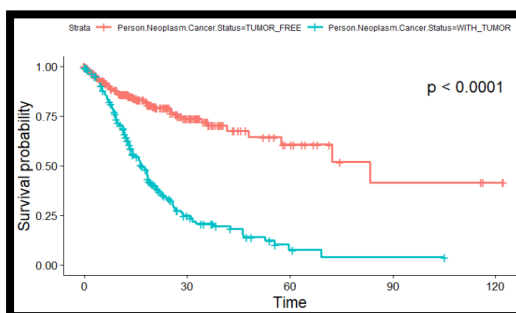
(圖十三)



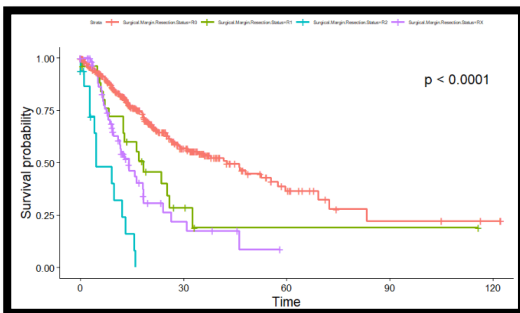
(圖十四)



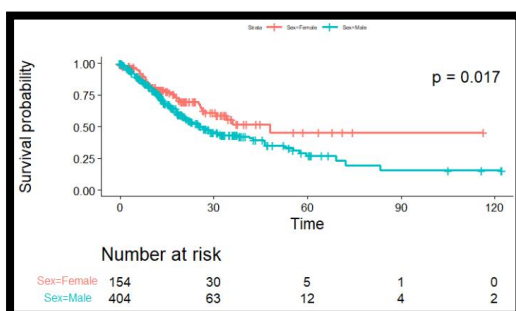
(圖十五)



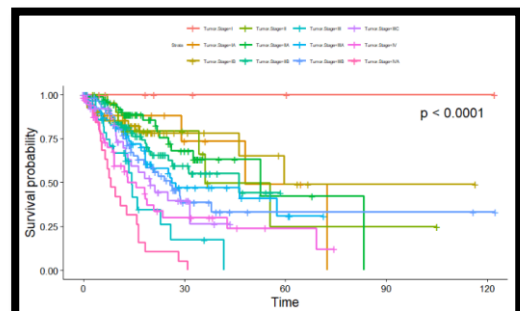
(圖十六)



(圖十七)



(圖十八)



(圖十九)

## 五、Cox Proportional Hazard (PH) Model

不考慮任何變數下的 Kaplan-Meier estimate 和累積風險函數(圖二十)、(圖二一)

### (一) Schoenfeld residuals test (Time varying cox model)

H0:係數隨時間變化，(圖二二)

額外對下面兩個變數個別增加新變數，並搭配  $\log(t+1)$  擬合模型，處理時間對係數的影響。

1. New.Neoplasm.Event.Post.Initial.Therapy.Indicator,  $p = 5.5e-05 < 0.05$ ，變數係數隨時間變化，(圖二三)
2. Person.Neoplasm.Cancer.Status,  $p = 9.3e-05 < 0.05$ ，拒絕 H0，變數係數隨時間變化，(圖二四)

### (二) AIC 自動挑選變數

最後模型，(圖二五)、(圖三六)

### (三) Model Check

Check 的模型為:最後資料集，Data6，559 obs. of 14 variables，(圖二六)，因為 Time varying cox model 在算 Martingale residual 時，自己新增的變數他不能順便去做處理。

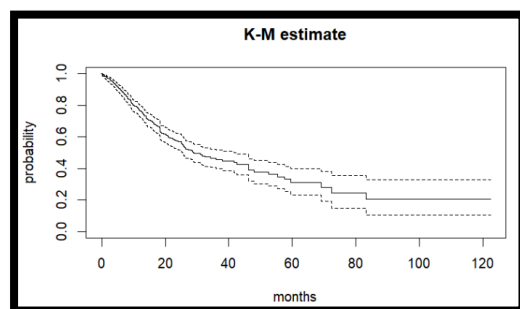
此模型的 Cox baseline cumulative hazard function 和 Kaplan-Meier estimate (圖二七)、(圖二八)

#### 1. Cox snell's residual

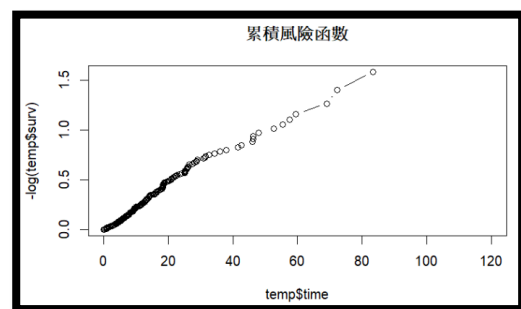
用(snell 殘差, event)估計累積風險函數，接近 45 度線，模型應該足夠(圖二九)

#### 2. Martingale residual

對連續變數做殘插圖，看起來皆沒有特別的趨勢，模型可能足夠(圖三十)至(圖三五)



(圖二十)

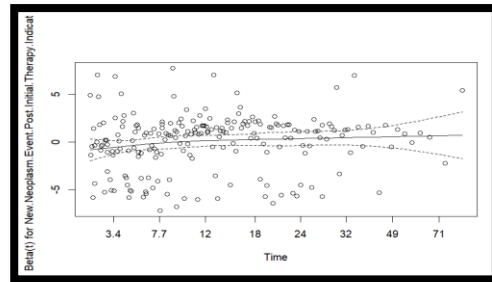


(圖二一)

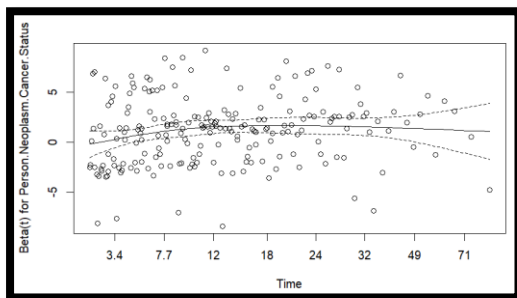


	chisq	df	p
Diagnosis.Age	2.10e+01	1	0.64713
Country	1.72e+01	12	0.14233
Estimated.Leukocyte.Percentage	2.89e+02	1	0.86506
Fraction.Genome.Altered	4.36e+01	1	0.50909
Genome.Duplications	3.61e+02	1	0.84925
Neoplasm.Histologic.Grade	4.22e+00	3	0.23846
MLH1.Methylation	8.34e+04	1	0.97697
New.Neoplasm.Event.Post.Initial.Therapy.Indicator	1.63e+01	1	5.5e-05
Person.Neoplasm.Cancer.Status	1.53e+01	1	9.3e-05
Surgical.Margin.Resection.Status	2.07e+00	3	0.55795
Sex	3.06e+03	1	0.95588
Tumor.Stage	2.73e+00	3	0.43528
GLOBAL	6.20e+01	29	0.00035

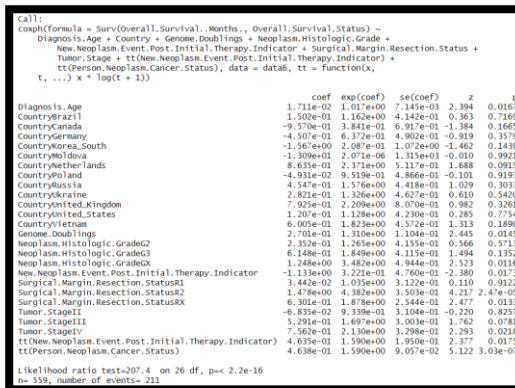
(圖二二)



(圖二三)



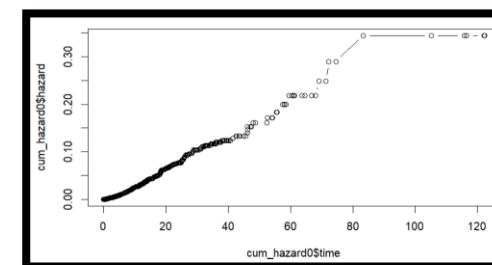
(圖二四)



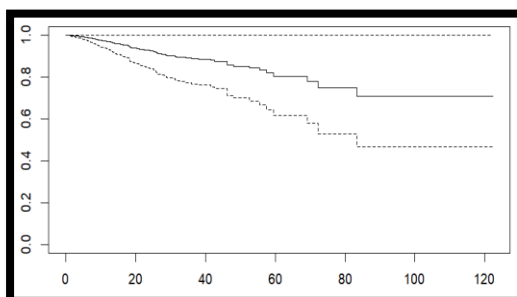
(圖二五)

	coef	exp(coef)	se(coef)	z	Pr(> z )
Diagnosis.Age	1.808e-02	1.018e+00	7.351e-03	2.460	0.01391 *
Country[Brazil]	2.398e-01	1.271e+00	4.316e-01	0.556	0.57848
Country[Canada]	-8.436e-01	4.301e-01	7.014e-01	-1.203	0.22908
Country[Germany]	-4.149e-01	6.604e-01	2.016e-01	-0.837	0.40805
Country[Korea_South]	-1.500e+00	2.232e-01	1.086e+00	-1.381	0.16730
Country[Moldova]	-1.314e+01	1.967e+06	1.317e+03	-0.010	0.99204
Country[Netherlands]	9.738e-01	2.653e+00	5.348e-01	1.825	0.06804
Country[Poland]	-3.476e-03	9.965e-01	4.968e-01	-0.007	0.99442
Country[Russia]	4.549e-01	1.576e+00	4.485e-01	1.014	0.31037
Country[Ukraine]	2.777e-01	1.320e+00	4.661e-01	0.596	0.55140
Country[United_Kingdom]	8.566e-01	2.355e+00	8.133e-01	1.053	0.29220
Country[United_States]	1.592e-01	1.175e+00	4.313e-01	0.369	0.71210
Country[Vietnam]	6.382e-01	1.891e+00	4.656e-01	1.371	0.17050
Estimated.Leukocyte.Percentage	-6.688e-01	5.123e-01	7.369e-01	-0.908	0.36409
Fraction.Genome.Altered	-8.142e-01	4.430e-01	3.306e-01	-1.534	0.12493
Genome.Duplications	3.340e-01	1.397e+00	1.222e-01	2.734	0.00626 **
Neoplasm.Histologic.Grade2	2.763e-01	1.254e+00	4.199e-01	0.539	0.58986
Neoplasm.Histologic.Grade3	6.107e-01	1.842e+00	4.182e-01	1.460	0.14423
Neoplasm.Histologic.GradeX	1.224e+00	3.402e+00	5.042e-01	2.429	0.01515 *
MLH1.Methylation	-1.240e-01	8.834e-01	2.853e-01	-0.434	0.66398
New.Neoplasm.Event.Post.Initial.Therapy.Indicator	-1.004e+00	3.662e-01	5.790e-01	-1.735	0.08278 *
Person.Neoplasm.Cancer.Status	1.701e-01	8.455e-01	5.718e-01	-0.288	0.76604
Surgical.Margin.Resection.Status1	1.955e-01	1.020e+00	3.132e-01	0.062	0.95023
Surgical.Margin.Resection.Status2	1.495e+00	4.459e+00	1.631e-01	4.113	1.91e-05 ***
Surgical.Margin.Resection.StatusX	6.348e-01	1.887e+00	2.598e-01	2.443	0.01455 *
Sex[Male]	-4.893e-02	9.523e-01	1.887e-01	-0.259	0.79346
Tumor.StageII	-6.411e-02	9.379e-01	3.108e-01	-0.206	0.83657
Tumor.StageIII	5.305e-01	1.700e+00	3.012e-01	1.761	0.07824 *
Tumor.StageIV	7.852e-01	2.193e+00	1.371e-01	2.364	0.01807 *
tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator)	4.227e-01	1.526e+00	2.314e-01	1.827	0.06770 *
tt(Person.Neoplasm.Cancer.Status)	5.201e-01	1.682e+00	2.286e-01	2.275	0.02290 *

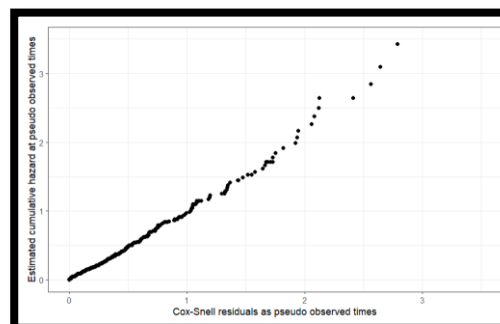
(圖二六)



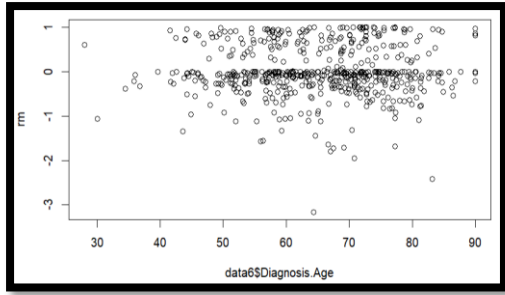
(圖二七)



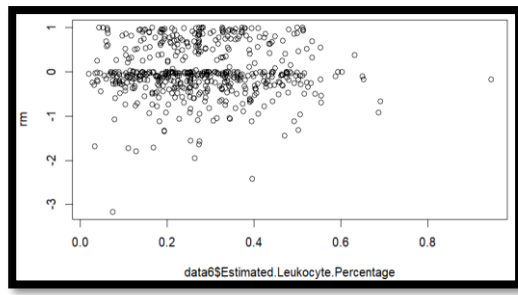
(圖二八)



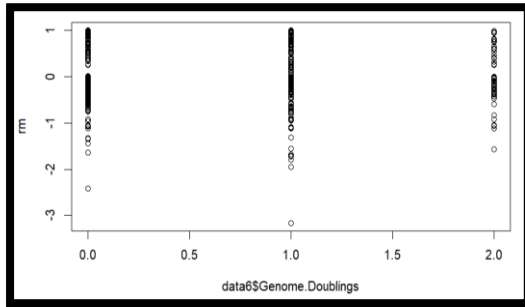
(圖二九)



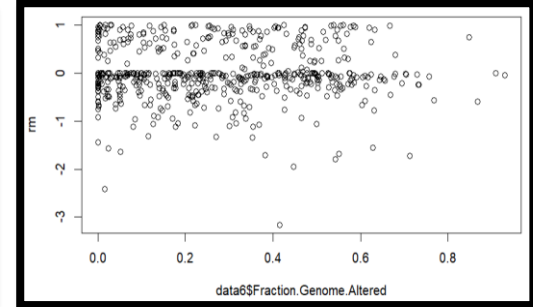
(圖三十)



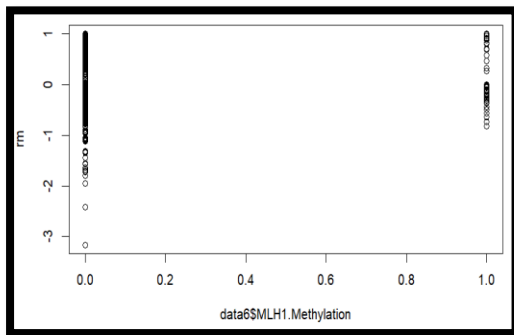
(圖三一)



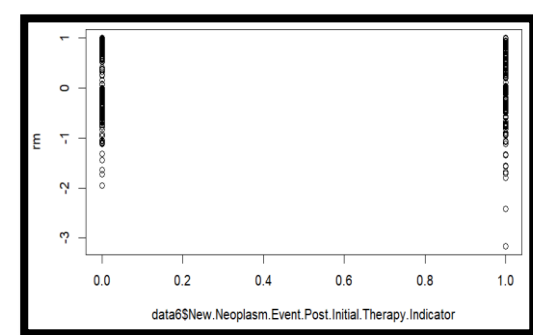
(圖三二)



(圖三三)



(圖三四)



(圖三五)

$$h(t|z) = h_0(t)e^{0.017 \times \text{Diagnosis.Age} + 0.1901 \times \text{Country.Brazil} - 0.957 \times \text{Country.Canada} - 0.451 \times \text{Country.Germany} - 1.567 \times \text{Country.Korea.south} - 13.087 \times \text{Country.Moldova} + 0.864 \times \text{Country.Netherlands} + \dots + 0.4635 \times \text{tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator)}}$$

(圖三六)

	coef	exp(coef)	se(coef)	z	p-value
Diagnosis.Age	1.711e-02	1.017e+00	7.145e-03	2.394	0.0167
Genome.Doublings	2.701e-01	1.310e+00	1.104e-01	2.445	0.0145
Neoplasm.Histologic.GradeGX	1.248e+00	3.482e+00	4.944e-01	2.523	0.0116
New.Neoplasm.Event.Post.Initial.Therapy.Indicator	-1.133e+00	3.221e-01	4.760e-01	-2.380	0.0173
Surgical.Margin.Resection.StatusR2	1.478e+00	4.382e+00	3.503e-01	4.217	2.47e-05
Surgical.Margin.Resection.StatusRX	6.301e-01	1.878e+00	2.544e-01	2.477	0.0133
Tumor.StageIV	7.562e-01	2.130e+00	3.298e-01	2.293	0.0218
tt(New.Neoplasm.Event.Post.Initial.Therapy.Indicator)	4.635e-01	1.590e+00	1.950e-01	2.377	0.0175

(圖三七)

## 六、分析結果及解釋

主要挑有顯著的變數做解釋(圖三七):

1. 診斷年齡每增加一歲，食道癌死亡的風險增加 1.017 倍
2. 基因體倍增(細胞的染色體數目增加為正常倍數的兩倍)<sup>1</sup>，食道癌死亡的風險增加 1.310 倍  
基因體倍增(細胞的染色體數目增加為正常倍數的四倍)<sup>2</sup>，食道癌死亡的風險增加 1.310 平方倍
3. 腫瘤分級 GX(無法判斷)比分級 G1(輕微)，食道癌死亡的風險增加 3.482 倍
4. 初次治療後出現新的腫瘤事件比沒事的，食道癌死亡的風險增加  $0.3221 * t^{1.590}$  倍，當時間越久， $t$  越大，出現事件比沒事的食道癌死亡的風險， $0.3221 * t^{1.590}$ ，越來越大
  - R0：原發腫瘤部位顯微鏡下未見癌細胞
  - R1：癌細胞在顯微鏡下出現在原發腫瘤部位
  - R2：區域淋巴結處有殘留腫瘤
5. 手術邊緣切除狀態 R2(區域淋巴結處有殘留腫瘤)比 R0，食道癌死亡的風險增加 4.382 倍
6. 手術邊緣切除狀態 RX(無法評估)比 R0，食道癌死亡的風險增加 1.878 倍
7. 腫瘤階段 IV 比階段 I(輕微)，食道癌死亡的風險增加 2.130 倍

## 七、結論

資料中其實有很多跟基因相關的變數，但各基因主要影響的器官不同，雖然基因突變容易造成癌症，但能直接影響食道癌的還是需要再進一步分析才能發現，從模型可以看到顯著變數大部分都跟基因沒有關係，主要還是跟患者本身年紀、腫瘤狀況等有較大關係。

## Reference

- <https://www.rdocumentation.org/packages/VIM/versions/6.2.2/topics/aggr>
- <https://blog.yjtseng.info/post/2020-05-13-survival-curve/>
- [https://rpubs.com/kaz\\_yos/resid\\_cox](https://rpubs.com/kaz_yos/resid_cox)
- <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/step>