



# 以臺灣癌症登記資料發展結腸癌預測模型

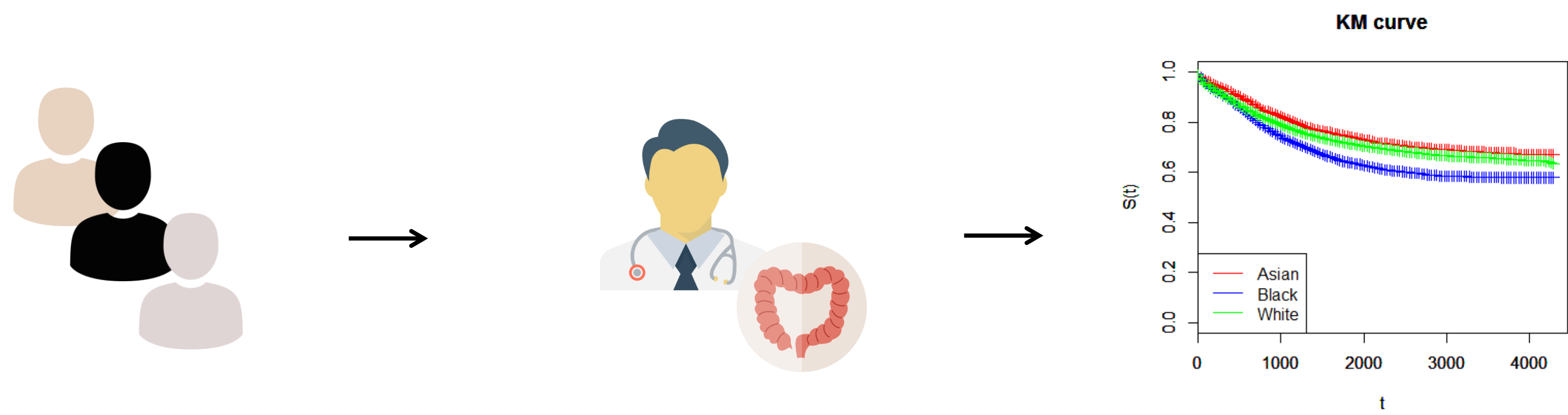
Using National Cancer Registry Data to Develop Prediction Model for Colon Cancer in Taiwan.

學生姓名: 葉冠宏 指導老師: 盧子彬副教授、黃慶杰學長

## 前言

結直腸癌(Colorectal Cancer)為世界第四高致死的惡性腫瘤，僅次於肺、肝、胃癌。我國衛生福利部國民健康署2017年統計，每年發生人數已超過15,000人，為我國發生人數最多的癌症。然而結直腸癌患者預後模型發展的研究大部分集中於西方國家。

因此，本篇研究的主要是以建構台灣人為中心的結腸癌存活模型用以預測台灣人結腸癌特定存活率，並實行於臨床結果存在顯著差異的亞洲和非亞洲患者之間。其次是透過模型區辨以及使用美國國家癌症研究中心(Surveillance, Epidemiology, and End Results Program，以下簡稱SEER)資料進行外部驗證評估不同種族的影響，進而校準模型。



## 研究方法

此研究使用台灣癌症登記中心的資料患者從2008年1月1日蒐集到2015年12月31日，最後追蹤到2017年，一共蒐集到 55,048 位國際疾病傷害及死因分類標準第九版 (ICD-9) 準則判斷為結腸癌病患並透過文獻探討設計排除條件將結腸癌病患區別。

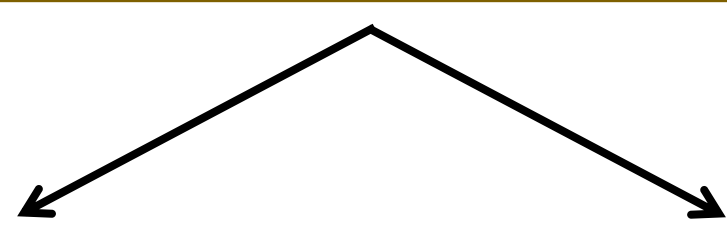
- Without engaging in any treatment at the reporting hospital (N=3,800)
- Suffering from other malignant tumors (N=9,837)
- Pathologic stage 0 (N=1,212)
- Pathologic stage IV (N=4,762)
- Without surgery (N=1)
- Transferred to another hospital (N=362)
- Age at diagnosis under 20 or over 80 (N=4,110)
- Rectosigmoid or rectum (N=11,927)
- Tumor size  $\geq 50$  mm or a cancer of unknown primary (CUP) (N=7,207)
- Number of positive lymph nodes  $\geq 4$  (N=1,561)
- Diagnosis not confirmed by histology (N=0)
- Incomplete survival data (N=245)



### Multivariate Cox Proportional Hazards Model

$HR(x)$  表示在給定  $x$  下會發生事件的風險比  
 $x$  即是自變項的數值  
 $h(t|x)$  表示在第  $t$  個時間點時，給定  $x$  值時的風險  
 $h_0(t)$  表示在第  $t$  個時間點時的基礎風險

$$\log(HR(x)) = \log\left(\frac{h(t|x)}{h_0(t)}\right) = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$



分割資料進行內部驗證  
10,024 位納入分析  
6,683 訓練資料(Training data)  
3,341 測試資料(Testing data)

使用SEER資料進行外部驗證  
12,873 位白人資料(White)  
2,056 位黑人資料(Black)  
1,061 亞洲人資料(Asian)

### 辨別力 (Discrimination)

使用ROC曲線下面積 (AUCs) 做為模型鑑別力的指標

### 校準 (Calibration)

診斷後X年，觀察死亡人數跟預測死亡人數的差異

## 研究結果

	TCR data	Asian in SEER	Black in SEER	White in SEER
Age	62.6 (10.84)	64.4 (10.51)	61.7 (10.81)	64.3 (10.81)
Tumor size, mm	30.5 (10.93)	28.5 (11.88)	29.2 (12.07)	28.3 (12.22)
Sex				
Male	5,284 (0.53)	499 (0.47)	933 (0.45)	6,574 (0.51)
Female	4,740 (0.47)	562 (0.53)	1,123 (0.55)	6,299 (0.49)
Pathological stage				
1	3,337 (0.333)	365 (0.344)	741 (0.36)	5,038 (0.39)
2	3,483 (0.347)	370 (0.349)	672 (0.327)	4,341 (0.34)
3	3,204 (0.32)	326 (0.307)	643 (0.313)	3,494 (0.27)
Chemotherapy				
without	5,273 (0.53)	747 (0.7)	1,483 (0.72)	9,413 (0.73)
with	4,751 (0.47)	314 (0.3)	573 (0.28)	3,460 (0.27)
Grade				
1	903 (0.09)	93 (0.09)	224 (0.11)	1,494 (0.12)
2	8,657 (0.86)	885 (0.83)	1,626 (0.79)	9,810 (0.76)
3	464 (0.05)	83 (0.08)	206 (0.10)	1,569 (0.12)
Primary site				
Cecum	713 (0.07)	156 (0.15)	460 (0.22)	2,679 (0.21)
Ascending colon (right colon)	1,519 (0.15)	161 (0.15)	455 (0.22)	2,587 (0.2)
Transverse colon	1,099 (0.11)	83 (0.08)	212 (0.10)	1,231 (0.10)
Descending colon (left colon)	1,097 (0.11)	95 (0.09)	196 (0.1)	779 (0.06)
Sigmoid colon	4,772 (0.48)	456 (0.43)	503 (0.25)	4,157 (0.32)

Table.1 Baseline Characteristics of Colon Cancer Patients

不同資料經過排除條件後所呈現的基準線極為相似。

Colon cancer-specific survival								
	Calibration year	Observed	Predicted	<i>P</i> value	Calibration year	Observed	Predicted	<i>P</i> value
Training data	1	78	78.6	0.946	6	375	372.1	0.879
	2	178	181.5	0.794	7	326	329.1	0.864
	3	282	286.7	0.782	8	248	253.8	0.718
	4	347	359.2	0.519	9	170	163.6	0.618
	5	377	388.1	0.572	—	—	—	—
Testing data	1	42	40.2	0.781	6	174	187.7	0.316
	2	91	92.8	0.849	7	154	163.5	0.456
	3	146	148.2	0.858	8	124	125.5	0.891
	4	165	184.4	0.153	9	79	85.5	0.485
	5	184	196.7	0.366	—	—	—	—
White	1	200	154.0	0.0002	6	661	789.7	<0.0001
	2	357	324.3	0.0697	7	630	753.9	<0.0001
	3	521	529.6	0.7087	8	565	665.2	0.0001
	4	611	679.0	0.0090	9	430	521.4	0.0001
	5	673	767.5	0.0006	—	—	—	—
Black	1	33	23.5	0.0503	6	152	115.5	0.0007
	2	82	48.8	<0.0001	7	143	110.5	0.0020
	3	123	78.8	<0.0001	8	124	96.7	0.0056
	4	150	101.8	<0.0001	9	94	73.1	0.0147
	5	160	112.8	<0.0001	—	—	—	—
Asian	1	12	13.3	0.724	6	60	69.9	0.234
	2	28	28.1	0.983	7	53	66.7	0.094
	3	42	46.0	0.553	8	47	57.4	0.168
	4	52	60.6	0.267	9	34	45.3	0.092
	5	55	67.3	0.132	—	—	—	—

Table 2. Model Calibration of Colon Cancer-specific Mortality

	Colon cancer-specific survival	對於內部驗證，訓練和測試資料都得到良好的校準與AUC。外部驗證方面雖然AUC在所有種族中表現良好，但非亞洲組缺乏模型校準顯示預測死亡人數跟觀察死亡人數有顯著不同，對結腸癌生存預測有顯著差異。
Variables in SEER data	Training	0.709
	Testing	0.713
	White	0.726
	Black	0.683
	Asian	0.750

Table 3. Comparison of Discrimination (AUCs)

## 結論

我們開發了一種開創性的預後模型，特別有利於亞洲的結腸癌患者  
本研究希望能夠成為相關研究的重要參考，並期望作為臨床診斷中幫助醫病共享決策的重要依據。

## 參考資料

- Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, Caldas C, Pharoah PD: PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res 2010, 12(1):R1.
- Usher-Smith, J. A., et al. (2016). "Risk Prediction Models for Colorectal Cancer: A Systematic Review." Cancer Prev Res (Phila) 9(1): 13-26.
- Altman DG, Vergouwe Y, Royston P, Moons KG: Prognosis and prognostic research: validating a prognostic model. Bmj 2009, 338:b605.
- Fedewa, S. A., et al. (2017). "Racial and Ethnic Disparities in Interval Colorectal Cancer Incidence: A Population-Based Cohort Study." Ann Intern Med 166(12): 857-866.