

肺癌主題式資料庫使用手冊

一、檔案內容說明

檔案代號	DATA:資料名稱 Y1:資料年度	資料筆數	肺癌研究主檔(1979-2022)： 387,472 筆
中文檔名	肺癌主題式資料庫	英文檔名	Lung Cancer Health Database
屬性	歷史檔	週期	年
檔案大小	197GB	譯碼簿 更新日期	2025/12/30
檔案列表 及其年 度、欄位 數	<p>本資料庫包含以下 14 個檔：</p> <ol style="list-style-type: none"> 1. 肺癌門急診明細檔 (2000-2023)：51 2. 肺癌住院明細檔 (2000-2023)：80 3. 肺癌藥局明細檔 (2000-2023)：44 4. 肺癌門急診醫令檔 (2000-2023)：21 5. 肺癌住院醫令檔 (2000-2023)：22 6. 肺癌藥局醫令檔 (2000-2023)：17 7. 肺癌承保檔 (2000-2023)：30 8. 肺癌重大傷病檔 (2001-2023)：27 9. 肺癌死因統計合併檔 (1985-2023)：18 (合併 1985-2023 年的死因統計檔) 10. 肺癌癌登檔 LF (2007-2022)：118 11. 肺癌癌登檔 SF (2007-2022)：42 12. 肺癌癌登檔 TCDB (2002-2006)：65 13. 肺癌癌登年報檔 (1979-2022)：13 14. 肺癌研究主檔 (1979-2022)：51 		
資料描述	<p>肺癌主題式資料庫包含了收錄在 1979-2022 年癌症登記年報檔 (H_BHP_CRSSC) 中之每一肺癌患者在癌症登記檔(包括長表 LF、短表 SF 和 TCDB)及健保檔、死因統計合併檔，共 14 個資料檔之所有資訊。詳細定義如下：</p> <ol style="list-style-type: none"> 1. 定義肺癌個案： 肺癌個案為癌症登記年報檔(H_BHP_CRSSC)1979-2022 年(經 IDNO 清理後)中，曾有一筆其原發部位(CASITE)的前三碼為 162 		

	<p>(癌症登記年報檔均依據 ICD-9 譯碼)，但其組織類型與性態碼 (HISTOLOG) 不為血液腫瘤等(詳見欄位(變項)譯碼(一)癌症類別 (newca2)對照表)。</p> <ol style="list-style-type: none"> 2. 癌症登記年報檔及死因統計檔均經 ID、性別、出生年及死亡年月加以串接清理，建立了癌症研究主檔及死因統計合併檔。其中並依據 SEER 之死因分類法，重新定義癌症病人之死因；以及利用串接健保檔提供了癌症病人罹癌前之共病資訊及罹癌後之存活狀態。 3. 以肺癌個案的 IDNO 為鍵值，串聯全民健保處方及治療明細檔_門急診 (H_NHI_OPDTE)、全民健保處方及治療明細檔_住院 (H_NHI_IPDTE)、全民健保處方及治療明細檔_藥局 (H_NHI_DRUGE)、全民健保處方及治療醫令明細檔_門急診 (H_NHI_OPDTO)、全民健保處方及治療醫令明細檔_住院 (H_NHI_IPDTO)、全民健保處方及治療醫令明細檔_藥局 (H_NHI_DRUGO)、全民健保承保檔 (H_NHI_ENROL)、全民健保重大傷病檔 (H_NHI_CATAS)、癌症登記檔長表 LF (H_BHP_CRFLF)、癌症登記檔短表 SF (H_BHP_CRFSF)、癌症登記檔 TCDB (H_BHP_CRSLF)、癌症登記年報檔 (H_BHP_CRSSC)、癌症研究主檔、死因統計合併檔，共 14 個資料檔。 4. 上述經由 ID 清理，並從其中擷取出肺癌個案的資料檔，說明如下： <ol style="list-style-type: none"> (1) 肺癌研究主檔：從癌症研究主檔中，擷取出肺癌個案的所有紀錄。 (2) 肺癌死因統計合併檔：從死因統計合併檔中，擷取出肺癌個案。
注意事項	<ol style="list-style-type: none"> 1. 肺癌患者若為多發癌症(一 IDNO 有多筆癌症年報檔的紀錄)的個案，除了肺癌的紀錄，肺癌研究主檔也包含了該個案在癌登檔中的所有罹癌紀錄。例如：個案有一筆肺癌紀錄，和一筆其他癌症(如乳癌或血癌等)紀錄，則兩筆紀錄皆會收錄。又例如，若要研究第一筆癌症診斷紀錄為肺癌的個案，建議從肺癌研究主檔中，以癌症類別(newca2)為 a1620，且記錄順序(rank_record)≤1，以及紀錄檢誤(accuracy_record)=1 篩選。 2. 申請本資料庫仍應考量所需研究之主題特性，應用時應注意資料使用限制，歡迎回饋使用意見。 <p>※本中心資料庫業經去識別化處理，刪除姓名、地址等可直接識別欄位，如有身分證字號、院所代碼、投保單位代號、統一編號、出生日期、就醫日期、入院日期等欄位均經屏蔽處理，且相關資料僅能於本部設置之獨立作業區內使用，並攜出經審核通過之統計結果，以確保資料應用及其揭露方式已無從識別特定當事人。</p>
主鍵與比	<ol style="list-style-type: none"> 1. 身分證字號(IDNO)、醫師身分證字號(PRSN_ID)及醫療機構代碼

對欄位	<p>(HOSP_ID)已經加密處理。</p> <ol style="list-style-type: none">個人歸戶及個人資料對照的鍵值：IDNO、ID(死因統計檔的個人資料對照之鍵值)。醫療機構歸戶及醫療機構資料對照的鍵值：HOSP_ID。
-----	--

二、檔案格式及資料描述

(一) 以肺癌個案的 IDNO 為鍵值，串聯健保資料庫、癌症登記檔之紀錄如

下，各資料檔的詳細欄位格式請參閱各資料庫使用手冊：

1. 全民健保處方及治療明細檔_門急診(H_NHI_OPDTE)
2. 全民健保處方及治療明細檔_住院(H_NHI_IPDTE)
3. 全民健保處方及治療明細檔_藥局(H_NHI_DRUGE)
4. 全民健保處方及治療醫令明細檔_門急診(H_NHI_OPDTO)
5. 全民健保處方及治療醫令明細檔_住院(H_NHI_IPDTO)
6. 全民健保處方及治療醫令明細檔_藥局(H_NHI_DRUGO)
7. 全民健保承保檔(H_NHI_ENROL)
8. 全民健保重大傷病檔(H_NHI_CATAS)
9. 癌症登記檔長表 LF (H_BHP_CRFLF)
10. 癌症登記檔短表 SF (H_BHP_CRFSF)
11. 癌症登記檔 TCDB (H_BHP_CRSLF)
12. 癌症登記年報檔(H_BHP_CRSSC)

(二) 肺癌死因統計合併檔之欄位格式請參照死因統計檔(H_OST_DEATH)。

(三) 肺癌研究主檔之欄位格式如下：

序號	欄位	型態	來源	內容	描述
1	IDNO	Char	癌登 年報 檔	身分證字 號	已加密轉碼
2	SEX	Char	癌登 年報 檔	原始檔性 別	1：男性，2：女性，4：變 性人
3	BIRTH_Y	Char	癌登 年報 檔	出生年	YYYY
4	DIAGDAY	Char	癌登	診斷日期	YYYYMMDD

序號	欄位	型態	來源	內容	描述
			年報 檔		
5	HISTOLOG	Char	癌登 年報 檔	組織類型 與性態碼	組織病理：用 ICD-O 譯碼 第 5 碼為 2：原位癌 第 5 碼為 3：侵襲癌
6	CASITE	Char	癌登 年報 檔	原發部位	第 1 碼為 H 者：表示為血液腫瘤個案 其他第 1 碼非 H 者：詳見原發部位(CASITE)對照表 血液腫瘤個案若要再細分 NHL(非何杰金氏淋巴瘤)、HL(何杰金氏淋巴瘤)及 LEU(白血病)等血液腫瘤個案，在此欄位無法辨識，請用組織類型與性態碼 (HISTOLOG) 區分。若是要分析各癌症部分(CASITE 欄位)，也請先把 NHL、HL、LEU 等血液腫瘤排除後再計算。
7	CURECOND	Char	癌登 年報 檔	治療情形	合併個案所有申報醫院治療狀況 0：未治療 1：手術治療 2：放射線治療 3：化學藥物治療 6：免疫治療， 7：內分泌（荷爾蒙）治療 8：支持治療 9：中藥（中醫療法） B：骨髓/幹細胞移植 O：其他治療 P：局部性化學治療 I：局部免疫治療 T：標靶治療 X：其他(包含局部肝動脈栓塞治療 TACE)，或為新的

序號	欄位	型態	來源	內容	描述
					治療方法，非指他院治療
8	AGE1	Num	癌登年報檔	診斷年齡	
9	RESIDCD	Char	癌登年報檔	戶籍地代碼	詳戶籍地代碼(RESIDCD)對照表 100年起開始使用新五都代碼
10	DIAGBASE	Char	癌登年報檔	癌症確診方式	0：臨床病狀 1：實驗診斷 2：影像檢查 3：內視鏡 4：探查手術 5：細胞診斷 6：血液病理 7：組織病理 8：屍體解剖 9：診斷未明
11	GRADE	Char	癌登年報檔	分級/分化	2002年起收錄 1：Well differentiated Differentiated, NOS 2：Moderately differentiated Moderately well differentiated Intermediate differentiated 3：Poorly differentiated 4：Undifferentiated Anaplastic 9：Grade or differentiation not determined, not stated or not applicable 注意：因各癌別腫瘤分級/分化程度有二、三、四分法，詳細說明請參考100年版之前的台灣癌症登記長短表摘錄手冊內容，可逕至

序號	欄位	型態	來源	內容	描述
					<p>台灣癌症登記中心網站[癌症登記申報相關參考資料]處下載，網址如下： http://tcr.cph.ntu.edu.tw/main.php?Page=A6</p> <p>For Lymphomas and Leukemias 5：T cell 6：B cell；pre-B；B-precursor 7：Null cell；non T-non B 8：NK (natural killer) cell 9：Cell type not determined, not stated or not applicable</p> <p>For Ovary borderline tumor B：Ovary borderline tumor (Low malignant potential)合併 intraepithelial carcinoma 或/及 microinvasion</p> <p>For Colon-Rectum tumor B：Colon-Rectum High grade dysplasia (severe dysplasia)</p> <p>2018 年起，新增編碼 A/B/C/D/E/H/L/M/S/X (詳見台灣癌症登記中心網站[癌症登記申報相關參考資料]_107 年版癌症登記摘錄手冊短表 45 欄位與長表 115 欄位內容)</p>
12	IDNO_S	Char	癌登年報檔	性別	1：男性 2：女性 9：不詳
13	IDNO_ROC	Char	癌登年報	身分證字號檢誤	0：身分證字號符合編碼原則

序號	欄位	型態	來源	內容	描述
			檔		1：不符合
14	Death_ID	Char	死因檔	身分證字號	(無死因資料者，此欄空白)
15	Death_SEX	Char	死因檔	性別	1：男性，2：女性 (無死因資料者，此欄空白)
16	Death_BIRTH_YM	Char	死因檔	出生年月	西元年月 YYYYMM (無死因資料者，此欄空白)
17	Death_D_DATE	Char	死因檔	死亡日期	西元 YYYYMMDD (無死因資料者，此欄空白)
18	Death_D_CODE	Char	死因檔	死因分類	70-96 年為 ICD9 編碼 (無死因資料者，此欄空白)
19	Death_ICD10	Char	死因檔	死因分類	97 年以後死因以 ICD10 編碼 (無死因資料者，此欄空白)
20	newca2	Char	本檔自訂	癌症類別	詳欄位(變項)譯碼(一)癌症類別
21	rank_record	Num	本檔自訂	紀錄順序	單一癌症者，此欄為 0；多發者為順序。
22	n_record	Num	本檔自訂	紀錄次數	1：單一；2 以上為多發
23	rank_ca_record	Num	本檔自訂	癌症類別內的診斷順序	各癌症類別(newca2)內的順序，單一癌症者，此欄為 0；多發者為順序。
24	n_cancer	Num	本檔自訂	癌症類別次數	依 newca2 統計每一 IDNO 的癌症類別次數。
25	rank_newca2	Num	本檔自訂	癌症類別的診斷順序	癌症類別(newca2)之間的順序，單一癌症者此欄為 0；多發者為順序。
26	new_d_code	Char	本檔自訂	SEER 死因	依 SEER 死因分類所得新死因，若是死於其所得的癌症，此欄與 newca2 相同；若非死於所得的癌症，則與死因檔之死因分類(Death_D_CODE 或 Death_ICD10)相同。詳欄位(變項)譯碼(二)SEER 死因。

序號	欄位	型態	來源	內容	描述
27	cs_event	Num	本檔自訂	是否為 cause-specific survival event	0：死於其他原因或仍存活； 1：死於其所得到的第一個癌症； 9：未知(沒有串到死因檔、門急診、住院、承保檔)； 空白表示不是第一個癌症。 詳欄位(變項)譯碼(二)SEER死因。
28	os_event	Num	本檔自訂	是否為 overall survival event	0：存活(無死因紀錄)； 1：死亡(有死因紀錄)； 9：未知(沒有串到死因檔、門急診、住院、承保檔)。
29	accuracy_record	Num	本檔自訂	紀錄檢誤	0：不符合；1：符合。詳欄位(變項)譯碼(三)紀錄檢誤。
30	adj_diag_dt	Char	本檔自訂	填補後的癌症診斷日期	當診斷日期(DIAGDAY)的日有 missing 的情況，考慮以死亡日期(Death_D_DATE)或隨機數字填補的方式，以便與追蹤日期計算出存活時間。詳欄位(變項)譯碼(四)填補後的癌症診斷日期。
31	flw_up_date	Char	本檔自訂	最後追蹤時間	有串到死因檔的病患為死亡時間(Death_D_DATE)；沒有串到死因檔的病患為門急診、住院、承保檔最後一次紀錄的日期。詳欄位(變項)譯碼(五)最後追蹤時間。
32	flw_up_type	Char	本檔自訂	最後追蹤日期的資料庫	最後追蹤時間(flw_up_date)來自哪一個資料庫。 death：死因檔； opdte：門急診； ipdte：住院； enrol：承保檔。
33	flw_up_day	Num	本檔自訂	存活時間(單位：天)	從 adj_diag_dt 到 flw_up_date 的天數。 - 9：不符合紀錄檢誤的情

序號	欄位	型態	來源	內容	描述
					況。詳欄位(變項)譯碼(四)填補後的癌症診斷日期和(五)最後追蹤時間。
34	ComD1_AMI	Num	本檔自訂	是否有 Acute Myocardial Infraction (AMI)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
35	ComD2_OMI	Num	本檔自訂	是否有 Old Myocardial Infraction (Old MI)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
36	ComD3_CHF	Num	本檔自訂	是否有 Congestive heart failure (CHF)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
37	ComD4_PVD	Num	本檔自訂	是否有 Peripheral vascular disease (PVD)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
38	ComD5_CVD	Num	本檔自訂	是否有 Cerebro-vascular disease (CVD)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
39	ComD6_COPD	Num	本檔自訂	是否有 Chronic obstructive	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個

序號	欄位	型態	來源	內容	描述
				pulmonary disease (COPD)	侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
40	ComD7_Demen	Num	本檔自訂	是否有 Dementia	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
41	ComD8_Paraly	Num	本檔自訂	是否有 Paralysis	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
42	ComD9_DMWOC	Num	本檔自訂	是否有 Diabetes without chronic complication (DM W/O CC)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
43	ComD10_DMC	Num	本檔自訂	是否有 Diabetes with chronic complication (DM W CC)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。
44	ComD11_CRF	Num	本檔自訂	是否有 Chronic renal failure (renal	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個侵襲性癌症，詳欄位(變項)譯碼(六)共病症。

序號	欄位	型態	來源	內容	描述
				disease) (CRF)	
45	ComD12_MLD	Num	本檔 自訂	是否有 Cirrhosis/ chronic hepatitis (Mild liver disease) (HBV, HCV, Mild LD)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個 侵襲性癌症，詳欄位(變項) 譯碼(六)共病症。
46	ComD13_MSLD	Num	本檔 自訂	是否有 Moderate -Server Liver Disease (M S LD)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個 侵襲性癌症，詳欄位(變項) 譯碼(六)共病症。
47	ComD14_Ulcer	Num	本檔 自訂	是否有 Ulcer (Peptic ulcer disease)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個 侵襲性癌症，詳欄位(變項) 譯碼(六)共病症。
48	ComD15_RD	Num	本檔 自訂	是否有 Rheumatic disease (RD)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個 侵襲性癌症，詳欄位(變項) 譯碼(六)共病症。
49	ComD16_AIDS	Num	本檔 自訂	是否為 AIDS	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個 侵襲性癌症，詳欄位(變項) 譯碼(六)共病症。
50	ComD17_HTU	Num	本檔 自訂	是否有 Hyperten-	0：沒有； 1：有；

序號	欄位	型態	來源	內容	描述
				sion, uncompli cated (HT UC)	-9：未知； 空白表示不是該病患第一個 侵襲性癌症，詳欄位(變項) 譯碼(六)共病症。
51	ComD18_HTC	Num	本檔 自訂	是否有 Hyperten- sion, compli- cated (HT C)	0：沒有； 1：有； -9：未知； 空白表示不是該病患第一個 侵襲性癌症，詳欄位(變項) 譯碼(六)共病症。

三、欄位(變項)譯碼

(一) 癌症類別(newca2)對照表

※ 癌登年報檔的原發部位(CASITE)目前均依據 ICD-9 譯碼。

† Histology code 參照 SEER Site Recode B ICD-O-3 (12/19/2003)

Definition，網址：

https://seer.cancer.gov/siterecode_b/icdo3_d12192003/

Site	ICD-9 cancer site / histology	newca2
Histology code ≠ 9050-9055,9140,9590-9993[†]		
Oral Cavity and Pharynx	140-149	a1401
Esophagus	150	a1500
Stomach	151	a1510
Small Intestine	152	a1520
Colon and Rectum	153,154.0-154.1,159.0	a1531
Anus, Anal Canal and Anorectum	154.2-154.3,154.8	a1541
Liver and Intrahepatic Bile Duct	155	a1550
Gallbladder	156.0	a1561
Other Biliary	156.1,156.2,156.8,156.9	a1562
Pancreas	157	a1570
Retroperitoneum	158.0	a1581
Peritoneum, Omentum and Mesentery	158.8,158.9	a1582
Other Digestive Organs	159.8	a1591
Nose, Nasal Cavity and Middle Ear	160	a1600
Larynx	161	a1610
Lung and Bronchus	162	a1620
Pleura	163	a1630
Trachea, Mediastinum and Other Respiratory Organs	164.2-164.3,164.8-164.9,165.0	a1641
Bones and Joints	170	a1700
Soft Tissue including Heart	171,164.1	a1711
Melanoma of the Skin	173 and Histology= 8720-8970	a1721
Other Non-Epithelial Skin	173 and Histology≠8000-8005,8010-8046,8050-8084,	a1730

Site	ICD-9 cancer site / histology	newca2
	8090-8110,8720-8790	
Breast	174,175	a1741
Cervix Uteri	180	a1800
Corpus ans Uterus, NOS	179, 182	a1821
Ovary	183	a1830
Vagina	184.0	a1841
Vulva	184.1-184.4	a1842
Other Female Genital Organs	184.8-184.9	a1843
Prostate	185	a1850
Testis	186	a1860
Penis	187.1-187.4	a1871
Other Male Genital Organs	187.5-187.9	a1872
Urinary Bladder	188	a1880
Kidney and Renal Pelvis	189.0-189.1	a1891
Ureter	189.2	a1892
Other Urinary Organs	189.3-189.4,189.8-189.9	a1893
Eye and Orbit	190	a1900
Brain and Other Nervous System	191-192	a1911
Thyroid	193	a1930
Other Endocrine including Thymus	194,164.0	a1941
Other	other	a9999
Histology code = 9050-9055,9140,9590-9993[†]		
Hodgkin Lymphoma	96503,96513,96523,96533,96593,96613,96623,96633,96643,96653,96673	a3001
Non-Hodgkin Lymphoma	96893,96993,96713,97613,96753,96903,96913,96953,96983,96733,96873,96263,96803,96843,96793,95963,96703,96783,97283,98333,99403,97193,97003,97013,97093,97183,97173,97163,97083,97023,97053,97143,98273,97293,98343,99483	a3002
Leukemia	98663,98713,98963,98973,98953,99203,98613,98403,98673,98703,98723,98733,98743,98913,99103,99313,99303,97273,98353,98363,98373,98633,98753,98763,	a3003

Site	ICD-9 cancer site / histology	newca2
	99453,99463,99633,99643,98233,98013, 97403,97413,99503,99603,99613,99623, 99753,99803,99823,99833,99843,99853, 99863,99873,99893,98053,98203,98323, 98603,98003	
Myeloma	9731-9732,9734	a3004
Mesothelioma	9050-9055	a3005
Kaposi Sarcoma	9140	a1760
Miscellaneous	Cancer site=195,196,199 & Histology=9740-9741,9750- 9769,9950,9960-9962,9965-9967,9970- 9971,9975,9980,9982-9987,9989,9991- 9992	a3006
Others	others	a3999

(二) SEER 死因

1. 根據癌症登記報告之前言，癌症資料已參照國際癌症研究總署 (International Agency for Research on Cancer, IARC) 之多重原發性腫瘤分類標準，進行歸類與計算。以下「多發癌症」之定義為，一 IDNO 有多筆癌症年報檔的紀錄。
2. 由於 SEER 的死因分類方法是根據癌症類別(newca2)和死因檔的死因分類(D_CODE 或 ICD10)，判斷癌症病患是否死於其所得到的第一個癌症 (cause-specific)，且單發癌症和多發癌症在死因分類的方法上，依癌症而有所不同(詳細請參照下列 5. 網站)。以下說明中的死因，指死因檔中的死因分類(D_CODE 或 ICD10)。
 - (1) 針對單發癌症病患的死因分類說明，以下各種情形都將死因歸於該單發癌症：
 - i. 死因為癌症，且與該單發癌症在同一癌症部位(cancer site)。
 - ii. 死因為癌症，且與該單發癌症在同一器官系統(general organ system as specified by ICD-O-3)。
 - iii. 死因為其他惡性腫瘤(malignant cancers)。
 - iv. 死因為伴有癌症的 AIDS，或 HIV 相關，或與該單發癌症有關之疾病。
 - (2) 如研究目標為多發癌症的第一個癌症，針對多發癌症病患的第一個癌症死因分類說明，以下各種情形都將死因歸於該多發癌症的第一個癌症：

- i. 死因為癌症，且與第一個癌症在同一癌症部位(cancer site)。
 - ii. 死因為癌症，且與第一個癌症在同一器官系統(general organ system of the study site)。
 - iii. 死因為未知原發的多發癌症(in multiple cancers with unknown primary)。
 - iv. 死因為伴有癌症的 AIDS，或 HIV 相關，或與第一個癌症部位相關的非癌症疾病(Noncancer disease related to the site of first cancer)。
- (3) 死因為其他的惡性腫瘤(all other malignant cancers)，則針對第一個原發癌的存活分析而言，其他惡性腫瘤訂為設限事件(censored events)。
3. 根據上述定義，建立欄位「是否為 cause-specific survival event(cs_event)」，cs_event=0 表示死於其他原因或仍存活；cs_event=1 表示死於其所得到的第一個癌症；cs_event=9 表示未知(沒有串到死因檔、門急診、住院、承保檔)；cs_event 為空白，表示不是第一個癌症。
 4. 參照文獻：Howlader N, Ries LA, Mariotto AB, Reichman ME, Ruhl J, Cronin KA. Improved estimates of cancer-specific survival rates from population-based data. *J Natl Cancer Inst.* 2010;102(20):1584-1598. doi:10.1093/jnci/djq366
 5. SEER cause-specific 死因分類法為依據下述的網站做分類：
<https://seer.cancer.gov/causespecific/> (版本日期：November 2010)

(三) 紀錄檢誤

為判斷該筆紀錄的資訊是否合理，當 accuracy_record=1 時，表示符合以下 4 個檢誤項目；當 accuracy_record=0 時，表示其中至少 1 個檢誤項目不符合。檢誤項目如下：

1. 癌症登記年報檔的出生年(BIRTH_Y)與死因統計檔的出生年相差正負三年內的資料(限有死因資料者)
2. 年份合理性(出生年 \leq 診斷年月日(DIAGDAY) \leq 死亡年月日(D_DATE))(限有死因資料者)
3. 癌症登記年報檔的出生年(BIRTH_Y) \neq 9999
4. 癌症登記年報檔的癌症確診方式(DIAGBASE)不為「屍體解剖」

(四) 填補後的癌症診斷日期

為了與最後追蹤日期(flw_up_date)計算出存活時間，依據下列狀態填補年報檔的診斷日期(DIAGDAY)的月份和日期為 99(不詳)的情況，建立欄位「填補

後的癌症診斷日期(adj_diag_dt)」，填補方法如下：

1. 當診斷日期(DIAGDAY)有確切的月日時，adj_diag_dt=診斷日期(DIAGDAY)；
2. 診斷日期(DIAGDAY)只在日期有 missing，且與死亡日期(Death_D_DATE)同年同月時，日以小於死亡日期的日之數字隨機補上；若與死亡日期(Death_D_DATE)不為同年同月時，日以1、2、...、28之間的數字隨機補上；
3. 診斷日期(DIAGDAY)只在日期有 missing，但沒有死亡日期(Death_D_DATE)時，日以1、2、...、28之間的數字隨機補上；
4. 診斷日期(DIAGDAY)的月份有 missing 時，則維持 adj_diag_dt 的月份為 99，不填補。

(五) 最後追蹤時間

1. 定義每個 ID 在健保檔中最後一個追蹤日期如下：
 - (1) 全民健保處方及治療明細檔_門急診(西醫、牙醫、中醫)：
 - 資料期間：2000-2023 年
 - 以每個 ID 最後一筆就醫日期(FUNC_DATE)為其追蹤日期。
 - (2) 全民健保處方及治療明細檔_住院：
 - 資料期間：2000-2023 年
 - 使用出院日期(OUT_DATE)作為追蹤日期，若出院日期(OUT_DATE)為空白，則使用申報期間-迄(APPL_END_DATE)取代，最後找出每個 ID 最後一筆出院日期(OUT_DATE)或申報期間-迄(APPL_END_DATE)。
 - (3) 全民健保承保檔：
 - 資料期間：2000-2023 年
 - 使用投保狀態(ID_STATUS)為「在保」的紀錄，找出每個 ID 最後一筆保費計費年月(PREM_YM)，為其追蹤日期。
2. 對每一癌症病患 IDNO，使用上述三個健保檔和清理後的死因檔，建立欄位「最後追蹤日期(flw_up_date)」，定義如下：
 - (1) 如果這個 IDNO 有出現在清理後的死因檔，則使用其死亡日期(Death_D_DATE)作為追蹤日期；
 - (2) 如果這個 IDNO 沒有出現在清理後的死因檔，則選擇上述三個健保檔中最後一個紀錄的日期，作為該 IDNO 的最後追蹤日期；若為承保檔的保費計費年月(PREM_YM)，則調整追蹤日期為該月份的最後一日；
 - (3) 如沒有出現在清理後的死因檔，也沒有在上述三個健保檔找到紀錄，則維持空白，表示不知是否存活。

(六) 共病症

1. 目標：根據病患在年報檔的第一筆侵襲性癌症診斷紀錄，於診斷日期(DIAGDAY)前第二個半年~第五個半年，判斷在這個期間中是否患有共病症(comorbidity)。考慮了 18 種共病症，詳欄位(變項)譯碼(七)、(八)共病症(comD1-18)之定義。(舉例：假如病患在 2015 年 03 月 23 日診斷出第一個侵襲性癌症是肺癌，則 2015 年 03 月 23 日至 2014 年 09 月 24 日為第一個半年，2014 年 09 月 23 日至 2014 年 03 月 24 日為第二個半年，以此類推，2013 年 03 月 23 日至 2012 年 09 月 24 日為第五個半年。)，詳細請參閱論文 Chien LH, Tseng TJ, Chen TY, Chen CH, Chen CY, et al. Prevalence of comorbidities and their impact on survival among older adults with the five most common cancers in Taiwan: a population study. Sci Rep. 2023;13(1):6727. Epub 2023/05/15. doi: 10.1038/s41598-023-29582-0. PubMed PMID: 37185775; PubMed Central PMCID: PMCPCMC10130096.
2. 所需要的健保檔：
 - (1) 全民健保處方及治療明細檔_門急診(西醫、牙醫、中醫)：
 - 資料期間：2000-2022 年
 - 使用欄位：就醫日期(FUNC_DATE_OLD)、診斷欄位(國際疾病分類號一(ICD9CM_CODE)、國際疾病分類號二(ICD9CM_CODE1)、國際疾病分類號三(ICD9CM_CODE2)、主手術代碼(OP_CODE)、次手術代碼(OP_CODE1)。
 - (2) 全民健保處方及治療明細檔_住院：
 - 資料期間：2000-2022 年
 - 使用欄位：入院日期(IN_DATE_OLD)、診斷欄位(主診斷代碼(ICD9CM_CODE)、次診斷代碼一(ICD9CM_CODE1)、次診斷代碼二(ICD9CM_CODE2)、次診斷代碼三(ICD9CM_CODE3)、次診斷代碼四(ICD9CM_CODE4)、主手術(處置)(OP_CODE)、次手術(處置)一(OP_CODE1)、次手術(處置)二(OP_CODE2)、次手術(處置)三(OP_CODE3)、次手術(處置)四(OP_CODE4)。
 - (3) 全民健保承保檔：
 - 資料期間：2000-2022 年
 - 使用欄位：保費計費年月(PREM_YM)。
3. 步驟：
 - (1) 按加密後的個人身分證字號(IDNO)將癌登檔診斷日期(DIAGDAY)、原發部位(CASITE)串接整理全民健保處方及治療明細檔_門急診(西醫、牙醫、中醫)、全民健保處方及治療明細檔_住院、全民健保承保檔，判斷欄位就醫日期、入院日期及保費計費年月在診斷日期

(DIAGDAY) 前第幾個半年，保留前第二個半年~第五個半年期間內診斷紀錄以診斷欄位判斷是否有這 18 種共病症。

- (2) 延續步驟 1.整理各共病症在癌症診斷日前最早日期(以已知的就醫日期和入院日期中挑選出最早日期)、最晚日期(以已知的就醫日期和入院日期中挑選出最晚日期)、以及在門診檔跟住院檔中各共有幾筆紀錄。
- (3) 新增 comDN，N=1~18 分別代表 18 項共病症的有無。(舉例：comD5_CVD=0 為沒有 Cerebrovascular disease 共病症；comD5_CVD=1 為有 Cerebrovascular disease 共病症；comD5_CVD=-9 為無法判斷。) 每個共病症以下列步驟各自判斷：
 - i. 在癌症診斷日前第二個半年~第五個半年有共病症住院紀錄，則 comDN=1。無共病症住院紀錄但有共病症門急診紀錄，若只有一筆紀錄或是兩筆以上就醫紀錄間隔不超過一個月，則 comDN=0；反之有兩筆以上就醫紀錄且間隔超過一個月，則 comDN=1。
 - ii. 若在癌症診斷日前第二個半年~第五個半年無門急診及住院的共病症紀錄，進一步從門急診及住院於診斷日期往前看之前是否有任何就醫紀錄，有就醫紀錄則 comDN=0，無就醫紀錄則再看承保檔在癌症診斷日期前第二個半年~第五個半年是否有在保，有在保則 comDN=0，無在保則 comDN=-9。

(七) Definitions of comorbidities+ ICD-9-CM

No.	Disease	ICD-9-CM	癌症主檔之欄位名稱
1	Acute Myocardial Infraction (AMI)	410.x	ComD1_AMI
2	Old Myocardial Infraction (Old MI)	412.x	ComD2_OMI
3	Congestive heart failure (CHF)	428.x	ComD3_CHF
4	Peripheral vascular disease (PVD)	441.x,443.9,785.4,V43.4, Procedure 38.48	ComD4_PVD
5	Cerebrovascular disease (CVD)	430.x-438.x	ComD5_CVD
6	Chronic obstructive pulmonary disease (COPD)	490.x-505x,506.4	ComD6_COPD

No.	Disease	ICD-9-CM	癌症主檔之欄位名稱
7	Dementia	290.x	ComD7_Demen
8	Paralysis	342.x,344.1	ComD8_Paraly
9	Diabetes without chronic complication (DM W/O CC)	250.0-250.3,250.7	ComD9_DMWOC
10	Diabetes with chronic complication (DM W CC)	250.4-250.6,250.8-250.9	ComD10_DMC
11	Chronic renal failure (renal disease) (CRF)	582.x, 583.x, 585.x, 586.x, 588.x	ComD11_CRF
12*	Cirrhosis/chronic hepatitis (Mild liver disease) (HBV, HCV, Mild LD)	571.2,571.4,571.5,571.6 070.2-070.7(excluded 070.41,070.51),070.9, V026.1	ComD12_MLD
13	Moderate-Severe Liver Disease (M S LD)	572.2-572.8,456.0-456.21	ComD13_MSLD
14	Ulcer (Peptic ulcer disease)	531.x-534.x	ComD14_Ulcer
15	Rheumatic disease (RD)	710.0,710.1,710.4,714.0- 714.2, 714.81, 725.x	ComD15_RD
16	AIDS	042.x-044x	ComD16_AIDS
17**	Hypertension, uncomplicated (HT UC)	401.1, 401.9, 642.0	ComD17_HTU
18**	Hypertension, complicated (HT C)	401.0, 402.x-405.x, 642.1, 642.2, 642.7, 642.9	ComD18_HTC

+Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-8-CM administrative databases. J Clin Epidemiol.1992; 45:613-9

Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining Comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005 Nov; 43(11): 1130-9.

*The definition of the Cirrhosis/chronic hepatitis (Mild liver disease) comorbidity in this study includes viral hepatitis B and viral hepatitis C, which differs from the one that used by Charlson in 1987.

**Elixhauser AHRQ-Web ICD-9-CM

(八) Definitions of comorbidities+ ICD-10-CM

2015 年起健保檔之國際疾病分類從 ICD-9-CM 轉換成 ICD-10-CM，所以須將 comorbidity 的 ICD-9-CM 代碼按照健保署公告轉碼成 ICD-10-CM。其中對應不確定的部分均以下列依規則處理：

ICD-10-CM	ICD-9-CM	Comorbidity (ICD-9-CM)	Comorbidity (ICD-10-CM)
E10.65	250.0	comD9_DMWOC	comD9_DMWOC
	250.8	comD10_DMC	
E11.65	250.0	comD9_DMWOC	comD9_DMWOC
	250.1	comD9_DMWOC	
	250.8	comD10_DMC	
E11.69+E11.65*	250.1	comD9_DMWOC	comD10_DMC
	250.8	comD10_DMC	
E10.69,	250.2	comD9_DMWOC	comD9_DMWOC,
E10.69+E10.65*	250.8	comD10_DMC	comD10_DMC
E10.21, E11.21	250.4, 583.x	comD10_DMC, comD11_CRF	<p>1.有 comD10_DMC 其他代碼而沒有 comD11_CRF 其他代碼的人分類到 comD10_DMC。</p> <p>2.有 comD11_CRF 其他代碼而沒有 comD10_DMC 其他代碼的人分類到 comD11_CRF。</p> <p>3-1.皆有 comD10_DMC 和 comD11_CRF 其他代碼的人分類到最早被記載到的其他代碼。</p> <p>3-2.最早被記載到的其他代碼因日期相同無法區別先後順序者皆視為有 comD10_DMC 和 comD11_CRF。</p> <p>4.只有 E10.21、E11.21 而沒有 comD10_DMC 其他代碼,而也沒有 comD11_CRF 其他代碼分類到 comD11_CRF。</p>
I10	401.1, 401.9	comD17_HTU	只有 I10 而沒有 comD17_HTU 其他代碼,而也沒有 comD18_HTC 其他代碼分類到 comD17_HTU。
	401.0	comD18_HTC	

*"+"代表前後之代碼在某一次的門診檔或住院檔中同時出現。前面代碼為組合式中的第一診斷，後面代碼為此組合式中的第二診斷。順序上之說明請詳衛生福利部中央健康保險署中 2014 年版_ICD-9-CM2001 年版與 ICD-10-CM/PCS 對

應檔(112.05.09 更新)，網址：

https://www.nhi.gov.tw/Content_List.aspx?n=C1C92AB9ED30A9FD&topn=23C660CAACAA159D。

四、檔案名稱格式說明範例

(一) 串聯後的健保資料庫、癌症登記檔資料檔的檔案名稱之說明範例

以下述的排列設定檔案名稱：

被串聯的檔案名稱年度(月份)_癌症名稱_癌登年報檔年度

範例：

	檔案名稱(Filename)	檔案內容
範例 1	OPDTE11001_10_LUNG_111	使用癌症研究主檔 1979-2022 年版本，篩選出肺癌個案後，擷取肺癌個案在 110 年 1 月全民健保處方及治療明細檔_門急診(西醫)(OPDTE11001_10)的紀錄。
範例 2	IPDTE11001_LUNG_111	使用癌症研究主檔 1979-2022 年版本，篩選出肺癌個案後，擷取肺癌個案在 110 年 1 月全民健保處方及治療明細檔_住院的紀錄。

(二) 經 ID 清理的資料檔之檔案名稱

以下述的排列設定檔案名稱：

資料檔名稱_癌症名稱_癌登年報檔年度

檔案名稱(Filename)	對應檔案中文名稱
DEATH_QC_LUNG_111	肺癌死因統計合併檔
MASTER_TABLE_LUNG_111	肺癌研究主檔