

乳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	乳癌研究主檔(1979-2020)： 304,550 筆
中文檔名	乳癌主題式資料庫	英文檔名	Breast Cancer Health Database
屬性	歷史檔	週期	年
檔案大小	184GB	譯碼簿 更新日期	2024/10/30
檔案列表及其年度、欄位數	<p>本資料庫包含以下 15 個檔：</p> <ol style="list-style-type: none"> 1. 乳癌門急診明細檔 (1998-2021)：51 2. 乳癌住院明細檔 (1998-2021)：81 3. 乳癌藥局明細檔 (2000-2021)：44 4. 乳癌門急診醫令檔 (2000-2021)：22 5. 乳癌住院醫令檔 (2000-2021)：22 6. 乳癌藥局醫令檔 (2000-2021)：17 7. 乳癌承保檔 (2000-2021)：30 8. 乳癌重大傷病檔 (2001-2022)：27 9. 乳癌死因統計合併檔 (1985-2021)：16(合併 1985-2021 年的死因統計檔) 10. 乳癌癌登檔 LF (2007-2020)：118 11. 乳癌癌登檔 SF (2007-2020)：42 12. 乳癌癌登檔 TCDB (2002-2006)：65 13. 乳癌癌登年報檔 (1979-2020)：13 14. 乳癌研究主檔 (1979-2020)：51 15. 癌症篩檢-乳癌篩檢清理檔 (2004-2021)：71 		
資料描述	<p>乳癌主題式資料庫包含了收錄在 1979~2020 年癌登年報檔中之每一乳癌患者在癌登檔(包括長表及短表)及健保檔、死因統計合併檔，共 14 個資料檔之所有資訊。另外我們也清理了乳癌篩檢檔，以便串聯乳癌主題式資料庫聯合使用。詳細定義如下</p>		

	<ol style="list-style-type: none"> 1. 定義乳癌個案： 乳癌個案為癌症登記年報檔 (H_BHP_CRSSC)1979-2020 年(經 ID 清理後)中，曾有一筆其原發部位(CASITE)的前三碼為 174(癌症登記年報檔均依據 ICD-9 譯碼)，但其組織類型與性態碼(HISTBEH)不為血液腫瘤的個案。 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其存活狀態等。 3. 以乳癌個案的 ID 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)、癌症登記檔長表 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，並從其中擷取出乳癌個案的資料檔，說明如下： <ol style="list-style-type: none"> (1) 乳癌研究主檔：從癌症研究主檔中，擷取出乳癌個案的所有紀錄。 (2) 乳癌死因統計合併檔：從死因統計合併檔中，擷取出乳癌個案。 5. 「癌症篩檢-乳癌篩檢檔」為衛生福利部國民健康署提供之乳癌篩檢資料。「癌症篩檢-乳癌篩檢清理檔」為「乳癌篩檢檔」經資料清理後之檔案。
注意事項	<ol style="list-style-type: none"> 1. 本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，身分證字號、院所代碼、投保單位代號、統一編號等欄位均經屏蔽處理，屏蔽強度符合 FIPS140-2 Level 3 國際資安標準，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。 2. 乳癌患者若為多發癌症(一 ID 有多筆癌症年報檔的紀錄)的個案，除了乳癌的紀錄，乳癌研究主檔也包含了多發癌症個案的其他非乳癌的紀錄，例如：個案有一筆乳癌紀錄，和一筆某癌症(如肺癌或血癌等)紀錄，則兩筆紀錄皆會收錄。若要分析第一筆診斷紀錄

	<p>為乳癌的個案，建議從乳癌研究主檔中，以癌症類別(newca2)為a1741，且記錄順序(rank_record)≤1，以及紀錄檢誤(accuracy_record)=1篩選。</p> <p>3. 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> <p>4. 「申請欄位數小計」請填寫所勾選的欄位數，並依「衛生福利統計資料整合應用服務收費標準」全人口不分年度檔收費。</p>
主鍵與比對欄位	<p>1. 身分證字號(ID)、醫師身分證字號(PRSN_ID)及醫療機構代碼(HOSP_ID)已經加密處理。</p> <p>2. 個人歸戶及個人資料對照的鍵值：ID。</p> <p>3. 醫療機構歸戶及醫療機構資料對照的鍵值：HOSP_ID。</p>

二、檔案格式及資料描述

(一) 以乳癌個案的 ID 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. 癌症登記檔長表 TCDB (H_BHP_CRSLF)
12. 癌症登記年報檔(H_BHP_CRSSC)

(二) 乳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) 癌症篩檢-乳癌篩檢清理檔之欄位格式與說明另列於後。

(四) 乳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	ID	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	DIAG_DT	Char	癌登年報檔	診斷日期	YYYYMMDD
5	HISTBEH	Char	癌登年報檔	組織類型與	組織病理：用 ICD-O

序號	欄位	型態	來源	內容	描述
			報檔	性態碼	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	用 ICD-O 譯碼。 NHL(非何杰金氏淋巴瘤)、HL(何杰金氏淋巴瘤)及 LEU(白血病)血液腫瘤個案，在此欄位無法辨識，請用組織類型與性態碼(HISTBEH)區分。若是要分析各癌症部分(CASITE 欄位)，也請先把 NHL、HL、LEU 排除後再計算。
7	TX	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)，或為新的治療方法，非指他院治療
8	DIAGAGE	Num	癌登年報檔	診斷年齡	

序號	欄位	型態	來源	內容	描述
9	RESID	Char	癌登年報檔	戶籍地代碼	詳戶籍地代碼(RESID)對照表 100年起開始使用新五都代碼
10	CONFIRM	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，可逕至台灣癌症登記

序號	欄位	型態	來源	內容	描述
					<p>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</p>

序號	欄位	型態	來源	內容	描述
					短表 45 欄位與長表 115 欄位內容)
12	ID_S	Char	癌登年報檔	性別	1：男性 2：女性 9：不詳
13	ID_ROC	Char	癌登年報檔	身分證字號檢誤	0：身分證字號符合編碼原則 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_COD E	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	癌別類別內的診斷順序	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。

序號	欄位	型態	來源	內容	描述
24	n_cancer	Num	本檔自訂	癌症類別次數	依 newca2 統計每一 ID 的癌症類別次數。
25	rank_newca2	Num	本檔自訂	癌別類別的診斷順序	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。
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。
28	os_event	Num	本檔自訂	是否為 overall survival event	0=存活； 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。符合與否的依據，詳資料整理(四)綜合癌登檔與死因檔做紀錄檢誤。
30	adj_diag_dt	Char	本檔自訂	填補後的癌症診斷日期	當診斷日期 (DIAG_DT) 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，

序號	欄位	型態	來源	內容	描述
					以便與追蹤日期計算出存活時間。填補方法詳資料整理(七) 追蹤日期與存活時間。
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。詳資料整理(七) 追蹤日期與存活時間。
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況。詳資料整理(四)綜合癌登檔與死因檔做紀錄檢誤和(七) 追蹤日期與存活時間。
34	ComD1_AMI	Num	本檔自訂	是否有 ComD1_AMI	0：沒有；1：有； -9：未知，詳資料整理(八)共病症； 空白表示不是該病患第一個侵襲性癌症。
35	ComD2_OMI	Num	本檔自訂	是否有 ComD2_OMI	0：沒有；1：有； -9：未知，詳同序號34； 空白表示不是該病患第一個侵襲性癌症。

序號	欄位	型態	來源	內容	描述
36	ComD3_CHF	Num	本檔自訂	是否有 ComD3_ CHF	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
37	ComD4_PVD	Num	本檔自訂	是否有 ComD4_ PVD	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
38	ComD5_CVD	Num	本檔自訂	是否有 ComD5_ CVD	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
39	ComD6_COPD	Num	本檔自訂	是否有 ComD6_ COPD	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
40	ComD7_Demen	Num	本檔自訂	是否有 ComD7_ Demen	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
41	ComD8_Paraly	Num	本檔自訂	是否有 ComD8_ Paraly	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
42	ComD9_ DMWOC	Num	本檔自訂	是否有 ComD9_ DMWOC	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
43	ComD10_DMC	Num	本檔自訂	是否有 ComD10_	0：沒有；1：有； -9：未知，詳同序號

序號	欄位	型態	來源	內容	描述
				DMC	34； 空白表示不是該病患 第一個侵襲性癌症。
44	ComD11_CRF	Num	本檔自 訂	是否有 ComD11_ CRF	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
45	ComD12_MLD	Num	本檔自 訂	是否有 ComD12_ MLD	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
46	ComD13_ MSLD	Num	本檔自 訂	是否有 ComD13_ MSLD	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
47	ComD14_Ulcer	Num	本檔自 訂	是否有 ComD14_ Ulcer	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
48	ComD15_RD	Num	本檔自 訂	是否有 ComD15_ RD	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
49	ComD16_AIDS	Num	本檔自 訂	是否為 AIDS	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。
50	ComD17_HTU	Num	本檔自 訂	是否有 ComD17_ HTU	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患

序號	欄位	型態	來源	內容	描述
					第一個侵襲性癌症。
51	ComD18_HTC	Num	本檔自訂	是否有 ComD18_ HTC	0：沒有；1：有； -9：未知，詳同序號 34； 空白表示不是該病患 第一個侵襲性癌症。

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。

Site	ICD-9 cancer site / histology	newca2
Histology code ≠ 9050-9055,9140,9590-9992		
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≠8000-8005,8010-8046,8050-8084,8090-8110,8720-8790	a1730
Breast	174,175	a1741
Cervix Uteri	180	a1800

Site	ICD-9 cancer site / histology	newca2
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 = 9050-9055,9140,9590-9992		
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	98663,98713,98963,98973,98953,99203,98613,98403,98673,98703,98723,98733,98743,98913,99103,99313,99303,97273,98353,98363,98373,98633,98753,98763,99453,99463,99633,	a3003

Site	ICD-9 cancer site / histology	newca2
	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) Definitions of comorbidities+ ICD-9-CM

No.	Disease	ICD-9-CM	癌症主檔之代碼
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
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	250.4-250.6,250.8-250.9	ComD10_DMC

	complication (DM W CC)		
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

(三) 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, E10.69+E10.65*	250.2	comD9_DMWOC	comD9_DMWOC, comD10_DMC
	250.8	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	OPDTE10901_10_BREAST_108	使用癌症研究主檔 1979-2019 年版本，篩選出乳癌個案後，擷取乳癌個案在 109 年 1 月全民健保處方及治療明細檔_門急診(西醫)(OPDTE10901_10)的紀錄。
範例 2	IPDTE109_BREAST_108	使用癌症研究主檔 1979-2019 年版本，篩選出乳癌個案後，擷取乳癌個案在 109 年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_109_BREAST_108	乳癌死因統計合併檔
MASTER_TABLE_BREAST_108	乳癌研究主檔

五、癌症篩檢—乳癌篩檢清理檔

根據 2022 年 01 月 28 日修訂修訂版本製作

(一)、檔案內容說明

檔案代號	H_BHP_BCS_CL	資料筆數	H_BHP_BCS_PD：4,435,634筆 H_BHP_BCS_DATA_CL： 9,063,986筆 H_BHP_BCS_UNIT：65,561筆		
中文檔名	1. 乳癌篩檢-個案基本資料檔 (H_BHP_BCS_PD) 2. 乳癌篩檢-乳攝資料清理檔 (H_BHP_BCS_DATA_CL) 3. 乳癌篩檢-篩檢單位檔 (H_BHP_BCS_UNIT)	英文檔名			
檔案大小	1. H_BHP_BCS_PD：111MB 2. H_BHP_BCS_DATA_CL： 3GB 3. H_BHP_BCS_UNIT： 1.62MB	欄位數	1. H_BHP_BCS_PD：7 2. H_BHP_BCS_DATA_CL：55 3. H_BHP_BCS_UNIT：9		
屬性	歷史資料檔	週期	無	譯碼簿 更新日期	2024/10/30
資料描述	<p>1. 國民健康署自民國99年起擴大辦理4項癌症（大腸癌、口腔癌、乳癌及子宮頸癌）篩檢服務，透過全民健保預防保健服務特約醫事服務機構全面提供大腸癌、乳癌、子宮頸癌與口腔癌篩檢，並將大腸癌、口腔癌篩檢納入預防保健服務項目，及擴大乳癌篩檢對象至40歲以上至未滿45歲二等親乳癌高危險群（即其母親、女兒、姊妹、祖母或外祖母曾患有乳癌）之婦女。</p> <p>2. 政府補助45-69歲婦女每2年一次乳房攝影檢查；40-44歲婦女，二等親內具乳癌家族史（指祖母、外婆、母親、女兒、姊妹曾有人罹患乳癌）婦女亦提供篩檢服務。</p> <p>3. 乳房X光攝影檢查：45-69歲婦女、40-44歲二等血親內曾罹患乳癌之婦女，每2年1次。</p> <p>4. 篩檢日期 包含 2004 年 7 月 到 2021 年 1 月 。</p> <p>5. 「癌症篩檢-乳癌篩檢清理檔」新增欄位：身高2、體重2、身體質量指數及資料保留。</p>				
注意事項	1. 本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，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，屏蔽強度符合FIPS140-2 Level 3 國際資				

	<p>安標準，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> <ol style="list-style-type: none"> 2. 身高及體重欄位有數字、文字、符號及其組合。 3. 乳攝資料清理檔 (H_BHP_BCS_PD)中，放射科醫師代碼(DR)及放射技術師代碼(XDR)，在 醫護人員檔(H_BHP_BCS_DR) 中皆為醫護人員代碼(DR)。 4. 篩檢單位包含各縣市衛生局及國外機構。
主鍵與比對欄位	<ol style="list-style-type: none"> 1. 個案基本資料檔(H_BHP_PD) 與乳攝資料清理檔 (H_BHP_BCS_DATA_CL) 比對之唯一鍵值為 ID。 2. 乳癌篩檢-篩檢單位檔(H_BHP_BCS_UNIT)與乳攝資料清理檔 (H_BHP_BCS_DATA_CL)比對之唯一鍵值為HOSP_ID。 3. 個人歸戶及個人資料比對的鍵值： ID。 4. 醫事機構歸戶與資料比對的鍵值： HOSP_ID。

(二)、檔案格式及資料描述

乳癌篩檢-個案基本資料檔(H_BHP_BCS_PD)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
1	身分證字號	ID	Char	12	經加密處理。國民身分證統一編號，或外籍居留證號碼（如無居留證號碼請填護照號碼），左靠不足補空白。
2	身分證字號檢誤	ID_ROC	Char	1	0：身分證字號符合編碼原則； 1：不符合
3	證件來源	ID_TYPE	Char	1	1：身分證；2：護照
4	性別	ID_S	Char	1	1：男；2：女；9：不詳
5	原始檔性別	SEX	Char	1	1：男；2：女；9：不詳
6	出生年月	BIRTH_YM	Char	6	提供西元年月YYYYMM
7	現居地代碼	LIVING	Char	4	代碼參照死因統計檔。

乳癌篩檢-乳房攝影資料清理檔(H_BHP_BCS_DATA_CL)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
1	支付方式	PAYWAY	Char	1	詳譯碼說明
2	疾病史-有無得過下列疾病	HIS_DIS	Char	1	0：無；1：乳房良性相關疾病；2：乳癌；3：其他
3	疾病史-其他癌症敘述	HIS_OC	Char	60	
4	家族史-與您有血緣的家屬中，有無人得過癌症	FAM_CANCE R	Char	1	0：無；1：有
5	家族史-母親數	FAM_MOM	Num	8	
6	家族史-姊妹	FAM_SIS	Num	8	
7	家族史-女兒	FAM_DAU	Num	8	
8	家族史-祖母	FAM_GMA1	Num	8	
9	家族史-外祖母	FAM_GMA2	Num	8	
10	家族史-姑姑	FAM_AUNT1	Num	8	
11	家族史-阿姨	FAM_AUNT2	Num	8	
12	月經史-初經年齡	MC_FAGE	Num	8	
13	月經史-是否停經	MC_END	Char	1	0：否；1：是
14	月經史-停經年齡	MC_EAGE	Num	8	
15	月經史-停經原因	MC_CAU	Char	10	詳譯碼說明
16	月經史-停經原因其他敘述	MC_CAU_O	Char	60	
17	生育史-生育次數	FER_TIMES	Num	8	
18	生育史-有無哺餵母乳	FER_FEEDIN	Char	1	0：無；1：有
19	生育史-第一胎生產年齡	FER_FAGE	Num	8	
20	用藥史-是否服用荷爾蒙補充劑	MED_HUSE	Char	1	0：否；1：是
21	用藥史-幾歲開始(荷)	MED_HAGE	Num	8	
22	用藥史-服用幾年(荷)	MED_HYEAR	Num	8	
23	用藥史-是否服用避孕藥	MED_CUSE	Char	1	0：否；1：是
24	用藥史-幾歲開始(避)	MED_CAGE	Num	8	
25	用藥史-服用幾年(避)	MED_CYEAR	Num	8	
26	是否有感覺到異常腫塊	TOUCH	Char	1	詳譯碼說明
27	其他-最近二年內是否做過-觸診	O_HAND	Char	1	1：有；Null：沒有勾選

乳癌篩檢-乳房攝影資料清理檔(H_BHP_BCS_DATA_CL)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
28	其他-最近二年內是否做過-乳攝	O_MG	Char	1	1：有；Null：沒有勾選
29	其他-最近二年內是否做過-超音波	O_ULTRA	Char	1	1：有；Null：沒有勾選
30	其他-最近二年內是否做過-皆未做過	O_NON	Char	1	1：有；Null：沒有勾選
31	其他-是否接受過乳房手術	O_SUR	Char	1	T：是；F：否
32	其他-乳房手術左側	O_LSUR	Char	1	0：無；1：有
33	其他-乳房手術右側	O_RSUR	Char	1	0：無；1：有
34	身分證字號	ID	Char	12	經加密處理。國民身分證統一編號，或外籍居留證號碼（如無居留證號碼請填護照號碼），左靠不足補空白。
35	身分證字號檢誤	ID_ROC	Char	1	0：身分證字號符合編碼原則；1：不符合
36	性別	ID_S	Char	1	1：男；2：女；9：不詳
37	教育程度	EDU	Char	1	詳譯碼說明
38	身高	HIGHET	Num	8	
39	體重	WEIGHT	Num	8	
40	乳攝醫院代碼	HOSP_ID	Char	9	經加密處理，為辨別院所鍵值。
41	乳攝日期	MG_DATE	Char	8	西元年YYYYMMDD
42	乳攝車	CAR	Char	1	0：否；1：是
43	攝影機機型	MACHINE	Char	1	詳譯碼說明
44	觸診結果	HRESULT	Char	1	詳譯碼說明
45	乳腺組成	PDATA	Char	1	詳譯碼說明
46	與舊片比較	COMPARE	Char	1	F：否；T：是
47	無舊片的原因	NO_COMPARE	Char	1	詳譯碼說明
48	乳攝結果	RESULT	Char	1	詳譯碼說明
49	乳攝結果4的次選項	RESULT4	Char	1	詳譯碼說明
50	放射科醫師代碼	DR	Char	5	不提供
51	放射技術師代碼	XDR	Char	5	不提供

乳癌篩檢-乳房攝影資料清理檔(H_BHP_BCS_DATA_CL)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
52	醫療機構縣市鄉鎮市區代碼	CITY	Char	4	代碼參照死因統計檔。
53	醫療機構權屬別	HOS	Char	2	代碼參照醫療機構現況檔
54	身高 2	HIGHET2	Num	8	乳房攝影資料清理檔新增欄位經過整理的身高(公分)
55	體重 2	WEIGHT2	Num	8	乳房攝影資料清理檔新增欄位經過整理的體重(公斤)
56	身體質量指數	BMI	Num	8	乳房攝影資料清理檔新增欄位體重 2(kg)/身高 2(m)^2
57	資料保留	DATA_QC	Char	1	乳房攝影資料清理檔新增欄位 N:資料不符合保留條件 Y:資料符合保留條件 保留條件詳譯碼簿說明

乳癌篩檢-篩檢單位檔(H_BHP_BCS_UNIT)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
1	單位代碼	HOSP_ID	Char	9	經加密處理，為辨別院所鍵值。
2	單位等級	U_LEVEL	Char	1	詳譯碼說明
3	單位所在區碼	U_AREA	Char	4	
4	醫院等級	U_HLEVEL	Char	1	詳譯碼說明
5	是否為乳攝合約醫院	U_MGCH	Char	1	T：是；Null：否
6	是否限制乳攝車	U_CAR	Char	1	0：否；1：是
7	報表用歸屬縣市	REPORT_CITY	Char	1	T：是；Null：否
8	醫療機構縣市鄉鎮市區代碼	CITY	Char	4	代碼參照死因統計檔。
9	醫療機構權屬別	HOS	Char	2	代碼參照醫療機構現況檔。

(三)、欄位(變項)譯碼

1. 支付方式(PAYWAY)

支付方式代碼	支付方式說明
1	婦女醫療給付
2	預防保健
3	其他公務預算補助
4	其他
5	自費健康檢查

2. 月經史-停經原因(MC_CAU)

停經原因代碼	停經原因說明
Null	未停經
2	自然停經
3	子宮切除
4	卵巢切除
5	其他

3. 是否有感覺到異常腫塊(TOUCH)

異常腫塊代碼	異常腫塊說明
1	有摸到硬塊或疼痛、壓痛
2	沒有症狀
3	未做過乳房自我檢查

4. 教育程度(EDU)

教育程度代碼	教育程度說明
1	不識字
2	小學
3	國(初)中
4	高中/專科
5	大學以上

5. 攝影機機型(MACHINE)

攝影機機型代碼	攝影機機型說明
0	不詳
1	傳統乳房攝影儀
2	數位乳房攝影儀(CR)
3	數位乳房攝影儀(DR)

6. 乳腺組成(PDATA)

乳腺組成代碼	乳腺組成說明
0	不詳
1	Fatty breast
2	Mildly dense
3	Moderate dense
4	Extremely dense

7. 觸診結果(HRESULT)

觸診結果代碼	觸診結果說明
1	有腫塊或其他異常
2	無異常
3	未做觸診檢查

8. 無舊片的原因(NO_COMPARE)

無舊片的原因代碼	無舊片的原因說明
1	第一次篩檢
2	無法取得舊片
3	無法獲知是否有舊片

9. 乳攝結果(RESULT)

乳攝結果代碼	乳攝結果說明
0	需附加其他影像檢查再評估
1	無異常發現
2	良性發現
3	可能是良性發現-需六個月追蹤檢查
4	可疑異常需考慮組織生檢
5	高度疑似為惡性腫瘤必須採取適當的措施

10. 乳攝結果4的次選項(RESULT4)

乳攝結果4的次選項代碼	乳攝結果4的次選項說明
a	Low suspicion
b	Intermediate suspicion
c	Moderate suspicion
d	不詳

11. 單位等級(U_LEVEL)

醫院等級代碼	醫院等級說明
1	國民健康署
2	監測中心
3	衛生局
4	衛生所
5	醫療檢驗所

12. 醫院等級(HLEVEL)

醫院等級代碼	醫院等級說明
1	教學醫院
2	區域醫院
3	地區醫院
4	診所
其他代碼	不詳

13. 資料保留(DATA_QC)

乳篩檔中之一筆乳篩資料如果有任何的離群值或矛盾情形則以 N 註記該筆資料。詳細處理依下列表格說明。

資料欄位	欄位內容的保留條件
家族史-母親數	0, 1
家族史-姊妹	0-10
家族史-女兒	0-10
家族史-祖母	0, 1
家族史-外祖母	0, 1
月經史-初經年齡	5-25 歲
月經史-停經年齡	當月經史-是否停經內容為 0，則月經史-停經年齡內容為 0 或空白的可以保留
	當月經史-是否停經內容為 1 且月經史-停經原因內容為 2，則月經史-停經年齡之內容為 35-70 可以保留
	當月經史-是否停經內容為 1 且月經史-停經原因內容為 3 或 4 或 5，則月經史-停經年齡之內容全部保留(空白除外)
生育史-生育次數	0-15
生育史-有無哺乳	生育史-生育次數內容為 0，則生育史-有無哺乳內容為 0 或空白的可以保留
	生育史-生育次數內容為 1-15，則生育史-有無哺乳內容為 0 或 1 的可以保留

生育史-第一胎生產年齡	生育史-生育次數內容為 0，則生育史-第一胎生產年齡內容為 0 或空白的可以保留
	生育史-生育次數內容為 1-15，則第一胎生產年齡內容為 10-60 的可以保留
是否服用賀爾蒙補充劑	0, 1
是否服用避孕藥	0, 1
身高 2	110-210
體重 2	35-120

-檢查時間(年齡)合理性:

初經年齡 <= 第一胎生產年齡 < 停經年齡(若篩檢時已經停經) <= 篩檢年齡

初經年齡、第一胎生產年齡、停經年齡為個案提供。

篩檢年齡為利用資料欄位計算：篩檢年-出生年

初經年齡	第一次生育年齡	停經年齡	篩檢年齡	保留條件
有	無	無	有	初經年齡<篩檢年齡
有	有	無	有	初經年齡 <= 第一胎生產年齡 <= 篩檢年齡 第一胎生產年齡大篩檢年齡兩歲亦算合理*
有	無	有 (含非自然停經)	有	初經年齡 < 停經年齡 <= 篩檢年齡 停經年齡大篩檢年齡兩歲亦算合理*
有	有	有 (含非自然停經)	有	初經年齡<=第一胎生產年齡<停經年齡<= 篩檢年齡 停經年齡大篩檢年齡兩歲亦算合理*

*考量到虛歲問題