



2023/8/23

請先掃描QR Code，完成上課簽到表

健保資料處理操作實務

SAS-2

Working with Your Data

- Setting Up Your Dataset

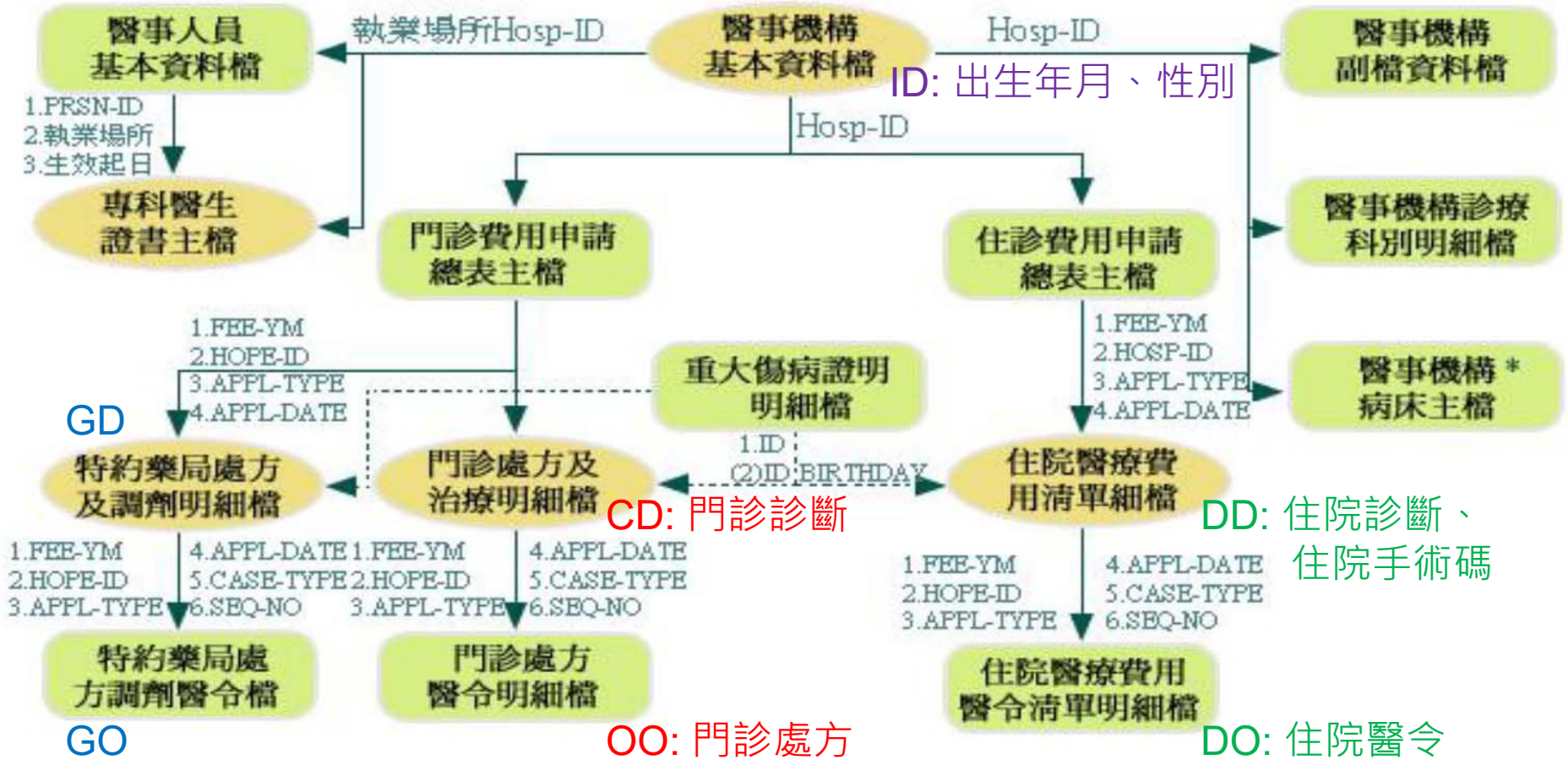
Definition of Medical History

高級助理研究員陳韻仔博士

簡稱	檔案全名	常用欄位
ID	HEALTH-07：H_NHI_ENROL (全民健保承保檔)	ID_S 性別、ID_BIRTH_YM 出生年月
CD	HEALTH-01：H_NHI_OPDTE (全民健保處方及治療明細檔_門急診)	ICD9CM_1-ICD9CM_3 門診診斷碼、 DRUG_DAY 給藥日份
OO	Health-04：H_NHI_OPDTO (全民健保處方及治療醫令明細檔_門急診)	DRUG_NO 藥碼
DD	HEALTH-02：H_NHI_IPDTE (全民健保處方及治療明細檔-西醫住院)	ICD9CM_1-ICD9CM_5 住院診斷碼、 ICD_OP_CODE1-ICD_OP_CODE5 住院手術碼
DO	HEALTH-05：H_NHI_IPDTO (全民健保處方及治療醫令明細檔-住院)	ORDER_CODE 醫令代碼(用藥、處置)
GD	特約藥局處方及調劑檔	
GO	特約藥局處方醫令檔	
Death	Health-10：H_OST_DEATH (分析區使用-死因統計檔)	D_DATE 死亡日期、 D_CODE (ICD8 or ICD9)、ICD10 死因

- CD檔, DD檔, GD檔有承保人ID欄位
- OO檔, DO檔, GO檔沒有承保人ID欄位

各檔案間串檔變項說明



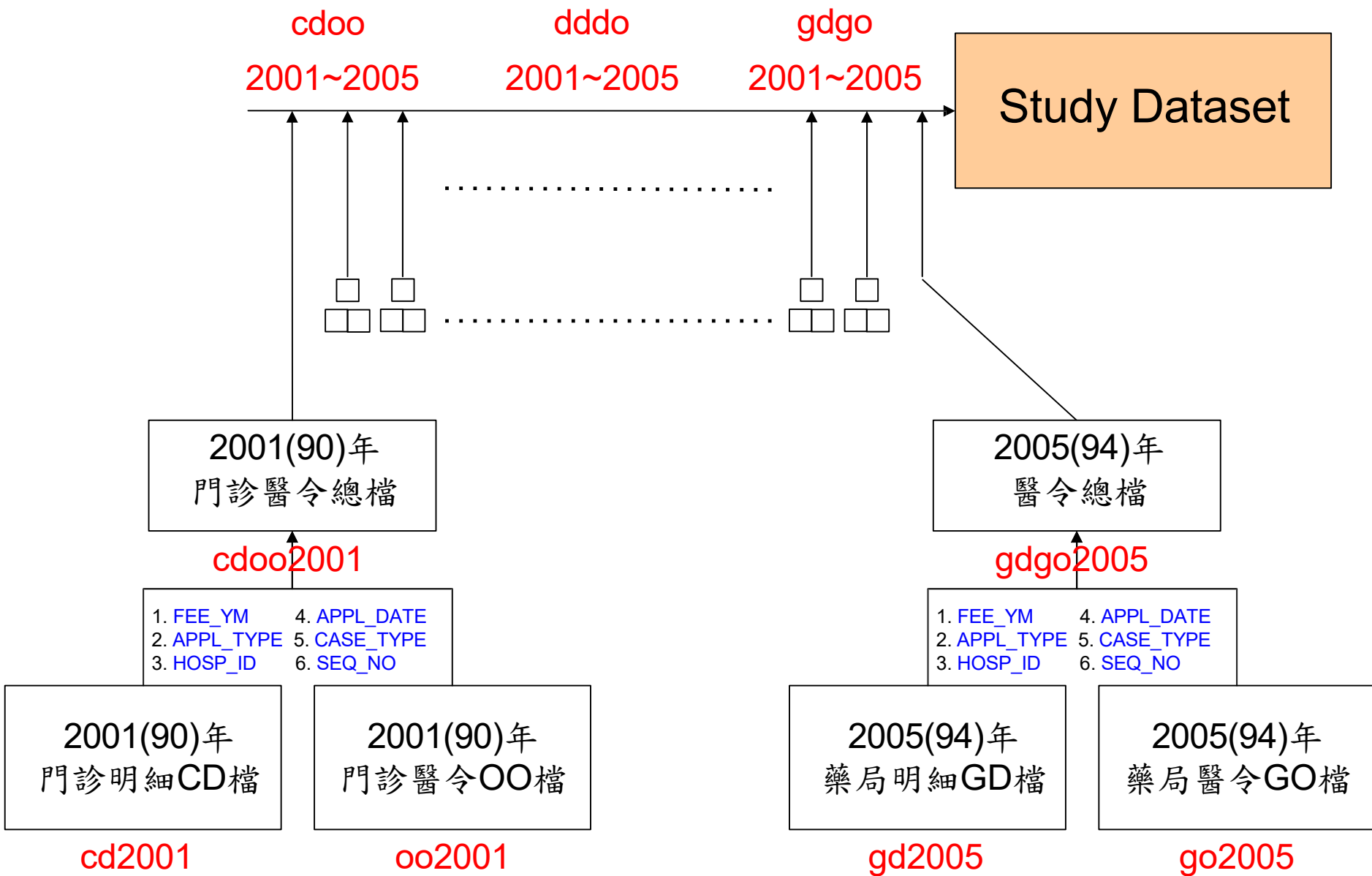
註:*須注意生效起訖日期

(2)可由ID+BIRTHDAY串檔

→ 各檔案間由所註明變項串檔可獲得對應資訊

→ 各檔案間可由所註明變項串檔,但未必獲得對應資料

串檔模式 1



資料編排方式與相對應的使用指令

2001(90)年
門診明細CD檔

Merge



1. FEE_YM
2. APPL_TYPE
3. HOSP_ID
4. APPL_DATE
5. CASE_TYPE
6. SEQ_NO

2001(90)年
門診醫令OO檔

Set ↓ **Append**

2002(91)年
門診明細CD檔

Merge



1. FEE_YM
2. APPL_TYPE
3. HOSP_ID
4. APPL_DATE
5. CASE_TYPE
6. SEQ_NO

2002(91)年
門診醫令OO檔

Set ↓ **Append**

SAS Date

- A SAS date is a numeric value equal to the number of days since 1/1/1960. For example,

Date	SAS date value
1/1/1959	-365
1/1/1960	0
1/1/1961	366
1/1/2003	15706

標準的健保資料檔

11	FUNC_TYPE	就醫科別	C	2	66	67	請參考代碼說明之「就醫科別及細分科」
12	FUNC_DATE	就醫日期	C	8	68	75	YYYYMMDD 請參考代碼說明之「就醫日期及治療結束日期補充說明」
13	TREAT_END_DATE	治療結束日期	C	8	76	83	YYYYMMDD 請參考代碼說明之「就醫日期及治療結束日期補充說明」，同一療程、排程檢查或慢性病連續處方籤調劑案件，本欄必填。
14	ID_BIRTHDAY	出生日期	C	8	84	91	YYYYMMDD

Age (1)

```
data work.cd1997a ;  
  set work.cd1997 ;
```

從FUNC_DATE的
第一位開始,抽取4位
文字變項乘以1轉為
數值變項

```
diagyear = substr (FUNC_DATE,1,4) * 1 ;  
diagmonth = substr (FUNC_DATE,5,2) * 1 ;  
diagday = substr (FUNC_DATE,7,2) * 1 ;  
diagdate = mdy ( diagmonth , diagday , diagyear ) ;
```

```
birthyear = substr (ID_BIRTH_YM,1,4) * 1 ;  
birthmonth = substr (ID_BIRTH_YM,5,2) * 1 ;  
birthday = 15 ;
```

```
birthdate = mdy ( birthmonth , birthday , birthyear ) ;
```

```
diagage = (diagdate - birthdate)/365.25;
```

```
run;
```

SAS date value
(the No. of days
since 1/1/1960)

可以使用ID或CD檔

Conversion between Numeric Values and Character

PUT and INPUT

- Character to Numeric

newvar = INPUT (*oldvar*, *informat*);
數值 文字 數值格式

- Numeric to Character

newvar = PUT (*oldvar*, *format*);
文字 數值 數值格式

Age (2)

```
data work.cd1997a ;
```

```
set work.cd1997 ;
```

```
diagdate = input(FUNC_DATE, YYMMDD8.) ;
```

```
birthdate = input(ID_BIRTHDAY, YYMMDD8.);
```

```
diagage = (diagdate - birthdate)/365.25;
```

```
run;
```

直接把文字讀成
SAS date value

Age Group

```
**set age group;  
data work.Q4;  
set work.Q3;  
if          diagage = . then age_gp = . ;  
else if     diagage < 10 then age_gp = 1;  
else if 10 =< diagage < 20 then age_gp = 2;  
else if 20 =< diagage < 30 then age_gp = 3;  
else if 30 =< diagage < 40 then age_gp = 4;  
else if 40 =< diagage < 50 then age_gp = 5;  
else if 50 =< diagage < 60 then age_gp = 6;  
else if 60 =< diagage < 70 then age_gp = 7;  
else if 70 =< diagage < 80 then age_gp = 8;  
else          age_gp = 9;  
run;
```

Symbolic

Mnemonic

Meaning

=

EQ

equals

^= or ~=

NE

not equal

>

GT

greater than

<

LT

less than

>=

GE

greater than or equal

<=

LE

less than or equal

- You can either use symbolic or mnemonic operators

Extraction According to Diagnosis

****Subjects with diagnosis of diabetes mellitus ;**

data DM;

set NHI.cd2000;

if substr (ACODE_ICD9_1, **1**, **3**) = '250' or

substr (ACODE_ICD9_2, **1**, **3**) = '250' or

substr (ACODE_ICD9_3, **1**, **3**) = '250' ;

run;

門診檔與住院檔:診斷與
手術之變項名稱不同

First / Last Observation

```
proc sort data = work.cd1999 out = a1;  
by ID;  
run;
```

一定要先排序

```
data work.firstvisit;  
set work.a1;  
by ID;  
if first.ID;  
run;
```

抓出同一個人的第一筆資料

```
data work.lastvisit;  
set work.a1;  
by ID;  
if last.ID;  
run;
```

抓出同一個人的最後一筆資料

```
/*2023/8/23 Medical History Identification*/  
/* Age & Sex */
```

```
data study.id2006 ; set study.h_nhi_enrol95 ;  
diagdate = mdy (12 , 31 , 2022 ) ;
```

```
birthyear = substr (ID_BIRTH_YM,1,4) * 1 ;  
birthmonth = substr (ID_BIRTH_YM,5,2) * 1 ;  
birthday = 15 ;  
birthdate = mdy ( birthmonth , birthday , birthyear ) ;  
age = (diagdate - birthdate)/365.25;
```

```
agegp=.;  
if age <= 65 then agegp=0;  
else agegp=1;
```

```
sex=.;  
if id_s=1 then sex=1; else if id_s=9 then sex=.;else sex=0;  
run;
```

```
Proc sort data=study.id2006; by id id_in_date; run;  
data id2006; set study.id2006; by id ; if first.id; run;
```

```
126 Proc sort data=study.id2006; by id id_in_date; run;  
NOTE: 已從資料集 STUDY.ID2006. 讀取 92091 個觀測值  
NOTE: 資料集 STUDY.ID2006 有 92091 個觀測值和 25 個變數。  
NOTE: 已使用 PROCEDURE SORT (總處理時間):  
即時 0.06 秒  
cpu 時間 0.07 秒  
  
127 data id2006; set study.id2006; by id ; if first.id; run;  
NOTE: 已從資料集 STUDY.ID2006. 讀取 92091 個觀測值  
NOTE: 資料集 WORK.ID2006 有 91581 個觀測值和 25 個變數。  
NOTE: 已使用 DATA 陳述式 (總處理時間):  
即時 0.02 秒  
cpu 時間 0.01 秒
```

HEALTH-01 : H_NHI_OPDTE (全民健保處方及治療明細檔_門急診) CD

欄位序號	欄位名稱	型態	長度	欄位中文說明	備註
1	FEE_YM	Char	6	費用年月	<常用欄位>
2	APPL_TYPE	Char	1	申報類別	<常用欄位>
3	APPL_DATE	Char	8	申報日期	<常用欄位>
4	CASE_TYPE	Char	2	案件分類	<常用欄位>
5	SEQ_NO	Char	8	流水號	<常用欄位>
6	CURE_ITEM_NO1	Char	2	特定治療項目代號(一)	
7	CURE_ITEM_NO2	Char	2	特定治療項目代號(二)	
8	CURE_ITEM_NO3	Char	2	特定治療項目代號(三)	
9	CURE_ITEM_NO4	Char	2	特定治療項目代號(四)	
10	FUNC_TYPE	Char	2	就醫科別	<常用欄位>
11	FUNC_DATE	Char	8	就醫日期	<常用欄位>
12	TREAT_END_DATE	Char	8	治療結束日期	
13	BIRTH_YM	Char	6	出生日期	提供出生年月 YYYYMM <常用欄位>
14	CARD_SEQ_NO	Char	4	健保卡就醫序號	
15	GAVE_KIND	Char	1	給付類別	
16	PART_NO	Char	3	部分負擔代號	<常用欄位>
17	ICD9CM_1	Char	5	國際疾病分類號一	<常用欄位>
18	ICD9CM_2	Char	5	國際疾病分類號二	<常用欄位>
19	ICD9CM_3	Char	5	國際疾病分類號三	<常用欄位>

門診病史



門診處方及治療明細檔 (適用 93 年以後資料)

CD

英文檔案名稱： CD		長度 301					
中文檔案名稱： 門診處方及治療明細檔							
序號	英文欄位	中文欄位	資料型態	長度	起始位置	迄末位置	資料描述
22	ICD_OP_CODE	主手術代碼	C	4	147	150	ICD-9-CM之主手術代號；預防保健、慢性病連續處方調劑及中醫本欄免填
23	DRUG_DAY	給藥日份	N	2	151	152	右靠不足補0，本次給藥日份最高之天數(註一)：除單次給藥為「非口服藥」外，餘請填「口服藥」之最高給藥日份(89.06)；慢性病連續處方調劑，請依連續處方箋上所列填寫；預防保健則本欄為非必要欄位
24	MED_TYPE	處方調劑方式	C	1	153	153	請參考代碼說明之「處方調劑方式」
25	PRSN_ID	醫師代號	C	32	154	185	已轉碼。醫師或原處方醫師之國民身分證統一編號或外籍居留證號。左靠不足補空白。
26	PHAR_ID	藥師代號	C	32	186	217	已轉碼。調劑藥師之國民身分證統一編號，外籍居留證號，左靠不足補空白。診所聘有藥師本欄為必要欄位，否則留空白。
27	DRUG_AMT	用藥明細金額小計	N	8	218	225	右靠不足補0，此項費用為OO檔之醫令類別為1之用藥明細金額(序號13)之小計，如為交付調劑之藥品費用，不得併入用藥明細金額小計內計算；西醫簡表及中醫請依給藥日份計算用藥申報金額
28	TREAT_AMT	診療明細金額小計	N	8	226	233	右靠不足補0，此項費用為OO檔之除用藥之外(醫令類別為2或3)，各項診療及特殊材料明細金額(序號13)之小計；慢性病連續處方調劑本欄免填
29	TREAT_CODE	診察費項目代號	C	12	234	245	填寫全民健保醫療費用支付標準碼。預防保健、洗腎、慢性病連續處方調劑本欄免填，補空白
30	DIAG_AMT	診察費	N	8	246	253	診察費金額，右靠不足補0。預防保健、洗腎、慢性病連續處方調劑本欄免填，請補0

門診手術



							計。
--	--	--	--	--	--	--	----

住院醫療費用清單明細檔 (適用 93~95 年資料)

DD

英文檔案名稱： DD		長度 494					
中文檔案名稱： 住院醫療費用清單明細檔							
序號	英文欄位	中文欄位	資料型態	長度	起始位置	迄末位置	資料描述
18	S_BED_DAY	慢性病床天數	N	3	140	142	病人住慢性病床之總計天數，以入院之日起計，出院之日不計。
19	PRSN_ID	主治醫師代碼	C	32	143	174	被保險人入院後病房主治醫師之身分證號或外籍居留證號，如住院中有前後兩位主治醫師時，以出院時之主治醫師填寫。
20	DRG_CODE	DRG參考碼	C	5	175	179	詳情請自健保局之醫事機構專區網頁查詢 (http://www.nhi.gov.tw/) 若案件分類為「2」或「A2」者，本欄不得空白。
21	EXT_CODE_1	外因分類一	C	5	180	184	汽機車交通事故為「Y」時，本欄不得空白 ICD-9-CM 之E碼
22	EXT_CODE_2	外因分類二	C	5	185	189	ICD-9-CM 之E碼
23	TRAN_CODE	轉歸代碼	C	1	190	190	1：治療出院 2：繼續住院 3：改門診治療 4：死亡 5：自動出院 6：轉院 7：身份變更 8：潛逃 9：自殺 0：其他 A：病危自動出院
24	ICD9CM_CODE	主診斷代碼	C	5	191	195	ICD-9-CM 之代碼 以1992年版為準 (94年12月止)，自95年1月起(費用年月)以2001年版為準。
25	ICD9CM_CODE_1	次診斷代碼一	C	5	196	200	ICD-9-CM 之代碼
26	ICD9CM_CODE_2	次診斷代碼二	C	5	201	205	ICD-9-CM 之代碼
27	ICD9CM_CODE_3	次診斷代碼三	C	5	206	210	ICD-9-CM 之代碼
28	ICD9CM_CODE_4	次診斷代碼四	C	5	211	215	ICD-9-CM 之代碼
29	ICD_OP_CODE	主手術 (處置)	C	4	216	219	ICD-9-CM 之手術代碼
30	ICD_OP_CODE_1	主手術 (處置) 一	C	4	220	223	ICD-9-CM 之手術代碼
31	ICD_OP_CODE_2	主手術 (處置) 二	C	4	224	227	ICD-9-CM 之手術代碼
32	ICD_OP_CODE_3	主手術 (處置) 三	C	4	228	231	ICD-9-CM 之手術代碼
33	ICD_OP_CODE_4	主手術 (處置) 四	C	4	232	235	ICD-9-CM 之手術代碼

住院病史

英文檔案名稱： 00		長度 120					
中文檔案名稱： 門診處方醫令明細檔							
序號	英文欄位	中文欄位	資料型態	長度	起始位置	迄末位置	資料描述
1	FEE_YM	費用年月	C	6	1	6	YYYYMM
2	APPL_TYPE	申報類別	C	1	7	7	1：送核 2：補報
3	HOSP_ID	醫事機構代號	C	34	8	41	已轉碼，後 2 碼為其權屬別
4	APPL_DATE	申報日期	C	8	42	49	YYYYMMDD (西元年)
5	CASE_TYPE	案件分類	C	2	50	51	請參考代碼說明之「案件分類」
6	SEQ_NO	流水號	N	6	52	57	右靠不足補0，依案件分類之類別分別連續編號，本欄與門診處方及治療同筆資料之流水號一致
7	ORDER_TYPE	醫令類別	C	1	58	58	0：診察費 1：用藥明細 2：診療明細 3：特殊材料 4：不得另計價之藥品或診療項目，請參考代碼說明之「特定項目治療代號」。(5：EPO注射 6：HCT檢驗) 7：代檢(95/7刪除) 8：器官捐贈
8	DRUG_NO	藥品(項目)代號	C	12	59	70	填寫全民健保用藥品代碼或醫療費用支付標準碼或特殊材料代碼 詳情請自健保局之醫事機構專區網頁查詢 (http://www.nhi.gov.tw/)。

門診用藥

- 投保服務
- 網路申辦及查詢
- 保險費計算與繳納
- 欠費催繳異議
- 申辦健保卡
- 經濟弱勢協助措施
- 健保醫療服務
- 常見就醫自費項目
- 自墊醫療費用核退
- 就醫申訴服務
- 常見問答
- 投保單位**
- 投保單位成立與異動
- 網路申辦及查詢
- 投保異動與申辦
- 保險費計算與繳納
- 常見問答
- 醫事機構**
- 特約由請函變更

藥材專區

藥材專區

- 藥品
- 特殊材料
- 網路查詢

健保用藥品項查詢

資料更新日期：1020826 共21858307人

查詢條件

藥品名稱(中文)：	請輸入藥品名稱(中文)
藥品名稱(英文)：	請輸入藥品名稱(英文)
健保代碼：	請輸入健保代碼
劑型：	請輸入劑型
藥品分類：	請輸入藥品分類
品質條件分類：	請輸入品質條件分類
成分名稱：	請輸入成分名稱
成分含量(數量)：	請輸入成分含量(數量)
成分含量(單位)：	請輸入成分含量(單位)
規格量(數量)：	請輸入規格量(數量)
規格量(單位)：	請輸入規格量(單位)
藥商：	請輸入藥商
單複方：	請輸入單複方
生效日期：	年 月
不適合吞服口服錠劑、膠囊劑之製劑：	請輸入製劑
異動：	請輸入異動類別

查詢顯示方式

- 只顯示最後一筆資料
- 顯示歷次異動的詳細資料

每次顯示 10 筆
另開視窗

[開始查詢](#) [重新選擇](#)

*注意：
「成分含量(數量) 規格量(數量)」限輸入數字
「成分含量(單位)」、「規格量(單位)」限輸入英文

查詢使用說明

使用說明：
字串：當您不確定所要查詢的字串全名時，可輸入部分字串，即可查到符合此部分字串的所有資料。例如：膠囊可查到"膠囊劑"的相關資料。
編號：若您所要查詢的條件屬於含有英文字母的編號，當您不確定查詢條件時可輸入部分條件*或*部分條件。例如：A*或*5的查詢條件即可查到A107090015的相關資料。

比較

- 藥品代碼**10**碼

Crestor: B024129100

- 特材代碼**12**碼

**"MEDTRONIC"TALENT AAA STENT
GRAFT W THE XCELERANT HYDRO
DELIVERY SYS 腹主動脈支架系統**

CBC03AAA01M4

- 如果只想要挑出用藥記錄

if Length (Drug_No) = 10;

共有8筆 (友善列印另開視窗)

異動	藥品代碼	藥品名稱(英文)	藥品名稱(中文)	成分	成分含量	規格量	單複方	價格	起迄	藥
	B022671100	AMARYL 2.0 TABLETS	瑪爾胰 2 · 0 錠	GLIMEPIRIDE	2 MG		單方	10.5	089.07.01 095.10.31	賽份
	B022671100	AMARYL 2.0 TABLETS	瑪爾胰 2 · 0 錠	GLIMEPIRIDE	2 MG		單方	10.1	095.11.01 096.08.31	賽份
	B022671100	AMARYL 2.0 TABLETS	瑪爾胰 2 · 0 錠	GLIMEPIRIDE	2 MG		單方	9.8	096.09.01 098.09.30	賽份
	B022671100	AMARYL 2.0 TABLETS	瑪爾胰 2 · 0 錠	GLIMEPIRIDE	2 MG		單方	7.8	098.10.01 100.11.30	賽份

[B024876100](#)

[B024876100](#)

[B024876100](#)

AMARYL M TABLETS 2/500MG	美爾胰錠 2/500 公 絲	GLIMEPIRIDE 2 MG METFORMIN 500 MG HCL	複方 9.8	098.07.01 098.09.30	賽 份
AMARYL M TABLETS 2/500MG	美爾胰錠 2/500 公 絲	GLIMEPIRIDE 2 MG METFORMIN 500 MG HCL	複方 7.8	098.10.01 100.11.30	賽 份
AMARYL M TABLETS 2/500MG	美爾胰錠 2/500 公 絲	GLIMEPIRIDE 2 MG METFORMIN 500 MG HCL	複方 5.3	100.12.01 迄今	賽 份

Extraction According to Prescription

****extract subjects with the prescription of
Amaryl (either single pill or combo);**

data work.Amaryl;

set NHI.go2000;

IF substr (DRUG_NO, **1**, **10**) = 'B022671100' or
substr (DRUG_NO, **1**, **10**) = 'B024876100' ;

run;

MACRO & ARRAY

参考SAS PRACTICE CODE

```
/*MACRO*/  
%macro Sas_test;  
%do i=2006 %to 2007;  
data Study.cd&i; set cd&i; run;  
data Study.oo&i; set cd&i; run;  
data Study.cdoo&i; set cdoo&i;  
run;  
%end;  
%mend;  
%Sas_test
```

```
/*另一種方法*/  
%macro test (i);  
data Study.cd&i; set cd&i; run;  
data Study.oo&i; set cd&i; run;  
data Study.cdoo&i; set cdoo&i;  
run;  
%end;  
%mend;  
%test(2006) %test(2007)
```

```
/*ARRAY*/
```

```
*define diseases*;
```

```
▫ data hw1; set study.cd2006(keep=id icd9cm_1-icd9cm_3);
```

```
array icd $ icd9cm_1-icd9cm_3;
```

```
do over icd;
```

```
  if substr(icd,1,3) in ('401','402','403','404','405') then htn=1;*hypertension*;
```

```
  if substr(icd,1,3) in ('433','434','435','436','437','438') then stroke=1;*stroke*;
```

```
  if substr(icd,1,3) in ('410','411','412','413','414') then ihd=1;*ischemic heart disease*;
```

```
  if substr(icd,1,3) in ('584','585') then rf=1;*renal failure*;
```

```
  if substr(icd,1,3) in ('714') then ra=1;*rheumatoid arthritis*;
```

```
  if substr(icd,1,3) in ('571') then liver=1;*chronic liver disease*;
```

```
  if substr(icd,1,4) in ('2900','2901','2902','2903','3310') then dementia=1;*dementia*;
```

```
  if substr(icd,1,4) in ('2962','2963','3004') or substr(icd,1,3) in ('311') then dep=1;*depression*;
```

```
end;
```

```
if htn=1 or stroke=1 or ihd=1 or rf=1 or ra=1 or liver=1 or dementia=1 or dep=1;
```

```
run;
```

感謝您的聆聽！



生統小組：統計方法教育訓練

**for home computer;

libname Study "D:\助理研究員\心血管大數據中心\研究助理授課\2023-教學資料\20230816-健保資料-SAS 1\健保資料庫sas使用

/*2023/8/16: Merge data*/

**cd2006 and oo2006;

DATA cd2006;

SET Study.H_nhi_opdte95;

KEEP FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO
ID FUNC_DATE
ICD9CM_1 - ICD9CM_3;

RUN;

DATA oo2006;

SET Study.H_nhi_opdto95;

KEEP FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO
DRUG_NO;

RUN;

PROC SORT DATA=cd2006;

BY FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO;

RUN;

PROC SORT DATA=oo2006;

BY FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO;

RUN;

DATA cdoo2006;

MERGE cd2006 oo2006 ;

BY FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO;

RUN;

**cd2007 and oo2007;

DATA cd2007;

SET Study.H_nhi_opdte96;

KEEP FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO
ID FUNC_DATE
ICD9CM_1 - ICD9CM_3;

RUN;

DATA oo2007;

SET Study.H_nhi_opdto96;

KEEP FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO
DRUG_NO;

RUN;

PROC SORT DATA=cd2007;

BY FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO;

RUN;

PROC SORT DATA=oo2007;

BY FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO;

RUN;

DATA cdoo2007;

MERGE cd2007 oo2007 ;

BY FEE_YM APPL_TYPE HOSP_ID
APPL_DATE CASE_TYPE SEQ_NO;

RUN;

```

/*stacking cdoo2006 and cdoo2007*/
data myset ;
set cdoo2006 cdoo2007 ;
run;

proc sort data = myset;
by ID FUNC_DATE;
run;

/*2023/8/23 Medical History Identification*/
/* Age & Sex */
data study.id2006 ; set study.h_nhi_enrol95 ;
diagdate = mdy ( 12 , 31 , 2022 ) ;

birthyear = substr (ID_BIRTH_YM,1,4) * 1 ;
birthmonth = substr (ID_BIRTH_YM,5,2) * 1 ;
birthday = 15 ;
birthdate = mdy ( birthmonth , birthday , birthyear ) ;
age = (diagdate - birthdate)/365.25;

agegp=.;
if age <= 65 then agegp=0;
else agegp=1;

sex=.;
if id_s=1 then sex=1; else if id_s=9 then sex=.;else sex=0;
run;

Proc sort data=study.id2006; by id id_in_date; run;
data id2006; set study.id2006; by id ; if first.id; run;

/*Extracting subjects with any diagnosis of DM*/
data DM;
set myset;
if substr ( ICD9CM_1, 1, 3) = '250' or
substr ( ICD9CM_2, 1, 3) = '250' or
substr ( ICD9CM_3, 1, 3) = '250' ;
dm=1;
run;

/*number of subjects with any diagnosis of DM, N = 46*/
proc sort data = DM nodupkey out = DM_ID;
by ID;
run;

/*identify subjects with DM*/
data DM_ID_1;
set DM_ID;
DM = 1;
keep ID DM;
run;

data myset_2;
merge DM_ID_1 myset ;
by ID;
run;

data myset_3;
set myset_2;
if DM =1;
run;

/*the cumulative number of drugs prescribed within 2 years */
proc sort data = myset_3 nodupkey out = DM_drug_1;
by ID Drug_No;
run;

```

```
data DM_drug_2;
set DM_drug_1;
if Length (Drug_No) = 10;
run;
```

```
data DM_count;
set DM_drug_2;
by ID;
if first.ID then count=0;
count+1;
if last.ID;
keep ID count;
run; *算次數;
```

```
proc sort data = DM_count;
by count;
run;
```

```
/*export data*/
```

```
PROC EXPORT DATA= WORK.DM_COUNT
```

```
OUTFILE= "D:\助理研究員\心血管大數據中心\研究助理授課\2023-教學資料\20230815-健保資料-SAS 1\健保資料庫sas
```

```
DBMS=EXCEL5 REPLACE;
```

```
RUN;
```

```
/*MACRO*/
```

```
%macro Sas_test;
```

```
%do i=2006 %to 2007;
```

```
data Study.cd&i; set cd&i; run;
```

```
data Study.oo&i; set cd&i; run;
```

```
data Study.cdoo&i; set cdoo&i; run;
```

```
%end;
```

```
%mend;
```

```
%Sas_test
```

```
/*另一種方法*/
```

```
%macro test (i);
```

```
data Study.cd&i; set cd&i; run;
```

```
data Study.oo&i; set cd&i; run;
```

```
data Study.cdoo&i; set cdoo&i; run;
```

```
%end;
```

```
%mend;
```

```
%test(2006) %test(2007)
```

```
/*ARRAY*/
```

```
*define diseases*;
```

```
data hw1; set study.cd2006(keep=id icd9cm_1-icd9cm_3);
```

```
array icd $ icd9cm_1-icd9cm_3;
```

```
do over icd;
```

```
if substr(icd,1,3) in ('401','402','403','404','405') then htn=1;*hypertension*;
```

```
if substr(icd,1,3) in ('433','434','435','436','437', '438') then stroke=1;*stroke*;
```

```
if substr(icd,1,3) in ('410','411','412','413','414') then ihd=1;*ischemic heart disease*;
```

```
if substr(icd,1,3) in ('584','585') then rf=1;*renal failure*;
```

```
if substr(icd,1,3) in ('714') then ra=1; *rheumatoid arthritis*;
```

```
if substr(icd,1,3) in ('571') then liver=1; *chronic liver disease*;
```

```
if substr(icd,1,4) in ('2900','2901','2902','2903','3310') then dementia=1; *dementia*;
```

```
if substr(icd,1,4) in ('2962','2963','3004') or substr(icd,1,3) in ('311') then dep=1;*depression*;
```

```
end;
```

```
if htn=1 or stroke=1 or ihd=1 or rf=1 or ra=1 or liver=1 or dementia=1 or dep=1;
```

```
run;
```

```
*計算comobility總次數，並化整為同一人;
```

```
proc univariate data=hw1 noprint;
```

```
var htn stroke ihd rf ra liver dementia dep;
```

```
class id;
```

```
output out=hw2 sum= htn1 stroke1 ihd1 rf1 ra1 liver1 dementia1 dep1;
run;

*select study subjects: dm patients, n=31*;
data dm; set study.cd2006;
if substr(icd9cm_1,1,3) in ('250') or substr(icd9cm_2,1,3) in ('250') or substr(icd9cm_3,1,3) in ('250');
keep id;
proc sort nodupkey; by id;
run;

*add dz information to dm cohort*;
proc sort data=hw2; by id;
data hw3;
merge dm (in=a) hw2;
by id;
if a=1;
run;

*create new indicator variables of diseases*;
data study.hw4; set hw3;
array dz1 htn1 stroke1 ihd1 rf1 ra1 liver1 dementia1 dep1;
array dz htn stroke ihd rf ra liver dementia dep;
do over dz1;
  if dz1>=1 then dz=1; else dz=0;
end;
sumdz=htn+stroke+ihd+rf+ra+liver+dementia+dep;
keep id htn stroke ihd rf ra liver dementia dep sumdz;
run;

data study.hw5; set study.hw4;
if sumdz>=2;
proc print;
run;
```